Automatic Dialog Acts Recognition based on Words Clusters

Pavel Kral, Jana Kleckova, Christophe Cerisara

WESPAC IX 2006
The 9th Western Pacific Acoustics Conference
Seoul, Korea, June 26-28, 2006

AUTOMATIC DIALOG ACTS RECOGNITION BASED ON WORD CLUSTERS

Pavel KRÁL, Jana KLEČKOVÁ
Dept. Informatics & Computer Science
University of West Bohemia
Univerzitni 8, 306 14, Plzen, Czech Republic
E-mail: {kral, kleckova}@kiv.zcu.cz

Christophe CERISARA
LORIA UMR 7503
Campus scientifique BP 239, 545 06, Vandoeuvre-les-Nancy
E-mail: cerisara@loria.fr

ABSTRACT

This paper deals with automatic dialog acts (DAs) recognition in Czech. A Dialog act is defined by J. L. Austin [1] as a meaning of an utterance at the level of illocutionary force. The four following DAs are considered: statements, orders, yes/no questions and other questions. In our previous works, we proposed, implemented and evaluated two new approaches to automatic DAs recognition based on sentence structure. These methods have been validated on a Czech corpus that simulates a task of train tickets reservation. The main goal of this paper is to propose a new approach to solve the problem of lack of training data for automatic DA recognition. This approach clusters the words in the sentence into several groups using maximization of mutual information between two neighbor word classes. The classification accuracy of the unigram model (our baseline approach) is 91 %. The proposed method, a clustered unigram model, reduces the DA error rate by 12 %.

KEYWORDS: Dialog Act, Word Clustering, Unigram Model

INTRODUCTION

The ability to model and automatically detect discourse structure is a crucial factor to interpret and guarantee a conversation in natural language. However, exact description of discourse structure is generally impossible to obtain. Therefore, this general problem is usually reduced to detection of *dialog acts (DAs)*. A dialog act is a meaning of an utterance at the level of illocutionary force [1]. There are several different DAs. For example ”question”, “answer”, “request” or “agreement” are all dialog acts.

We proposed in [2] several methods to combine prosodic and lexical classifiers and we compared them in order to improve DA recognition score. In [3], we proposed two new DA recognition approaches based on sentence structure. The main goal of this paper is to propose a new approach to solve the problem of lack of training data for automatic DA recognition. This approach clusters the words in any sentence into several groups using maximization of mutual information (MMI) between two neighbor word classes.

Section 2 presents a short review of dialog acts recognition approaches. Section 3 presents our new method, a clustered unigram model. Section 4 evaluates and compares this method with a unigram model. In the last section, we discuss the research results and we propose some future research directions.

SHORT REVIEW OF DIALOG ACTS RECOGNITION APPROACHES

To the best of our knowledge, there is very little existing work on automatic modeling and recognition of dialog acts in the Czech language. Alternatively, a number of studies have been published for other languages, and particularly for English and German.

In most of these works, the first step consists in defining the set of dialog acts to be recognized. In [4], [5], 42 dialog acts classes are defined for English, based on the Discourse Annotation and Markup System of Labeling (DAMSL) tag-set [6]. The MALTUS (Multidimensional Abstract Layered Tagset for Utterances) [7] is another DAs tag set based on DAMSL.

Automatic recognition of dialog acts is usually realized using one or a combination of the three following models:

1. DA-specific language models
2. dialog grammar
3. DA-specific prosodic models

The first class of models infers the DA from the word sequence. Usually, probabilistic approaches are represented by language models such as n-gram [5], [8], or knowledge based approaches such as semantic classification trees [8].

The methods based on probabilistic language models exploit the fact that different DAs use distinctive words. Some clue words and phrases can serve as explicit indicators of dialogue structure. For example, 88.4 % of the trigrams "<start> do you" occurs in English in *yes/no questions* [9].

Semantic classification trees are decision trees that operate on word sequence with a rule-based decision. These rules are trained automatically on a corpus. Alternatively, in classical rule-based systems, these rules can be coded manually.

A dialog grammar is used to predict the most probable next dialog act based on the previous ones. It can be modeled by hidden Markov models (HMMs) [5], Bayesian Networks [10], Discriminative Dynamic Bayesian Networks (DBNs) [11], or n-gram language models [12].

