Sentiment Lexicons for Arabic Social Media

Saif M. Mohammad[1], Mohammad Salameh[2], Svetlana Kiritchenko[1]
[1]National Research Council Canada, [2]University of Alberta

Sentiment Analysis in Resource-Poor Languages

- Sentiment analysis research: predominantly on English texts
- Fewer resources for other languages
- Approaches to improve sentiment analysis in a resource-poor language, that leverage these English resources can be of two kinds:
  a. translate the focus language text to English; apply a powerful English sentiment analysis system.
  b. translate resources such as sentiment labeled corpora and sentiment lexicons from English into the focus language; use them as additional resources in the focus-language sentiment analysis system.

Sentiment Analysis in Resource-Poor Languages: Our Broader Work

- We show that sentiment analysis of English translations of Arabic texts produces competitive results, compared to direct sentiment analysis of Arabic texts.
  - manual annotation study
    - how often and why the sentiment of a translation is different from the sentiment of the source text
- We show that Arabic sentiment analysis systems benefit from the use of automatically translated English sentiment lexicons.
  - manual annotation study
    - how often and why the sentiment of a translation is different from the sentiment of the source word

Manual studies are useful for building better automatic translation systems.

Sentiment Analysis in Resource-Poor Languages: This Talk

- We show that sentiment analysis of English translations of Arabic texts produces competitive results, compared to direct sentiment analysis of Arabic texts.
  - manual annotation study
    - how often and why the sentiment of a translation is different from the sentiment of the source text
- We show that Arabic sentiment analysis systems benefit from the use of automatically translated English sentiment lexicons.
  - manual annotation study
    - how often and why the sentiment of a translation is different from the sentiment of the source word

Manual studies are useful for building better automatic translation systems.

Sentiment Analysis in Resource-Poor Languages: Our Work

- created a state-of-the-art Arabic sentiment analysis system
- the first Arabic–English parallel corpus, where:
  - the Arabic text annotated for sentiment by Arabic speakers
  - the English text annotated for sentiment by English speakers

Sentiment After Translation: A Case-Study on Arabic Social Media Posts. Mohammad Salameh, Saif M Mohammad and Svetlana Kiritchenko, In Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL-2015), June 2015, Denver, Colorado.

How Translation Alters Sentiment. Saif M. Mohammad, Mohammad Salameh, and Svetlana Kiritchenko, Journal of Artificial Intelligence Research, 2016, Volume 55, pages 95-130.

Sentimen Lexicons

capturing word-sentiment associations

Word-Sentiment Associations

- Adjectives
  - reliable / موثوق and stunning / مذهل are typically associated with positive sentiment
  - rude / وقح and broken / مكسور are typically associated with negative sentiment
- Nouns and verbs
  - holiday / عطلة and smiling / يبتسم are typically associated positive sentiment
  - death / موت and crying / يبكي are typically associated with negative sentiment

Sentiment Lexicons

Lists of positive and negative words.

