PhD Report – Spring 2014

Wladimir Sidorenko
uladzimir.sidarenka@uni-potsdam.de

University of Potsdam

April 24, 2014

Table of Contents

Composition Criteria

For developing and testing our sentiment analysis system, we have created a corpus of 3996 Twitter messages. This corpus consists of four major topic parts (two political and two non-political ones) each of which was sampled using three disjoint selection criteria.

Composition Criteria

The covered topics are:

1. Political:
   - Tweets containing political terms (March 27 – May 25, 2013);
   - Tweets pertaining to the federal election 2013 (June 15 – September 30, 2013);
2. Non-political:
   - General tweets with no particular topic (March 31 – April 30, 2013);
   - Tweets pertaining to the pope election 2013 (March 13 – March 14, 2013).

Composition Criteria

The selection criteria for each of the topics are:

1. Presence of polar terms (SentiWS [4]);
2. Presence of smileys and exclamation marks;
3. Others.

For each of the above criteria, we sampled 333 messages for each topic. All messages were sampled disjointly so that tweets which fell into one of the preceding categories were excluded from the next ones.

Annotation Scheme

Emo-expressions (*expressive subjective elements* [7]) - lexical items with polar evaluative sense, e.g. *gut, schrecklich, kritisieren, zum Besten halten etc.*;
Diminishers (*down-turners* [5]) - words or phrases which decrease the intensity of an emo-expression term, e.g. *weniger, bisschen, kaum etc.*
Intensifiers - lexical elements which strengthen the polar evaluative sense of an emo-expression, e.g. *recht, super, außerordentlich etc.*
Negations - language elements which reverse the polarity of subjective meaning expressed by an ESE, e.g. *nicht, kein, etc.*

Annotation Scheme

Sentiment - minimal complete coherent syntactic or discourse-level unit that expresses a polar evaluative opinion of a person or organization about some particular subject, topic, or event, e.g. *Ich hasse diese Reform*, *ein ausgezeichneter Film*, *Meine Mutter ruft mich heulend an. Man hat einen Argentinier zum Papst gewählt.*;
Source - the immediate originator of a polar evaluative opinion who either directly expresses her opinion or whose opinion is being cited;
Target - subject or event which is being evaluated in a sentiment.

Annotation Scheme

Example

[[Ich]$_{\textbf{source}}$[hasse]$_{\textbf{emo-expression}}$[Merkel]$_{\textbf{target}}$]$_{\textbf{sentiment}}$.

Preliminary Statistics

Table: Distribution of emotional expressions across topics and selection criteria in corpus.

Selection Criterion	Politics		Non-politics	
	General Politics	Federal Election	General Discussions	Pope Election
Polar Terms	225	199	270	163
Emoticons	426	415	457	364
Other	76	75	82	54

Preliminary Statistics

Table: Distribution of sentiments across topics and selection criteria in corpus.

Selection Criterion	Politics		Non-politics	
	General Politics	Federal Election	General Discussions	Pope Election
Polar Terms	90	105	79	83
Emoticons	68	71	35	50
Other	54	46	17	30

Classifiers

Table: Classification results for automatic sentiment analysis (token-based).

ML-System	Sentiment	Source	Target	Other
MLN	na	na	na	na
SVM	3.4	10.7	0	94.5
Bayes Net	15.7	9.4	5.8	89
NB	15.9	7.5	8.9	78.4
Multinomial NB	17.5	9.8	11	85.6
CRF	16.53	17.65	7.89	94.47

CRF

Features:

- Formal:
  - Initial three characters of word form;
  - Final three characters of word form;
  - Character class of word (title, upper, lower, alphabetic mixed, alnum, digit, punct, mixed);
- Morphological:
  - Case;
  - Gender;
  - Degree of Comparison;
  - Mood;
  - Tense;
  - Person;
- Lexical:
  - Word Form;
  - Polarity Score (SentiWS* [4] and GermanPolarityClues [6]);
  - Class of modal verb (lexical or true modal);
- Syntactical:
  - Dependency relation of preceding and current word;
  - Dependency relation of current word;
  - Dependency relation of current and next word;
  - Lemma of parent;
  - PoS-Tag of grandmother;
  - Form of grandmother;
  - Polarity class of grandmother;
  - Child Lemma + Dependency Relation;
  - Child Lemma + Dependency Relation + Lemma;
  - Child PoS-Tag + Dependency Relation + PoS-Tag;
  - Cummulative polarity class for children (polarity class of the sum of children's scores);

