Prior Learning Assessment with Latent Semantic Analysis

Marco Kalz, Jan van Bruggen, Bas Giesbers, and Rob Koper, *Educational Technology Expertise Centre, Open University of the Netherlands*

Abstract—**Until now most approaches in technology enhanced learning that take into account prior learning stem from learner modeling. In the context of the TENCompetence project we are exploring alternatives to this top down approach for Prior Learning Assessment. We explore Latent Semantic Analysis as a technique to assess prior learning by correlating documents in a learner portfolio with documents in target learning activities.**

I. Introduction

PRIOR learning experiences are important for learning. In some European Countries like the Netherlands or the UK the process of Accreditation of Prior Learning (APL) is a standard procedure to assess a student and allow exemptions for a study program [1]. The result of this process is an individualized curriculum. In a traditional APL procedure students apply for exemptions with a portfolio that is subsequently assessed by experts from the domain who then decide about exemptions. The drawback of this procedure is that is very time and cost consuming.

In the TENCompetence project we are aiming at the development of an infrastructure for lifelong competence development [2]. In this context we explore approaches to assess prior learning experiences and to offer individualized learning paths through a collection of learning activities in a learning network. Traditionally this problem has been addresses by adaptive hypermedia research on learner modeling [3]. But the solutions from learner modeling have several limitations: On the one hand they only work in one adaptive system that "learns" over time about the learner's preferences and behavior. On the other hand a static pre-designed model of a learner does not fit to the dynamics in lifelong learning.

To overcome the limitations of existing approaches we are developing content-based approach to prior learning assessment in learning networks.

II. Prior Learning Assessment in Learning Networks

Our project is based on the assumption that content can be taken as a proxy to estimate prior knowledge of a learner. The rational behind the project is discussed in [4]. The estimation of prior knowledge is calculated through the similarity of content in the learner's portfolio and the content that is connected to his/her target learning activities. To calculate this similarity we use Latent-Semantic-Analysis [5].

The results of such a similarity analysis is an ordered list of correlations between documents in the learner portfolio/profile and the target learning activities. High correlations with target activities may, depending on the policies of the learning environment entitle the learner to exemptions. Since the number of documents for solving this educational problem will be quite small compared to an information retrieval scenario, Van Bruggen et al. conducted an exploratory study into the usability of LSA in small scale corpora and reported promising results [6].

III. The Case Study

To test our model we collected data in an introductory psychology course at the Open University of the Netherlands. The online course consisted of 18 learning activities based on a textbook. Every chapter covers a subtopic of the psychology domain. We asked participants of this course in advance to comment on prior learning experiences that they considered relevant to the course. We invited them to substantiate this by, wherever possible, uploading files they had produced or read during their prior education. Since we could not expect students to know exactly what topics were presented in the chapters we also questioned them after completion of each chapter on the novelty of the presented material. We also constructed some additional cases to reach a sufficient variety of profiles. Latent Semantic Analysis was used to analyze this material and to calculate correlations between the learner documents and the target learning activities.

To evaluate these results we will use an expert validation. Domain experts will analyze the material and decide about exemptions under a strict exemptions policy and under a more lenient policy. Another measure we are interested in is the time that experts spend to come to a decision because one of the main reasons for our project is to make the APL procedure more efficient. The decisions and the time needed for analysis of the portfolios will be compared to LSA results.

IV. The Corpus and the Software

The final corpus contains 800 documents selected from the course book, other psychology books and Wikipedia articles from the Dutch Wikipedia. Textstat [7] reports 35742 words. The corpus was filtered using a modified Dutch newspaper stop list [8].

For the analysis we followed the optimization procedure described in [6] and decided to use 20 singular values for the analysis, corresponding with 90% of the variance being explained. Visual inspection ("Scree test") of the singular values revealed a steep drop in the size of the singular values as well. We compared the results of analyses using 10, 20 and 40 singular values and found that the analysis with 20 singular values resulted in a.) a sufficient discrimination between the chapters; b.) a high correlation between the chapter and the learner portfolio when there is sufficient thematic overlap and c.) a low correlation when there is no or only a little overlap. For the analysis we used the GTP application by Giles, Wo & Berry [9].

