



# Sentiment Analysis

Xiaodan Zhu

**National Research Council Canada, Ottawa**



# Sentiment Analysis of Social Media Texts

Xiaodan Zhu

**National Research Council Canada, Ottawa**

# Sentiment Analysis

- Is a given piece of text **positive, negative, or neutral?**

# Sentiment Analysis

- Is a given piece of text **positive, negative, or neutral**?
- Semantic differential (Osgood et al., 1957)
  - Three main factors accounted for most of the variation in the connotative meaning of adjectives
    - evaluative: good-bad
    - potency: strong-weak
    - activity: active-passive

# Sentiment Analysis

- Tracking sentiment towards politicians, movies, products
- Security applications
- Detecting happiness and well-being
- Improving customer relation models
- Measuring the impact of activist movements through text generated in social media.
- Identifying what evokes strong sentiment in people
- Improving automatic dialogue systems
- Improving automatic tutoring systems
- Detecting how people use emotion-bearing-words and metaphors to persuade and coerce others



We discuss

- sentiment analysis,
- social media text processing,

and review the following technologies/components:

- lexical semantics,
- classification models,
- sequence labeling models,
- syntactic parsing,
- semantic composition,

through a **cool** application, and several state-of-the-art models.

# Sentiment Analysis

- Is a given piece of text **positive, negative, or neutral**?

# Emotion Analysis

- What emotion is being expressed in a given piece of text?
  - Basic emotions: joy, sadness, fear, anger, surprise...
  - Other emotions: guilt, pride, optimism, frustration,...



# Can a machine feel *love*?

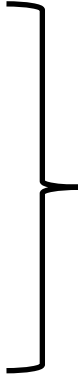
— “*The Emotion Machine*”, Marvin Minsky.



# Social Media Texts

- Large volume: 500 million tweets posted every day!

# Social Media Texts

- Large volume: 500 million tweets posted every day!
  - SMS messages
  - Customer reviews
  - Blog posts
  - Tweets
  - Facebook posts
  - ...and so on.
- 
- Short, informal pieces of text.

# Social Media Texts

- Informal
- Abbreviations and shortenings
- Large vocabulary & wide array of topics
- Spelling mistakes

# Social Media Texts

- Informal
- Abbreviations and shortenings
- Large vocabulary & wide array of topics
- Spelling mistakes

On the other hand:

- Rich information and (noisy) human annotation are freely available.
  - Emoticons: 😊 :-p
  - Hashtags: #loveobama
  - Capital information: *that's really what you MUST TRY*

## Problems

- Message-level sentiment analysis
- Phrase(term)-level sentiment analysis
- Aspect-level sentiment analysis

## Problems

- **Message-level sentiment analysis**
- Phrase(term)-level sentiment analysis
- Aspect-level sentiment analysis

# Message-Level Sentiment: The Task

**Tweet:** Happy birthday, Hank Williams. In honor of the Hank turning 88, we'll play 88 Hank songs in a row tonite @The\_ZOO\_Bar. #honkytonk  
positive

**Tweet:** #Londonriots is trending 3rd worldwide .....  
This is NOT something to be proud of United Kingdom!!! Sort it out!!!!  
negative

**Tweet:** On the night Hank Williams came to town.  
neutral

# Message-Level Sentiment: The Task

**Tweet:** Happy birthday, Hank Williams. In honor of the Hank turning 88, we'll play 88 Hank songs in a row tonite @The\_ZOO\_Bar. #honkytonk  
positive

**Tweet:** #Londonriots is trending 3rd worldwide .....  
This is NOT something to be proud of United Kingdom!!! Sort it out!!!!  
negative

**Tweet:** On the night Hank Williams came to town.  
neutral

(conflicting sentiments vs. target-based sentiment)





# How to decide message-level sentiment?

# You can write rules (Reckman et al., 2013)

- Develop lexicalized hand-written rules: each rule is a pattern that matches words or sequences of words.

- Examples:

**Negative:** `_def{Negation} _def{PositiveAdjectives}`  
`(SENT, (DIST_4, “_a{_def{HigherIsBetter}}”,`  
`“_a{_def{Lowering}}”))`

**Positive:** `(ORDDIST_7, “_def{PositiveContext}”,`  
`“_a{_def{PositiveAmbig}}”)`

- Background data: use blogs, forums, news, and tweets to develop the rules.

# Remarks

- Carefully developed rule-based systems can sometimes achieve complete performance on the data/domains they are created for.
- Advantages: explicit knowledge representation, so intuitive to develop and maintain.
- Problems
  - Coverage: hand-written rules often have limited coverage, so recall is often low. This can impact the overall performance.
  - Extensibility: not easy to be extended to new data/domains; rule-based models have inherent difficulty in automatically acquiring knowledge.
  - Modeling capability, feature interactions, rule conflicts, uncertainty, etc.

## Remarks (continued)

- The main stream is statistical approaches, which achieve top performance across different tasks and data sets.
  - Note that knowledge acquired by applying rules can often be easily incorporated as features into statistical approaches.

# Message-Level Sentiment: The Approach

- Classification
  - Pick your classifier: SVM
  - Pick you kernels?

# How to decide message-level sentiment?

- Features

# Social Media Texts

- Informal
- Abbreviations and shortenings
- Large vocabulary & wide array of topics
- Spelling mistakes

On the other hand:

- Rich information and (noisy) human annotation are freely available.
  - Emoticons: 😊 :-p
  - Hashtags: #loveobama
  - Capital information: *that's really what you MUST TRY*

# How to decide message-level sentiment?

---

Features	Examples
word n-grams	spectacular, like documentary
char n-grams	un, dis, ...
part of speech	#N: 5, #V: 2, #A:1; just; like
word clusters	probably, definitely, def; good; bad;
all-caps	YES, COOL
punctuation	#!+: 1, #?+: 0, #!?: 0
emoticons	:D, >:(
elongated words	coooooool, yaayyy
<b>sentiment lexicon</b>	#positive: 3, scorePositive: 2.2; maxPositive: 1.3; last: 0.6, scoreNegative: 0.8, scorePositive_neg: 0.4
<b>negation</b>	#Neg: 1; ngram:perfect → ngram:perfect_neg, polarity:positive → polarity:positive_neg

