Towards Unsupervised Recognition of Dialogue Acts

Nicole Novielli
Dept. of Informatics, University of Bari
via Orabona 4
I-70125 Bari, Italy
novielli@di.uniba.it

Carlo Strapparava
FBK-irst
via Sommarive, Povo
I-38050 Trento, Italy
strappa@fbk.eu

Abstract

When engaged in dialogues, people perform communicative actions to pursue specific communicative goals. Speech acts recognition attracted computational linguistics since long time and could impact considerably a huge variety of application domains. We study the task of automatic labeling dialogues with the proper dialogue acts, relying on empirical methods and simply exploiting lexical semantics of the utterances. In particular, we present some experiments in supervised and unsupervised framework on both an English and an Italian corpus of dialogue transcriptions. The evaluation displays encouraging results in both languages, especially in the unsupervised version of the methodology.

1 Introduction

People proceed in their conversations through a sequence of dialogue acts to yield some specific communicative goal. They can ask for information, agree or disagree with their partner, state some facts and express opinions.

Dialogue Acts (DA) attracted linguistics (Austin, 1962; Searle, 1969) and computational linguistics research (Core and Allen, 1997; Traum, 2000) since long time. With the advent of the Web, a large amount of material about natural language interactions (e.g. blogs, chats, conversation transcripts) has become available, raising the attractiveness of empirical methods analyses on this field. There is a large number of application domains that could benefit from automatically labeling DAs: e.g. conversational agents for monitoring and supporting human-human remote conversations, blogs, forums and chat logs analysis for opinion mining, interpersonal stances modeling by mean of conversational analysis, automatic meeting summarizations and so on. These applications require a deep understanding of the conversational structure and the ability of the system to understand who is telling what to whom.

This study defines a method for automatically labeling dialogues with the proper speech acts by relying on empirical methods. Even if prosody and intonation surely play a role (e.g. (Stolcke et al., 2000; Warnke et al., 1997)), nonetheless language and words are what the speaker uses to convey the communicative message and are just what we have at disposal when we consider texts found on the Web. Hence, we decided to simply exploit lexical semantics of the sentences. We performed some experiments in a supervised and unsupervised framework on both an English and an Italian corpora of dialogue transcriptions, achieving good results in all settings. Unsupervised performance is particularly encouraging, independently from the used language.

The paper is organized as follows. Section 2 gives a brief sketch of the NLP background on Dialogue Acts recognition. In Section 3 we introduce the English and Italian corpora of dialogues, their characteristics and DA labeling. In Section 4 we describe the preprocessing of the data sets. Then Section 5 explains the supervised and unsupervised settings, showing the experimental results obtained on the two corpora and providing an error analysis. Finally, in Section 6 we conclude the paper with a brief discussion and some directions for future work.

Proceedings of the NAACL HLT Student Research Workshop and Doctoral Consortium, pages 84–89,
Boulder, Colorado, June 2009.

Speaker	Dialogue Act	Utterance
A	OPENING	*Hello Ann.*
B	OPENING	Hello Chuck.
A	STATEMENT	*Uh, the other day, I attended a conference here at Utah State University on recycling*
A	STATEMENT	*and, uh, I was kind of interested to hear cause they had some people from the EPA and lots of different places, and, uh, there is going to be a real problem on solid waste.*
B	OPINION	Uh, I didn't think that was a new revelation.
A	AGREE /ACCEPT	*Well, it's not too new.*
B	INFO-REQUEST	So what is the EPA recommending now?

Table 1: An excerpt from the Switchboard corpus

2 Background

A DA can be identified with the communicative goal of a given utterance (Austin, 1962). Researchers use different labels and definitions to address this concept: *speech act* (Searle, 1969), *adjacency pair part* (Schegloff, 1968) (Sacks et al., 1974), *game move* (Power, 1979)

Traditionally, the NLP community has employed DA definitions with the drawback of being domain or application oriented. Recently some efforts have been made towards unifying the DA annotation (Traum, 2000). In the present study we refer to a domain-independent framework for DA annotation, the DAMSL architecture (Dialogue Act Markup in Several Layers) by (Core and Allen, 1997).

