Sentiment Analysis and Opinion Mining

Furu Wei

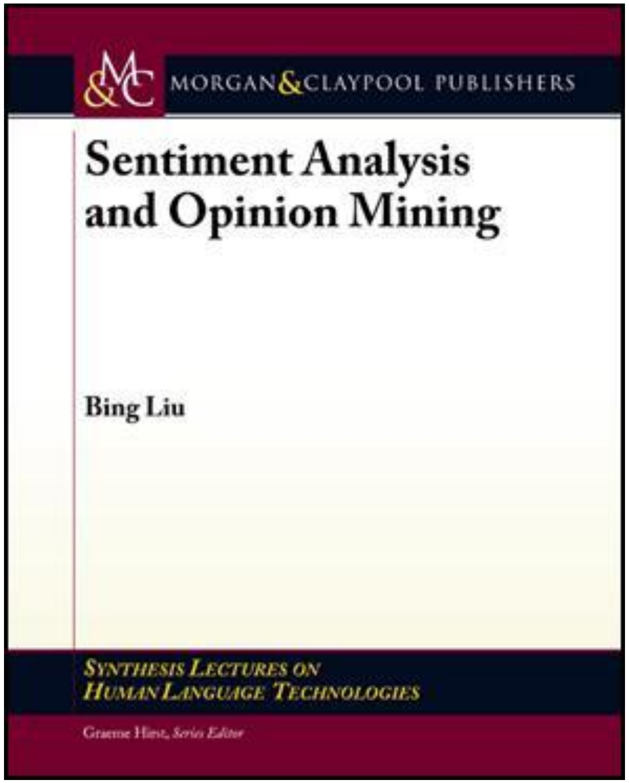
Natural Language Computing Group,

Microsoft Research Asia

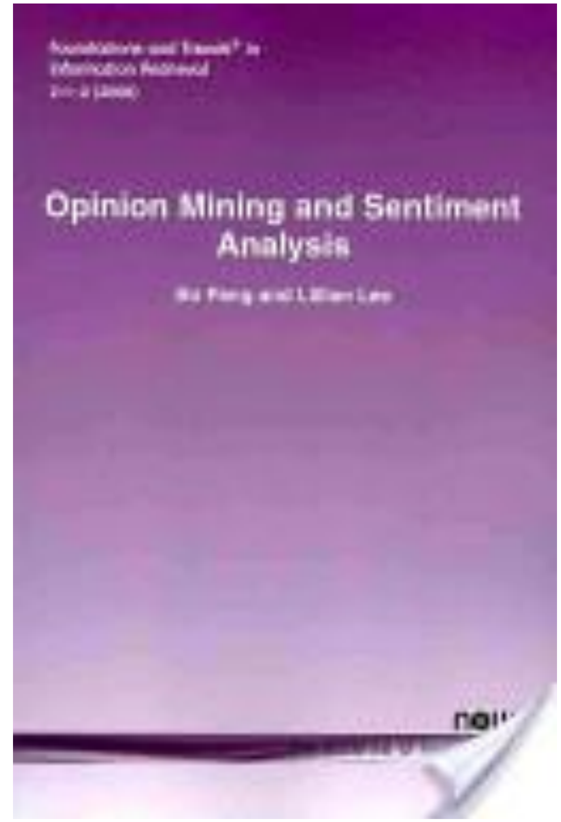
fuwei@microsoft.com

Sentiment Analysis Tutorials

- Sentiment Analysis in Practice Tutorial by Yongzheng Zhang et al. (ebay Research Lab | ICDM | 2011)
- Sentiment Symposium Tutorial by Christopher Potts (Stanford Linguistics | 2011)
- Sentiment Analysis Tutorial by Bing Liu (University of Illinois at Chicago | AAAI | 2011)
- Opinion Mining and Sentiment Analysis: NLP Meets Social Sciences by Bing Liu (University of Illinois at Chicago | 2010)
- Opinion Mining and Summarization by Bing Liu (University of Illinois at Chicago | WWW | 2008)
- Sentiment analysis and opinion mining (survey) by Bo Pang and Lillian Lee (Cornell University | 2008)



Bing Liu. *Sentiment Analysis and Opinion Mining* (Introduction and Survey), Morgan & Claypool Publishers, May 2012.



