

A DATA-ASSISTED APPROACH TO SUPPORTING
INSTRUCTIONAL INTERVENTIONS IN TECHNOLOGY
ENHANCED LEARNING ENVIRONMENTS

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of PhD
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Christopher Brooks

©Christopher Brooks, December 2012. All rights reserved.

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5C9

ABSTRACT

The design of intelligent learning environments requires significant up-front resources and expertise. These environments generally maintain complex and comprehensive knowledge bases describing pedagogical approaches, learner traits, and content models. This has limited the influence of these technologies in higher education, which instead largely uses learning content management systems in order to deliver non-classroom instruction to learners.

This dissertation puts forth a *data-assisted approach* to embedding intelligence within learning environments. In this approach, instructional experts are provided with summaries of the activities of learners who interact with technology enhanced learning tools. These experts, which may include instructors, instructional designers, educational technologists, and others, use this data to gain insight into the activities of their learners. These insights lead experts to form instructional interventions which can be used to enhance the learning experience. The novel aspect of this approach is that the actions of the intelligent learning environment are now not just those of the learners and software constructs, but also those of the educational experts who may be supporting the learning process.

The kinds of insights and interventions that come from application of the data-assisted approach vary with the domain being taught, the epistemology and pedagogical techniques being employed, and the particulars of the cohort being instructed. In this dissertation, three investigations using the data-assisted approach are described. The first of these demonstrates the effects of making available to instructors novel sociogram-based visualizations of online asynchronous discourse. By making instructors aware of the discussion habits of both themselves and learners, the instructors are better able to measure the effect of their teaching practice. This enables them to change their activities in response to the social networks that form between their learners, allowing them to react to deficiencies in the learning environment. Through these visualizations it is demonstrated that instructors can effectively change their pedagogy based on seeing data of their students' interactions.

The second investigation described in this dissertation is the application of unsupervised machine learning to the viewing habits of learners using lecture capture facilities. By clustering learners into groups based on behaviour and correlating groups with academic outcome, a model of positive learning activity can be described. This is particularly useful for instructional designers who are evaluating the role of learning technologies in programs as it contextualizes how technologies enable success in learners. Through this investigation it is demonstrated that the viewership data of learners can be used to assist designers in building higher level models of learning that can be used for evaluating the use of specific tools in blended learning situations.

Finally, the results of applying supervised machine learning to the indexing of lecture video is described. Usage data collected from software is increasingly being used by software engineers to make technologies that are more customizable and adaptable. In this dissertation, it is demonstrated that supervised machine

learning can provide human-like indexing of lecture videos that is more accurate than current techniques. Further, these indices can be customized for groups of learners, increasing the level of personalization in the learning environment. This investigation demonstrates that the *data-assisted approach* can also be used by application developers who are building software features for personalization into intelligent learning environments.

Through this work, it is shown that a data-assisted approach to supporting instructional interventions in technology enhanced learning environments is both possible and can positively impact the teaching and learning process. By making available to instructional experts the online activities of learners, experts can better understand and react to patterns of use that develop, making for a more effective and personalized learning environment. This approach differs from traditional methods of building intelligent learning environments, which apply learning theories *a priori* to instructional design, and do not leverage the *in situ* data collected about learners.

ACKNOWLEDGEMENTS

I would not have been able to complete this work without the dedication and support of others. First and foremost, I would like to thank my supervisors, Dr. Jim Greer and Dr. Carl Gutwin, who provided morale, financial, and intellectual support for my academic endeavours. I was once given the advice that you only get one chance at a doctorate degree, so make sure it is the doctorate degree you want. Over the years, Jim and Carl have provided to me liberal amounts advice but always left up to me the choice of which direction I would take my research. Thank you both for letting me make this doctorate into the one I wanted it to be.

I would also like to thank Dr. Gordon McCalla who, beyond being an outstanding committee member, acted in many ways as a third supervisor for me. Much of my work comes from insights gleaned while talking with Gord, and his ability to holistically consider technology enhanced learning research has had a profound impact on how I approach my work. I look forward to our continued discussions and collaborations in this and other areas.

As committee members, Dr. Mike Horsch and Dr. Jay Wilson provided comments and challenges which have made my work stronger and more well rounded. Dr. Judy Kay served as external examiner and provided thoughtful remarks which have helped promote in me further insight on the area. I'm thankful to each of these faculty members for their involvement.

Through my time at the University of Saskatchewan I have had the opportunity to supervise, collaborate with, and work beside many excellent students. Without these collaborations, I would not have been able to gain deep insight into my research nor make the impact on students that I have. I want to specifically thank Lori Kettel and Ryan Silk for their work on the iHelp Discussion system that was used for the basis of the visualizations in Chapter 5, Graham Erikson and Carrie Demmans Epp for their extensive help in analysis of the behaviours of learners using lecture capture, much of which is described in Chapter 6, Kristofor Amundson, and G. Scott Johnston for their work in helping to bring a proof of concept to life in Chapter 7. I would also like to thank Greg Logan, Stephen Damm, Kristofor Amundson, and Adam McKenzie for their helping building the Recollect and Opencast systems. These systems have been pivotal to my research in the last five years, and I would not have been able to provide such broad results without their help.

I'd like to offer special thanks to my colleague Craig Thompson who has helped sustain my enthusiasm for all of these subjects (and more) over the last five years with his willingness to dive into the details of my research. My work has attained a new of rigour based on our discussions, and I look forward to pursuing more joint research with Craig in the future.

Lastly, I would like to thank the faculty and students of the University of Saskatchewan who have enabled my work through their willingness to use research software in their day-to-day educational practice. This includes those who used the iHelp suite of tools (more than 8,500 students), the Recollect lecture capture system (more than 4,300 students), and the Opencast Matterhorn system (more than 3,500 students). Without their acceptance of trying new methods for teaching and learning, my work would not have been possible.

for my grandfathers, Carl Hansen (1923–) and Arthur Brooks (1917–1962)
who, in the building of this university both physically and academically,
provided for me world class facilities and a strong sense of place

CONTENTS

| | |
|--|-----------|
| Permission to Use | i |
| Abstract | ii |
| Acknowledgements | iv |
| Contents | vi |
| List of Tables | viii |
| List of Figures | x |
| 1 Introduction | 1 |
| 1.1 Problem | 2 |
| 1.2 Solution | 3 |
| 1.3 Contributions | 5 |
| 1.4 Dissertation Outline | 6 |
| 2 Background | 7 |
| 2.1 Traditional Learner Modelling | 7 |
| 2.1.1 Overlay Models | 7 |
| 2.1.2 Stereotypes | 8 |
| 2.1.3 Knowledge Tracing | 9 |
| 2.1.4 Constraint-Based Models | 10 |
| 2.1.5 Episodic Models | 12 |
| 2.1.6 Active Modelling | 12 |
| 2.1.7 Scrutable and Open Modelling | 13 |
| 2.1.8 Group Modelling | 15 |
| 2.1.9 Conclusions | 18 |
| 2.2 Data-driven Learner Modelling | 19 |
| 2.2.1 Educational Data Mining and Learning Analytics | 20 |
| 2.2.2 Motivations and Techniques | 21 |
| 2.2.3 Differentiating between EDM and LAK | 23 |
| 2.2.4 Ecological Approach | 25 |
| 2.2.5 Conclusions | 26 |
| 2.3 Instructional Remediation Through Adaptation | 27 |
| 2.3.1 Content Adaptation & Sequencing | 27 |
| 2.3.2 Expertise Recommendation | 29 |
| 2.3.3 Adaptive Testing | 29 |
| 2.3.4 Assessment for Learning | 30 |
| 2.3.5 Conclusions | 31 |
| 3 The Data-Assisted Approach to Supporting Instructional Interventions in Technology Enhanced Learning Environments | 32 |
| 3.1 What is Meant by <i>Data</i> in the Data Assisted Approach? | 33 |
| 3.2 What Does it Mean to <i>Assist</i> in the Data Assisted Approach? | 34 |
| 3.3 Contrasting with the Ecological Approach | 35 |
| 4 Realizing the Data-Assisted Approach | 37 |
| 4.1 Scenario One: Visualizing Community Interactions | 38 |

| | | |
|----------|---|------------|
| 4.2 | Scenario Two: Measuring Educational Impact | 42 |
| 4.3 | Scenario Three: Adapting Learning Environments to Tasks | 45 |
| 4.4 | Conclusion | 49 |
| 5 | Visualizing Community Interactions | 50 |
| 5.1 | Augmenting an Asynchronous Discussion Forum with Visualizations | 51 |
| 5.1.1 | iHelp Discussions | 51 |
| 5.1.2 | Visualization of Learner Activities | 52 |
| 5.2 | Effect of Visualization of Traces on Pedagogical Practice | 54 |
| 5.2.1 | Case study: Online Small Cohort Course | 54 |
| 5.2.2 | Case study: Face-to-Face Large Cohort Course | 58 |
| 5.2.3 | Case study: Comparison of Visualizations based on Granularity | 60 |
| 5.2.4 | Case study: Comparing Pedagogical Approaches | 61 |
| 5.3 | Conclusions | 62 |
| 6 | Measuring Educational Impact | 64 |
| 6.1 | The Recollect System | 64 |
| 6.2 | Formal Assessment | 65 |
| 6.3 | Actionability of Model | 72 |
| 6.4 | Learner Goals | 76 |
| 6.5 | Conclusion | 77 |
| 7 | Adapting Learning Environments to Tasks | 80 |
| 7.1 | Navigation in Recollect | 81 |
| 7.2 | Indexing Algorithms | 83 |
| 7.3 | Grouping Learners by Navigation Usage | 83 |
| 7.3.1 | Comparing Users to Traditional Algorithms | 83 |
| 7.3.2 | Comparing Users to One Another | 85 |
| 7.3.3 | Conclusions | 86 |
| 7.4 | Adapting Navigation based using Supervised Machine Learning | 87 |
| 7.5 | Conclusion | 90 |
| 8 | Discussion and Conclusions | 92 |
| 8.1 | Discussion of Findings | 92 |
| 8.1.1 | Forming Insight | 93 |
| 8.1.2 | Creating Instructional Interventions | 94 |
| 8.1.3 | Connecting Data to Insights and Insights to Interventions | 95 |
| 8.1.4 | Primary Contribution | 96 |
| 8.2 | Secondary Contributions | 96 |
| 8.3 | Future Work | 97 |
| 8.3.1 | The Data-Assisted Approach | 97 |
| 8.3.2 | Visualizing Community Interactions | 98 |
| 8.3.3 | Measuring Educational Impact | 98 |
| 8.3.4 | Adapting Learning Environments to Tasks | 99 |
| 8.4 | Conclusion | 99 |
| | References | 100 |
| | A Student Survey Instrument | 107 |
| | B Behavioural Ethics Certificates | 113 |

LIST OF TABLES

| | | |
|-----|---|----|
| 2.1 | Data mining techniques commonly found in EDM research, from [9] as cited in [10]. | 22 |
| 6.1 | K-means clustering of Chemistry 2010 Recollect data ($k = 5$). Clusters range in size from 8 to 103 participants, with mean error between 5% and 25%. Attributes used to form clusters were whether learners watched (w) or didn't watch (d) at least five minutes of lecture video. Week seven was omitted from the model due to the midterm break, and weeks 14 through 16 denote weeks between the end of lectures and the final exam. In this table, error refers to the difference between a particular centroid and the mean value of learners who best fit that cluster (e.g., a ratio of disagreement). | 67 |
| 6.2 | Comparison of mean disagreement between an ideal clustering and a pedagogical model within the Chemistry 2010 and 2011 data sets. The values can range between 0 (perfect fit) and 1 (complete disagreement), and are given by the equation $\frac{\sum_{x \in s} [\sum_{i=0}^w (x_i - cluster(x)_i)]}{n * w}$, where n is the number of students, w is the number of weeks data exists for, s is the set of students being considered, and $cluster(x)$ is the vector of cluster centroids that the student best fits. Ideal clusters were restricted by $k = 5$ | 69 |
| 6.3 | Midterm, final examination, and overall grade averages and standard deviations broken down by cluster in percentages for the Chemistry 2011 course. Students who did not use the system ($n = 216$) had midterm examination marks of $\bar{x} = 65.19, \sigma = 16.21$ and final examination marks of $\bar{x} = 61.75, \sigma = 21.29$ and overall examination marks of $\bar{x} = 68.59, \sigma = 15.34$. Group ANOVA provided support for significance for midterm marks ($f = 3.05, p = 0.02$) and overall marks ($f = 2.435, p = 0.0472$) but not final marks ($f = 1.71, p = 0.15$). | 69 |
| 6.4 | Tukey HSD confidence values between pairs of clusters using midterm examination marks. Strong support for differentiating the performance of <i>Minimal Activity</i> and <i>Deferred</i> groups of learners from the <i>High Activity</i> learners ($p = 0.021$ and $p = 0.057$ respectively). Weaker support exists when differentiating between <i>High Activity</i> learners and the groups of <i>Disillusioned</i> and <i>Just-In-Time</i> learners ($p = 0.160$ and $p = 0.240$ respectively). | 70 |
| 6.5 | Tukey HSD confidence values between pairs of clusters using final examination marks. Weak support for differentiating the performance of all groups of learners from the <i>High Activity</i> learners, with p values ranging between 0.103 and 0.190. | 70 |
| 6.6 | Tukey HSD confidence values between pairs of clusters using overall marks, which included examination marks, laboratory marks, and assignment marks. Strong support for differentiating the performance of <i>Disillusioned</i> and <i>Minimal Activity</i> learners from <i>High Activity</i> learners ($p = 0.085$ and $p = 0.031$ respectively). Weak support for differentiating the performance of <i>Just-In-Time</i> and <i>Deferred</i> learners with <i>High Activity</i> learners ($p = 0.161$ and $p = 0.157$ respectively). | 70 |
| 6.7 | Comparison of mean disagreement between an ideal clustering and a pedagogical model within the Biomolecules 2011 data set. The values can range between 0 (perfect fit) and 1 (complete disagreement), and are given by the equation described in Table 6.2. Ideal clusters were restricted by $k = 5$ | 71 |
| 6.8 | Overall grade averages and standard deviations broken down by cluster in percentages for the BMSC-200 2011 course. Group ANOVA provided does not show strong support for significance for overall marks ($f = 1.286, p = 0.2773$). | 71 |
| 6.9 | Tukey HSD confidence values between pairs of clusters using overall course marks for Biomolecules 2011. | 72 |

| | | |
|------|--|----|
| 6.10 | Actionability table for Chemistry 2011 data. At each week of instruction a learner could have either watched online video or not, and is classified into the appropriate cluster based on the model derived from Table 6.1. For each week of instruction, all learners can be classified as to which cluster they best fit based on their data thus far. For example, in week one a total of 70 learners best fit the <i>high activity</i> cluster (<i>true positive</i> + <i>false positive</i>), and 11 of these learners will end up in this cluster at the end of the term based on their interactions throughout the term (<i>true positive</i>). Fifty-nine of these learners will end up in another cluster by the end of the term (<i>false positive</i>). Further, a total of 262 learners do not fit this cluster based on their interactions thus far (<i>true negative</i> + <i>false negative</i>), with the vast majority (259) not fitting this cluster by the end of the term (<i>true negative</i>). Only three learners (<i>false negative</i>) do not fit this cluster after one week of data, but will by the end of term. Since it is impossible to differentiate at week one between the <i>high activity</i> and <i>disillusioned</i> clusters, they both have identical values for <i>borderline</i> | 73 |
| 6.11 | Student behaviour clusters based on the number of minutes watched and unique videos watched with $k = 2$. Cluster labels, <i>keen</i> and <i>less keen</i> were added by the author as descriptive elements only, and are arbitrary. The only strong correlation with questionnaire data existed for questions six, nine, 22, and 24. See appendix A for question details. | 78 |
| 7.1 | Inter-rater reliability measures using Fleiss' κ [53] for significant events in four lecture videos. κ is calculated with $n = 6$, and κ' with ($n = 5$) by removing the one outlying rater (semantic annotator). Higher agreement between raters is shown for Videos A through C, while lower agreement is shown for the handwritten Video D. Adding in Dickson's algorithm (adapted from [43]) lowers overall group agreement in all cases. | 84 |
| 7.2 | The series of visual attributes used to form supervised model. Each attribute was calculated for each image. | 88 |
| 7.3 | Comparison over one term of instruction from two classes, Course A and Course B, as well as the union of these datasets. Each video was encoded by a single rater, and J48 decision trees (κ_{J48}) provided with Weka using ten-fold testing to minimize chances of Type III errors. The result of applying Dickson's algorithm to this data set is provided for performance comparison (κ_D). | 89 |
| 7.4 | Group κ between all raters and predictors. <i>upper</i> and <i>lower</i> contextualize the range κ can take when a new algorithm or rater is included. See §7.2 for descriptions of the <i>Time</i> , <i>Opencast</i> , and <i>Dickson</i> algorithms. Six supervised machine learning methods were trained from data T_1 through T_6 , where T_1 indicates that in order for an index to be chosen only one rater must have indicated so, T_2 indicates that for an index to be chosen at least two raters must have indicated so, and so on. | 90 |
| 8.1 | Outline of data, data-processing techniques, and the insight and instructional interventions they might lead to when using the data-assisted approach. | 96 |

LIST OF FIGURES

| | | |
|-----|---|----|
| 4.1 | A mock up of a discussion forum information visualization system. The sociogram on the left is made up of three concentric rings of users: the outermost ring holds those users who have not used the system, the middle ring holds those users who have only read forums posts but not written any (<i>lurkers</i>), while the inner ring contains users who have written discussion forum posts. The ring containing lurkers puts users closer to the centre of the sociogram if they have read many forum postings, and puts users closer to the edge of the sociogram if they have read few forums posts. The inner ring connects users based on whether their post is a reply or a new thread. Instructional experts are shown using darker colours. Details about a particular user that might be select (e.g., the dark (red) node that represents Katheryn in the middle) are shown to the right of the sociogram. | 39 |
| 4.2 | A mock up of a discussion forum system similar to the one in figure 4.1. In this system, the circles representing users are sized depending upon how often a users forum postings are read. Large nodes represent users who are read by many other users, while small nodes represent users who are read by few users. In this example, only the one instructional expert node is large, suggesting that only that users forum postings are of interest to the community. Nodes that have a dotted outline represent users who have not logged into the system in the last two weeks. Customization of the sociogram is done via rules which are shown at the top right. In this case, there are three rules defined which control node colour, size, and line outline. Hyperlinks allow the instructional expert to modify rules, while new rules can be defined using the form controls at the bottom right portion of the application. | 41 |
| 4.3 | A mock up of a clustering-based learning analytics tool which allows instructional experts to explore the interactions learners have had within a learning environment. Instructional experts choose attributes to cluster on in the left hand window (in this case, lecture capture viewing habits), with the results of clustering being shown using a treemap of cluster sizes in the upper right. Each cluster can be labelled by the expert by clicking on the cluster, and details about the cluster are shown in the <i>membership tab</i> of the bottom window. In this case, five clusters were found ranging from four learners to 89 learners. | 43 |
| 4.4 | A mock up of the variables tab from the application shown in Figure 4.3. This tab allows the instruction expert to select those variables they are interested in correlating with the currently selected cluster. In this case, Michelle has selected to view the relationship between <i>Midterm 1</i> and <i>Midterm 2</i> variables for the current cluster. | 44 |
| 4.5 | A mock up of a lecture capture environment. The images down the left hand side control navigation through the video, and allow a user to seek to a particular <i>chapter</i> directly. The video playback component on the upper right includes video of both the instructor and the data projector, as well as a traditional scrubber that allows for non-chapter seeking. Additional tools such as note taking components, discussion forums, or suggested readings are available underneath the video. | 46 |
| 4.6 | An adaptive lecture capture system mock up nearly identical to the one shown in figure 4.5. In this system, Adam has customized the navigational indices at the left hand side to be different depending upon the tab that is active. Each tab contains different indices as appropriate for tasks that Adam has trained for using <i>supervised machine learning</i> . Users can change tasks by selecting a tab; by default they are shown the tab that most closely matches their system behaviours. | 48 |
| 5.1 | The iHelp Discussion forum system c.2006. The left side shows the different forums that a user has access to, in this case a number of course related forums as well as social and employment forums. Forums are heirarchical, and may include sub forums related to special topics. Each forum is made up of heirarchical <i>threads</i> of discussion as shown in the top window. Content for a given <i>posting</i> in a thread is shown in the middle of the application, and other e-learning tools (such as a synchronous chat system) can be shown at the bottom of the application. . . | 51 |

| | | |
|-----|---|----|
| 5.2 | Portion of a sociogram from an introductory computer science course which was taught in a blended fashion and had approximately 200 students and contained 254 discussion postings. Darker (red) nodes indicate facilitators, while lighter (grey) coloured nodes indicate learners. The inner circle is made up of participants, four of which are very important to the community (as shown by having a larger node size). A casual observation of this network indicates that, while some learners write a fair bit (many interconnected nodes in the middle), there are lots of learners who havent ever read anything (the outer ring of non-users), and many lurkers who read very little (as they tend to be closer to the outside of the middle ring instead of the inside of that ring). Note that the ring of non-users includes a disproportionately high number of facilitators as our currently deployment gives access to this forum to most staff and faculty in the department. | 53 |
| 5.3 | Visualizations of an introductory Computer Science course for non-majors. Figure 5.3a shows interactions amongst learners who attend traditional lectures, while figure 5.3b shows interactions amongst learners in a completely online environment. In the traditional course there are many lurkers who have read only a few messages, while the online course required participation of learners and thus has very few lurkers (only other instructional experts). In both cases the instructor is the principal actor in the social network (large dark (red) node), and learners rarely reply to anyone but the instructor. | 56 |
| 5.4 | A visualization of a subsequent offering of an online introductory Computer Science course for non-majors. This course was offered after the instructor made pedagogy changes aimed at having learners interact more with one another. In this example, there are several learners who are <i>important</i> to the community as shown by their large node size and high level of connectedness. | 57 |
| 5.5 | A visualization of an introductory course in Computer Science for majors. This visualization is made up of several sections of learners, each with their own instructor. The high degree of connectedness along with large student nodes (light grey in colour) supported the instructor's notion that course discussion would be sustainable if he reduced his level of involvement. . . . | 59 |
| 5.6 | Visualization of a general discussion forum (figure 5.6a) and a more specialized discussion forum which is unrelated to course content (figure 5.6b). The user cohort is the same, though qualitative remarks about participation in and importance of the discussions can be made by comparing visual aspects of the two renderings. | 60 |
| 5.7 | Visualization of two weeks of conversations, the first in debate style (figure 5.7a) and the second in role playing style (figure 5.7b). Learner nodes were named so that direct comparisons between the two sociograms could be made, though names were randomly assigned. Node size corresponds to quality of messages as determined by the instructor (directly related to the student grade), and a force directed layout was used to show interconnections between learners. | 62 |
| 6.1 | A screenshot of the Recollect lecture capture system. The system shows thumbnails of upcoming slides on the left of the window to allow for navigation. The right side of the player shows video of the classroom as well as video captured from the data projector. | 65 |
| 6.2 | Graphs plotting the prediction accuracy of learners versus week in the term. Generally, as more information is collected about learners, the correct predictions increase and the mispredictions decrease. Borderline learners are those who fit equally well in multiple clusters based on their activity. | 75 |
| 7.1 | The Recollect lecture capture system, showing navigational thumbnails on the left hand side. As users mouse over a given thumbnail, a small image opens up and shows what the data projector feed recorded at the corresponding time in the video. All interactions involving the thumbnails such as mousing over, clicking, or scrolling through the list, are recorded. | 82 |
| 7.2 | (a) contains pairwise κ between raters. (b) is an agreement map for 6 raters where distances between nodes are approximately inversely proportional to κ thus raters closer together have higher agreement. Node placement is done using a spring model layout. Edges are omitted for clarity as the graph is fully connected. | 86 |

CHAPTER 1

INTRODUCTION

One of the problems facing institutes of higher education is how to provide adequate support for learning that scales across departments, disciplines, and courses. Higher education increasingly makes use of courses with large cohorts of learners and smaller instructor-to-learner ratios. Bloom [17] demonstrated that learners who are taught in one-on-one learning have, on average, summative assessment marks two standard deviations higher than those taught in a traditional classroom setting. Given this finding, a move to more individualized learning seems appropriate.

One kind of technological solution aiming to provide personalized learning, and the result of decades of academic research, is *intelligent tutoring systems*. These systems form models of learners and adapt the learning environment based on learner performance, much the way a human tutor would. Despite several commercial successes (e.g., [34],[113]) and continued development of the research field¹, intelligent tutoring systems have been deployed only in a few isolated cases and have introduced minimal changes in day-to-day teaching and learning in higher education. In part this is an effect of the expense and labour required to build an intelligent tutor: it requires the work of a domain expert to outline how concepts relate to one another, a pedagogical expert to provide a method for determining when a learner has made a mistake and what should be done to mediate the issue, and a content expert to build initial and remedial content to be delivered to learners. Instructors in higher education rarely have such resources available to them, and instead have to rely upon their own understanding of each of the aforementioned areas, understanding which may be limited and which may be difficult to implement as an automated computational tutor.

Partly as a result of these difficulties, many institutions have taken a different approach to enhancing learning with technology. Over the last decade universities and colleges have rapidly deployed online learning technologies and *learning management systems* as an alternative to intelligent tutors, with cost effectiveness and scalability in mind. These learning management systems include content creation and delivery solutions, synchronous and asynchronous discussion forums, and multimedia streaming, capture, and playback systems, and provide infrastructure support for learning activities. They are widely used across higher education institutions, and augment both traditional teaching and learning experiences as well as distance and online learning.

¹In particular see the *Intelligent Tutoring Systems (ITS)* and *Artificial Intelligence in Education (AIED)* conference series, as well as the *Journal of User Modeling and User-Adapted Interaction (UMUAI)*.

In learning management systems the focus is not on personalization of content as much as it is on anticipating the general types of misconceptions that learners may have and planning for them. The environments are, in a sense, non-intelligent tutoring systems, where personalization of course content is difficult and the instructor is relied upon to make an *a priori* identification of the kinds of problems learners may have. Following this, the instructor forms some hypotheses as to how the problems would be best dealt with, builds content to address these concerns, and finally determines where the content should be put in the learning environment such that the appropriate learners will see it. Through successive offerings of the same course, the instructor refines the hypotheses and content as appropriate, until a change disrupts this process such as the course being taught by another instructor, the curriculum changes, or the backgrounds of the learners taking the course change.

These non-intelligent systems benefit from a reduction in the number of experienced programmers and knowledge engineers needed, which has helped scalability and thus led to greater adoption in higher education. Despite this, these systems suffer from some of the same resource issues that affect intelligent tutoring systems. In order to ensure that all learners are supported, a breadth of content needs to be developed that fits the expected needs and goals of those learners. In sufficiently large or diverse courses this results in a team of instructional design experts being used to build a comprehensive and pedagogically sound course offering. The end result is the same in terms of *a priori* effort: building instructional interventions requires significant up front modelling of learners, pedagogical approaches, and domain-specific content.

Learning management systems do afford some flexibility for instructors to make interventions if learners are having problems. These interventions are often in response to explicit problems that a learner encounters, such as a message requesting help being posted to an online discussion forum. Using their knowledge of the domain, the instructor can respond to these problems in an *ad hoc* manner, even seeking out new knowledge in order to respond to the learner, something intelligent tutoring systems do not do. Yet, the learning management system emphasizes up-front development of learning content, and does not encourage the personalization of this content to different learners. Further, instructors can only respond to that which they observe; as more interaction happens between the learner and the learning content management system, the instructor's ability to see the problems that the learners face and the ability to intervene in a timely manner is reduced.

1.1 Problem

The question I address in this dissertation is how insights based on the behaviours of learners can be formed, and how instructional experts might use those insights to form instructional interventions. In a traditional teaching situation, insights are gained *in situ* from the interaction the instructional experts (such as instructors, instructional designers, and tutorial assistants) have with learners. In technology enhanced learning situations the interaction is generally between the learner and the learning environment, and learning man-

agement systems do not make these interactions available to instructional experts, hampering their ability to make interventions. Instead of relying so heavily on a team of experts who have mapped out the space of possible challenges, difficulties, and misconceptions learners might encounter *a priori*, I believe that the instructional experts involved in delivering a course can use learner interaction data and make use of their contextual knowledge of the content, cohort, and pedagogy to provide a more individualized learning experience. By considering the issues that learners face while the course is being offered, a smaller highly contextual problem space (e.g., how to teach a specific concept that a specific group of learners are having issues with) need be considered instead of a broader general problem space (e.g., the design of a whole course for learners with a variety of backgrounds). This has potential to reduce efforts and costs in providing more individualized learning across institutional contexts.

Coming up with remedies for all conceptual problems (as in intelligent tutoring systems), or expert hypothesized problems (as in learning management systems) is time consuming. Instead, by focusing instructional activities on observed learner interactions and dealing with the actual problems a given cohort of learners is facing, the up-front planning of instructional remediation is minimized. This new approach, a data-assisted approach, is flexible with respect to both changes in curriculum as well as changes in learner demographics and pedagogical approaches. As courses evolve and learners with new characteristics interact with resources in new ways, the instructor can be made well-informed of learner activities, and use his or her expertise to identify and mediate problems that arise.

Instructors often employ pedagogical interventions as a part of their normal teaching practice in traditional lectures – it is well known that master teachers observe the behaviours of their students and adapt instruction on demand [33] – but as course sizes grow and interaction moves online, both the observation of social cues that inform the instructor as well as the ability of that instructor to intervene based on those behaviours becomes more difficult. In this dissertation, I show how making the hidden behaviours of learners visible to the instructional expert allows him or her to form *insights*, and enables him or her to react to these insights from technology enhanced learning environments with *instructional interventions*.

1.2 Solution

The solution I propose to the issue of improving personalized learning experiences in technology enhanced learning environments is to enable instructional experts with an *in situ data-assisted approach* based on the interactions of learners. Instead of pre-loading the learning environment with knowledge about the learning process (as in the intelligent tutoring system approach), this data-assisted approach sees the instructor and learning environment work together to augment the learner experience. As learners interact with the environment, their activities are summarized and presented to the instructional expert for consideration. With knowledge of the domains involved, the expert explores, labels, and acts on this knowledge by creating instructional interventions, either inside or outside of the learning environment. Thus the data-assisted

approach embraces and enables the contributions of instructional experts throughout the learning process, unlike many other solutions.

The key aspect of the data-assisted approach to supporting instructional interventions in technology enhanced learning environments is that both the insights and the interventions are the result of a dialogue between the intelligence in the system and intelligence of the instructional expert. Traditionally there are two forms of intelligent learning environments: those that are *adaptive*, and those that are *adaptable*. Adaptive environments are those described as intelligent systems that automatically change in response to user activity based on some pre-programmed knowledge base and rule set. This classification places intelligence in the software system itself. Adaptable environments are those which are personalizable based on direct requests from the learner. This represents a view that the intelligence exists in the learner alone. The continuum between these two approaches is referred to as the *locus of control*, and historically has placed artificial intelligence techniques at one end and human computer interaction techniques at the other. Any particular system can exist along this continuum, however, trading off control of the adaptation process to either the system or the end user.

In most higher education circumstances, such a diametric view of intelligence need not be accepted. Instructional experts are often directly responsible for maintaining the teaching and learning environment, and the knowledge of these experts can be leveraged by including them in the process of adaptation, fundamentally changing the locus of control. Whereas adaptive environments require much up-front planning, and adaptable environments are often limited in the depth of support they can provide, data-assisted environments use instructional expertise to provide deep pedagogical support as needed without additional upfront development costs. This inclusion of instructional experts as active participants in the intelligent educational environment is what makes the data-assisted approach novel. By including these experts, the data-assisted approach offers opportunity for minimal up-front design of the learning environment while providing deep and meaningful instructional interventions.

The data-assisted approach can be seen as being made up of two pieces. The first of these aims to generate pedagogical *insight* for instructional experts based on behaviours of learners. To do this, these behaviours are captured, aggregated, and represented in a way that experts can make sense of them. Sensemaking is not easily defined, as different kinds of instructional experts might form different kinds of conclusions based on a given visualization or statistical representation. However, through interaction with behaviour data, experts such as instructors can gain a sense of learners' progress and problems in a course, instructional designers can compare courses to one another and understand how they are similar or different, and educational technologists can understand how learners using a system are similar to one another.

The second piece of the data-assisted approach is the support of *instructional interventions* based upon these insights. To do this, the insights should be pedagogically relevant, and mechanisms should exist to tie insights to customizations of the learning environment. Not all learning scenarios call for a fine-grained instructional intervention such as the suggestion of new content (common in adaptive hypermedia solutions)

or the generation of a new problem set (common in intelligent tutoring systems). However, insights should lead to interventions within the teaching and learning process, and these interventions may be non-automated and high level (e.g., broad pedagogical changes made by an instructor) or automated and low level (e.g., the adaptation of content within a software system).

Satisfying these two goals enables a form of individualization of learning in higher education that is different from that of traditional adaptive or adaptable systems. The amount of up-front modelling and knowledge engineering required which is typically high in adaptive systems is reduced, while use of the deep pedagogical knowledge of the instructional expert (which is missing from adaptable systems) is increased. There is a cost in terms of time or effort on the part of the instructional experts themselves; however, these experts are already devoted to doing this kind of work in institutes of higher education.

In this dissertation I define the *data-assisted approach to supporting instructional interventions in technology enhanced learning environments*. I provide examples of this approach, and demonstrate how this approach can generate insights in different kinds of instructional experts. I describe how these insights can be turned into instructional interventions, which can positively impact the learning environment.

1.3 Contributions

Creating teaching and learning interventions in technology enhanced learning environments is difficult. This difficulty comes from a lack of support in making the interactions learners have with the learning environment available to instructional experts. The solution is to assist these experts with data relating to these interactions, making it easier for the expert to achieve *insight* into their learners and provide *instructional interventions* as appropriate.

This work is multi-disciplinary in nature, and uses techniques from Computer Science to make contributions to the disciplines of Computer Science, Education, and the emerging discipline of Technology Enhanced Learning. Sitting at the intersection of Computer Science and Education, Technology Enhanced Learning involves the creation, exploration, and understanding of new technologies and how these can be used for teaching and learning.

The techniques from Computer Science used in this dissertation come primarily from the fields of Artificial Intelligence and Human Computer Interaction. In particular, data mining, machine learning, and information visualization are used to demonstrate, discover, and verify claims of human and machine performance in forming insights from data and providing appropriate instructional interventions.

At a broad level, this work is a reformulation of *intelligent learning environments* to explicitly include the behaviours of learners and the sensemaking capabilities of instructional experts as they interact in the teaching and learning process. Whereas the traditional view of intelligent learning environments is one in which the software is imbued with artificial intelligence and replaces instructional experts at run time, this conceptualization takes a systems view of the technology enhanced teaching and learning process and includes

experts as *in situ* actors who can respond to learner behaviours. This reduces the need for *a priori* design and development of instructional interventions for situations that might not occur as the responses of the system are grounded in the behaviours of actual learners and their interacts with the learning environment.

In demonstrating this approach, several secondary contributions to areas relating to artificial intelligence and human computer interaction have been made. These contributions include:

- A demonstration that showing instructors the interactions of learners with technology-based learning environments support making high level pedagogical decisions. This dissertation provides evidence of how insight is generated using both online and blended teaching modalities, and is of interest to the communities of *computer supported collaborative learning* and *user modelling*.
- The improvement of methods for creating navigational indices in lecture video through the use of end-user data. This dissertation also provides a method by which to compare the suitability of a set of indices for a set of users. These results are an example of automated *instructional interventions*, and are relevant to the communities of *artificial intelligence in education* and *educational data mining*.
- The replicable discovery of a correlation between formal assessment outcomes and the patterns of behaviour of learners in lecture capture environments. This dissertation provides a method for model formulation and quantification of precision of said models, the results of which are useful in generating *insight* into the student use of learning technologies, a topic of interest to the communities of *educational data mining* and *learning analytics*.

1.4 Dissertation Outline

This dissertation is separated into eight chapters. A brief overview of related literature will be discussed in Chapter 2 with a focus on the area of *learner modelling*. Many different approaches have been explored to support a personalized learning experience, and the theories and techniques of learner modelling are at their core. After this, a more in-depth discussion of the data-assisted approach will be outlined in Chapter 3. This discussion will step through the details of the approach, focusing on how it might be implemented by software developers and how it relates to modern learner modelling theory.

The data-assisted approach offers a broad contribution to the field of technology enhanced learning, and can be realized in various ways. To motivate this, three scenarios involving different kinds of courses and instructional experts will be presented in Chapter 4. Each scenario will then be related to a chapter of detailed work showing how exposing learner behaviours generates *insight* (chapters 5–6) which is used to form *instructional interventions* (Chapter 7). The work concludes by looking at avenues for further investigation (Chapter 8).

CHAPTER 2

BACKGROUND

This chapter surveys the activities that have taken place over the last thirty-five years towards achieving personalized and adaptive learning in computational educational environments. The literature is both broad and deep, and the focus presented here is on overview of the field of *learner modelling* which, broadly described, aims to create computational structures or processes that describe the activities or mental states of learners. These models are then used in intelligent tutoring systems or other interactive learning environments to personalize the learning experience. These models are also important for the data-assisted approach where they are reflected back to instructional experts through visualizations or statistical representations.

Section 2.1 provides an overview of the field of learner modelling focusing on the historical and traditional techniques used to create models. A description of the more recent efforts towards data-driven modelling is presented in §2.2. Finally, this chapter ends with a brief discussion of some of the methods that have been used for the creation of adaptive learning experiences in §2.3.

2.1 Traditional Learner Modelling

2.1.1 Overlay Models

One of the earliest approaches to learner modelling which is still in use today is that of the *knowledge overlay* [35]. In this method, student behaviour is compared directly to known “correct” behaviour of an expert. As a learner makes a choice in the learning environment, a computational *coach* determines whether the decision corresponds to a portion of the expert model and updates the learner model as appropriate. The learner model then acts as a hypothesis of what the learner knows, and other computational processes can use this model to customize interaction with the learner.