Positive

spectacular

okay

Negative

lousy

unpredictable

Sentiment Lexicons

Lists of positive and negative words

- optionally, with scores indicating the degree of association

Positive

spectacular 0.91

okay 0.3

Negative

lousy -0.84

unpredictable -0.17

Sentiment Analysis in Twitter
SemEval-2013, Task 2

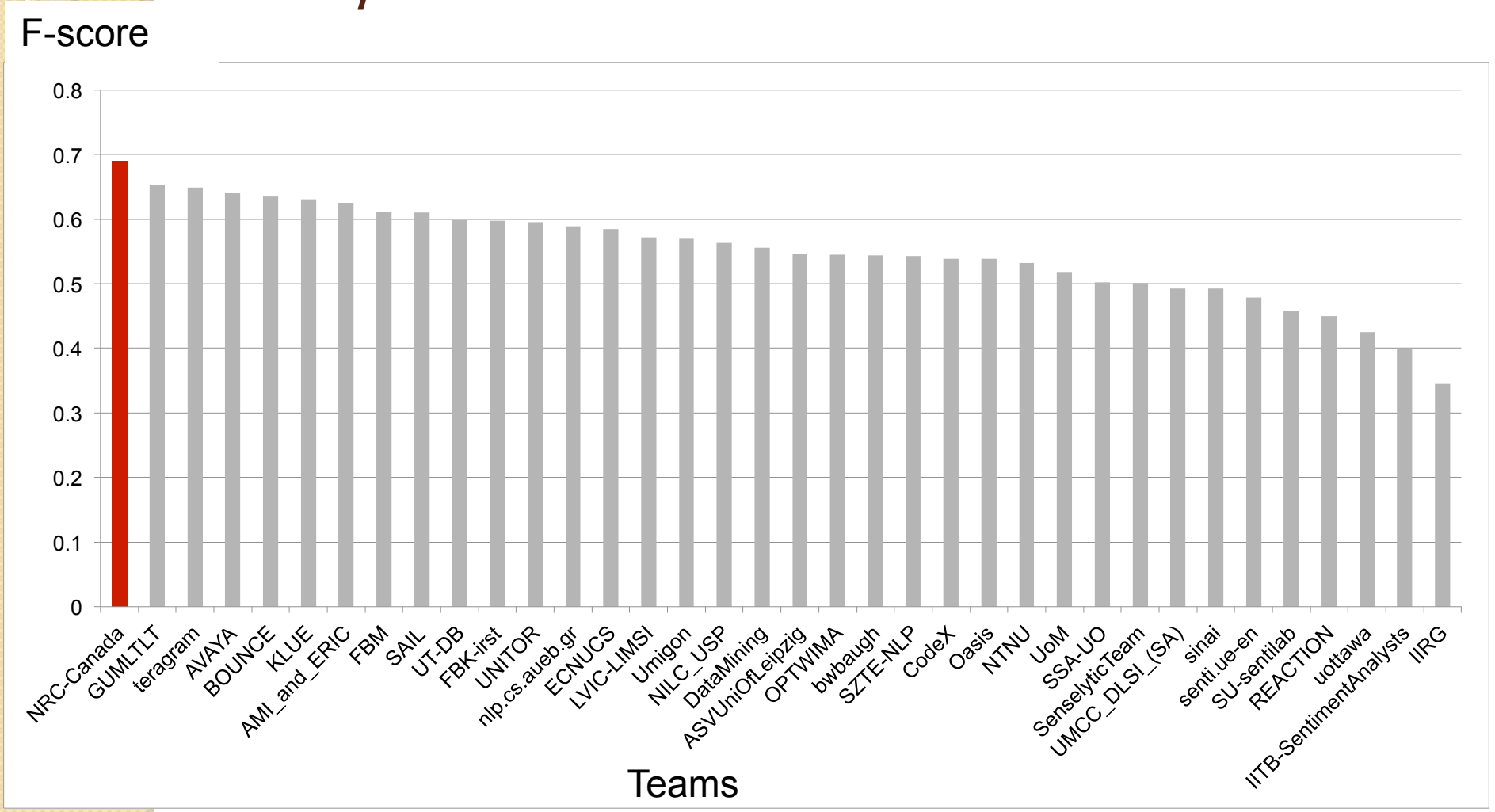
- Is a given message positive, negative, or neutral?
  - tweet or SMS
- Is a given term within a message positive, negative, or neutral?

Team NRC-Canada: SVM + array of surface-form and lexical features

NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu, In Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013), June 2013, Atlanta, USA.

