Joint Segmentation and Classification of Dialog Acts using Conditional Random Fields

M. Zimmermann
matthias.zimmermann@xbrain.ch

xbrain.ch, Switzerland

[xbrain.ch]

Abstract

We investigate the use of Conditional Random Fields (CRF) for joint segmentation and classification of Dialog Acts (DA) exploiting information extracted from the stream of words and durational prosodic features.

To validate the approach experiments are conducted under a variety of setups:

- Reference, and Speech-to-Text (STT) conditions
- 5 Different coding schemes
- 2 Different class maps with 5, resp. 17 DA types

Task

The task considered in this work is to segment a sequence of words into individual DAs and assign mutually exclusive DA labels to the segments at the same time.

Examples for class map Map1 with five DA types:

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

Class map Map5b further breaks down some of the groups: Questions are split into yes/no-questions, wh-questions, etc. In the case of statements, responses to questions (rejection, acceptance, and uncertain) as well as action motivators (combining suggestions, commands, and commitments) are separately identified.

Acknowledgment: This work is partly based on prior research at the International Computer Science Institute (ICSI), USA.

Method

Five different coding schemes are used to map the joint task to the CRF framework. Class labels that encode both DA type and boundary information are assigned to the individual words of the word stream**:**

Reference	| S S | Q Q Q Q | D D D | B |
Coding E	n S> n n n Q> n n D> B>
Coding B	<S n <Q n n n <D n n <B
Coding EI	s S> q q q Q> d d D> B>
Coding BI	<S s <Q q q q <D d d <B
Coding BIE	<S S> <Q q q Q> <D d D> <B>

From coding E to coding BIE an increasing amount of information is preserved projecting the 5(17) DA types into a 6(18)-class problem for coding E, and to a 20(68)-class problem for coding BIE respectively.

Definition of performance metrics used to tune and evaluate the systems:

Reference	|S S|Q Q Q Q|D D D|B|
System Output	|S|S|Q Q Q|S S S S|B|
Words	E E E E E E E E E C
DAs	E | E | E |C|

Metric	Counts	Reference	Rate
F Measure	2 RP/(R + P)	R, P	22.2
Recall R	1 correct DA	4 DAs	25.0
Precision P	1 correct DA	5 DAs	20.0
Strict	9 match errors	10 words	90.0
DER*	3 match errors	4 DAs	75.0

*: DER = 1 - Recall

Experiments

Results for some of the different coding schemes:

Scheme	F Measure	Strict	DER
E	52.3	61.9	52.0
B	50.1	63.7	54.7
EI	52.6	61.6	51.6
BIE	53.8	60.0	50.4

Results under different conditions, different class maps, and comparison to previous work:

Cond.	Class Map	System	F Measure	Strict	DER
Ref	Map1 (5)	[2]	n/a	62.8	51.0
Ref	Map1 (5)	[3]	n/a	n/a	62.7
Ref	Map1 (5)	CRF	53.8	60.0	50.4
Ref	Map5b (17)	CRF	46.9	64.7	53.7
STT	Map 1 (5)	[2]	n/a	73.6	62.6
STT	Map 1 (5)	CRF	44.5	69.9	59.9
STT	Map5b (17)	CRF	39.2	72.4	62.0

Conclusion

The proposed methodology based on CRF is conceptually simple and outperforms all previous systems for the task based on the ICSI (MRDA) corpus.

However, detection of disruptions need further attention. To improve detection of responses and action motivators (as defined in the detailed class map) it might help to consider speaker interactions.

[1] E. Shriberg et al., “The ICSI meeting recorder dialog act (MRDA) corpus”, in Proc. SIGDIAL, pages. 97–100, Cambridge, USA , 2004.

[2] M. Zimmermann, A. Stolcke, E. Shriberg, “Joint Segmentation and Classification of Dialog Acts in Multi-party Meetings”, ICASSP, pages 581-584, Toulouse, France, 2006

[3] A. Dielmann and S. Renals, “Recognition of dialogue acts in multiparty meetings using a switching DBN,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 7, 2008.