This approach is effective in well-defined domains, but requires known correct solutions which often entails ontological modelling of the domain (which can be time consuming). The strictness of overlay models introduces tractability problems in ill-defined domains where the model is not a set of possibilities known *a priori*. The lack of an explicit representation of the assumptions leading to the formulation of the model can lead to inconsistencies as a learner may appear to understand a concept (for example, by guessing on an exam) that contains prerequisites the learner has demonstrated a lack of understanding in (by answering a related question on an exam incorrectly).

Most often, overlay models do not maintain the *how* of the user modelling activity, only the *what*. This tends to make them less suitable for dynamic learner modelling and probabilistic learner modelling as updates to the model limit the ability for future iterations of reasoning because the context leading to the formation of the model has been lost.

2.1.2 Stereotypes

In general, learner modelling suffers from many of the cold-start problems that other model-based artificial intelligence techniques suffer from: an accurate and detailed model requires a bulk of observation that is difficult to obtain before a model is already in place. The use of expert defined knowledge bases (*stereotypes*) for a canonical set of users is often used to try and alleviate these problems.

Rich [104] is one of the first to introduce the use of stereotypes in learner modelling through a book recommendation system, *Grundy*. In this system, a single-rooted directed acyclic graph represents classes of traits of users with increasing levels of specificity. Classes in the earlier levels of the graph might be very broad (e.g., the learner is a “religious person”), while classes further down the hierarchy a more specific class would override the traits of the general class (e.g., the learner is a “christian”). *Triggers* denote an action that determines when a stereotype should be applied to a learner.

Grundy disambiguates between the user model formed through inheriting stereotype information and the model that is formed through direct observation or inference from user action. Together the two models join to form the *user synopsis (USS)*, which is probabilistic in nature.

Stereotypes can be built either by soliciting expert input or through machine learning, and even the early work on Grundy was done with an eye towards adapting stereotype attributes based on how well they fit actual learners. In Grundy, stereotypes are a set of triples containing a string description (*facets*), a relative strength by which the facet applies to the stereotype (*values*, a number¹ from -5 to 5), and the level of certainty with which the facet applies (*ratings*, a number from 0 to 1,000). Stereotypes are updated by modifying both the value and rating for a particular facet. In the case of reinforcement of a facet, the stereotype value is changed using equation 2.1, while the stereotype rating is changed using equation 2.2.

$$new_value = \frac{old_value \times const + new_info_value}{const + 1} \quad (2.1)$$

$$new_rating = old_rating + \frac{new_info_rating}{old_rating} \quad (2.2)$$

More recent learner modelling approaches use automated clustering techniques such as k-means, hierarchical clustering, or fuzzy clustering, to form stereotypes based on bottom-up data. For instance, [55] uses a combination of all three clustering techniques on both behaviour and perception data from learners. Some of the resultant clusters created by these techniques are then correlated with a high significance to cognitive

¹Grundy also supported facets that were string literals, but did not provide a means for updating stereotypes based on these.

style (in particular, field dependance as outlined by [123]) and level of expertise.

It should be noted that the nature of a stereotype as an incomplete body of knowledge about a user makes it more appropriate to use stereotypes with inductive reasoning approaches [78]. The output of applying a stereotype in the reasoning process is a model (or piece of a model) that has some probability of being correct, and computational processes acting upon this model should consider this uncertainty appropriately.

2.1.3 Knowledge Tracing

The goal of knowledge tracing is to build a cognitive model of a learner’s skill acquisition over the learning process. This technique seeks to “monitor students’ changing knowledge state as they practice a complex cognitive skill, to individualize the practice sequence to enable students to master the skill efficiently and to accurately predict student performance” page 254 of [39].

The foundation of knowledge tracing is Anderson’s ACT-R theory of skill knowledge [2], refined from the ACT* production system work done earlier [3]. In brief, the theory suggests that cognitive skills are composed of production rules: small operations of data manipulation organized around atomic goals. Resolution of rules is done repeatedly with the facts available to a learner, causing them to demonstrate a particular higher level cognitive skill. Inability to resolve these rules in such a way that a skill is demonstrated indicates a lack of having the correct rules, and suggests a need for modification to the rules (learning).

Classical knowledge tracing applies this idea directly, and considers a learner model as a set of these production rules with each rule in one of two states, either learned or unlearned. State for a given rule can only change from unlearned to learned; there is no procedure to model forgetting a rule once learned. Learners are compared against the ideal learner model at each step in a problem solving situation, and the probability that a learner has learned a given rule is the sum of the probability that a rule was learned previously and the probability that the rule will transition to a learned state given the current opportunity.

The comparison of a learner’s current state to that of an expert bears remarkable similarities to overlay models, with the *ideal model* of the former directly equivalent to the *expert model* of the latter. The distinctive difference between these two methods is the process by which the model is constructed; overlay models consider learner knowledge as a declarative artifact and do not make any assumptions about the way a learner’s knowledge can change given a particular model, while knowledge tracing observes learners transitioning through a procedural space similar to a state transition graph.

In some of the early work of knowledge tracing, applications included incorrect models of production rules as *buggy models* [3] to be used in building intelligent tutoring capabilities. This has been expanded upon to cover situations where learners make mistakes (called *slips*) or guess [2], to include individualized learning rates [3, 100] for these and other parameters to support personalization, and to include support for evidence obtained from noisy environments [13].

As a technique, knowledge tracing has proven to be very robust, and various large scale systems have been built with this approach (e.g., [34]). Its principal weakness is that it suffers from many of the same

downsides as overlay models; e.g., that models must be complete, correct, and thorough, and are ideally suited for well-defined domains where domain experts are able to perform knowledge engineering tasks.

2.1.4 Constraint-Based Models

Constraint-based modelling encodes domain knowledge into a set of rules that describe the appropriateness of a particular solution state. These rules include both a *relevance* clause and a *satisfaction* clause, with the former determining whether the rule is applicable to the given situation and the latter determining whether the solution is correct or not [87]. A learner model is then represented as the set of solution states that the individual has violated.

This form of modelling is described in Ohlsson’s theory of learning based on performance errors. In this work, he argues that it is through making mistakes and correcting them that we demonstrate learning [98]. Providing a correct answer does not signify the learner understands; instead, the learner may just not yet have made a mistake and may have inadvertently answered correctly. It is the times the learner demonstrates mistakes that indicate learning is happening.

Ohlsson provides further insight with respect to remediation based on learner models:

“[i]t is not useful for a tutoring system to know that student S has bug A rather than bug B, unless the system has in its repertoire some instructional action that is more relevant for A than for B. If all its instructional actions are equally appropriate (or inappropriate) for both bugs, then the distinctions between the two bugs does not matter...” (From [97])

In this he argues that the support a tutoring system is capable of giving is a factor in determining the appropriate learner modelling approach.² The constraint-based approach more easily captures the vastness of false knowledge: a constraint designer doesn’t need to consider all of the steps a learner has taken to get to their current understanding, just whether the learner is violating any particular constraint.

The direct comparison of constraint-based modelling and knowledge tracing has been done by several scholars [91, 82, 92, 81] as both techniques aim to capture learner knowledge, and both techniques have been the root of successful tutoring implementations (see [87, 113, 91] for a partial list). There is no consensus over which method is more appropriate for a given situation, but the two methods can be compared more generally:

- Constraint-based modelling takes a declarative approach and describes a lack of knowledge, while model tracing takes a procedural approach and formalizes the steps of learning.
- Constraint-based modelling does not require an expert model as model tracing does, but requires expertise in the formulation of domain constraints. It is unclear, and likely both application and domain

²Ohlsson is not the only scholar to consider the question of what a tutoring system should be using learner models for. John Self, in his seminal work summarizing the challenges of learner modelling, provides four ‘slogans’ for learner model usage by intelligent tutors; (1) avoid guessing – get the learner to tell you what you need to know; (2) don’t diagnose what you can’t treat; (3) empathize with the learner’s beliefs, don’t label them as bugs; (4) don’t feign omniscience – adopt a ‘fallible collaborator’ role [108]. Slogan two in particular reaffirms Ohlsson’s statement as to whether diagnosis is a worthwhile endeavour, while the other slogans become a sort of genesis for the *open and scrutable* modelling techniques described in §2.1.7.

specific, whether the creation of constraints is as difficult to obtain as an expert model, suggesting that in different circumstances one approach may be more suitable than the other.

- Constraint-based modelling is not affected by the size of the task state space, but by the size of the domain solution space. A model tracing system is affected by the task space size but not the domain solution space. For procedural domains where the task space and the domain solution space are the same, both methods have the same scalability issues though may have different computational tractabilities depending on how they are implemented.
- Both constraint-based modelling and model tracing can recognize misconceptions and provide remediation. Model tracing uses buggy models [3] while constraint-based modelling uses buggy constraints [82], though the latter approach has been criticized as not being required, and that a violation of a constraint is demonstration enough of a theoretical bug [92].
- Feedback at a fine grained level (both fine grained in a domain sense as well as in a temporal sense) is more natural for model tracing systems while less common in constraint-based modelling systems [91].

The most significant critique of constraint-based modelling is that while it can be used to determine whether a learner is at a solution state, it does not necessarily identify how the learner arrived there, and thus may not be able to remediate appropriately [82]. Whether remediation can be provided (or the detail to which it can be provided) depends on how the domain is modelled, and whether the constraint authors have used *buggy* or *path* constraints in their modelling.

Some authors have suggested that the creation of model tracing production rules is more difficult than constraints [82]. The difference in perceptions of difficulty may lie in the level of expertise and cognitive load required to consider a domain holistically when authoring constraints or production rules. Creating constraints requires a broad view of the domain and an understanding of how concepts in it relate to one another at a deep level. This suggests that the knowledge engineer must have the ability to reflect on the domain, generally seen as a high order meta-cognitive skill. A procedural understanding of the domain, which is lower in educational taxonomies such as [16, 4] and may be more ingrained from practitioners' day-to-day experiences, aligns better with the knowledge tracing approach. The question of which approach is most suitable might then be considered from a resource availability perspective: if domain experts who have deep meta-cognitive and conceptual knowledge about the domain are available it may be better to use constraint-based modelling. Conversely, if domain practitioners with more procedural experience are available it may be more appropriate to use the knowledge tracing approach.

The question of scope and knowledge engineering process when building a learner modelling system might also be important to considering which approach to use, and the constraint-based approach offers a quick way to provide basic modelling. By creating broad general constraints, a system can be iteratively built (perhaps as knowledge tracing information is being collected) and refined. This tradeoff allows for a balance between simplicity and tractability versus precision.

2.1.5 Episodic Models

Episodic learner modelling [120] is based on Levine’s investigations in concept learning [85]. In this model it is suggested that learners follow set hypotheses, modifying them only when demonstrated incorrectly, and only then modifying them such that the hypothesis is satisfiable. Learners don’t change their method of enquiry until they have exhausted the search for their currently held hypothesis.

The learner model in these systems takes the form of a series of *episodes* made up of examples learners may have studied in course material as well as solutions they have provided through exercises. These cases are built up through the use of a diagnostic system which has access to task descriptions, domain knowledge, and learner artifacts created during learning (in Weber’s original work on the topic these were LISP code fragments [120]). Domain knowledge is represented as a hierarchical arrangement of concepts and generalizations or specializations of these concepts. Rules are provided with the concepts, and can capture optimal answers, suboptimal but correct answers, or incorrect answers. The first of these two sets of rules align roughly with knowledge tracing production rules, while the last is equivalent to a buggy model. Tasks in the episodic system are represented as algorithms that can be applied to solve a particular problem. These algorithms are hierarchical arrangements of concept rules that describe the task, and the learner model is thus a series of instantiations of concepts describing a path through problem tasks.

Episodic learner modelling can be thought of as a superset of knowledge tracing, where the knowledge gained about individuals is not lost between tutoring sessions. At any given stage in problem solving, the historic trace of the learner can be examined to customize feedback or perform evaluation of the learner’s knowledge. Episodic learner modelling differentiates itself from overlay models in that it may relate to but does not stress the expert model. An expert model can be used to evaluate the correctness of a given episode, but feedback for the learner is compiled both from this as well as from the historical perspective of user actions.

Individualization done by tutors using this approach is broader than in the case of knowledge tracing or constraint based modelling, because of this historic record. For instance, two learners who demonstrate an identical conceptual understanding of content may be tutored in different ways depending upon previous attempts those learners have made to understand that content. This could be included in both constraint-based and knowledge tracing methods, but would require explicitly representing the historical attempts of learners as either constraint satisfaction conditions or production rules, and would not be in keeping with the underlying theories for these approaches ([98] and [2] respectively).

2.1.6 Active Modelling

The active learner modelling paradigm proposed by McCalla et al. [89] leverages many of the ideas present in episodic modelling, but emphasises the need to create ad hoc representations of learners for specific purposes. The paradigm is best described in the authors’ own words:

“It is common to think of a ‘learner model’ as a global description of a student’s understanding of domain content. We propose a notion of learner model where the emphasis is on the modelling process rather than the global description. In this re-formulation there is no one single learner model in the traditional sense, but a virtual infinity of potential models, computed ‘just in time’ about one or more individuals by a particular computational agent to the breadth and depth needed for a specific purpose. Learner models are thus fragmented, relativized, local, and often shallow.”
(Page 53 of [89])

In addition to reconsidering the trend of monolithic model creation, the active learner modelling approach was used in large scale systems that demonstrated learner modelling in non-tutoring environments. For instance, in the I-Help environment, computational agents are used to match learners to one another to receive help for particular problems they may be encountering. Each agent maintained a model of the learners’ knowledge, interests, cognitive style, eagerness, helpfulness, interaction preferences, opinions of peers, and user actions [28]. A decentralized negotiation between agents was used to determine who should be helping whom in an on demand fashion [118]. The S/UM learner modelling environment [25] is tutoring-based, but similar to I-Help it focuses on matching human tutors to one another for collaborative and reciprocal writing sessions. The learner models that result from these situations are used only to match peers together and not to provide automated remediation to learner misconceptions as was the case of constraint-based tutoring and knowledge tracing paradigms.

The active modelling approach embraces diverse sources of information as well as the involvement of non-computational entities in the learner modelling process.³ In both the I-Help and S/UM systems, learners could self identify their own knowledge level as well as both the knowledge level and how willing a peer was to help. The breadth of the sources of data makes contradictory statements both possible and probable. Similar to episodic systems, contradictions must be solved during the reasoning process itself.

The domains in which the active modelling approach has been deployed makes evaluation of the approach uncertain. Unlike the rigorous and quantitative experimentation used in knowledge tracing experiments, systems built on the active learner modelling paradigm tend to be qualitatively evaluated without control populations. While there doesn’t exist evidence to suggest that the active learner modelling approach is limited to non-traditional cognitive tutoring, the flexibility of the input data makes it particularly appropriate in these kinds of situations.

2.1.7 Scrutable and Open Modelling

Most user models are hidden from end users, and are created from evidence only indirectly manipulated by the learner being modelled. Scrutable and Open modelling change this and include the user as a first class participant in both the production and consumption of learner models.

³Neither of these features are expressly forbidden in the previous outlined approaches, but implementations of other forms of learner modelling did not typically take this information into account. While it might be tempting to attribute this change to the active approach, it is more likely the result of increased interest in building learner models for non-tutoring environments as is described more in §2.1.7, as well as the increased level and diversity of technology used by learners.

Scrutable user modelling and the foundations of making models available to learners comes from the seminal work of Judy Kay (notably [75], also [38] and [77]). In these she argues that the end user should be able to know the process by which a user model is formed, the terms and nomenclature that the user model uses, and that the user should be able to modify the model and observe how their actions change the model. She goes on to argue that in doing so privacy is better achieved, accountability is strengthened, accuracy and correctness of the model can be better determined, and that reflective learning and meta-cognitive understanding can be better supported. This is supported by statements by Paiva et al. and their work on externalizing learner models [99], where they suggest that supporting the inspection of learner models, by computational or human agents, also provides important self assessment mechanisms and allows the user to aid the system in preventing misconceptions.

Kay identifies three actors in a user modelling system: the end user, the programmer (focusing on the intent the programmer had in developing the system), and the system itself (referred to as the machine). Each of these actors has knowledge which they share with others (public knowledge), and each has knowledge that they keep to themselves (private knowledge). She argues that for a model to be scrutable, each aspect of it must be attributed to a programmer who is responsible for providing an explanation of how the tool works. This requirement necessarily ties the scrutability of a model with the tool being used to update, generate, and evaluate that model; a requirement that is not present in other learner modelling paradigms.

At first consideration the use of stereotypes as used in Grundy [104] in overlay modelling doesn't seem as relevant to scrutable outcomes. A stereotype is a sort of shortcut to fit users into different coarse grained categories, and this doesn't correspond with the high individualization requirements of scrutability. Interestingly, Kay outlines that stereotypes are useful, and beyond the creation of the user model itself. Instead, stereotypes have value in the reflective activity learners undergo when they interrogate their learner models for information. For instance, when answering the broad questions of a learner as to whether they are "...a beginner?" or "What would be different if I were an expert?", stereotypes of other learners can be useful in understanding how one's own model differs from experts [76]. This offloads both the identification and remediation of the novice-expert gap onto the learner as opposed to onto a computational tutor found in knowledge tracing or constraint based modelling systems.

While scrutable user models were being developed, other researchers were working on a similar concept dubbed *open learner modelling* [26]. This approach relaxes the conditions set out by Kay in scrutable modelling, and argues that it is the content of the model that users need to access, and not necessarily the method by which the content was determined. Thus the goal of open learner modelling is on displaying the model and letting users interact with it, making the models *inspectable* [27]. The display of models is largely done graphically, and interaction can take many different forms for different purposes including awareness and reflection (which has developed as a whole sub-field of open learner modelling [18]), argumentation [126, 19], evaluation [19], or summarization.

To whom a learner model is made available is another question that open learner modellers have studied.

While the vast majority of open models are intended directly for use by the learner, work has been done making models available to other learners [32], instructors [31], and parents [84].⁴For instance, Bull et al. [30] allowed learners to control whether others could see the learners model, and allowed the learner to make the models available in a pseudonymous form (e.g., “User 65’s model”) or a form that revealed their identity. These learners could then share their models with instructors or peers, and experimental results demonstrated that learners were more willing to open their models to other learners in their course, especially when those learners were also their friends. Hansen [66] goes further and gives fine grained access to learner models in the iHelp Courses learning content management system. These models are created by the instructor using a query tool which allows for the inclusion of both fine grained dynamic information (e.g., page views, number of postings in discussion forums, etc.) and static information (e.g., name, email address, etc.). Models can then be shared with different cohorts of learners depending on the purpose; online learners might be able to see detailed models in order to increase a sense of community, while in class learners may be limited to more privacy preserving models. Different stakeholders in the academic process may have different levels of access as well; tutors may not be able to see learner email addresses or names, while instructors may be able to see everything (if they so choose). This approach blends the open learner modelling and active learner modelling approaches [65].

Providing learner models to individuals on the periphery of the learning activity has also been investigated. Most notably models of young learners where parents are involved ensuring that motivation and outcomes remain high. Lee and Bull [84] have involved parents by creating separate language and visualizations for them in order to summarize more quickly the kinds of misconceptions their child may have. Zapata-Rivera et al. [127] provide similar summarized performance visualizations for parents of learners who are taking part in national testing, while learners received more in-depth comparisons of their performance compared to their peers.

A discussion of open and scrutable learner modelling is not complete without acknowledging the burgeoning field of learning analytics. Siemens refers to learning analytics as the “use of data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning” [110]. Researchers working in the field of learning analytics rarely identify themselves as doing learner modelling *per se*, but both the techniques and outcomes are similar to those who are doing open learner modelling. A more full discussion of learning analytics as a field is addressed in §2.2.1.

2.1.8 Group Modelling

Modelling groups of people has been broadly investigated in the fields of management, human resources, and psychology. Even within the fields of education and computer mediated communication, there are dozens of conferences and journals that describe how people interact in groups. Instead of providing an overview

⁴A comprehensive overview of the many open learner modelling systems can be found in [29].

of this literature, this section will focus on specific efforts to characterize learning groups within the field of learner modelling, where the result of the modelling effort is to enable an intelligent learning environment.⁵

Hoppe [71] approaches the intelligent tutoring task as a question of how to best form groups of individuals in an effort to support cooperative learning and social constructivism. In this work, he posits that the cognitive models of learners can be used to generate three interesting situations where groups of learners are useful:

- Where a competency imbalance allows one learner to take the role of a tutor, thus leading to an increase in competency of the second learner.
- Where a competency balance between two learners encourages competitive behaviour, and thus an increase in competency for both learners.
- Where a problem requires specialized knowledge that no single learner can provide, and thus can only be solved by creating a group of learners with complementary competencies.

Hoppe continues to explicitly describe these situations and inputs as parameters in a formal sense using overlay models; given s as a particular student, T as a particular skill or topic that mastery has been achieved in, and K as the set of all possible knowledge elements (where $T \subseteq K$). A student model for student s is $SM(s)$ where $SM(s) \subseteq K$. He then defines primitive functions describing whether a student knows some topic (equation 2.3) has difficulties with the topic (equation 2.4), and whether a particular problem P is solvable by a group of learners (equation 2.5).

$$knows(s, T) \Leftrightarrow T \subseteq SM(s) \quad (2.3)$$

$$has_difficulties(s, T) \Leftrightarrow T - SM(s) \neq \emptyset \quad (2.4)$$

$$solvable(P, s) \Leftrightarrow \text{for each } t \in P : knows(s, t) \text{ where } p \subseteq K \quad (2.5)$$

A competency imbalance is then described as:

$$can_help(s1, s2, T) \Leftarrow knows(s1, T) \wedge has_difficulties(s2, T) \quad (2.6)$$

A competency balance, and thus competitive situation, is described as:

$$competes(s1, s2) \Leftarrow \forall t \in T : knows(s1, T) \Leftrightarrow knows(s2, T) \quad (2.7)$$

⁵For an overview of the field from an educational psychology perspective, the interested reader is directed to the work of Dillenbourg et al. [44]

And the specialization knowledge situation as:

$$\begin{aligned} \text{Given group } G = \{s_1, \dots, s_n\}, P \text{ is adequate for } G \Leftarrow \text{not exists } i \in \{1..n\} \\ \text{such that } solvable(P, G_i) \wedge solvable(P, \cup_{SM}(G - G_i)) \end{aligned} \quad (2.8)$$

Winter [122] broadens the investigation of group modelling, and looks at larger sizes of collaborative learners and domains beyond the cognitive. In his work, teams of nine learners are put together and asked to do software development activities over an academic term. He models the cognitive knowledge and skills of each person, as well as teamwork skills (using Belbin’s team role taxonomy [14]), and a measure of personality characteristics which he refers to as “social intelligence”. He includes in his model a characterization of the task, showing strong relationship to the active learner modelling paradigm described in §2.1.6. Winter goes on to use C4.5 data mining to develop a decision tree which can be used to predict group success based on individual learner models.

The work on the I-Help 1-to-1 system [117] leveraged many of the ideas of Hoppe’s working in a running tutoring system. In it, learner models included a self-diagnosed cognitive element, as well as the willingness of a learner to help and their availability. Learners could search the system for peers and a matchmaker process would determine those peers who were best suitable to help in real time. Like Hoppe’s work, teams were limited to two individuals, and the implementation of the matchmaking algorithm favoured the unbalanced criteria where one learner was in the role of the tutor. Later iterations of the tool included a market-based incentive mechanism to provide extrinsic motivation for learners to help their peers.

Various authors have taken the approach of modelling groups of learners as an aggregation of individual learner models. This is most prevalent in the open modelling research (e.g. in [30, 127, 79, 66]), where learners or tutors are often able to graphically compare their knowledge or progress through curriculum to other learners. In addition to work done by learner modelling researchers, the use of graphical methods to compare learners has also been well studied in the fields of computer supported collaborative learning and social network analysis.

Groups of learners in social networks have also been modelled using visual techniques. Notable here is the work of Erickson and Kellogg [49] on Social Translucence. In this work, graphical proxies are used to represent participants in situational interactions. For instance, a wedge might be used to represent a lecture, with the lecturer identified by a small circle at the tip of the wedge, and the audience members each identified by small circles at the broad portion of the wedge. As audience members interject with questions or comments, their circular proxy moves towards the lecturer, indicating a change of status. Erickson and Kellogg indicate such social proxies have four basic characteristics [from [49]]:

1. Figure-Ground. A social proxy typically consists of two components: a relatively large geometric shape with an inside and an outside and sometimes other features that represent the online situation or context (e.g. a circle), and much smaller shapes positioned relative to the larger shape (e.g. small

colored dots) that represent participants.

2. **Relative Movement.** The presence and activities of participants in an online context are reflected in the location and movement of the smaller shapes relative to the larger one. Most often, the relationships and movements of the proxys visual elements have a metaphoric correspondence to the position and movement of peoples bodies in face-to-face analogs of the online situation
3. **Public Not Personal.** Social proxies are public representations. That is, everyone who looks at a social proxy for a given situation, sees the same thing. It is not possible for participants to customize their views of a social proxy. This is important because it is what supports mutual awareness and accountability: I know that if I see something in the social proxy, that all other viewers can see it as well.
4. **Third Person Perspective.** Social proxies are represented from a third-person point of view. When I look at a social proxy, I see myself represented in it in the same way that other participants are represented. This opens an important avenue for learning. As I act within system, I can see how my actions are reflected in my personal representation, and thus I can begin to make inferences about the activities of others.

This is similar to the work of Upton and Kay in Narcissus [116], which builds a visual group model of learner interactions with software development systems. Like the social translucence modelling of Erickson and Kellogg, the group model is shown as an aggregation of individual models of learners. Unlike the social translucence work, each participant in Narcissus has a detailed individual model (in the open/scrutable sense) which is inspectable when observing the group model. This allows for learners to compare themselves to their team members when considering the overall group model.

Finally, measuring a group of learners over time as a multivariate entity has been the goal of researchers interested in quantifying *social capital*. This measurement is often done through analysis of artifacts of interaction (for instance, discussion forum messages or collaboratively authored documents). The work of Daniel et al. [42, 41], for example, uses data mining of learning artifacts to determine initial probability values for an expert-generated Bayesian Belief Network (BBN). This BBN models elements of the community such as trust, shared understanding, hospitality, and task awareness. When loaded with values, the effect of the actions of learners on the community can be observed. As the learners' progress introduce new actions, the values in the group model can be updated giving a constant understanding of the changing social capital.

2.1.9 Conclusions

Initially, learner modelling activities were focused on building models in the cognitive domain for intelligent tutoring systems. The two dominant modelling techniques for tutoring systems were and continue to be the *knowledge tracing* and *constraint-based modelling* approaches. The community has been unable to clearly

distinguish which method is a better approach and, despite attempts by several authors to do so, there remains a lack of clarity as to when one approach should be used instead of another [82, 92, 81].

The differences between the two approaches relate to the educational theories which they are based upon; one could create a constraint-based model which effectively described the set of production rules of a knowledge tracing system, and such a system would be functionally equivalent to knowledge tracing system. Similarly, one could invoke production rules that describe only the set of errors a learner could make and is thus functionally equivalent to the constraint-based satisfaction clauses.

Perhaps the more useful consideration between the two approaches is in the initial formalization of the production rules or satisfaction clauses. Consider for a moment educational taxonomies such as Bloom's [16] or Anderson's [4] reformulation; if subject matter experts with meta-cognitive knowledge of the domain are available, then it may be more efficient to have them use the constraint-based approach over the knowledge tracing approach, while the reverse maybe be true if the experts available have hands-on practical experience with the domain but not meta-cognitive knowledge of it. Thus a distinguishing feature between the two approaches is on how they are positioned towards different groups of modellers, and not so much as on how they can be used to describe learners. This consideration has yet to be explored in the literature.

The storage and computational ability of modern computers has pushed both the boundaries of techniques and the purposes for which a model was needed beyond the intelligent tutoring system; it has become possible to store every interaction a learner had with a system such that the purpose for modelling could be dynamic (as in the *episodic* and *active* approaches), and interaction with the learner about their model became a desired featured (as in either the *open* or *scrutable* approaches).

It is unclear if these newer techniques are as appropriate for the original goal of cognitive modelling inside of intelligent tutoring systems. For instance, if one is creating an intelligent tutor for the domain of atmospheric physics, is using an active and episodic approach appropriate, or is a constraint-based model more accurate and efficient? Are multiple episodes compatible with production rules and, if so, how does one include reasoning over previous interactions with respect to the current state of the learner? These issues are poorly addressed by the current body of literature; instead, each technique is typically introduced with new purposes for which that technique may be well suited.

2.2 Data-driven Learner Modelling

Data-driven approaches use a statistical analysis of the behaviours of learners to form or inform descriptive models. This approach is a significant divergence from the techniques described previously in §2.1. Compared to cognitive modelling methods (such as knowledge tracing and constraint-based modelling) which tend to have deep roots in cognitive psychology, data-driven approaches provide for learner models which see their structure emerge from end-user interaction and not theory.

The division between the theory-driven and data-driven approaches does not have to be a hard

boundary, and a growing amount of work in the area is recognizing this. Learner modelling researchers can use data-driven modelling techniques to verify theory and, conversely, use theory to make tractable the problem of what data and techniques should be chosen for understanding learner behaviour. In this way data-driven approaches can work with theoretical approaches in a sort of feedback loop, constantly validating theories yet providing new hypothesis for further investigation.

The move to a descriptive understanding of learners based on their actions prompts another interesting shift for the field. Previously, theory drove implementation and implementations were used to change learner behaviours. Achieving statistical significance has been problematic for researchers in the field, but the end goal was either a positive increase in learning or one of the many indicators linked to learning.⁶ The descriptive understanding made available by data-driven approaches is now much more akin to categorization activities in other sciences, such as early taxonomy creation in the field of Biology. Instead of proactively determining the cause and effects of interventions through experimentation, the learner modelling researcher can act in a completely observational capacity and characterize learners as they go through the learning process. A side effect is that it is relatively easy to outfit educational tools with data collection features, and many data-driven models have risen from large cross-domain studies where statistical significance has been achieved.

Research in this subfield of e-learning primarily comes largely out of a series of workshops, conferences, and journals on *educational data mining* and *learning analytics* which are described in more detail in the remainder of this section.

2.2.1 Educational Data Mining and Learning Analytics

The term Educational Data Mining (EDM) has been recently popularized in learner modelling research circles as a focal-point for data-driven approaches to learner modelling. The kinds of data used are varied and purpose-driven, and Ryan Baker, generally seen as one of the founders of the area, provides a definition:

“Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in.” (Page 4 of [10])

Baker goes on to an analysis of the literature to date [10] and indicates four different significant purposes which EDM researchers are aiming to understand: student modelling, domain knowledge modelling, pedagogical support, and validation of educational theory.

Romero [105] provides an insight into the target of EDM research by adding academics (administrators) and educators as principal entities e-learning systems must interact with. Whereas the bulk of traditional learner modelling research is aimed at learner-facing systems such as intelligent tutoring systems, it is common

⁶Although some researchers have managed to provide large cross-domain investigations of their technologies, most results are limited to within a domain or even within a cohort of learners.

to see EDM researchers building systems to be used by instructors, instructional designers, and administrators.

Learning Analytics (LAK) has much overlap with EDM. Formed independently, and made up of a body of researchers that includes some EDM researchers as well as others, learning analytics has been defined by Siemens as:

“...the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs.”
(Page 1 of [112])

Siemens goes on to differentiate learning analytics from institutional analytics or business intelligence, by suggesting that the target audience for learning analytics is explicitly not administrators, but learners and educators [112]. This differentiation isn’t universally agreed upon; other definitions of learning analytics, such as the following by Educause Next Generation Learning Challenges program [46], include administrators as a participant in the system:

“[T]he real-time use of learning analytics by students, instructors, and academics advisors to improve student success.”
(From [47] as cited by [48], page 4)

2.2.2 Motivations and Techniques

In a review of the literature, Baker [10] distills a number of different techniques commonly used in the field of EDM (summarized in Table 2.1). Frias-Martinez et al. provide a more comprehensive description of these techniques, though they aim their work specifically at the User Modelling and Adaptive Hypermedia communities [54]. A high level summary of these works is that there are two broad approaches to data mining, unsupervised and supervised, and that these approaches correspond to techniques for the labelling of data (*clustering*) or the discovery of rules between variables in the data (*association rule mining, classifying*) respectively.⁷

Unsupervised methods, such as k-means, fuzzy k-means, expectation maximization, and hierarchical clustering can be used to take data and group instances into categories. Kardan and Conati [74] used k-means to group learners into categories of “high” and “low” clusters depending on whether or not there were significant learning gains after they interacted with in an online courseware system. HersHKovitz and Nachmias [68] used between-groups linkage methods (a hierarchical clustering technique) to group learners into three motivation-based groups corresponding to “engagement” if learners are highly motivated, “source” if learners have extrinsic motivation, or “energization” if learners were quick and highly performant. Brooks et al. [22] use k-means to group learners into six different categories based on their amount of usage of a lecture playback system, corresponding to learning strategies the students may be employing over an academic term. Mercer on and Yacef [90] use a number of data mining techniques, including k-means clustering, to cluster

⁷Frias-Martinez et al. provide much more than this high level discussion, and those interested in understanding how specific techniques work is encouraged to read [54] for details.

| Category of Method | Goal of Method | Key applications |
|--|---|--|
| Prediction | Develop a model which can infer a single aspect of the data (predicted variable) from some combination of other aspects of the data (predictor variables) | Detecting student behaviors (e.g., gaming the system, off- task behavior, slipping); Developing domain models; Predicting and understanding student educational outcomes |
| Clustering | Find data points that naturally group together, splitting the full data set into a set of categories | Discovery of new student behavior patterns; Investigating similarities and differences between schools |
| Relationship Mining | Discover relationships between variables | Discovery of curricular associations in course sequences; Discovering which pedagogical strategies lead to more effective/robust learning |
| Discovery with Models | A model of a phenomenon developed with prediction, clustering, or knowledge engineering, is used as a component in further prediction or relationship mining. | Discovery of relationships between student behaviors, and student characteristics or contextual variables; Analysis of research question across wide variety of contexts |
| Distillation of Data for Human Judgement | Data is distilled to enable a human to quickly identify or classify features of the data. | Human identification of patterns in student learning, behavior, or collaboration; Labeling data for use in later development of prediction model |

Table 2.1: Data mining techniques commonly found in EDM research, from [9] as cited in [10].

learners based on the mistakes they made in academic exercised. Regardless of the method used to cluster learners, labels are applied to the clusters by the investigators and are often motivated by qualitative or theoretical research.

Supervised methods require annotated data, usually provided by human experts. This data is made up of a set of instances, each of which is made up of a set of attributes and a category label. Machine learning algorithms, such as C4.5, k-nearest neighbour, or support vector machines, use this data to create models describing relationships between the attributes and the categories that exist. These models can then be used to predict the category of future instances without requiring manual labelling. Breşfelean [51] provides a fairly simple example of this, where learners were given questionnaires about their background (such as secondary school enrolment characteristics, gender, etc.) and goals (an undergraduate diploma, a post graduate degree, etc.). Using the C4.5 algorithm, a J48 decision tree was created that was able to model which background attributes would affect learner goals, to an accuracy of 88.68%. Rus et al. [106] provide a more elaborate use of supervised learning which is more indicative of the current state of the field. In this work a variety of techniques were used to create learner mental models as either “high” “intermediate” or “low” based on a set of 284 learner-provided essays. Each essay was analyzed using three different methods; an expert-defined taxonomy, n-grams, and expert-defined descriptions, and each method corresponds to a different set of attributes to be considered (from seven taxonomical entries for the first method, to over 1,038 weighted words for the last method). The investigators showed that an accuracy of 76.31% ($\kappa = 0.63$) could be achieved using Bayesian belief networks when considering expert-defined descriptions.

It is common for educational data mining researchers to cluster or classify data with multiple methods using the WEKA toolkit [69]. Consumption of analytics information tends to be done through different forms of information visualization. It is not uncommon to compare learner data through a myriad of bar charts, line graphs, or other tools common in data mining suites such as pentaho [102] and Tableau [114], and various researchers have built platforms on these methods, such as the Student Activity Monitor [59]. Novel visualizations of learner data are often purpose driven and more limited in their scope. The Purdue University project Course Signals [115] provides an example of a very simple visualization backed by 20 indicators of success from the Blackboard learning content management system. In this system, learners are shown a traffic-light inspired image for each course they are involved with: green indicating the learner is on track, yellow indicating the learner could do better, and red indicating that the learner is in danger of not meeting criteria.

2.2.3 Differentiating between EDM and LAK

Differentiating the fields of educational data mining and learning analytics has been a concern of several researchers. Siemens takes the position that educational data mining encompasses both learning analytics and academic analytics, [112] the former of which is aimed at governments, funding agencies, and administrators instead of learners and faculty. This distinction is not one that all agree with: Baepler and Murdoch, for

instance, define academic analytics as an area that “...combines select institutional data, statistical analysis, and predictive modeling to create intelligence upon which learners, instructors, or administrators can change academic behavior” page 3 of [8]. Baepler and Murdoch go on to attempt to disambiguate educational data mining from academic analytics based on whether the process is hypothesis driven or not:

“Analytics is associated with a scientific, hypothesis-driven approach, while data mining has a legacy with strategic business techniques and marketing. The fact that the latter method – data mining – typically lacks a hypothesis to drive an investigation can seem troubling, but it’s a distinction that might be rendered immaterial when it produces insights. That is to say, if a model works, even if one does not understand exactly how it works, the results may still be valuable even if they lack an originating hypothesis.” (Page 2 of [8])

The authors also claim that the work done in the EDM field is exploratory in nature, but do not provide specifics about how they come to this conclusion. A brief review of past EDM conferences suggests that much of the published work in the field is scientific work iterating on previous results [96, 58, 60, 95], making the distinction of EDM as exploratory in comparison to LAK questionable. Further, the level of statistical rigour required for publication at the EDM conference series is much higher than the LAK conference series, while the latter provides venues for a more diverse form of publication.⁸

Perhaps a better distinction between the EDM and LAK communities is in the roots of where each community originated. Authorship at the EDM annual conference is dominated by researchers coming from the intelligent tutoring paradigm. These researchers are most often associated with computer science departments, and the community has strong roots in publishing at the Artificial Intelligence in Education (AIED) and Intelligent Tutoring Systems (ITS) conferences.⁹ These communities model learners with the end goal of creating intelligent learning environments; the models are runnable, prescriptive, and papers often include a discussion of interventions or meditations that take place because of aspects of the learner models.

