Learning to Paraphrase: Using Paraphrase Detection of Spoken Utterances to Predict Learner Expertise

Abstract

Predicting student knowledge from text has become an increasingly common for automatic grading and assessment. Much of this work, however, hinges on natural language processing techniques that tend to neglect the relative locations of individual words by using a bag of words model. Instead of using this technique we segment text into a bag of utterances. We then use those utterances to examine the extent to which individuals of different levels of expertise paraphrase one another. Results indicate clear distinctions among paraphrase frequencies of the different levels of expertise. Furthermore, results suggest that experts and novices have many of the same intuitions, but that the expert's knowledge is more accurately applied.

Introduction

Predicting student knowledge from written text has become an increasingly prevalent approach for automatically grading student essays and, occasionally, student speech. Much of this work, however, hinges on natural language processing techniques that tend to neglect the relative locations of individual words by using a bag of words model. Using such an approach overlooks the rich contextual information that can be gleaned from studying the student's work in a more complete form. In order to better utilize this contextual information we use a technique that looks at similarity of student statements at the utterance or sentence level. This is to say that instead of breaking up a student's work into a bag of words, we segment their work into a bag of sentences. In this paper we describe how to use an advanced technique from machine learning and natural language processing, in order to map students utterances into a vector space mapping that can be used to identify similarities in student's work. More importantly we show how identifying utterance similarity across levels of expertise demonstrates meaningful underlying differences in the intuitions that students have about designing. Accordingly, in this paper we will present a summary of the data, methodology and results concerning of our work. Furthermore we will highlight the implications that this work may have on using advanced machine learning and natural language processing techniques for predicting and understanding student cognition and expertise, as well as opportunities to use these forms of analysis for scalable analysis of non-traditional, process-oriented data.

Theoretical Framework

This work is informed by diSessa's (2002) ecology of conceptual change. In his work he describes how students' intuitions from the physical world are often times evidenced in their verbal descriptions of concepts in science, technology, engineering and mathematics. Furthermore, his *knowledge in pieces* framework can be utilized for understanding how novices transition from being a novice to being an expert. Based on his work, becoming an expert is not necessarily about replacing previous intuitions and knowledge, but is, instead, about better understanding when to apply those intuitions, and how to connect them.

In terms of analytical techniques, this work follows in the tradition of a variety of text-based analytics techniques including: discourse analysis (Litman et al 2009, Forbes-Riley et al 2009), content word extraction (Chi et al 2010, Purandare and Litman 2008, Litman et al 2009), sentiment analysis (Craig et al 2008, D'Mello et al 2008, Conati 2009), linguistic analysis (Litman et al 2009, Forbes-Riley and Litman 2010) and automatic essay grading (Chen et al 2010, Rus et al 2009). However, we extend text analytics to a far more complex and higher dimensional space that permits more contextual analysis of the data.

Methods

The data for this study comes from interviews with 18 students from a tier-1 research university. Of the 18 students, 8 were women, 10 were men; 10 were from technical majors, 3 were undergraduates, and 15 were graduate students. There were 3 novices, 9 intermediates, and 6 experts and each interview took approximately 30 minutes. Participants were asked to draw and think aloud (Ericsson and Simon 1980) about how to build various electronic and mechanical devices. The questions were posed in a semi-structured clinical interview format. The main question challenged the student to design a device that could automatically separate, glass, paper, plastic and metal.

The data consisted of audio files, transcriptions of the interviews, and digitized drawings that the students produced during the interview. For the purposes of this paper, we will only be used the transcriptions.

Prior to the interviews, the subjects were labeled as being experts, intermediates or novices in engineering and robotics. This classification was based on previous formal technical training either through a degree program or through a lab course on physical computing. This classification is in accordance with theory that suggests that experts are those that have had extended time practicing their skill (Ericsson, Krampe and Tesch-Romer 1993).

