IJICT International Journal of Information & Communication Technology

ITRC

Volume 2- Number 1- May 2010 (pp.65-71)

Research Note

Speech Acts Classification of Persian Language Texts Using Three Machine Learning Methods

Mohammad Mehdi Homayounpour
Lab. for Intelligent Signal and Speech Proc.
Department of Computer Engineering and IT
Amirkabir University of Technology
Tehran, Iran
homayoun@aut.ac.ir

Arezou Soltani Panah
Lab. for Intelligent Signal and Speech Proc.
Department of Computer Engineering and IT
Amirkabir University of Technology
Tehran, Iran
a_soltani_p@yahoo.com

Received: December 14, 2008- Accepted: January 19, 2010

Abstract- **The objective of this paper is to design a system to classify Persian speech acts. The driving vision for this work is to provide intelligent systems such as text to speech, machine translation, text summarization, etc. that are sensitive to the speech acts of the input texts and can pronounce the corresponding intonation correctly. Seven speech acts were considered and 3 classification methods including (1) Naive Bayes, (2) K-Nearest Neighbors (KNN), and (3) Tree learner were used. The performance of speech act classification was evaluated using these methods including 10-Fold Cross-Validation, 70-30 Random Sampling and Area under ROC. KNN with an accuracy of 72% was shown to be the best classifier for the classification of Persian speech acts. It was observed that the amount of labeled training data had an important role in the classification performance.**

Keywords- Speech act, Persian language, text processing, Text To Speech, Naive Bayes, K-Nearest Neighbors, Tree learner

I. INTRODUCTION

The idea of a *speech act* can be captured by emphasizing that "by saying something, we *do* something", as when a minister joins two people in marriage saying, "I now pronounce you husband and wife" . We use speech acts in our conversations every day, when greeting, compliment, request, invitation, apology, threaten and so on. In other words, speech act is the basic unit of language used to express meaning, an utterance that expresses an intention and exists in all languages with different cultures. Most misunderstandings by learners of second languages are related to speech acts, because the natural tendency for language learners is to fall back on what they know to be appropriate in their first language. Normally, the speech act is a sentence, but it can be a word like “Sorry!” to perform an apology, or a phrase as long as it follows the rules necessary to accomplish the intention like "I’m sorry for my behavior."

In this paper, attempt has been made to classify speech acts in Persian texts and to draw out features of speech acts in Persian and implement a speech act classifier to be used in systems such as text to speech, machine translation, information retrieval, text summarization, text categorization, etc. It is clear that

the benefit of this work is for example to have TTS systems that can pronounce the speech intonation more correctly or machine translation systems that translate more accurately. Seven categories of speech acts were considered in this study and three different approaches were used to classify them including corpus-based, knowledge-based and hybrid approaches. Corpus-based methods may be supervised or unsupervised. Supervised methods usually outperform unsupervised methods, but they need tagged corpora to be used as training data. Supervised methods that are based on the machine learning techniques are divided into three methods including probabilistic, instance-based and rule-based methods. In this paper, three methods including naïve Bayes, K-nearest neighbor and decision trees are evaluated for speech act classification, knowing that many other machine learning techniques can also be used and assessed. Knowledge-based methods need knowledge sources including corpora with semantic knowledge, thesaurus, etc. Unfortunately, no labeled corpus including speech act tags, exists for Persian. Since we had no access to Persian knowledge sources such as Persian WordNet, SenseNet, tagged corpora, etc., we used a raw corpus, partially labeled it manually and used it in our experiments.

This paper is structured as follows. First, in Section 2, it describes related works on this subject and compares our classification with other speech act taxonomies. Then in Section 3, implementation for the identification of speech acts is described, and in Section 4, results are discussed. Section 5 presents some concluding remarks and pointers to future works and deduction to our work on this subject.

II. RELATED WORKS

No previous work has attempted to identify speech act categories for Persian but there are, of course, distinct taxonomies for different application of speech acts in other languages, especially English. Some of these taxonomies are as follows:

John Searle was one of the people who worked much on the speech act theory. He used the following classification of speech acts: assertive, directive, commisive, expressive and declaratives. He believed that to understand language, one must understand the speaker's intention. Thus, Searle refers to statements as speech acts and then presents a program for the analysis of indirect speech act performances as it follows [1]:

Step1: Understand the facts of the conversation.