The authorship at the LAK annual conference is also heavily participated in by researchers from computer science departments, but includes a mix of researchers from philosophy and education departments as well. Papers presented at the last LAK conference did not focus on intelligent learning environments, and topics such as visualization for human intervention and the analysis of analytic information received more attention.¹⁰ Thus the modelling of learners is often to gain deeper insight into their behaviours, but not to build any particular software system which automatically adapts or mediates based on this behaviour.

Regardless of the differences between the LAK and EDM communities, the two areas have significant overlap both in the objectives of investigators as well as in the methods and techniques that are used in the investigation.

⁸The 2011 EDM conference provided traditional peer-reviewed, short and long papers, and a demonstration track. The 2012 LAK conference included these forums, as well as tutorials, workshops, panels, and a design briefing track.

⁹Before the conference series was founded in 2008 workshops run at a mixture of predominantly computer science conferences, including several AIED and ITS conferences, as well as the User Modelling and Adaptive Hypermedia (UMAP) and the Association for the Advancement of Artificial Intelligence (AAAI) conferences.

¹⁰No papers from the 2011 proceedings for this conference mentioned intelligent tutoring systems by name, and only four of the papers (including this author’s) mentioned artificial intelligence explicitly as an area of study, though techniques commonly attributed to the area of artificial intelligence were commonly used.

2.2.4 Ecological Approach

The ecological approach [88] developed by McCalla builds on previous work including active modelling [89] and purpose-based [94] modelling. The approach considers the codification of the *interaction* of learners with learning resources to be the key element in creating usable and situated learner models. This process is best described by the author himself:

“In a phrase, the approach involves attaching models of users to the information they interact with, and then mining these models for patterns that are useful for various purposes. The information and the data mining algorithms interact with one another in an ecosystem where the relevance and usefulness of information is always being adjusted to suit the changing needs of learners and teachers and to fit changes in the external environment and the system’s perception.”

(Page 1 of [88])

McCalla divides the learner model into two parts; the first representing the *characteristics* of a learner that may be true for multiple learning activities. These characteristics include items such as gender, affective state, learning style, etc. The second portion of the model is the *episodic* history, which covers a range of information relating to a single learning activity, such as learner evaluation of content, results of interaction with the content, or even keystroke level details captured during interaction with the content. Both portions of the model are associated with content after it has been interacted with by a learner, and these learner model attachments become the basis for high level activities of an intelligent learning environment such as content adaptation, recommendation, or customization.

These models are then associated with learning objects in the system, and reasoned over for particular purposes in an on-demand manner:

“This gradual accumulation of information and the focus on end use are two key aspects of the ecological approach. The third key aspect of the ecological approach is the purpose - based use (in the sense of McCalla, Vassileva, Greer and Bull, 2000) of the information associated with the content, to achieve a particular goal...it is the purpose that determines what information to use and how it is to be used. Further this determination is made actively (in the sense of McCalla, Vassileva, Greer and Bull, 2000) at the time the purpose is invoked: no apriori interpretation needs to be given to the information.

Thus, the ecological approach for educational domains boils down to this: each time a learner is interacting with a learning object, the learner model for that learner (in its current instantiation, of course) is attached to the learning object. Over time this means that each object collects many learner models and these can be mined for patterns of particular use to a given application.”

(Page 5 of [88])

McCalla suggests that the ecological approach is most suited for learner-centric constructivist pedagogies, although he provides references to his own work with others that includes elements of social constructivism and situated cognition [117]. A criticism of the ecological approach presented by Frielick [56] paints it as a “...neo-Darwinian attempt to apply artificial intelligence to education...” where human traits are “...reduced to other kinds of disembodied ‘objects’ that, like the earth, can be ‘mined’ for useful resources...”. Frielick goes on to suggest this is an antithesis of ecological¹¹ and espouses the work of Bateson [12] suggesting that the

mind is an emergent property of the larger mental systems in which we exist. He then poses a redefinition of what an ecological approach might be, as an “ecosystemic process of transforming information into knowledge, in which teacher, subject and student relationships are embedded or situated in a context where complex interacting influences shape the quality of learning outcomes” [56]. This description shares a similar quality with McCalla’s in that context is an important factor for determining success of learning. It differentiates from McCalla’s in that it only considers interaction between persons to be valuable: content in the form of intelligent learning environments or static learning objects plays little to no value in forming expertise, which he claims “can only be gained in the real presence of intercorporeal engagement with a master or expert” [56].

Several researchers have outlined systems similar to the ecological approach¹². The Massive User Modelling System [23] by this author and others describes a decentralized semantic-web system for collecting learner interactions. Each interaction of significance is represented in an *opinion*, which is temporal in nature and may be described using any number of ontologies.¹³ Opinions are routed to software systems called *consumers* who either reason over opinions for immediate use (e.g., content recommendation) or enrich the opinion with more information and rebroadcast it. In this way, a collection of different e-learning tools can share information about learners for on-demand interventions. Similar to this is the attention metadata effort from Najjar [93], Wolpers [124], and Duval [45] which, while currently only applied to e-learning, is aimed at the broader effort of realizing interaction-based data-driven approaches in a general case. Compared to the MUMS approach, the attention metadata schema is more rigid and has several predefined categories for technical, semantic, and contextual information surrounding the content use. Several kinds of learning environments (blogs, instant messaging applications, browsers, content management systems, etc.) have been modified to produce attention metadata.

2.2.5 Conclusions

Data-driven approaches have had an impact on how and why learner modelling is practised. In the traditional approach, a learner model was created by a pedagogical expert for a specific purpose and perhaps even a specific group of learners. The model was a codification of important attributes of the learner as envisioned by the expert, and was tied to the pedagogical theory the expert practised. Data-driven approaches change this in that the model emerges from the interaction data of the learners with learning systems: this interaction is often multi-modal, and can be with content, software applications, or with other learners and instructors.

¹¹Frielick further attacks this as part of the “...modernistic reductionism that plague much of the literature on instructional design, learning objects, and e-learning environments” [56] In this statement he is not wrong; the ecological approach, and all of educational data mining, views learners as complex entities that can be modelled by considering various sources of evidence, usually learner-system interactions. Learning objects are also reductionist attempts to describe content and its structure.

¹²It might be useful to note that while authors of these systems seem to share much in mind with McCalla’s approach, none of the systems actually attach learner models directly to the learning content. Instead of being learning object focused, they are interaction metadata focused, and use the metadata that forms the learner models as the principal artifact of interest. Further, neither work explicitly indicates their goal to adhere with the ecological approach.

¹³*Opinions* are codified as RDF fragments.

The diminishing role of the content expert is particularly stark given that the exciting work in the field a decade ago was largely based on the semantic web, a technology that required human expertise.

Data-driven approaches have largely been made more tractable by the increased use of technology in daily lives: in the most extreme cases, all interactions except for internal processes of a learner are mediated and monitored through computational devices.¹⁴ The broad deployment of learning management systems within institutes of higher education has furthered the ability of researchers to investigate data-driven approaches. Whereas early learner modelling activities required a dedicated cohort of students willing to use the system (and a researcher willing to support this use), many data-driven activities can be done on the activities of students in whole programs, departments, or universities with minimal extra effort.

2.3 Instructional Remediation Through Adaptation

Remediation of the learning process by adapting content or the presentation thereof is commonplace throughout many forms of education, formal and informal. In higher education this is largely done by the instructor who notes the behaviours of learners as they interact in a course and changes lectures, assignments, examinations, or other learning activities as appropriate. Instructional designers can also play a critical role in determining appropriate remediation, and use their knowledge of educational theory, pedagogy, and content to build adaptations that increase the motivation and performance of learners. The focus of this section of the background work is on a brief survey of automated adaptation techniques – those interventions which are done by intelligent systems instead of by the instructor or instructional designer. While traditional remediation can be used with the data-assisted approach, and will briefly be discussed in Chapter 4, the emphasis here is on describing the automatic or semi-automatic methods that are currently used in conjunction with learner modelling.

2.3.1 Content Adaptation & Sequencing

By far the most common instructional intervention done in intelligent learning systems is content adaptation and re-sequencing of content. The area of *Adaptive Hypermedia* in particular provides many different examples of content adaptation based on learner models. Popular techniques include adapting the form of content to fit the learning style or personality style of the learner [7, 24, 70, 61], the level of granularity of the content to appropriately address the particular issues a learner may be facing [119], or the sequence of content pieces that change how the learner navigates between various topics in the system [73, 52]. In a survey on the area, Conlan describes four axis of adaptation [37]:

- Adaptive Navigation: Customization of the link structure of content based on a learner model. The navigation presented by the system determines the amount of guidance and freedom a learner is given.

¹⁴With the work being done in correlating learner behaviours with neurological indicators such as [67], it's perhaps a question of *when* and not *if* we might be able to broadly data mine learners' internal processes.

Evidence has been shown that depth of knowledge in a topic leads to less linear navigation.

- **Structural Adaptation:** Customization of the spatial structure of content based on a learner model. Techniques such as maps, fisheye lenses, filters, and pivot browsing can be used to adapt the selections that are available to the learner.
- **Historical Adaptation:** Adapting content to show the temporal nature of access throughout the system. Techniques such as footprints, breadcrumbs, or landmarks that are made by the learner as they progress through content can be used. More recently, and outside the description given in [37], collaborative historical adaptations can be shown (e.g., [11]).
- **Adaptive Presentation:** Customizing content based on a learner model. Conlan suggests that artifacts of adaptation can begin at the word level, and be aggregated up into sentences, paragraphs, and *pagelets* which describe an atomic topic. Different pagelets can be created for different learners based on various aspects of the learner model.

One of the larger activities that has been done to support the delivery and adaptation of content in learning content management systems has been the work on *learning objects*. This has led to a standardization of the format of content (IMS Content Packaging), the metadata that describes this content (IEEE Learning Object Metadata), and the order in which content should be presented (IMS Learning Design).¹⁵ All significant learning content management systems now adhere to these standards (or derivatives thereof, such as SCORM). These standards focus on interoperability issues of online learning content, and provide some minimal learner modelling and adaptive sequencing opportunities. A more full description of this area can be found in this authors' previous work [20].

Comparing the adaptation capabilities of adaptive hypermedia systems versus learning content management systems is similar to the discussion of how intelligent tutoring systems differ from learning content management systems; adaptive hypermedia systems are largely small scale systems that have a high specificity to some context (domain, particular pedagogical approach, etc.) while learning content management systems are widely deployed across contexts but provide minimal flexibility or adaptation. In both cases, significant *a priori* work is required to encode instructional information into the learning environment.

One technique of content sequencing and adaptation that reduces the need for extensive knowledge engineering is *collaborative filtering* such as in [125]. This technique is used heavily in product selection for e-commerce, the quintessential examples being Amazon book recommendations and Netflix movie recommendations. In this approach learners are suggested content that was consumed by peers they are most similar with. Several challenges exist with collaborative filtering as an adaptive tools for e-learning: First, collaborative filtering algorithms must determine which learning artifacts and learner behaviours adaptation takes place on. There is minimal exploration capability. Second, the instructional expert is largely removed

¹⁵See [103] for an overview of these various e-learning standards.

from the system and unable to correct poor recommendations. For instance, if a common negative behaviour pattern manifests itself (such as the misidentification of a core concept and the reading of content unrelated to the problem) this will be reinforced by recommending non-optimal content to other learners. Third, collaborative filtering techniques only help to identify what content should be recommended, not when it should be recommended. Thus, techniques to identify when an adaptation should take place are required as well. Finally, collaborative filtering is susceptible to the *cold start* problem, where a lack of historical information early on leads to poor or incomplete recommendations.

Making content navigation more adaptable has also been well explored. Common methods of doing this include *faceted browsing* and *pivot browsing*. In the former, content is tailored through a set of filtering criteria the user provides, and results of searches or navigation indices are pruned accordingly. Pivot browsing is similar, but the list of criteria on which to filter is made up of a set of folksonometric tags that users have put on content. When viewing a list of content that is tagged with multiple values (for instance, one that is tagged with the values “social”, “constructivism”, and “newton’s method”), the learner can *pivot* their view of recommendations by clicking on a particular tag (for instance, to see more pieces of content about newton’s method, they may click on the last tag). These forms of adaptation are almost entirely end-user driven, though one could imagine hybrid systems that use both collaborative filtering as well as faceted browsing or pivot browsing to form accurate yet flexible navigation for learners.

2.3.2 Expertise Recommendation

Recommending experts or peer learners to interact with has also been explored in the area of artificial intelligence in education and adaptive hypermedia. Notably, the I-Help system [28, 117] was designed to model learners and recommend to them peers with similar interests. Recommendation was user-driven, and a variety of social and economic mechanisms were employed to encourage participation. Learners could chat with one another in real-time, and leave feedback about the expertise and helpfulness of their peer when help sessions ended. The I-Help system then included this feedback in future recommendations.

Cristea and Ghali [40] suggest a system that hybridizes expertise and content recommendation into one tool, the My Online Teacher (MOT) system. Similar to the I-Help system, the intent of the recommendation is to find a peer who can act as a tutor for the learner seeking help. MOT also includes a domain model, unlike I-Help, and uses it to identify social groups that might be appropriate for the learner, as well pieces of content. Formal evaluations, such as learner results on quiz questions, are included in the MOT pedagogical model, which allows it to “unlock” new content when a sufficient level of achievement has been reached.

2.3.3 Adaptive Testing

While not done within the same research communities as content and expertise recommendation, adaptive testing is a widely used technique amongst educational psychologists. In adaptive testing systems, the results of asking one question are used to determine which question to ask next. It is often used for physiological

or mental state testing (e.g., [64, 57]) in order to reduce the overall burden on the subject being tested. It can also be used across domains quite easily, and the results of adaptive testing have been shown to correlate with the results obtained from traditional testing [86].

Adaptation in adaptive testing is most often hidden from the end user, intending not to be of obvious benefit to themselves.

2.3.4 Assessment for Learning

Institutes of higher education typically use grades to assess student performance. These grades are ordinal in form, and are often expressed through numbers (e.g., 1 through 4, where 4 is a better grade), letters (e.g., A through D, where A is a better grade) or percentages (0–100%, where 100% is a better grade). Grades can be assigned to learners in many different ways, with two common methods being criterion assessment, where better grades indicate mastery of the subject, and normative assessment, where learners are ranked against one another through their grades. Grades are then used throughout the institution (and beyond) to determine the qualifications of individuals for various purposes, often relating to certification (e.g., being awarded a degree) or acceptance for studies (e.g., indicating that prerequisite courses have been completed satisfactorily).

The assessment for learning movement differentiates itself from this typical use of grades and instead emphasises the role assessment can have in enhancing the learning process. It has its roots in an extensive literature review conducted by Black and William [15] which conclude that using assessment mechanisms to aid learning instead of to evaluate learning can lead to sizable learning gains (an effect size between 0.4 and 0.7). The breadth of their review describes many different strategies used by teachers to incorporate assessment into the learning process and, from this, the Assessment Reform Group¹⁶ identifies the core of the area as “...the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there” [5]. In the same publication, this group identified ten key principles that should be followed in the classroom in order to make learning gains:

- Assessment for learning should be part of effective planning of teaching and learning
- Assessment for learning should focus on how students learn
- Assessment for learning should be recognised as central to classroom practice
- Assessment for learning should be regarded as a key professional skill for teachers
- Assessment for learning should be sensitive and constructive because any assessment has an emotional impact

¹⁶The Assessment Reform Group [6] was an informal group of 14 researchers who were involved in a number of British and European collaborative partnerships between 1989 and 2010.

- Assessment should take account of the importance of learner motivation
- Assessment for learning should promote commitment to learning goals and a shared understanding of the criteria by which they are assessed
- Learners should receive constructive guidance about how to improve
- Assessment for learning develops learners capacity for self-assessment so that they can become reflective and self-managing
- Assessment for learning should recognise the full range of achievements of all learners

Research on assessment for learning has focused activities instructors and learners can do to change the learning process by using assessment mechanisms. It seems reasonable for technology to contribute to this as well and, while the area of learner modelling has little mention of its compatibility with the assessment for learning field, many of the principles described here show up as explicit core activities of specific learner modelling paradigms. Scrutable and open learner modelling in particular focus on communicating to learners details of their progress and goals, focusing on building meta-cognitive awareness of learning activities. Similarly, some of the activities of intelligent tutoring systems can be considered as approaches compatible with assessment for learning, depending on whether feedback delivered aim to promote meta-cognitive skills. Methods from educational data mining are highly relevant as well, as long as they are used to inform instructional activities instead of just characterizing cohorts of learners. Finally, Shute et al. [109] describe an adaptive testing system that not only maintains a quality assessment of learning (e.g., for ranking learners), but also contributes to individual learning by tailoring user feedback.

2.3.5 Conclusions

Adaptable and adaptive learning environments are built from a broad range of techniques, and while content adaptation is most common, there are other aspects of the learning process that have been more adaptive. Broader still is the area of instructional interventions, which includes non-electronic adaptations such as question and answer sessions with an instructor or one-on-one interaction with tutors. The goal of this chapter of the work was to provide a brief overview of some of the more common adaptive methods used in current-generation intelligent learning environments.

The data-assisted approach aims to empower an instructional expert with knowledge of learners based on learner behaviours. The intent of this empowerment is that instructional interventions can be formed that are customized to the groups of learners that the instructor has discovered. This dissertation does not make the claim that such interventions must be electronic in nature, and Chapter 4 will provide some motivating examples of how the data-assisted approach can be used with traditional face-to-face forms of instruction. Nonetheless, the investigation provided in Chapters 5 through 7 apply the data-assisted approach, including instructional interventions in the form of customized navigation, in online educational environments.

CHAPTER 3

THE DATA-ASSISTED APPROACH TO SUPPORTING INSTRUCTIONAL INTERVENTIONS IN TECHNOLOGY ENHANCED LEARNING ENVIRONMENTS

The data-assisted approach is intended to generate *insight* which leads to and supports *instructional interventions*. It is different from traditional methods of building intelligent educational systems in that it explicitly acknowledges the role of instructional experts. It is in these experts that insight is generated, and from these experts that instructional interventions come.

Before talking about methods of generating insight or interventions, it is useful to talk about the experts themselves. Instructional experts for a particular environment can be many different people with different tasks. These experts can be instructors, instructional designers, content creators, tutors, markers, or peer learners. The experts can have a formal role in the course, or can be informal actors that engage with the environment through happenstance. There is already a body of work that examines how learners gain insights about their own actions (e.g., open or scrutable learner modelling), as well as how learners can adapt an environment to their own needs (e.g., adaptable systems). But how instructors tune learning environments through a sensemaking process has been largely unaddressed by the technology enhanced learning environment community.

The data-assisted approach starts with the collection of information about learners as they interact with tools in the learning environment. This information describes the behaviours of learners, and may be collected through a variety of means across a variety of different tools (see §3.1). The data must then be summarized and correlated to identify groups of learners based on some educational attributes, tasks, or goals.

The process of aggregating and correlating data with pedagogical goals requires that the instructional experts interact with computational elements of the system. A number of different computational approaches can be used, and this dissertation will demonstrate how two approaches in particular, *information visualization* and *unsupervised machine learning*, can be successful in identifying aggregations of learners. Regardless of the computational technique involved, it is the instructional expert who gives meaning to the groups, and identifies or parameterizes appropriate instructional interventions.

One of the differences between the data-assisted approach and traditional intelligent learning environments (e.g., intelligent tutoring systems) is that the interventions are largely based on the expertise of the

instructional expert and not on *a priori* domain, curriculum, or pedagogical knowledge that has been formalized and loaded into the system. The intelligence is in the overall system which includes the learners, the software, and the instructional expert, and not in software alone. This does not forbid intelligent software from interacting inside a data-assisted approach, but contextualizes the main focus of this work.

The nature of interventions can be varied, and this work will demonstrate both broad high-level interventions (e.g., the decision by an instructor to change assignment details) as well as narrow low-level interventions (e.g., the visibility of navigational cues shown to learners).

As the data-assisted approach relies upon the instructional expert to provide pedagogical knowledge, it is compatible with many learning theories and instructional design approaches. For instance, an instructor who observes a deficiency of learning in one set of learners might decide that a social constructivist approach is appropriate as a particular intervention, and perhaps configure an online learning environment to send learners to a blogging activity. Another instructor might, after seeing the same grouping of learners, modify the curriculum to include another assignment that focuses on experiential learning. Regardless, both instructors can use the data-assisted approach to get insights into students in their courses and to customize the course for those learners.

3.1 What is Meant by *Data* in the Data Assisted Approach?

As learners interact with educational environments, traces of their activities can be logged. These traces link actors (e.g., learners, instructors, instructional assistants) and artifacts (e.g., videos, tests, web pages) with interaction behaviours (e.g., watching, answering, clicking). Most educational environments already have some form of trace logging, although many don't separate functional traces (those that pertain to the correct operation of the system) from informational traces (those that are useful specifically for data analysis).

The granularity of the data collected is an important factor to consider when creating an educational environment. One approach is to capture low level interactions such as key presses or mouse movements (sometimes referred to as clickstreams [88, 101]), while another is to use more coarse-grained events such as whether or not an answer to a question was correct. The choice between a fine or coarse level of granularity affects the types of behaviours that are available for analysis of learners. A fine level of granularity (e.g., individual key presses) may require the end-user to aggregate data in order to form meaningful insights that they believe are useful (e.g., concepts in a course), while a coarse level of granularity (e.g., a grade on an online examination) may not be decomposable in order to relate to the same meaningful insight. A number of pragmatic concerns around the capture and storage of user data also exists – collecting mouse pointer data as a user interacts with an online content management system, for instance, would create a significant amount of data that would need to be transferred back to a central location for storage. In addition, this level of trace data would almost certainly need to be summarized in order to be useful as an attribute in either automated reasoning or visualization techniques.

A number of authors have considered how semantics can be added to learner trace data to make analysis easier, including [93, 23]. Building tools to support the data-assisted approach requires a consideration of how data is collected and labelled such that it is meaningful, but this activity takes place outside of the adaptation process, and is largely one of traditional knowledge engineering and system design.

3.2 What Does it Mean to *Assist* in the Data Assisted Approach?

The summarization of learner traces into meaningful insights can be considered a form of learner modelling, where the underlying data for the modelling process comes from the interactions learners have had with the learning environment, and the model itself is represented by the insights that are generated from this data. An important differentiator between the data-assisted approach and other forms of learner modelling is that in the data-assisted approach the instructional expert is considered the key actor in the modelling processes. Instead of loading instructional intelligence into the software *a priori*, there is a reliance on the instructional expert to form hypotheses of learning activity, validate the pedagogical relevance of patterns, and form instructional interventions as appropriate. Thus, software to support data-assisted investigation must collect data about learners, allow instructors to parameterize the analysis of this data as needed, summarize and present patterns of behaviours to the instructors, and map instructional interventions to groups of learners as instructed. The intelligence in such software may be quite limited depending on the data being collected and the mechanisms being used to present this to instructors; on one end of the spectrum, a data-assisted system may rely totally on information visualization techniques with minimal filtering to provide a summary to the instructor, while at the other end of the spectrum a data-assisted approach may provide sophisticated artificial intelligence techniques such as unsupervised machine learning mechanisms to aggregate learners into cliques. It is through the application of these techniques that assistance is given to instructional experts to complete the modelling process.

Embedding the instructor in the learner modelling process aims to increase the scalability of adaptive e-learning systems. While current intelligent learning environments scale well to many learners, they do not scale well between domains, and require a significant amount of domain and pedagogy modelling when being applied to new curricula. By helping instructors to form groups of learners based on behaviours, the data-assisted approach seeks to be a generalizable approach to building adaptive learning environments. Human intelligence in the form of the instructor is used to provide the labelling of groups and their relationships to adaptive components, while artificial intelligence and information visualization can be used to provide a statistical and graphical understanding of the relationships between observed behaviours and groups of learners.

The use of human intelligence in the data-assisted approach fits well with the notion of *sensemaking* as a principle goal of the field of learning analytics as described by Siemens [111]. Klein et al. [80] describe sensemaking as a process by which events can be understood with consideration of perspectives, which they

refer to as a *frames*:

“We can express frames in various meaningful forms, including stories, maps, organizational diagrams, or scripts, and can use them in subsequent and parallel processes. Even though frames define what count as data, they themselves actually shape the data (for example, a house fire will be perceived differently by the homeowner, the firefighters, and the arson investigators). Furthermore, frames change as we acquire data. In other words, this is a two way street: Frames shape and define the relevant data, and data mandate that frames change in nontrivial ways.”
(Page 88 of [80])

The process of the instructional expert interacting with the data collection and analysis aspects of the learning environment can be thought of as a dialogue; as the instructional expert elicits the formation of groups of learners from the system, he or she can form new hypotheses as to the state of learning happening in each group, and modify how course attributes such as content, sequencing, activities, or tools are made available. The dialogue is bi-directional – the instructor uses the environment to better understand learners, and shares this understanding of learners with the system through labelling. The system takes these labels, and the associated adaptations provided, and applies them to learners who fall within the groups based on learner behaviours.

The dialogue between system and instructional expert is also an ongoing one. As learners continue to interact, or as new learners begin to interact, with the tools in a learning environment, new data and new behaviours can be presented to the instructional expert. This expert creates, deletes, and modifies instructional interventions as they apply to learning objectives and pedagogical approach. Thus, the modelling process is intended to fit within the day-to-day activities of an instructional expert such as a course instructor, and not be an activity that takes place before learners are present, as much traditional instructional design activity does. This is not to say there is no place for *a priori* design in electronic learning environments or that, when available, resources to build intelligent tutoring systems should not be utilized. Instead, the suggestion here is that the data-assisted approach is a complementary method to both instructional design activities and intelligent tutoring methods, and requires a different kind of resource that may be more readily available in the context of higher education (e.g., the instructional expert).

3.3 Contrasting with the Ecological Approach

The data-assisted approach most closely resembles elements of McCalla’s ecological approach [88] described in §2.2.4. There are several aspects of the two approaches that are different. First, and most importantly, the ecological approach does not explicitly position instructional experts as a key actor in the sensemaking process. Instead, the emphasis is on automated “data-mining algorithms” interacting with information. The data-assisted approach, however, is focused on leveraging instructional experts in the adaptation process in order to gain scalability between domains. The vision is to engage all forms of instruction expertise, including not only the instructor, but teams of persons such as tutorial assistants, instructors, instructional designers, content experts, or even the students themselves if the resources are available.

An important aspect of the ecological approach is the notion of *purpose-based* adaptation:

“That is, contextual information such as the purpose, the particular learner(s) involved, the application goals, etc., determine how (and even whether) the information in the learner models and learner model instances is used. The choice of clustering algorithm and/or data mining algorithm, and the particular constraints put on each such algorithm is highly contextualized. Many other algorithms are also similarly contextualized, such as the experiential summarization algorithms in the recommender system example. Thus, much research is needed into what algorithms work, where, and for what purposes.” (Page 10 of [88])

In the data-assisted approach the purpose for which adaptation is taking place is held completely within the instructional expert. The software system never needs to know the semantics of the purpose, and the amount of explicit pedagogical modelling that needs to happen is diminished. The model of the learner that results from the software/expert dialogue is the only formalized evidence of a particular purpose. While the ecological approach calls for explicit description of purposes to determine which elements of the learner model should be used, the data-assisted approach integrates the purpose for the learner model within the model itself through the activity of the instructional expert during the creation of insights. New purposes require new consideration of the data by instructional experts, and thus generate new learner models.

Where the two approaches are similar is most readily seen in the use of interaction data. Both the ecological and data-assisted approaches use the learner interactions within the learning environment as the principal artifact around which modelling and adaptation takes place. They are both approaches that focus on modelling learners *in situ* instead of *a priori*. The on-demand nature of these perspectives leads to a more data-driven process of creating learner models which has implications for the scalability of software solutions. Learning environments which must make customizations of the environment to a large number of learners within a single domain may be better served through traditional learner modelling techniques. Learning environments which must make customizations across a variety of domains may find scalability gains with the ecological or data-assisted approaches.

CHAPTER 4

REALIZING THE DATA-ASSISTED APPROACH

The goal of this chapter of the dissertation is to elaborate on how the data-assisted approach could be used in real-world contexts. This chapter will follow the investigations and decisions instructional experts might make as they use the data-assisted approach throughout an academic term. Each of the scenarios presented explores what tool support for the data-assisted approach might look like through the use of low fidelity mock-ups, and a number of different parameters affecting the size of the student cohort, domain, pedagogical approach, and instructional expert role are considered.

The first of these scenarios (§4.1) will consider an introductory Computer Science course that is taught completely online and at a distance to a small cohort. This course makes use of a number of educational support tools, the most important of which is a learning content management system that has an asynchronous discussion forum. In this scenario, the instructor adapts her pedagogical approach based on visualization tools that describe the hidden traces students leave behind when they interact with the learning environment.

The second scenario (§4.2) describes the perspective of an instructional designer as she supports a second year Chemistry course. The course is taught by multiple instructors, and one of the instructors has elected to have his lectures recorded and made available to learners regardless of the section they are in. In this scenario, the instructional designer will be focused on both the immediate support of the learners as well as the longer term trends of how learners use technology to support their learning.

The final scenario (§4.3) will follow an educational technologist as he seeks to build adaptivity features into an educational environment. Based upon experiences with lecture capture tools, the technologist believes that different learners approach navigation in video differently. Using the data-assisted approach, the technologist will test this hypothesis and then parametrize an adaptive learning environment based on groups of similar learners.

The scenarios presented here are fabricated but rooted in actual investigations. Each will be paired with a full chapter of work (chapters 5 through 7) that provides a more detailed investigation into one or more of the success conditions for that scenario. The emphasis of these chapters is not on completely satisfying the requirements of the scenario, but in providing evidence which demonstrates that the data-assisted approach is a satisfactory method of solving a particular problem.

4.1 Scenario One: Visualizing Community Interactions

Katheryn is a faculty member who regularly teaches introductory Computer Science for non-majors. This course is typically moderate in size (50-100 students) and made up of learners from a wide variety of disciplines. This year, Katheryn is teaching the course in an online capacity instead of in a traditional lecture format. Learners have access to an online content management system which includes sets of web pages describing content as well as an asynchronous discussion forum. There are 20 learners enrolled, and because of the distance component the majority of evaluation is weighted on the final examination and assignments.

The content for the course is broad in nature, and Katheryn feels that encouraging group discussion will be one key to keeping learners engaged. She is concerned that the distance modality of the course will cause learners to shy away from interpersonal interactions, and scaffolds this aspect of the course by making weekly reading assignments. Each week, she will post questions to the discussion forums and learners will have to reply with their thoughts on the issues. By having learners read one another's posts, she hopes they will form a shared sense of community, resulting in greater engagement and deeper learning.

In a traditional discussion forum system, Katheryn can see only that students have written messages, as well as the content of those messages. All other interactions, such as traces related to the reading of postings, are inaccessible to her. With tools built for using the data-assisted approach, Katheryn can understand more deeply how learners are collaborating by showing these hidden interactions. Consider a mockup of a visualization tool for discussion forums showing in figure 4.1. This mockup categorizes forum participants into three different levels of engagement; those who have not used the system but were able to, those who only read messages but do not post (*lurkers*), and those who interact with one another by posting messages. The tool differentiates learners from instructional experts (which might include tutorial assistants or lab assistants) using lighter or darker colours for nodes. Katheryn can click on individual learners of interest in the course to find out more about them, including a list of topics they have posted in the discussion forum.

An important aspect of the data-assisted approach is that the underlying data is summarized as it is presented to the instructional expert, and the expert is actively involved in the summarization process. The intelligence in the system is thus a *mediated* and *emergent* property; it exists because of the interaction between the domain and pedagogy expert and the data-rich learning environment. In this scenario, Katheryn has some specific thoughts as to what active engagement means; i.e. that learners are reading and considering one another's messages. Katheryn can aid in the summary of the data by modifying the visualization to render traces differently based on her goals, the learning environment provides a clearer understanding of collaboration between learners in her course.

Figure 4.2 shows how Katheryn might explore the sociogram-based visualization. Katheryn chooses the *customize* tab in the right hand pane, and creates a new rule (bottom right). In this mock up, rules are described in equation 4.1. The set of objects includes nodes and edges, the set of attributes depends on whether a node or edge is selected, and includes items such as colour, line width, outline form, or

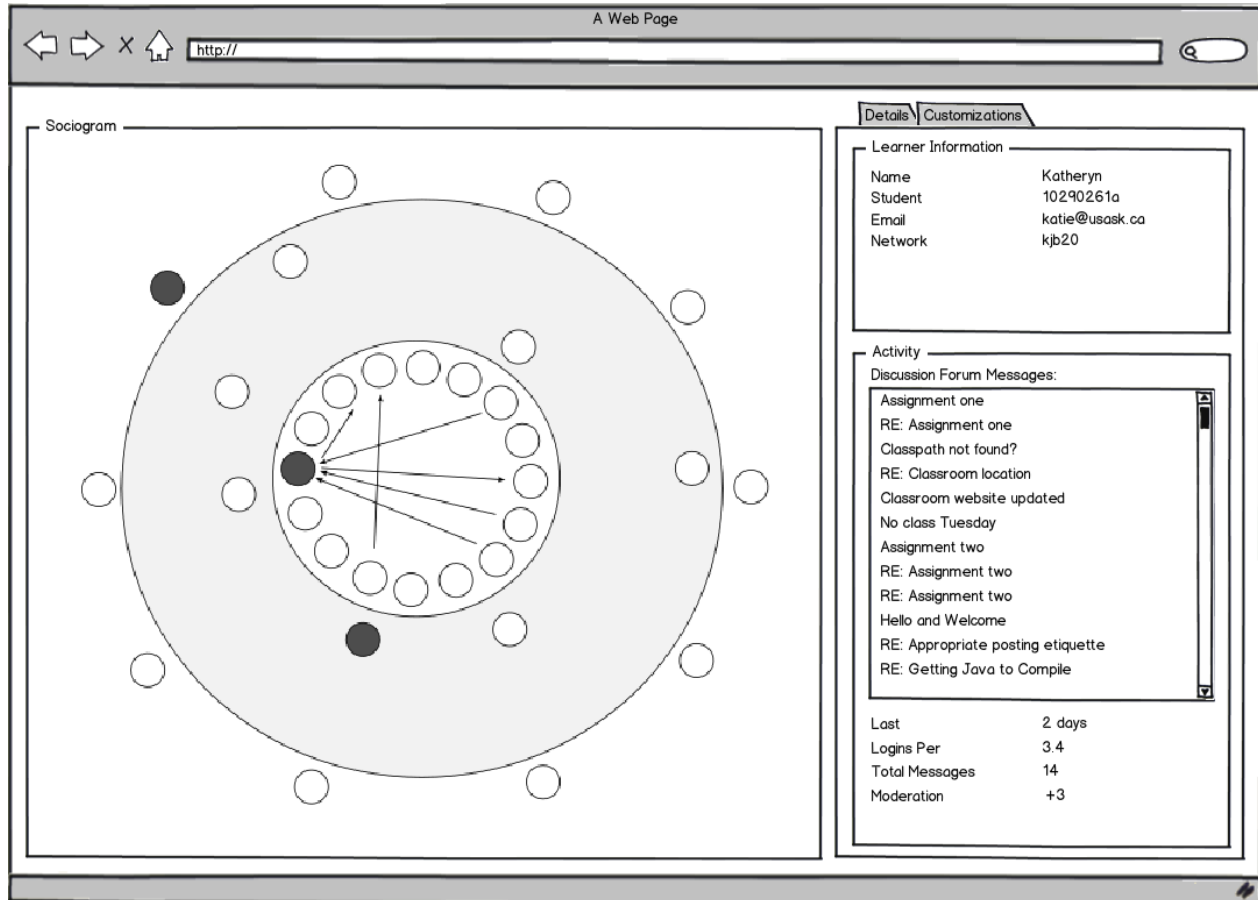


Figure 4.1: A mock up of a discussion forum information visualization system. The sociogram on the left is made up of three concentric rings of users: the outermost ring holds those users who have not used the system, the middle ring holds those users who have only read forums posts but not written any (*lurkers*), while the inner ring contains users who have written discussion forum posts. The ring containing lurkers puts users closer to the centre of the sociogram if they have read many forum postings, and puts users closer to the edge of the sociogram if they have read few forums posts. The inner ring connects users based on whether their post is a reply or a new thread. Instructional experts are shown using darker colours. Details about a particular user that might be select (e.g., the dark (red) node that represents Katheryn in the middle) are shown to the right of the sociogram.

directionality. Values depend on the object attribute, and include changing of colours, shapes, line forms, and other visual representations. Many values will be from closed vocabularies, but some might be free-form and allow Katheryn to use variables the system maintains. In this example, Katheryn wants to change the size of the node based on whether that person has had many of their messages read. The system might maintain a variable such as *others_read_number* to represent this, and the final rule would look like the one in equation 4.2. The complexity of the rule creation depends in part on the level of expressiveness the information visualization tool gives, and usability analysis would be required in order to build such a tool for a particular group of users.

$$\langle \text{rule} \rangle ::= \langle \text{object} \rangle \langle \text{attribute} \rangle \text{ is } \langle \text{value} \rangle \text{ if person } \langle \text{conditional} \rangle \quad (4.1)$$

$$\langle \text{rule} \rangle ::= \text{node size is } 2 * \text{others_read_number} \text{ if person true} \quad (4.2)$$

In the visualization environment, Katheryn has indicated that users who had written messages that have been read a lot should show up as bigger circles in the visualization, and those learners who have not logged into the system within the last two weeks should show up as dotted circles. By adding more of these graphical rules to the system, Katheryn is able to see that many of her learners are still quite active (very few dotted circles), but that only her posts (the dark grey circle) are read by learners in the class (as most of the learner circles are still very small, Figure 4.2 left hand side). This insight based on hidden traces suggests to Katheryn that she must do something else to increase collaboration between her learners. Katheryn decides to change the format of her weekly assignments; instead of just posting a synopsis of their thoughts on an issue, learners are also required to reply to at least one other student message. By saving her graphical rules and loading them at a later date, she can see the effects her changes have made on the classroom dynamics.

