

Automated Detection of Mentors and Players in an Educational Game

Fazel Keshkar
University of Memphis
Institute for Intelligent Systems
Memphis, TN, USA
fkshkar@memphis.edu

Brent Morgan
University of Memphis
Institute for Intelligent Systems
Memphis, TN, USA
bmorgan2@memphis.edu

Arthur Graesser
University of Memphis
Institute for Intelligent Systems
Memphis, TN, USA
graesser@memphis.edu

ABSTRACT

Automatically identifying the various roles (e.g., mentor, player) in multi-party collaborative chat is a challenging task. To better understand the conversational demands of mentors and players, this paper investigates the dynamics and linguistic features of multi-party chat in the context of an online educational game. In this paper we introduce a novel computational linguistics method using a machine learning algorithm to automatically classify utterances of players and mentors in a serious game, where players act as interns in an urban planning firm and discuss their ideas about urban planning and environmental science in written natural language. Our results are promising and our model can be extended to any multi-party environment that leaders (Mentors) are needed to be distinguished based on their conversation.

Keywords

serious game, natural language processing, machine learning, multi-party chat

1. INTRODUCTION

Individuals in a collaborative learning environment often adapt to specific roles, whether organically or directly assigned. Automatic role detection would provide a crucial step in understanding the impact these roles have in collaborative learning dynamics [2]. Multi-party chat presents an especially challenging task, as the tone is conversational, and distinguishing between roles is relatively difficult (as opposed to email, for example).

Previous research suggests that humans quickly infer a speaker's intentions, often within the first few words of an utterance [3]. Thus, we selected UNIGRAMS and BIGRAMS as our units of analysis. Our results also show that UNIGRAMS perform much better than BIGRAMS. In our automated approach, we investigated the problem as supervised learning in set of features and then we used machine learning algorithms to learn the parameters of the model from annotated training data.

In this paper, we analyzed four chat room conversations between 21 players and two mentors as they interacted within the epistemic game *Urban Science* [4].

In this research, we introduce an automated method to explore a component of natural language processing, using ma-

chine learning techniques, to classify online chat room utterances into one of two categories, player or mentor. The proposed automated method relies on a model that emphasizes the use of the UNIGRAMS or tokens in an utterance to decide: players vs mentors. For instance, chat utterances such as; "What should I do?" and "Please check your inbox.", our automatic approach can detect that first chat said by player and second was posted by mentor.

The resulting models and content feature sets (i.e., UNIGRAMS and BIGRAMS) were tested against an intercept-only model (i.e., the baserate) to determine if it was possible to accurately classify utterances according to the role of the speaker (player or mentor).

2. METHODOLOGY

Participants and Procedure

Twenty-one high school-aged participants and two mentors played Urban Science for ten hours over three days as a part of a week-long Conservation Leadership Pro-gram. Players had no prior experience in urban planning and were recruited by out-reach specialists at the Massachusetts Audubon Society's Drumlin Farm Wildlife Sanctuary. Players conversed with their planning team and a human mentor via a chat window. The chat room corpus contained 1963 total utterances, with 972 and 991 utterances from players and mentors, respectively.

Data Processing and Feature Extraction

Before extracting features, we first addressed the issue of chat-specific terms and emoticons (e.g., "lol", ":)"). These Chat-specific terms and emoticons were treated as individual tokens, as they were intended to convey emotion. Also obvious misspellings (e.g., "Helo") were corrected prior to analysis.

The actual feature vectors were then generated on the basis of this linguistic information by using a "bag of n-grams" approach, i.e. by constructing n-grams (UNIGRAMS and BIGRAMS). In addition to these n-gram counts, we also included punctuation counts, average word length and average utterance length. The part of speech (POS) tagging used the Penn Treebank tagset with some additions specific to the problems related to a chat corpus.

Automated Classification

We used three automated approaches to classify the utterances of players and mentors, each of which utilizes clas-

Table 1: Automated classifier performance for three approaches based on 10-fold cross-validation experiments. Reported: Accuracy, Precision, Recall and F-measure (baseline \cong 50.48%).

