Speech Act Classification with Cross Training and Domain Adaptation

Atul Mohan
Graduate Student
Indiana University
atumohan@iu.edu

Chitesh Tewani
Graduate Student
Indiana University
ctewani@iu.edu

Prashanth Balasubramani
Graduate Student
Indiana University
prasbala@iu.edu

Abstract

It is often useful to classify conversational data into speech acts, as speech act analysis is important for digital assistant systems where the speech act of an utterance is linked with the intention of the end user. It is also pivotal in identifying the character of a person based on his conversations on the social media. This paper presents experimental results of speech act classification on two major dialog act corpus using the Support Vector Machine (SVM) representative model. We also perform cross validation of data and draw a comparison between the results of the various scenarios tested.

1 Introduction

There are many forms of dialogues in which the speakers involve in goal oriented utterances. Each of these utterance can symbolize the intention of the speaker in that particular context. These utterances can be categorized and represented in total as speech acts. We consider three separate instances of conversations where each of them is tagged with its corresponding speech act:

Do you think it is possible? - Question
I may be wrong but I feel that the process would not complete - Hedge
Sorry for the trouble caused - Apology

Speech act classification finds itself a major role in digital assistant systems such as SIRI where the system can effectively respond to user commands only if it understands the intention of the user, which in turn can be identified from the speech act of the utterance. Speech act classification can be challenging based on the fact that a large percentage of conversations are context based, due to which classification of a specific utterance may be dependent on the speech acts of the previous or succeeding conversations. The corpus used for this paper are not audio conversations, but rather their transcribed form. This may not capture the intention of the every utterance in the conversation as certain utterances can be categorized only by analysing the speaker tone and body language. The textual conversations provides a good amount of data to work with and the analysis performed has more relevance to identifying speech acts in social media conversations such as Twitter.
In the following chapters, we present both the datasets used as part of our analysis followed by the methodology used for classification and the results of our experiment.

2 Literature review

Research in Speech Act or Dialogue Act recognition techniques is popular in conversation analysis. Major work in this field is based on the conversation corpus annotated in start of 1990. The work of "Dialogue act modeling for automatic tagging and recognition of conversational speech" by Andreas et.al includes stochastic sequential models like Hidden Markov Models (HMM) treating conversation utterances as Markov Chains. The domain to recognize speech acts have expanded in recent years. In the paper, "Semi-supervised Speech Act Recognition in Emails and Forums" by Jeong and et.al, discusses about extracted features such as phrases and dependency trees, called sub- tree features, for semi-supervised learning. Until recently, dialogue act recognition is experimented over Twitter conversation. A. Ritter and et.al in "Unsupervised Modeling of Twitter Conversations" extended a conversation sequence model to separate topic and dialogue words, resulting in an interpretable set of automatically generated dialogue acts. These discovered acts have interesting differences from those found in other domains, and reflect Twitters nature as a micro-blog. Using different classification model, Riloff and et.al in “Classifying Sentences as Speech Acts in Message Board Posts”, modeled a speech act classifier using lexical and syntactic features, speech act word lists from external resources, and domain-specific semantic class features.

3 Dataset review

For the purpose of the analysis, we have identified and used two dialog act corpora: The ICSI Meeting Recorder Dialog Act(MRDA) corpus and The Switchboard Dialog Act corpus.

MRDA:

```
Example 6: Bmr027
    2049.340-2051.730   c5   qy^rt      did i say that ?
Example 7: Bmr027
    1836.000-1838.580   c4   qy^bu^rt   didn't they want to do language
                                        modeling on you know recognition
                                        compatible transcripts ?
Example 8: Bmr012
    6.805-17.875        c1   qy^rt      is this channel one ?
```

Figure 1: Example of MRDA corpus

The MRDA corpus has over 180,000 hand annotated dialog act tags for approximately 72 hours of speech which has been extracted from over 75 naturally occurring meetings.

SWBD:

```
sv    A.35 utt1: I think that's part of it too. /
sv    A.35 utt2: {C But } I do think, -/
qy    B.36 utt1: {E I mean } do you think,
x     A.37 utt1: .
+     B.38 utt1: people really need two cars and --
nn    A.39 utt1: No, /
nn^r  A.39 utt2: no. /
sd^e  A.39 utt3: # I don't. # /
+     B.40 utt1: -- # a house # in the suburbs {C or, } -/
nn    A.41 utt1: No, /
sd^e  A.41 utt2: I don't think that. /
```

Figure 2: Example of SWBD corpus

The SWBD corpus consists of over 2400 two sided telephone conversations among 543 speakers. The entire corpus has approximately 220,000 conversation lines all of which are hand annotated with the corresponding dialog act tags. Comparison of the tags across both the corpora shows that there are several overlapping tags as well as independent tags specific to each corpus. Following table shows the correspondence between the SWBD and the MRDA dialog tags.