Step2: Assume cooperation and relevance on behalf of the participants.

Step3: Establish factual background information pertinent to the conversation.

Step4: Make assumptions about the conversation based on steps 1–3.

Step5: If steps 1–4 do not yield a consequential meaning, then infer that there are two illocutionary forces at work.

Step6: Assume the hearer has the ability to perform the act the speaker suggests. The act that the speaker is asking to be performed must be something that would make sense for one to ask. For example, the hearer might have the ability to pass the salt when asked to do so by a speaker who is at the same table, but not have the ability to pass the salt to a speaker who is asking the hearer to pass the salt during a telephone conversation.

Step 7: Make inferences from steps 1–6 regarding possible primary illocutions.

Step 8: Use background information to establish the primary illocution [1].

With this process, John Searle found a method that acceptably reconstructs what happens when an indirect speech act is performed.

An interesting Taxonomy is related to Dore who breaks down speech acts for utterances of Turkish child to nine categories: labeling (e.g., "This is a book"), repeating (when a child frequently requests something), answering (e.g., ""This is a book" in response to the question: "What is this?"), requesting an action (e.g., "Give me my doll"), requesting an answer (e.g., "tell me where is dad"), calling (e.g., "Daddy"), greeting (e.g., "hi"), protesting, and practicing (like "when she sings"). For this purpose, he chose a Turkish girl named Didem and her utterances were recorded every week for thirty minutes [2].

Verbal Response Mode (VRM) is a principled taxonomy of speech acts that is used by Andrew Lampert and Robert Dale and Cécile Paris. This taxonomy categories on two dimensions, characterized as literal meaning and pragmatic meaning. VRM is categorized in these 8 following categories: disclosure, edification, advisement, confirmation, question, acknowledgment, interpretation, and reflection. They achieved an accuracy of 60.8% using linguistic features derived from VRM' human annotation guidelines and could improve it to 79.8% using additional features. For implementation they used a VRM coder training application for training human annotators. This software included transcripts of spoken dialogues from various domains segmented into utterances, with each utterance annotated with two VRM categories that classify both its literal and pragmatic meaning [3].

In [2], speech acts in naturalistic tutorial dialog is classified into four categories: Assertions, Yes/No Questions, WH-Questions, and Frozen Expressions

and it is pointed that of course that there are other categories of speech acts, such as promises, declarations, and expressive evaluations, but since the goal of their research is to develop a speech act classification system that can operate as part of a robust computerized dialogue system in tutoring, they have chosen this taxonomy.

In an AutoTutor system, the primary goal is to model human-human tutorial dialogues. In such dialogues, the tutor normally takes control of the dialogue by presenting problems to the student, getting answers and other contributions back, and providing help when needed. Sometimes, however, the student takes the initiative, as in the case of student questions. For example, the student might ask for the definition of a technical term (e.g., "What does ROM mean?") or a confirmation that information is correct (e.g. "Isn't ROM Read Only Memory?"). A student might ask the tutor to repeat something (e.g. "What?", "Could you say that again?").

For the computerized tutor to be able to appropriately respond in such mixed-initiative dialogue, the system must understand not only the information content of the speech act, but also the speaker's intentions. The intentions are to some extent made clear by the categorization of the speech act. They tested three models including a feed-forward neural network, a parsing model, and a simple rule-based classifier and concluded that the parsing model had the best performance of 79% accuracy [2].

Authors of [4] describe a probabilistic method for dialogue act (DA) extraction for concept-based multilingual translation systems. A DA is a unit of a semantic interlingua which consists of speaker information, speech act, concept and argument. In [5] a semi-supervised method for automatic speech act recognition in email and forums is presented. The major challenge of this task is due to lack of labeled data in these two genres. The method used in this paper uses automatically extracted features such as phrases and dependency trees, called subtree features, for semi-supervised learning. Empirical results demonstrate that the model is effective in email and forum speech act recognition.

Many other studies have been done on speech acts. Some of these are: classification of speech acts in tutorial dialog [6], tagging of speech acts and dialogue game in call home [7], mining and assessing discussions on the Web through speech act analysis [8], dialog act modeling for conversational Speech [9], using speech act theory to model conversations for automated classification and retrieval [10], dialogue act classification using language models [11], applying machine learning to discourse processing [12], and dialogue act modeling for automatic tagging and recognition of conversational speech [13].