---



# How to decide message-level sentiment?

---

Features	Examples
word n-grams	spectacular, like documentary
char n-grams	un, dis, ...
part of speech	#N: 5, #V: 2, #A:1; just; like
word clusters	probably, definitely, def; good; bad;
all-caps	YES, COOL
punctuation	#!+: 1, #?+: 0, #!?: 0
emoticons	:D, >:(
elongated words	coooooool, yaayyy
<b>sentiment lexicon</b>	#positive: 3, scorePositive: 2.2; maxPositive: 1.3; last: 0.6, scoreNegative: 0.8, scorePositive_neg: 0.4
<b>negation</b>	#Neg: 1; ngram:perfect → ngram:perfect_neg, polarity:positive → polarity:positive_neg

---

# Manual Sentiment Lexicons

Lists of positive and negative words:

## **Positive**

spectacular

okay

## **Negative**

lousy

bad

# Sentiment Lexicons: **Manually** Created

- General Inquirer (Stone, Dunphy, Smith, Ogilvie, & associates, 1966): ~3,600 words
- MPQA (Wilson, Wiebe, & Hoffmann, 2005): ~8,000 words
- Hu and Liu Lexicon (Hu and Liu, 2004): ~6,800 words
- NRC Emotion Lexicon (Mohammad & Turney, 2010): ~14,000 words and ~25,000 word senses
  - senses are based on categories in a thesaurus
  - has emotion associations in addition to sentiment
- AFINN (by Finn Årup Nielsen in 2009–2011): ~2400 words
- MaxDiff Sentiment Lexicon (Kiritchenko, Zhu, and Mohammad, 2014): about 1,500 terms
  - has intensity scores



# Sentiment Lexicons

Two major issues:

# Sentiment Lexicons

Two major issues: (1) coverage;

# Sentiment Lexicons

Two major issues: (1) coverage; (2) detailed sentiment scale.

## Positive

spectacular **0.91**

okay **0.30**

## Negative

lousy **-0.84**

bad **-0.97**

# Turney and Littman (2003) Method

- Created a list of **seed** sentiment words:
  - positive seeds (Pwords): good, nice, excellent, positive, fortunate, correct, superior
  - negative seeds (Nwords): bad, nasty, poor, negative, unfortunate, wrong, inferior

# Turney and Littman (2003) Method

- Pointwise Mutual Information (PMI) based measure
- PMI between two words,  $w_1$  and  $w_2$  (Church and Hanks 1989):

$$\text{PMI}(w_1, w_2) = \log_2(p(w_1 \text{ and } w_2) / p(w_1)p(w_2))$$

$p(w_1 \text{ and } w_2)$  is probability of how often  $w_1$  and  $w_2$  co-occur

$p(w_1)$  is probability of occurrence of  $w_1$

$p(w_2)$  is probability of occurrence of  $w_2$



## Turney and Littman (2003) Method (continued)

- For every word  $w$  a sentiment association score is generated:

$$\text{score}(w) = \text{PMI}(w, \text{positive}) - \text{PMI}(w, \text{negative})$$

PMI = pointwise mutual information

$$\text{PMI}(w, \text{positive}) = \sum_{pword \in Pwords} \text{PMI}(w, Pword)$$

If  $\text{score}(w) > 0$ , then word  $w$  is positive

If  $\text{score}(w) < 0$ , then word  $w$  is negative

# Hashtagged Tweets

- Hashtagged words are good labels of sentiments and emotions

Can' t wait to have my own Google glasses #awesome  
Some jerk just stole my photo on #tumblr. #grr #anger

# Automatically Generated New Lexicons

- Polled the Twitter API for tweets with seed-word hashtags
  - A set of 775,000 tweets was compiled from April to December 2012
- Sentiment lexicons can be generated from sentiment-labeled data
  - Emoticons and hashtag words can be used as labels

# PMI-based Lexicons

- Hashtag Sentiment Lexicon
  - created from a large collection of hashtagged tweets
  - has entries for ~215,000 unigrams and bigrams
- Sentiment140 Lexicon
  - created from a large collection of tweets with emoticons
    - Sentiment140 corpus (Alec Go, Richa Bhayani, and Lei Huang, 2009)  
<http://help.sentiment140.com/for-students/>
  - has entries for ~330,000 unigrams and bigrams

# SemEval: International Workshop on Semantic Evaluation



# SemEval\*-2013 Task 2: Sentiment Analysis in Twitter

- Message-level task (44 teams)
  - tweets set: 1<sup>st</sup>
  - SMS set: 1<sup>st</sup>
- Performance
  - Tweets: Macro-averaged F: 69.02
  - Tweets: Macro-averaged F: 68.42