Both these corpora has various other intersecting and non intersecting tags, but for our analysis, the tags documented below are used.

TAG	SWBD	MRDA
Yes-No question	qy	qy
Statement-opinion	sv	s
Wh-question	qw	qw
Open-question	qo	qo
Yes answers	ny	aa
Hedge	ny	Undefined
No answers	nn	ar
Thanks	ft	ft

Figure 3: Comparison of relevant tags between SWBD and MRDA

4 Classification process

This paper majorly aimed to develop and analyze models to classify speech acts in Switch Board Corpus and Meeting Recorder Dialogue Acts datasets. Using Supervised Learning, a machine learning technique, for inferring a function from labeled training data, we modeled our classifiers. The classifiers we built used Support Vector Machines.

4.1 Support Vector Machines

Support vector machines (SVMs) are a set of supervised learning are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. A support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

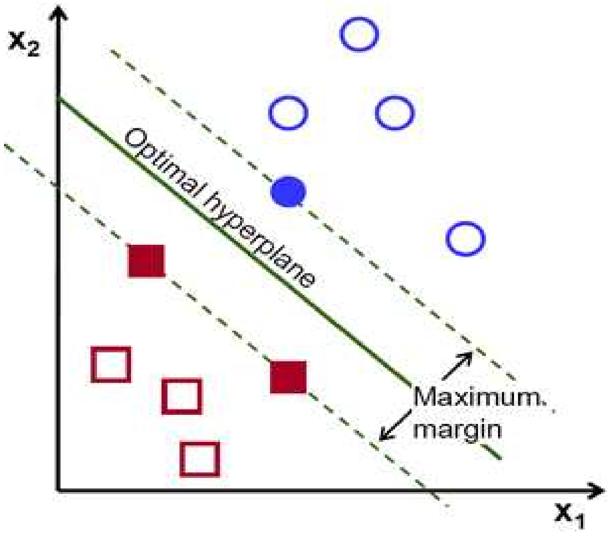


Figure 4: SVM hyper plane for binary classification

Inherently, SVM is a two class classifier, i.e. it builds a model on binary classification of data. For classifying Speech Acts in the corpuses, we need to use Multi-class classification. The traditional way to do multiclass classification with SVMs is to use the general strategies by reducing into multiple binary classification problems. The following are the two methods:

4.1.1 One vs One

In the one-vs.-one reduction, one trains $K(K-1)/2$ binary classifiers for a K-way multiclass problem; each receives the samples of a pair of classes from the original training set, and must learn to distinguish these two classes.

4.1.2 One vs Rest

The one-vs.-rest (or one-vs.-all, OvA or OvR) strategy involves training a single classifier per class, with the samples of that class as positive samples and all other samples as negatives. This strategy requires the base classifiers to produce a real-valued confidence score for its decision, rather than just a class label; discrete class labels alone can lead to ambiguities, where multiple classes are predicted for a single sample.

In this paper we used both strategies of multiclass classification with varying parameters. The results are described in the following section.

4.2 Feature Selection for Support Vector Machines

In context of Machine Learning, a feature in machine learning task is considered an individual measurable property of a phenomenon being observed. A feature vector is an n-dimensional vector of numerical features that represent some object. Many algorithms in machine learning require a numerical representation of objects, since such representations facilitate processing and statistical analysis.
In this paper, we developed an incremental model for selecting features Feature selection is based on the following categories

4.2.1 Bag of Words (BoW)

In this model, a text (such as a sentence or a document) is represented as the bag (multiset) of its words, disregarding grammar. In context of Text mining, Bag of Words strategy is also known as Bag of n- grams. Steps

1. Parse all the conversation and collect words in the sentences

2. Create a word space of all the words

3. Remove stop words from the word space. Feature vector now would be of length of remaining words in the space.

4. For each conversation, create a feature vector of all zeros

5. For each words in the conversation, set the count in the feature vector in its position in the word space

4.2.2 Hand crafted features

In this paper, along with Bag of Words, we trained the Support Vector Model on our set of additional hand crafted features. The hand crafted features were chosen by finding patterns in the conversations and intuition of improving classification for few speech acts. The hand crafted features were:

1. Presence of Tri-grams and Bi-grams Few tri-grams like “I don‘t know”, “I don‘t think”, “It may not” etc. and bi-grams like “may be”, “not sure”, “don‘t know” etc. boost the classification process and avoid misclassification.