III. IMPLEMENTATION

In this paper, we used a raw corpus created by the Research Center of Intelligent Signal Processing (RCISP) [14]. This is the most important comprehensive corpus for the Persian language, and contains about one hundred million words. The texts of this corpus are gathered from different sources like newspapers, magazines, journals, Internet, books, plays, itineraries, diaries, calendars, and letters. This corpus includes various domains such as economy, export, culture, sciences, etc. This corpus does not include speech act category tags and is a raw corpus for our problem. Ten millions words of this corpus include Part of Speech (POS) Tags [14].

Seven speech act categories were considered in our experiments. These are as follows:

1- Questions. These are usual questions.
2- Requests. We use these speech acts to ask somebody for something in a polite way.
3- Directives. With these speech acts we cause the hearer to take a particular action perforce.
4- Threats. With these speech acts we can promise hurting somebody or doing something if the hearer does not do what we want.
5- Quotations. These are speech acts that another person said or wrote before.
6- Declaratives. With these speech acts we can transfer information to hearer.
7- Narratives. With these speech acts we tell what has happened.

Supervised machine learning algorithms were used in this study to classify speech acts. Hence, a sufficient amount of Persian sentences corresponding to the concerned speech acts were extracted and labeled.

Persian language is a complex language and it is difficult to elicit some rules for each category of speech acts. So it seems that supervised classification methods may present a better performance for speech act classification.

A lot of exceptions were observed during labeling the sentences and experiments. For example for the category of "question" speech act, the following words are usually used in question sentences, while these sentences are usually ended with a question mark:

"چه" , "چه وقت" , "چند" ,"کی" , "چه موقع" ,"چه کسی" , "آیا" ,"چه حد" , "چه" ,"چگونه" , "چه طور" , "چرا" , "کجا" , هنگام" "چه میزان" , "چیه" and so on.

But there are a lot of question sentences that do not contain these words. Some of these sentences are:

- "درس می‌خونی ؟ " (Do you study?)
- " برای کار خاصی به تهران آمدید ؟ " (Did you come toTehran for a certain work?)
- "فکر اینرو کردی ؟ " (Have you thought about it?)

Also there are sentences that have question marks but they are not appertain to "question" speech acts, e.g. "ممکنه خواهش کنم اون نمک را بدید به من ؟" (Could you pass me the salt please?). Moreover, there are

sentences that contain question words but they are not question sentences really, e.g., " او می‌دانست پشت پرده چه خبر است " (He knew what was going on behind the curtain).

It is really difficult to extract features that help us understand whether a sentence belongs to Threat speech act category or not. Indeed it is necessary to subtilize in the meaning of the sentence. Look at the following sentences:

- " اگه دست به من بزنی خودمو می‌کشم." (I will kill myself if you touch me.)
- "از اینجا که برم از دستتون شکایت می‌کنم." (When I leave here, I will sue you.)
- "نگهدار و گرنه خودمو پرت می‌کنم بیرون." (stop the car or I will jump out of the car)
- "الان زنگ می‌زنم کلانتری تا به تو نشان بدهم، که در این مملکت چاقوکشیدن تاوان دارد." (I call commissariat now to show you ruffianism has a penalty in this country.)
- "یه حالی ازت بگیرم که اسم خودت یادت بره." (I turn you off so that you will forget your name.)

Alike there is no special features for declarative and narrative speech acts. Hence, supervised methods were used.

9145 sentences containing different speech acts were tagged and used in our experiments. These sentences were extracted from the Persian corpus described before. The *Orange* software was used to implementing our experiments. This software, which is freely available online[1], includes machine learning methods for classification. These methods start from the data that incorporates classified instances (e.g. a collection of attribute values that are classified to a certain class), and attempts to develop models that would, given the set of attribute-values, predict a class for such instances.