# Message-Level Sentiment: The Data (Semeval-2013 Task 2)

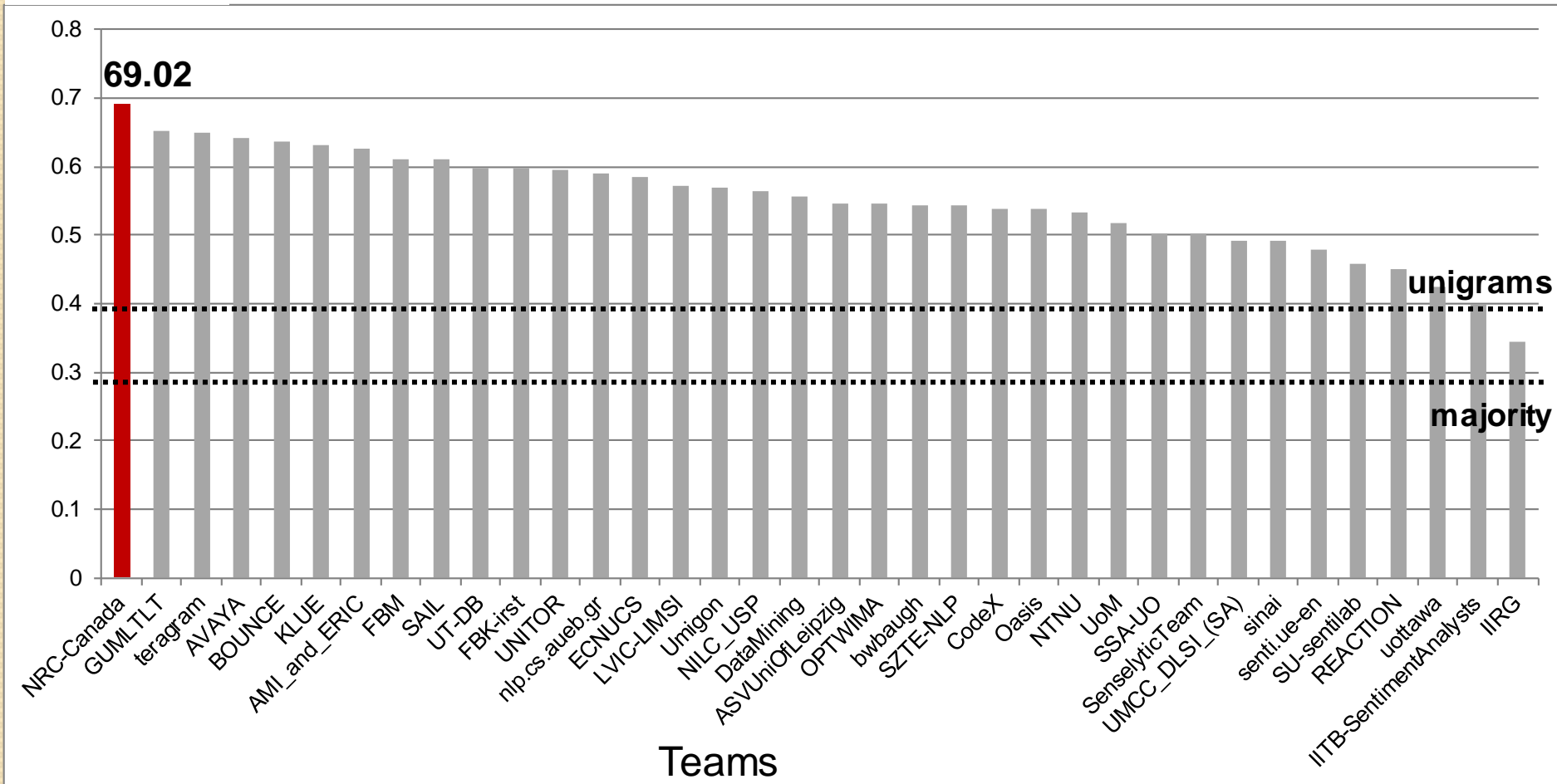
- Training: ~ 10,000 labeled tweets
  - positive: 40%
  - negative: 15%
  - neutral: 45%

