A Tool for Speech Act Classification Using Interactive Machine Learning

Borhan Samei
The University of Memphis
365 Innovation Dr.
Memphis, TN, USA
bsamei@memphis.edu

Fazel Keshtkar
The University of Memphis
365 Innovation Dr.
Memphis, TN, USA
fkshtkar@memphis.edu

Arthur C. Graesser
The University of Memphis
365 Innovation Dr.
Memphis, TN, USA
a-graesser@memphis.edu

Vasile Rus
The University of Memphis
365 Innovation Dr.
Memphis, TN, USA
vrus@memphis.edu

ABSTRACT
In this Demo, we introduce a tool that provides a GUI interface to a previously designed to Speech Act Classifier. The tool also provides features to manually annotate data by human and evaluate and improve the automated classifier. We describe the interface and evaluate our model with results from two human judges and Computer.

Keywords
Speech act, interactive machine learning, Speech act classifier

1. INTRODUCTION
Speech act classification is the task of classifying a sentence, utterance or any other discourse contribution, into a speech act category which is selected from a set of predefined categories. Each category represents a particular social discourse function. "What is your name?" for example, is classified as a Question. There are other examples of speech act categories, such as Statement, Greeting, etc.

Our tool is designed to offer annotation facilities in order to improve a previously developed automated speech act classification (SAC; the SAC is available online at www.cs.memphis.edu/~vrus/SAC/) [2]. Furthermore, we provide a GUI-based interface to the speech act classifier [2]. We use an interactive machine learning model for this task that allows for manual classification by human judges which is used to improve the accuracy of our machine learning model. The tool is created and written in Java. The SAC relies on decision tree (J48) that has proved to provide based performance on training data from human annotated utterances [2].

The decision tree is a machine learning approach that requires a feature set to be designed. The feature set is an important part of machine learning algorithms. Moldovan, Rus, and Graesser [2] designed a feature set and used it in order to automatically classify chat utterances.

They used eight speech act categories which are shown in Table 1. According to analyzes on a variety of corpora, such as chat and multiparty games we can converge on a set of speech act categories that are both theoretically justified and can be used by trained judges [3]. For the feature set we tokenize the chat utterances based on basic regular expressions and for each utterance five features are extracted: the first three words, the last word, and the length of the sentence in words. Many other feature sets have been experimented with but the five features just mentioned proved to lead to highest performance in conjunction with decision trees and naïve Bayes methods [2]. The model has been trained and experimented with on data sets from intelligent tutoring systems as well as chat data [2]. Our model is a J48 model built on the training the data using Weka toolkit.

The model and training data can be updated and improved independently. This tool can be used with several training data based on the domain. For example, if we are looking at the dialogs in a movie we can use a different training data model based on this domain.

Table 1. Set of speech act categories and an example of each category.

Speech act category	Example
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
Other	ed is tough, no doubt.

2. THE INTERFACE
We have designed a Graphical User Interface (GUI) for the tool. It can be used on any machine, since it is implemented in Java. Figure 1 shows a snapshot of the starting interface of the tool. Use is able to "Run" or "Annotate" the input data (see Figure 1).

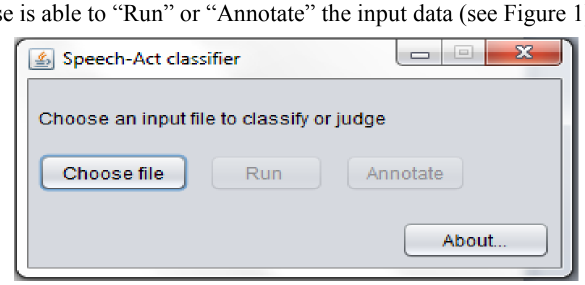

Figure 1. A snapshot of the starting window of the tool's interface.

By clicking on "Run" the tool will start to classify the input file which contains a collection of utterances. After classifying the utterances, the output will be saved as an excel file. The users can also click on "Annotate" to annotate the data manually. By clicking on "Annotate" a new GUI will appear (Figure 2) which contains the utterances. The user will see 10 utterances in each step and for each utterance there is a drop down list of categories from which the user selects one. After annotating by user can go to next utterances or save the current annotation any time and do the rest after getting back.

