Brick: Mining Pedagogically Interesting Sequential Patterns

ANJO ANJEWIERDEN and HANNIE GIJLERS, University of Twente
NADIRA SAAB, University of Leiden
ROBERT de HOOG, University of Twente

One of the goals of the SCY project (www.scy-net.eu) is to make (inquiry) learning environments adaptive. The idea is to develop "pedagogical agents" that monitor learner behaviour through the actions they perform and identify patterns that point to systematic behaviour, or lack thereof. To achieve these aims, it is necessary to develop methods to identify relevant patterns, and to associate these patterns with feedback to the learner. Here, we present a tool, called Brick, in which pedagogical researchers can interactively specify global patterns (e.g., "which actions occur between running a simulation and providing a correct answer"). The tool then responds with the matching sub-sequences in the data along with various metrics of interestingness (such as frequency and likelihood). Researchers can then determine which patterns are relevant to include in a pedagogical agent and how the agent should respond when a pattern occurs. We describe the tool, the pattern language and metrics used, and provide examples of applying these to log files of a simulation-based inquiry learning environment.

1. INTRODUCTION

We provide a brief illustrated example of how Brick is used to explore patterns from action sequences derived from a simulation-based inquiry learning environment in which learners collaborated in dyads, using a chat channel for communication [Saab 2005]. The environment provided *assignments* that consisted of a question, a list of possible answers and a simulation learners could manipulate. Action types that are extracted directly from the log files are: Ⓢ **start** (an assignment), **simulation** (running a simulation), **correct** (correct answer to a question), **wrong** and ⓔ **end** (of an assignment). The symbol in front of the action type is the "brick" (character) used to build patterns from. Chat communication within dyads has been manually coded capturing the inquiry process [Kuhn et al. 2000]: **planning** (domain-related chat about planning a simulation), **interpretation** (interpreting results of an experiment), **emotional** (non-domain chat about the results), **answer** (discussing what answer to give), **next** (which assignment is next), **off-task**.

2. PATTERN LANGUAGE

Patterns are specified using the following standard notation of regular expressions: **wildcard** (matches precisely a single action), * **any sequence** (any sequence of actions including none), | **alternation**, ? + * **quantification**, and ( ) **grouping**. Negation (!) is often required to filter unwanted matches. For example, finding all sequences in which the first answer provided by the learner is correct, require that there should not be any intermediate wrong answers: Ⓢ (* !) . See the figure for the matching sequences for this pattern.

3. METRICS

Running a pattern often results in many different matching sequences. In order to filter "interesting" short sequences several metrics have been proposed in the literature. In statistical natural language processing sequences of two or more words that correspond to some conventional way of expression are called collocations [Manning and Schütze 1999].

This study was conducted in the context of Science Created by You (SCY), which is funded by the European Community under the Information and Communication Technologies (ICT) theme of the 7th Framework Programme for R&D (Grant agreement 212814). This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content. Author's address: A. Anjewierden, H. Gijlers, R. de Hoog, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands; email a.a.anjewierden@utwente.nl; N. Saab, Leiden University, PO Box 9555, 2300 RB Leiden, The Netherlands; email nsaab@fsw.leidenuniv.nl.

Recently, D'Mello and colleagues [2010] have proposed a *likelihood metric* for the study of behavioural sequences. The general idea behind both these metrics is that sequences that occur much more (or less) likely than chance are the most interesting.

4. RESULTS AND CONCLUSIONS

The figure illustrates the interface pedagogical researchers can use to explore sequences. The pattern , for example, can be used to find the action learners perform after planning an experiment. As the figure shows both the likelihood metric **L(X)** and collocation **Coll** give a high value to the most "logical" matching pattern: . The pattern provides a typical example of learner behaviour: after giving a wrong answer, learners don't enter the inquiry learning cycle again, but either give another answer or start discussing what answer to give next. As with other data mining approaches, for example association rules, the output of the pattern search is not always clear cut and requires interpretation by the researcher.

REFERENCES

D'MELLO, S., OLNEY, A., AND PERSON, N. 2010. Mining collaborative patterns in tutorial dialogues. *Journal of Educational Data Mining 2*, 1–37.

KUHN, D., BLACK, J., KESELMAN, A., AND KAPLAN, D. 2000. The development of cognitive skills to support inquiry learning. *Cognition and Instruction 9*, 285–327.

MANNING, C. D. AND SCHÜTZE, H. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts.

SAAB, N. 2005. Chat and explore: The role of support and motivation in collaborative scientific discovery learning. Ph.D. thesis, University of Amsterdam.

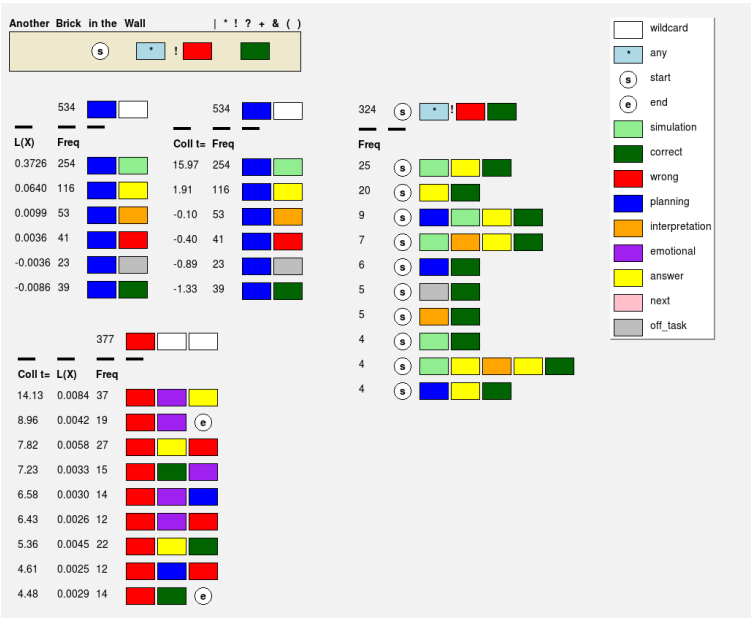

Fig. 1. The Brick user interface consists of three parts. On the right is a palette with bricks representing action types for the given corpus (see Section 1). At the top left is a search box, and above the search box are controls and the meta-characters to create regular expressions from (see Section 2). The largely area displays search results as lists of matching sequences preceded by the values for the metrics. Queries are specified by dragging bricks and meta-characters into the search box.