## **Imbalanced categories!!**

- Test:
  - tweets: ~ 4,000
  - SMS: ~ 2,000

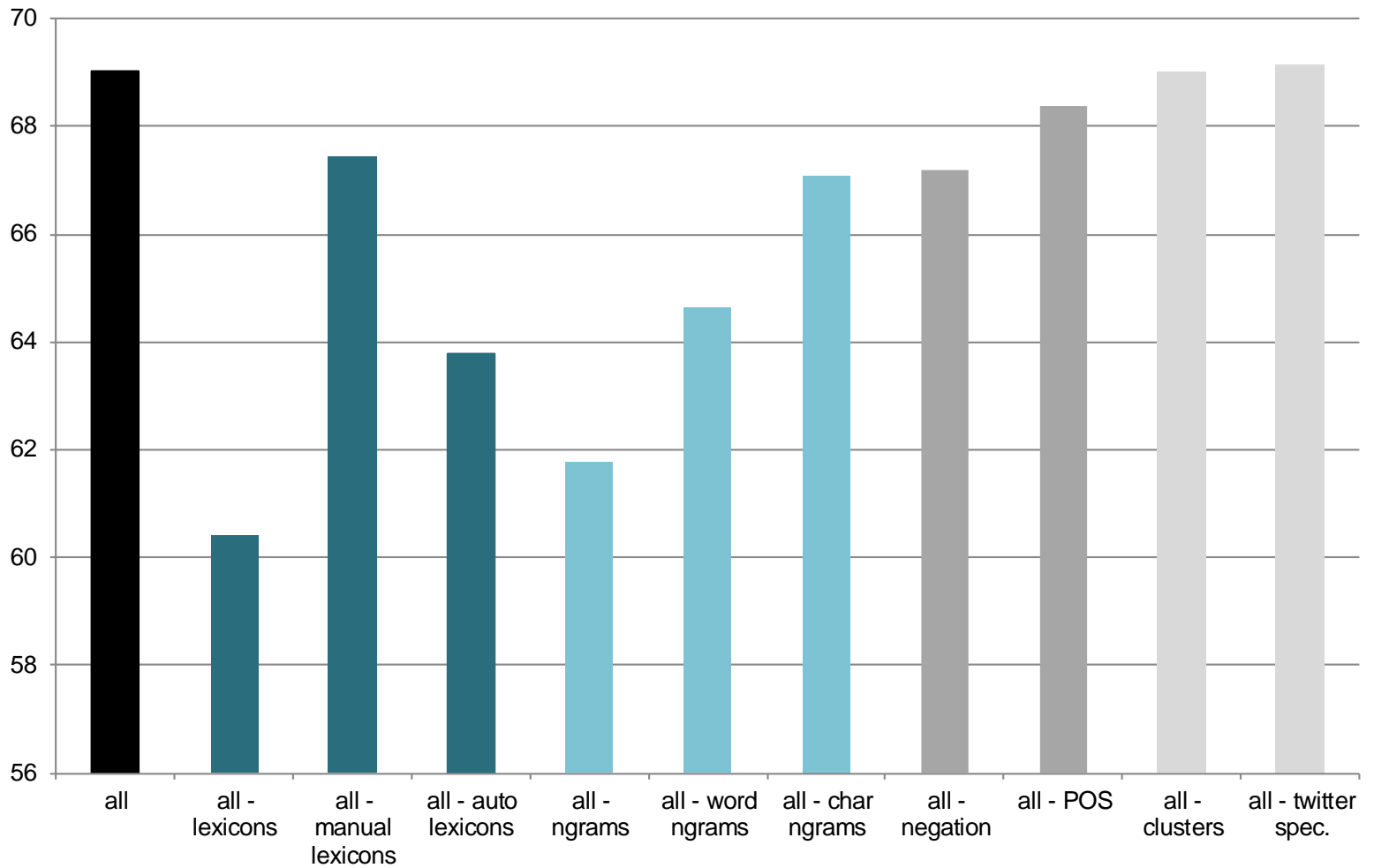
# Detailed Results on Tweets

F-score



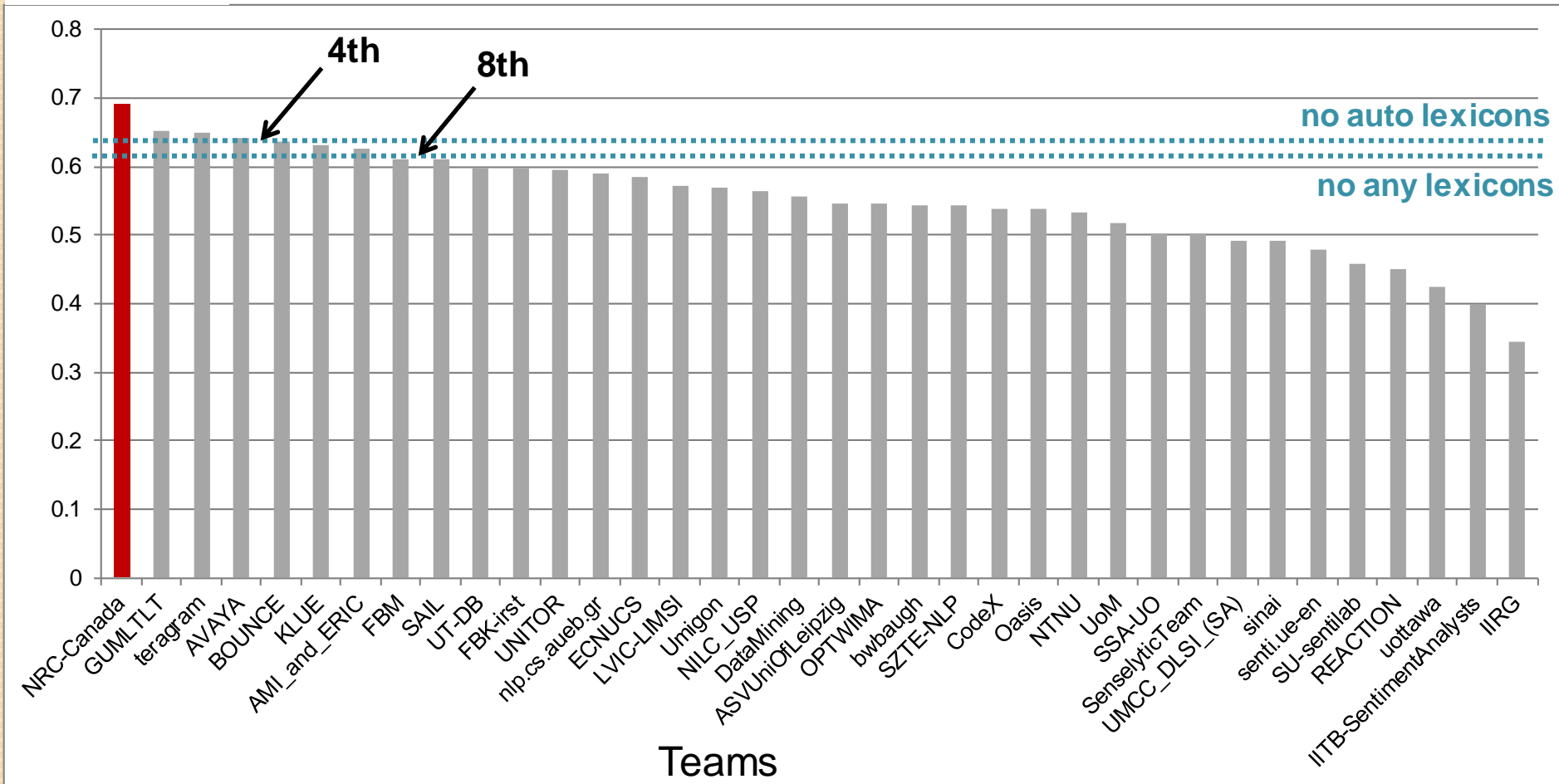


# Feature Contributions on Tweets



# Detailed Results on Tweets

F-score

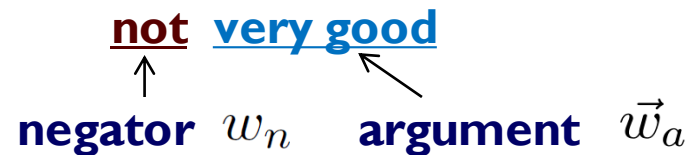




# Negation

# Negation

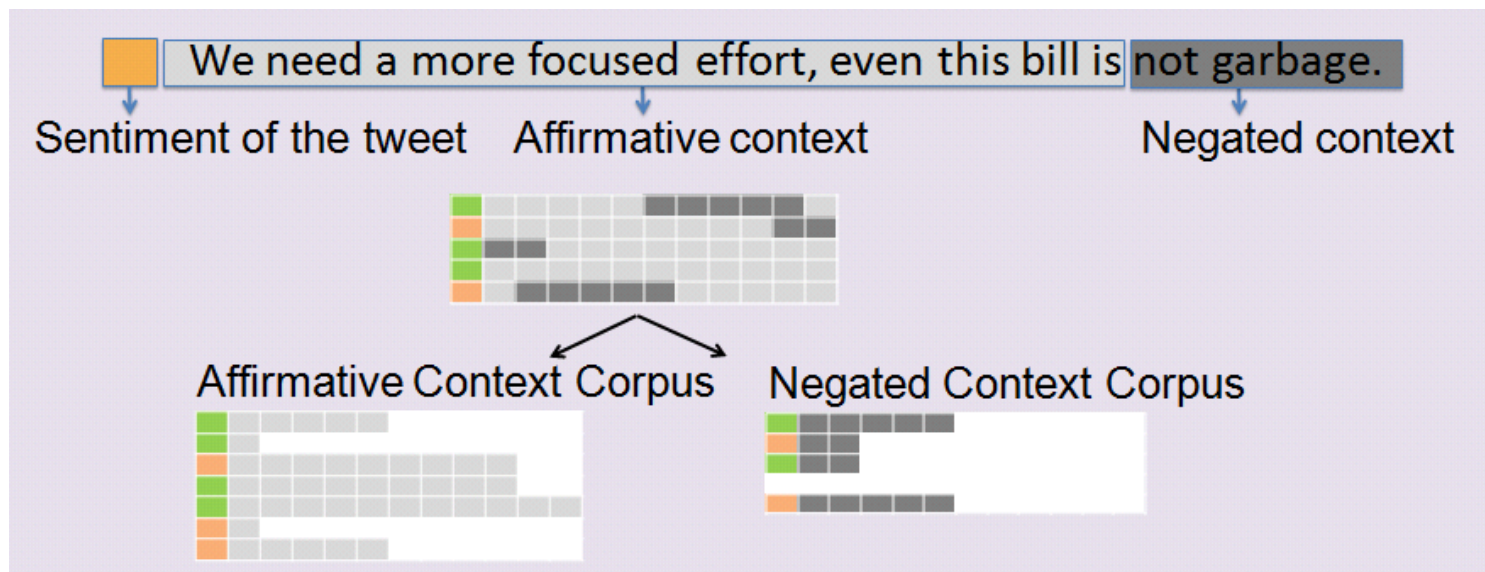
