Semantic Features for Dialogue Act Recognition

Pavel Král[1,2], Ladislav Lenc[1,2], and Christophe Cerisara[3]

[1] Dept. of Computer Science & Engineering
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic
[2] NTIS - New Technologies for the Information Society
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic
[3] LORIA-UMR7503, Nancy, France
{pkral,llenc}@kiv.zcu.cz, cerisara@loria.fr

Abstract. Dialogue act recognition commonly relies on lexical, syntactic, prosodic and/or dialogue history based features. However, few approaches exploit semantic information. The main goal of this paper is thus to propose semantic features and integrate them into a dialogue act recognition task to improve the recognition score. Three different feature computation approaches are proposed, evaluated and compared: Latent Dirichlet Allocation and the HAL and COALS semantic spaces. An interesting contribution is that all the features are created without any supervision. These approaches are evaluated on a Czech dialogue corpus. We experimentally show that all proposed approaches significantly improve the recognition accuracy.

Keywords: COALS, dialogue act, HAL, language model, LDA, semantic spaces, semantics, speech act, syntax

1 Introduction

Automatic Dialogue act (DA) recognition has received much attention in the last years, because this task is fundamental for many emerging dialogue systems. Many approaches have been proposed and evaluated on different corpora containing several dialogue acts. These methods use different type of information coming from the user input.

The features for dialogue act recognition are usually computed from lexical, syntactic, prosodic and/or dialogue history information. However, few approaches consider semantic features, while such features may bring additional information and prove useful to improve the accuracy of our dialogue act recognition system. For instance, because DA recognition systems are typically trained in a supervised way, a frequent cause of recognition errors are "unknown" words in the testing corpus that never occur in the training sentences. Lexical semantic similarity may partly address this issue by grouping words into coherent classes. Depending on how these semantic vectors are computed, these classes, or more generally "semantic distances", can also include some syntactic information, e.g., related to the relative position or degree of proximity of pairs of words within a sentence. This additional information can be used to improve DA recognition, in particular when the training and test conditions differ, or when the size of training corpus is relatively small.

The main goal of this paper is thus to propose semantic features for dialogue act recognition to improve DA recognition results. We describe next three different approaches, which respectively use Latent Dirichlet Allocation (LDA) [5], the Hyperspace Analogue to Language (HAL) [25] and Correlated Occurrence Analogue to Lexical Semantics (COALS) [33] semantic spaces to compute these features. The dialogue act recognition is further done by a supervised classification algorithm, which takes as input both the semantic and the baseline lexical features.

An interesting contribution is that all of these features are computed without any supervision. Another contribution is the proposal to use semantic space models (i.e. HAL and COALS), which, to the best of our knowledge, have never been used for dialogue act recognition. These models will further be compared. The last contribution consists in the evaluation of the proposed approaches on Czech, as a representative of morphologically rich language.

Our target application for the proposed dialogue act recognition approaches is a dialogue system that handles ticket reservation tasks. This system can exploit dialogue acts to better interpret the user's inputs. Our main interest is question (and order) detection, because these sentence modalities constitute an important clue for dialogue management. For example, when our system detects an explicit question (or an order), it has to treat it immediately and react accordingly.

The rest of the paper is organized as follows. Section 2 summarizes important DA recognition approaches with a particular focus on the recent methods using semantic features. Section 3 describes the different models we propose. Section 4 presents experimental results on the Czech Railways dialogue corpus. In the last section, we discuss the research results and we propose some future research directions.

2 Related Work

Relatively few studies on dialogue act modelling and automatic recognition have been published for the Czech language. Conversely, there are many works for other languages, especially for English and German.

Different sets of dialogue acts are defined in these works, depending on the target application and the available corpora. In [37], 42 dialogue act classes are defined for English, based on the Discourse Annotation and Markup System of Labeling (DAMSL) tag-set [1]. Switchboard-DAMSL tag-set [14] (SWBD-DAMSL) is an adaptation of DAMSL in the domain of telephone conversation. The Meeting Recorder DA (MRDA) tag-set [7] is another very popular tag-set, which is based on the SWBD-DAMSL taxonomy. MRDA contains 11 general DA labels and 39 specific labels. Jekat [11] defines 42 DAs for German and Japanese, with 18 DAs at the illocutionary level, in the context of the VERBMOBIL corpus.