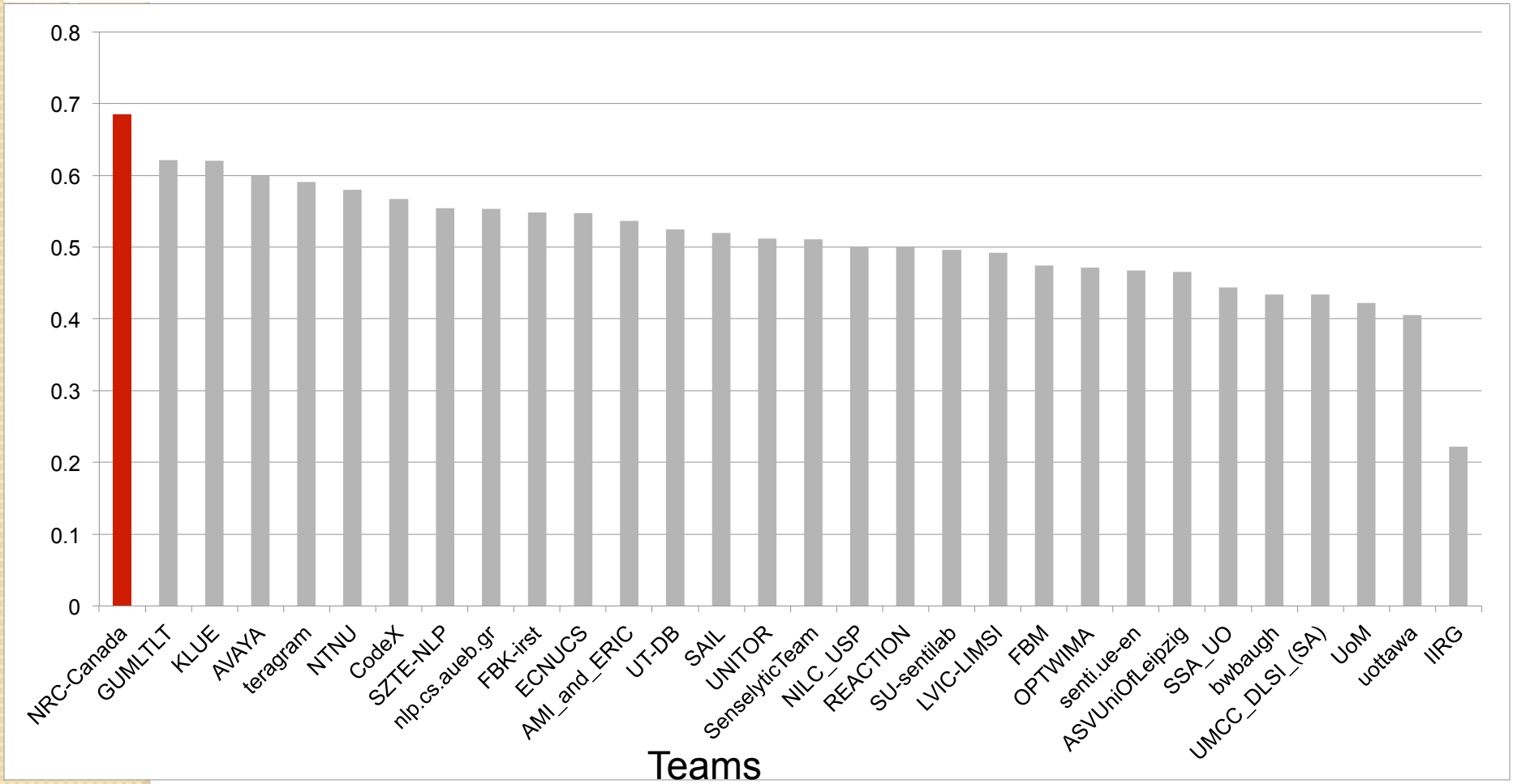
Sentiment Analysis Competition

Classify Tweets

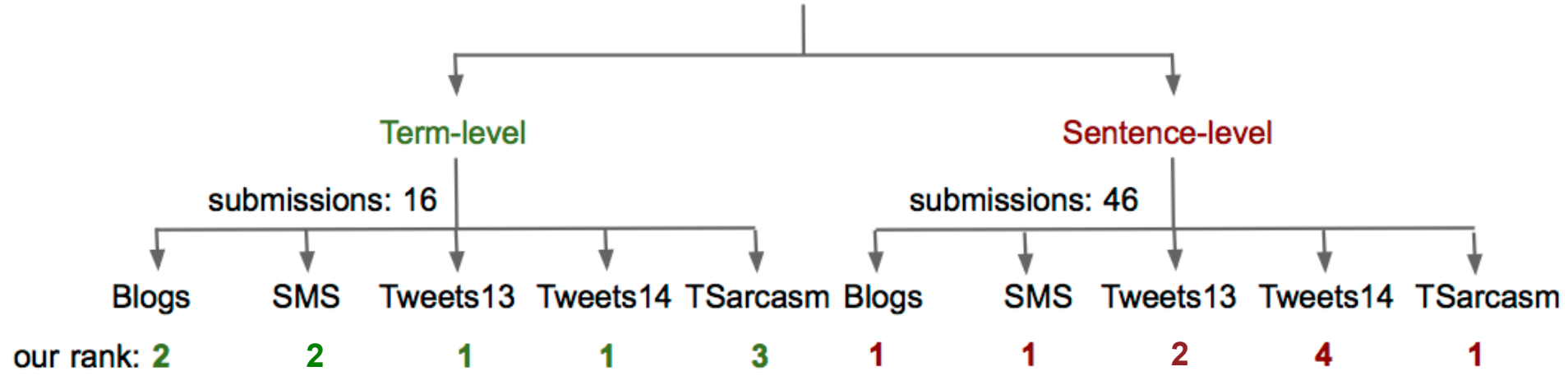