Bo Pang, Lillian Lee. *Opinion Mining and Sentiment Analysis*. Foundations and Trends in Information Retrieval 2(1-2): 1–135. 2008.

Outline

- Sentiment analysis
- Twitter sentiment analysis
- Multilingual sentiment analysis

Thumb up or down

Positive or negative

Love or hate

Sentiment

- Sentiment := <Holder, <u>Target</u>, **Polarity**, Auxiliary>
  - Holder: who expresses the sentiment
  - Target: what/whom the sentiment is expressed to
  - Polarity: the nature of the sentiment (positive, negative, or neutral)
  - Auxiliary: strength, summary, confidence, time

Sentiment

- Sentiment := <**Holder**, Target, Polarity, Auxiliary>
  - **Holder:** who expresses the sentiment
  - Target: what/whom the sentiment is expressed to
  - Polarity: the nature of the sentiment (e.g., positive/negative)
- *In his recent State of the Union address,* **US President Bush** *quite unexpectedly labeled Iran, Iraq, and the DPRK as an "axis of evil".*

Negative

Sentiment

- The games in iPhone 4s are pretty funny!

Feature/Aspect

Target

Positive

Polarity

Holder = the user/reviewer

Sentiment Analysis

- Computational study of opinions, sentiments, appraisal, and emotions expressed in text
  - Reviews, blogs, discussions, microblogs, social networks

- Also known as Opinion Mining

Google

hp printer

HP Officejet 6500A Plus e-All-in-One E710n Color Ink-jet - Fax / copier / printer / scanner - English, French, Spanish / Canada, United States

$152 online

★★★★☆ 1,251 reviews

September 2010 - Printer - Copier - Fax - Scanner - HP - Inkjet - Office - Wireless - Color - Duplex - 250 sheet - 64 MB memory

Want an all-in-one that's fast, economical, and eco-conscious? The Officejet 6500 All-in-One will save your color page as well as let you use less energy. You'll reduce intervention with the 35-page automatic document feeder. And you can network this versatile machine for group use via built-in Ethernet networking.

Reviews

Summary - Based on 1,251 reviews

1 | 2 | 3 | 4 stars | 5 stars

What people are saying

ease of use	"Easy to set up, makes great copies, works very well"
setup	"Setup, including wireless was quick."
value	"Great value and super easy set up."
design/style	"Nice sleek look, nice features."
customer service	"customer service was unable to help resolve the issue."
picture/video	"Good quality prints and photos."
size	"Pretty Paper weight."

Google Product Search (10/01/2012)