In this paper, three classifiers were used: Naive Bayes, K-Nearest Neighbors or KNN and a tree learner. Of course, there are many other machine learning methods such as Neural Networks, Support Vector Machine (SVM), etc. that can be used for speech act classification. We aim to use these classifiers as future works. In machine learning, one should not learn and test classifiers on the same data set. For this reason, three methods for training and testing were used:

1-In n-fold cross-validation method, data set is spilt into n equally sized subsets, and then in the i-th iteration (i=1 to n), the i-th subset is used for testing the classifier that has been built on all other remaining subsets (n=10 in our experiments).

2- In random sampling method, the data set is spilt to 70% and 30%, use the first part of the data (70%) to learn the model and obtain a classifier that will be tested on the remaining data (30%).

3- The third method is much used in medicine and is called as area under ROC curve. It is a discrimination measure that ranges from 0.5 (random guessing) to 1.0 (a clear margin exists in probability that divides the two classes). This algorithm will investigate all pairs of data items. Those pairs where the outcome was original will be termed valid pairs and will be checked. The area under ROC is then the proportion of correct pairs in the set of valid pairs of instances and it is rather fast than n-fold cross validation because there exist a better algorithm with complexity O (n.log$n$) instead of O (n^2).

IV. EXPERIMENTAL RESULTS

In order to achieve general results, as it was said before, the training texts were extracted from a 10 million words corpus which has been collected from multiple subjects and sources. The speech acts were compiled into a database of 1734 Questions, 928 Requests, 1113 Directives, 544 Threats, 850 Quotations, 2000 Declaratives and 1976 Narratives. Each method of speech act classification was subsequently tested on the database. Figure 1 depicts the results of our experiments.

A. *Naive Bayes*

As shown in Fig. 1, the accuracy of this baseline varies from 66.5% to 68%, depending on the method used for training and testing. Totally for all three machine learning algorithms accuracy is improved using "Area under ROC" method for training but as the amount of data increases, accuracy will decrease with this method. Speed is the privilege of this machine learning algorithm. It takes only 12 or 13 seconds to execute. This model's performance will be increased to 77% accuracy by omitting declarative and narrative speech act categories.

B. *K-Nearest Neighbors*

In this algorithm, K is number of neighbors. Through our experiment, if K is set to 40, accuracy will improve. As shown in Fig. 1, using 9145 labeled samples (sentences), the accuracy of KNN varies from 58.3% to 71.6%, depending on the method used for training and testing and it takes nearly 15 minutes to execute. By omitting sixth and seventh classes, 78% accuracy will be obtained.

[1] Orange is available to download from http://magix.fri.uni-lj.si/orange/

C. Decision tree

One important problem of decision tree method is its high computation complexity and its execution time. As training data increases, the tree grows extraordinarily. For 8676 labeled data forenamed, the tree could not be made completely in 12 hours! With 10 fold cross validation method for training, accuracy is 50% with 1071 training samples. Finally by third method for training and test (Area under ROC) and using 2822 samples, 67% accuracy could achieve.

It can be seen in Fig. 1 that KNN identifies speech acts with the best accuracy which is about 72% when training and test method is the area under ROC. It seems that one reason for the decrease in the classification performance is related to the fact that there is no major difference in expressing the declarative and narrative speech acts. Hence, these two classes were merged and new tests were conducted. The results are shown in Table 1.

More training data improves the performance. Parameters of training data will be more unbiased if more training data are used. Thus, more generality will be achieved and the performance will be improved.

The corpus used to extract speech acts includes articles from different subjects such as story, theater, weblogs, books, different newspapers and journals, etc. Therefore, a sufficient variety of subjects has been considered for the collection of training and test corpora in our experiments.

Table1. Accuracy results with merging declarative and narrative speech acts in %

Training & testing method / Classifier	10-fold cross validation	70-30 random sampling	Area under ROC
Naive Bayes	73.9	67.3	65.8
KNN	72.0	66.0	73.9

Of course the size of training data is an effective parameter. More training data means more information on different speech acts and this can improve the performance. Performance improvement versus the size of training data is depicted in Figure 1. Based on Figure 1, increasing the amount of tagged samples, increases the performance. Of course the slope of increase in the performance versus the amount of tagged training samples is slow. For example for naïve Bays classifier, for a doubling of training samples, only 10% of performance improvement was observed. This may be due to the high number of speech acts (7 speech acts) in our experiments and due to the similarity between speech acts. For example there is no distinctive difference between question and request speech acts, or between declaratives and narratives. One way to increase the performance is to reduce the number of classes. We observed that the dominant speech act included in educational, economical and political texts is "declaratives" (information transfer), and the most