Sentiment Analysis Competition

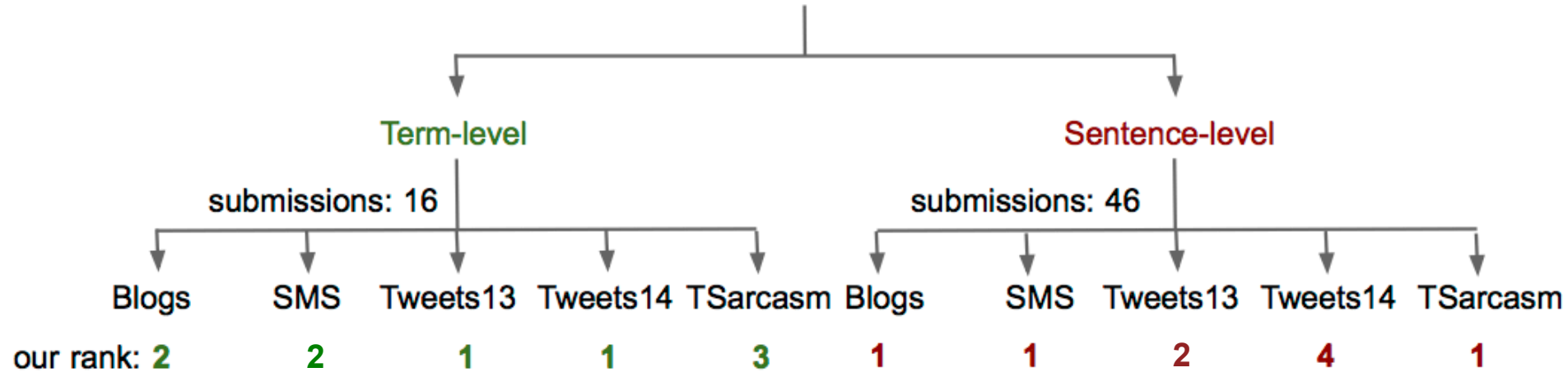
Classify SMS



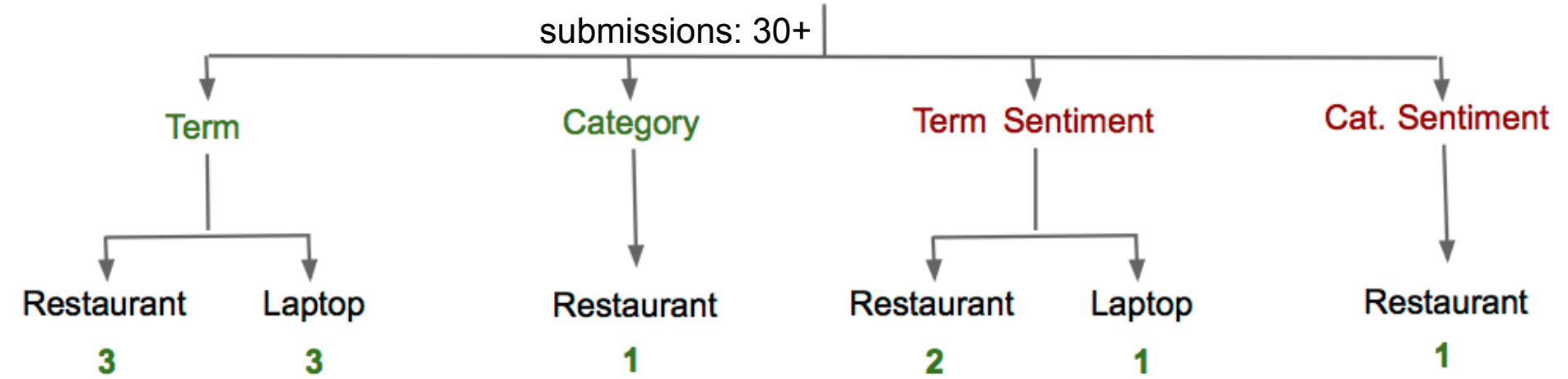
Sentiment Analysis of Social