University of Zagreb
Faculty of Electrical Engineering and Computing
Text Analysis and Knowledge Engineering Lab

take[lab];

Speech Act Based Classification of Email Messages in Croatian Language

Tin Franović, Jan Šnajder
{tin.franovic,jan.snajder}@fer.hr

Eighth Language Technologies Conference (LTC IS-2012)
Ljubljana, October 8th, 2012

Background & motivation

take[lab];

- Increase in popularity of email as means of communication
- Recent surveys – up to 2 hours a day spent on emails
- Automated **email classification** can reduce the amount of time users spend reading and sorting emails

Speech acts (Searle, 1965)

Speech acts are illocutionary acts that attempt to convey meaning from the speaker (or writer) to the listener (or reader)

- Speech acts are effective way of **summarizing the intended purpose** of the message

Background & motivation

take[lab];

- Increase in popularity of email as means of communication
- Recent surveys – up to 2 hours a day spent on emails
- Automated **email classification** can reduce the amount of time users spend reading and sorting emails

Speech acts (Searle, 1965)

Speech acts are illocutionary acts that attempt to convey meaning from the speaker (or writer) to the listener (or reader)

- Speech acts are effective way of **summarizing the intended purpose** of the message

Goal & methodology

take[lab];

Our goal

Develop and evaluate **speech act classification** of email messageg in Croatian language using **supervised machine learning**

- Task framed as a **multilabel text classification problem**
- Thorough evaluation using six machine learning algorithms
- Evaluated using message-level, paragraph-level, and sentence-level features

Goal & methodology

Our goal

Develop and evaluate **speech act classification** of email messageg in Croatian language using **supervised machine learning**

- Task framed as a **multilabel text classification problem**
- Thorough evaluation using six machine learning algorithms
- Evaluated using message-level, paragraph-level, and sentence-level features

Coming up next...

1. Message classification
   - Dataset
   - Message preprocessing
   - Training classifiers
2. Evaluation
3. Conclusion and future work

Dataset annotation

take[lab];

- Several publicly available email datasets, however none in Croatian
- We compiled a dataset using 1337 messages from five sources
- Annotated using 13 different speech acts [Searle, 1965]
  - Assertives (AMEND, PREDICT, CONCLUDE);
  - Directives (REQUEST, REMIND, SUGGEST);
  - Expressives (APOLOGIZE, GREET, THANK);
  - Commisives (COMMIT, REFUSE, WARN);
  - Declarations (DELIVER).

Dataset annotation

take[lab];

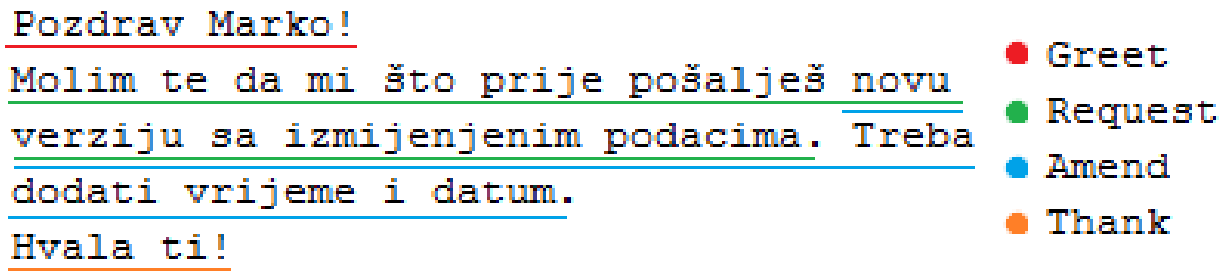
- Two annotators, 15% of dataset double-annotated

Speech act	κ	Speech act	κ
AMEND	0.714	REFUSE	0.000
APOLOGIZE	0.856	REMIND	0.747
COMMIT	0.851	REQUEST	0.589
CONCLUDE	0.005	SUGGEST	0.544
DELIVER	0.792	THANK	0.949
GREET	0.779	WARN	0.174
PREDICT	0.267		

Dataset annotation

take[lab];

- Infrequent and low-IAA speech acts removed:
  - APOLOGIZE, CONCLUDE, GREET, PREDICT, REFUSE, THANK, WARN
- Speech acts used:
  - DELIVER, AMEND, COMMIT, REMIND, SUGGEST, REQUEST