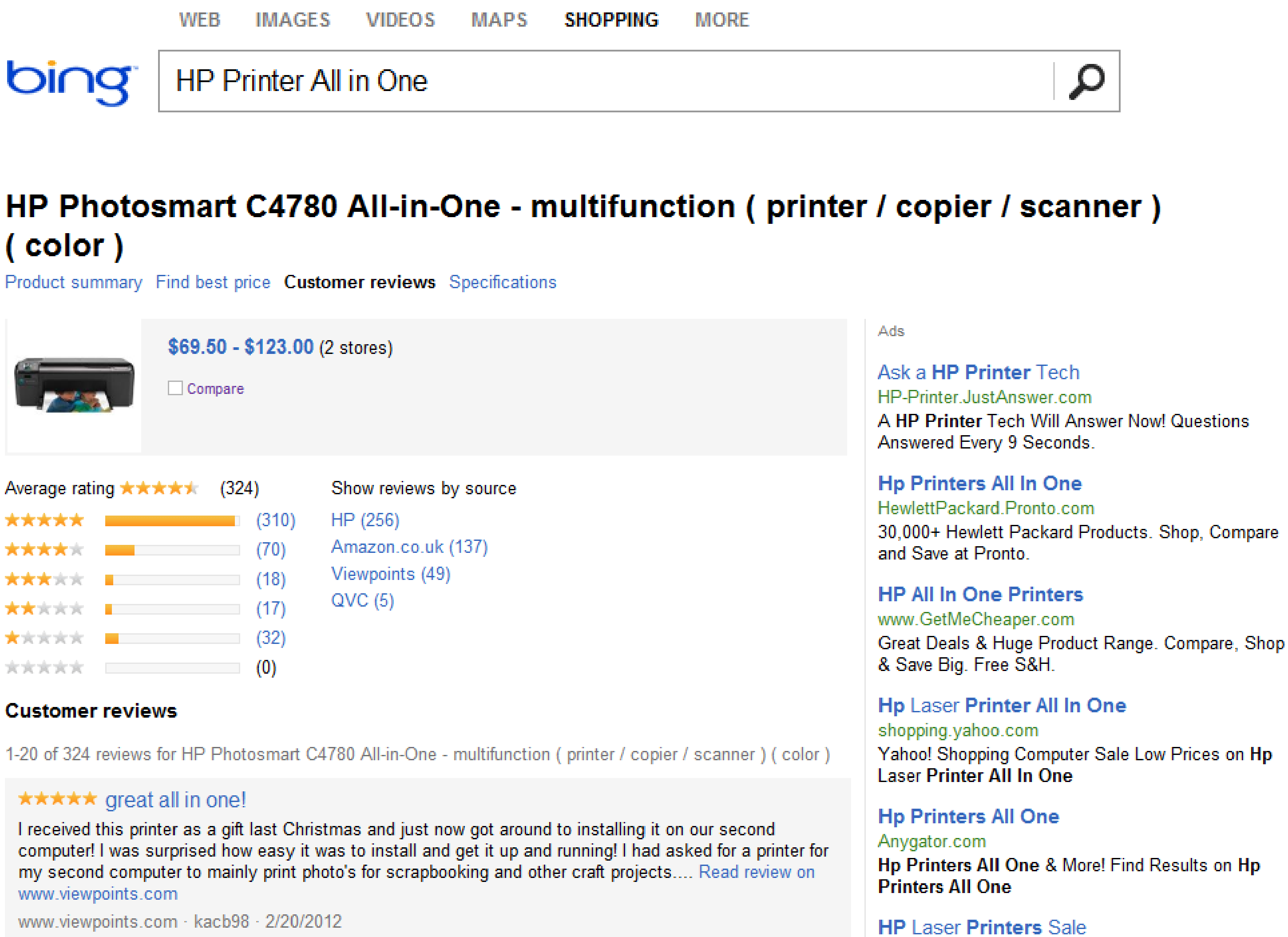
Great Printer!