2. Presence of Thanking and Apology words Tagging words which expresses thanking and apology

3. Interposition of utterance in the conversation Every conversation closes with sentences expressing “Good Bye” and wishing well. Computing the inter-position of the utterances within the conversation improves the classification of such speech acts.

4. Dependence on Parts-of-speech Patterns and sequence of parts of speech in the conversation can help in tagging speech acts.

5. Dependence on the previous speech act Some speech acts are highly likely to follow a previous speech act. E.g. Words like “Welcome” are usually followed by “Thanking words”.

An overview of the classification process:

1. Read training and test data

2. Normalize the data into classes needed to classify (as in the input they are classified in one of the 226 tags)

3. Create feature vectors for each conversation utterance

4. Run the One vs. Rest classifier SVM with fine tuned parameters

5. From the trained model, test on the test data

5 Experiment

Cross training is a methodology to train the classifier over data in a domain and test on the data from a very similar domain. To compare classifiers for Speech Acts, we created different experimental setting.

CASES	TRAIN SET	TEST SET
1	SWBD	SWBD
2	MRDA	MRDA
3	MRDA TRAIN + SWBD	SWBD
4	SWBD TRAIN + MRDA	MRDA
5	MRDA TRAIN + SWBD TRAIN $< SHUFFLE >$	MRDA TEST + SWBD TEST $< SHUFFLE >$

Figure 5: Data scenarios that were tested on the developed model

6 Results and Analysis

6.1 Classification on Switchboard corpus

By using the modelled SVM classifier, we were able to achieve higher accuracy with the SWDA dataset. When the classifier was run using the BOW (Bag of Words) feature, the accuracy range went up to 85%. This accuracy score was directly related to the training set because larger the training set, larger the size of BOW feature. In order to improve the accuracy even further, as explained we included our own set of crafted features and added the same with BOW feature. This custom setting yielded us an accuracy of 89% which is an increase of 4% from the initial accuracy. As a result, we were able to infer that adding custom features to a static feature increases the classification accuracy. The classification of individual classes can be seen below:

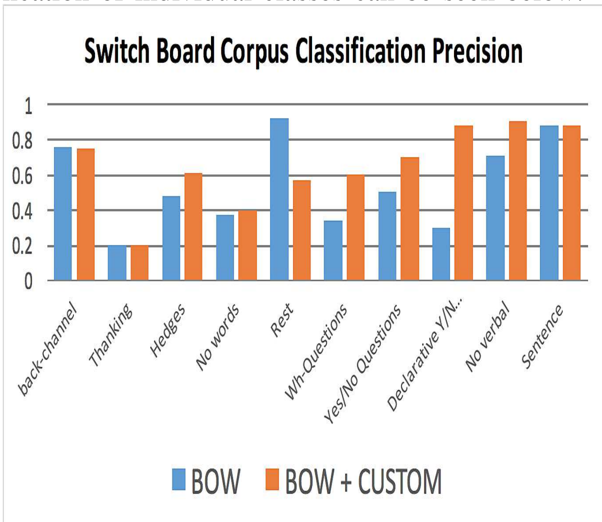


Figure 6: Results on SWBD corpus

ments across different feature set is shown below:

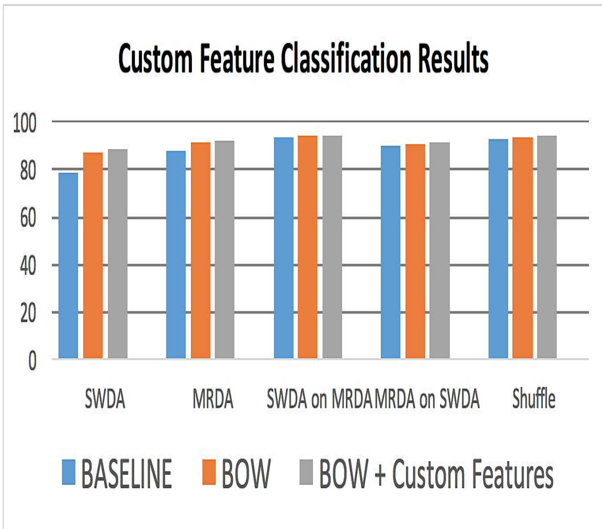


Figure 8: Results on custom training sets

6.2 Classification on Meeting Recorder Corpus

In the case of the MRDA dataset, the same SVM classifier with BOW feature alone was able to classify at an accuracy of 91%. There is a certain relation between the classes of SWDA and MRDA datasets. The relation is that all the classes designed for the MRDA corpus were derived from that of the SWDA corpus. As the MRDA corpus followed the same class structure as that of SWDA, including our own set of crafted features yielded the same effect as it had in SWDA. The accuracy of the classification increased by a considerable amount. The resultant distribution of individual classes in the MRDA Classification is given as follows:

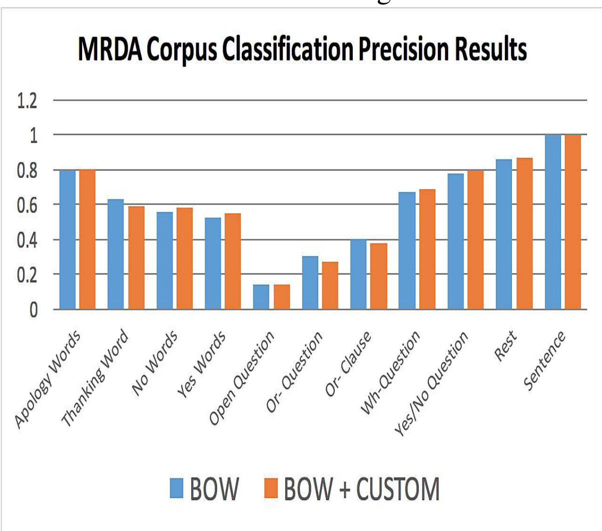


Figure 7: Results on MRDA corpus

6.3 Custom training sets

Based on the experiments, we were able to make our mode Domain Adaptive. Domain Adaptation is a term in machine learning where you train a model on a certain domain and test it on different domain. As explained, we did experiments by trying different combinations of training sets which helped us in building our classifier to be domain adaptive. Some of the training combinations such as including MRDA training data with SWDA training data and building a model based on that. Similarly, we tested the same by including SWDA training set with MRDA data and building the classifier on the same. The results on the custom experiments across different feature set is shown below:

6.4 Overall analysis

In summary, our experiments confirmed that BOW and custom crafted feature based modeling can improve Speech act classification accuracy quite substantially in principle, at least for certain speech act types, but that the skewed distribution of speech acts limits the usefulness of the approach on the Switchboard corpus. The benefits of custom feature modeling might therefore be more pronounced on corpora with more even speech act distribution, as is typically the case for task-oriented dialogues. This suggests that even in task- oriented domains more research is needed to realize the potential of Speech Act modeling.

7 Future Work

7.1 Domain Adaptation

In Machine learning, the term Domain Adaptation refers to creating a model in a certain domain and testing the same on a different domain. This term is very useful if you want to build a generic classifier which will work across different domains. The most important aspect of making a classifier Domain Adaptive, the training data should vary and should include data across several domains. In our experiment, we have built a SVM based classifier through training data across 2 different domains (Switch Board and Meeting Corpus). This certain setting has a positive effect on the classifier and makes it Domain Adaptive. Further, the extension of our work is to test our custom classifier on a new Social Media Domain(Twitter).

7.2 Twitter Domain

The Extension of our work is to implement the same model and test it on a different domain. If we can build a classifier with good domain adaptability capacity, it would help out the classification of several domains using a single classifier which will be domain adaptive to a greater extent. We selected the twitter domain to be our future work for classification.

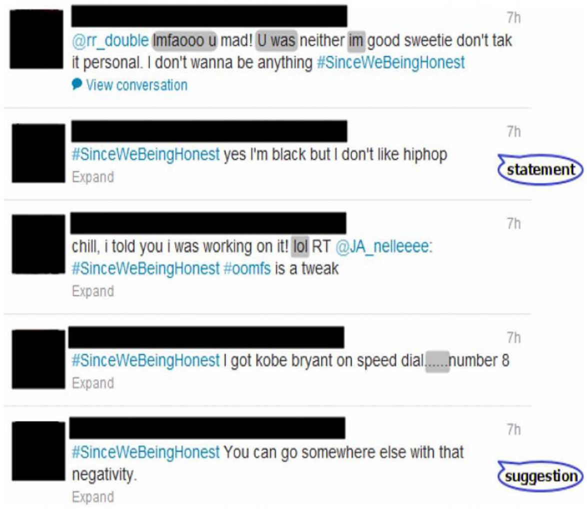


Figure 9:Speech act in twitter conversations

We propose a novel speech act-guided summarization approach in this work. Speech acts characterize tweeters communicative behavior and provide an organized view of their messages. The recognized speech acts in tweets are then used to direct the extraction of key words and phrases to fill in templates designed for speech acts. Leveraging high-ranking words and phrases as well as topic information for major speech acts, we propose a round- robin algorithm to generate template-based summaries. Different from the extractive method adopted in most previous works, our summarization method is abstractive. Evaluated on two 100-topic datasets, the summaries generated by our method outperform two kinds of representative extractive summaries and rival human-written summaries in terms of explanatories and in formativeness.