Recently, the problem of DA recognition has been addressed with promising results: Poesio and Mikheev (1998) combine expectations about the next likely dialogue 'move' with information derived from the speech signal features; Stolcke et al. (2000) employ a discourse grammar, formalized in terms of Hidden Markov Models, combining also evidences about lexicon and prosody; Keizer et al. (2002) make use of Bayesian networks for DA recognition in dutch dialogues; Grau et al. (2004) consider naive Bayes classifiers as a suitable approach to the DA classification problem; a partially supervised framework has also been explored by Venkataraman et al. (2005)

Regardless of the model they use (discourse grammars, models based on word sequences or on the acoustic features or a combination of all these) the mentioned studies are developed in a supervised framework. In this paper, one goal is to explore also the use of a fully unsupervised methodology.

3 Data Sets

In the experiments of the present paper we exploit two corpora, both annotated with DAs labels. We aim at developing a recognition methodology as general as possible, so we selected corpora which are different in content and language: the Switchboard corpus (Godfrey et al., 1992), a collection of transcriptions of spoken English telephone conversations about general interest topics, and an Italian corpus of dialogues in the healthy-eating domain (Clarizio et al., 2006).

In this section we describe the two corpora, their features, the set of labels used for annotating the dialogue acts with their distributions and the data preprocessing.

3.1 Description

The Switchboard corpus is a collection of English human-human telephone conversations (Godfrey et al., 1992) between couples of randomly selected strangers. They were asked to choose one general interest topic and to talk informally about it. Full transcripts of these dialogues are distributed by the Linguistic Data Consortium. A part of this corpus is annotated (Jurafsky et al., 1997) with DA labels (overall 1155 conversations, for a total of 205,000 utterances and 1.4 million words)[1]. Table 1 shows a short sample fragments of dialogues from the Switchboard corpus.

The Italian corpus had been collected in the scope of some previous research about Human-ECA interaction. A Wizard of Oz tool was employed (Clarizio et al., 2006) and during the interaction, a conversational agent (i.e. the 'wizard') played the role of

[1] `ftp.ldc.upenn.edu/pub/ldc/public\_data/swb1\_dialogact\_annot.tar.gz`

Label	Description	Example	Italian	English
INFO-REQUEST	Utterances that are pragmatically, semantically, and syntactically questions	*'What did you do when your kids were growing up?'*	34%	7%
STATEMENT	Descriptive, narrative, personal statements	*'I usually eat a lot of fruit'*	37%	57%
S-OPINION	Directed opinion statements	*'I think he deserves it.'*	6%	20%
AGREE-ACCEPT	Acceptance of a proposal, plan or opinion	*'That's right'*	5%	9%
REJECT	Disagreement with a proposal, plan, or opinion	*'I'm sorry no'*	7%	.3%
OPENING	Dialogue opening or self-introduction	*'Hello, my name is Imma'*	2%	.2%
CLOSING	Dialogue closing (e.g. farewell and wishes)	*'It's been nice talking to you.'*	2%	2%
KIND-ATT	Kind attitude (e.g. thanking and apology)	*'Thank you very much.'*	9%	.1%
GEN-ANS	Generic answers to an Info-Request	*'Yes', 'No', 'I don't know'*	4%	4%
total cases			1448	131,265

Table 2: The set of labels employed for Dialogue Acts annotation and their distribution in the two corpora

an artificial therapist. The users were free to interact with it in natural language, without any particular constraint. This corpus is about healthy eating and contains (overall 60 dialogues, 1448 users' utterances and 15,500 words).

3.2 Labelling

Both corpora are annotated following the Dialogue Act Markup in Several Layers (DAMSL) annotation scheme (Core and Allen, 1997). In particular the Switchboard corpus employs a revision (Jurafsky et al., 1997).[2]

Table 2 shows the set of labels employed with their definitions, examples and distributions in the two data sets. The categories maintain the DAMSL main characteristic of being domain-independent and can be easily mapped back into SWBD-DAMSL ones, and maintain their original semantics. Thus, the original SWBD-DAMSL annotation had been automatically converted into the categories included in our markup language.[3]

4 Data preprocessing

To reduce the data sparseness, we used a POS-tagger and morphological analyzer (Pianta et al., 2008) for preprocessing both corpora. So we considered lemmata instead of tokens in the format *lemma#POS*. In addition, we augment the features of each sentence with a set of linguistic markers, defined according to the semantic of the DA categories. We hypothesize, in fact, these features could play an important role in defining the linguistic profile of each DA. The addition of these markers is performed automatically, by just exploiting the output of the POS-tagger and of the morphological analyzer, according to the following rules:

- **WH-QTN**, used whenever an interrogative determiner (e.g. 'what') is found, according to the output of the POS-tagger;
- **ASK-IF**, used whenever an utterance presents the pattern of a 'Yes/No' question. ASK-IF and WH-QTN markers are supposed to be relevant for the INFO-REQUEST category;
- **I-PERS**, used for all declarative utterances whenever a verb is in the first person form, singular or plural (relevant for the STATEMENT);
- **COND**, used for conditional form is detected.
- **SUPER**, used for superlative adjectives.
- **AGR-EX**, used whenever an agreement expression (e.g.'You're right', 'I agree') is detected (relevant for AGREE-ACCEPT);
- **NAME**, used whenever a proper name follows a self-introduction expression (e.g. 'My name is') (relevant for the OPENING);
- **OR-CLAUSE**, used for or-clauses, that is utterance starting by 'or' (should be helpful for the characterization of the INFO-REQUEST);
- **VB**, used only for the Italian, when a dialectal form of agreement expression is detected.

5 Dialogue Acts Recognition

We conducted some experiments both in a supervised and unsupervised settings.

[2]The SWBD-DAMSL modifies the original DAMSL framework by further specifying some categories or by adding extra features (mainly prosodic) which were not originally included in the scheme.

[3]Also we did not consider the utterances formed only by non-verbal material (e.g. laughter).

5.1 Supervised

Regarding the supervised experiments, we used Support Vector Machines (Vapnik, 1995), in particular SVM-light package (Joachims, 1998) under its default configuration. We randomly split the two corpora into 80/20 training/test partitions. SVMs have been used in a large range of problems, including text classification, image recognition tasks, bioinformatics and medical applications, and they are regarded as the state-of-the-art in supervised learning. We got .71 and .77 of F1 measures respectively for the Italian and English corpus. Table 4 reports the performance for each direct act.

5.2 Unsupervised

It is not always easy to collect large training, partly because of manual labeling effort and moreover because often it is not possible to find it.

Schematically, our unsupervised methodology is: (i) building a semantic similarity space in which words, set of words, text fragments can be represented homogeneously, (ii) finding seeds that properly represent dialogue acts and considering their representations in the similarity space, and (iii) checking the similarity of the utterances.

To get a similarity space with the required characteristics, we used Latent Semantic Analysis (LSA), a corpus-based measure of semantic similarity proposed by Landauer (Landauer et al., 1998). In LSA, term co-occurrences in a corpus are captured by means of a dimensionality reduction operated by a singular value decomposition (SVD) on the term-by-document matrix $\mathbf{T}$ representing the corpus.

SVD decomposes the term-by-document matrix $\mathbf{T}$ into three matrices $\mathbf{T} = \mathbf{U}\mathbf{\Sigma_k}\mathbf{V}^T$ where $\mathbf{\Sigma_k}$ is the diagonal $k \times k$ matrix containing the k singular values of $\mathbf{T}$, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k$, and $\mathbf{U}$ and $\mathbf{V}$ are column-orthogonal matrices. When the three matrices are multiplied together the original term-by-document matrix is re-composed. Typically we can choose $k' \ll k$ obtaining the approximation $\mathbf{T} \simeq \mathbf{U}\mathbf{\Sigma_{k'}}\mathbf{V}^T$.

LSA can be viewed as a way to overcome some of the drawbacks of the standard vector space model (sparseness and high dimensionality). In fact, the LSA similarity is computed in a lower dimensional space, in which second-order relations among terms and texts are exploited. The similarity in the resulting vector space is then measured with the standard cosine similarity. Note also that LSA yields a vector space model that allows for a *homogeneous* representation (and hence comparison) of words, sentences, and texts. For representing a word set or a sentence in the LSA space we use the *pseudo-document* representation technique, as described by Berry (1992). In practice, each text segment is represented in the LSA space by summing up the normalized LSA vectors of all the constituent words, using also a *tf.idf* weighting scheme (Gliozzo and Strapparava, 2005).