Media Texts (SemEval-2014 Task 9)



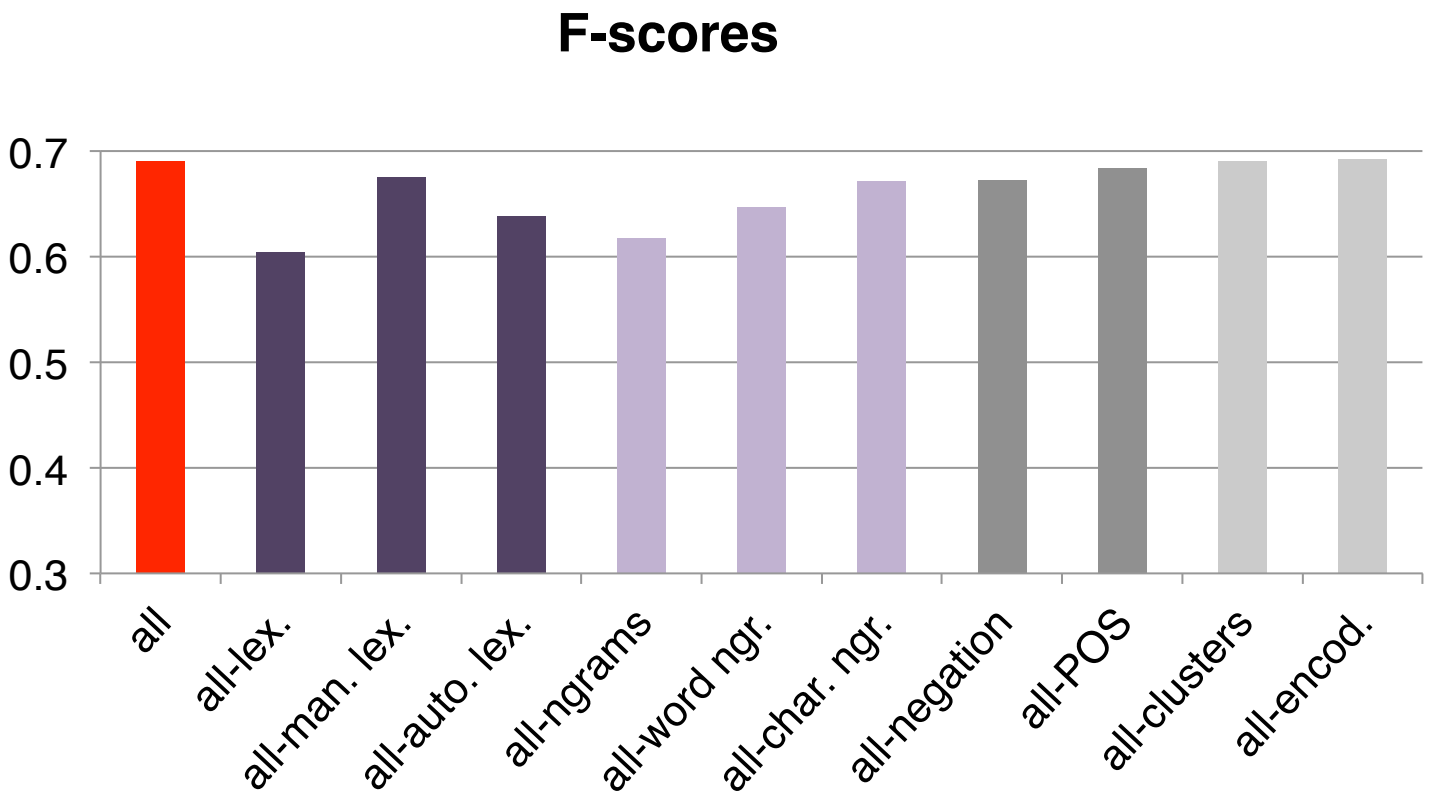
Sentiment Analysis of Social Media Texts (SemEval-2014 Task 9)