- Why negation? Negation often significantly affects the sentiment of its scopes.



- Negation has a complex effect on sentiment (Zhu et al. '14; Socher et al. '12)

# Improving the Systems for SemEval-2014 Task 9

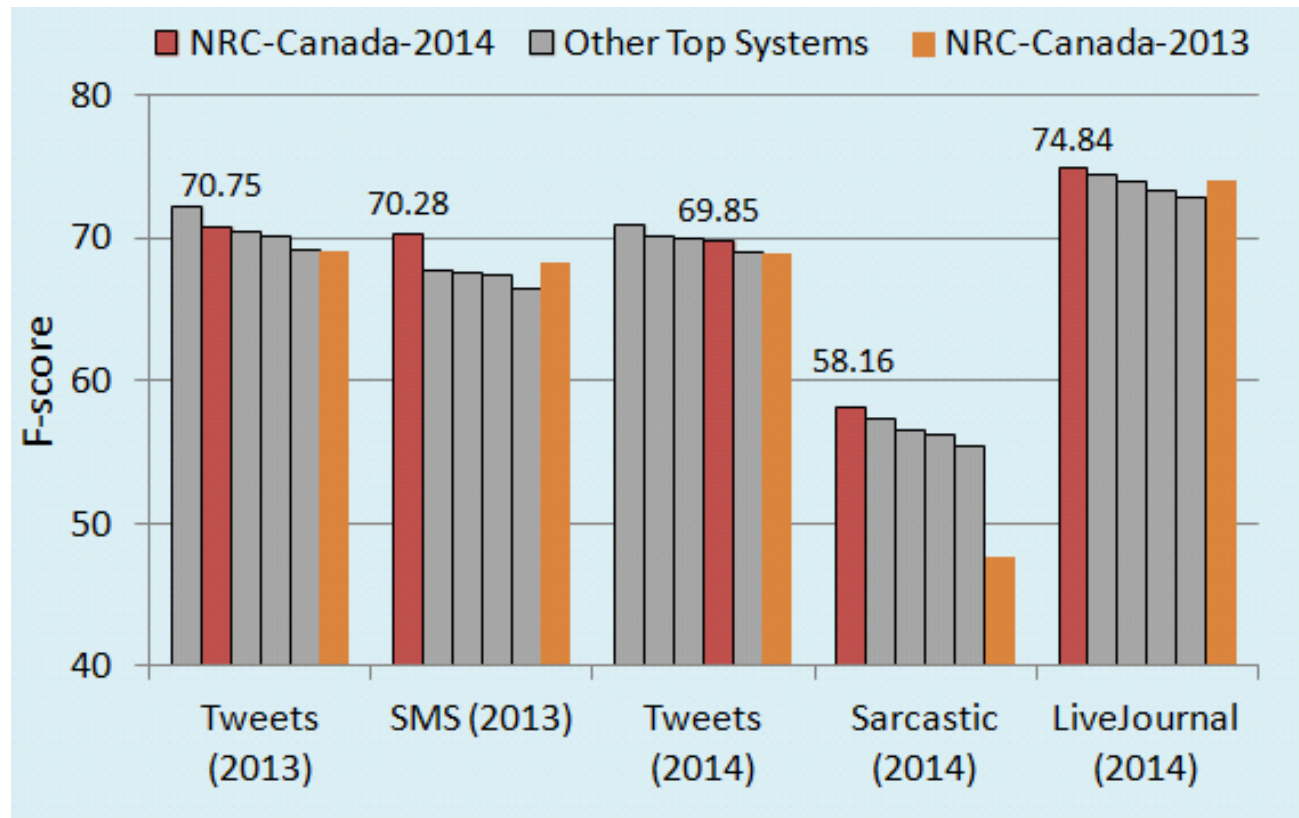
- In our SemEval-2014 system, we adopted a lexicon-based approach to determine the sentiment of words in affirmative and negated context.