7.3 Speech acts in Twitter Domain

Searle's Types	Our Types	Example Tweets
Assertive	Statement	*Libya Releases 4 Times Journalists - http://www.photozz.com/?104k*
Directive	Question	*#sincewebeinghonest why u so obsessed with what me n her do?? Don't u got ya own man???? Oh wait.....*
	Suggestion	*RT @NaonkaMixon: I will donate 10 $ to the Red Cross Japan Earthquake fund for every person that retweets this! #PRAYFORJAPAN*
Expressive	Comment	*is enjoying this new season of #CelebrityApprentice.... Nikki Taylor = Yum!!*
Commissive	Miscellaneous	*65. I want to get married to someone i meet in highschool. #100factsaboutme*
Declarative		

Figure 10: Classification of speech acts of twitter utterances

Our approach is the use of speech acts, which capture the common grounds of tweets from a communicative perspective. Each tweet is associated with a type of speech act, like the statement and suggestion for the two tweets. Unless in a few cases (e.g., using a speech act hashtag like #question), users do not report the speech acts they are performing when twittering. So before using speech acts for summarization, we need to recognize them in tweets automatically. Then, guided by the recognized speech acts in the tweets, we can proceed to extract key words and phrases from the tweets. Leveraging the linguistic knowledge of speech acts, we generate summaries that integrate the extracted language materials into speech act-sentences.

Category	Topic	# Tweets
News	Japan Earthquake	1742
	Libya Releases	1408
Entity	Dallas Lovato	677
	Nikki Taylor	786
LST	#100factsaboutme	2000
	#sincewebeinghonest	2000

Figure 11: Categorization of tweets based on Topics

7.4 Twitter Profiling

Extending the work from Speech Act recognition on Twitter Conversation, we plan to profile Tweeters in terms of Speech Acts. Future steps for Twitter Profiling:

1. Extract all tweets from Tweeters profile
2. Clean and normalize tweets
3. Run classifiers over all the tweets
4. Generate a time series of the classification of tweets

References

[1] Ashequl Qadir and Ellen Riloff. 2011. *Classifying Sentences as Speech Acts in Message Board Posts*, volume 1. University of Utah, Salt Lake City, UT.

[2] Rachele De Felice, Jeannique Darby, Anthony Fisher, David Peplow. 2013. *A classification scheme for annotating speech acts in a business email corpus* . University of Nottingham.

[3] Cortes, C.; Vapnik. 1995. *Support-vector networks*

[4] Bishop, Christopher. 2006 *Pattern recognition and machine learning*,

[5] Ritter A., Cherry, C., Dolan, B. 2010 *Unsupervised modeling of twitter conversations*,

[6] Renxian Zhang, Wenjie Li, Dehong Gao, and You Ouyang 2013 *Automatic Twitter Topic Summarization With Speech Acts*,

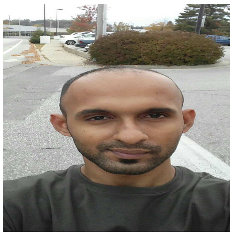
[7] Renxian Zhang,Dehong Gao,Wenjie Li 2011 *Towards Scalable Speech Act Recognition in Twitter: Tackling Insufficient Training Data*,

[8] Jeong, M., Lin, C. Y., Lee, G. G. 2009 *Semi-supervised speech act recognition in emails and forums. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3*,

[9] Stolcke, Andreas, Ries, Klaus, Coccaro, Noah, Shriberg, Elizabeth, Bates, Rebecca, Jurafsky, Daniel, Taylor, Paul, Martin, Rachel, Meteer, Marie, and Van Ess-Dykema, Carol 2000 *Dialogue act modeling for automatic tagging and recognition of conversational speech. Computational Linguistics*,

8 Appendix

8.1 Member contributions



Atul Mohan: Developed code for identification of majority classes, creation of various data sets as required by the cross training requirements.

Chitesh Tewani: Developed the bag of words and hand crafted features for SVM. Normalized the tags across both the datasets.

Prashanth Balasubramani: Involved in identifying the relevant tags for the hand crafted features. Tabulated the various results obtained from the experiments conducted.