Aspect-Based Sentiment Analysis (SemEval-2014 Task 4)



Feature Contributions (on Tweets)



Goals of this work

- generate Arabic sentiment lexicons
  - new lexicons created from Arabic texts
    - tens of thousands of entries
    - with scores of association to sentiment
  - translations of English lexicons

- apply the lexicons in a supervised sentiment classification task on sentences/tweets from Arabic social media
  - evaluate usefulness of these lexicons

- conduct qualitative study on translated entries
  - how often and why the sentiment of a translation is different from the sentiment of the source word

New Arabic Sentiment Lexicons

- the emoticons and hashtag words in a tweet can often act as sentiment labels for the rest of the tweet.

  Super awesome to be here at LREC :)

  Some jerk just stole my photo on #tumblr #grrr #anger

- hashtags and emoticons are not always good labels:
  - used sarcastically

    The reviewers want me to re-annotate the data. #joy
  - affect not present in the rest of the message

    Mika used my photo on tumblr. #anger

Papers:

Twitter sentiment classification using distant supervision. Go, A., Bhayani, R., and Huang, L. Technical report, Stanford University, 2009.

#Emotional Tweets. Saif Mohammad, In Proceedings of the First Joint Conference on Lexical and Computational Semantics (*Sem), June 2012, Montreal, Canada.

New Arabic Sentiment Lexicons (continued)

- Arabic Emoticon Lexicon:
  - collected close to one million Arabic tweets that had emoticons :) (positive) or :((negative)
  - For each word w, a sentiment score was calculated (Mohammad et al. (2013) and Kiritchenko et al. (2014b)):

$$score(w) = PMI(w,positive) - PMI(w,negative)$$

where, PMI = pointwise mutual information

If $score(w) > 0$, then word w is positive

If $score(w) < 0$, then word w is negative

New Arabic Sentiment Lexicons (continued)

- Arabic Hashtag Lexicon:
  - generated using similar method
  - using translations of English seed words
    - which were used to create the NRC Hashtag Lexicon
- Arabic Hashtag Lexicon (Dialectal):
  - generated using similar method
  - used dialectal Arabic words compiled by Refaee and Rieser (2014) as seeds

Generating Arabic Translations of English Sentiment Lexicons

- used Google Translate to translate into Arabic these English sentiment lexicons:
  - AFINN (Nielsen, 2011)
  - Bing Liu Lexicon (Hu and Liu, 2004)
  - MPQA Subjectivity Lexicon (Wilson et al., 2005)
  - NRC Emotion Lexicon (Mohammad and Turney, 2010; 2013)
  - NRC Emoticon Lexicon aka Sentiment140 Lexicon (Mohammad et al., 2013; Kiritchenko et al., 2014b)
  - NRC Hashtag Sentiment Lexicon (Mohammad et al., 2013; Kiritchenko et al., 2014b).

Application

To determine the usefulness of the Arabic sentiment lexicons, we apply them in a sentence-level sentiment analysis system.