Message preprocessing

take[lab];

- Reduce the dimensionality and morphological variation
- Stemming
  - Suffix of each word after last vowel removed
  - Number of terms reduced from 15,100 to 11,856
- Stop-word removal
  - Filtered out words with little semantic information
  - List of 2,024 Croatian stop-words

Message preprocessing (2)

take[lab];

- Separate training set created for each speech act using annotated data
- Text segments extracted at corresponding discourse levels
  - Sentence and paragraph levels – segments that enclose start and end point of annotation
  - Message level – complete message
- Negative examples sampled from the set of segments not annotated with the corresponding speech act

Training classifiers

take[lab];

- Rapid Miner implementation
- **Six different models**:
  - SVMs (Support Vector Machines), naive Bayes (NB), k-NN (k-Nearest Neighbors), Decision Stump (DS), AdaBoost (with Decision Stump as the weaker learner), and RDR (Ripple Down Rule)
- **Three term weighting schemes**:
  - TF (Term Frequency) and TF-IDF (Term Frequency – Inverted Document Frequency) - all models except RDR
  - Binary weights - only RDR
- Separate classifier trained for every speech act, term weighting scheme, and discourse level (198 models)
- Re-trained using stop-word removal

Training classifiers (2)

take[lab];

- Parameter optimization
  - Grid-search
  - 10-fold cross-validation for every parameter combination
  - Optimal parameter chosen based on averaged F1 score
- Optimal model re-trained using whole training set and tested on held-out set
- 70% for training/validation, 30% held-out test set

Classifier performance

take[lab];

- F1 performance for best feature/discourse level combinations:

	NB	k-NN	SVM	DS	AB	RDR
DELIVER	69.70	83.72	88.16	85.71	87.50	**88.51**
AMEND	**79.31**	71.43	77.97	72.29	74.63	77.27
COMMIT	62.45	67.44	78.61	79.37	81.97	**83.75**
REMIND	60.87	63.64	75.00	76.92	**94.74**	76.92
SUGGEST	67.06	70.27	**76.84**	76.27	75.12	71.50
REQUEST	69.69	75.44	**78.76**	70.57	75.23	74.46

Discourse level

take[lab];

- F1 performance for best classifier/feature combinations:

	Message	Paragraph	Sentence
DELIVER	86.59	83.64	**88.51**
AMEND	**79.31**	77.27	72.38
COMMIT	**83.75**	81.97	78.93
REMIND	**94.74**	76.92	69.57
SUGGEST	71.88	**76.84**	69.74
REQUEST	70.09	**78.76**	72.19
Overall	94.74	83.64	78.93

Feature types

take[lab];

- F1 performance for best classifier/discourse level combinations:

	With stop-words			Without stop-words		
	Binary	TF	TF-IDF	Binary	TF	TF-IDF
DELIVER	**88.51**	87.50	88.00	**88.51**	88.16	87.96
AMEND	70.07	77.19	**79.31**	77.27	75.86	**77.19**
COMMIT	**83.75**	79.37	81.63	78.82	79.76	**81.97**
REMIND	76.92	76.92	**77.78**	75.00	**94.74**	77.78
SUGGEST	71.50	**76.84**	76.27	68.40	73.08	**73.68**
REQUEST	61.90	**78.76**	78.10	74.46	**78.08**	77.53

Overall performance

take[lab];

- F1 performance with optimal feature sets for each classifier, averaged over speech acts:

	Message	Paragraph	Sentence
NB	**79.31**	69.70	72.38
k-NN	72.73	75.44	**83.72**
SVM	83.87	81.55	**88.16**
DS	78.65	79.37	**85.71**
AB	**94.74**	83.54	87.50
RDR	86.59	83.64	**88.51**

Conclusion

take[lab];

- Addressed multilabel speech act classification for Croatian
- Thorough evaluation using six machine learning algorithms and three feature types
- Discourse level and feature type do not influence significantly classification performance
- Certain speech acts more accurately classified on particular levels
- Obtained F1 scores notably higher than reported in previous work [Cohen, 2004; Carvalho, 2006]

Future work

- Future work
  - Explore relationship between discourse level and speech acts
  - Employ information extraction methods to augment speech acts
  - Impact of speech acts on importance-based classification

Thank you for your attention

take[lab];

Let's keep in touch. . .

www.takelab.hr

info@takelab.hr