The data-assisted approach for supporting instructional interventions is made up of two activities: the summarization of usage data to generate insight, and the support for making instructional interventions based on this insight. This scenario has focused on the first of these, and demonstrates that by revealing to instructors the the hidden behaviours of learners (such as the reading of a message), an instructor can understand the affects of their pedagogical practice. The data-assisted approach does not require that instructional interventions take place directly in the technology enhanced learning environment; insights can be leveraged to form traditional instructional interventions, such as curriculum changes or changes to assignment activities (as Katheryn did). By providing mechanisms to compare visualizations over time, instructional experts can compare the effects of their actions and deepen their understanding of both the learning cohort as well as their pedagogical practice.

A number of questions remain as to whether such an insight is truly practical: is it possible to augment a discussion forum system to capture the hidden traces that were described in this scenario? Can an instructor

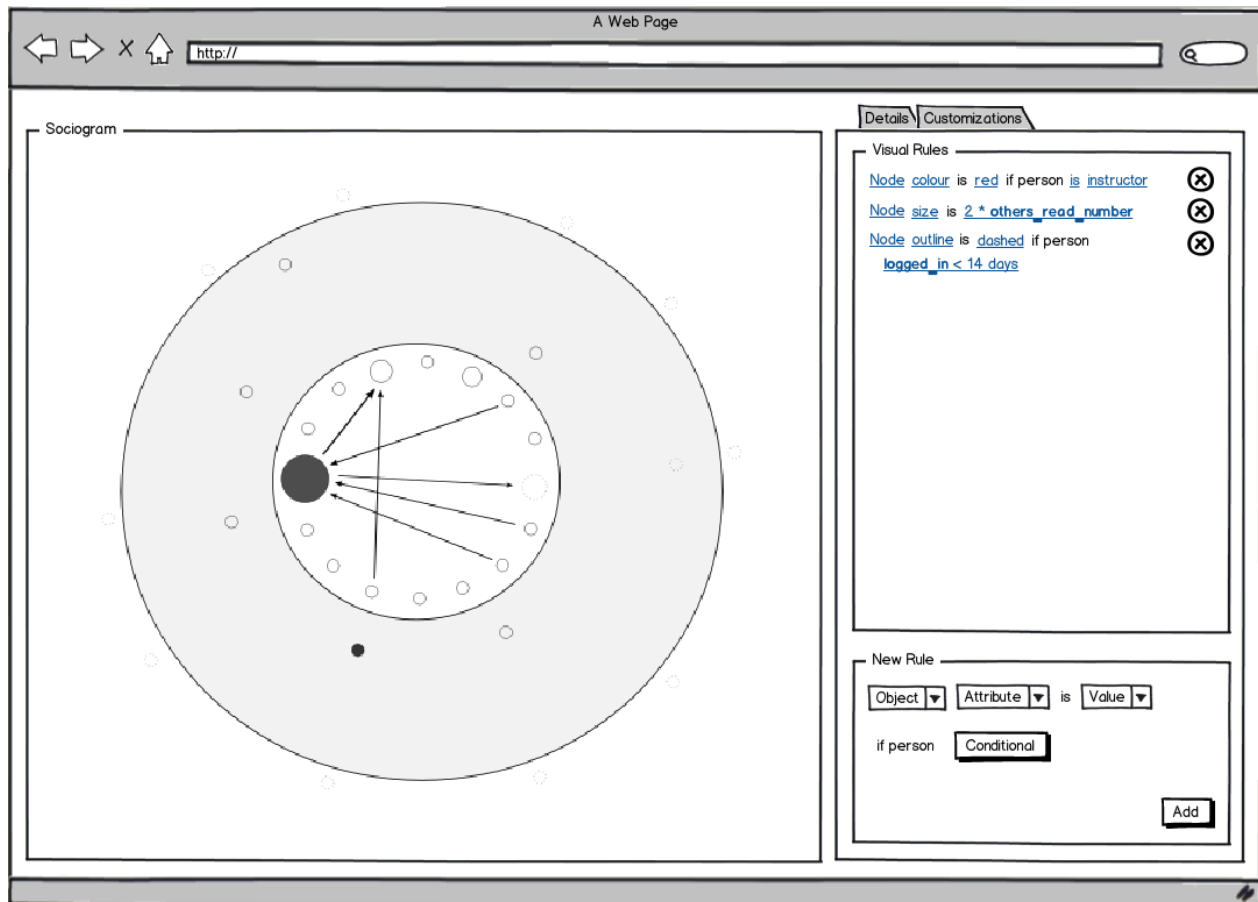


Figure 4.2: A mock up of a discussion forum system similar to the one in figure 4.1. In this system, the circles representing users are sized depending upon how often a users forum postings are read. Large nodes represent users who are read by many other users, while small nodes represent users who are read by few users. In this example, only the one instructional expert node is large, suggesting that only that users forum postings are of interest to the community. Nodes that have a dotted outline represent users who have not logged into the system in the last two weeks. Customization of the sociogram is done via rules which are shown at the top right. In this case, there are three rules defined which control node colour, size, and line outline. Hyperlinks allow the instructional expert to modify rules, while new rules can be defined using the form controls at the bottom right portion of the application.

derive enough meaning from a visualization that they are able to modify their pedagogical practice? These questions will be considered in Chapter 5 in more depth by exploring a visualization system similar to the one discussed in this scenario.

4.2 Scenario Two: Measuring Educational Impact

Michelle is an instructional designer supporting a second year undergraduate Chemistry course. This course serves as both a service course for other colleges (Agriculture, Engineering, Medicine) as well as for the core Chemistry program. The course is taught by five different instructors with a common curriculum and common set of examinations. The course reaches a total of six hundred learners in a given semester. One instructor has agreed to have his lectures recorded using a lecture capture and playback system. These recordings are then made available (asynchronously) to all students enrolled in the course.

Michelle's responsibility with the course is to manage both the short-term and long-term instructional design. The lecture recording system is new this year, and Michelle is interested in better understanding how learners use recorded lectures and the effect they have on performance. In particular, Michelle is interested in understanding whether the students who use the lecture capture system incorporate it regularly into their study habits, or whether they just use it only during the examination period. If there is a difference in performance between these groups she will adjust the course content as appropriate, guiding the learners to more effective study habits.

The data-assisted approach is ideal for Michelle as she is interested in both exploring the effect of introducing a new tool as well as tuning the course to take best advantage of that tool. Michelle has deep knowledge of different instructional interventions, but lacks an awareness of the the underlying activities of learners on which to base these. In a traditional classroom Michelle could go to lectures and observe attendance, a time consuming process. Further, the lack of time shifting tools in a traditional setting (such as lecture recording) can cause learners to rely on traditional methods for consuming content (e.g., notes or textbooks) which Michelle can not easily observe. With the data-assisted approach, learning tools are augmented with tracking features that log the activities of learners and make them available to experts like Michelle. While there may still exist modalities of learning outside these, the set of tools that collect learner traces along with methods of summarizing and acting upon these traces form the basis of the learning environment. Michelle is now able to watch in real-time as learners use the system, and leverage her pedagogical expertise to change the way in which the course is taught.

Michelle's interest in seeing groups of related students fits well with unsupervised machine learning techniques (*clustering*). Using an interface like the one shown in Figure 4.3, Michelle can choose the attributes to which she is interested in applying machine learning. In this case, she is interested in grouping learners by their weekly access patterns in the lecture capture system, so she selects the first nine weeks of accesses in the left hand pane. A number of other attributes from other tools are available (e.g., comments posted in

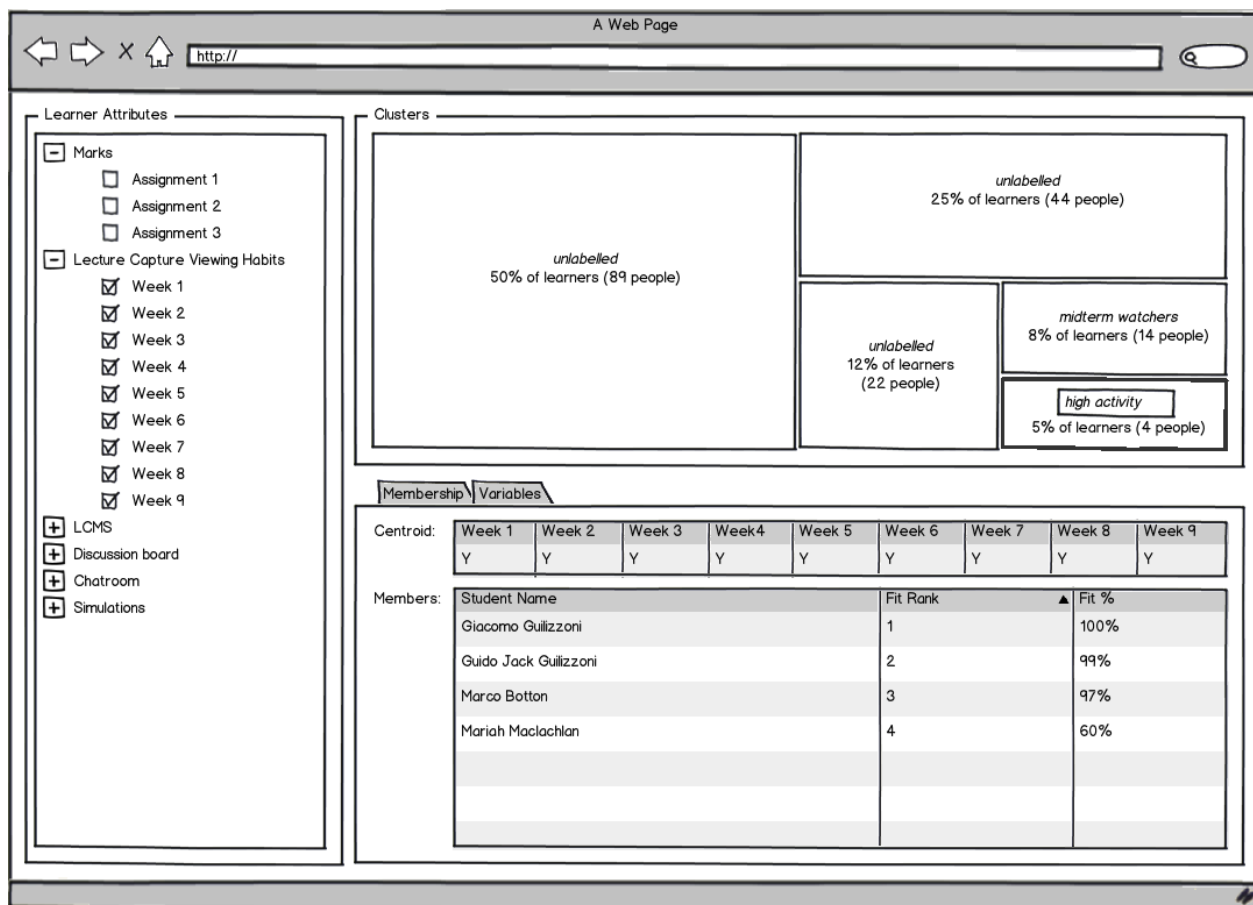


Figure 4.3: A mock up of a clustering-based learning analytics tool which allows instructional experts to explore the interactions learners have had within a learning environment. Instructional experts choose attributes to cluster on in the left hand window (in this case, lecture capture viewing habits), with the results of clustering being shown using a treemap of cluster sizes in the upper right. Each cluster can be labelled by the expert by clicking on the cluster, and details about the cluster are shown in the *membership* tab of the bottom window. In this case, five clusters were found ranging from four learners to 89 learners.

discussion boards, blog posts in the content management system, or interactive simulations finished), but for her current investigation they are unnecessary.

As attributes are selected, they are clustered by the system and the results are returned to Michelle. A treemap visualization (shown in the upper right of Figure 4.3) is particularly useful for clustering because it identifies not only the number of clusters found but also the relative sizes of clusters. Larger squares indicate a cluster with more learners, and the number of clusters returned can be controlled by the underlying clustering algorithm. In this example, every learner is placed into one cluster only, and a total of five clusters are returned by the clustering tool.

Michelle is able to explore the membership of a cluster and the relationship the cluster has with other variables interactively using the *explorer* in the bottom right pane. This pane shows the *centroid* of the cluster (sometimes called a *prototype*), which describes what the ideal learner in this cluster would look like.

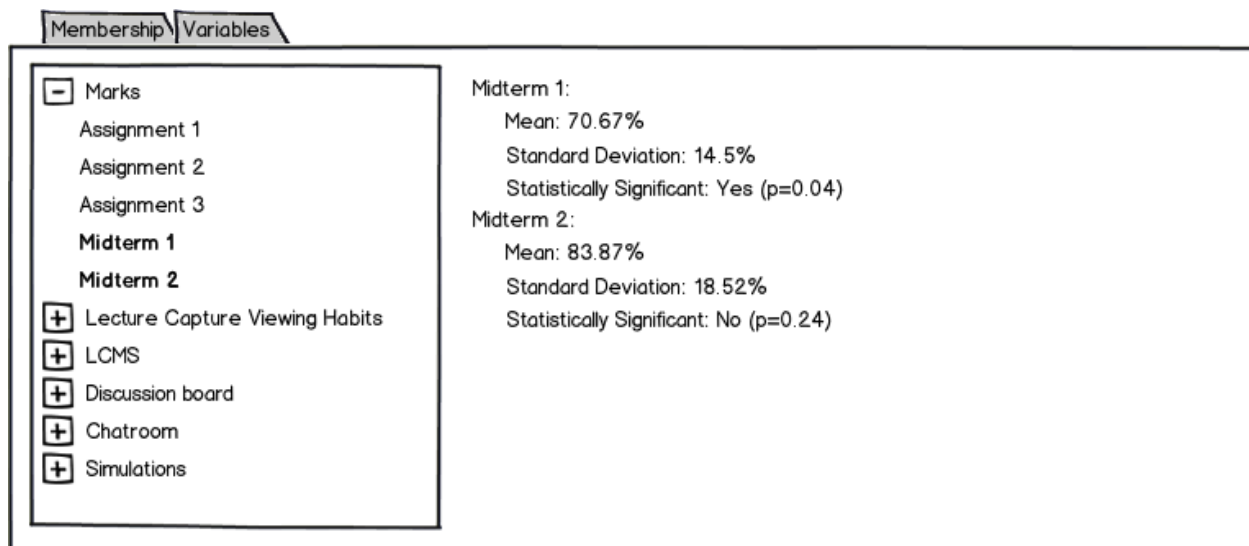


Figure 4.4: A mock up of the variables tab from the application shown in Figure 4.3. This tab allows the instruction expert to select those variables they are interested in correlating with the currently selected cluster. In this case, Michelle has selected to view the relationship between *Midterm 1* and *Midterm 2* variables for the current cluster.

The pane also includes a list of all of the learners who fit in the cluster, and details about how well they fit. From this, Michelle can determine how appropriate the clustering was, and consider the centroid of the cluster with respect to her pedagogical goals. Labels for clusters are not able to be determined automatically; they instead require that the instructional expert use their knowledge of the domain to describe the content of the cluster. In this example, Michelle has already labelled one cluster as *midterm watchers*, and is in the process of labelling another as *high activity*, since the centroid for the cluster involves the viewing of lecture at a weekly level. Once labelled, clusters can be used in automated interventions as will be described in more detail in the §4.3.

Michelle's goal in this scenario is to compare the results of clustering to other educational variables such as assessment. By clicking on the *variables* tab (Figure 4.4) in the cluster explorer, Michelle is able to see correlations between variables and the current cluster, as well as whether those correlations are statistically significant or not. Designing a tool to do this kind of analysis requires paying close attention to the roles and use cases of users. For instance, it is not unreasonable for Michelle to have some basic statistical knowledge, but if the tool were to be used by faculty or system designers, such an expectation might not be reasonable. To facilitate external statistical analysis export capabilities to encourage off-the-shelf tool use.

This scenario has demonstrated a number of attributes of teaching and learning in higher education that are well addressed by the data-assisted approach. First, courses are often large multi-section offerings made up of hundreds of learners where a visual in-person analysis of activity is either difficult or impossible. As learning opportunities increasingly happen outside of the classroom (e.g., through time shifting of lectures using lecture capture), the need to capture interaction between learners and technology grows. The data-

assisted approach is built on the premise that these interactions allow instructional experts to gain valuable insights into how learners interact with the educational environment.

Second, the data-assisted approach puts the instructional expert at the centre of the *sensemaking* process. In this scenario, it is Michelle who determines whether the clusters fit with her pedagogical approach and are relevant to her investigation. The educational environment helps aggregate and correlate behaviours, but it is the expert who validates the results and forms interventions. How these findings might be tied to interventions is largely omitted from this scenario, but will be described in the following section (§4.3).

A number of assumptions about technological approaches have been made to illustrate this scenario. In particular, it has been assumed that by clustering learners based on their viewing habits, pedagogically relevant groups can be discovered. Is this a reasonable assumption? If so, do these groups relate to formal assessment as Michelle was interested in, or are other variables, such as self-reported satisfaction of learners, a more useful indicator of learning? Is it reasonable to do this kind of investigation in mid-semester, as Michelle is attempting to, or is this form of clustering only useful in multi-year comparisons? Each of these questions will be addressed in section Chapter 6, where data from an actual lecture capture system is analysed.

4.3 Scenario Three: Adapting Learning Environments to Tasks

The previous scenarios have described how an instructor and an instructional designer might use data-assisted approach tools in their daily activities. As put forth in the previous chapter, intelligent learning environments typically offer automatic adaptivity in response to learner behaviours. The data-assisted approach is appropriate for configuring and building of these systems by bringing technologists into the sensemaking process. Whereas in a traditional intelligent learning environment the sensemaking is done *a priori* by technologists, a data-assisted intelligent learning environment encourages the creation of adaptive features, or the configuration of existing adaptive features, in response to new patterns of behaviours observed.

This scenario follows Adam, an educational technologist with a software engineering background, who is supporting a lecture capture system similar in nature to the one Michelle is using. Adam has received a lot of feedback from instructors that the navigation in the application is hard for students, especially since the principal method of navigation through the video is by selecting from a series of index thumbnails that are generated from the captured data projector every five minutes of recording (Figure 4.5). Some instructors want index thumbnails more often, while others want fewer thumbnails that are more like chapter markers in traditional DVD media.

Adam understands the perspectives the instructors have shared, and he believes that students navigate through the videos according to their needs. He examines the behaviours of learners using visualization tools for the lecture capture system that have similar capabilities as the ones Kathryn used in §4.1. Students rarely use the same indices when navigating, but he uses clustering tools (such as described in §4.2) to find three clusters of activity which he labels as *reviewing* for frequent use of many indices, *watching* for those

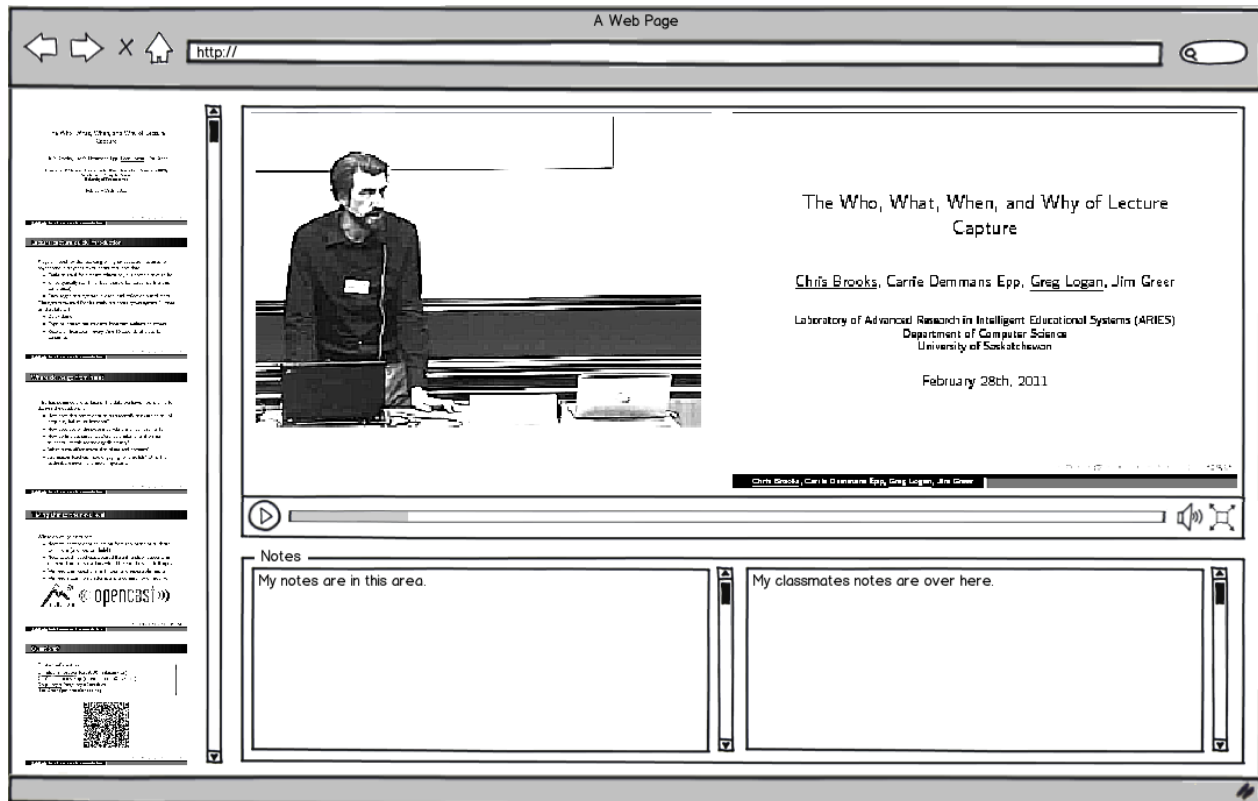


Figure 4.5: A mock up of a lecture capture environment. The images down the left hand side control navigation through the video, and allow a user to seek to a particular *chapter* directly. The video playback component on the upper right includes video of both the instructor and the data projector, as well as a traditional scrubber that allows for non-chapter seeking. Additional tools such as note taking components, discussion forums, or suggested readings are available underneath the video.

students who seem to only use a few indices that refer to significant breaks in the lecture, and *images* for a cluster of access that appears to focus on images in slide content (e.g., chemical drawings in material sciences courses, paintings shown in art history courses, or diagrams used in paediatric nursing courses).

Adam has identified that it is possible to parameterize the indexing functionality of the lecture capture system to present indexes in different ways. At a high level, the indexing functionality works by converting the lecture video to a set of images, then chooses an image to act as an index marker when content has changed significantly. Adam modifies this functionality by providing a new indexing algorithm that outputs three different sets of indices, one appropriate for each of the groups of learners he has identified. This algorithm works using supervised machine learning (*classification*). In short, each group of learners is represented by a set of index markers from their behaviour data, similar to the *centroid* described in the previous scenario. Adam *trains* the classification algorithms to provide customized results based on these centroids. The result of training is a set of *content models* (one for each of the three circumstances which are represented by centroids) which can be applied to new lecture videos in order to form sets of indices. Adam then programs the system to provide different groups of navigational indices to learners, with the default being those indices the learner has most often used. The end result is that learners are shown a new interface with three tabs on the left-hand side, one for each kind of indexing corresponding to the centroids Adam has discovered. Students are automatically shown the tab that their behaviour most closely represents, but are free to choose from the alternative indexing schemes as they see fit (Figure 4.6).

In this scenario, Adam has used the data-assisted approach to understand how learners are using the lecture recording tools. He has identified patterns of behaviour using tools similar to those used by Katheryn and Michelle. As shown in this section, the data-assisted approach does not have to end with just the detection of patterns. While Katheryn and Michelle were able to leverage insights to change their pedagogical approach and instructional designs respectively, Adam is making the results of the patterns available to the educational environment by labelling clusters. With labels attached to prototypical behaviours (the student *centroids*), learners can be classified into groups and the appropriate indices can be shown automatically.

This scenario prompts a number of questions both theoretical and pragmatic. On the theoretical side do groups of learners really agree on where an index should be placed, or are their preferences for navigational aides more varied? If the former is true, then Adam's approach may well yield a more personalized and efficient navigation structure, where the latter would suggest individualized navigation would be more appropriate. On the pragmatic side it is worth considering whether it is appropriate to use supervised machine learning to build indices, and how such an approach might compare to the algorithms that already exist. Supervised machine learning is strong technique when dealing with complex data which has a clear set of attributes. However, is it appropriate to deal with lecture videos in this manner? If so, can such a technique outperform hand-tuned algorithms which already exist? Each of these questions will be investigated in Chapter 7 through laboratory studies involving learners and actual recordings available from a production lecture capture system.

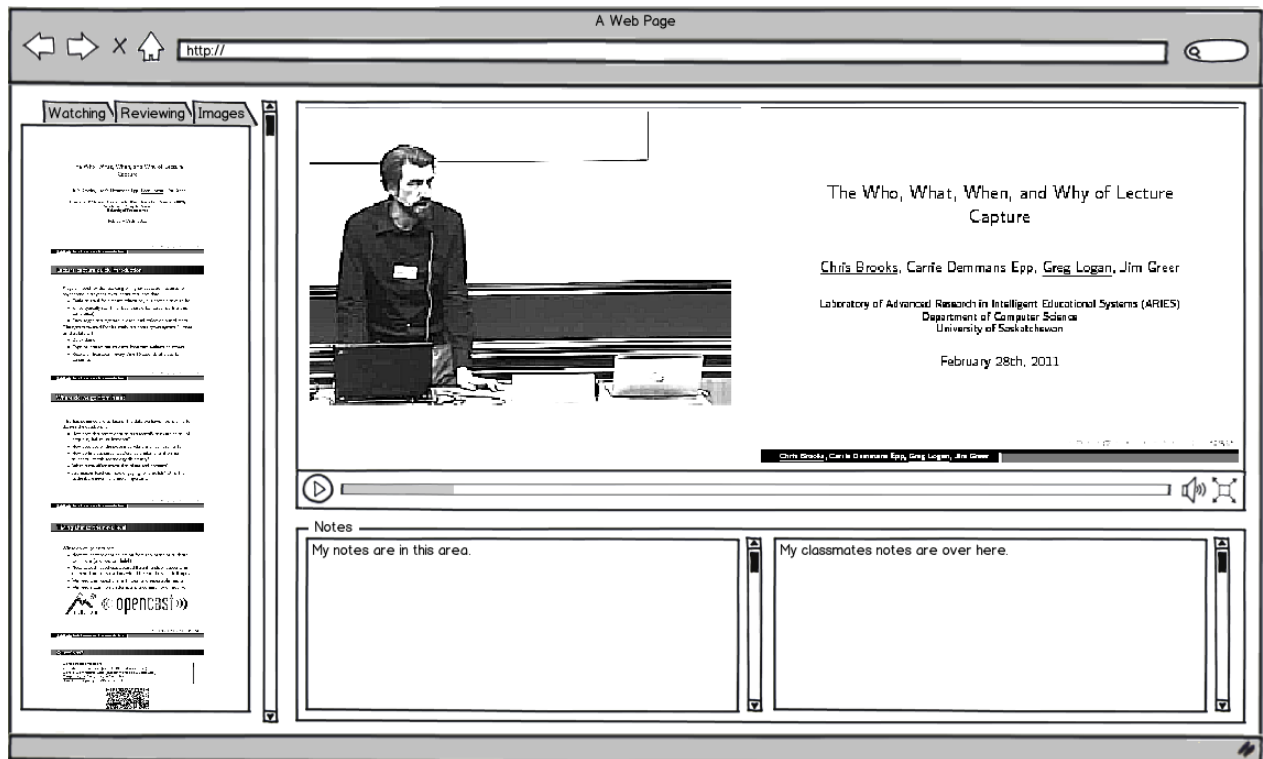


Figure 4.6: An adaptive lecture capture system mock up nearly identical to the one shown in figure 4.5. In this system, Adam has customized the navigational indices at the left hand side to be different depending upon the tab that is active. Each tab contains different indices as appropriate for tasks that Adam has trained for using *supervised machine learning*. Users can change tasks by selecting a tab; by default they are shown the tab that most closely matches their system behaviours.

4.4 Conclusion

This chapter has provided three scenarios designed to demonstrate how the *data-assisted approach* can be used in real world contexts. The scenarios do not fully describe the capabilities of the approach, but are instead intended to motivate how different instructional experts might interact with learning environments that collect end-user behaviour data. To this end, a number of different actors (instructors, instructional designers, and educational technologists) with different use cases (supporting learners, analyzing effectiveness of tools, and customizing the learning environment) using different instructional modalities (online distance learning and traditional face-to-face lecture learning) teaching to different sized class are considered.

Each of the scenarios has introduced questions as to the feasibility and utility of the data-assisted approach as a solution for further enhancing technology enhanced learning environments. While the scenarios have not been implemented in full, the key questions raised have been answered through laboratory experiments and wide scale field deployments. Chapters 5 through 7 take each of these scenarios presented here in turn, and address the key issues that have been raised.

CHAPTER 5

VISUALIZING COMMUNITY INTERACTIONS

The data-assisted approach for supporting instructional interventions is made up of two activities: using traces of learners to generate insight, and supporting instructional experts as they create instructional interventions from this insight.

One way the data-assisted approach is different from other methods of creating intelligent learning environments is in the way insight is generated. In the data-assisted approach, insight is created through an interaction between an instructional expert and data that represents a particular cohort of interest (a group of learners). A more traditional intelligent learning environment approach would be to fully describe the learning space before the system is deployed, then categorize learners and react accordingly. The data-assisted approach is more reactive, and focuses on *in situ* exploration of learner activities.

Similarly, instructional interventions can be constructed and delivered through multiple methods. A traditional intelligent learning environment such as an intelligent tutoring system or adaptive hypermedia system changes content delivered to a learner in reaction to their activity. These systems will not often react by changing pedagogical approach, however, unless they have been specifically pre-programed to do so. In contrast, human instructors often react by considering how they might restructure aspects of a course. By giving these experts control over how instructional interventions are formed, larger pedagogical changes can be made with minimal *a priori* consideration.

This chapter explores how the traces learners leave as they interact with the learning environment can be made available to instructors using information visualization techniques. The scenario presented in §4.1 raised two questions which will be considered in more depth, namely:

1. Is it possible to augment a discussion forum system to capture the hidden traces that were described in this scenario?
2. Can an instructor derive enough meaning from a visualization that they are able to modify or improve upon their pedagogical practice?

These questions will be addressed in §5.1 and §5.2 respectively.

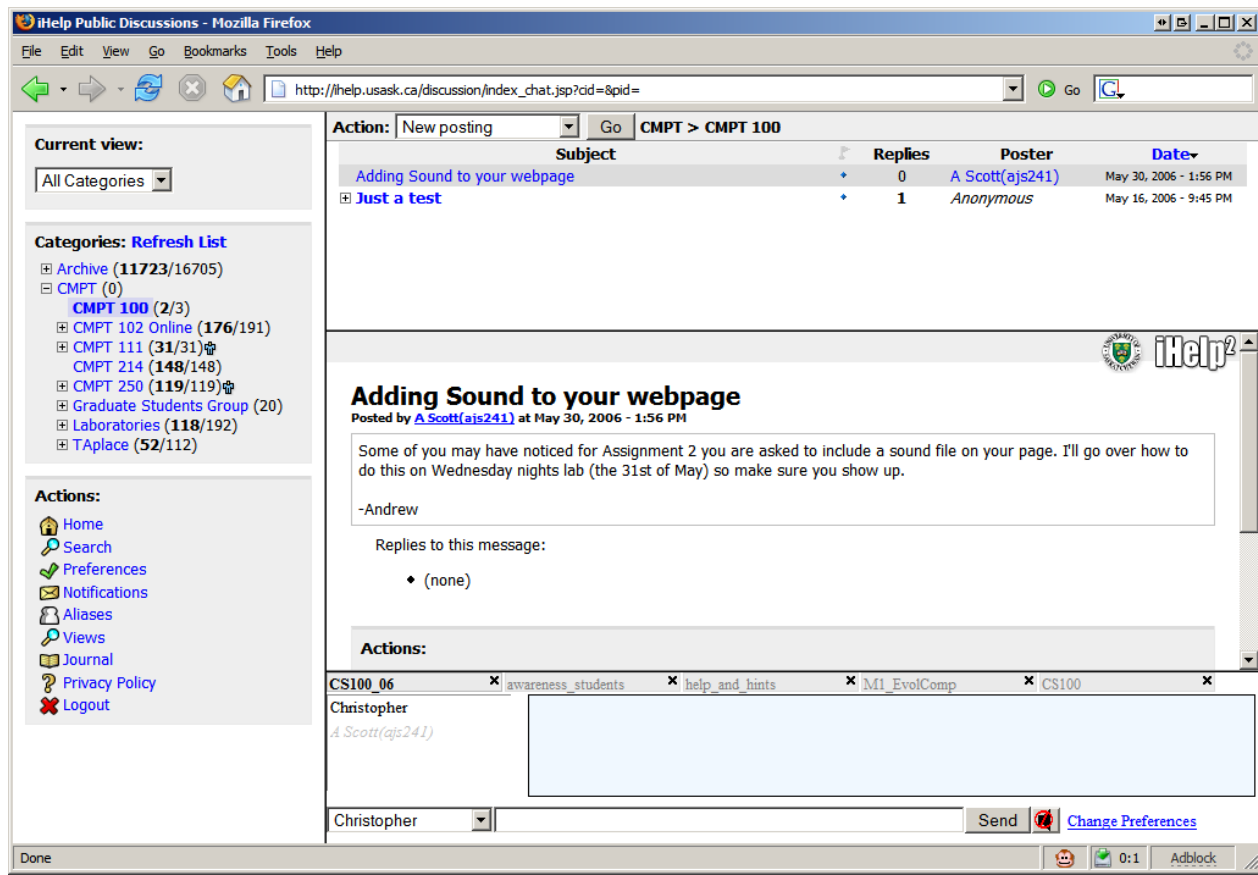


Figure 5.1: The iHelp Discussion forum system c.2006. The left side shows the different forums that a user has access to, in this case a number of course related forums as well as social and employment forums. Forums are heirarchical, and may include sub forums related to special topics. Each forum is made up of heirarchical *threads* of discussion as shown in the top window. Content for a given *posting* in a thread is shown in the middle of the application, and other e-learning tools (such as a synchronous chat system) can be shown at the bottom of the application.

5.1 Augmenting an Asynchronous Discussion Forum with Visualizations

5.1.1 iHelp Discussions

The iHelp Discussions learning environment (Figure 5.1) is a Web 2.0 asynchronous discussion forum intended for use within higher education. It was developed out of the I-Help research project ([62]) with the goal of increasing usability and scalability of a technology enhanced learning environment supporting peer help. The system was deployed within the Department of Computer Science from 2004 – 2010, and was used by thousands of students in dozens of courses annually. When it was retired from general use by the Department in 2010, it contained over 75,000 messages in over 2,200 forums created by over 3,300 users.

This environment is different from other web discussion forums in that it has been augmented to record

learner interactions at a fine-grained level. Where other systems typically would show all of the messages within a thread at one time, iHelp Discussions has nested lists of messages which allow the system to record when a user requests a message and how long the user stays with the message open. Further, the forums are heirarchical in nature, and access to forums are logged using similar mechanisms. Over the six years it was deployed, more than 3,000,000 read requests were issued by users, with the top message being read over 1,700 times.

The use of iHelp Discussions as a forum ranged broadly by instructor and course. It was commonly used in large cohort undergraduate courses, and there were several department-wide “off topic” forums that students used to discuss issues of technology, philosophy, and politics (amongst others). Access to the forums was restricted through an institutional username and passwords, and public access to messages was not available. In some forums learners were permitted to post anonymously, though the back end system is still able to disambiguate usernames if needed. Instructors were free to structure sub-forums however they wanted, and it was common for topic-based, course structure-based, or a flat single forum environment to be used.

One unique aspect of the iHelp Discussions deployment situation is how access is granted to instructional experts. It was not uncommon for instructional experts such as instructors and tutorial/lab assistants to be routinely given access to all of the undergraduate forums. This, along with the detailed usage tracking, makes it possible to study the behaviour of instructional experts as well as learners

5.1.2 Visualization of Learner Activities

Sociograms are a common method of visualizing asynchronous discussions, and have been used by a number of researchers to visualize email correspondence in particular (e.g., [121]). These visualizations are graph-based structures where nodes represent individuals in the community and arcs between nodes represent the creation of replies to a message. Nodes are typically rendered using different sizes to represent the status of an individual in the community or the amount of discussion the individual has contributed (e.g., [121]). Nodes can be arranged in a number of ways: a strict hierarchy which outlines the abilities or status of groups of nodes is common, as is a physics-based “force graph” which moves nodes closer to one another depending on the characteristics they share.

A representation for the iHelp Discussion forums has been formulated using a sociogram where nodes indicate particular persons involved in a course with directed edges between those nodes indicate that a person has replied to another person. To clearly indicate the difference between learners and instructional experts (e.g., Instructors, Tutorial Assistants, and Markers), each node is colour coded to be either a learner (light grey), or an expert (red). As the iHelp Discussion forums are available to many instructional experts (faculty and assistants assigned to other courses) regardless of the content of the forum, there are typically many red circles in the sociogram.