Arabic sentiment analysis system

Ported the NRC-Canada English system (Mohammad et al., 2013; Kiritchenko et al., 2014b) to Arabic

- linear-kernel Support Vector Machine (Chang and Lin, 2011)
- features:
  - baseline features
    - the presence/absence of word and character ngrams
    - the presence/absence of all-cap words, hashtags, and punctuation marks
  - lexicon features
    - the number of sentiment words with non-zero sentiment score
    - the sum of sentiment scores of positive words (and separately negative words)
    - the sentiment score of the last token

Sentiment-Labeled Arabic Social Media Data

- BBN posts
  - the BBN Arabic Dialectal Text (Zbib et al., 2012)
    - blog posts
    - mixture of expressions from the Levantine dialect of Arabic as well as Modern Standard Arabic
      https://catalog.ldc.upenn.edu/LDC2012T09
  - we randomly selected a subset of 1200 sentences
    - annotated them for sentiment via crowdsourcing
      http://www.saifmohammad.com/WebPages/ArabicSA.html

Ran cross-validation experiments on this dataset with versions of the sentiment system that used different Arabic lexicons.

System	Accuracy (in percentage)
a. Baseline (uses word ngrams and other surface form features)	62.0
b. Baseline + Arabic lexicon	
Manual lexicons:	
i. Abdul-Mageed et al. (2011) Lexicon	62.2
ii. Refaee and Rieser (2014) Lexicon	63.0
iii. Kiritchenko et al. (2016) Lexicon	62.7

Table 3: Sentiment classification accuracies on the BBN sentences. Highest scores in b., c., and d. are shown in bold.

System	Accuracy (in percentage)
a. Baseline (uses word ngrams and other surface form features)	62.0
b. Baseline + Arabic lexicon	
Manual lexicons:	
i. Abdul-Mageed et al. (2011) Lexicon	62.2
ii. Refaee and Rieser (2014) Lexicon	63.0
iii. Kiritchenko et al. (2016) Lexicon	62.7
Automatic lexicons:	
iv. the Arabic Emoticon Lexicon	62.4
v. the Arabic Hashtag Lexicon	63.0
vi. the Arabic Hashtag Lexicon (dialectal)	**65.3**
vii. lexicon features from iv., v., and vi.	63.5

Table 3: Sentiment classification accuracies on the BBN sentences. Highest scores in b., c., and d. are shown in bold.

System	Accuracy (in percentage)
a. Baseline (uses word ngrams and other surface form features)	62.0
b. Baseline + Arabic lexicon	
Manual lexicons:	
i. Abdul-Mageed et al. (2011) Lexicon	62.2
ii. Refaee and Rieser (2014) Lexicon	63.0
iii. Kiritchenko et al. (2016) Lexicon	62.7
Automatic lexicons:	
iv. the Arabic Emoticon Lexicon	62.4
v. the Arabic Hashtag Lexicon	63.0
vi. the Arabic Hashtag Lexicon (dialectal)	**65.3**
vii. lexicon features from iv., v., and vi.	63.5
c. Baseline + Arabic translation of English lexicon	
Manual lexicons:	
i. English lexicon: AFINN	63.4
ii. English lexicon: Bing Liu Lexicon	63.0
iii. English lexicon: MPQA	61.9
iv. English lexicon: NRC Emotion Lexicon	**63.5**

Table 3: Sentiment classification accuracies on the BBN sentences. Highest scores in b., c., and d. are shown in bold.

System	Accuracy (in percentage)
a. Baseline (uses word ngrams and other surface form features)	62.0
b. Baseline + Arabic lexicon	
Manual lexicons:	
i. Abdul-Mageed et al. (2011) Lexicon	62.2
ii. Refaee and Rieser (2014) Lexicon	63.0
iii. Kiritchenko et al. (2016) Lexicon	62.7
Automatic lexicons:	
iv. the Arabic Emoticon Lexicon	62.4
v. the Arabic Hashtag Lexicon	63.0
vi. the Arabic Hashtag Lexicon (dialectal)	**65.3**
vii. lexicon features from iv., v., and vi.	63.5
c. Baseline + Arabic translation of English lexicon	
Manual lexicons:	
i. English lexicon: AFINN	63.4
ii. English lexicon: Bing Liu Lexicon	63.0
iii. English lexicon: MPQA	61.9
iv. English lexicon: NRC Emotion Lexicon	**63.5**
Automatic lexicons:	
v. English lexicon: NRC Emoticon Lexicon	62.4
vi. English lexicon: NRC Hashtag Lexicon	61.7
d. Baseline + Arabic Hashtag Lexicon (dialectal) + Arabic translation of NRC Emotion Lexicon	**66.6**