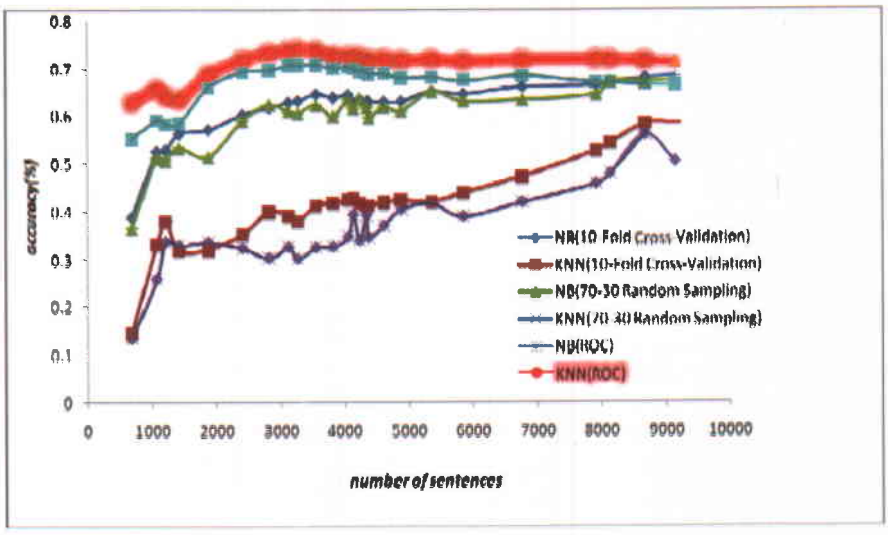

Fig. 1- Performance comparison of different machine learning methods for Persian speech act classification vs. the number of training samples (sentences).

frequent speech act for stories and historical texts is "narratives". In an experiment, we reduced the number of classes to 4 classes based on the above idea and this improved the classification accuracy to about 88%.

The principle used in this paper for speech act classification supposes that some words are more frequent in the contexts relating to each speech act. The evidence of this claim was shown in this paper by obtaining more than 72% of speech act classification performance for seven Persian speech acts. It seems that this principle can be generalized to most of other languages. So, it is not far from the reality that the principles assumed in this paper can be generalized to other languages. Of course sometimes, a given speech act is expressed using some prosodic and intonational cues. For example, in Persian, one sentence may be expressed in an interrogative or declarative form only by changing the speech intonation. But in most cases some words such as do, does, am, is, are and Wh-question words are used for interrogative sentences, while they are less likely to be used in declarative sentences.

In this paper, Persian words are sued as components of feature vectors used for machine learning techniques. In other words all potential Persian words can be an element of each feature vector. Functional words such as conjunctions, propositions, and numbers, surnames, etc, do not usually have an important role in determination of the kind of speech act included in a given text. These kinds of words were discarded once in our experiments. This raised speech act performance to about 80%.

Another idea to improve the performance is to use and morphological analyzer or a verb stemmer to find the root of spent verbs and derivational words. This can both reduce the size of feature vector and increase the performance.

V. CONCLUSIONS

In this paper, the classification of seven speech acts in Persian language texts were studied. Naive Bayes as a fast classifier showed a performance of only 68.4%. The accuracy of KNN as the best classifier was 71.2%. Decision tree method with 67% accuracy is computationally very expensive. It was observed that increasing the labeled corpus to produce a larger and more evenly distributed training set, usually improves the classifier performance. For example increasing the training data from 5484 sentences to 9145 sentences, improved the accuracy of Naïve Bayes method from 64.6% to 68.4%. In speech act classification, some information about the text may improve the classification accuracy. For example if the input text has a label of "didactic" or "inform" or "economic", most of its subjects belong to Declarative speech act, or if the input text has labels such as "story" or "historic", clearly most sentences appertain to Narrative speech act. This idea was studied and a classification performance of 81% was achieved.

As it was mentioned, the RCISP text corpus was the only available knowledge source we used in our experiments. Better performances could be obtained if more appropriate Persian knowledge sources similar to WordNet, SenseNet, etc. exist and could be used.

