Context-based Speech Act Classification in Intelligent Tutoring Systems

Borhan Samei[1], Haiying Li[1], Fazel Keshtkar[2], Vasile Rus[1], Arthur C. Graesser[1]

[1]University of Memphis, Institute for Intelligent Systems, TN, USA
{bsamei,hli5,vrus,graesser}@memphis.edu
[2]Southeast Missouri State University, MO, USA
{fkeshtkar}@semo.edu

Abstract. In intelligent tutoring systems with natural language dialogue, speech act classification, the task of detecting learners' intentions, informs the system's response mechanism. In this paper, we propose supervised machine learning models for speech act classification in the context of an online collaborative learning game environment. We explore the role of context (i.e. speech acts of previous utterances) for speech act classification. We compare speech act classification models trained and tested with contextual and non-contextual features (contents of the current utterance). The accuracy of the proposed models is high. A surprising finding is the modest role of context in automatically predicting the speech acts.

Keywords: speech act· machine learning· intelligent tutoring systems

1 Introduction

Speech act classification is one of the indispensable components of dialogue-based intelligent tutoring systems (ITS) because speech act categories dramatically constrain the system's response [1, 2]. For example, when a student asks a question, the system should respond very differently than when the student asserts a fact or expresses being lost. Speech act classification is used for detecting students' intentions (Is the student asking a question or asserting a fact?). More precisely, speech act classification is framed as a classification task in which the goal is to detect the speech act categories of a given utterance from a predefined set of categories that together form the speech act taxonomy. The speech act taxonomy is usually predefined by researchers although attempts to automatically discover it from data are emerging [3]. We used a predefined taxonomy in the present paper [4].

The models in this paper will be incorporated in a multiparty simulation game on urban planning, called Land Science, an expansion of Urban Science [5]. The previous model of speech act classification in Land Science relied entirely on the lexical, semantic, and discourse features of the individual utterances without considering previous utterances within the context [3,4]. However, conversation progresses dependent on the previous utterances or context. For instance, after a greeting a greeting is more likely. Therefore, this study aims to investigate the role of context in speech act classification.

Speech act classification has theoretical roots in Austin's language as action theory [6] and subsequent work by Searle [7,8]. Different speech act taxonomies have been used in different domains of application. D'Andrade and Wish proposed seven categories of speech acts with high inter-annotator agreement among human judges: assertions, questions, requests and directives, reactions, expressive evaluations, commitment, and declaration [9].

Researchers have proposed several other taxonomies that are sensitive to various tasks and knowledge domains. Rus et al (2012) developed a data-driven method for automatically discovering speech act categories from online chats that were extracted from educational games, Urban Science and Land Science [3]. They applied utterance clustering methods based on the content of utterances and tried to find the natural groupings of the utterances in a fully automatic approach. The clusters were then deemed as speech act categories by assigning semantic names to the automatically discovered clusters.

Rasor et al (2011) proposed a machine learning approach using decision trees to automate the speech act classification in student chat interactions [10]. Olney et al. (2003) proposed a rule-based approach to classify speech acts by focusing on 16 categories of questions [11]. The Question category is important in an ITS because the tutor/mentor is expected to give answer to students' questions. Therefore, the first step is to identify questions in student utterances.

Moldovan et al. (2011) developed automated speech act classification for Land Science epistemic game [4]. The categories of their taxonomy included the same seven categories as Rus et al. [3]: *Statement, Request, Reaction, Metastatement, Greeting, ExpressiveEvaluation, and Question.* Using a supervised machine learning approach, they examined several models with feature sets containing the 2-8 leading tokens of the utterance and found that using 3 leading tokens achieves more accurate results. Based on their approach, our model uses the two leading tokens, the last token, and the length of utterance as features and we used the same taxonomy.

2 Method

Our approach to speech act classification is a supervised machine learning approach. In supervised machine learning approach, models of the tasks are proposed as sets of features. Parameters of these models are learned/trained from annotated data and the performance of the learned models is then assessed on new, test data. The parameters of the proposed models are learned using several machine learning algorithms, i.e. decision trees and naïve Bayes.

The feature set used in our models was designed based on two principles: first, it is intuitively inferred and tested that human identified the speech act of an utterance as soon as they heard the first few words [4], namely, the first leading tokens. However, the context of an utterance is assumed to improve the accuracy. Thus, another feature set included the contextual information, e.g. speech act category of the last few utterances. Our model adds context to previous models that relied merely on the contents of current student utterance [4].