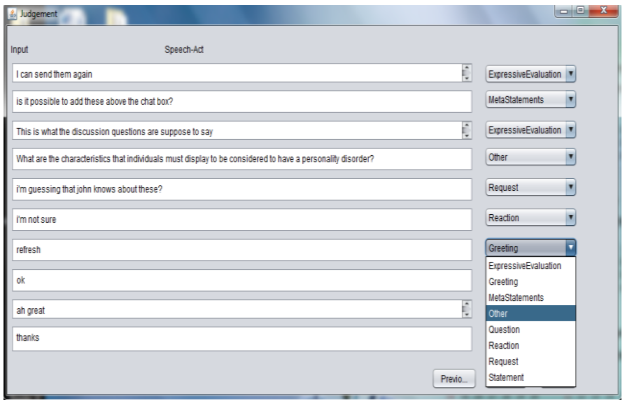

Figure 2. A snapshot of the manual annotation tool.

3. RESULTS

A collection of chat utterances are used as the data set to evaluate the algorithm. The system is trained by a collection of datasets derived from Auto-Mentor frame board dataset. The test data is chosen from a collection of new chat data. We have chosen a hundred chat utterances from the data and tried to maintain a normal distribution on the speech act categories so each category has 10 to12 utterances in our test data. The test data is also annotated by two human judges.

The system runs the algorithm on the test data and for each utterance we show top three speech-act categories based on their probability distribution in the decision tree. Top three categories are the ones with the highest probability and we represent these 3 categories as Comp1, Comp2, and Comp3. Figure 3 represents the agreement of Human judges with Comp1, Comp2, and Comp3.

As our model is based on interactive machine learning we tend to compare automated classification to human judges in order to improve the model and retrain with new enhanced training data. To do this, we have calculated agreement among humans and computer. For each utterance we have five output categories, the top three assigned by our model, and the annotations by human judges. These five outputs are compared to investigate their agreement and evaluate the current model. We have looked at agreement both by Speech-Act categories and overall among the dataset.

Table 2 shows the overall agreement among the classifiers. Our human judges agree on 70.00% of the utterances. Agreement of our model and human judges is about 50%. The agreement of judges with Comp2 and Comp3 are less than 5% of the human agree with the second and third category computed by our model. As it was mentioned earlier Comp2 and Comp3 are actually the two categories that our model would propose based in their probability in the decision tree. This result helps improve our model by increasing the probability of the right categories based on human judge annotations, in the decision tree. We will also look at the agreement by.

Table 2. Agreement among judges and classifiers.

Judges	Agreement
Human1 – Human2	70%
Human1 – Comp1	50%
Human2 – Comp1	47%
Human1 – Comp2	4%
Human2 – Comp2	3%
Human1 – Comp3	2%
Human2 – Comp3	3%

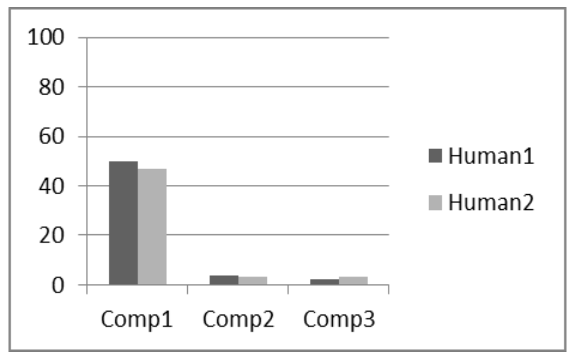

Figure 3. Overall agreement of human judges with automated classifier.

4. CONCLUSION

As the results show, our model is working close to human judges. The tool can be used to improve the model by taking both human and computer annotations and enhance the training data. The main goal of this tool is not only to automate the classification task, but also provide more features to improve the classifier. Both automated and manual annotations are easy to use by the interface and this can be used in several applications and domains. The training data and J48 models can be externally changed for different domains.

5. References

[1] Bagley, E., 2011 , Stop Talking and Type: Mentoring in a Virtual and Face-to-Face Environmental Education Environment. Ph.D. thesis, University of Wisconsin-Madison

[2] Moldovan, C., Rus, V., Graesser, A.C. 2011, Automated speech act classification for on-line chat. In: Proceeding of FLAIRS 2011

[3] R.G. DAndrade, M.W., 1985, Speech act theory in quantitative research on inter-personal behavior. In: Discourse Processes. 8 (2), 229-259