Evaluation

Evaluation schemes:

- Binary Overlap [1]:
  Precision $= \frac{|\{p|p \in P \wedge \exists c \in C \text{ s.t. } f(c,p)\}|}{|P|}$; Recall $= \frac{|\{c|c \in C \wedge \exists p \in P \text{ s.t. } f(c,p)\}|}{|C|}$;
  where C is the set of correct spans, P is the set of predicted spans, and $f(c, p)$ is a function which yields "true" if the spans overlap and "false" otherwise;

Evaluation

Evaluation schemes:

- Binary Overlap [1]:
  Precision $= \frac{|\{p|p \in P \wedge \exists c \in C \text{ s.t. } f(c,p)\}|}{|P|}$; Recall $= \frac{|\{c|c \in C \wedge \exists p \in P \text{ s.t. } f(c,p)\}|}{|C|}$;
  where C is the set of correct spans, P is the set of predicted spans, and $f(c, p)$ is a function which yields "true" if the spans overlap and "false" otherwise;
- Proportional Overlap [3]:
  Precision $= \frac{\text{Score}(C,P)}{|P|}$; Recall $= \frac{\text{Score}(P,C)}{|C|}$;
  where $\text{Score}(S, S') = \sum_{s \in S} \sum_{s' \in S'} f(s, s')$ and $f(s, s') = \frac{|s \cap s'|}{|s'|}$;

Evaluation

Evaluation schemes:

- Binary Overlap [1]:
  Precision $= \frac{|\{p|p \in P \wedge \exists c \in C \text{ s.t. } f(c,p)\}|}{|P|}$; Recall $= \frac{|\{c|c \in C \wedge \exists p \in P \text{ s.t. } f(c,p)\}|}{|C|}$;
  where C is the set of correct spans, P is the set of predicted spans, and $f(c, p)$ is a function which yields "true" if the spans overlap and "false" otherwise;
- Proportional Overlap [3]:
  Precision $= \frac{\text{Score}(C,P)}{|P|}$; Recall $= \frac{\text{Score}(P,C)}{|C|}$;
  where $\text{Score}(S, S') = \sum_{s \in S} \sum_{s' \in S'} f(s, s')$ and $f(s, s') = \frac{|s \cap s'|}{|s'|}$;
- Exact Match [1]:
  the same as binary overlap except that $f(c, p)$ yields "true" iff the compared spans completely agree on their boundaries.

Evaluation

Table: Classification results for automatic sentiment analysis (binary overlap).

Classification Element	Precision	Recall	F-Measure
Training Set			
Sentiment	99.23	86.27	92.29
Source	91.56	75.55	82.78
Target	95.99	75.69	84.64
Test Set			
Sentiment	25	16.04	19.55
Source	47.06	25	32.65
Target	31.51	18.11	23

Evaluation

Table: Classification results for automatic sentiment analysis (binary overlap).
Sentiment is ESE

Classification Element	Precision	Recall	F-Measure
Training Set			
Sentiment	94.38	81.43	87.43
Source	92.31	48.54	63.62
Target	96.95	56.83	71.66
Test Set			
Sentiment	76.54	68.5	72.29
Source	25	18.75	21.43
Target	15.46	11.81	13.39

Evaluation

Table: Classification results for automatic sentiment analysis (proportional overlap).

Classification Element	Precision	Recall	F-Measure
Training Set			
Sentiment	97.62	84.94	90.84
Source	90.4	73.71	81.21
Target	93.55	74.02	82.65
Test Set			
Sentiment	21.31	14.53	17.28
Source	40	25	30.77
Target	26.06	13.75	18

Evaluation

Table: Classification results for automatic sentiment analysis (proportional overlap). Sentiment is ESE

Classification Element	Precision	Recall	F-Measure
Training Set			
Sentiment	93.62	80.5	86.57
Source	92.07	48.26	63.33
Target	94.39	55.58	69.96
Test Set			
Sentiment	74.38	67.27	70.65
Source	22.22	18.75	20.34
Target	12.16	10.56	11.3

Evaluation

Table: Classification results for automatic sentiment analysis (exact match).