In large courses (e.g., those with more than 100 students) this formulation quickly became unwieldy. To address this, individuals are further broken up into membership of one of three categories:

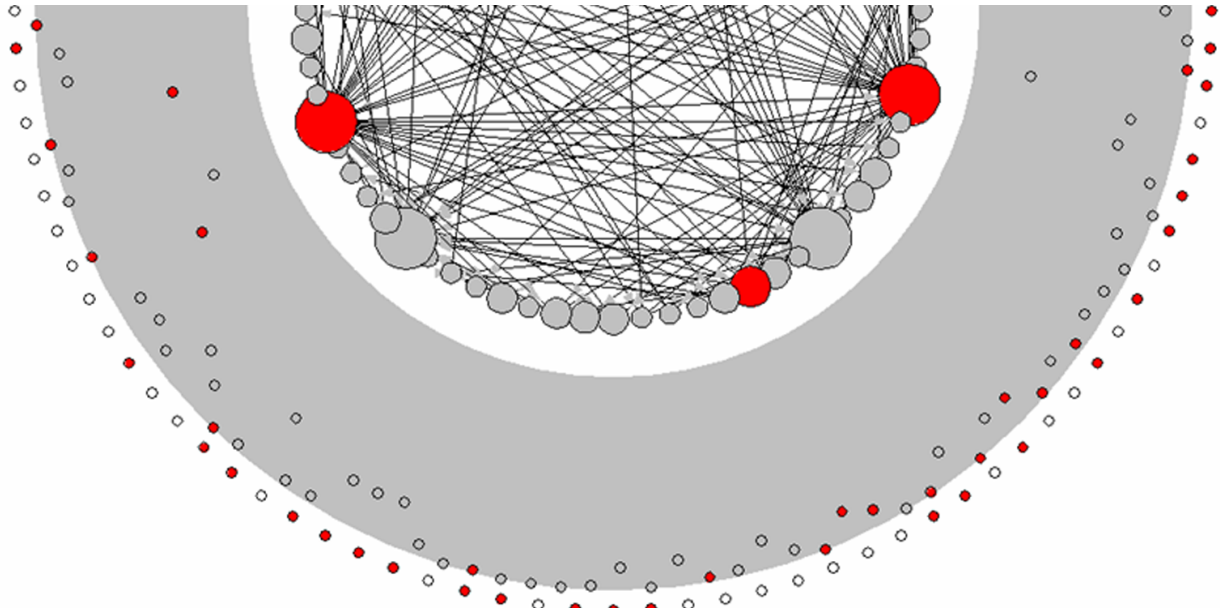


Figure 5.2: Portion of a sociogram from an introductory computer science course which was taught in a blended fashion and had approximately 200 students and contained 254 discussion postings. Darker (red) nodes indicate facilitators, while lighter (grey) coloured nodes indicate learners. The inner circle is made up of participants, four of which are very important to the community (as shown by having a larger node size). A casual observation of this network indicates that, while some learners write a fair bit (many interconnected nodes in the middle), there are lots of learners who haven't ever read anything (the outer ring of non-users), and many lurkers who read very little (as they tend to be closer to the outside of the middle ring instead of the inside of that ring). Note that the ring of non-users includes a disproportionately high number of facilitators as our currently deployment gives access to this forum to most staff and faculty in the department.

- Participants: Those individuals who have written messages, either on their own or as replies to other messages.
- Lurkers: Those individuals who have read postings but have not written any.
- Non-users: Those individuals who have never read nor written a posting.

Each category of users is put into their own sociogram that aligns nodes along the exterior of a circle. The different sociograms are then layered on top of one another such that the non-users are farthest from the centre of the sociogram, the participants are closest to the centre of the sociogram, and the lurkers were in between (Figure 5.2). This corresponds well both with the perceived participation rate of individuals (*participants* are more central than *lurkers* who in turn are more central than *non-users*), as well as with the sizes of the different categories of individuals (generally there are more *non-users* than there are *lurkers*, and more *lurkers* than there are *participants*).

Lurking is modelled as a continuous variable – one individual can lurk more than another by reading more forum postings. To support this in the visualization, lurking ratios are calculated and node distances are varied from the outer edge of the lurker region to the inner edge, where lurkers who are closer to the inner

edge of the sociogram have read more content. This further reinforces the idea that users who are central in the overall visualization are participating more in the course than learners that are close to the edge of the screen.

Initial feedback from instructors indicates that the act of posting a message does not mean that a user has contributed in a meaningful way. To represent a measure of *importance* in a course, the size of an individual node is varied by the number of persons who have read a users postings. The calculation for an individual's *importance* is given in equation 5.1.

$$importance = \frac{\text{number of people who read my postings}}{(\text{number of participants} + \text{number of lurkers}) * \text{number of postings}} \quad (5.1)$$

A number of observations about the sociograms can be made based on informal interactions with instructors who reviewed the prototype. In particular, it was observed that:

- A highly connected graph indicates that learners are communicating with one another, while a graph where many nodes are connected only to an instructional expert indicate little peer collaboration.
- The degree of edges coming out of expert (red) nodes indicates how much direct control an instructor has over conversations. Instructors who wait to answer questions have very few arrow heads pointing at their node, while instructors who provoke discussion have many arrow heads point at their node.
- Lurking rates are highly variable, and the majority of lurkers read fewer than 30% of the postings. This includes non-learners (e.g., tutorial assistants or other instructors), a result that surprised many of the instructors who saw the visualizations.

The next four sections provide specific examples of how this visualization has been used by instructors, and demonstrate the effect of making learner traces more readily available.

5.2 Effect of Visualization of Traces on Pedagogical Practice

This section of the work describes the different investigations that have been conducted in order to understand better how instructors can make pedagogical decisions based on visualizations of learner interactions. While the results here are largely anecdotal, they each describe how actual cohorts were visualized and the instructional interventions that were executed by instructors.

5.2.1 Case study: Online Small Cohort Course

Many instructors change their pedagogical approach when teaching in online environments, in order to match the needs of learners. In these situations, learners can more easily lose a sense of shared community or shared purpose as the activities and actions of their classmates may be hidden from them. The instructor feels this

change too, and many of the consequential awareness indicators that they would normally get from an in-person teaching environment (e.g., attendance and interaction in the lecture, or students coming to tutorials and office hours for assistance) are missing. By revealing indicators of activity in discussion forum systems, the data-assisted approach can help instructors to understand what the virtual classroom interaction looks like.

The Department of Computer Science offered an introductory course for non-majors on the topic of basic Computer Science history, principles, and techniques. In c2006 this course was taught simultaneously to a large cohort of over 100 learners in a blended manner, and to a small online group of 20 learners who had no face-to-face instruction. While taught by different instructors, the online instructor was also the creator of content, assignments, and examinations for both sections..

The online instructor was a firm believer in using the iHelp Discussion forums to build a sense of community amongst the students. Especially in the online course, she saw the discussion forums as the main method of engaging with learners. To this end, one of the weekly requirements was to write an online forum posting about a prescribed course topic and this requirement was mandatory for the online learners only. She would post the initial message indicating what the weekly topic was, and learners were expected to respond with details of content they found on the Internet. Her reasoning was that by sharing the results of this activity publicly, learners would form a sense of community with one another. A number of weeks into the course, the instructor was shown visualizations for both the in-class and online discussion forums (Figure 5.3).

The visualization was explained to her, and she showed particular interest to the way size of nodes was generated. She was bothered by the fact that she could easily identify herself as the large dark (red) node who was connected to most students. She wanted students to communicate and form a community with one another, and not just herself. That the pattern existed in both the online and the traditional cohort didn't surprise her, but she felt that traditional learners had other mechanisms by which they form a shared sense of community and thought that making the assignment mandatory for the online learners would be enough to encourage broader use of the discussion forums. In a subsequent offering of the course, the instructor changed the weekly assignments to be more problem solving based, and set the evaluation criteria such that learners would interact more. The result was a discussion graph that was more fully connected and where learner nodes varied in size (thus *importance*) as the instructor (Figure 5.4).

The use of the data-assisted approach provided insight that was otherwise lost to the instructor. By visualizing the hidden traces learners leave, the instructor became able to understand the effect of her pedagogy, and make interventions that were reflected through subsequent visualizations. Further, the instructor was able to compare the familiar traditional face-to-face cohort to a new mode of teaching, and create the kind of learning environment she thought was important for online learners.

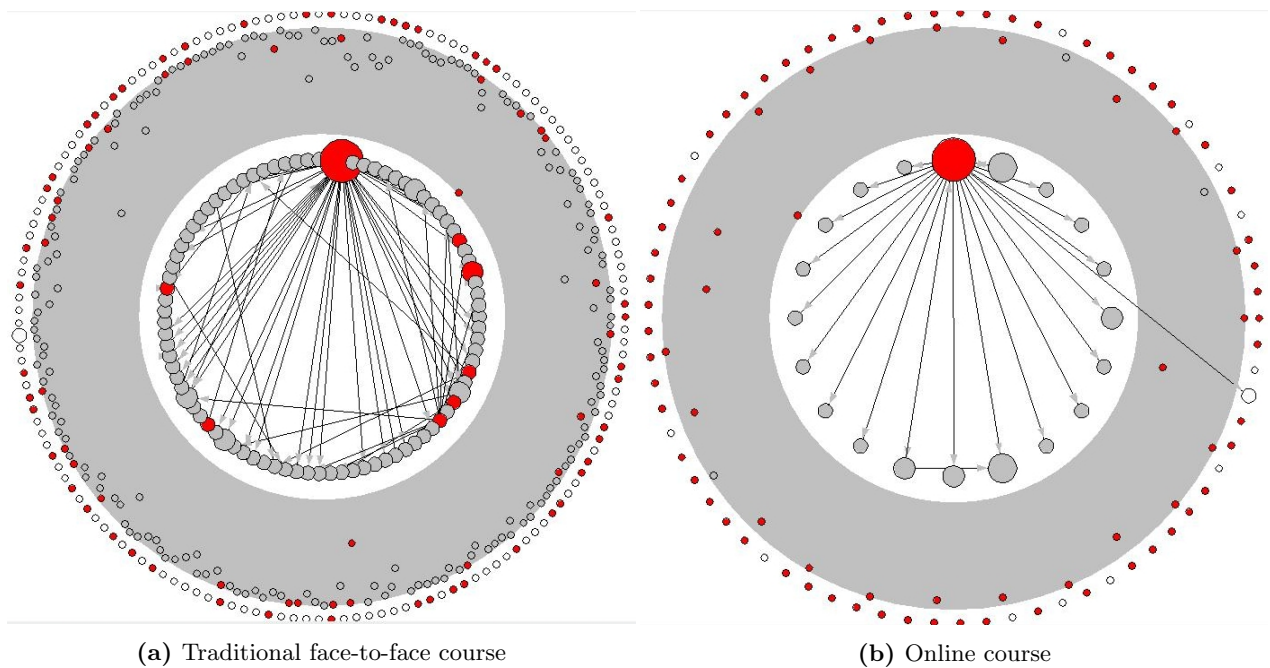


Figure 5.3: Visualizations of an introductory Computer Science course for non-majors. Figure 5.3a shows interactions amongst learners who attend traditional lectures, while figure 5.3b shows interactions amongst learners in a completely online environment. In the traditional course there are many lurkers who have read only a few messages, while the online course required participation of learners and thus has very few lurkers (only other instructional experts). In both cases the instructor is the principal actor in the social network (large dark (red) node), and learners rarely reply to anyone but the instructor.

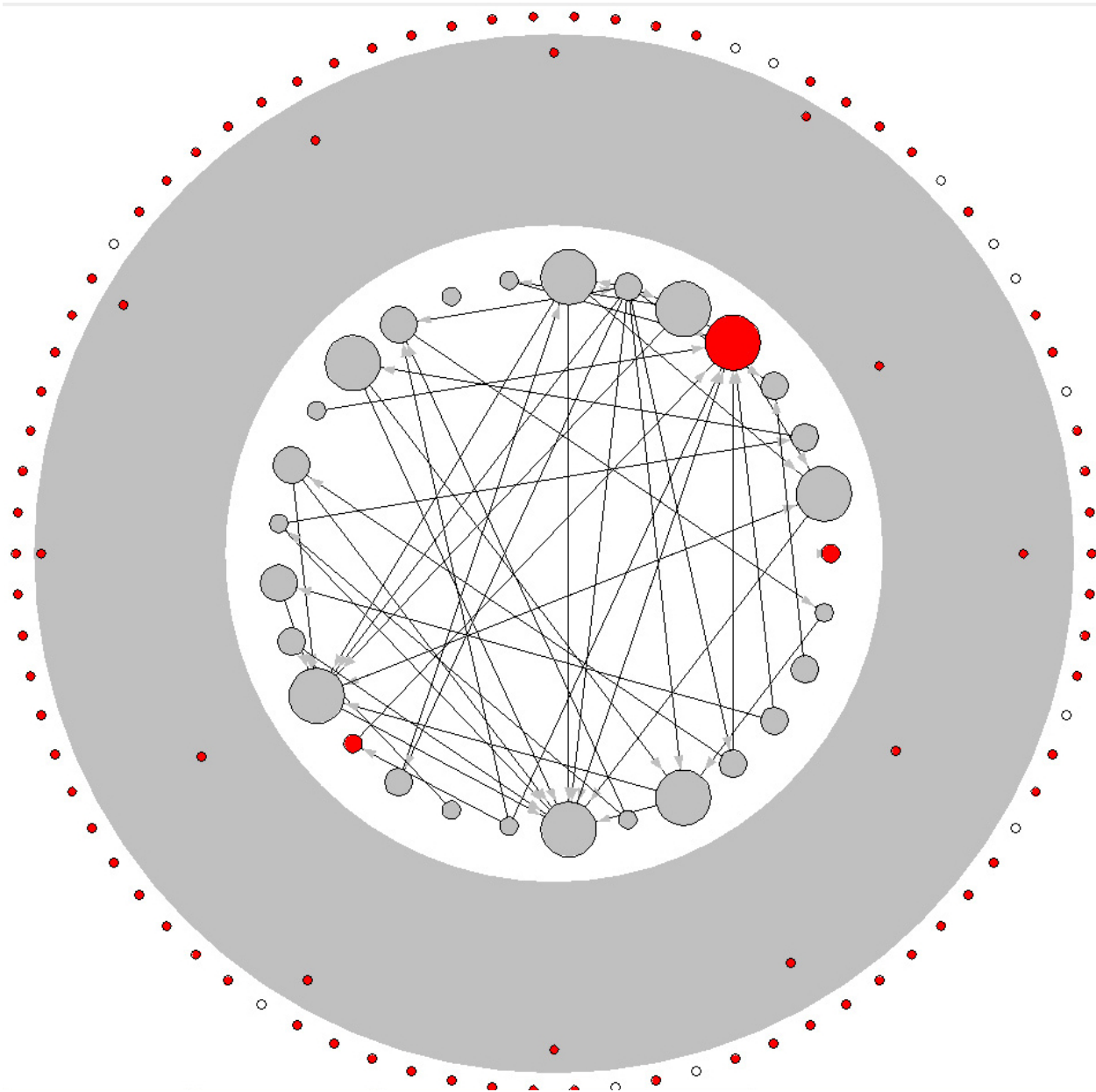


Figure 5.4: A visualization of a subsequent offering of an online introductory Computer Science course for non-majors. This course was offered after the instructor made pedagogy changes aimed at having learners interact more with one another. In this example, there are several learners who are *important* to the community as shown by their large node size and high level of connectedness.

5.2.2 Case study: Face-to-Face Large Cohort Course

In traditional face-to-face courses, instructors have the ability to make direct observations of learners. But these observations are minimal in bandwidth, especially when dealing with large cohorts, and tend to only give instructors a broad understanding of the issues learners face. Especially in introductory courses where many of the learners are new to both the discipline and higher education, there is often a hesitancy to speak up in the classroom. The hidden traces of online activities can be leveraged in these circumstances if the traditional course is augmented with technology and taught in a blended mode. Blending a course generally changes the lectures minimally, but offers other avenues for help and exploration (such as discussion forums).

A visualization (Figure 5.5) of a large cohort of learners involved in introductory Computer Science courses for majors was shown to one of the instructors who teaches in a blended fashion. Unlike the study described in §5.2.1, all learners taking this course (regardless of section) were shown one discussion forum. The instructor shown the visualization was both a lecturer as well as the overall administrative organizer for the course, and he would often leave the discussion forums open throughout the day so he could answer student questions. He immediately identified the nodes that represented himself and one of his tutorial assistants (large dark (red) nodes), and theorized on the identities of several of the learner nodes using the size and in-degree. He seemed comfortable with the visualization as an interpretation of the community formed in his course, and felt that the results he had achieved were those he set out to achieve.

A follow-up interview with the instructor provided surprising results; despite being pleased with the interpretation that resulted from the visualization of his course, he still modified his pedagogical approach. He indicated that for the first portion of his course he would typically answer questions as soon as they appeared in the discussion forums. He knew from the volume of questions being asked that learners used the system heavily, and he wanted to address those questions publicly as fast as possible in order to minimize their waiting. However, after seeing several large learner nodes in the visualization (e.g., several learners were read regularly by their peers), he felt that the community had reached a level where peer help would be sustainable without his intervention. In short, he felt he could reduce the speed at which he replied to learners without negatively affecting their learning experience, as their peers were active both in writing responses to questions and in having those responses read by classmates. He was shown a follow up image of the social network for his course after some time, and remained satisfied that interactions were still happening at a good pace.

In this case, the data-assisted approach was used to confirm a belief an instructor held which would have been otherwise untestable. By aggregating the low level read events captured by the forum system into a simple metric (*importance*), a small feature of the visualization became enough to convince the instructor of the self-sustaining nature of his course discussions.

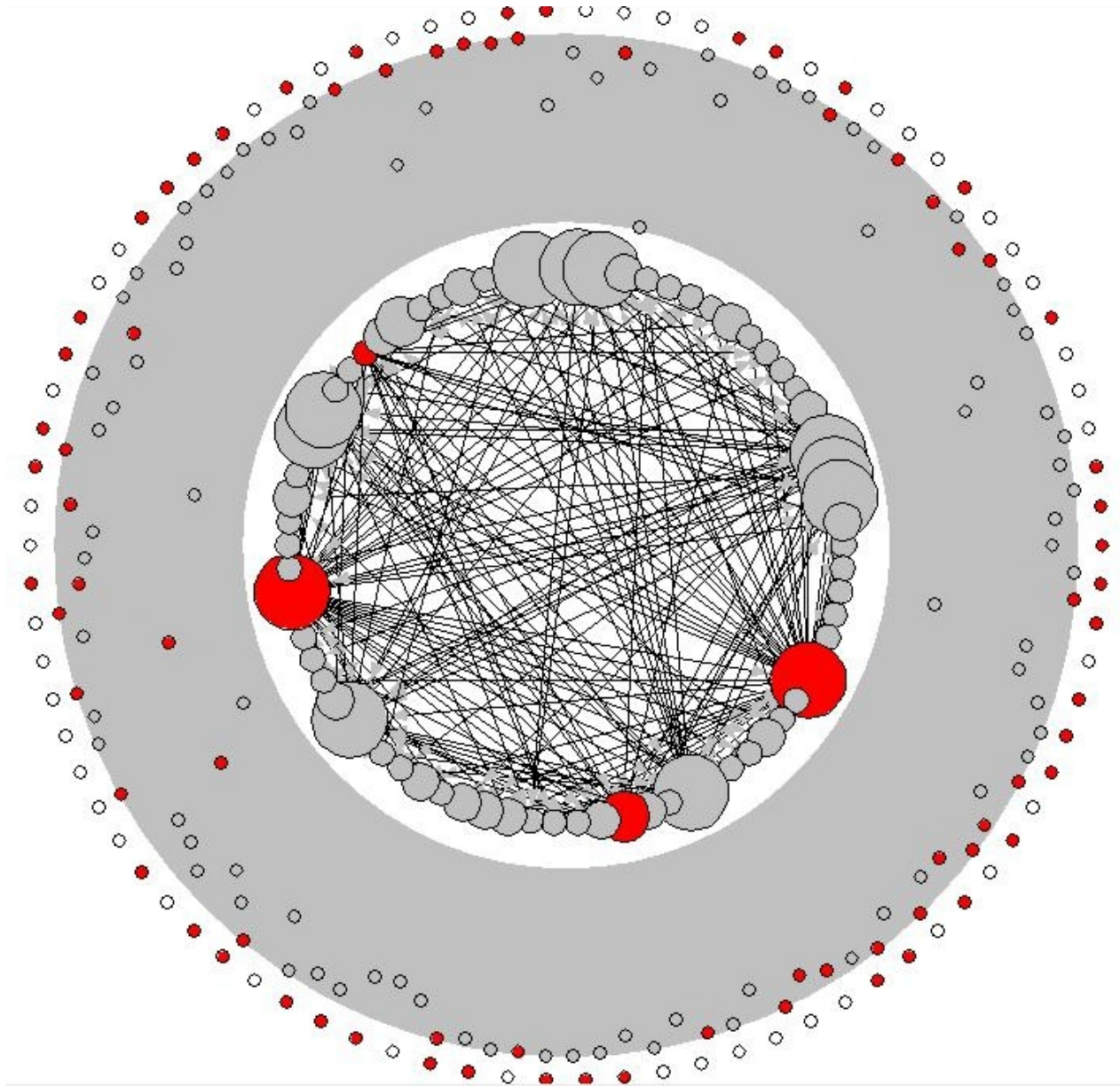


Figure 5.5: A visualization of an introductory course in Computer Science for majors. This visualization is made up of several sections of learners, each with their own instructor. The high degree of connectedness along with large student nodes (light grey in colour) supported the instructor's notion that course discussion would be sustainable if he reduced his level of involvement.

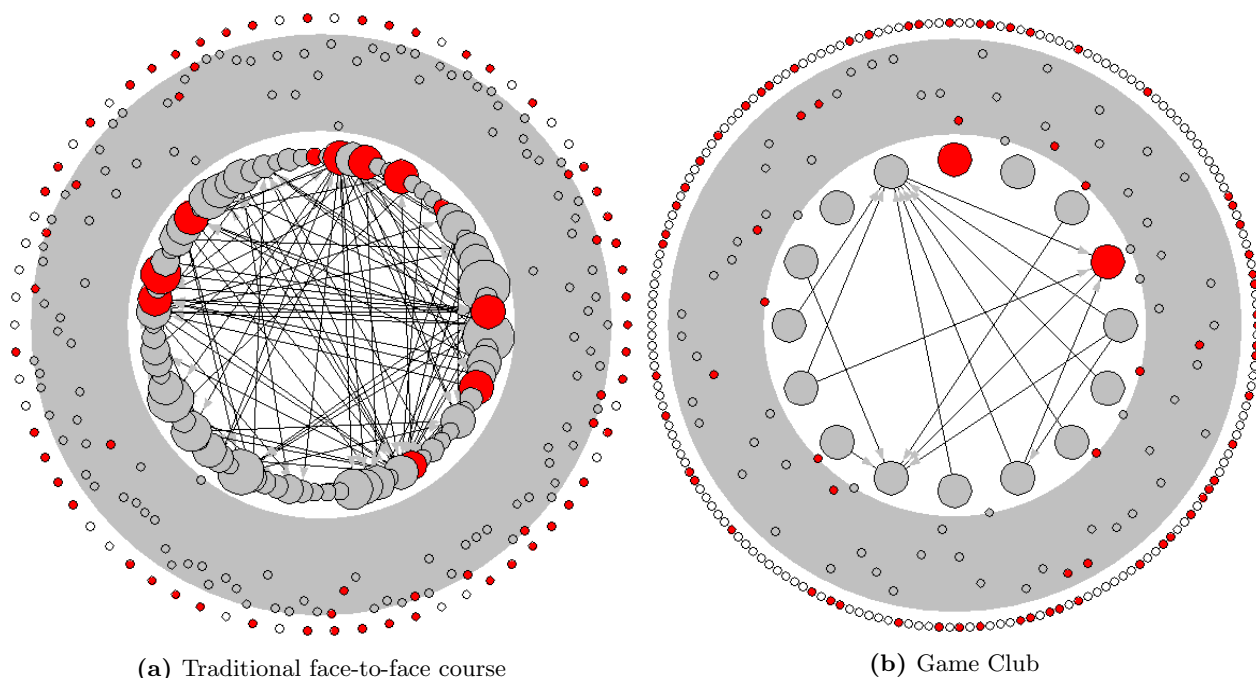


Figure 5.6: Visualization of a general discussion forum (figure 5.6a) and a more specialized discussion forum which is unrelated to course content (figure 5.6b). The user cohort is the same, though qualitative remarks about participation in and importance of the discussions can be made by comparing visual aspects of the two renderings.

5.2.3 Case study: Comparison of Visualizations based on Granularity

The hierarchical nature of the iHelp Discussion forum system allows instructors to create niche topics for their courses. One such example of this comes from an introductory course in Computer Science for majors, where most of the discussion happened in a general discussion forum (726 of 901 messages), but some happened in topic-specific forums. This course taught general principles of Computer Science, but based on the interests of the instructor and the students a subforum called the *Game Club* was created to talk about video game related issues. These two discussion forums can be visualized individually (Figure 5.6) and, while no discussion with the instructors about this forum took place, some high level qualitative comments can be made.

First, it is easy to see that there is much more activity in the general discussion forum than the game forum by looking at the in-degree of participants (higher in general forum) and the number of persons in the non-user sociogram (higher in game forum). Second, the rate of lurking is higher in the more specialized game forum, though the cause of this is unknown (it might be that there are just fewer postings, so it is easier for the keen learners to read them all, or there is an initial surge of activity where the community has promise but then dies out). Finally, it is worth noting that a number of instructional experts sit very close to the participation ring in the game forum, but do not write new messages or reply to messages to become participants. The precise role of the instructional experts here is not clear and it is difficult to draw specific conclusions; they may be instructors who are keeping an eye on developments, or tutorial assistants who are

peer students and very much interested in the content of the discussions.

Insight coming out of use of the data-assisted approach is very much about making visible the invisible – instructional experts cannot regularly see interactions learners make and thus do not involve them when making pedagogical decisions. In this example, an end user looking at discussion forum postings without the visualization would assume that only the couple of instructional experts who have written messages in the game forum are interested in the topic. The visualization, however, makes it clear that this is not the case, in that there are eight other instructional experts who have read almost every message posted in that forum. By making visible the hidden traces users leave as they interact with the learning environment, instructional experts are better positioned to make both short term and long term pedagogical decisions. Given the keenness of learners to view this game related forum, it might not be unreasonable for similar content to find its way into the standard curriculum for the course in future offerings.

5.2.4 Case study: Comparing Pedagogical Approaches

A modified version of the sociogram visualization tool was prepared for a specific investigation into how different kinds of conversation scaffolding can affect the engagement of students. In this version of the sociogram the non-users were excluded as all students were required to post, and node size was varied by the *perceived quality* of the messages a learner wrote (no instructors were involved in posting to the discussion forums). Our subjects included twenty senior level undergraduate education majors, and conversations were scaffolded either as a debate or as a role play on alternating weeks of the course.

Through a discussion with the instructor, a number of observations of these visualizations were made. First, there is a greater quality variation among students' postings in debates as opposed to role-plays. For instance, the large nodes in Figure 5.7a represent students who made several good contributions, while small nodes indicate students who made poorer contributions. Comparing with Figure 5.7b, we see that the differences in quality of contributions are much larger in the debate. Note that quality is subjective; however, node size was based directly on the information the instructor supplied about the quality of each student contribution. Using an evaluation rubric, the instructor calculated a weekly grade for each student's participation. These grades were converted to normalized values for each student for each week of the course.¹

Second, the rate of lurking was measured over time, and regardless of the pedagogy being applied this rate stayed relatively constant. The top third of active lurkers for any particular week were generally in the top third of lurkers throughout the whole course, supporting the notion that lurking is a learning style.

Finally, there is less interconnectedness of nodes in the role play approach versus the debate approach. In the debate, many learners replied to one another, which in the role play two learners controlled the conversation (*Christopher* and *Brian*). These two learners also received high marks for their contributions, a potentially interesting relationship for further study.

¹All names in this set of visualizations were fabricated and are not those of the learners being represented.

their course. In many institutions, program administrators or instructional designers are interested in similar understandings but across different offerings of a course. The following chapter will describe how data-assisted tools have been used to generate insights into learner behaviours across courses and disciplines.

CHAPTER 6

MEASURING EDUCATIONAL IMPACT

As shown in the previous chapter, the data-assisted approach can be used to generate insight which instructors can use to modify their teaching activity. Each of the case studies described broad changes in pedagogy that could be made based on understanding the hidden behaviours of learners. But the data-assisted approach is not limited to just instructors nor broad changes; it is possible for instructional designers, for instance, to leverage similar techniques to make changes that affect only a portion of the learner population. This chapter¹ describes how learner behaviour data can be statistically described and related directly to educational outcomes. It further provides methods to model learners and associate them with pedagogically sound groups where more individualized interactions can take place. In doing this, this chapter will address three questions that were raised by the motivation scenario in §4.2, namely:

- Can learners be clustered based on their viewing habits into pedagogically relevant groups?
- If so, do these groups differ in their formal assessment measure, or are other variables, such as self-reported learner satisfaction, more useful indicators of learning?
- Is it reasonable to do this kind of clustering investigation in mid-semester, or is forming of clusters of like learners only useful in multi-year comparisons?

These questions will be addressed in §§6.2–6.4 respectively.

6.1 The Recollect System

A key consideration of the data-assisted approach is whether learner interactions within the learning environment can be correlated with pedagogical goals and measures of learning outcomes. These interactions can be either explicitly made by learners (e.g., through the filling out of a survey) or implicitly made as a by-product of the learning activity itself (e.g., navigating through content). Sometimes referred to as *clickstream* data or *traces*, these interactions are difficult to understand on their own in part because of the large amount of data collected (potentially millions of data points) and the low level meaning that the data represents (e.g.,

¹Portions of this chapter appear in [22]

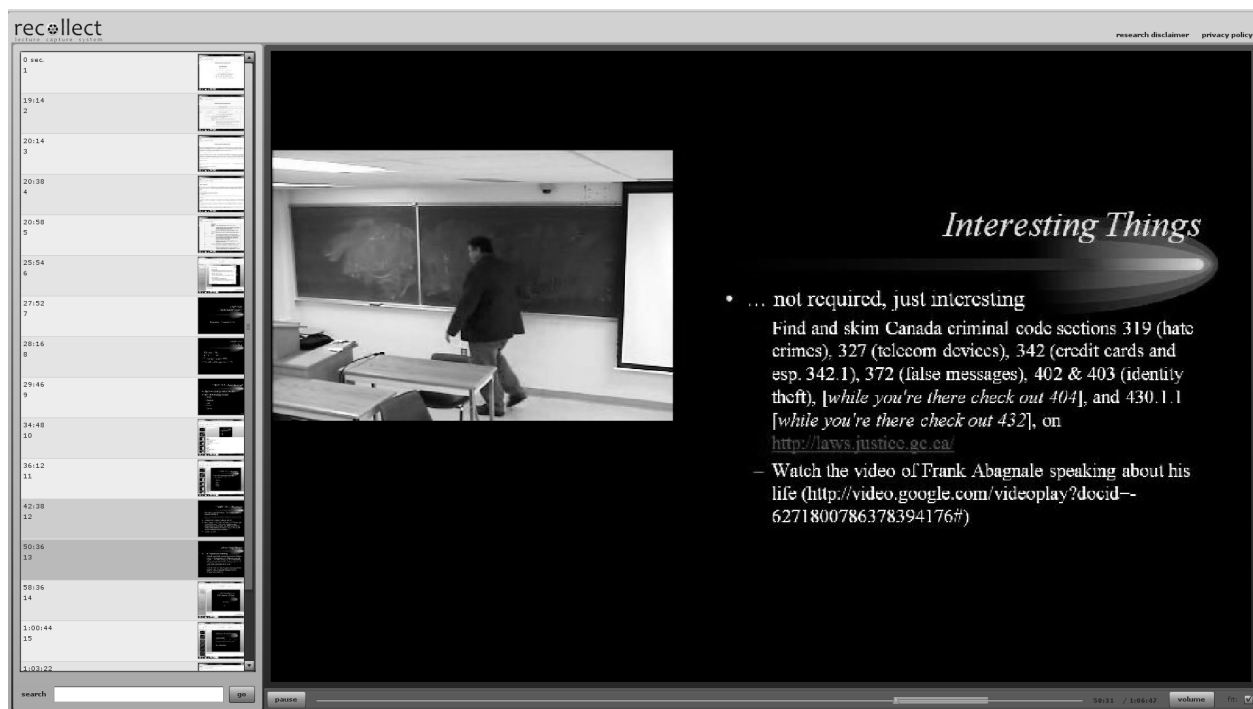


Figure 6.1: A screenshot of the Recollect lecture capture system. The system shows thumbnails of upcoming slides on the left of the window to allow for navigation. The right side of the player shows video of the classroom as well as video captured from the data projector.

the clicking of a single link or keystroke on the keyboard). The data must first be summarized, and then linked to learner goals in order to be made actionable.

One learning environment that collects this low level learner behaviour data is the Recollect lecture capture solution (Figure 6.1), developed at the University of Saskatchewan in part by the author. This system records in-person classroom lectures and stores them for playback by students for both the initial viewing of and reviewing of content. Recollect records a number of user behaviours, including the time spent streaming a lecture (discretized into thirty second intervals), clicks on any buttons in the user interface (e.g., volume change), searching through lecture slide content, seeking within the video using the video scrubber, and navigating within the video using section thumbnails. Each of these behaviours is linked to the time in which they were observed, the student who initiated the behaviour, and the particular video that was being watched.

6.2 Formal Assessment

The Recollect system was deployed for a number of sections of a second year Chemistry course in both the 2010 and 2011 academic years. Students were allowed to use the system how they saw fit, and every lecture from a single section taught by one professor was shared with students in all sections of the course. Instructors did not change grading criteria based on the presence of the recorded lectures, and midterm and

final examinations were common across all sections of the course.

Lecture capture is only one resource learners had available to them, and consistent patterns are difficult to see from the raw data. The viewing behaviours for Chemistry 2010 learners were cleaned² and summarized into viewing habits broken down by calendar week. In this model, learners were deemed to have either watched or not watched lecture content during the twelve weeks of the course. Out of 636 learners registered in the course, 133 were included in the study (participation rate of 36.6%) by virtue of their use of the Recollect system.

The weekly viewing rates of learners were used as attributes with k-means clustering ($k = 5$) to create a model of learner behaviours. Five clusters were chosen based on preconceived hypotheses of how learners might use the system (see the authors' previous work at [22] for more details). The form of the clusters is summarized in Table 6.1, and error for clusters ranged from 5% to 25% ($\bar{x} = 5.37\%$).

The results of the clustering activity provide a number of insights into learner activities. For instance, some learners only watch lectures during the week of the midterm, while others watch fairly regularly. Regardless of viewing patterns, the last two weeks of the course (corresponding to the time between the end of classes and the final examination) tended to have a high amount of disagreement between participants and the centroids. The disagreement for these weeks, ranging from 19% to 40%, suggests that activity throughout the term isn't indicative of behaviours between the end of term and the final exam, and thus only data during the teaching portion of the term was used for further analysis.

²A threshold of at least five minutes of viewing was arbitrarily chosen to remove behaviours that were deemed to be tool experimentation over tool use for learning. As the time period for this course was in the second semester of the academic year, the one week of data over midterm break was excluded from analysis.

| | | Academic Week | | | | | | | | | | | | | | | |
|------------------|--|---------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|--|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | |
| High Activity | Centroids | w | w | w | w | w | w | w | w | w | w | w | w | w | d | d | |
| | $n = 10, \bar{x}_{error} = 0.25$ Error | 0.50 | 0.30 | 0.20 | 0.20 | 0.10 | 0.30 | 0.00 | 0.10 | 0.30 | 0.20 | 0.20 | 0.40 | 0.30 | 0.20 | 0.40 | |
| Just-In-Time | Centroids | d | d | d | d | d | d | d | w | d | d | d | d | d | d | d | |
| | $n = 98, \bar{x}_{error} = 0.05$ Error | 0.03 | 0.08 | 0.08 | 0.04 | 0.05 | 0.01 | 0.10 | 0.00 | 0.02 | 0.05 | 0.04 | 0.04 | 0.05 | 0.05 | 0.15 | |
| Disillusioned | Centroids | w | w | w | w | w | d | d | d | d | d | d | d | d | d | d | |
| | $n = 8, \bar{x}_{error} = 0.15$ Error | 0.38 | 0.00 | 0.25 | 0.12 | 0.38 | 0.12 | 0.00 | 0.25 | 0.00 | 0.12 | 0.00 | 0.00 | 0.00 | 0.25 | 0.38 | |
| Deferred | Centroids | d | d | d | d | d | d | w | w | d | w | w | w | w | d | d | |
| | $n = 13, \bar{x}_{error} = 0.21$ Error | 0.08 | 0.31 | 0.15 | 0.00 | 0.08 | 0.08 | 0.38 | 0.23 | 0.38 | 0.31 | 0.08 | 0.15 | 0.31 | 0.23 | 0.31 | |
| Minimal Activity | Centroids | d | d | d | d | d | d | d | d | d | d | d | d | d | d | d | |
| | $n = 103, \bar{x}_{error} = 0.05$ Error | 0.02 | 0.06 | 0.04 | 0.02 | 0.06 | 0.02 | 0.08 | 0.00 | 0.07 | 0.08 | 0.05 | 0.02 | 0.03 | 0.02 | 0.19 | |

Table 6.1: K-means clustering of Chemistry 2010 Recollect data ($k = 5$). Clusters range in size from 8 to 103 participants, with mean error between 5% and 25%. Attributes used to form clusters were whether learners watched (w) or didn't watch (d) at least five minutes of lecture video. Week seven was omitted from the model due to the midterm break, and weeks 14 through 16 denote weeks between the end of lectures and the final exam. In this table, error refers to the difference between a particular centroid and the mean value of learners who best fit that cluster (e.g., a ratio of disagreement).

From this data, a high level model of viewing habits can be considered. The goal of developing such a model is to be able to accurately partition learners in a course into groups with behaviours corresponding to pedagogically meaningful patterns. Models can then be validated across multiple offerings of a course, disciplines, and year of study to understand the effect each of these clusters has on how learners use learning tools. The models can also be correlated with learning outcomes, such as final grade or perceived value of a course.

Using the data from Table 6.1, a model for five idealized clusters was developed. For instance, the first cluster has learners who habitually watch lectures throughout the term, so they were labelled *high activity learners*. The second cluster is made up of learners who observed the lecture the week before the midterm examination, so they were labelled as *just-in-time learners*. The third and fourth cluster appear to correspond to (roughly), the first and second half of the course, so they were labelled as *disillusioned learners* and *deferred learners* respectively^{3, 4}. Finally, the last cluster is made up of learners who habitually did not watch lectures. To be included in the study learners in this group must have watched at least five minutes of video in a given week, but the overall pattern of behaviour from these *minimal activity learners* suggests that the tool is used sparingly throughout the course instead of primary study aid, perhaps to catch up on missed lectures. As discussed, the relevance of the time between the end of lectures and the final exam did not seem indicative of the clusters as a whole, and was discarded from the general model.