Briefly, our feature set consists of content (non-contextual) features of the current utterance and contextual features (speech acts and speaker of previous utterances). The non-contextual features include the first two tokens and the last token which were represented as the actual string of characters (tokens) and the length of the utterance in words. The contextual features captured contextual information with the five prior utterances (the speech acts and actual speakers of these utterances). Our taxonomy consisted of seven categories. Table 1 shows examples extracted from the actual utterances for each category.

Table 1. Speech act taxonomy of seven categories with examples.

Speech act category	***Example from dataset***
ExpressiveEvaluation	Your stakeholders will be grateful!
Greeting	Hello!
MetaStatements	oh yeah, last thing.
Statement	a physical representation of data.
Question	What should we do?
Reaction	Thank you
Request	Please check your inbox

Our training data was extracted from a dataset of mentor-student chat utterances from seven Land Science games. A total number of 26,148 chat utterances were generated by the players and the mentor. We randomly extracted chat utterances to form our training data and adjusted the training data to include an even distribution of 30 instances per speech act category.

This data set was annotated by one human expert within the context of the chats. The human expert had access to the whole dialogue and context of the conversation. This annotated data set is deemed as the reference annotation and includes 30 utterances per speech act category.

In order to examine the impact of the limited contextual information defined in our automated models (speech acts of previous five utterances), the data set was further annotated by a second human judge in two forms. First, the utterances were randomly ordered and the rater annotated them without considering the limited context. Second, each utterance was accompanied by the speech act category (not the content) of five prior utterances and rater annotated the data considering the contents of the current utterance and prior context.

In the first form of annotations, the rater showed a kappa of 0.55 in agreement with reference annotations. The agreement with reference annotations was improved to 0.75 kappa when the rater was provided with contextual information. On the other hand, the agreement of the rater with himself on the two forms of annotations (with/without context) was about 0.6 *kappa* which implies that having some sort of information about context, changes human’s judgments and improves their accuracy compared to reference annotations.

Using the reference annotation data set, we applied J48 decision trees and Naïve Bayes machine learning models to create the automated speech act classifier with different feature sets of contextual and semantic information to examine the role of context. The performance of our models is presented in next section.

3 Results

Based on the human annotation, having contextual information improves the accuracy of human judgments. In fact, the more we know about context the better we can make decisions. Our feature set consists of two types of features: A set of 10 features which represent the context of the utterance by looking at the speech act category and speaker of five prior utterances (contextual features), and 4 features representing the semantic information of the individual utterances including the first two tokens, last token, and the length of the utterance (semantic features). The performance of proposed models was tested with feature sets of contextual, semantic and both.

Using the reference annotations as our training data, we created J48 decision trees and Naïve Bayes learning models using WEKA [12] and we tested our models with 10-fold cross validation. The overall performance of models was evaluated with the three feature sets (contextual, semantic, and semantic & context).

Table 2. Overall Accuracy and Kappa statistics of Naïve Bayes and J48 decision tree models with different feature sets.

	J48 decision tree		***Naïve Bayes***	
Feature set	Accuracy (%)	Kappa	Accuracy (%)	Kappa
Contextual	23.80	0.11	37.14	0.26
Semantic	55.71	0.48	53.80	0.29
Contexual & Semantic	56.19	0.48	54.76	0.47

As seen in Table 2, using only contextual features provides enough clue to predict the speech act categories with an accuracy of about 37% with Naïve Bayes model. The semantic features improve the accuracy of J48 model to 55%, with 0.48 kappa. Using both kinds of features together, surprisingly, showed a low impact on the performance. Adding context to semantic feature set improved Naïve Bayes algorithm while the performance of the J48 model did not change by adding contextual features.

Overall, J48 model had better performance. To take a closer look at the role of context in our models, we examined the performance of J48 models on predicting each of the speech act categories. Table 3 shows the precision and recall on each category for models with different feature sets.

Table 3. The performance of J48 models on predicting each speech act category with different feature sets.