★★★★★ By ttmae12 - Jun 7, 2011 - Staples

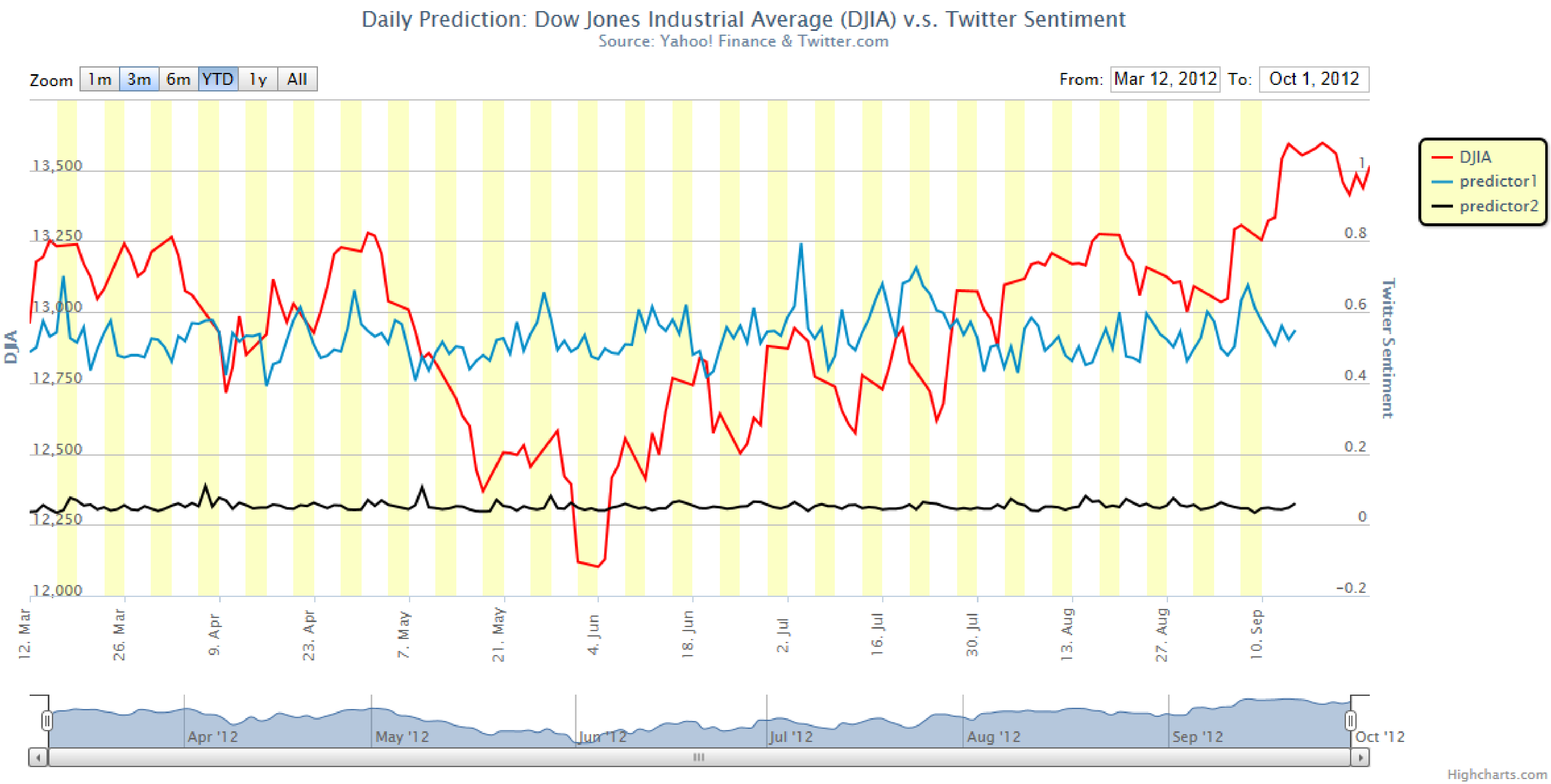
Pros: Great Print Quality; Fast Operation; Easy To Use; Quiet; Easy To Set Up; Compact Design; Reliable

This is an excellent printer for the price. I am amazed at how nice it prints. Machine was easy to setup. Printer is wired to my desktop and wireless to my iPad. Prints beautifully!
The only problem I encountered was printing photos. Once the correct paper quality is selected, there were no issues. I have always had good luck with HP products and the company has good customer support as well. In today's economy customer service and a great product means a lot.
Thanks

Full review provided by: STAPLES



Bing Shopping (10/01/2012)



Daily [Live Stock Market Prediction and Tracking] using Twitter Sentiment

Basic Tasks for Sentiment Analysis

- Holder detection
  - Find who express the sentiment
- Target recognition
  - Find whom/what the sentiment is expressed towards
- **Sentiment (Polarity) classification**
  - Positive, negative, neutral
- Opinion summarization
- Opinion spam detection