These complete DA tag-sets are usually reduced for recognition into a few broad classes, because some classes occur rarely, or because other DAs are not useful for the target application. One typical regrouping may be [36]:

- statements
- questions
- backchannels
- incomplete utterance
- agreements
- appreciations
- other

Automatic recognition of dialogue acts is usually achieved using one of, or a combination of the following types of information:

1. lexical (and syntactic) information
2. prosodic information
3. context of each dialogue act

Lexical information (i.e. word sequence in the utterance) is useful for automatic DA recognition, because different DAs are usually composed from different word sequences. Some cue words and phrases can thus serve as explicit indicators of dialogue structure. For example, 88.4% of the trigrams "<start> do you" occur in English in *yes/no questions* [15].

Several methods are used to represent lexical information [37]. Syntactic information is related to the *order* of the words in the utterance. For instance, in French and Czech, the relative order of the *subject* and *verb* occurrences might be used to discriminate between declarations and questions.

Words n-grams are often used to model some local syntactic information. Král et al. propose in [21] to represent word position in the utterance in order to take into account global syntactic information. Another type of syntactic information recently used for DA recognition are "cue phrases". These can be modelled with a subset of specific n-grams. The n value may vary from 1 to 4, which are selected based on their capacity to predict a specific DA and on their occurrence frequency [39]. A recent work in the dialogue act recognition field [20] also successfully uses a set of syntactic features derived from a deep parse tree.

Unfortunately, there are only few works that incorporate semantic features into the dialogue act recognition task. An interesting DA recognition approach using semantic information is presented in [24]. Sentence parse trees are computed on top of speech recognition output. Semantic information and the derivation rules of the partial trees are extracted and used to model the relationship between the DAs and the derivation rules. The resulting model is then used to generate a semantic score for dialogue act recognition when audio input is given.

The authors of [18] use for DA recognition syntactic and semantic relations acquired by information extraction methods. These features are successfully used as an input to a Bayesian network classifier. They use structured semantic features in the form of semantic predicate classes and semantic roles.

The authors of [28] study lexical semantics to recognize dialogue acts. They compare an unsupervised DA recognition approach based on the Latent Semantic Analysis

(LSA) with another supervised one based on the Support Vector Machines (SVM). The authors show that the unsupervised method brings very good recognition results.

Ritter et al. propose in [32] another interesting unsupervised approach for dialogue act modelling and recognition. This method uses an LDA topic model together with clustering for DA recognition in twitter conversations. The LDA model is used to separate content words from dialogue indicators.

Prosodic information [36], particularly the melody of the utterance, is often used to provide additional clues to classify sentences in terms of DAs. The last useful information is the "dialogue history" which represents the sequence of recognized dialogue acts [37].

Dialogue act recognition is usually based on supervised machine learning as for instance Bayesian Networks [17], Discriminative Dynamic Bayesian Networks [13], BayesNet [29], Memory-based [22] and Transformation-based Learning [34], Decision Trees [26], Neural Networks [23], but also more advanced approaches such as Boosting [38], Latent Semantic Analysis [35], Hidden Backoff Models [4], Maximum Entropy Models [2], Conditional Random Fields [30, 8] and Triangular-chain CRF [12].

3 Semantics for Dialogue Act Recognition

3.1 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) [5] is a popular topic model that assigns a topic to each word in the sentence. The semantically close words are usually represented by similar topics (e.g. synonyms). In this approach, we use a standard LDA model to compute a sentence topic for each word. These features will be used together with word labels for DA recognition.

3.2 Semantic Spaces

Semantic spaces are approaches that derive from word co-occurrence matrices a high dimensional semantic vector to represent every word in the vocabulary. The matrix is computed on a large unlabelled text corpus. Semantically close words are thus usually represented by similar vectors, according to some distance between vectors such as the cosine distance. Moreover, the vector space allows to create clusters of semantically close words by a clustering approach.