Table 3: Sentiment classification accuracies on the BBN sentences. Highest scores in b., c., and d. are shown in bold.

A Manual Study of
Automatically Translated Sentiment Entries

Lexicons created by translating entries in existing English lexicons into Arabic improve sentiment analysis; however,

- to what extent are such translated sentiment entries appropriate?
- in what ways does automatic translation alter the sentiment conveyed by the source word?

Small Manual Annotation Study

- Who:
  - a native speaker of Arabic, who is also fluent in English
- Given:
  - 300 entries from the NRC Emotion Lexicon
    - 100 positive, 100 negative, and 100 neutral
  - Arabic translations of the words (through Google Translate)
- Asked:
  - is the sentiment associated with the English word, also appropriate for the Arabic translation?
  - if not, then specify reason (coarse categories)

Reasons for Altered Sentiment

1. The word is completely mistranslated.
2. The translation is not perfect, but the English word is translated into a word related to the correct translation. The Arabic word provided has a different sentiment than the English source word.
3. The translation is correct, but the Arabic word has a different sentiment than the English source word.
   a) The dominant sense of the Arabic word is different from the dominant sense of the English source word, and they have different sentiments.
   b) Cultural and lifestyle differences between Arabic and English speakers lead to different sentiment associations of the English word and its translation.
   c) Some other reason (specify).

Results of the study

- Appropriate translated sentiment entries
  - 88% of all words
  - 85% of all positive English words
  - 92% of all negative English words
  - 88% of all neutral English words:

- Most of the errors were caused not by gross mistranslations, but by differences in how the word is used in Arabic
  - because a different sense is more dominant

Summary

- Created new Arabic sentiment lexicons
  - using techniques of distant supervision
  - by translating existing English sentiment lexicons into Arabic using Google Translate.
- Showed usefulness in sentiment analysis of social media posts
  - the Arabic Dialectal Hashtag Lexicon
  - Arabic translation of the NRC Emotion Lexicon
- Analyzed automatically translated sentiment entries
  - showed the extent to which sentiment is preserved
  - identified the different reasons that can lead to erroneous entries in the translated lexicon

Arabic Sentiment Analysis Project Homepage

http://saifmohammad.com/WebPages/ArabicSA.html

- Sentiment Lexicons
  - Arabic Emoticon Lexicon
  - Arabic Hashtag Lexicon
  - Arabic Hashtag Lexicon (dialectal)
  - Arabic translation of NRC Emotion Lexicon
  - Arabic translation of NRC Emoticon Lexicon
  - Arabic translation of NRC Hashtag Sentiment Lexicon
  - Arabic translation of Bing Liu's Lexicon
- Sentiment Corpora
  - BBN Blog Posts Sentiment Corpus
  - Syria Tweets Sentiment Corpus

SemEval-2016 Task #7: Determining Sentiment Intensity of English and Arabic Phrases

http://alt.qcri.org/semeval2016/task7/

- Test data
- Development data

Other Related Work

- A subset of the entries from the automatically generated Arabic lexicons were manually annotated for sentiment
  - by Best-Worst Scaling
  - highly reliable, real-valued, manual sentiment association scores
- Semeval-2016 Task #7
  - Determining sentiment intensity of Arabic words and phrases

Semeval-2016 Task 7: Determining Sentiment Intensity of English and Arabic Phrases.
Svetlana Kiritchenko, Saif M. Mohammad, and Mohammad Salameh. In *Proceedings of the International Workshop on Semantic Evaluation (SemEval '16)*. June 2016. San Diego, California.