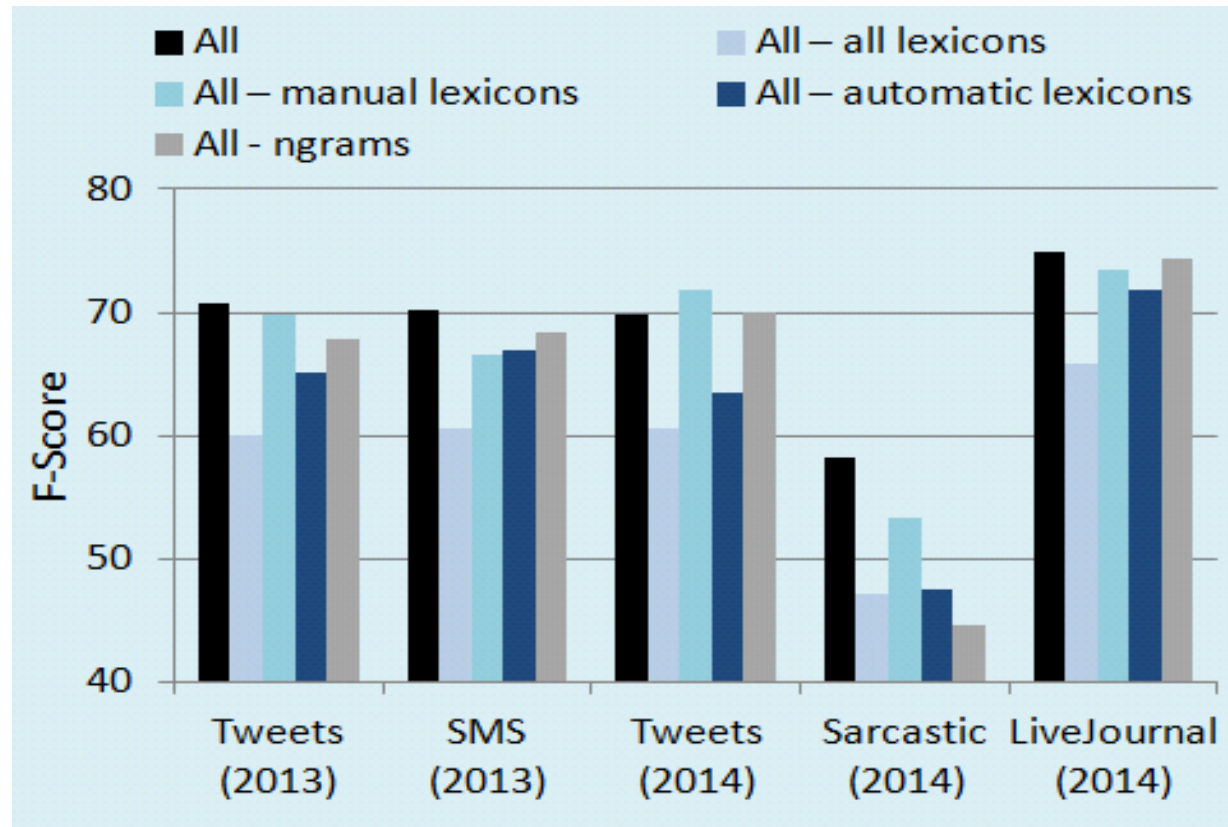
# Message-Level Sentiment : The Data (Semeval-2014 Task 9)

- Training (same as in SemEval-2013): ~ 10,000 labeled tweets
  - positive: 40%
  - negative: 15%
  - neutral: 45%
- Test
  - Official 2014 data:
    - tweets: ~ 2,000
    - sarcastic tweets: ~ 100
    - LiveJournal blogs (sentences): ~ 1,000
  - Progress (SemEval-2013 test data):
    - tweets: ~ 4,000
    - SMS: ~ 2,000

# Official Performance/Rankings



# Ablation Effects of Features





# Message-Level Sentiment: Summary

- No deep analysis; utilized big data and free (noisy) human annotation
- Automatically built lexicon and better negation handling improve the performance significantly.
- Best micro- and macro-averaged results on all 5 datasets
- System trained on tweets showed similar performance on SMS and LiveJournal blog sentences
- Strong performance on sarcastic tweets
- Most useful features on all datasets:
  - sentiment lexicons, especially automatic tweet-specific lexicons (**free available!**)

## Problems

- Message-level sentiment analysis
- **Phrase(term)-level sentiment analysis**
- Aspect-level sentiment analysis

# Term-Level Sentiment : The Problem

**Tweet:** plot of this movie is quite unpredictable, which is what I like. target is positive

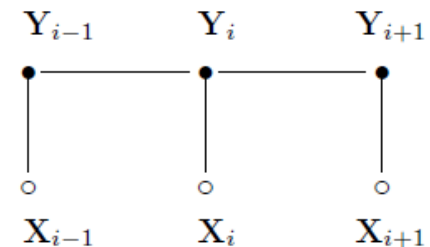
**Tweet:** the performance of our team is unpredictable, making me nervous. target is negative

# Further Clarification of the Problem

- The task is not defined as a sequence labeling problem:

Tweet:  $\frac{w_1}{obj} \frac{w_2}{pos} \frac{w_3}{neu} \frac{w_4}{obj} \frac{w_5}{neg} \frac{w_6}{neg} \frac{w_7}{neg} \frac{w_8}{neg} \frac{w_9}{neg} .$

- no boundary detection is required
- no need to label all expressions in a tweet.



- It is an independent classification problem for each sentiment term.

Tweet:  $w_1 \frac{w_2}{pos} \frac{w_3}{neu} \frac{w_4}{neu} \frac{w_5}{neg} \frac{w_6}{neg} \frac{w_7}{neg} \frac{w_8}{neg} \frac{w_9}{neg} .$

- Term-level sentiment (within tweets, blogs, SMS)
  - SemEval-2013 Task 2, SemEval-2014 Task 9

# Basic Feature Categories

Features	Description
term features	extracted from the target terms, including all the features discussed above.
context features	extracted from a window of words around a target term or the entire tweet, depending on features.

# Official Performance/Rankings

- Tweets
  - Macro-averaged F: 89.10
  - 1<sup>st</sup> place
- SMS
  - Macro-averaged F: 88.34
  - 2<sup>st</sup> place

# Term Features vs. Context Features

- Are contexts helpful? How much?