It has been shown in [6] that semantic spaces improve the results of language modelling. We assume in this work that these models can also bring relevant information for dialogue act recognition.

We use the HAL and COALS semantic spaces [25, 33]. In the following, sentence-level semantic vectors are computed by additive composition of word-level semantic vectors, which are themselves computed with either the HAL or COALS methods. The resulting sentence-level vector is then used as an additional semantic information for DA recognition. It is worth noting that these two semantic spaces have never been used for dialogue act recognition before.

Note that HAL and COALS are computed on relatively shorter context window, as compared to LDA, which takes in our case the full sentence as context. Hence, HAL and

COALS will capture local dependencies between words while LDA will capture longer dependencies. Intuitively, the local word dependencies shall play a more important role than the distant ones for dialogue act recognition, because longer dependencies are more likely to capture information about the topic of the conversation rather than information about the possible substitutions of words within syntactic structures. Therefore, we expect HAL and COALS to give better results than LDA in the following experiments.

3.3 Dialogue Act Recognition

Let W be a sequence of n words w_i in the sentence, F be a sequence of semantic features f_i ($i \in [1; n]$) and C be a dialogue act class. We use two classifiers for dialogue act recognition:

- The first one is Naive Bayes [31] (also referred as an *unigram* when only word features available). This classifier which models simply $P(W|C)$ is used only as the first baseline.
- The second one is the Maximum Entropy (ME) [3] classifier. This classification algorithm is used to represent $P(C|W)$ or $P(C|W, F)$ in the semantic case. We use this classification approach as being very popular in the natural language processing field, because it has high recognition score. This approach is further also used (with lexical features only) as another baseline to show the impact of the proposed semantic features.

4 Evaluation

4.1 Corpus

The Czech Railways dialogue corpus, which contains human-human conversations, is used to validate the proposed approaches. The number of sentences of this corpus is shown in the second column of Table 1.

This corpus is divided into a training part, described in the first/top part of this table, and a testing part, described in the second/bottom part of the table. The training section of the corpus is used to train our LASER speech recognizer [9]. The testing part, composed of 2173 sentences pronounced by different speakers (see second part of Table 1) is used for testing the DA recognition systems. The sentences in the testing part of the corpus have been annotated manually with the following dialogue acts: statements (S), orders (O), yes/no questions (Q[y/n]) and other questions (Q). Note that in this corpus one utterance corresponds to one DA.

The automatic words transcription obtained with the LASER recognizer (1-best hypothesis) is used to compare the performances of our DA recognition systems on both manual and automatic speech transcriptions. Utterance recognition accuracy is 39.78% and word recognition accuracy is 83.36%.

DA	No.	Example	English translation
1. Training part			
Sent.	6234		
2. Testing part (annotated by DAs)			
S	566	Chtěl bych jet do Písku.	I would like to go to Písek.
O	125	Najdi další vlak do Plzně!	Look at for the next train to Plzeň!
Q[y/n]	282	Řekl byste nám další spojení?	Do you say next connection?
Q	1200	Jak se dostanu do Šumperka?	How can I go to Šumperk?
Sent.	2173		

Table 1. Composition of Czech Railways dialogue corpus

4.2 Tools & Model Configuration

We use the LDA implementation from the MALLET [27] tool-kit for the following experiments. The LDA is trained with 1,000 iterations of the Gibbs sampling. The hyperparameters of the Dirichlet distributions are (as in [10]) initially set to $\alpha = 50/K$, where K is the number of topics and $\beta = 0.1$.

We use the S-Space package [16] for implementation of the HAL and COALS semantic space models and the whole above described corpus for training. For each semantic space, we use a four-words context window in both directions. Both semantic spaces use a matrix composed of 1,000 columns. Dimensionality reduction is not used in our experiments.

The LDA and both semantic space models are trained on the training part of the Railways corpus which is not annotated with the DAs (see first part of Table 1).

For DA recognition itself we use the Brainy [19] implementation of Naive Bayes and Maximum Entropy classifiers.