Label	Seeds
INFO-REQ	WH-QTN, Question_Mark, ASK-IF, huh
STATEMENT	I-PERS, I
S-OPINION	Verbs which directly express opinion or evaluation (guess, think, suppose, affect)
AGREE-ACC	AGR-EX, yep, yeah, absolutely, correct
REJECT	Verbs which directly express disagreement (disagree, refute)
OPENING	Greetings (hi, hello), words and markers related to self-introduction (name, NAME)
CLOSING	Interjections/exclamations ending discourse (alright, okeydoke), Expressions of thanking (thank) and farewell (bye, bye-bye, goodnight, goodbye)
KIND-ATT	Wishes (wish), apologies (apologize), thanking (thank) and sorry-for (sorry, excuse)
GEN-ANS	no, yes, uh-huh, nope

Table 3: The seeds for the unsupervised experiment

The methodology is completely unsupervised. We run the LSA using 400 dimensions (i.e. k', as suggested by (Landauer et al., 1998)) respectively on the English and Italian corpus, without any DA label information. Starting from a set of seeds (words) representing the communicative acts (see the complete sets in Table 3), we build the corresponding vectors in the LSA space and then we compare the utterances to find the communicative act with higher similarity. To compare with SVM, the performance is measured on the same test set partition used in the supervised experiment (Table 4).

We defined seeds by only considering the communicative goal and the specific semantic of every single DA, just avoiding as much as possible the overlapping between seeds groups. We wanted to design

	Italian						English					
	SVM			LSA			SVM			LSA		
Label	prec	rec	f1	prec	rec	f1	prec	rec	f1	prec	rec	f1
INFO-REQ	.92	.99	.95	.96	.88	.92	.92	.84	.88	.93	.70	.80
STATEMENT	.85	.68	.69	.76	.66	.71	.79	.92	.85	.70	.95	.81
S-OPINION	.28	.42	.33	.24	.42	.30	.66	.44	.53	.41	.07	.12
AGREE-ACC	.50	.80	.62	.56	.50	.53	.69	.74	.71	.68	.63	.65
REJECT	-	-	-	.09	.25	.13	-	-	-	.01	.01	.01
OPENING	.60	1.00	.75	.55	1.00	.71	.96	.55	.70	.20	.43	.27
CLOSING	.67	.40	.50	.25	.40	.31	.83	.59	.69	.76	.34	.47
KIND-ATT	.82	.53	.64	.43	.18	.25	.85	.34	.49	.09	.47	.15
GEN-ANS	.20	.63	.30	.27	.38	.32	.56	.25	.35	.54	.33	.41
micro	.71	.71	.71	.66	.66	.66	.77	.77	.77	.69	.69	.69

Table 4: Evaluation of the two methods on both corpora

an approach which is as general as possible, so we did not consider domain words. The seeds are the same for both languages, which is coherent with our goal of defining a language-independent method.

5.3 Experimental Results and Discussion

We evaluate the performance of our method in terms of precision, recall and f1-measure (see Table 4) according to the DA labels given by annotators in the datasets. As baselines we consider (i) most-frequent label assignment (respectively 37% for Italian, 57% for English) for the supervised setting, and (ii) random DA selection (11%) for the unsupervised one.

Results are quite satisfying (Table 4). In particular, the unsupervised technique is largely above the baselines, for both the Italian and the English experiments. The methodology is independent from the language and the domain: the Italian corpus is a collection of dialogue about a very restricted domain while the Switchboard conversations are essentially task-free. Moreover, in the unsupervised setting we use in practice the same seed definitions. Secondly, it is independent on the differences in the linguistic style due to the specific interaction scenario and input modality. Finally, the performance is not affected by the difference in size of the two data sets.

Error analysis. After conducting an error analysis, we noted that many utterances are misclassified as STATEMENT. One possible reason is that statements usually are quite long and there is a high chance that some linguistic markers that characterize other dialogue acts are present in those sentences. On the other hand, looking at the corpora we observed that many utterances which appear to be linguistically consistent with the typical structure of statements have been annotated differently, according to the actual communicative role they play. For similar reasons, we observed some misclassification of S-OPINION as STATEMENT. The only significative difference between the two labels seems to be the wider usage of 'slanted' and affectively loaded lexicon when conveying an opinion. Another cause of confounding is the confusion among the backchannel labels (GEN-ANS, AGREE-ACC and REJECT) due to the inherent ambiguity of common words like *yes*, *no*, *yeah*, *ok*.

Recognition of such cases could be improved (i) by enabling the classifiers to consider not only the lexical semantics of the given utterance (local context) but also the knowledge about a wider context window (e.g. the previous *n* utterances), (ii) by enriching the data preprocessing (e.g. by exploiting information about lexicon polarity and subjectivity parameters). We intend to follow both these directions in our future research.