Data Analysis

Transcripts were automatically segmented into individual utterances using a sentence segmentation algorithm in the Natural Language Toolkit (NLTK). They were then subjected to dependency parsing (Klein & Manning 2003) before being passed into a recursive auto-encoder (a type of machine learning algorithm that is based on neural networks). The recursive auto-encoder mapped each sentence into a complex vector space of 100 dimensions (Socher et al 2011). Using this vector space representation of each sentence we computed similarity scores between all utterances in the transcript. Similarity scoring was done using both Euclidean distance and Cosine distance. A given utterance was matched with its closest paraphrase if both the Euclidean distance and Cosine distance metrics identified the same utterance as being the closest. Use of two distance metrics was done to put a lower bound on the quality of the paraphrase. Finally, after we matched each utterance with its closest paraphrase, we computed the relative probability with which utterances from novices, intermediates and experts mapped to one another. In the following section we describe the results of this analysis.

Results

The primary result of this study can be seen in Table 1, which reports the relative probability with which individuals of different levels of expertise paraphrased one another. Recall that the probabilities that we report are for the level of expertise associated with the *most* similar paraphrase for a given utterance. So, to understand the table, one can read across the top row and observe that comments made by novices will most closely match an utterance by another novice 28.9% of the time; most closely match an intermediate's utterance 45.8% of the time; and most closely match an expert 's utterance 25.3% of the time. One will note that the sum of the probabilities for each row is one. Additionally, one should realize that the table is not symmetric because while a given utterance, X, by a novice may be most closely paraphrased by an utterance, Y, by an intermediate, the closest paraphrase for Y, may in fact be some other utterance Z.

Table 1 - Normalized Probability of Paraphrasing Among Differing Levels of Expertise

Expertise	Novice	Intermediate	Expert
Novice	0.289	0.458	0.253
Intermediate	0.268	0.503	0.228
Expert	0.081	0.369	0.551

Discussion

While Table 1 is relatively small in terms of size, the numbers reported represent a number of key concepts from the learning sciences literature. First of all, the reader notes that a novice was more likely to paraphrase an intermediate than they were to paraphrase another novice. This points to the idea that novices tend to possess a broad range of knowledge, that may not always be consistent across all novices. Nonetheless, the ideas expressed by novices appear to be a part of the larger body of knowledge that both intermediates and experts drawn upon. That said, there is a distinction between the knowledge of novices and experts as evidenced in the fact that novices are more likely to paraphrase other novices than an expert.

Among intermediates and experts we see that knowledge, in the form of utterances, follows the ideas of shared referentials (Heath, 2012). More specifically, it is clear that intermediates are most likely to paraphrase other intermediates and experts are most likely to paraphrase other experts. According to Heath (2012), diSessa (2002), and Wenger (1999), as people become more experienced within a domain or community of practice they begin to utilize similar references, intuitions and ideas. They employ a shared body of knowledge, and this is, in part, what defines a community. This is directly observed in the paraphrase based analysis that we completed with this study.

Looking forward, we see a number of potential opportunities to expand and develop this work. First, we think that this work may help motivate more automated analysis of student utterances. More specifically, we envision both supervised and unsupervised machine learning algorithm that can be used to study comparisons between utterances of students. For example, one can consider using the technique in this paper on unlabeled, and then automate the process of finding the correct distribution of paraphrases in order to fit the expected model.

Additionally, we plan to expand our analysis to also look at other engineering design tasks, as well as explore verbal data derived from students as they are physically building. Finally, we will do some human analysis of the paraphrase results in order to validate that the paraphrases that were identified using the combined recursive auto-encoder and dependency parser yielded accurate results.

Conclusion

In this paper we have presented a technique for comparing utterances between students of differing level of expertise. This technique has been adapted from machine learning and natural language processing for looking at student language beyond the traditional bag of words approach. In so doing we identified that this analysis mirrored previous work on expertise by diSessa (2002), Heath (2012) and others. To this effect we found that novices tend to reference a broad set of intuitions, ideas and utterances, a number of which are shared by intermediates and experts. In the same way we observed that both intermediates and experts tend to converge towards shared sets of utterances that are closer to other individuals of similar expertise, as opposed to being most similar to individuals of different levels of expertise. Ultimately we have undertaken this work with the hope of further motivating the use of advanced computational techniques that can provide meaningful analysis of non-traditional data. Moreover, by using these advanced techniques we believe that we can begin to transform the types of scalable assessments being used in today’s varied learning environments, since we see these changes in assessments may be a way to create more equitable learning opportunities for the population at large.