With this general model in defined, a learner from any cohort can be placed into a particular group based on similarity. Put another way, a learner can be fitted into the cluster which minimizes the difference between the weekly list of learner interactions and the cluster centroids. For instance, a learner who watches video every week except for the first and sixth weeks will be placed into the *high activity* cluster. Despite this learner not fitting perfectly with this cluster, his or her activity patterns are most closely related to it. Similarly, a learner who watches during the midterm week and week 12 will be placed in the *just-in-time* cluster. That the learner watched lectures during week 12 contributes to the error the cluster centroids have (in this case, 0.04). Thus the centroids are not only the interesting aspects of the clusters, but also the amount of error for any given week of that cluster. Table 6.1 shows the amount of disagreement learners in a cluster have with the centroid of that cluster (*error*). Some weeks have large disagreement (e.g., week 1 of the *high activity* cluster, where 50% of the learners in that cluster did not watch video, or weeks eight and nine of the *deferred* cluster where 38% and 23% of the learners in that cluster did not watch video), while others have perfect agreement (e.g., week eight in the *high activity* cluster).

³It is useful to note that there is some overlap with the learners that used the lecture capture tool in the second half of the semester and those who used the tool only for preparation for the midterm examination. The discretization of the data into week-long time periods makes analysis sensitive to alignment issues with respect to the day of the week. There is some evidence to suggest that access to Recollect by learners is cyclical with respect to day of week. Previous work [22] ignored day of week, but for comparison and generalization of models between academic terms it is may be more principled to consider weekly accesses as starting on the first working day of the calendar week (as is done here).

⁴It is somewhat unclear where the division boundary between the disillusioned learners and the deferred learners should be. For the purpose of this work the division is considered to be the overlap of the midterm examination week which is included in both categories, but it is highly likely that the academic withdrawal schedule would impact learner decisions. Regardless, such a labelling is to be done by the educational expert, and they could experiment themselves with determining the most appropriate range of values as will be described in in §4 of the final dissertation.

| | n | Ideal ($k = 5$) | Model | Model-Ideal |
|----------------|-----|-------------------|-------|-------------|
| Chemistry 2010 | 232 | 0.063 | 0.067 | 0.005 |
| Chemistry 2011 | 333 | 0.099 | 0.109 | 0.010 |

Table 6.2: Comparison of mean disagreement between an ideal clustering and a pedagogical model within the Chemistry 2010 and 2011 data sets. The values can range between 0 (perfect fit) and 1 (complete disagreement), and are given by the equation $\frac{\sum_{x \in s} [\sum_{i=0}^w (|x_i - cluster(x)_i|)]}{n * w}$, where n is the number of students, w is the number of weeks data exists for, s is the set of students being considered, and $cluster(x)$ is the vector of cluster centroids that the student best fits. Ideal clusters were restricted by $k = 5$.

| Cluster Label | $Midterm_{\bar{x}}$ | $Midterm_{\sigma}$ | $Final_{\bar{x}}$ | $Final_{\sigma}$ | $Overall_{\bar{x}}$ | $Overall_{\sigma}$ |
|---------------------------|---------------------|--------------------|-------------------|------------------|---------------------|--------------------|
| High Activity Learners | 77.32 | 15.71 | 75.43 | 22.19 | 80.14 | 14.49 |
| Disillusioned Learners | 64.71 | 15.72 | 58.98 | 17.81 | 66.82 | 13.85 |
| Just-In-Time Learners | 68.14 | 15.92 | 60.49 | 23.68 | 70.69 | 14.39 |
| Minimal Activity Learners | 64.30 | 15.36 | 58.98 | 22.89 | 68.41 | 14.71 |
| Deferred Learners | 63.33 | 12.20 | 59.83 | 20.21 | 69.04 | 12.43 |

Table 6.3: Midterm, final examination, and overall grade averages and standard deviations broken down by cluster in percentages for the Chemistry 2011 course. Students who did not use the system ($n = 216$) had midterm examination marks of $\bar{x} = 65.19, \sigma = 16.21$ and final examination marks of $\bar{x} = 61.75, \sigma = 21.29$ and overall examination marks of $\bar{x} = 68.59, \sigma = 15.34$. Group ANOVA provided support for significance for midterm marks ($f = 3.05, p = 0.02$) and overall marks ($f = 2.435, p = 0.0472$) but not final marks ($f = 1.71, p = 0.15$).

An overall description of how well the model fits a particular course can be obtained by applying the model to all learners in that course who have used the system, and summing and normalizing the error for each learner. Table 6.2 demonstrates this using the original cohort upon which the model was formed (Chemistry 2010), and a cohort from the following year (Chemistry 2011). While overall disagreement went up between the two offerings of the course this was mostly an effect of an increasing number of learners, and the model introduced extremely small error in comparison to the ideal clusters (0.010).

Instructional goals are often represented by midterm and final examinations and, while marks are not a complete measure of learning, they are often used by learners, instructors, and adaptive systems as a proxy for learning. Correlating patterns of behaviours with differences in marks provides evidence of detecting learning from activity. Learners use lecture capture as one tool to aid in learning, but many other tools and methods contribute to learning (e.g., online quizzes, in class lectures, textbooks, study groups) and make identifying the effect of any single tool difficult. A pairwise tukey test for the midterm examination, final examination, and overall marks for the 2011 cohort (Tables 6.4 through 6.6) demonstrates there is an effect on marks for one cluster of learners in particular, the *high activity learners*, and that the effect's significance ranges between the levels of $p = 0.021$ and $p = 0.240$.

| | High Activity | Disillusioned | Just-In-Time | Minimal Activity |
|------------------|---------------|---------------|--------------|------------------|
| Disillusioned | $p = 0.160$ | * | | |
| Just-In-Time | $p = 0.240$ | $p = 0.919$ | * | |
| Minimal Activity | $p = 0.021$ | $p = 1.000$ | $p = 0.317$ | * |
| Deferred | $p = 0.057$ | $p = 1.000$ | $p = 0.662$ | $p = 0.998$ |

Table 6.4: Tukey HSD confidence values between pairs of clusters using midterm examination marks. Strong support for differentiating the performance of *Minimal Activity* and *Deferred* groups of learners from the *High Activity* learners ($p = 0.021$ and $p = 0.057$ respectively). Weaker support exists when differentiating between *High Activity* learners and the groups of *Disillusioned* and *Just-In-Time* learners ($p = 0.160$ and $p = 0.240$ respectively).

| | High Activity | Disillusioned | Just-In-Time | Minimal Activity |
|------------------|---------------|---------------|--------------|------------------|
| Disillusioned | $p = 0.164$ | * | | |
| Just-In-Time | $p = 0.190$ | $p = 0.960$ | * | |
| Minimal Activity | $p = 0.103$ | $p = 0.986$ | $p = 0.995$ | * |
| Deferred | $p = 0.151$ | $p = 1.000$ | $p = 0.974$ | $p = 0.994$ |

Table 6.5: Tukey HSD confidence values between pairs of clusters using final examination marks. Weak support for differentiating the performance of all groups of learners from the *High Activity* learners, with p values ranging between 0.103 and 0.190.

| | High Activity | Disillusioned | Just-In-Time | Minimal Activity |
|------------------|---------------|---------------|--------------|------------------|
| Disillusioned | $p = 0.085$ | * | | |
| Just-In-Time | $p = 0.161$ | $p = 0.855$ | * | |
| Minimal Activity | $p = 0.031$ | $p = 0.993$ | $p = 0.749$ | * |
| Deferred | $p = 0.157$ | $p = 0.989$ | $p = 0.988$ | $p = 1.000$ |

Table 6.6: Tukey HSD confidence values between pairs of clusters using overall marks, which included examination marks, laboratory marks, and assignment marks. Strong support for differentiating the performance of *Disillusioned* and *Minimal Activity* learners from *High Activity* learners ($p = 0.085$ and $p = 0.031$ respectively). Weak support for differentiating the performance of *Just-In-Time* and *Deferred* learners with *High Activity* learners ($p = 0.161$ and $p = 0.157$ respectively).

| | n | Ideal ($k = 5$) | Model | Model-Ideal |
|-------------------|-----|-------------------|-------|-------------|
| Biomolecules 2011 | 190 | 0.133 | 0.163 | 0.0298 |

Table 6.7: Comparison of mean disagreement between an ideal clustering and a pedagogical model within the Biomolecules 2011 data set. The values can range between 0 (perfect fit) and 1 (complete disagreement), and are given by the equation described in Table 6.2. Ideal clusters were restricted by $k = 5$.

| Cluster Label | $Overall_{\bar{x}}$ | $Overall_{\sigma}$ |
|---------------------------|---------------------|--------------------|
| High Activity Learners | 78.53 | 12.20 |
| Disillusioned Learners | 66.5 | 15.67 |
| Just-In-Time Learners | 68.58 | 17.82 |
| Minimal Activity Learners | 67.47 | 18.82 |
| Deferred Learners | 67.6 | 16.05 |

Table 6.8: Overall grade averages and standard deviations broken down by cluster in percentages for the BMSC-200 2011 course. Group ANOVA provided does not show strong support for significance for overall marks ($f = 1.286, p = 0.2773$).

Forming a model on the Chemistry 2010 dataset and verifying its utility on the Chemistry 2011 dataset provides a baseline that might be adapted to other courses. To show cross-course validity of this model, data was collected from a second year Biomolecules course in 2011. This course has many of the same attributes as the Chemistry 2011 course: it was made up of a large cohort of learners ($n = 190$), had a single midterm, and required some of the same prerequisites as Chemistry 2011. The course was taught by a different instructor, and all students were in a single section.

Quantification of the amount of error when applying the model to the data set is shown in table 6.7. It is useful to note that the Biomolecules 2011 data set does not cluster as well to five clusters compared to the Chemistry 2010 or Chemistry 2011 values. Whereas the Chemistry courses had ideal values with under 0.1 error (0.063 and 0.099 respectively), the Biomolecules ideal $k = 5$ cluster had 0.133 error, which corresponds to roughly one misclassified week of interactions per learner. Despite this, the application of the five high level cluster descriptions in the model to the Biomolecules courses introduced minimal new error (0.0298), suggesting a reasonable fit.

While the Biomolecules 2011 course had midterm and final examples similar to the Chemistry 2011 course, only the overall grade of learners was available for analysis. This grade includes the aggregate of examinations, assignments, and laboratory exercises, and may be curved or scaled. Nonetheless, it is worthwhile considering Tukey honest significant difference [72] values to differentiate clusters to see how well the model performs (Table 6.9). The pattern of p-values (e.g., lower for the cluster of *high activity* versus all other clusters) is similar to that of the the Chemistry 2011 final marks (Table 6.6), though with less confidence.

This section of the work has demonstrated how interaction data can be used to form a model of interactivity for a cohort, how this model ties to formal evaluation goals, and how this model can be applied to similar courses with relative success. An important question around the pragmatics of this model still exists: Can the model be applied throughout the academic term to predict cluster membership with accuracy? The

| | High Activity | Disillusioned | Just-In-Time | Minimal Activity |
|------------------|---------------|---------------|--------------|------------------|
| Disillusioned | $p = 0.472$ | * | | |
| Just-In-Time | $p = 0.335$ | $p = 0.997$ | * | |
| Minimal Activity | $p = 0.179$ | $p = 0.9998$ | $p = 0.997$ | * |
| Deferred | $p = 0.456$ | $p = 1.000$ | $p = 0.9997$ | $p = 1.000$ |

Table 6.9: Tukey HSD confidence values between pairs of clusters using overall course marks for Biomolecules 2011.

answer to this question depends on what accuracy is necessary for a particular goal, and how long data collection can take place before a prediction of membership is necessary. This is the topic of the following section.

6.3 Actionability of Model

In the design of a tool to make use of interaction data that has been fitted to a complete model it is important to consider of how that model can be applied when only partial data exists. Consider the previously described model of user lecture video accesses throughout an academic term; if the intent is to make a prediction engine such as in [115], it is important to know what categorizational power the model holds at a particular time. Is it appropriate to use this model only at the end of the course for reflecting on learning, or can it be used by an instructor during the teaching period? Is it useful to use this model in the first week of a course, or is it only useful after midterm examinations? This section of the work will investigate how prediction accuracy of the model changes as more interaction data about learners is collected.

Table 6.10 shows the prediction accuracy of cluster membership using the Chemistry 2011 cohort. At each week five values are given for each cluster; a) the number of learners who are predicted to be in a given cluster by the end of the term and who, from their activity, already best fit this cluster (*true positives*), b) the number of learners who are predicted to be in another cluster by the end of the term and whose data thus far suggests they do not fit this cluster well (*true negatives*), c) the number of learners who are predicted to be in this cluster by the end of the term but whose data thus far suggests they best fit another cluster (*false positives*), d) the number of learners who are predicted to be put into this cluster by the end of the term but whose data suggests they do not fit this cluster the best thus far (*false negatives*), and e) the number of learners who are predicted as best fitting into this cluster (either as a *true positive* or a *false positive*) but who could optimally fit into at least one other cluster (*borderline*).

| | | Academic Week | | | | | | | | | | | |
|------------------|----------------|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| High Activity | True Positive | 11 | 12 | 10 | 14 | 13 | 14 | 14 | 8 | 12 | 12 | 14 | 14 |
| | True Negative | 259 | 238 | 292 | 282 | 297 | 293 | 300 | 314 | 309 | 313 | 311 | 318 |
| | False Positive | 59 | 80 | 26 | 36 | 21 | 25 | 18 | 4 | 9 | 5 | 7 | 0 |
| | False Negative | 3 | 2 | 4 | 0 | 1 | 0 | 0 | 6 | 2 | 2 | 0 | 0 |
| | Borderline | 70 | 92 | 36 | 50 | 34 | 39 | 32 | 4 | 14 | 4 | 11 | 3 |
| Deferred | True Positive | 22 | 26 | 25 | 27 | 25 | 27 | 15 | 8 | 14 | 18 | 21 | 27 |
| | True Negative | 65 | 20 | 34 | 26 | 32 | 22 | 213 | 289 | 288 | 294 | 297 | 305 |
| | False Positive | 240 | 285 | 271 | 279 | 273 | 283 | 92 | 16 | 17 | 11 | 8 | 0 |
| | False Negative | 5 | 1 | 2 | 0 | 2 | 0 | 12 | 19 | 13 | 9 | 6 | 0 |
| | Borderline | 262 | 311 | 296 | 306 | 298 | 310 | 107 | 15 | 23 | 14 | 15 | 14 |
| Minimal Activity | True Positive | 164 | 195 | 191 | 197 | 196 | 202 | 202 | 202 | 199 | 202 | 197 | 202 |
| | True Negative | 32 | 14 | 25 | 21 | 28 | 22 | 129 | 129 | 129 | 129 | 130 | 130 |
| | False Positive | 98 | 116 | 105 | 109 | 102 | 108 | 1 | 1 | 1 | 1 | 0 | 0 |
| | False Negative | 38 | 7 | 11 | 5 | 6 | 0 | 0 | 0 | 3 | 0 | 5 | 0 |
| | Borderline | 262 | 311 | 296 | 306 | 298 | 310 | 0 | 12 | 0 | 10 | 0 | 11 |
| Disillusioned | True Positive | 11 | 15 | 13 | 16 | 15 | 18 | 18 | 15 | 18 | 13 | 18 | 18 |
| | True Negative | 255 | 237 | 291 | 280 | 295 | 293 | 300 | 308 | 308 | 312 | 311 | 314 |
| | False Positive | 59 | 77 | 23 | 34 | 19 | 21 | 14 | 6 | 6 | 2 | 3 | 0 |
| | False Negative | 7 | 3 | 5 | 2 | 3 | 0 | 0 | 3 | 0 | 5 | 0 | 0 |
| | Borderline | 70 | 92 | 36 | 50 | 34 | 39 | 32 | 7 | 17 | 6 | 14 | 7 |
| Just-In-Time | True Positive | 82 | 91 | 88 | 90 | 88 | 92 | 92 | 86 | 92 | 85 | 92 | 92 |
| | True Negative | 60 | 20 | 32 | 24 | 30 | 22 | 225 | 231 | 231 | 237 | 234 | 240 |
| | False Positive | 180 | 220 | 208 | 216 | 210 | 218 | 15 | 9 | 9 | 3 | 6 | 0 |
| | False Negative | 10 | 1 | 4 | 2 | 4 | 0 | 0 | 6 | 0 | 7 | 0 | 0 |
| | Borderline | 262 | 311 | 296 | 306 | 298 | 310 | 107 | 7 | 26 | 6 | 18 | 7 |

Table 6.10: Actionability table for Chemistry 2011 data. At each week of instruction a learner could have either watched online video or not, and is classified into the appropriate cluster based on the model derived from Table 6.1. For each week of instruction, all learners can be classified as to which cluster they best fit based on their data thus far. For example, in week one a total of 70 learners best fit the *high activity* cluster (*true positive* + *false positive*), and 11 of these learners will end up in this cluster at the end of the term based on their interactions throughout the term (*true positive*). Fifty-nine of these learners will end up in another cluster by the end of the term (*false positive*). Further, a total of 262 learners do not fit this cluster based on their interactions thus far (*true negative* + *false negative*), with the vast majority (259) not fitting this cluster by the end of the term (*true negative*). Only three learners (*false negative* + *false negative*), with the vast majority (259) but will by the end of term. Since it is impossible to differentiate at week one between the *high activity* and *disillusioned* clusters, they both have identical values for *borderline*.

The clustering model developed uses a series of attributes relating to the week of lecture with a binary value of either watched or not-watched. It is impossible to differentiate clusters after the first week of data alone, as several clusters have the same pattern of usage (e.g., the pattern for the *high activity* and *disillusioned* learners is denoted by watching the first week of lecture, while the activity pattern for the *just-in-time*, *deferred*, and *minimal activity* learners are all denoted by not watching the first week of lecture). At the seventh week, it is possible to differentiate the *just-in-time* learners from the *minimal activity* learners, and by week eight it is possible to differentiate all of the clusters from one another.

Depending on the instructional intervention being instigated, different values of the actionability table may be interest. For instance, if it is in week five and the instructor is planning to e-mail learners who do not fit into the *high activity* cluster to encourage them to watch more video, then the instructor should be aware that a portion of the learners they are not emailing (in this case, twenty-one, which is the value for week five for the *high activity false positives*) will actually end up in the *disillusioned* cluster. Similarly, if it is week nine and the instructor has asked tutorial assistants to get in contact with all of the *disillusioned* learners, they should be aware that several of these learners (in this case six, given by the value for the *disillusioned false positives*) will not best fit this cluster by the end of the term. Thus it is important to contextualize that, with this model as a predictor, there is error that depends in part on the purpose for which an intervention is being made.

Although this error cannot be measured accurately until the end of the term, there are some broad comments that can be made. The *false* values drop significantly after clusters can be disambiguated from one another (table 6.2). As mentioned, in week seven it is possible to differentiate the *just-in-time learners* from the *minimal activity learners*, and the total *false positives* drop dramatically. In week eight it is possible to differentiate between the *just-in-time* and *deferred* learners, which reduces the *borderline* value for the *just-in-time* cluster. These are two different kinds of error; in the first case, the error is a mis-prediction caused by inaccuracy in the model. The model does not fit a particular learner well, and the solution may be to include more attributes in the model or form a larger set of models (increase k). In the second case, the error is caused by cluster ambiguity; the clusters are too similar to one another to distinguish where a particular learner goes. The solution for this problem is the same as the first; ambiguity is not necessarily a problem depending on the instructional intervention being taken. For instance, in the first example given previously where an instructor wanted to e-mail all learners who are not in the *high activity* cluster, it does not matter if he or she is unable to distinguish between the *minimal activity* cluster or the *just-in-time* cluster as both clusters contain learners of interest.

The previous section described how a pedagogically relevant model can be built based on learner interactions by correlating these actions with assessment. This section extends the understanding of the model by describing how it could be applied to data collected throughout the term, and how the prediction accuracy changes over time. However, marks are not the only indicator of learning, and learner sentiment into the perceived usefulness of a learning environment and how it impacts their activities may be useful as well. The

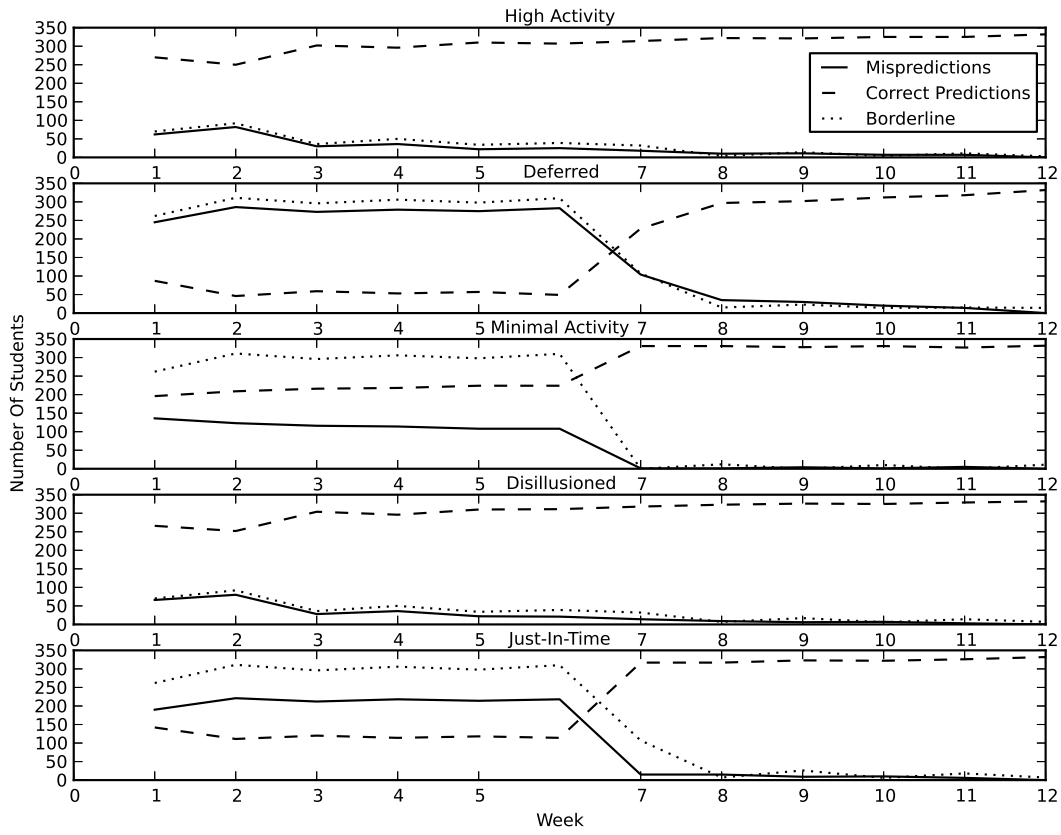


Figure 6.2: Graphs plotting the prediction accuracy of learners versus week in the term. Generally, as more information is collected about learners, the correct predictions increase and the mispredictions decrease. Borderline learners are those who fit equally well in multiple clusters based on their activity.

following section will apply the same methods described previously to subjective self-reported views of the Recollect lecture capture system with the goal of forming a model by which these views can be predicted from interaction data.

6.4 Learner Goals

Forming a relationship between behaviours and subjective questionnaire data allows for future instructional interventions to be made based on the behaviours alone. For instance, if a relationship is discovered between low use of lecture capture and negative opinions of the technology, the instructional expert may be able to change the learning environment in the future to accommodate learners who would prefer alternative tools based on their activity alone. While this section demonstrates some statistically significant differences were found, the differences between clusters is limited and (in this case) may not be strong enough for an instructor to act on. Nonetheless, this section acts as a further example of how unsupervised machine learning might be applied to interaction data through the *data-assisted approach*.

Formal assessment of educational activities through examination marks has long been criticized as not being completely indicative of learning. Learning itself is difficult to measure, and a number of proxies or determinants for learning have been considered by others including collaboration and discussion between learners, self-assessment of the value of activities, or affective state of the learner. The data-assisted approach encourages the exploration of the data by instructional experts. Armed with their knowledge of the domain and pedagogy, the experts can query the system to gain a deeper understand of how learners are acting within groups. Such queries are likely to be driven by hypotheses based on curiosity, preconceptions based on training, and instincts based on years of practice. To emulate this investigation, seven questions on the usefulness of the system and perceived workload were examined with respect to how well they fit clusters ($k = 2$)⁵ based on two behaviours: The number of minutes the learner watched and the number of unique videos the learner watched. The goal in doing this was to see if activity could be linked to statistically significant differences in learner opinions. If so, an instructor could, from behaviour alone, differentiate between the mental state of learners and form better adaptations for the two groups.

Learners ($n = 636$) in the Chemistry 2011 cohort were surveyed as to the relevance and usefulness of the Recollect system in this class. The questions asked covered a mixture of technical, pedagogical, and policy questions issues, and a number of these questions were designed to elicit beliefs learners had about their learning (response rate of $n = 229$, %30). The full survey instrument is included in Appendix A. The seven questions examined in more detail included questions four and five which asked learners to identify the final grade they were trying to obtain and to estimate the grade they thought they would actually obtain. In additions, questions six, nine, 21, 22, and 24 were asked, all dealing with perceptions on workload and

⁵The choice of the number of clusters (i.e. the value of k) to make affects outcomes greatly. This was an initial investigation to determine if unsupervised machine learning approaches can be used for clustering of subjective responses response data. Given the results shown here it is reasonable to continue exploration with an aim to find ideal values for k .

satisfaction. These questions were all zero indexed likert-based, with zero representing the strong affirmative (e.g., “Very High” or “Very Important”) and four representing the strong negative (e.g., “Very Low” or “Not Important”). From the survey instrument, the questions text were:

- Q4: I am working in this class to try and get a mark in range:
- Q5: Reflecting on my performance in the class so far, I think my mark will actually be in the range:
- Q6: My workload this term including all of the courses I am in as well as other commitments is:
- Q9: How important do you feel that watching the recording of the lecture was for your success in this class?
- Q21: If you used the lecture capture system, how important was it for reviewing content you hadn’t seen (e.g., missed classes).
- Q22: If you used the lecture capture system, how important was it for reviewing content you saw but didn’t understand or couldn’t remember?
- Q24: If you used the lecture capture system, how important was it for studying for examinations?

The analysis of these questions with the clusters formed is shown in Table 6.11. Labels on clusters were chosen by the author, and learners were segmented into a smaller cluster ($n = 12$) of users who watched a large number of videos (on average, 28) for a mean time of 19 hours and three minutes. A larger number of learners ($n = 115$) watched fewer videos (on average, six) for a mean time of three hours and 56 minutes. Only questions six, nine, 22, and 24 showed statistically significant results ($p \leq 0.05$), though the means between clusters for question six were of little meaningful difference.

This section of the work has shown how qualitative opinions from learners can be correlated with their behaviours. By clustering learners into groups based on their coarse grained usage data, the results of survey questions can be linked to activity. An instructional expert can then use these clusters as a proxy for survey results, and provide interventions without having to poll learners directly. Unfortunately, the particulars of the answers given by learners did not provide overly deep insight. While it was shown that learners who used the system more rated the importance of using lecture capture as higher, this is a minor (and perhaps common sense) result. Nonetheless, this section acts as a demonstration of how traces of learner behaviour can be linked to qualitative data that is not based on performance.

6.5 Conclusion

The data-assisted approach enables adaptivity in learning environments by presenting the traces of learner interactions to instructional experts. The instructional experts then investigate, interpret, and label collections of traces which are linked to instructional interventions. The interaction traces that learners leave behind are

| Attribute | Clusters | | | | ANOVA p-value |
|-----------------|-------------------|-----------|-------------------------|-----------|---------------|
| | Keen ($n = 12$) | | Less Keen ($n = 115$) | | |
| | \bar{x} | σ | \bar{x} | σ | |
| Minutes Watched | 1143.25 | 414.3705 | 187.3043 | 236.5078 | ≤ 0.001 |
| Unique Videos | 28.25 | 7.085517 | 6.930435 | 6.425991 | ≤ 0.001 |
| q4 | 1 | 1.044466 | 1.156522 | 1.005173 | 0.610 |
| q5 | 1.5 | 1.314257 | 1.695652 | 1.077472 | 0.559 |
| q6 | 0.5833333 | 0.7929615 | 0.7913043 | 0.7665788 | 0.037 |
| q9 | 0.4166667 | 0.5149287 | 1.191304 | 1.033608 | 0.012 |
| q21 | 0.25 | 0.8660254 | 0.4434783 | 0.9660128 | 0.507 |
| q22 | 0.1666667 | 0.3892495 | 0.9913043 | 1.301117 | 0.031 |
| q24 | 0.5 | 0.797724 | 1.252174 | 1.276257 | 0.048 |
| Midterm | 72.08333 | 18.45859 | 71.32609 | 15.13942 | 0.872 |
| Final | 72.22222 | 21.15137 | 70.4116 | 20.00128 | 0.767 |
| Overall | 77 | 15.32081 | 75.74783 | 13.15017 | 0.758 |

Table 6.11: Student behaviour clusters based on the number of minutes watched and unique videos watched with $k = 2$. Cluster labels, *keen* and *less keen* were added by the author as descriptive elements only, and are arbitrary. The only strong correlation with questionnaire data existed for questions six, nine, 22, and 24. See appendix A for question details.

numerous, and finding ways to summarize these interactions to instructional experts is an important problem in making the data-assisted approach tractable. This section of the dissertation demonstrates that the traces learners leave behind when using a lecture capture and playback system can be data-mined and related to educational outcomes and goals.

The first of these demonstrations (§6.2) showed how learners could be clustered based on coarse-grained viewing habits. These clusters were the basis for an abstract model, which was then verified on two additional cohorts of learners. Clusters in this model represented different learning strategies, and one group in particular, the *high activity learners*, correlated well with an increased achievement compared to other groups. With this knowledge, the instructor could apply clustering to future students and build instructional interventions aimed at particular groups. For instance, if it is the instructor’s belief that the correlation relationship between regular lecture video watching and higher marks is a causal relationship, he or she might send out an alert to all learners who are not watching videos to encourage them to watch more consistently.

The second demonstration (§6.3) extended the clustering of learners by describing how the model’s predictive power changes over time. This is an important practical factor when using the model to form real time instructional interventions for learners. This section also contextualized the kinds of errors (*false* and *borderline*) that exist in the model, and how they might affect intervention decisions.

Finally, §6.4 demonstrated how learner’s subjective opinions of the usefulness of the learning environment can be correlated with their behaviour patterns. While the particular patterns identified in this study couldn’t be confirmed through replication, they serve to illustrate a second method of how a pedagogically interesting model can be built using data-assisted approaches.

In all of these examples the parameters of and meaning of clusters needs to be defined or interpreted by the instructional expert. Relying on this expert for domain, content, and pedagogical reasoning increases

the generalizability of the data-assisted approach. Instead of exploring these domains ahead of time through a knowledge engineering process (for instance, through semantic web ontology creation or intelligent tutor development), the instructional expert can investigate data *in situ* and prune away aspects of the search space that have limited relevance to their context.

By providing labels for discovered groups back to the system along with some instructions on how adaptation should take place, future interactions with learners can be customized. In this way, the system and the instructional expert work together to form a personalized experience for learners. The next chapter will explore how this activity might happen from the perspective of an educational technologist.

CHAPTER 7

ADAPTING LEARNING ENVIRONMENTS TO TASKS

The scenario presented in §4.3 followed an educational technologist, Adam, who used the data-assisted approach to understand how learners are using the lecture recording tools. More than just gaining *insight*, Adam was interested in building *instructional interventions* of an automated manner. The previous chapter demonstrated how clusters of learners could be formed from their behaviour data. Using this, it is possible for an instructional expert like Adam to identify interesting groups of learners, and create an intervention. Thus far, however, only broad pedagogical interventions executed by instructors or instructional designers have been described. One of the interesting aspects of traditional intelligent learning environments (such as intelligent tutoring systems) is that they respond automatically to learner actions. It is worth considering whether this ability is lost in a data-assisted approach, where instructional experts are expected to be involved in the sensemaking process.

This chapter¹ describes how the data-assisted approach can be used to form low-level *instructional interventions* in lecture capture systems. To provide similar adaptations as one might find in intelligent learning environments, this chapter will focus on a very small technical adaptation, the customization of navigational elements in the Recollect system. In a production system these navigation element adaptations could grow out of the sensemaking process described previously, and directly link the insight (in the form of labelled clusters of learners) to the intervention (a change in navigational indexing scheme). The chapter goes further, and suggests how the underlying data collected by the learning environment might be used to further customize the indexing scheme. In doing so, specific learner behaviours not only influence insight, but also influence the form of instructional intervention that is delivered.

The studies in this chapter use opinions from learners as a proxy for interaction data. Using learner behaviours to determine significance would be preferable, however, measuring this would require significant re-tooling of the Recollect learning environment. Further, there are also many competing methods by which this might be done (a research question in its own right). For instance, one could imagine deploying a system with naïve indexing and collecting the frequency by which learners use different index points. Over time, the system could drop rarely used indices, and re-index portions of the lecture content. It may also be possible to use the scrubber navigation data, or to bootstrapping the system with data from instructional experts. Regardless, the laboratory studies described in this chapter focus on how behaviours can be turned into

¹Portions of this chapter appear in [21].

instructional interventions.

Considering again the scenario of Adam in §4.3 there were three main questions that needed to be addressed:

- Do groups of learners really agree on where indices should be placed, or are their preferences for navigational aids more varied? If the former is true, then the clustering methods described previously may well yield a more personalized and efficient navigation structure.
- Is it appropriate to use supervised machine learning to build indices, and how might such an approach compare to algorithms that already exist? Supervised machine learning is a strong technique when dealing with complex data containing a clear set of attributes. However, is it appropriate to deal with lecture videos in this manner? If so, can such a technique out-perform hand-tuned algorithms that already exist?

Each of these questions will be investigated in §7.3 and §7.4 respectively, using laboratory studies involving learners and actual recordings available from the Recollect lecture capture system.

7.1 Navigation in Recollect

The Recollect lecture capture environment (Figure 7.1) has multiple methods a learner can employ to navigate through content. For instance, thumbnails across the left hand side of the environment allow for quick “chaptering” of the content with image preview, while the scrubber along the bottom allows for precise navigation throughout the video based on time. It is not unreasonable to think that navigational style might differ depending on learning goal; a learner watching a lecture for the first time might not use either of these navigational aids; a learner who is searching for a particular topic in the lecture might use the thumbnails provided; a third kind of learner might use the scrubber to quickly replay video about a critical concept they missed while watching.

Thumbnails in the Recollect environment were originally generated using a naïve algorithm – every five minutes of video a still image would be copied from the video and metadata for the video would be updated linking the image and its position in the video. More sophisticated methods have been proposed for the same purpose; for instance, the Opencast Matterhorn system² uses a frame differencing algorithm with thresholds for RGB colour values, while Dickson’s algorithm [43] is a multi-pass image processing function that examines both pixel and block characteristics of video to determine stable events. Both of these algorithms were designed to work with lecture video captured by similar hardware as used by the Recollect system, making the potential for comparative study possible.

²In 2009 the University of Saskatchewan, through this author, joined the Opencast community and contributed ideas and software from the Recollect system. This relationship has deepened over the years, and the University of Saskatchewan now uses the Opencast Matterhorn product as its principal lecture capture system. While the author has worked on development of Opencast Matterhorn, no contributions were made to the indexing algorithms described here at the time of writing.

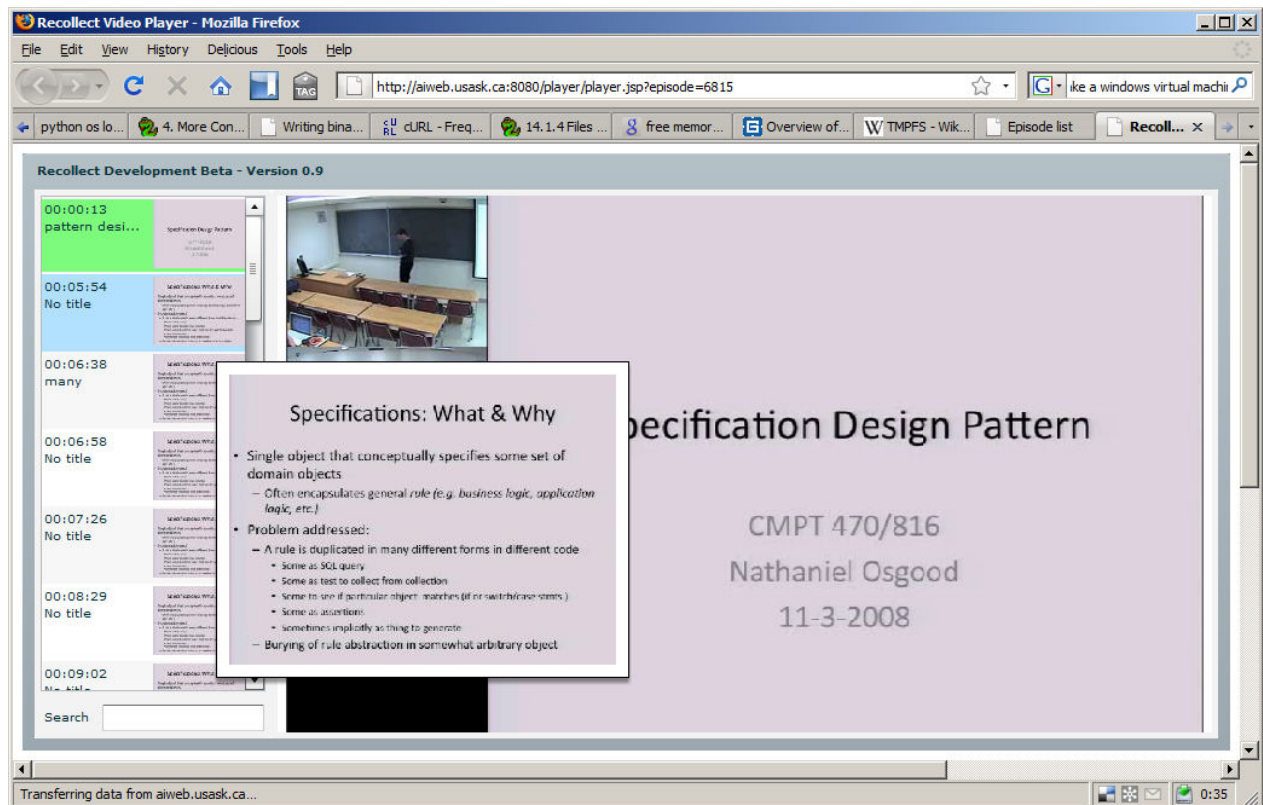


Figure 7.1: The Recollect lecture capture system, showing navigational thumbnails on the left hand side. As users mouse over a given thumbnail, a small image opens up and shows what the data projector feed recorded at the corresponding time in the video. All interactions involving the thumbnails such as mousing over, clicking, or scrolling through the list, are recorded.