<b>Experiment</b>	<b>Tweets</b>	<b>SMS</b>
all features	89.10	88.34
all - target	72.97 ( <b>-16.13</b> )	68.96 ( <b>-19.38</b> )
all - context	85.02 (-4.08)	85.93 (-2.41)

- By large, sentiment of terms can be judged by the target terms themselves.
- The contextual features can additionally yield 2-4 points improvement on F-scores.



# Improving the Systems for SemEval-2014 Task 9

- Improving sentiment lexicons (as in message-level models)
  - Using a lexicon-based approach (Kiritchenko et al., '14) to determining the sentiment of words in affirmative and negated context.

- Discriminating negation words
  - Different negation words, e.g. never and didn't, can affect sentiment (Zhu et al., 2014) differently.

- We made a simple, lexicalized modification to our system

*This is never acceptable*

The word acceptable is marked as acceptable\_not in our old system but as acceptable\_beNever in our new system.

# Summary

- Better handling of negation words is helpful.
- Effect of lexicon features
  - Sentiment lexicons automatically built from tweets are particularly effective in our models.

## Problems

- Message-level sentiment analysis
- Phrase(term)-level sentiment analysis
- **Aspect-level sentiment analysis**

# Aspect-Level Sentiment

- Sub-Task 1: Aspect term extraction
  - Find terms in a given sentence that are related to aspects of the products.
- Sub-Task 2: Aspect term polarity
  - Determine whether the polarity of each aspect term is positive, negative, neutral or conflict.
- Sub-Task 3: Aspect category detection
  - Identify aspect categories discussed in a given sentence (e.g., food, service)
- Sub-Task 4: Aspect category polarity
  - Determine the polarity of each aspect category.

# Aspect-Level Sentiment

- Sub-Task 1: Aspect term extraction
  - Find terms in a given sentence that are related to aspects of the products.
- Sub-Task 2: **Aspect term polarity**
  - Determine whether the polarity of each aspect term is positive, negative, neutral or conflict.
- Sub-Task 3: Aspect category detection
  - Identify aspect categories discussed in a given sentence (e.g., food, service)
- Sub-Task 4: Aspect category polarity
  - Determine the polarity of each aspect category.

# Aspect Term Polarity: The Task

*The asian salad of Great Asian is barely eatable.*

Task: in the sentence above, what's the sentiment expressed towards the target term "*asian salad*"?

# Aspect Term Polarity: The Task

- This is different from the “term-level” sentiment analysis.  
The asian salad of Great Asian is barely eatable.  
aspect terms sentiment terms
- How the task is different from the previous two?

# Aspect Term Polarity: The Features

- Consider two examples:
  - Long-distance sentiment phrases  
*The ma-po tofu, though not as spicy as what we had last time, is actually great too.*
  - Local ambiguity  
*a serious sushi lover*



# Aspect Term Polarity: The Features

- Syntactic features
    - Consider long-distance sentiment phrases  
*The ma-po tofu, though not as spicy as what we had last time, is actually great too.*
    - Consider local syntax  
*a serious sushi lover*
- 
- Word- and POS-ngrams in the parse context
  - Context-target bigrams, i.e., bigrams composed of a word from the parse context and a word from the target term
  - All paths that start or end with the root of the target terms
  - Sentiment terms in parse context

# Aspect Term Polarity: The Features

- Surface features
  - Unigrams
  - Context-target bigrams (formed by a word from the surface context and a word from the target term itself)
- Lexicon features
  - Number of positive/negative tokens
  - Sum/maximum of the tokens' sentiment scores