All experiments are realized using a cross-validation procedure, where 10% of the corpus is reserved for the test. The resulting global accuracy has a confidence interval of $\pm$ 1%.

4.3 Dialogue Act Recognition with Manual Speech Transcripts

Table 2 shows the results of a series of dialogue act recognition experiments with manual speech transcription. These experiments are realized in order to show the performance of the proposed approaches in "ideal" condition, i. e. without errors from Automatic Speech Recognition (ASR).

The first part of this table shows the results of our baseline approaches which use lexical word features with Naive Bayes (NB) and Maximum Entropy (ME) classifiers. The second part shows the results when the semantic features are combined with the baseline word features.

Only the maximum entropy classifier is used with semantic features, because of two main reasons: 1) The maximum entropy classifier has the best performances in terms of DA recognition accuracy; 2) More importantly, the Naive Bayes classifier assumes independence between the input features, and this assumption is clearly broken with the semantic features, because of their strong dependencies with the lexical features.

This table shows that every type of semantic features significantly improves the dialogue act recognition accuracy. We can also see that both semantic spaces outperform the LDA topic model. This result was expected, as already discussed in Section 3.2, because the semantic spaces model more local word dependencies than LDA, which are intuitively more important to characterize dialogue acts. The best recognition accuracy is obtained by the more sophisticated COALS model. Using the semantic features thus increases the recognition accuracy by 7.4% over the baseline Naive Bayes approach and by 3.8% over the discriminative Maximum Entropy model.

	Accuracy in [%]				
Approach/ Classifier	S	O	Q[y/n]	Q	Glob.
1. Lexical information (baselines)					
NB	93.5	77.6	96.5	89.9	**91.0**
ME	90.3	88.0	97.2	96.5	**94.6**
2. Semantic information					
LDA + ME	93.3	87.2	96.5	98.5	**96.4**
HAL + ME	95.1	96.0	97.9	97.9	**97.2**
COALS + ME	96.1	97.6	99.3	99.2	**98.4**

Table 2. Dialogue acts recognition accuracy for different approaches/classifiers and their combination with manual word transcription

4.4 Dialogue Act Recognition with Automatic Speech Recognition

Table 3 shows the dialogue act recognition scores, when word transcriptions are estimated by the LASER speech recognizer. The results are obtained with a word class based trigram language model. The sentence speech recognition accuracy is 39.78% and the word recognition accuracy is 83.36%.

These experiments are done in order to show the robustness of our DA recognition approaches to the ASR errors. The structure of Table 3 is similar as in the previous case. This table shows that the DA recognition accuracy only slightly decreases, when word sequences are estimated automatically from the recognizer. The absolute decrease of the recognition score is very small and varies from 0.4% to 3.9% depending on the approach used. This confirms that our DA recognition system is quite robust to low and moderate ASR recognition errors.

	Accuracy in [%]				
Approach/ Classifier	S	O	Q[y/n]	Q	Glob.
1. Lexical information (baselines)					
NB	93.1	68.8	94.7	86.3	**88.2**
ME	87.5	77.6	89.7	95.2	**91.6**
2. Semantic information					
LDA + ME	88.3	80.8	89.0	96.3	**92.5**
HAL + ME	92.2	86.4	93.6	96.9	**94.8**
COALS + ME	95.9	96.8	97.5	99.0	**98.0**

Table 3. Dialogue acts recognition accuracy for different approaches/classifiers and their combination with word transcription by ASR

5 Conclusions & Future Work

In this paper, we have shown the impact of semantic features for automatic dialogue act recognition. Three approaches to create semantic features have been proposed, implemented and evaluated on the Czech Railways dialogue act corpus. The dialogue act recognition model itself is a maximum entropy classifier, which has been trained in a supervised way. The experimental results confirm that the computed semantic features improve the dialogue act recognition score. We have further shown that, for this task, the semantic spaces significantly outperform the LDA model. This observation can be explained by the fact that these semantic spaces model more local dependencies between words than the LDA model.