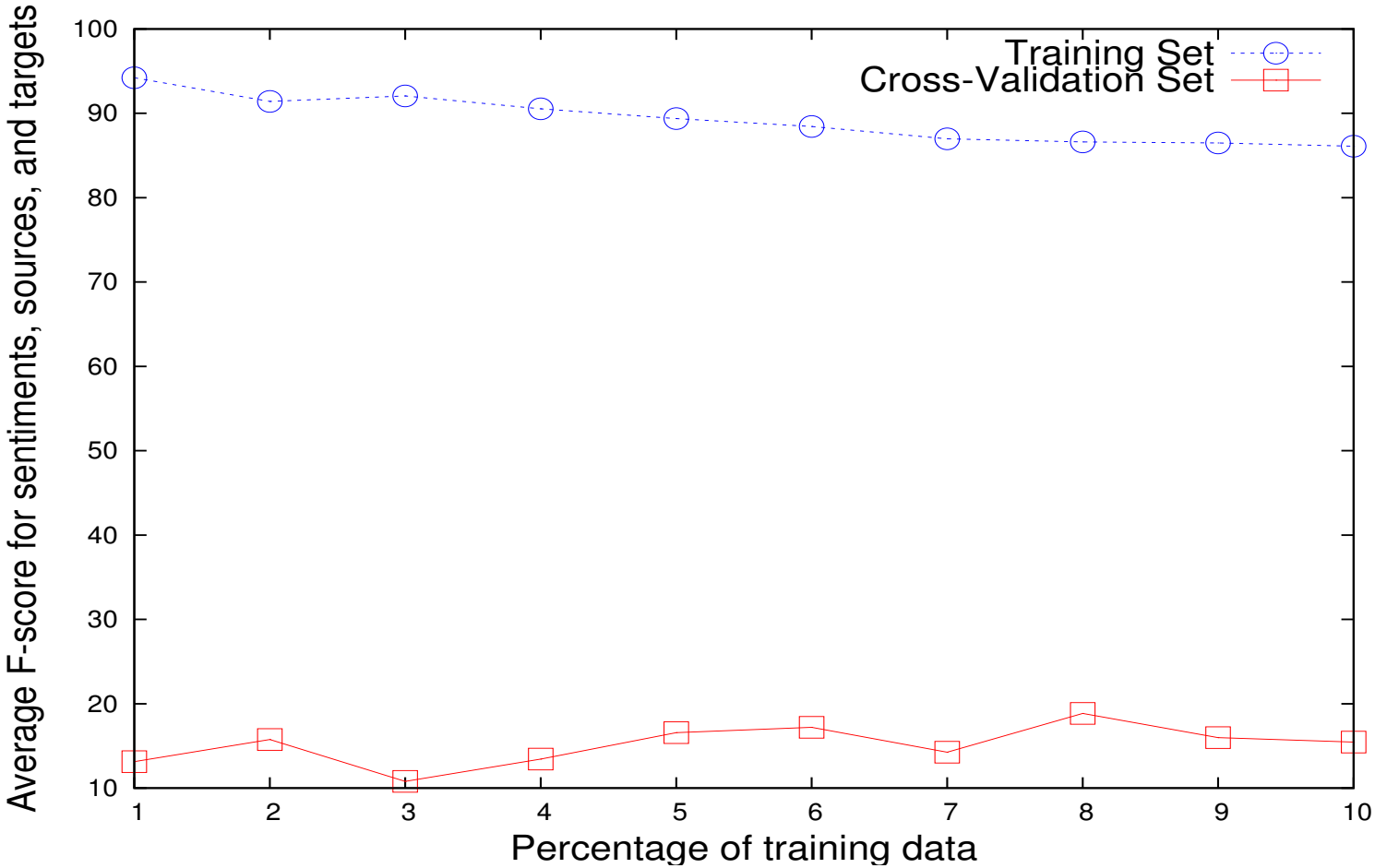
Classification Element	Precision	Recall	F-Measure
Training Set			
Sentiment	87.37	72.7	79.36
Source	88.24	71.17	78.79
Target	85.54	66.44	74.79
Test Set			
Sentiment	13.95	9.09	11.01
Source	40	25	30.77
Target	14.67	8.66	10.89

Evaluation

Table: Classification results for automatic sentiment analysis (exact match). Sentiment is ESE

Classification Element	Precision	Recall	F-Measure
Training Set			
Sentiment	90.9	78.39	84.18
Source	89.51	46.72	61.39
Target	80.08	45.6	58.11
Test Set			
Sentiment	70.84	63.21	66.81
Source	20.83	15.62	17.86
Target	8.25	6.3	7.14

Learning Curve



Problems and Open Questions

- Bad overfitting;
- Polarity detection;
- Flat tagging scheme;
- Relation linking.