Prosody can be used to provide additional clues to recognize DAs [4]. Prosodic features (fundamental frequency, energy, duration, pause and speaking rate) are extracted automatically for each dialog act and are then used for DA classification.

DIALOG ACT RECOGNITION APPROACHES

Performance of classical approaches, such as n-gram models, depends on the size of the DA corpus. They are working especially well with large DA corpus. However, when the corpus size is small, the number of words per DA class is insufficient for a correct estimation of word probabilities. Our approach, a clustered n-gram model, handles this issue.

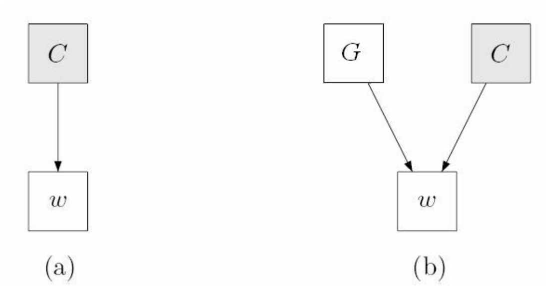

Figure 1. Graphical model of dialog act recognition approaches: grayed nodes are hidden and white ones are observed

Unigram model The general objective of automatic DA recognition is to compute the probability that a sentence belongs to a DA class C, given the lexical and syntactic information, i.e. the words sequence $w_1,..,w_T$. Assuming that the words of the sentence are independent, the probability of the sentence is given by equation 1.

$$P(w_1,...,w_T \mid C) = \prod_{i=1}^{T} P(w_i \mid C) \tag{1}$$

This decomposition corresponds to the left model of figure 1.

Clustered unigram model The words of the application vocabulary are clustered into several groups, in order to reduce the number of parameters to estimate in the unigram models.

During recognition, this approach can be modeled by a very simple Bayesian network with three variables, as shown in the right part of figure 1. In this figure, C encodes the dialog act class of the test sentence, w represents a word and G its cluster.

Words with a similar functional position in the sentence are clustered into the same group. Mutual information between two neighbor word classes is maximized (by MMI method [13]) to find word clusters G. Clustering of all the words of the vocabulary is realized hierarchically, as shown in figure 2. The root of this tree (node G in figure 2) contains all the words, and each leave of the tree (nodes w_1, .., w_n) contains a single word. Nodes G_{11}, ..., G_{1m} illustrate word clusters after the first step of the clustering. Many levels exist between these nodes and the root of the tree.

During training of groups unigram models, group probabilities $P(G_{ij}|C)$ are estimated for each group from the training part of the corpus. During recognition, sentences are classified into DA classes using group models. The optimal group model in the tree is not known *a priori* and must be found empirically.

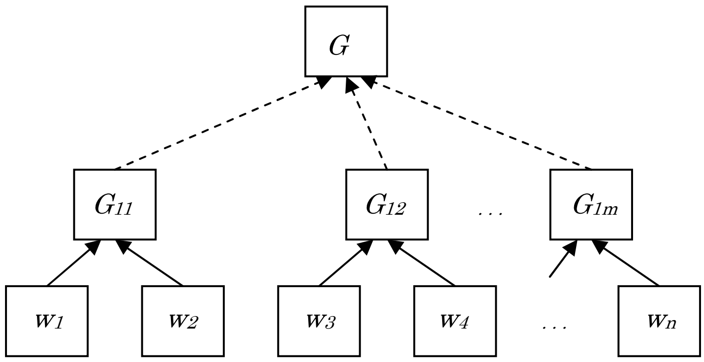

Figure 2. Word clusters hierarchy

EXPERIMENTS

Dialog acts corpus A subset of the Czech Railways corpus, which contains human-human dialogs, is used to validate the proposed methods. It has been labeled manually with the following set of dialog acts: statements, orders, investigation questions and other questions. The corpus contains 2173 utterances (566 statements (S), 125 orders (O), 282 investigation questions (Q[y/n]) and 1200 others questions (Q)). All the following experiments are realized using a cross-validation procedure, where 10 % of the corpus is reserved for the test, and another 10 % for the development set.