	Contextual		Semantic		Cont. & Sem.	
Category	*Precision*	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*
Expressive Evaluation	0.22	0.30	0.35	0.93	0.35	0.93
Greeting	0.36	0.43	0.73	0.63	0.73	0.63
Metastatement	0.30	0.36	0.64	0.60	0.60	0.56
Question	0.20	0.20	0.76	0.63	0.63	0.70
Reaction	0.08	0.06	0.50	0.13	0.13	0.21
Request	0.18	0.13	0.70	0.46	0.50	0.60
Statement	0.23	0.16	0.62	0.50	0.53	0.58

As shown in Table 3, adding contextual features to the semantic feature set improves the recall on some categories, such as Question, Reaction, Request, and Statement, whereas the precision on the categories gets lower by adding context. Overall adding context to the feature set had a modest impact on the performance of models.

Conclusion

In this paper, we examined the role of context (i.e., prior speech act categories and speakers, but not the actual content) in the performance of automated speech act classification. Contextual features seem to not have a significant impact on the overall performance of models; however adding context improves the performance on certain categories.

The results presented in previous sections showed that having some sort of contextual information has a positive impact on the accuracy of speech act classification for both human and computer. The models presented in this paper can be improved with having a larger training data and adjusting the features sets. The taxonomy also can be modified to multi-layer structure which enables the use of multiple feature sets to maximize the accuracy on certain categories.

For future work, we plan to test our model on different and new data sets once available. The models can be applied to different domains to explore the possible improvements. We will also investigate different types and representations of contextual features which can be used in the System to improve the accuracy.

Acknowledgements. This work was supported by the National Science Foundation BCS 0904909 and DRK-12-0918409. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these funding agencies, cooperating institutions, or other individuals.

References

1. Graesser, A. C., D'Mello, S. K., Hu. X., Cai, Z., Olney, A., Morgan, B.: AutoTutor. In: McCarthy P., Boonthum-Denecke. C. (eds.), Applied natural language processing: Identification, investigation, and resolution, pp. 169-187. IGI Global, Harshey PA (2012).
2. Rus, V., D'Mello, S., Hu, X., Graesser, A.C.: Recent Advances in Intelligent Systems with Conversational Dialogue., vol. 34 , pp. 42-54. AI Magazine (2013).
3. Rus, V., Graesser, A. C., Moldovan, C., Niraula, N.: Automated Discovery of Speech Act Categories in Educational Games. In: Yacef, K., Zaïane, O., Hershkovitz, H., Yudelson, M., Stamper, J. (eds.) Proceedings of the 5th International Conference on Educational Data Mining, pp. 25-32. (2012).
4. Moldovan, C., Rus, V., and Graesser, A. C.: Automated Speech Act Classification For Online Chat. In: The 22nd Midwest Artificial Intelligence and Cognitive Science Conference. pp. 23-29. (2011).
5. Shaffer, D.W., Gee, J.P.: Epistemic Games as Education for Innovation. BJEP Monograph Series II, Number 5-Learning through Digital Technologies. vol. 1(1), pp. 71-82. (2007).
6. Austin, J. L.: How To Do Things With Words, Cambridge, Harvard University Press (1962).
7. Searle, J. R.: Speech acts: An essay in the philosophy of language, Vol. 626, Cambridge, Cambridge University press (1969).
8. Searle, J. R.: A Taxonomy of Illocutionary Acts, In: Gunderson, K. (ed.) Language, Mind and Knowledge, pp. 344-369., Minneapolis (1975).
9. D'Andrade, R.G., Wish, M.: Speech Act Theory in Quantitative Research on Interpersonal Behavior. In: Discourse Processes, vol. 8(2), pp. 229-258. (1985).
10. Rasor, T., Olney, A., D'Mello, S. K.: Student Speech Act Classification Using Machine Learning. In: Proceedings of the International Florida Artificial Intelligence Research Society Conference, AAAI Press. (2011).
11. Olney, A., Louwerse, M. M., Mathews, E. C., Marineau, J., Mitchell, H. H., Graesser, A. C.: Utterance classification in AutoTutor. Building Educational Applications using Natural Language Processing. In: Burstein, J., Leacock, C. (eds.) Proceedings of the Human Language Technology - North American Chapter of the Association for Computational Linguistics Conference Workshop, pp. 1-8. Association for Computational Linguistics, Philadelphia (2003).
12. Hall, Mark, Frank, Eibe, Holmes, Geoffrey, Pfahringer, Bernhard, Reutemann, Peter, Witten, Ian H.:The WEKA Data Mining Software: An Update, SIGKDD Explorations, Volume 11, Issue 1 (2009).