Preliminary Conclusions and Perspectives

Conclusions:

- Preprocessing matters (w/25.067 vs. wo/18.277);
- Quality of polarity dictionaries is important (sentiws/25.067 vs. gpc/23.903);

Perspectives:

- Different classifiers (higher order CRFs, structural SVMs, etc.);
- Experiments with polarity dictionaries and ontologies;

Connectives

Table: Approximate distribution of discourse connectives in corpus.

Top-10 Connectives	# of tweets	
	Non-Answers	Answers
und	2,787,798	769,301
für	1,734,911	267,975
auch	596,142	599,445
wie	559,449	287,471
aber	366,489	397,896
oder	318,181	172,969
dass	301,296	131,624
ja	290,461	414,543
dann	288,707	241,499
da	269,446	228,490
Percent of tweets with connectives	37%	49%

Discourse Relations

Table: Approximate distribution of discourse relations in corpus.

Top-10 Discourse Relations	# of tweets	
	Non-Answers	Answers
Elaboration*	4,620,913	1,917,685
Sequence	1,397,751	735,492
Concession*	106,0244	894,494
Circumstance	928,997	400,505
Contrast*	918,498	777,630
Cause	895,181	624,868
Anthithese*	471,187	518,792
Joint	401,916	285,196
Substitution	104,074	25,067
Unconditional	99,904	41,589

Discourse Segmenter

! @DJVBB Wulff hat nicht getäuscht, sondern wegen der Mehrfachbelastung als junger Familienvater den Überblick über die Geldquellen verloren.

(**"sondern"**
(**intern**: wegen/APPR der/ART Mehrfachbelastung/NN als/KOKOM junger/ADJA Familienvater/NN den/ART Überblick/NN über/APPR die/ART Geldquellen/NN verloren/VVFIN)
(**extern**: Wulff/NE hat/VAFIN nicht/PTKNEG getäuscht/VVPP ,/$,))

(**"wegen"**
(**intern**: der/ART Mehrfachbelastung/NN)
(**extern**: Wulff/NE hat/VAFIN nicht/PTKNEG getäuscht/VVPP ,/$, sondern/KON wegen/APPR der/ART Mehrfachbelastung/NN als/KOKOM junger/ADJA Familienvater/NN den/ART Überblick/NN ber/APPR die/ART Geldquellen/NN verloren/VVFIN))

Perspectives and Future Work

- Disambiguation of discourse connectives and relations;
- Analysis of discourse relations within and among the tweets (inter- and intra-tweet relations);
- Analysis of discourse relations and connectives in tweets expressing sentiments;
- Analysis of tweets expressing CONTRAST and ELEBORATION relations for sentiments.


Eric Breck, Yejin Choi, and Claire Cardie.
Identifying expressions of opinion in context.
In Manuela M. Veloso, editor, *IJCAI*, pages 2683–2688, 2007.


Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors.
Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta. European Language Resources Association, 2010.


Richard Johansson and Alessandro Moschitti.
Reranking models in fine-grained opinion analysis.
In Chu-Ren Huang and Dan Jurafsky, editors, *COLING*, pages 519–527. Tsinghua University Press, 2010.

Robert Remus, Uwe Quasthoff, and Gerhard Heyer.
Sentiws - a publicly available german-language resource for sentiment analysis.
In Calzolari et al. [2].

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly D. Voll, and Manfred Stede.
Lexicon-based methods for sentiment analysis.
Computational Linguistics, 37(2):267–307, 2011.

Ulli Waltinger.
Germanpolarityclues: A lexical resource for german sentiment analysis.
In Calzolari et al. [2].

Janyce Wiebe, Theresa Wilson, and Claire Cardie.
Annotating expressions of opinions and emotions in language.

Language Resources and Evaluation, 39(2-3):165–210, 2005.