Text based Dialog Act Classification for Multiparty Meetings

Matthias Zimmermann[1], Dilek Hakkani-Tür[1], Elizabeth Shriberg[1,2], Andreas Stolcke[1,2]

[1]International Computer Science Institute (ICSI), Berkeley, USA, [2]SRI International, Menlo Park, USA
{zimmerma, dilek, ees, stolcke}@icsi.berkeley.edu

Abstract

We compare four machine learning approaches and their combination for the classification of dialog acts (DAs) on the ICSI Meeting Corpus:

- mini language models (Mini LMs) [1]
- cue phrase selection (Cue Phrases) [2]
- maximum entropy (MaxEnt) [3]
- boosting based classifier (BoosTexter) [4].

We implemented the following combination schemes:

- simple voting
- linear combination
- multi-layer perceptron (MLP).

Task

The task considered in this work is to assign one of the following five mutually exclusive DA labels to an isolated DA:

Backchannel	*uhhuh / right / yeah*
Disruption	*it's just it's / i mean you'd y-*
Floor grabber	*um / so / yeah*
Question	*right / what does the p. stand for anyway*
Statement	*yeah / so here's the thing*

Note, that we use Disruptions for all DAs that have been marked to be incomplete. All DAs that are related to the management of the floor are labeled as Floor grabbers.

We thank Özgür Çetin and Luke Gottlieb for their contributions in the preparation of the experimental setup. This work was partly supported by the European Union 6th FWP IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811 publication), by DARPA Contract NBCHD030010 through the SRI CALO project (approved for public release, distribution unlimited), NSF Awards IIS-0121396 and IRI-9619921, and the Swiss National Science Foundation through the research network IM2.

Method

For the **Mini LM** approach [1] a LM is first trained on all available examples of a specific DA. To classify an uknown DA the utterance probabilities are computed for all available DA-specific LMs and weighted with the corresponding DA prior probabilities.

In the case of the **Cue Phrases** approach [2] all word n-grams up to a fixed *n* are considered to be potential cue phrase candidates. For cue phrase its predictivity for each DA type is then estimated on the training data. To classify an unknown DA the cue phrase with the highest predictivity determines the DA label.

The features used by the best **MaxEnt** classification scheme [3] are the number of words in a DA, the first two and the last two words , as well as the first word bigram and the last word bigram (after removing initial filler words). The use of more than two words at the beginning or end of DA did degrade the performance as well as the use of all the words (and word n-grams).

Example: The creation of a MaxEnt feature vector.

so **it's** **at** **friday** **at** **three**

```
Question: n=5, 1=it's, 2=at, 3=it's_at, \
          4=at_three, 5=at, 6=three
```

For **BoosTexter** [4] the length of the DA as well as all word n-grams up to trigrams were used as features. Training BoosTexter on the feature vector of the MaxEnt approach increases the error only by 0.2%, confirming the importance of the beginning and the end of a DA.

Simple Voting was used to combine the output of all four classifiers. For the **Linear Combination** and **MLP** based combination schemes, all but the Cue Phrases approaches were combined.

Experiments

The experimental setup of [3] for the classification of the DAs was used.

System		Ref	STT Manual	STT Auto
WER		NA	35.4%	38.2%
Chance		45.0%	42.0%	35.5%
Mini LM	[1]	26.7%	29.8%	27.3%
Cue Phrases	[2]	26.6%	29.6%	28.5%
MaxEnt	[3]	22.5%	26.5%	23.8%
BoosTexter	[4]	21.7%	26.9%	24.2%
Simple Voting		23.8%	27.3%	24.4%
Linear Comb.		21.7%	26.3%	23.6%
MLP		21.3%	26.2%	23.3%

Reference conditions assume the true words while STT Manual uses the words generated by a speech-to-text (STT) system using a manual segmentation of the audio input. STT Auto refers to a setup where the segmentation of the audio input is done automatically.

Conclusion

For the investigated task we find that both BoosTexter and MaxEnt significantly outperform the Mini LM and the Cue Phrases approaches. Interestingly, the best classification methods do only take into account the first two and the last two words of a DA.

[1] M. Nagata and T. Morimoto, “First steps toward statistical modeling of dialogue to predict the speech act type of the next utterance”, Speech Communication, vol. 15, 1994

[2] N. Webb et al. “Dialog act classification based on intra-utterance features”, CS-05-01, University of Sheffield, UK, 2005

[3] J. Ang et al. “Automatic dialog act segmentation and classification in multiparty meetings”, ICASSP, 2005

[4] R. E. Schapire and Y. Singer, “BoosTexter: A boosting-based system for text categorization”, Machine Learning, vol. 39, no. 2-3, 2000