References

CHEN, Y., LIU, C., LEE, C., AND CHANG, T. 2010, An Unsupervised Automated Essay Scoring System. Intelligent Systems, IEEE , vol.25, no.5, pp.61-67, Sept.-Oct. 2010

CHI, M., VANLEHN, K., LITMAN, D., AND JORDAN, P. 2010. Inducing Effective Pedagogical Strategies Using Learning Context Features. In: Proc. of the 18th Int. Conference on User Modeling.

CONATI, C. AND MACLAREN, H. 2009. Empirically building and evaluating a probabilistic model of user affect. User Modeling and User-Adapted Interaction, August, 2009 267-303.

CRAIG, S. D., D'MELLO,S., WITHERSPOON, A. AND GRAESSER, A. 2008. 'Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive-affective states during learning', Cognition & Emotion, 22: 5, 777 — 788.

D'MELLO, S. K., CRAIG, S. D., WITHERSPOON, A., MCDANIEL, B., AND GRAESSER, A. 2008. Automatic detection of learner's affect from conversational cues. User Modeling and User-Adapted Interaction 18, 1-2 (Feb. 2008), 45-80.

DISESSA, A.A. 2002. Why “conceptual ecology” is a good idea. In M.Limón & L.Mason (Eds.), Reconsidering conceptual change: Issues in theory and practice (pp. 29–60). Dordrecht: Kluwer.

ERICSSON, K. A., KRAMPE, R. TH., & TESCH-RÖMER, C. 1993. The role of deliberate practice in the acquisition of expert performance. Psychological Review, 100(3), 363-406.

ERICSSON, K. A. & SIMON, H. A. 1980. Verbal reports as data. Psychological Review, 87, 215-251

FORBES-RILEY, K., AND LITMAN, D. 2010. Metacognition and Learning in Spoken Dialogue Computer Tutoring. Proceedings 10th International Conference on Intelligent Tutoring Systems (ITS), Pittsburgh, PA.

FORBES-RILEY, K., ROTARU, M., AND LITMAN, J. 2009. The Relative Impact of Student Affect on Performance Models in a Spoken Dialogue Tutoring System. User Modeling and User-Adapted Interaction (Special Issue on Affective Modeling and Adaptation), 18(1-2), February, 11-43.

HEATH, S. 2012. Seeing our way into learning science in informal environments. In *Research on schools, neighborhoods, and communities: Toward civic responsibility.* W. F. Tate, Ed. Published for the American Educational Research Association. New York: Rowman & Littlefield Publishers. Pp. 249-267.

KLEIN, D AND MANNING, C. 2003. Accurate Unlexicalized Parsing. Proceedings of the 41st Meeting of the Association for Computational Linguistics, pp. 423-430.

LITMAN, D., MOORE, J., DZIKOVSKA, M., AND FARROW. E. 2009. Using Natural Language Processing to Analyze Tutorial Dialogue Corpora Across Domains and Modalities. Proceedings 14th International Conference on Artificial Intelligence in Education (AIED), Brighton, UK, July.

PURANDARE, A., AND LITMAN. D. 2008. Content-Learning Correlations in Spoken Tutoring Dialogs at Word, Turn and Discourse Levels. Proceedings 21st International FLAIRS Conference, Coconut Grove, Florida, May.

RUS, V., LINTEAN, M., AND AZEVEDO, R. 2009. Automatic Detection of Student Mental Models During Prior Knowledge Activation in MetaTutor. In Proceedings of the 2nd International Conference on Educational Data Mining (Jul. 1-3, 2009). Pages 161-170

SOCHER, R., HUANG, E., PENNINGTON, J., NG, ANDREW, MANNING, C. 2011. Dynamic Pooling and Unfolding Recursive. In Proceedings of the 2011 Advances in Neural Information Processing Systems Conference.

WENGER, E. 1999. Communities of practice: Learning, meaning and identity. New York, NY: Cambridge University.