	Features	Acc.	Player			Mentor		
			P	R	F	P	R	F
CLASSIFIERS	UNIGRAMS _{svm}	81.03%	0.776	0.872	0.821	0.854	0.749	0.798
	BIGRAMS _{svm}	76.82%	0.723	0.862	0.786	0.833	0.676	0.747
	UNIGRAMS _{nb}	75.58%	0.721	0.835	0.774	0.804	0.677	0.735
	BIGRAMS _{nb}	75.60%	0.711	0.854	0.776	0.822	0.66	0.732
	UNIGRAMS _{j48}	75.65%	0.727	0.814	0.768	0.793	0.70	0.744
	BIGRAMS _{j48}	60.98%	0.562	0.961	0.709	0.874	0.265	0.407

sifiers trained on the dataset. Our classification approach provides us to model both content and context with n-gram features. Specifically, we consider the following two n-gram feature sets, with the corresponding features lowercased and unstemmed: UNIGRAMS and BIGRAMS. Features from the our approaches are used to train Naïve Bayes, Support Vector Machine classifiers, and Decision Tree.

3. RESULTS AND DISCUSSION

The accuracy of the induced classifiers were measured using a 10-fold cross-validation method, under its default setting in Weka [1]. The parameters for our model were chosen for each test fold based on standard cross-validation experiments on the training dataset. Table 1 shows the results of the top scores that we managed to achieve with each of the three classifiers. We also used the combination of features and learner parameters that were determined to give the best accuracy by the classifiers. However the results indicated that feature combination had the highest performance. In Table 1, the ‘‘Features’’ column indicates which features were used, and the following columns indicate the results based on Accuracy, Precision, Recall, and F-measure (Acc., P, R, F) for the two roles, player and mentor.

The resulting models from the automated classifiers were quite successful, as each outperformed the 50.48% accuracy attained by the intercept-only model (i.e., selecting ‘mentor’ for every utterance). However one exception would be the j48 model with BIGRAMS features. Here, we have reduction in performance under the Mentor Recall column in Table 1. The models based on the SVM method performed the best overall, with 81.03% accuracy on UNIGRAMS and 76.82% accuracy on the BIGRAMS feature set.

In our educational game, the conversation between players and mentor are short, an average three tokens. It proves us that BIGRAMS features performed more significant in short conversations, since BIGRAMS features maybe performs lower due to same features in both classes (however, SVM performs better even in BIGRAMS).

Interestingly, the Naïve Bayes approach performs almost 25% more accurately than baseline (\cong 50.48%) on both UNIGRAMS and BIGRAMS feature set. It also performed higher than j48, but only on BIGRAMS features.

Overall, all the standard text categorization approach proposed in Section 2 performed between 10% and 31% more accurately than baseline. However, the best performance overall was achieved by SVM on both UNIGRAMS and BIGRAMS features with accuracy of 81.06% and 76.82%, respectively.

Overall, we have presented some preliminary evidence that

it is possible to automatically classify individuals’ roles in multi-party chat within the context of the serious game Urban Science.

However, it remains to be seen whether these findings will generalize to other serious games. Furthermore, the specific individual roles in this context were preassigned; when peers interact in a collaborative learning environment without predefined roles, certain roles or personalities may arise organically.

Although this presents a significantly more challenging task, it represents a critical step in understanding interaction dynamics in a collaborative learning environment.

Acknowledgments

This work was funded in part by the Macarthur Foundation, National Science Foundation (REC-0347000, DUE-091934, DRL-0918409, DRL-0946372, BCS 0904909, DRK-12-0918409), the Gates Foundation, and U.S. Department of Home-land Security (Z934002/UTAA08-063).

4. REFERENCES

- [1] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: An update; SIGKDD explorations, , volume 11, issue 1. 2009.
- [2] F. Keshtkar, C. Burkett, A. Graesser, and H. Li. Detecting players personality behavior with any effort of concealment. In *A. Gelbukh (Ed.): CICLing 2012, Part II, LNCS 7182*, pages 503–515, Springer-Verlag Berlin Heidelberg 2012, 2012.
- [3] C. Moldovan, V. Rus, and A. C. Graesser. Automated speech act classification for online chat. In *Proceeding of FLAIRS 2011*, 2011.
- [4] D. Shaffer, N. Chesler, G. Arastoopour, and DAngelo. Nephrotex: Teaching first year students how to think like engineers. Washington D.C., 2011. Laboratory Improvement (CCLI) PI Conference.