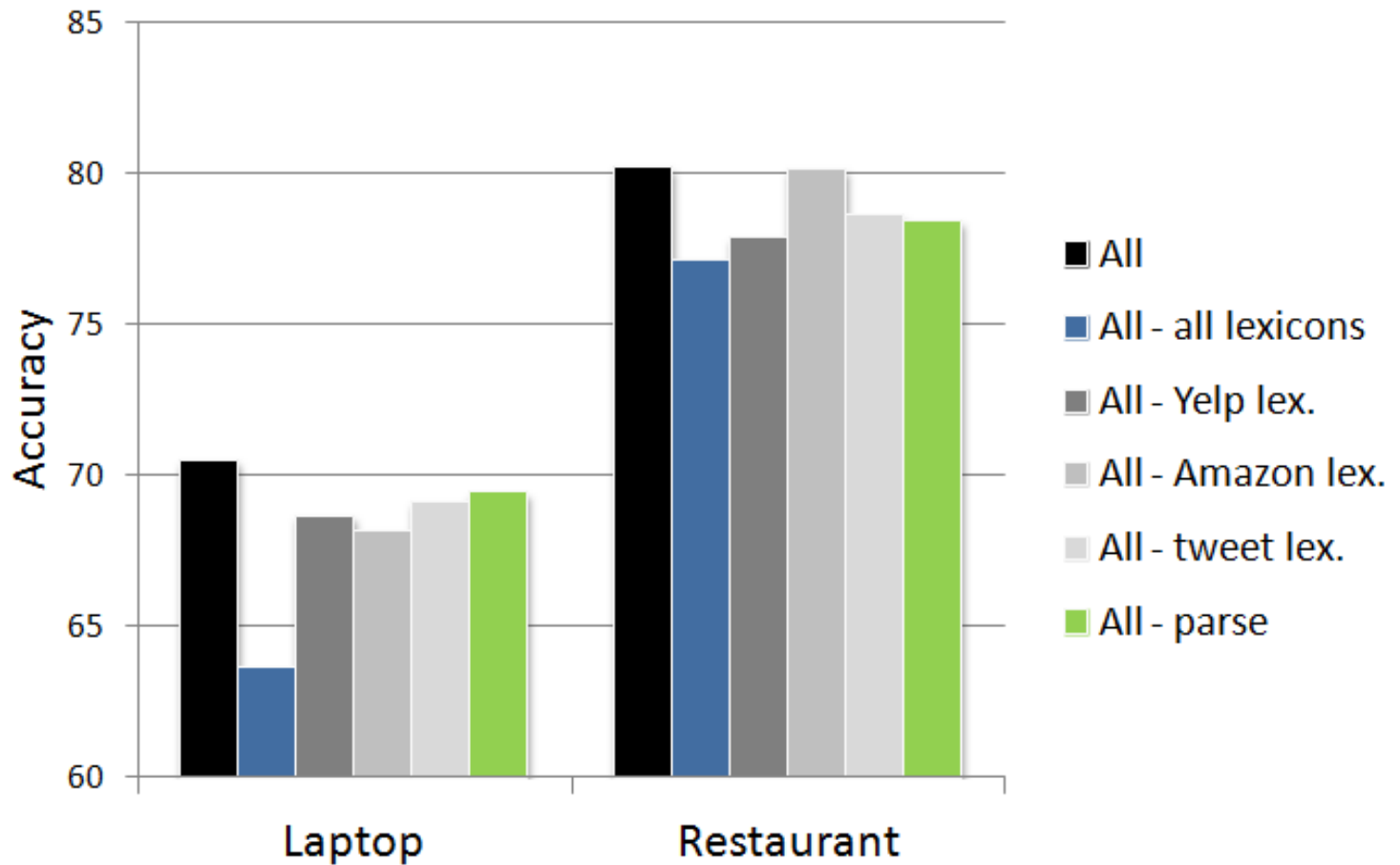
# Aspect Term Polarity: The Data

- Customer reviews
  - Laptop data
    - Training: 2358 terms
    - Test: 654 terms
  - Restaurant data
    - Training: 3693 target terms
    - Test: 1134 terms
- Pre-processing
  - We tokenized and parsed the provided data with Stanford CoreNLP Toolkits to obtain (collapsed) typed dependency parse trees (de Marneffe et al., 2006).

# Aspect Term Polarity: Results

- Laptop reviews
  - Accuracy: 70.49
  - 1<sup>st</sup> among 32 submissions from 29 teams
- Restaurant reviews
  - Accuracy: 80.16
  - 2<sup>nd</sup> among 36 submissions from 29 teams

# Aspect Term Polarity: Contributions of Features



# Sentiment Analysis of Social Media Texts

- Message-level sentiment analysis
- Phrase(term)-level sentiment analysis
- Aspect-level sentiment analysis

Use your NLP “tools” (skills) you have learned in this class to solve research or/and application problems.

# Sentiment Analysis of Social Media Texts

- Message-level sentiment analysis
- Phrase(term)-level sentiment analysis
- Aspect-level sentiment analysis

Use your NLP “tools” (skills) you have learned in this class to solve research or/and application problems.

NLP is not just a tool sets ...



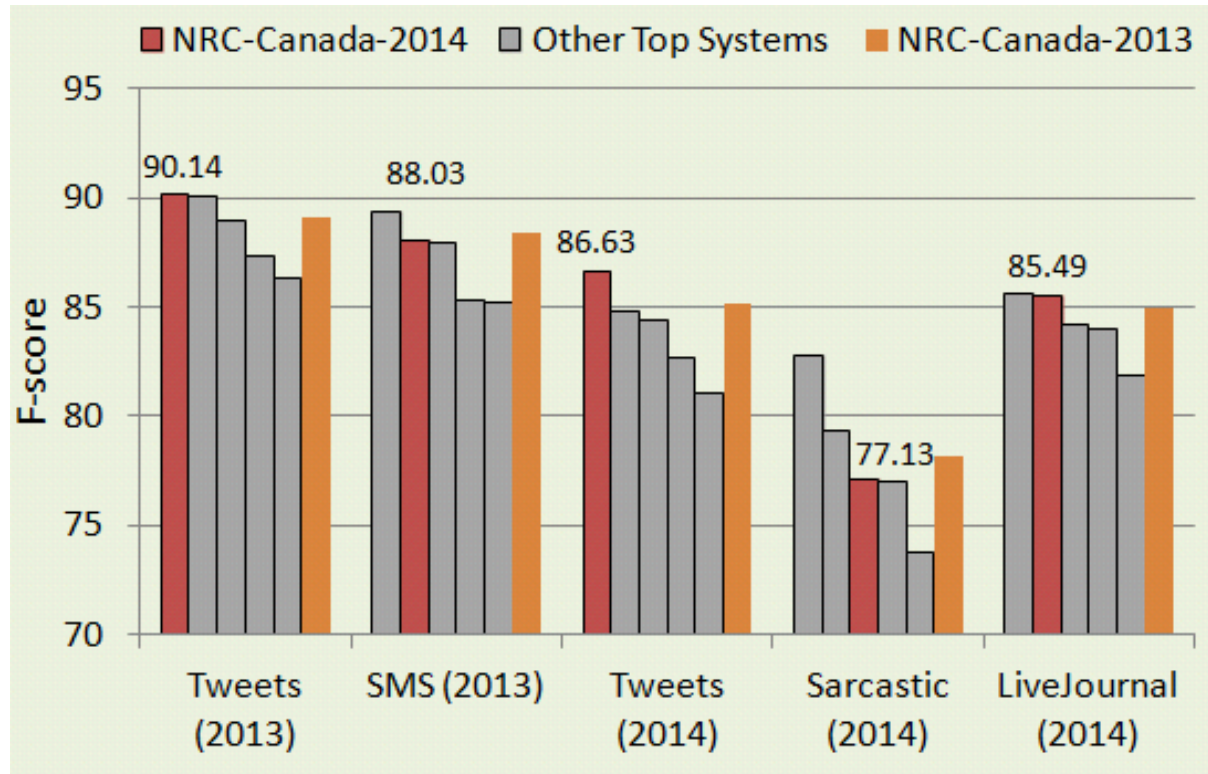
Questions?



# Term-Level Sentiment: The Data (Semeval-2014 Task 9)

- Training (same as in SemEval-2013): 8,891 terms
  - positive: 62%; negative: 35%; neutral: 3%
- Test
  - Official 2014 data:
    - tweets: 2,473 terms
    - sarcastic tweets: 124
    - LiveJournal blogs: 1,315
  - Progress (SemEval-2013 test data):
    - tweets: 4,435
    - SMS: 2,334

# Official Performance/Rankings



# Ablation Effects of Features