As a future work, some more speech acts such as poems and proverbs will be considered in our experiments. In order to verify the dependency of results on the language, we intend to continue our research by conducting similar experiments using the same approach and classification techniques on English texts. New classifiers such as Support Vector Machine and Neural nets will also be evaluated for Persian speech act classification. Weighting each word for a given speech act based on its contribution to discriminate that speech act from other speech acts is the next idea to be considered in our future works.

ACKNOLWLEDGMENT

The authors would like to thank Iranian Supreme Council of ICT (SCICT) for supporting this work.

REFERENCES

[1] J. Searle, "Speech Acts: An Essay in the Philosophy of language", pp. 178-184, 1969.

[2] J. Dore, "Children's illocutionary acts", In R. Freedle (ed.), Discourse Comprehension and Production. Hillsdale, NJ: Erlbaum, 1977

[3] A. Lampert, R. Dale, C. Paris, "Classifying Speech Acts using Verbal Response Mode", in Australasian Language Technology Workshop (ALTW), 2006.

[4] T. Fukadayz, D. Koll, A. Waibely, K. Tanigakiz, "Probanilistic speech act extraction for concept based multilingual translation systems", http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.39.5392.

[5] [2] M. Jeong, C. Lin, G.G. Lee, "Semi-supervised Speech act recognition in emails and forums", in Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 1250–1259, 2009.

[6] J. Marineau, P. Wiemer-Hastings, D. Harter, B. Olde, P. Chipman, A. Karnavat, V. Pomeroy, A. Graesser, S. Rajan, "Classification of speech acts in tutorial dialog", in Proceedings of the workshop on modeling human teaching tactics and strategies at the Intelligent Tutoring Systems 2000 conference, pp. 65–71, 2000.

[7] A. Levin, K. Ries, A. Thyme-Gobbel, A. Lavie, "Tagging of speech acts and dialogue game in Spanish Call Home", in Proceedings of ACL-99 Workshop on Discourse Tagging, College Park, 1999.

[8] K. Jihie, C. Grace, F. Donghui, S. Erin; H. Eduard, "Mining and assessing discussions on the Web through speech act analysis", in Proceedings of the Workshop on Web Content Mining with Human Language Technologies at the 5th International Semantic Web Conference (ISWC), Athens, Georgia, 2006.

[9] A. Stolcke, E. Shriber, R. Bates et al. "Dialog act modeling for conversational speech". In Applying machine learning to discourse processing, AAAI Spring Symposium, Technical Report SS-98-01, pp. 98-105, 1998.

[10] D.P. Twitchell, M. Adkins, J.F. Nunamaker, K. Burgoon, "Using speech act theory to model conversations for automated classification and retrieval". in M. Aakus & M. Lind (Eds.), Proceedings of the 9th International Working Conference on the Language-Action Perspective on Communication Modeling. New Brunswick, NJ: Rutgers University Press, pp. 121-130, 2004.

[11] N. Reithinger, M. Klesen, "Dialogue act classification using language models", Spoken Language Technology Workshop, pp. 70-73, 1997.

[12] J. Chu-Carroll. "Applying machine learning to discourse processing", in AAAI Spring Symposium, pp. 128, 1998.
[13] A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, M. Meteer, and C. Van Ess-Dykema, "Dialogue act modeling for automatic tagging and recognition of conversational speech", 2000.
[14] M. Bijankhan, Persian text corpus, Technical Report, Research Center of Intelligent Signal Processing (RCISP), http://www.rcisp.com, 2008.

M. Mehdi Homayounpour was born in 19960 in Iran. He received his BSc degree in Electronics from Amirkabir University of Technology in 1986, MSc in Telecommunications from Khaje Nasireddin Toosi in 1989, and Ph.D. in Electrical Engineering from Paris-11 University, Paris, France. He has been a faculty member of Computer Engineering and IT Department at Amirkabir University of Technology (Tehran Polytechnics), Tehran, Iran, since 1995. His research interests include signal and speech processing, natural language processing, hardware design and multimedia.

Arezou Soltani Panah was born in 1986 in Iran. She received her BSc degree in Computer engineering from Amirkabir University of Technology in 2008, studing MSc in Computer networks at the same university. Since 2009, she is working in the area of peer-to-peer networks.