Clustered unigram model We test two alternatives of this model. In the first one, words are clustered independently of their DA class. Hence, word clusters are the same for all DA classes. The main advantage of this option is that the number of word occurrences within every word cluster is larger. A drawback is that word clusters do not take into account the specificities of each DA class. In the second implementation, a word cluster is created for each DA class. The unigram statistics are not estimated as robustly as in the previous solution, but they should be more accurate.

The optimal number of word clusters depends on the corpus characteristics. In our experiments, it is found empirically.

Table 1 shows the recognition accuracy of both variants of clustered unigram model. The baseline unigram model recognition accuracy is reported in the first row of this figure. The global recognition accuracy of the DA-independent clustered unigram model is 91.1 %, which is comparable with the unigram model. The DA-dependent clustered unigram model gives 92.1 % recognition accuracy, which slightly outperforms the unigram model, our baseline approach. The DA error rate is thus reduced by 12 %.

Table 1.
Dialog acts recognition accuracy for different approaches

Approach	Classification accuracy in [%]				
	S	O	Q[y/n]	Q	Global
Unigram	93.5	77.6	96.5	89.9	**91**
Common clusters	94	65.6	93.6	91.8	**91.1**
Clusters per DA	92.5	76	92.5	93.8	**92.1**

CONCLUSIONS

In this paper, we presented two variants of a new method for automatic dialog acts recognition based on word clusters. The first one, a DA-independent clustered unigram model gives 91.1 % of recognition accuracy. The recognition accuracy of the second variant, a DA-dependent clustered unigram model, is 92.1 %. Compared to the baseline system, the dialog acts error rate is reduced by 12 %.

The main perspective of our work is to add dialog history (c.f. section 2) to improve DAs recognition accuracy.

In real applications, other clues such as the current dialog state shall also be considered. However, we proposed in this work a DA recognition module that is independent from the task, and which can be easily retrained on another corpus. Another perspective is also to test these methods on another corpus (radio), another language (French) and with more DA classes.

ACKNOWLEDGEMENTS

This work has been partly supported by the European integrated project AmIGO (IST-004182), a project partly funded by the European Commission. The authors would like to thank the project consortium for their help.

REFERENCES

1. J. L. Austin, “How to do Things with Words,” Clarendon Press, Oxford, 1962.
2. P. Král, C. Cerisara, and J. Klečková, “Combination of Classifiers for Automatic Recognition of Dialog Acts,” in Interspeech’2005, Lisboa, Portugal, September 2005, pp. 825–828, ISCA.
3. P. Král, C. Cerisara, and J. Klečková, “Automatic Dialog Acts Recognition based on Sentence Structure,” in ICASSP’06, Toulouse, France, May 2006.
4. E. Shriberg et al., “Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech?,” in Language and Speech, 1998, vol. 41, pp. 439–487.
5. A. Stolcke et al., “Dialog Act Modeling for Automatic Tagging and Recognition of Conversational Speech,” in Computational Linguistics, 2000, vol. 26, pp. 339–373.
6. J. Allen and M. Core, “Draft of Damsl: Dialog Act Markup in Several Layers,” 1997.
7. A. Clark and A. Popescu-Belis, “Multi-level Dialogue Act Tags,” in 5th SIGdial Workshop on Discourse and Dialogue, Boston MA, 2004.
8. M. Mast et al., “Automatic Classification of Dialog Acts with Semantic Classification Trees and Polygrams,” in Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing, 1996, pp. 217–229.
9. D. Jurafsky et al., “Automatic Detection of Discourse Structure for Speech Recognition and Understanding,” in IEEE Workshop on Speech Recognition and Understanding, Santa Barbara, 1997.
10. S. Keizer, Akker. R., and A. Nijholt, “Dialogue Act Recognition with Bayesian Networks for Dutch Dialogues,” in 3rd ACL/SIGdial Workshop on Discourse and Dialogue, Philadelphia, PA, July 2002, pp. 88–94.
11. G. Ji and J. Bilmes, “Dialog Act Tagging Using Graphical Models,” in ICASSP’05, Philadelphia, March 2005.
12. N. Reithinger and E. Maier, “Utilizing Statistical Dialogue Act Processing in VERBMOBIL,” in 33rd annual meeting on Association for Computational Linguistics, Morristown, NJ, USA, 1995, pp. 116–121, Association for Computational Linguistics.
13. J.F. Allen, Natural Language Understanding, The Benjamin/Cummings Publishing Company, 1988.