7.2 Indexing Algorithms

There are several different approaches to indexing algorithms that have been taken by others. The most simple of these, *time-based indexing* is a naïve method that does not take into account the behaviours of learners or the content of the video. Index points are chosen every certain number of frames (e.g., every five minutes), and navigation with time-based indexing is of minimal utility when the video scrubber can be used as well. This method of generating indices is extremely simple and quick.

More sophisticated methods of generating indices tend to come by examining image content directly. The Opencast Matterhorn³ system uses compares frames of the video using the difference between RGB colour channels. A threshold is used to determine whether an image is stable or not in order to suppress indexes when moving content is detected. This is common in lecture capture situations, as an instructor might show a few web page items or play a youtube video, and an index should only be created at the start of these events and not for every frame through the events.

A method provided by Dickson et al. [43] is similar to that of the one Opencast Matterhorn uses but includes mechanisms to handle pixel shift. In this, the RGB differences between two frames are calculate only if the difference is greater than 100 (8 bit images, thus 255 bits per pixel) and if there is no immediate neighbouring pixel less than this amount. This handles issues of fuzziness resulting from the analog capture.

7.3 Grouping Learners by Navigation Usage

7.3.1 Comparing Users to Traditional Algorithms

A laboratory study was undertaken to determine whether indices of video could be created for lecture video based on learner opinions of relevance. indices are used by Recollect as thumbnails for navigating through a lecture, as shown in Figure 7.1, and correspond roughly with DVD chaptering.

Six human subjects who were unfamiliar with lecture capture systems were asked to go through four different lectures and identify where significant events occurred. For the purposes of this study, we described a significant event as “...the position in which a sub-chapter marker would be useful for students like [themselves] in quickly navigating through the lecture[s] for review or study”. A screenshot was provided (Figure 7.1) of the Recollect system to demonstrate the intended usage of these markers.

The tool provided to study subjects allowed for navigating through a video linearly both forwards and backwards in one, five, 10, and 30 frame increments (each frame was equivalent to one second of video). Subjects were also able watch the video at various speeds, though none used this option. No audio was provided as it was not being considered part of the candidate feature space. Subjects were both observed and given post-study questionnaires to elicit feedback.

³Version 1.2 of this system is described, see <http://www.opencastproject.org>

| Videos | κ ($n = 6$) | κ' ($n = 5$) | $\kappa'p_{pos}$ | $\kappa'p_{neg}$ | κ' with Dickson |
|---------|----------------------|-----------------------|------------------|------------------|------------------------|
| Video A | 0.65 | 0.86 | 0.87 | 0.99 | 0.55 |
| Video B | 0.49 | 0.65 | 0.68 | 0.99 | 0.42 |
| Video C | 0.51 | 0.65 | 0.66 | 0.99 | 0.44 |
| Video D | 0.14 | 0.18 | 0.21 | 0.99 | 0.11 |

Table 7.1: Inter-rater reliability measures using Fleiss’ κ [53] for significant events in four lecture videos. κ is calculated with $n = 6$, and κ' with ($n = 5$) by removing the one outlying rater (semantic annotator). Higher agreement between raters is shown for Videos A through C, while lower agreement is shown for the handwritten Video D. Adding in Dickson’s algorithm (adapted from [43] lowers overall group agreement in all cases.

The lecture videos chosen for this study included two senior-level undergraduate courses that used Microsoft PowerPoint as the main classroom presentation mechanism, and included either mostly plain black on white lettering (Video A) or more graphical slides with significant moments of Web-browsing during the presentation (Video B). One graduate course was also included (Video C), as well as an introductory undergraduate course which used handwritten lecture notes obtained by capturing instructor interaction with the SMART Sympodium system (Video D). All lectures were in the domains of Computer Science or Mathematics.

Observations and survey results from the subject participants identified that they used two distinct mechanisms for identifying significant events: visual structure (e.g., slide advancement in PowerPoint, or extending the canvas in the Sympodium) and semantics (topics being taught in the slides). Out of the six participants, five used primarily visual structure to identify events, while the sixth used the semantics of the lecture material. All participants indicated frustration at being unable to find good event times for a lecture that was handwritten (Video D in Table 7.1), in part because the video was constantly changing and because the ideal thumbnail for an index is not available until the end of the content it describes.⁴ All of the videos described in Table 7.1 were of content familiar to the subjects.

The effect of including the semantic rating used by the sixth annotator is clear throughout all videos. Using the categories provided by Landis and Koch [83], we see that for one of the videos (Video A) agreement increases to almost perfect ($\kappa=0.86$) when the semantic annotations of the sixth annotator are removed, and increases from moderate to substantial agreement for two of the other videos (Video B and Video C).

In addition to reporting the significance between human raters, values including Dickson’s algorithm [43] and positive and negative agreements are provided. Dickson’s algorithm reduces overall group agreement by a significant amount as the similarity between its output and any human rater is quite small. The positive and negative agreements were calculated using [107] on the advice of Feinstein and Cicchetti [50, 36] who argue that binary ratings can result in misleading low kappa values. As $\kappa'p_{pos}$ values (e.g., the amount of agreement that a slide is an index) are nearly equivalent to κ' , this supports the notion that κ' is a reasonable

⁴A thumbnail from time t best describes the content starting at time $t - \delta$ ending at t . This may be similar for videos which rely upon animations as well, depending on the purpose of the animations.

statistic to use.

7.3.2 Comparing Users to One Another

The previous section demonstrated that learners could create indexing points at locations similar to one another from a shared high level goal. A follow-up study was run to investigate why learners chose different indices, and six participants (three men and three women) were recruited, all of whom were or had been in the last two years undergraduate students but none of whom had used lecture capture in the past. Nine lectures from a single undergraduate course were used, resulting in a total of 43,770 video frames. Participants were asked to identify those images which indicated a new transition had taken place.

Study participants were instructed to mark index points (referred to in the study as *transitions*) in each video using a tool that allowed them to navigate through the video on a frame-by-frame basis. A purpose-based goal was used as a motivator for this task: learners were to “...mark all transitions as if [they] were building the left hand navigation window for a lecture video player,” and were shown the image similar to the one in Figure 7.1 to help better understand the task. Based on previous results, learners were dissuaded from semantically analyzing the content, and were asked to “mark transitions based on visual changes that [they thought] would be helpful for this lecture and other similar lectures.” Finally, subjects were asked to limit their index choices to between fifteen and thirty indices per lecture video.

Study participants also filled out questionnaires and met for a structured debriefing in which they were presented with collated results of the indexing activity, and participants shared with one another the methods they used to find indices. They described visual changes in images, charts, and graphs as being significant indicators, as well as changes in the titles on lecture slides, and the amount of time since they had last marked a transition. However, each participant did not indicate that these features had the same effect on their decisions, nor could they precisely agree on one heuristic that would best fit most cases. Coming to consensus within the group did not happen.

One finding that came out of the structured debriefing was that perceived educational task affected whether subjects thought a particular image was relevant as an index. This task was often described hypothetically based on the subject’s experience; for instance, when talking about the choice of an index that showed administrative information about the class, one subject said it was “useful, but only in the context of this as the first lecture...in another context, if this was a lecture in the middle of the year, I might not have caught it or cared.”.

This subject went on to distinguish between the task of reviewing a lecture versus viewing it the first time, indicating that for her this tool might be useful for some tasks but not all: “...if I’m going to be using this to review a lecture, it’s the context of the lecture that I’m getting, otherwise I’m just going to go back to the book, go back to the notes, I’m looking for the context of the lecture that he’s giving...”. Another subject followed this with a question about whether learners should “...be using this as a study instrument...”, while a third suggest the whole point of index thumbnails was to “...just quickly review”.

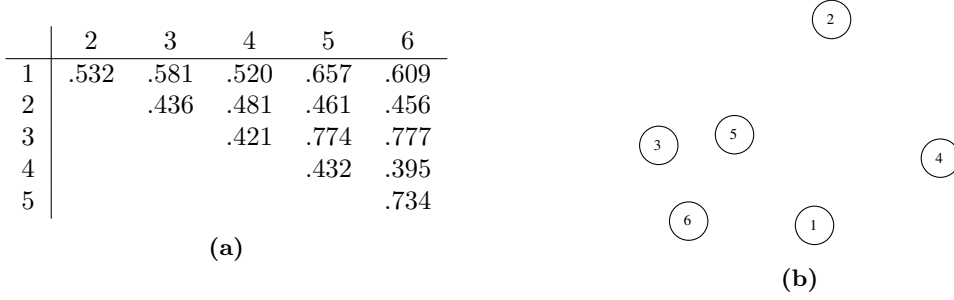


Figure 7.2: (a) contains pairwise κ between raters. (b) is an agreement map for 6 raters where distances between nodes are approximately inversely proportional to κ thus raters closer together have higher agreement. Node placement is done using a spring model layout. Edges are omitted for clarity as the graph is fully connected.

The first subject went on to suggest that there was some domain specificity as well, and that choosing an index might depend on the format of the learning that is happening in the class “...in an English class you’re looking for a lot of context...you’re not going to talk about how did you get that number from that example...if I’m taking a class where I’m going through say an accounting problem, and I don’t know how you made [an] entry I can’t get to the next one until I find the first [entry]...”.

Finally, a number of subjects suggested that their choices for rating came from intrinsic values, such as learning style: “...these seem important because they’re examples, so if someone’s a visual learner they’ll see it [as important]”

Fleiss’ κ statistic was used to measure agreement between these raters, similar to the first study. Recall that in the case where there are more than two raters, the value of κ measures overall group agreement, and any individual within the group may agree differently with each of the others. The amount of change any single rater can make to the overall level of agreement decreases with the number of raters, and one outlying participant could only marginally decrease an otherwise high kappa. Thus, it is useful to measure all pairwise κ ’s (for this study, Figure 7.2b) to understand the significance of each individual’s ratings.

Using terminology from Landis and Koch [83], pairwise agreement values range from fair ($\kappa = 0.395$) to substantial ($\kappa = 0.777$), confirming that human raters continue to maintain some reasonable overall degree of agreement when choosing index points; there may be room to subdivide the population into multiple groups that each have higher agreement individually.

7.3.3 Conclusions

This section has shown that learners are often unable to come to complete agreement on the choice of appropriate navigational indices, even in a structured discussion session. The agreement between learners, however, can be measured (Figure 7.2a). Measurements taken against current state-of-the-practice algorithms demonstrated that agreement between learners alone is higher than agreement between learners with the algorithm, suggesting that there is room for algorithm improvement (Table 7.1).

The amount of agreement between individual learners is not constant, and changes based on the perceived educational task, learning style, and end goal. In particular, the data from the second study shown in Figure 7.2a demonstrates that it is possible to form sub-groups of learners based on how alike they are. In that example, learners three, five, and six are much more similar to one another than learners two and four are. This suggests that forming a cluster to represent the first group of learners may be an appropriate method of aggregating opinions of indexing when many (hundreds) of learners are considered.

The question that remains is whether it is possible to make instructional interventions based on these results. The algorithms discussed here from Recollect, Opencast, and Dickson, all require no parametrization and return a single set of images. But learners do not share a constant level of agreement of what denotes an event worth indexing. The next section will provide a novel method for customizing indexing for a group of learners, and demonstrate its effectiveness using the data from the two studies described here.

7.4 Adapting Navigation based using Supervised Machine Learning

Methods such as the one in Opencast Matterhorn and Dickson's (described in §7.2) approach the issue of forming indices in lecture video as an image recognition problem. The goal of these methods is to measure the difference between two or more frames of video, and use this with some threshold value to determine when a significant change has occurred. The problem with this approach is in the selection of attributes that make up the difference function; they are not constant, and may need to be changed for different groups of users.

The image recognition approach can be augmented with a method for choosing and weighting appropriate attributes. By taking a representative set of indices for a group of learners, a selection criterion can be learned using supervised machine learning techniques. This criterion can then be applied to the broader set of videos, and indexing can be made more individualized.

Supervised machine learning methods take a set of instances, a set of attributes, and a set of classifications and build a model that can be used to predict new classifications for further instances with similar attributes. In the case described here, the set of instances are the video frames shown to subjects, the set of attributes are the image characteristics for these frames which are determined automatically (see Table 7.2) and the set of classifications is whether a given image is an index or is not. The output of a supervised method is a set of rules which can be applied to new images to determine if, based on this *training data*, those images are or are not indices.

To test the hypothesis that supervised machine learning can produce useful indexing, one human rater

⁵The GOCR framework was used, available at <http://jocr.sourceforge.net>

⁶All lectures were presented in English, though they may have contained non-English words, acronyms, domain-specific language, or related formula that may have been discarded through the text cleaning process. The *5 desk dictionary* word list was used, available at <http://wordlist.sourceforge.net>.

| Attribute | Description |
|---|--|
| Pixel Contrast | The difference of contrast between pixels in each image as per Rec. 601. [1] |
| Pixel Histogram Intersection | The minimum values of the histograms of two images summed across all bands of the image. |
| Block Contrast (16^2) | Similar to Pixel Contrast but the image is first partitioned in 16×16 pixel sized blocks. |
| Block Histogram Intersection (16^2) | Similar to Pixel Histogram Intersection but the image is first partitioned into 16×16 pixel sized blocks. |
| Block Contrast (32^2) | Similar to Block Contrast (16^2) but with 32×32 pixel sized blocks. |
| Block Histogram Intersection (32^2) | Similar to Block Histogram Intersection (16^2) but with 32×32 pixel sized blocks. |
| Block Contrast (64^2) | Similar to Block Contrast (16^2) but with 64×64 pixel sized blocks. |
| Block Histogram Intersection (64^2) | Similar to Block Histogram Intersection (16^2) but with 64×64 pixel sized blocks. |
| Difference | The mean of the difference of pixel values in each band of two images. |
| Difference ² | Similar to the above but it is the difference between the squares of the differences for each image band. |
| Mean | The average pixel value across all bands of one image compared to its neighbor. |
| Root Mean Squared | The difference between the root of the square of the means of each image across each band. |
| OCR Set Difference | Each image was run through an Optical Character Recognition (OCR) engine ⁵ . Previous experience suggested that non-English words and words less than three characters long are often the result of miss-recognitions by the OCR software. This attribute is the number of words included in the OCR output for this set that are longer than three characters, in the English dictionary, and not in an adjacent images word set. ⁶ |
| OCR Set Difference 2 | Similar to the above, but uses the previous images' word set. |
| OCR Intersection | The number of English words greater than three characters that exist in the intersection of two image word sets. |

Table 7.2: The series of visual attributes used to form supervised model. Each attribute was calculated for each image.

| Videos | κ_{J48} | $\kappa_{J48p_{pos}}$ | $\kappa_{J48p_{neg}}$ | κ_D | $\kappa_{Dp_{pos}}$ | $\kappa_{Dp_{neg}}$ |
|-------------------|----------------|-----------------------|-----------------------|------------|---------------------|---------------------|
| Course A | 0.72 | 0.75 | 0.99 | 0.36 | 0.01 | 0.99 |
| Course B | 0.66 | 0.63 | 0.99 | -0.28 | 0.01 | 0.99 |
| Course A \cup B | 0.67 | 0.66 | 0.99 | 0.37 | 0.01 | 0.99 |

Table 7.3: Comparison over one term of instruction from two classes, Course A and Course B, as well as the union of these datasets. Each video was encoded by a single rater, and J48 decision trees (κ_{J48}) provided with Weka using ten-fold testing to minimize chances of Type III errors. The result of applying Dickson’s algorithm to this data set is provided for performance comparison (κ_D).

was paid to annotate a term’s worth of video data for Video A as Course A and Video B as Course B using the same tool the subjects were given in the study described in §7.3.1. The rater was familiar with the project, and was instructed to use visual structure to do his annotations. He annotated roughly 59 hours of presentation.

The frames from these videos (210,637 in total) were encoded as high quality JPEG images, and the image characteristics described in §7.2 were calculated for each image. The annotated data and image characteristics were fed into the Weka toolkit, and a trained model was generated.

Table 7.3 summarizes the results of the data mining process using the J48 decision tree classifier provided with the Weka toolkit. While a number of classifiers were tried, the J48 classifier provided good results which are easily interpreted and easy to codify within the Recollect system. In addition to comparing the classifiers’ performance on each video individually, the union of both videos was also compared.

The results indicated a substantial agreement ($\kappa > 0.60$) for all tests, along with corresponding p_{pos} values, which indicates that the results of the classification correlated well with human subject values. While it is unfair to compare this agreement directly to κ_p values in Table 7.1 because of the difference in data set sizes (κ_p examined only a single lecture), it is interesting to note that values fall roughly in the same range of significance outlined by Landis and Koch [83]. Similar to previous results there was poor agreement between Dickson’s algorithm and the human rater. When slides included minimal interactivity (e.g., Course A) confidence was much higher, but still not significant. That $\kappa_{Dp_{neg}}$ is so high may be why in practice Dickson’s algorithm seems reasonable to most users - it doesn’t find “correct” indices, but is able to discount most of the “incorrect” frames appropriately.

With this demonstration, it seems reasonable to consider that navigation elements within lecture capture such as scene indices can be created using supervised machine learning of learners’ opinions, and that the accuracy of these indices (measured as agreement with the original rater) is competitive with other algorithms.

The second study described in this chapter had six human raters. The results of the three different algorithms described in §7.2 when added to the human rater values is shown in (Table 7.4), and only minimal changes in κ were observed. This is the result of the original group of raters being large in size (six persons) and largely in agreement; adding any new rater minimally affects the overall agreement. To more clearly see the effect of any particular algorithm then, the value for κ should be contextualized as to how it can change. Since it is more likely that a set of raters will disagree partially on any given instance than that they will

| Humans κ ($n = 6$) | κ Bounds | | Comparison Algorithms | | | Trained Algorithms | | | | | |
|--------------------------------|-----------------|--------------|-----------------------|----------|---------|--------------------|-------|-------|-------|-------|-------|
| | <i>upper</i> | <i>lower</i> | <i>Time</i> | Opencast | Dickson | T_1 | T_2 | T_3 | T_4 | T_5 | T_6 |
| 0.577 | 0.626 | -0.15 | 0.391 | 0.370 | 0.448 | 0.574 | 0.565 | 0.565 | 0.537 | 0.530 | 0.487 |

Table 7.4: Group κ between all raters and predictors. *upper* and *lower* contextualize the range κ can take when a new algorithm or rater is included. See §7.2 for descriptions of the *Time*, Opencast, and Dickson algorithms. Six supervised machine learning methods were trained from data T_1 through T_6 , where T_1 indicates that in order for an index to be chosen only one rater must have indicated so, T_2 indicates that for an index to be chosen at least two raters must have indicated so, and so on.

agree unanimously, the maximum agreement an added artificial set of ratings could achieve is determined by adding a rater who always agrees with the majority; the same is true for the minimum level of agreement, obtained by adding a rater who disagrees with the majority. After constructing these two artificial raters, a bound of possible κ values can be calculated for a group. Table 7.4 presents these lower and upper bounds on κ and shows that any given algorithm can at best raise the group κ to 0.626, or at worst lower the group κ to -0.15.

Six ten-fold cross-validated J48 decisions trees were formed using the data provided by the raters. The trees were formed on modified versions of the training set, adjusting the threshold for the minimum agreement among raters before an instance was considered an index. The thresholds were set such that T_1 was given a positive classification on all instances where at least one rater indicated there should be an index, T_2 is a tree trained on data where the threshold was two raters, and so on, until T_6 required perfect agreement between raters that an instance was an index. The goal was to see what effect different aggregation methods would have on the resulting level of agreement. As shown in 7.4, the trained algorithms all out perform the traditional methods of indexing, with the most stringent training method (T_6) providing the worst results ($\kappa = 0.487$) and the most lax training method (T_1) providing the best results ($\kappa = 0.574$).

7.5 Conclusion

Once interesting insights have been discovered in learner interaction data, instructional interventions to support learning can be created. These interventions can vary depending upon the pedagogical approach being taken and the resources available to the instructional expert. This chapter of the dissertation has explored the question of how a learning environment might be made adaptive based on the educational tasks of learners. In particular, it demonstrates that the kinds of low level adaptations found in traditional intelligent tutoring systems or adaptive hypermedia systems can be supported through the data-assisted approach. By having educational technologists embed their insights in a learning environment (e.g. labelling clusters as described in §4.3), fine grained instructional interventions can be created.

Using a mixture of qualitative and quantitative laboratory studies, a method for adapting the presentation of navigational indices in the user interface of a lecture capture system was explored. By combining

learner opinions of significance with supervised machine learning techniques, it has been demonstrated that substantially higher levels of accuracy of navigational indices can be achieved (§7.4).

This chapter has laid the groundwork for addressing the question of how navigational elements might be made even more personalized by considering educational task. In particular, this chapter has described how learners disagree with one another in index selection activities; different end uses, such as reviewing versus studying, as well as different contexts, such as the domain being reviewed or studied for, effected learner opinions about what made an particular time period appropriate for indexing. By identifying similarities between learners' behaviours (as in Chapter 6), learners can be partitioned into groups based on educational task, and opinions specific to the task can be used with supervised machine learning approaches to form even more appropriate indices for navigation. These indices are more accurate than traditional methods of index generation that ignore behaviour data.

CHAPTER 8

DISCUSSION AND CONCLUSIONS

Students learn better in more individualized tutoring situations [17], a result that has spawned two decades of intensive research in intelligent learning environments. These environments, such as intelligent tutoring systems and adaptive hypermedia systems, deliver content to learners and form models of them based on the *a priori* definition of pedagogical approaches, learner traits, and content semantics. This allows for personalization of the learning environment which can be realized by changing the content, navigation, or structure of the learning environment for a particular group of learners.

This method of personalizing learning environments is expensive. It requires up front cost in design and development and, as such, these methods are usually used within a single discipline or course. To scale across different domains, institutions of higher education use simplified learning content management systems. These systems offer instructors a thin technological wrapper around their existing content and learning activities, and provide only minimal support for personalization.

The data-assisted approach presented in this dissertation supports learning in technology enhanced learning environments by generating *insight* for instructional experts and enabling this insight to be used to form *instructional interventions*. Instead of replacing instructional experts, the data-assisted approach enables them to see the hidden traces learners leave behind as they interact with the learning environments, to understand these traces in light of educational goals, and to apply this insight to form instructional interventions.

8.1 Discussion of Findings

The principal contribution of this dissertation is a *data-assisted approach* to supporting instructional interventions in technology enhanced learning environments. This approach is novel: traditional support methods for technology enhanced learning environments require significant up-front design of learning situations (e.g., intelligent tutoring systems), provide minimal data on student interactions (e.g., learning content management systems), or focus attention on student-software interactions alone (e.g., adaptive hypermedia systems). The data-assisted approach, however, makes learner interaction data available to instructional experts with the aim of helping them to form *insights* into the learning process that will guide their *instructional interventions*. Thus the data-assisted approach explicitly embraces instructional experts as an on-going actor in

the intelligent learning environment.

This section of the dissertation discusses this principal contribution in more detail, outlining the results and impact of the various experiments described in chapters 5 through 7.

8.1.1 Forming Insight

There are many different kinds of instructional experts who might use data-assisted approaches: for example, instructors, instructional designers, tutorial assistants, and educational technologists. Each of these groups have different needs in order to generate insight. For instance, an instructor might need to be able to understand what problems are faced by a particular cohort of students they are instructing, while an instructional designer or an educational technologist might want to generalize trends across cohorts and obtain insight about particular approaches or tools.

The data-assisted approach is general enough to address these different cases. Chapter 5 described a situation where a variety of different instructors were shown visualizations of student interactions. These instructors taught different courses with different modalities (e.g., online versus blended in §5.2.1 and §5.2.2), different scopes (e.g., communities of interest in §5.2.3) and different pedagogical approaches (e.g., role-play versus debate as described in §5.2.4). In each of these cases instructors were able to form insights into how learners in their course were interacting, and were able to use this insight to change their teaching practice.

Instructional designers and education researchers are also actors that can engage with data-assisted approaches. Chapter 6 describes the use of unsupervised machine learning methods to discover clusters of learners based on their lecture video viewing habits. These clusters correlate well with both pedagogical expectations and educational outcomes. By making visible the hidden viewing habits of learners using a lecture capture system, statements about the efficacy of lecture capture as a study aid can be made. For instance, the evidence that learners who watch lectures regularly have higher outgoing grades suggests that lecture capture may have an impact on learning, an important consideration when designing support for large cohort courses.

The data-assisted approach creates insight with instructional experts through dialogue. In Chapter 5 this dialogue takes the form of information visualization, and instructors could see different discussion forums in their courses at different times. In Chapter 6 this dialogue was more interactive, and allows experts to parametrize clustering and select attributes of interest. Regardless, it is the method of explicitly including the instructional expert in the sensemaking process that makes the data-assisted approach suitable for building intelligent educational environments in higher education.

Having discussed some of the different cases the data-assisted approach has been applied to, it is worthwhile considering how general the approach is. In the work described in this dissertation, discussion forum and lecture capture tools were augmented with user tracking capabilities. Each of these tools has small fine-grained activities (e.g., reading a posting or watching a video) that were useful in building models of learners. Further, in each case the interactions being monitored were related to artifacts created by others –

in the case of the discussion forum tool these were messages posted by peers, while in the case of the lecture recording environment this was the instructors lecture. It seems reasonable to consider that tools with these two characteristics might be augmented similarly. For instance, learning content management systems are very similar to lecture capture systems in that the principal artifact is created by the instructor and that learner interaction with this artifact is through reading. There is some evidence then to suggest that once augmented, instructional experts may be able to draw insights from learner interactions here as well.

In the technology enhanced learning field there exist a number of tools that do not deal with collaborative artifacts. An example of this would be e-portfolio systems, where the software system exists to house a repository of a learner's previous work. Learners can use these systems for various purposes, with reflection and development of metacognitive skills being one well researched goal. Intelligent learning environments have a place here – intelligent tutors might query learners as to reflect on their work, and open learner modelling environments might share with the learner details about the tasks they have left to accomplish. In both circumstances, data-mining of the learner-created artifacts would be an important consideration. Are these kinds of environments suitable for the data-assisted approach? At first consideration there seem like several kinds of events that might be tracked – logging in of learners, adding new pieces of evidence, or writing a reflection on evidence. But is this enough to adequately inform instructional experts such that they might make an intervention? If not, is the data-assisted approach *necessarily* about the interactions learners have with artifacts, or should it include provisions for semantic analysis of the artifacts themselves? This latter approach suggests more up-front modelling; for instance, topics in a particular domain might need to be identified *a priori* in order to relate a learning artifact to that domain. Does such a broadening of this approach make it too general to effectively use, and are hybrid techniques between the data-assisted approach and others (e.g., knowledge engineering) a more appropriate description to make?

8.1.2 Creating Instructional Interventions

Once insight has been formed, instructional experts need a way to improve learning through instructional interventions. Here again the different roles of experts change the way that instructional interventions are made. For instance, in Chapter 5 instructors largely developed interventions outside of the technology-based learning environment. One instructor, for instance, changed her assignment requirements which caused learners to interact differently – an interaction pattern she saw as more pedagogically sound. Another instructor used the insight generated from the data-assisted approach to reduce his level of interaction in the class based on a perception that the current discussion environment was already sustainable. In both of these scenarios, the instructor made these interventions based on visualizations resulting from the application of the data-assisted approach.

Instructional interventions are not always broad pedagogical changes, and software systems such as adaptive hypermedia systems often focus on small customizations to the learning environment to improve learner experience for individuals or groups of learners. Chapter 7 demonstrated that the data-assisted approach

can be used to provide these forms of adaptations as well. Working from data representing ideal indices in lecture video, supervised machine learning was used to take prototypes of ideal indexing and apply these to different video content. Many different prototypes for various situations can be formed – those learners who want an overview might get one set of indices, while those who want visual navigation might get another. These prototypes do not have to be formed by an expert; instead, they can come directly from the learners themselves, and the expert (in this case an educational technologist) employs insight in designing a technology enhanced learning environment.

It is interesting to consider the level of granularity at which different kinds of experts create instructional interventions at. For instance, would it be appropriate for instructors to create the kind of indexing methods that Adam created in §4.3? In this dissertation instructors used the data-assisted approach for broad pedagogical changes (e.g., those in §4.1). Is it reasonable that they could “parameterize” the educational environment to take advantage of the insights they draw, or does this increase the demand on the instructor’s time more than appropriate?

There exists potential for novel hybrid techniques that use aspects of the data-assisted approach as well as other intelligent learning environment methods. For instance, constraint-based tutors can be iteratively built as new constraints are discovered. Here, the data-assisted approach offers the opportunity to identify cohorts of learners who perform well in a tutoring system but poorly on formative assessment. These learners may be holding misconceptions that the tutor does not handle, and an analysis of their actions could be useful in updating the constraint domain model. Thus a simple tutor can be designed initially (minimizing up-front modelling costs), and a more refined tutor can be developed over time. This refined tutor would have the advantages that it was built from the data of actual students – if these students are similar to future students, build a comprehensive domain model is not required.

8.1.3 Connecting Data to Insights and Insights to Interventions

While in this dissertation I demonstrate several different insights and interventions, it is less clear exactly which data and data-processing techniques lead to these insights and interventions. This question is particularly salient in light of the software engineering task of building personalized learning environments – to form repeatable design patterns (“recipes” for successful software development) that can be used by software developers to build data-assisted software, these developers need to understand what data and data-processing techniques will provide instructional insight.

This issue is multi-disciplinary in nature, and requires the consideration of education researchers, human-computer interaction designers, and information retrieval experts. Further, each set of data, data-processing technique, insight, and intervention can be considered at different levels of granularity, making the issue potentially more complex. Table 8.1 provides initial formulations of what such a taxonomy might look like, using the investigations provided in this dissertation. The spirit of the data-assisted approach is that the intelligence of the system is the result of a dialogue between software and the instructional expert. In keeping

| Data | Data-Processing Technique | Insight | Intervention |
|--|---|---|---|
| Student reading data of asynchronous discussion messages | Sociogram-based information visualization | Discover community of practice and level of social engagement | Scaffold discussion (instructor), contribute to communities of interest (instructor) |
| Student viewing of lecture videos | Clustering of students with K-Means and statistical treatment of assessment | Identify students with sub-optimal study habits | Prompt change in study habits (instructor), recommend lecture video usage (system) |
| Student navigation in lecture videos | Clustering of students with K-Means | Discover popular portions of lectures (instructional expert), identify segments of lecture video (system) | Review heavily studied concepts with extra material (instructor), Provide better indexing of video and adaptive navigation (system) |

Table 8.1: Outline of data, data-processing techniques, and the insight and instructional interventions they might lead to when using the data-assisted approach.

with this, table 8.1 should be seen as some general guidelines towards design patterns, and not an exhaustive list of which data and techniques lead to specific insights and interventions.

8.1.4 Primary Contribution

This dissertation has demonstrated that instructional experts are capable of gaining *insight* into their pedagogical practice and student cohorts which can be used to create various *instructional interventions*. The form of the insight and interventions varies greatly depending on the expert role, the learning situation, and the educational environment. By giving instructional experts an explicit role in the intelligent learning environment, minimal *a priori* effort is needed to create pedagogically valuable adaptations.

8.2 Secondary Contributions

In addition to the primary contribution of the data-assisted approach, this dissertation has made several three other meaningful contributions. First, several demonstrations that information visualizations of learner reading patterns can affect pedagogical approaches in both online and blended learning situations (§5.2.1 and §5.2.2) have been given. These demonstrations provide evidence that reading behaviours (e.g., lurking, or the importance metric), which are not often used in social network visualizations or learning content management system reports, are useful for instructional experts. Observing such behaviours in traditional classrooms is difficult, but the movement into technology enhanced learning environments opens a new set of behaviours instructional experts can use when trying to gain insight into their courses.

It was also shown that methods for creating navigational indices in lecture video (Chapter 7) could be improved upon. Not only was significant improvement demonstrated ($\kappa = 0.574$ versus the best alternative

of $\kappa = 0.448$, where $-0.15 \geq \kappa \leq 0.626$), but methods were described for quantifying the the bounds on the amount of improvement as well as qualitatively characterizing the expectations of learners. This has direct implications for the betterment of lecture capture systems. Further, there are methodological contributions when training machine learning algorithms from multiple raters in disagreement. As the agreement of raters change, so too does the amount of any possible change in κ .

Finally, there was the discovery of a correlation between formal assessment outcomes and the patterns of behaviour of learners in lecture capture environments (§6.2). This discovery was replicated and it was shown that similar trends in correlation exist in multiple cohorts. Further, a method for describing the actionability of such a learner model (§6.3) as well as initial results in correlating learner behaviour with self-reported goals were described (§6.4). These discoveries quantify how, in lecture capture environments, learner activities can be used to predict their educational outcomes. These correlations are the first step towards building predictive models for educational assessment based on lecture capture usage behaviours alone.

8.3 Future Work

8.3.1 The Data-Assisted Approach

When building an intelligent learning environment there are many paradigms and approaches that can be used. There is no clear guide as to which approach is best suited for a given purpose, and while Chapter 2 provides some modest suggestions of the suitability of different techniques, there have been very few comparison studies published. The biggest resource cost with the data-assisted approach is in the time and effort instructional experts put in. Quantifying this resource cost across different kinds of domains (e.g., ill-defined versus well-defined domains), pedagogical approaches (e.g., active learning where many tools might be used, versus environments that only use a few tools) and student types (e.g., higher education which is much more free-formed learning, versus the classroom learning of primary school) is important in understanding which situations the data-assisted approach excels in.

One growing method of instruction in higher education is the Massively Open Online Course (MOOC). To date, these courses involve hundreds to hundreds of thousands of learners at once, and are taught completely online, with the material for the course made up of a combination of lecture video, pre-recorded videos, live chat sessions, wiki content, and hypermedia. What makes these courses an interesting testbed for technology enhanced learning approaches is that they have some characteristics that intelligent tutoring systems excel at (e.g., teaching a single domain with well-defined steps or solutions) as well as some that intelligent tutoring systems are poor in supporting (e.g., a high diversity of learner background that is not known *a priori*). It may be that these kinds of courses (as well as others), are well suited to hybrid techniques, where data-assisted approaches are used for broad pedagogical direction and tutoring systems are used for particular topics.

8.3.2 Visualizing Community Interactions

One of the benefits of presenting visualizations as static images is that they can be shown quickly to instructors without requiring knowledge of particular interaction techniques for manipulating the visualizations. Some of the instructors who were showed the sociogram visualizations were interested in getting an idea of *momentum*. Were lurkers starting to read more, or did they quit reading in the second week of class and never come back to the tool? Was importance of participants growing uniformly, or were only a few individuals active at one time? There are various mechanisms that might be used to show this, with telepointer trails [63] being one promising direction.

Once there are tools in place to explore visualizations interactively (such as the tools Katheryn used in §4.1), it would be interesting to consider sharing visualization parameters between instructors. For instance, if Katheryn creates a visualization that demonstrates a deficiency in her course, is it useful for her to share these parameters through a repository with other instructors so they can apply it to their courses? If so, is it possible to create a set of antipatterns that describe various learning deficiencies, and form a community of practice around these experiences?

8.3.3 Measuring Educational Impact

Identifying determinants of learning in order to positively affect student experience is a key goal of much research in education. By demonstrating a correlative effect between regular usage of lecture capture and high achievement, the question that is left is whether a causal model can be formed. If using lecture capture regularly increases grades, learners should be encouraged to use the technology and deployment in more disciplines should be made a priority. Some work has already begun on this, by examining the incoming grades of learners in each cluster to determine if they are different. However, verification of the results across multiple cohorts studying different disciplines is important.

Verifying results across contexts is an important issue and, while this work involved some verification (three second year undergraduate STEM courses), it would be interesting to see if patterns of interaction hold for many different kinds of courses. Especially in first year courses it would be beneficial to verify that the clustering centroids chosen have minimal error with the ideal clusters. If not, are the clusters for these courses still pedagogically motivating, and do they lend themselves to forming instructional interventions? If so, do the patterns of interaction exist with the same amplitude (e.g., same proportion of learners per cluster), or does this change over the span of a learner's degree? Answers to these questions might help identify whether good study habits are learned, or intrinsic in the undergraduate population, and might help to support efforts in designing secondary school transition programs.

Lecture capture is just one technology that students have access to and, as shown in §6.3, acting on the results of lecture capture data at the beginning of a semester can involve a significant amount of false information (*false positives* and *false negatives*). By using multiple tools to form predictive models of student

success, it may be possible to make earlier interventions with more confidence. In particular, social network indicators (such as those from Chapter 5) may be useful in forming a more well rounded predictive student model, where both early and late predictions of educational outcomes could be made.

8.3.4 Adapting Learning Environments to Tasks

The method described in this work for generating indices in video is quite strong. It performs almost as well as the original human set of raters, and the process of generating indices is reasonably fast. Future work in this topic is split in three orthogonal directions. First, a reasonable method for getting training data from learner interactions needs to be created. While creating training data using a small group of learners is a reasonable first step, an automated method of getting training data from learners in a course would allow index selection models to be customized for a particular group and set of content without human intervention. This would add robustness to models and allow them to be used in new situations that the original model was not trained on (e.g., highly interactive content).

Second, extending the set of attributes beyond the initial measures should provide an iterative improvement to results described. In particular, only video data was considered, but lecture capture environments may also include audio data, clicker data, or copies of the original media (e.g., PowerPoint or PDF files). Including these as attributes may increase the quality of index markers. Further, speed is of concern and the processing of video content is generally slow, but processing audio content (for instance) is quite quick. It may be that some attributes are not needed for a given accuracy level, and can be omitted in order to speed up the indexing process.