Our approaches have been evaluated on the small Czech dialogue act corpus annotated with four dialogue acts. Our perspective thus consists in evaluation of the proposed methods on larger corpora and in other languages with more dialogue acts. The semantic features are particularly interesting for this task, because they can be computed without any supervision. Therefore, no additional annotation will be required when applying the proposed approach on another target language. We concentrated mainly on the creation of highly discriminative features. Another perspective thus consists in evaluation of the other classifiers for our task.

6 Acknowledgements

This work has been partly supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports.

References

1. Allen, J., Core, M.: Draft of Damsl: Dialog Act Markup in Several Layers. In: http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/RevisedManual.html (1997)

2. Ang, J., Liu, Y., Shriberg, E.: Automatic dialog act segmentation and classification in multiparty meetings. In: Proc. ICASSP. Philadelphia, USA (Mar 2005)
3. Berger, A.L., Pietra, V.J.D., Pietra, S.A.D.: A maximum entropy approach to natural language processing. Computational linguistics 22(1), 39–71 (1996)
4. Bilmes, J.: Backoff model training using partially observed data: Application to dialog act tagging. Tech. Rep. UWEETR-2005-0008, Department of Electrical Engineering, University of Washington (Aug 2005)
5. Blei, D.M., Ng, A.Y., Jordan, M.I., Lafferty, J.: Latent dirichlet allocation. Journal of Machine Learning Research 3, 2003 (2003)
6. Brychcín, T., Konopík, M.: Semantic spaces for improving language modeling. Computer Speech & Language 28(1), 192 – 209 (2014)
7. Dhillon, R., S., B., Carvey, H., E., S.: Meeting Recorder Project: Dialog Act Labeling Guide. Tech. Rep. TR-04-002, International Computer Science Institute (February 2004)
8. Dielmann, A., Renals, S.: Recognition of dialogue acts in multiparty meetings using a switching dbn. IEEE trans. on Audio, Speech, and Language Processing 16(7), 1303–1314 (Sep 2008)
9. Ekštein, K., Pavelka, T.: Lingvo/laser: Prototyping concept of dialogue information system with spreading knowledge. In: NLUCS'04. pp. 159–168. Porto, Portugal (April 2004)
10. Griffiths, T.L., Steyvers, M.: Finding scientific topics. Proceedings of the National Academy of Sciences of the United States of America 101(Suppl 1), 5228–5235 (Apr 2004)
11. Jekat *et al.*, S.: Dialogue Acts in VERBMOBIL. In: Verbmobil Report 65 (1995)
12. Jeong, M., Lee, G.G.: Triangular-chain conditional random fields. IEEE trans. on Audio, Speech, and Language Processing 16(7), 1287–1302 (Sep 2008)
13. Ji, G., Bilmes, J.: Dialog Act Tagging Using Graphical Models. In: Proc. ICASSP. vol. 1, pp. 33–36. Philadelphia, USA (Mar 2005)
14. Jurafsky, D., Shriberg, E., Biasca, D.: Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation (Coders Manual, Draft 13). Tech. Rep. 97-01, University of Colorado, Institute of Cognitive Science (1997)
15. Jurafsky *et al.*, D.: Automatic Detection of Discourse Structure for Speech Recognition and Understanding. In: IEEE Workshop on Speech Recognition and Understanding. Santa Barbara (1997)
16. Jurgens, D., Stevens, K.: The s-space package: An open source package for word space models. In System Papers of the Association of Computational Linguistics (2010)
17. Keizer, S., R., A., Nijholt, A.: Dialogue Act Recognition with Bayesian Networks for Dutch Dialogues. In: 3rd ACL/SIGdial Workshop on Discourse and Dialogue. pp. 88–94. Philadelphia, USA (July 2002)
18. Klüwer, T., Uszkoreit, H., Xu, F.: Using syntactic and semantic based relations for dialogue act recognition. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters. pp. 570–578. Association for Computational Linguistics (2010)
19. Konkol, M.: Brainy: A machine learning library. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) Artificial Intelligence and Soft Computing, Lecture Notes in Computer Science, vol. 8468. Springer Berlin Heidelberg (2014)
20. Král, P., Cerisara, C.: Automatic dialogue act recognition with syntactic features. Language Resources and Evaluation 48(3), 419–441 (8 February 2014)
21. Král, P., Cerisara, C., Klečková, J.: Lexical Structure for Dialogue Act Recognition. Journal of Multimedia (JMM) 2(3), 1–8 (June 2007)
22. Lendvai, P., van den Bosch, A., E., K.: Machine Learning for Shallow Interpretation of User Utterances in Spoken Dialogue Systems. In: EACL-03 Workshop on Dialogue Systems: Interaction, Adaptation and Styles Management. pp. 69–78. Budapest, Hungary (2003)