V. Preliminary Results and Outlook

The provisional results are encouraging: portfolios with 'popular psychology' content produced no match. A portfolio of a student who had already finished several psychology courses produced several matches for the subchapters of the book. On the other hand student portfolios with only prior knowledge for one of the chapters (e.g. the chapter about perception) showed only a high correlation to this specific chapter but low correlations for the other chapters.

The current results are limited and provisional in many ways. First, the results need to be validated against expert assessments, where the main question is whether LSA-based decisions are comparable to expert placement decisions. Here, as well as in essay rating, the reliability of expert judgments has to be taken into account. More interesting, however, is whether experts operate by matching documents. For example in one case, a technical description of an experiment, LSA returned no matches. A human expert is capable of inferring prior knowledge. Second, the current analyses are based on the assumption that a one-to-one match exists between a student document and a target document (here a chapter). A more realistic scenario would be that there are several partial matches between student documents and target documents. For example, a student paper that addresses one particular topic would partially match a target document that deals with other topics as well. The type of automatic topic recognition in combination with segmentation of the documents is beyond the scope of our current research.

In this part of the project we only focus on content analysis while we will widen the scope in the future also on the use of metadata and ontologies for prior knowledge assessment. The whole project plan is described in [10].

VI. Acknowledgement

The authors' efforts were partly funded by the European Commission in TENCompetence (IST-2004-02787).

References

[1] Koper, R. & Specht, M. (2007). TenCompetence: Lifelong Competence Development and Learning. In: Sicilia, M.-A. (Ed.). Competencies in Organizational E-Learning: Concepts and Tools. Idea Group.

[2] Merrifield, J., McIntyre, D., & Osaigbovo, R. (2000) Mapping APEL: Accreditation of Prior Experiential Learning in English Higher Education. London. Retrieved July 1, 2006 from http://www.dfes.gov.uk/dfee/heqe/let_final.pdf

[3] Aroyo, L. (Ed.) (2006). Learner Models for Web-based Personalised Adaptive Learning: Current Solutions and Open Issues. Deliverable 1.3. Prolearn Network of Excellence Professional Learning. Internal Document.

[4] Van Bruggen, J., Sloep, P., van Rosmalen, P., Brouns, F., Vogten, H., Koper, R., & Tattersall, C. (2004). Latent Semantic Analysis as a Tool for Learner Positioning in Learning Networks for Lifelong Learning, *British Journal of Educational Technology*, 35(6), 729-738.

[5] Deerwester, S., Dumais, S.T., Furnas, G. W., Landauer, T. & Harshman, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41, 391-407.

[6] Van Bruggen, J., Rusman, E., Giesbers, B. & Koper, R. (2006). Content Based Positioning in Learning Networks. Proceedings of the Sixth International Conference on Advanced Learning Technologies. IEEE Computer Society. 366 – 368.

[7] Hüning, M. (2007.) TextSTAT – Simples Text Analyse Tool. Concordance software for Windows, GNU/Linux and MacOS X. FU Berlin.

[8] Bouma, G., & Klein, E. H. (2001). Volkskrant database nad practicum [Data file]. Retrieved May 12, 2005 from http://hagen.let.rug.nl/~hklein/Volkskrant/Volkskrant

[9] Giles, J., Wo, L., & Berry, M. (2003). GTP (general text parser) software for text mining. In H. Bozdogan (Ed.). Software for text mining, in statistical data mining and knowledge discovery (pp. 455-471). Boca Raton, FL: CRC Press.

[10] Kalz, M., van Bruggen, J., Rusman, E., Giesbers, B. & Koper, R. (2007). Positioning of Learners in Learning Networks with Content, Metadata and Ontologies. *Interactive Learning Environments.* Special Issue on Lifelong Competence Development.