Finally, some work has already been done to apply this method to broader classification of video content. For instance, if portions of the video can be classified to high level descriptions such as “web surfing”, “video playback”, or “textual content”, these classifications may be useful in other features of the lecture capture system. Of particular interest is automatically providing relevant links that might be detected from web usage, dealing with potential copyright considerations by identifying when videos are playing images are shown, and increasing search capabilities in video through optical character recognition of textual content.

8.4 Conclusion

Unlike most other methods of building intelligent learning environments, the data-assisted approach does not seek to replace instructional experts but to actively engage with them. It does this in two ways; by generating *insight* from data through a dialogue with the expert, and supporting experts as they act on this insight to form *instructional interventions*. This allows institutions of higher education to leverage the intellectual support resources they already have (e.g., instructors, instructional designers, and educational technologies) to provide more personalized learning experiences in technology enhanced environments.

REFERENCES

- [1] International Telecommunications Union. ITU-R Recommendation BT.601, 1982.
- [2] J. R. Anderson. *Rules of the mind*. Erlbaum, 1993.
- [3] J. R. Anderson, C. F. Boyle, and G. Yost. The geometry tutor. In *Proceedings of the Ninth International Joint conference on Artificial intelligence*, volume 1, pages 1–7. Morgan Kaufmann, 1985.
- [4] L. W. Anderson, D. R. Krathwohl, P. W. Airasian, K. A. Cruikshank, R. E. Mayer, P. Pintrich, J. Raths, and M. Wittrock. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom’s Taxonomy of Educational Objectives, Abridged Edition*. Allyn & Bacon, 2000.
- [5] Assessment Reform Group. Assessment for Learning: 10 Principles, 2002.
- [6] Assessment Reform Group. Assessment Reform Group, 2010.
- [7] D. J. Ayersman and A. Von Minden. Individual differences, computers, and instruction. *Computers in Human Behavior*, 11(10):371–390, 1995.
- [8] P. Baepler and C. J. Murdoch. Academic Analytics and Data Mining in Higher Education. *International Journal for the Scholarship of Teaching and Learning*, 4(2), 2010.
- [9] R. S. J. d. Baker. Data Mining for Education. In B. McGaw, P. Peterson, and E. Baker, editors, *International Encyclopedia of Education*. Elsevier, Oxford, UK, 3rd edition.
- [10] R. S. J. D. Baker and K. Yacef. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1):3–17, 2009.
- [11] S. Bateman, C. Gutwin, and G. McCalla. WebWear: Social Navigation for Loosely-Coupled Information Seeking in Tightly-Knit Groups. In *In submission to the ACM 2013 Conference on Computer Supported Cooperative Work (CSCW’13)*, 2013.
- [12] G. Bateson. *Steps to an Ecology of Mind*, volume 26. Aronson, 1973.
- [13] J. E. Beck and J. Sison. Using knowledge tracing in a noisy environment to measure student reading proficiencies. *International Journal of Artificial Intelligence in Education*, 16(2):129–143, 2006.
- [14] R. M. Belbin. *Management Teams*. John Wiley and Sons, New York, 1981.
- [15] P. Black and D. Wiliam. Assessment and classroom learning. *Assessment in education*, (October 2012):37–41, 1998.
- [16] B. S. Bloom. *Taxonomy of educational objectives: Handbook I*. David McKay Co, Inc., 1956.
- [17] B. S. Bloom. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6):4–16, 1984.
- [18] P. Brna, S. Bull, and V. Dimitrova. Learner Modelling for Reflection (LeMoRe).
- [19] P. Brna, J. Self, S. Bull, and H. Pain. Negotiated collaborative assessment through collaborative student modelling. In *Proceedings of the Workshop on Open, Interactive and other Overt Approaches to Learner Modelling*, pages 35–42. Citeseer, 1999.

- [20] C. Brooks. Supporting Learning Object Versioning, 2005.
- [21] C. Brooks and K. Amundson. Detecting Significant Events in Lecture Video using Supervised Machine Learning. In V. Dimitrova, R. Mizoguchi, B. du Boulay, and A. Graesser, editors, *2009 conference on Artificial Intelligence in Education (AIED09)*, Brighton, UK, 2009. IOS Press Amsterdam, The Netherlands, The Netherlands.
- [22] C. Brooks, C. D. Epp, G. Logan, and J. Greer. The who, what, when, and why of lecture capture. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*, pages 86–92, New York, New York, USA, 2011. ACM Press.
- [23] C. Brooks, M. Winter, J. Greer, and G. McCalla. The massive user modelling system (MUMS). In *Intelligent Tutoring Systems 2004 (ITS04)*, pages 73–112, Maceiò, Alagoas, Brazil, 2004. Springer.
- [24] P. Brusilovsky. Adaptive Hypermedia. *User Modeling and UserAdapted Interaction*, 11(1):87–110, 2001.
- [25] S. Bull. A Multiple Student and User Modelling System for Peer Interaction. In R. Schäfer and M. Bauer, editors, *ABIS-97: 5 GI-Workshop, Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen, Universität des Saarlandes*. 1997.
- [26] S. Bull, P. Brna, and H. Pain. Extending the scope of the student model. *User Modelling and User Adapted Interaction*, 5(1):45–65, 1995.
- [27] S. Bull, V. Dimitrova, and G. Mccalla. Open Learner Models: Research Questions Part 1. *International Journal of Artificial Intelligence in Education*, 2007.
- [28] S. Bull, J. Greer, G. McCalla, L. Kettel, and J. Bowes. User Modelling in I-Help: What, Why, When and How. *User Modeling 2001*, pages 117–126, 2001.
- [29] S. Bull and J. Kay. Chapter 15 : Open Learner Models. In R. M. R. Nkambou, J. Bourdeau, editor, *Advances in Intelligent Tutoring Systems*, pages 301–322. Springer-Verlag, Berlin Heidelberg, 2010.
- [30] S. Bull, A. Mabbott, and A. Abu Issa. UMPTEEN: Named and Anonymous Learner Model Access for Instructors and Peers. *International Journal of Artificial Intelligence in Education*, 17(3):227–253, 2007.
- [31] S. Bull and M. McKay. An open learner model for children and teachers: inspecting knowledge level of individuals and peers. In *Intelligent Tutoring Systems*, pages 232–251. Springer, 2004.
- [32] S. Bull and T. Nghiem. Helping learners to understand themselves with a learner model open to students, peers and instructors. In *Proceedings of Workshop on Individual and Group Modelling Methods that Help Learners Understand Themselves, International Conference on Intelligent Tutoring Systems*, volume 2002, pages 5–13. Citeseer, 2002.
- [33] W. Buskist. Ways of the Master Teacher. *Observer*, 2004.
- [34] Carnegie Learning. The Cognitive Tutor: Applying Cognitive Science to Education. Technical report, Carnegie Learning, Inc., Pittsburgh, PA, USA, 1998.
- [35] B. Carr and I. P. Goldstein. Overlays: a Theory of Modelling for Computer Aided Instruction, AI Memos, <http://dspace.mit.edu/handle/1721.1/5772>. Technical report, Massachusetts Institute of Technology, 1977.
- [36] D. V. Cicchetti and A. R. Feinstein. High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, 43(6):551–558, 1990.
- [37] O. Conlan. State of the Art : Adaptive Hypermedia. (May):47–57, 2003.
- [38] R. Cook and J. Kay. The justified user model: a viewable, explained user model. *Fourth International Conference on User Modelling*, 1994.

- [39] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.
- [40] A. I. Cristea and F. Ghali. Towards Adaptation in E-Learning 2 . 0. 2010.
- [41] B. K. Daniel, G. McCalla, and R. A. Schwier. Mining Data and Modelling Social Capital in Virtual Learning Communities. In *Artificial Intelligence in Education 2005 (AIED2005)*, pages 200–2008, 2005.
- [42] B. K. Daniel, R. A. Schwier, and G. McCalla. Social Capital in Virtual Learning Communities and Distributed Communities of Practice. *Canadian Journal of Learning and Technology*, 29(3):113–139, 2003.
- [43] P. Dickson, W. Adrion, and A. Hanson. Automatic Capture of Significant Points in a Computer Based Presentation. In *Proceedings of the Eighth IEEE International Symposium on Multimedia*, pages 921–926, San Diego, CA, US, 2006. IEEE Computer Society.
- [44] P. Dillenbourg, M. Baker, A. Blaye, and C. O’Malley. The evolution of research on collaborative learning. *Learning in Humans and Machine: Towards an interdisciplinary learning science*, pages 189–211, 1996.
- [45] E. Duval. Attention please! Learning analytics for visualization and recommendation. *1st International Conference on Learning Analytics (LAK11)*, 2011.
- [46] Educause. Next Generation Learning Challenges, 2011.
- [47] Educause Next Generation Learning Challenges. Learning Analytics, 2010.
- [48] T. Elias. Learning Analytics: Definitions, Processes and Potential, 2011.
- [49] T. Erickson and W. A. Kellogg. Social Translucence : Using Minimalist Visualizations of Social Activity to Support Collective Interaction. *Designing Information Spaces The Social Navigation Approach*, 7(1):1–19, 2003.
- [50] A. R. Feinstein and D. V. Cicchetti. High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, 43(6):543–549, 1990.
- [51] V. P. B. Felean. Analysis and Predictions on Students’ Behavior Using Decision Trees in Weka Environment. *2007 29th International Conference on Information Technology Interfaces*, pages 51–56, June 2007.
- [52] S. Fischer. Course and exercise sequencing using metadata in adaptive hypermedia learning systems. *Journal of Educational Resources in Computing*, 1(1es):5–es, 2001.
- [53] J. L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382, 1971.
- [54] E. Frias-Martinez, S. Y. Chen, and X. Liu. Survey of Data Mining Approaches to User Modeling for Adaptive Hypermedia. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 36(6):734–749, Nov. 2006.
- [55] E. Frias-Martinez, S. Y. Chen, R. D. Macredie, and X. Liu. The role of human factors in stereotyping behavior and perception of digital library users: a robust clustering approach. *User Modeling and User-Adapted Interaction*, 17(3):305–337, Apr. 2007.
- [56] S. Frielick. Beyond constructivism : An ecological approach to e-learning. *Technology*, pages 328–332, 2004.
- [57] W. Gardner, K. J. Kelleher, and K. A. Pajer. Multidimensional adaptive testing for mental health problems in primary care. *Medical Care*, 40(9):812–823, 2002.

- [58] Y. Gong, J. E. Beck, and N. T. Heffernan. Using multiple Dirichlet distributions to improve parameter plausibility. *Knowledge Creation Diffusion Utilization*, pages 61–70.
- [59] S. Govaerts, K. Verbert, J. Klerkx, and E. Duval. Visualizing Activities for Self-reflection and Awareness. pages 1–10.
- [60] S. M. Gowda, J. P. Rowe, R. S. J. d. Baker, M. Chi, and K. R. Koedinger. Improving Models of Slipping, Guessing, And Moment-By-Moment Learning with Estimates of Skill Difficulty. In *4th International Conference on Educational Data Mining*, pages 199–208, 2011.
- [61] S. Graf, T.-C. Liu, N.-S. Chen, and S. J. H. Yang. Learning styles and cognitive traits Their relationship and its benefits in web-based educational systems. *Computers in Human Behavior*, 25(6):1280–1289, 2009.
- [62] J. Greer, G. McCalla, and J. Vassileva. Lessons learned in deploying a multi-agent learning support system: The I-Help experience. 2001.
- [63] C. Gutwin. Traces: Visualizing the immediate past to support group interaction. In *Graphics Interface*, pages 43–50, 2002.
- [64] S. M. Haley, P. Ni, M. A. Fragala-Pinkham, A. M. Skrinar, and D. Corzo. A computer adaptive testing approach for assessing physical functioning in children and adolescents. *Developmental Medicine and Child Neurology*, 47(2):113–120, 2005.
- [65] C. Hansen. Active open learner modelling. In *Proceedings of the Workshop on Learner Modelling for Reflection (Supplemental Proceedings Vol 5)*, 2003.
- [66] C. Hansen. *Active Support for Instructors and Students in an Online Learning Environment*. PhD thesis, 2007.
- [67] A. Heraz and C. Frasson. Towards a Brain-Sensitive Intelligent Tutoring System: Detecting Emotions from Brainwaves. *Advances in Artificial Intelligence*, 2011.
- [68] A. HersHKovitz. Developing a log-based motivation measuring tool. *Educational Data Mining 2008*.
- [69] G. Holmes, A. Donkin, and I. H. Witten. WEKA: a machine learning workbench. *Proceedings of ANZIIS 94 Australian New Zealand Intelligent Information Systems Conference*, 24(3):357–361, 1994.
- [70] H. Hong and Kinshuk. Adaptation to Student Learning Styles in Web Based Educational Systems. In *EdMedia 2004 World Conference on Educational Multimedia Hypermedia Telecommunications*, pages 491–496, 2004.
- [71] H. U. Hoppe. The use of multiple student modeling to parameterize group learning. In *7th World Conference on Artificial Intelligence in Education (AIED95)*, 1995.
- [72] J. Hsu. *Multiple Comparisons: Theory and Methods*. Chapman and Hall/CRC, 1996.
- [73] P. Karampiperis and D. Sampson. Adaptive Learning Resources Sequencing in Educational Hypermedia Systems. *Educational Technology & Society*, 8(4):128–147, 2005.
- [74] S. Kardan and C. Conati. A Framework for Capturing Distinguishing User Interaction Behaviours in Novel Interfaces. In *Educational Data Mining 2011*, number i, 2011.
- [75] J. Kay. *A Scrutable User Modelling Shell for User-Adapted Interaction*. PhD thesis, University of Sydney, July 1998.
- [76] J. Kay. Stereotypes, student models and scrutability. In *Intelligent Tutoring Systems*, pages 19–30, 2000.
- [77] J. Kay. Scrutable Adaptation: Because We Can and Must. In V. Wade, H. Ashman, and B. Smyth, editors, *Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 11 – 19. Springer Berlin / Heidelberg, 2006.

- [78] J. Kay and A. Lum. Ontology-based user modelling for the Semantic Web. In *Proceedings of the Workshop on Personalisation on the Semantic Web: Per-SWeb05*, pages 15–23, 2005.
- [79] J. Kay, N. Maisonneuve, K. Yacef, and P. Reimann. The Big Five and Visualisations of Team Work Activity. In *Intelligent Tutoring Systems 2006 (ITS06)*, pages 197 – 206, 2006.
- [80] G. Klein, B. Moon, R. R. Hoffman, and K. Associates. Making Sense of Sensemaking 2: A Macro-cognitive Model. *IEEE Intelligent Systems*, 21(5):88–92, 2006.
- [81] V. Kodaganallur, R. Weitz, and D. Rosenthal. An Assessment of Constraint-Based Tutors: A Response to Mitrovic and Ohlsson’s Critique of “A Comparison of Model-Tracing and Constraint-Based Intelligent Tutoring Paradigms”. *International Journal of Artificial Intelligence in Education*, 16(3):291–321, 2006.
- [82] V. Kodaganallur, R. R. Weitz, and D. Rosenthal. A Comparison of Model-Tracing and Constraint-Based Intelligent Tutoring Paradigms. *Journal of Artificial Intelligence in Education*, 15:117–144, 2005.
- [83] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.
- [84] S. Lee and S. Bull. An Open Learner Model to Help Parents Help their Children. *Technology Instruction Cognition and Learning*, 6(1):29, 2008.
- [85] M. Levine. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 71(3):331–338, 1966.
- [86] M. Lilley and T. Barker. An evaluation of a computer adaptive test in a UK university context. In *7th Annual conference on Computer Assisted Assessment*, pages 171–182, 2003.
- [87] B. Martin. Constraint-based modelling: Representing student knowledge. *New Zealand Journal of Computing*, 7(2):30–38, 1999.
- [88] G. McCalla. The ecological approach to the design of e-learning environments: Purpose-based capture and use of information about learners. *Journal of Interactive Media in Education*, 2004(May):1–23, 2004.
- [89] G. McCalla, J. Vassileva, and J. Greer. Active learner modelling. In *Intelligent tutoring systems*, pages 53–62, 2000.
- [90] A. Merceron and K. Yacef. Educational data mining: a case study. In *Proceedings of the 12th international Conference on Artificial Intelligence in Education AIED*, 2005.
- [91] A. Mitrovic, K. Koedinger, and B. Martin. A comparative analysis of cognitive tutoring and constraint-based modeling. *User Modeling 2003*, pages 147–147, 2003.
- [92] A. Mitrovic and S. Ohlsson. A Critique of Kodaganallur, Weitz and Rosenthal,” A Comparison of Model-Tracing and Constraint-Based Intelligent Tutoring Paradigms”. *International Journal of Artificial Intelligence in Education*, 16(3):277–289, 2006.
- [93] J. Najjar and M. Wolpers. Attention metadata: Collection and management. *Workshop on Logging Traces of Web Activity: The Mechanics of Data Collection*, 2006.
- [94] X. Niu, J. Vassileva, and G. McCalla. Purpose-based User Modelling in Decentralized Agent and Web-Service Based Environments. *Environments*.
- [95] B. Nooraei, Z. A. Pardos, N. T. Heffernan, and R. S. J. D. Baker. Less Is More: Improving the Speed and Prediction Power of Knowledge Tracing by Using Less Data. *nth.wpi.edu*.
- [96] A. F. Nwagwe and K. R. Koedinger. The Simple Location Heuristic is Better at Predicting Students Changes in Error Rate Over Time Compared to the Simple Temporal Heuristic. *Computer*, 2005.

- [97] S. Ohlsson. Constraint-Based Student Modelling. In J. Greer and G. McCalla, editors, *Student Modelling: The Key to Individualized Knowledge-Based Instruction*, pages 165–189. Springer–Verlag, 1993.
- [98] S. Ohlsson. Learning from performance errors. *Psychological Review*, 103(2):241–262, 1996.
- [99] A. Paiva, J. Self, and R. Hartley. Externalising learner models. *Proceedings of World Conference on Artificial Intelligence*, 1995.
- [100] Z. Pardos and N. Heffernan. Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing. *User Modeling, Adaptation, and Personalization*, pages 255–266, 2010.
- [101] T. Peckham and G. McCalla. Mining Student Behavior Patterns in Reading Comprehension Tasks. In *Proceedings of the 5th International Conference on Educational Data Mining*, pages 87–94, 2012.
- [102] Pentaho Corporation. pentaho. <http://www.pentaho.com/>, 2011. Accessed August 2012.
- [103] C. Prpitsch and P. Veith. Content and management standards: LOM , SCORM and Content Packaging. In U.-D. Ehlers and J. M. Pawlowski, editors, *Handbook on Quality and Standardisation in E-Learning*, chapter 15, pages 209–223. Springer Berlin Heidelberg, 2006.
- [104] E. Rich. User Modeling via Stereotypes. *Cognitive Science*, 3(4):329–354, Oct. 1979.
- [105] C. Romero and S. Ventura. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1):135–146, July 2007.
- [106] V. Rus, M. Lintean, and R. Azevedo. Automatic detection of student mental models during prior knowledge activation in MetaTutor. In *Proceedings of the 2nd International Conference on Educational Data Mining*, pages 1–3. Cordoba Spain, 2009.
- [107] G. P. Samsa. Sampling distributions of p(pos) and p(neg). *Journal of Clinical Epidemiology*, 49(8):917–919, 1996.
- [108] J. Self. Bypassing the Intractable Problem of Student Modelling. *Artificial Intelligence*, (41):1–26, 1990.
- [109] V. J. Shute, F. State, E. G. Hansen, and R. G. Almond. You Can’t Fatten A Hog by Weighing It Or Can You ? Evaluating an Assessment for Learning System Called ACED. *International Journal of Artificial Intelligence in Education*, 18:289–316, 2008.
- [110] G. Siemens. What are learning analytics?, 2010.
- [111] G. Siemens. Sensemaking: Beyond analytics as a technical activity, 2012.
- [112] G. Siemens, D. Gasevic, C. Haythornthwaite, S. Dawson, S. B. Shum, R. Ferguson, E. Duval, K. Verbert, and R. S. J. D. Baker. Open Learning Analytics: an integrated & modularized platform. 2011.
- [113] P. Suraweera and A. Mitrovic. KERMIT: a constraint-based tutor for database modeling. In *Intelligent Tutoring Systems: 6th International Conference, ITS 2002, Biarritz, France and San Sebastian, Spain, June 2-7, 2002. Proceedings*, pages 201–216. Springer, 2002.
- [114] Tableau Software. Tableau. <http://www.tableausoftware.com/>, 2011. Accessed March 2012.
- [115] University of Purdue. Course Signals - Stoplights for Student Success. <http://www.itap.purdue.edu/learning/tools/signals/>, 2012. Accessed October 28, 2012.
- [116] K. Upton and J. Kay. Narcissus: Group and Individual Models to Support Small Group Work. *User Modeling Adaptation and Personalization*, pages 54–65, 2009.
- [117] J. Vassileva, R. Deters, J. Greer, and G. McCalla. Lessons from deploying I-Help. *Multi-Agent Architectures*, pages 5–7.

- [118] J. Vassileva, J. Greer, G. McCalla, R. Deters, D. Zapata, C. Mudgal, and S. Grant. A Multi-agent Approach to the Design of Peer-help Environments A Multi-agent Approach to the Design of Peer-help Environments. In *Artificial Intelligence in Education '99*, 1999.
- [119] H. Von Brevern. Context aware e-learning objects and their types from the perspective of the object-oriented paradigm. *Ieee International Conference on Advanced Learning Technologies Proceedings*, pages 681–683, 2004.
- [120] G. Weber. Episodic learner modeling. *Cognitive Science*, 20(2):195–236, June 1996.
- [121] M. Weskamp. Social Circles. <http://www.marumushi.com/apps/socialcircle>, 2003. Accessed March 2006.
- [122] M. Winter. Developing a Group Model for Student Software Engineering Teams, 2004.
- [123] H. A. Witkin, C. A. Moore, D. R. Goodenough, and P. W. Cox. Field-Dependent and Field-Independent Cognitive Styles and Their Educational Implications. *Review of Educational Research*, 47(1):1, Jan. 1977.
- [124] M. Wolpers, J. Najjar, K. Verbert, and E. Duval. Tracking Actual Usage : the Attention Metadata Approach. *Educational Technology & Society*, 10:106–121, 2007.
- [125] O. Zaiane. Recommender systems for e-learning: towards non-intrusive web mining. In C. Romero and S. Bentura, editors, *Data Mining in E-Learning*, chapter 5, pages 79–96. WIT Press, 2008.
- [126] D. Zapata-rivera, E. Hansen, V. J. Shute, and J. S. Underwood. Evidence-based Approach to Interacting with Open Student Models.
- [127] D. Zapata-rivera, J. S. Underwood, and M. Bauer. Advanced reporting systems in assessment environments. In *Proceedings of the Workshop on Learner Modelling for Reflection, International Conference on Artificial Intelligence in Education 2005 (AIED 2005)*, pages 23–31, Amsterdam, Netherlands, 2005.

APPENDIX A

STUDENT SURVEY INSTRUMENT

Department of Computer Science, University of Saskatchewan Informed Consent Form

Research Project: Data mining implicit and explicit user-generated data for semantic reasoning of educational video

Investigators: Dr. Jim Greer, Professor, Department of Computer Science (966-2234), greer@cs.usask.ca
Dr. Carl Gutwin, Professor, Department of Computer Science (966-8646), gutwin@cs.usask.ca
Mr. Christopher Brooks, Department of Computer Science (966-1442), cab938@mail.usask.ca
Dr. Michel Gravel, Department of Chemistry
Ms. Jaymie Koroluk, University Learning Centre

This consent should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information is not included here, please contact one of the study investigators listed above. Please take the time to read this form carefully and to understand any accompanying information.

In this study you will be asked to answer questions about your experiences with the Recollect lecture capture system. The results of these questions will be correlated with your activity in the Recollect system, as well as anonymized university achievement data (marks), to give us deeper insight into the effectiveness of the system and how it is used.

The data collected from this study will be used in articles for publication in journals and conference proceedings. As one way of thanking you for your time, we will be pleased to make available to you a summary of the results of this study once they have been compiled. This summary will outline the research and discuss our findings and recommendations.

All of the information we collect from you will be stored so that your name, student number, nsid, or email address is not associated with it. Any write-ups of the data will not include any information that can be linked directly to you. The research materials will be stored with complete security throughout the entire investigation.

None of the information collected will be shared with your instructor in non-anonymized forms. None of the information collected will influence your grade in this or other courses.

If you have further questions concerning matters related to this research, please contact one of the investigators listed above. By choosing the “yes” option below, you indicate that you have understood to your satisfaction the information regarding participation in the research project and agree to participate. In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. If you have further questions about this study or your rights as a participant, please contact:

Mr. Christopher Brooks, PhD Student
Department of Computer Science
(306) 966-1442,
cab938@mail.usask.ca

Alternatively, you can contact the University of Saskatchewan Research Ethics Office directly at:

Research Ethics Office,
University of Saskatchewan
1607 - 110 Gymnasium Place
Box 5000 RPO University
Saskatoon, SK Canada S7N 4J8
ethics.office@usask.ca
(306) 966-2084

The remuneration for this study is an entry in a draw to win a prize. As of March 1st, roughly 600 persons are eligible to win. Prize winners will be contacted via email in mid to late April. The prize is 1 of 10 \$50 Computer Store Gift Certificates

If you would like a copy of this consent form please ask at the front of the room or contact one of the investigators listed above. This research has the ethical approval of the Office of Research Services at the University of Saskatchewan.

NSID: _____ E-Mail: _____

Date: _____ Signature: _____

Instructions

- Please write your NSID **as well as** fill in the student number bubbles at the top of your opscan sheet.
- Please answer multiple choice questions using the opscan sheet provided. If you do not understand a question, or do not have an opinion, please skip the question or ask for clarification
- When done, please hand in this booklet at the front, including the consent form. You may take a copy of the loose consent form (at the front) for your records if you wish.

Thank you for helping us to make our research more effective!

Part 1: Demographics

1. What year of study are you in?
 - a) 1st
 - b) 2nd
 - c) 3rd
 - d) 4th
 - e) 5th or higher
2. Which group of student do you most identify with:
 - a) I'm in an agriculture program
 - b) I'm in an engineering program
 - c) I'm trying to get into a professional program such as Medicine, Pharmacy, Veterinary Medicine, etc
 - d) I'm in a natural sciences program
 - e) I'm in an arts, social sciences, or humanities program
3. I am taking this class because (mark all that apply)..
 - a) This class (specifically) is required for my program of study or a program I wish to be in.
 - b) I found content from similar classes interesting.
 - c) I have friends in this class.
 - d) I need a science credit.
 - e) It fit with my schedule.
 - f) The course description sounded interesting.
4. I am working in this class to try and get a mark in the range:
 - a) 90-100
 - b) 80-89
 - c) 70-79
 - d) 60-69
 - e) 50-59
5. Reflecting on my performance in the class so far, I think my mark will actually be in the range:
 - a) 90-100
 - b) 80-89
 - c) 70-79
 - d) 60-69
 - e) 50-59

6. My workload this term including all of the courses I am in as well as other commitments is:
- a) Very High
 - b) High
 - c) Average
 - d) Low
 - e) Very Low
7. Every student has different learning goals. Please describe how you are approaching learning in this class (mark all that apply)?
- a) I want to be able to remember the content. For example, to be able to recall some of it at a later date, such as on a midterm, when asked about it.
 - b) I want to understand the content. For example, to infer, summarize, or explain pieces of it to people when asked by others.
 - c) I want to be able to apply the content. For example, to use the knowledge in other courses, laboratories, or jobs I might have in the future.
 - d) I want to be able to analyze the content. For example, to be able to break it down into smaller pieces or organize concepts based on my experiences to build a more comprehensive understanding of why it is important.
8. How important do you feel that attending the lecture in person was for your success in this class?
- a) Very important
 - b) Somewhat important
 - c) Neither important nor unimportant
 - d) Somewhat unimportant
 - e) Not at all important
9. How important do you feel that watching the recording of the lecture was for your success in this class?
- a) Very important
 - b) Somewhat important
 - c) Neither important nor unimportant
 - d) Somewhat unimportant
 - e) Not at all important
10. How important do you feel that reading the text book was for your success in this class?
- a) Very important
 - b) Somewhat important
 - c) Neither important nor unimportant
 - d) Somewhat unimportant
 - e) Not at all important
11. How important do you feel that attending peer study groups (formal or informal) was for your success in this class?
- a) Very important
 - b) Somewhat important
 - c) Neither important nor unimportant
 - d) Somewhat unimportant
 - e) Not at all important
12. How important do you feel that reading online notes or PowerPoints was for your success in this class?
- a) Very important
 - b) Somewhat important
 - c) Neither important nor unimportant
 - d) Somewhat not important
 - e) Not important

13. How important do you feel that using the online quizzing system provided by the textbook publisher was for your success in this class?

- a) Very important
- b) Somewhat important
- c) Neither important nor unimportant
- d) Somewhat not important
- e) Not important

14. How important were web searches and web content outside of that provided by your instructor for your success in this class (e.g. Googling for similar content or explanations)?

- a) Very important
- b) Somewhat important
- c) Neither important nor unimportant
- d) Somewhat not important
- e) Not important

15. Did you purchase access to the online quizzing tools provided by the textbook publisher?

- a) Yes
- b) No

Part 2: Lecture Capture

16. If you used the lecture capture system, how valuable was it for you in this course?

- a) Very valuable
- b) Valuable
- c) Not valuable
- d) I didn't use it

17. Do you feel that having lectures recorded has lowered your attendance in this class?

- a) Yes
- b) No

18. Do you feel that having lectures recorded has lowered other students' attendance in this class?

- a) Yes
- b) No

19. If you did not use the lecture capture system, why didn't you use it?

- a) I did not know that recorded lectures were available for this class
- b) I did not have easy access to a computer or internet to access the videos
- c) I thought the recorded lectures were not valuable
- d) I did not feel I had time to view them adequately
- e) Other

20. Do you think you will use the lecture recording system before the final exam in this course?

- a) Yes
- b) No
- c) Maybe

21. If you used the lecture capture system, how important was it for reviewing content you hadn't seen (e.g. missed classes).
- a) Very important
 - b) Somewhat important
 - c) Neither important nor unimportant
 - d) Somewhat not important
 - e) Not important
22. If you used the lecture capture system, how important was it for reviewing content you saw but didn't understand or couldn't remember?
- a) Very important
 - b) Somewhat important
 - c) Neither important nor unimportant
 - d) Somewhat not important
 - e) Not important
23. If you used the lecture capture system, how important was it for completing assignment questions?
- a) Very important
 - b) Somewhat important
 - c) Neither important nor unimportant
 - d) Somewhat not important
 - e) Not important
24. If you used the lecture capture system, how important was it for studying for examinations?
- a) Very important
 - b) Somewhat important
 - c) Neither important nor unimportant
 - d) Somewhat not important
 - e) Not important
25. Check any of the following reasons you might have used the lecture system if they apply to you:
- a) I missed a lecture because of weather.
 - b) I missed a lecture because of illness.
 - c) I missed a lecture because of work.
26. If you were able to watch a one hour "highlights" video that was made up of pieces of video from different lectures on topics relevant to the midterm, would you?
- a) Yes
 - b) No
27. What features do you think would be useful with recorded lectures (check all that apply):
- a) High speed playback (e.g. 150% speed)
 - b) Annotation or note taking during video content
 - c) Live streaming of lectures as they are being taught
 - d) Access to recorded lectures on mobile devices (e.g. iPod, smart phone, etc)
 - e) Linking of video content to other web resources of the same topic
28. Have you used lecture capture in any other courses?
- a) Yes
 - b) No

29. Do you think recorded lectures are something you would like to see in your other courses?

- a) Yes
- b) No

30. If you were unable to take a course because of scheduling conflict or because of course enrolment restrictions, would you be interested in taking it online using recorded lectures and other online tools (e.g. discussion forums, blackboard, online advisers, etc.)

- a) Yes
- b) No

31. If the university had to increase your tuition by \$25 per course for each course that had recorded lectures, do you think this would be a reasonable expense?

- a) Yes
- b) No

APPENDIX B

BEHAVIOURAL ETHICS CERTIFICATES



Behavioural Research Ethics Board (Beh-REB)

Certificate of Approval

PRINCIPAL INVESTIGATOR
Jim Greer

DEPARTMENT
Computer Science

BEH#
09-265

INSTITUTION(S) WHERE RESEARCH WILL BE CONDUCTED
University of Saskatchewan
Saskatoon SK

SUB-INVESTIGATOR(S)
Carl Gutwin

STUDENT RESEARCHERS
Sonia Adams, Kris Amundson, Chris Brooks, Greg Logan

SPONSOR
NATURAL SCIENCES & ENGINEERING RESEARCH COUNCIL OF CANADA (NSERC)

TITLE
Data Mining Implicit and Explicit User-Generated Data for Semantic Reasoning of Educational Video

ORIGINAL REVIEW DATE
30-Nov-2009

APPROVAL ON
05-Jan-2010

APPROVAL OF:
Application for Approval of Research inclusive
of Appendix A, B & C (rev'd 23-Nov-2009)
Appendix D, Consent Form (rev'd 03-Jan-2010)
Appendix E: Follow-up Questionnaire

EXPIRY DATE
04-Jan-2011

Full Board Meeting ☐

Delegated Review ☒

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: http://www.usask.ca/research/ethics_review/


John Rigby, Chair
University of Saskatchewan
Behavioural Research Ethics Board

Please send all correspondence to:

Research Ethics Office
University of Saskatchewan
Box 5000 RPO University, 1602-110 Gymnasium Place
Saskatoon SK S7N 4J8



Behavioural Research Ethics Board (Beh-REB)

Certificate of Re-Approval

| | | |
|------------------------|------------------|--------|
| PRINCIPAL INVESTIGATOR | DEPARTMENT | Beh # |
| Jim Greer | Computer Science | 09-265 |

INSTITUTION (S) WHERE RESEARCH WILL BE CARRIED OUT

University of Saskatchewan
Saskatoon SK

SUB-INVESTIGATOR(S)

Carl Gutwin

STUDENT RESEARCHER(S)

Sonia Adams, Kris Amundson, Chris Brooks, Greg Logan

SPONSORING AGENCIES

NATURAL SCIENCES & ENGINEERING RESEARCH COUNCIL OF CANADA (NSERC)

TITLE:

Data Mining Implicit and Explicit User-Generated Data for Semantic Reasoning of Educational Video

RE-APPROVED ON

EXPIRY DATE

04-Jan-2011

04-Jan-2012

Full Board Meeting ☐

Delegated Review ☒

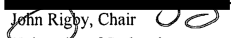
CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: http://www.usask.ca/research/ethics_review/


John Rigby, Chair
University of Saskatchewan
Behavioural Research Ethics Board

Please send all correspondence to:

Research Ethics Office
University of Saskatchewan
Box 5000 RPO University, 1607 – 110 Gymnasium Place
Saskatoon, SK S7N 4J8
Phone: (306) 966-2975 Fax: (306) 966-2069



UNIVERSITY OF
SASKATCHEWAN

Behavioural Research Ethics Board (Beh-REB)

**Certificate of Approval
Study Amendment**

PRINCIPAL INVESTIGATOR
Jim Greer

DEPARTMENT
Computer Science

Beh #
09-265

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT
University of Saskatchewan

SUB-INVESTIGATOR(S)
Carl Gutwin, Jaymie Koroluk, Michel Gravel

STUDENT RESEARCHER(S)
Sonia Adams, Kris Amundson, Chris Brooks, Greg Logan

FUNDER(S)
NATURAL SCIENCES & ENGINEERING RESEARCH
COUNCIL OF CANADA (NSERC)

TITLE
Data Mining Implicit and Explicit User-Generated Data for Semantic Reasoning of Educational Video

APPROVAL OF
Revised Student Survey
Revised Consent Form
Revised Remuneration Strategy
Addition of Jaymie Koroluk and Michel Gravel to Research Team

APPROVED ON
11-Mar-2011

CURRENT EXPIRY DATE
04-Jan-2012

Full Board Meeting ☐

Date of Full Board Meeting:

Delegated Review ☒

Expedited Review ☐

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: http://www.usask.ca/research/ethics_review/


John Rigby, Chair
University of Saskatchewan
Behavioural Research Ethics Board

Please send all correspondence to:

Ethics Office
University of Saskatchewan
Room 302 Kirk Hall, 117 Science Place
Saskatoon SK S7N 5C8
Telephone: (306) 966-2975 Fax: (306) 966-2069



UNIVERSITY OF
SASKATCHEWAN

Behavioural Research Ethics Board (Beh-REB)

Certificate of Re-Approval

| | | |
|------------------------|------------------|--------|
| PRINCIPAL INVESTIGATOR | DEPARTMENT | Beh # |
| Jim Greer | Computer Science | 09-265 |

INSTITUTION (S) WHERE RESEARCH WILL BE CARRIED OUT

University of Saskatchewan
Saskatoon SK

SUB-INVESTIGATOR(S)

Carl Gutwin, Jaymie Koroluk, Michel Gravel

STUDENT RESEARCHER(S)

Sonia Adams, Kris Amundson, Chris Brooks, Greg Logan

FUNDER(S)

NATURAL SCIENCES & ENGINEERING RESEARCH
COUNCIL OF CANADA (NSERC)

TITLE:

Data Mining Implicit and Explicit User-Generated Data for Semantic Reasoning of Educational Video

| | |
|----------------|-------------|
| RE-APPROVED ON | EXPIRY DATE |
| 04-Jan-2012 | 03-Jan-2013 |

Full Board Meeting ☐

Delegated Review ☒

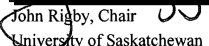
CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: http://www.usask.ca/research/ethics_review/


John Rigby, Chair
University of Saskatchewan
Behavioural Research Ethics Board

Please send all correspondence to:

Research Ethics Office
University of Saskatchewan
Box 5000 RPO University, 1607 – 110 Gymnasium Place
Saskatoon, SK S7N 4J8
Phone: (306) 966-2975 Fax: (306) 966-2069