6 Conclusions and Future Work

This study aims at defining a method for Dialogue Acts recognition by simply exploiting the lexical semantics of dialogue turns. The technique had to be independent from some important features of the corpus being used such as domain, language, size, interaction scenario. In a long-term perspective, we will employ the technique in conversational analysis for user attitude classification (Martalo et al., 2008).

The methodology starts with automatically en-

riching the corpus with additional features, such as linguistic markers. Then the unsupervised case consists of defining a very simple and intuitive set of seeds that profiles the specific dialogue acts, and subsequently performing a similarity analysis in a latent semantic space. The performance of the unsupervised experiment has been compared with a supervised state-of-art technique such as Support Vector Machines, and the results are quite encouraging.

Regarding future developments, we will investigate how to include in the framework a wider context (e.g. the previous *n* utterances), and the introduction of new linguistic markers by enriching the preprocessing techniques. In particular, it would be interesting to exploit the role of slanted or affective-loaded lexicon to deal with the misclassification of opinions as statements. Along this perspective, DA recognition could serve also as a basis for conversational analysis aimed at improving a fine-grained opinion mining in dialogues.

References

J. Austin. 1962. *How to do Things with Words*. Oxford University Press, New York.

M. Berry. 1992. Large-scale sparse singular value computations. *International Journal of Supercomputer Applications*, 6(1).

G. Clarizio, I. Mazzotta, N. Novielli, and F. deRosis. 2006. Social attitude towards a conversational character. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 2–7, Hatfield, UK, September.

M. Core and J. Allen. 1997. Coding dialogs with the DAMSL annotation scheme. In *Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*, Cambridge, MA.

A. Gliozzo and C. Strapparava. 2005. Domains kernels for text categorization. In *Proceedengs of (CoNLL-2005)*, University of Michigan, Ann Arbor, June.

J. Godfrey, E. Holliman, and J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of ICASSP-92*, pages 517–520, San Francisco, CA. IEEE.

S. Grau, E. Sanchis, M. J. Castro, and D. Vilar. 2004. Dialogue act classification using a bayesian approach. In *Proceedings of SPECOM-04*, pages 495–499, Saint-Petersburg, Russia, September.

T. Joachims. 1998. Text categorization with Support Vector Machines: learning with many relevant features. In *Proceedings of the European Conference on Machine Learning*.

D. Jurafsky, E. Shriberg, and D. Biasca. 1997. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual, draft 13. Technical Report 97-01, University of Colorado.

S. Keizer, R. op den Akker, and A. Nijholt. 2002. Dialogue act recognition with bayesian networks for dutch dialogues. In K. Jokinen and S. McRoy, editors, *Proceedings 3rd SIGdial Workshop on Discourse and Dialogue*, pages 88–94, Philadelphia, PA, July.

T. K. Landauer, P. Foltz, and D. Laham. 1998. Introduction to latent semantic analysis. *Discourse Processes*, 25.

A. Martalo, N. Novielli, and F. deRosis. 2008. Attitude display in dialogue patterns. In *AISB 2008 Convention on Communication, Interaction and Social Intelligence*, Aberdeen, Scotland, April.

E. Pianta, C. Girardi, and R. Zanoli. 2008. The TextPro tool suite. In *Proceedings of LREC*, Marrakech (Morocco), May.

M. Poesio and A. Mikheev. 1998. The predictive power of game structure in dialogue act recognition: Experimental results using maximum entropy estimation. In *Proceedings of ICSLP-98*, Sydney, December.

R. Power. 1979. The organisation of purposeful dialogues. *Linguistics*, 17:107–152.

H. Sacks, E. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

E. Schegloff. 1968. Sequencing in conversational openings. *American Anthropologist*, 70:1075–1095.

J. Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, London.

A. Stolcke, N. Coccaro, R. Bates, P. Taylor, C. Van Ess-Dykema, K. Ries, E. Shriberg, D. Jurafsky, R. Martin, and M. Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–373.

D. Traum. 2000. 20 questions for dialogue act taxonomies. *Journal of Semantics*, 17(1):7–30.

V. Vapnik. 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag.

A. Venkataraman, Y. Liu, E. Shriberg, and A. Stolcke. 2005. Does active learning help automatic dialog act tagging in meeting data? In *Proceedings of EUROSPEECH-05*, Lisbon, Portugal.

V. Warnke, R. Kompe, H. Niemann, and E. Nöth. 1997. Integrated dialog act segmentation and classification using prosodic features and language models. In *Proceedings of 5th European Conference on Speech Communication and Technology*, volume 1, pages 207–210, Rhodes, Greece.