23. Levin, L., Langley, C., Lavie, A., Gates, D., Wallace, D., Peterson, K.: Domain Specific Speech Acts for Spoken Language Translation. In: 4th SIGdial Workshop on Discourse and Dialogue. Sapporo, Japan (2003)
24. Liang, W.B., Wu, C.H., Chen, C.P.: Semantic information and derivation rules for robust dialogue act detection in a spoken dialogue system. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2. pp. 603–608. Association for Computational Linguistics (2011)
25. Lund, K., Burgess, C.: Hyperspace analogue to language (hal): A general model semantic representation. In: Brain and Cognition. vol. 30, pp. 5–5. ACADEMIC PRESS INC JNL-COMP SUBSCRIPTIONS 525 B ST, STE 1900, SAN DIEGO, CA 92101-4495 (1996)
26. Mast *et al.*, M.: Automatic Classification of Dialog Acts with Semantic Classification Trees and Polygrams. In: Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing. pp. 217–229 (1996)
27. McCallum, A.K.: Mallet: A machine learning for language toolkit (2002), `http://mallet.cs.umass.edu`
28. Novielli, N., Strapparava, C.: Exploring the lexical semantics of dialogue acts. Journal of Computational Linguistics and Applications 1(1-2), 9–26 (2010)
29. Petukhova, V., Bunt, H.: Incremental dialogue act understanding. In: Proc. of the 9th International Conference on Computational Semantics (IWCS-9). Oxford (Jan 2011)
30. Quarteroni, S., Ivanov, A.V., Riccardi, G.: Simultaneous dialog act segmentation and classification from human-human spoken conversations. In: Proc. ICASSP. Prague, Czech Republic (May 2011)
31. Rish, I.: An empirical study of the naive bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence. vol. 3, pp. 41–46. IBM New York (2001)
32. Ritter, A., Cherry, C., Dolan, B.: Unsupervised modeling of twitter conversations. In: NAACL HLT 2010 - Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Main Conference. pp. 172–180 (2010)
33. Rohde, D.L., Gonnerman, L.M., Plaut, D.C.: An improved model of semantic similarity based on lexical co-occurrence. Communications of the ACM 8, 627–633 (2006)
34. Samuel, K., Carberry, S., Vijay-Shanker, K.: Dialogue Act Tagging with Transformation-Based Learning. In: 17th international conference on Computational linguistics. vol. 2, pp. 1150–1156. Association for Computational Linguistics, Morristown, NJ, USA, Montreal, Quebec, Canada (10-14 August 1998)
35. Serafin, R., Di Eugenio, B.: LSA: Extending latent semantic analysis with features for dialogue act classification. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. Spain (2004)
36. Shriberg, E., Bates, R., Stolcke, A., Taylor, P., Jurafsky, D., Ries, K., Coccaro, N., Martin, R., Meteer, M., Van Ess-Dykema, C.: Language and Speech, Special Double Issue on Prosody and Conversation, vol. 41, chap. Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech?, pp. 439–487 (1998)
37. Stolcke *et al.*, A.: Dialog Act Modeling for Automatic Tagging and Recognition of Conversational Speech. In: Computational Linguistics. vol. 26, pp. 339–373 (2000)
38. Tur, G., Guz, U., Hakkani-Tur, D.: Model adaptation for dialogue act tagging. In: Proceedings of the IEEE Spoken Language Technology Workshop (2006)
39. Webb, N., Hepple, M., Wilks, Y.: Dialog act classification based on intra-utterance features. Tech. Rep. CS-05-01, Dept of Comp. Science, University of Sheffield (2005)