Sentiment Classification

Lexicon Based Sentiment Classification

- Basic idea
  - Use the dominant polarity of the opinion words in the sentence to determine its polarity
  - If positive/negative opinion prevails, the opinion sentence is regarded as positive/negative
- **Lexicon + Counting**
- **Lexicon + Grammar Rule + Inference Method**

Minqing Hu and Bing Liu. **Mining and summarizing customer reviews**. *KDD*: 168-177, 2004.
Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. **Lexicon-Based Methods for Sentiment Analysis**. *Computational Linguistics*: 37(2), 267-307. 2011.

The General Inquirer Lexicon

Philip J. Stone, Dexter C Dunphy, Marshall S. Smith, Daniel M. Ogilvie. 1966. The General Inquirer: A Computer Approach to Content Analysis. MIT Press.

- Homepage: http://www.wjh.harvard.edu/~inquirer
- Categories
  - Positive (1,915 words) and Negative (2,291 words)
  - Strong vs Weak, Active vs Passive, Overstated versus Understated
  - Pleasure, Pain, Virtue, Vice, Motivation, Cognitive Orientation, etc
- Free for research use

LIWC (Linguistic Inquiry and Word Count)

Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007). Linguistic Inquiry and Word Count: LIWC 2007.

- Homepage: http://www.liwc.net/
- 2,300 words, > 70 classes
  - Affective Processes
    - negative emotion (bad, weird, hate, problem, tough)
    - positive emotion (love, nice, sweet)
  - Cognitive Processes
    - Tentative (maybe, perhaps, guess), Inhibition (block, constraint)
  - Pronouns, Negation (no, never), Quantifiers (few, many)
- $30 or $90 fee

MPQA Subjectivity Cues Lexicon

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. HLT-EMNLP-2005.
Riloff and Wiebe (2003). Learning extraction patterns for subjective expressions. EMNLP-2003.

- Homepage: http://www.cs.pitt.edu/mpqa/subj_lexicon.html
- 6,885 words from 8,221 lemmas
  - 2,718 positive
  - 4,912 negative
- Each word annotated for intensity (strong, weak)
- GNU GPL

Opinion Lexicon

Minqing Hu and Bing Liu. Mining and Summarizing Customer Reviews. ACM SIGKDD-2004.

- Homepage: http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar
- 6,786 words
  - 2,006 positive
  - 4,783 negative

- Bing Liu's Page on Opinion Mining

SentiWordNet

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010 SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. LREC-2010.

- Homepage: http://sentiwordnet.isti.cnr.it/
- All WordNet synsets automatically annotated for degrees of positivity, negativity, and neutrality/objectiveness
  - [estimable(J,3)] “may be computed or estimated”
    - Pos 0 Neg 0 Obj 1
  - [estimable(J,1)] “deserving of respect or high regard”
    - Pos .75 Neg 0 Obj .25

Disagreements between polarity lexicons

Christopher Potts, Sentiment Tutorial, 2011

	Opinion Lexicon	General Inquirer	SentiWordNet	LIWC
MPQA	33/5402 **(0.6%)**	49/2867 **(2%)**	1127/4214 **(27%)**	12/363 **(3%)**
Opinion Lexicon		32/2411 **(1%)**	1004/3994 **(25%)**	9/403 **(2%)**
General Inquirer			520/2306 **(23%)**	1/204 **(0.5%)**
SentiWordNet				174/694 **(25%)**
LIWC				

Experiment

Table 10
Comparison of performance using different dictionaries with SO-CAL.

Dictionary	Percent correct by corpus				
	Epinions 1	**Epinions 2**	**Movie**	**Camera**	**Overall**
Google-Full	62.00	58.50	66.31	61.25	62.98
Google-Basic	53.25	53.50	67.42	51.40	59.25
Maryland-Full-NoW	58.00	63.75	67.42	59.46	62.65
Maryland-Basic	56.50	56.00	62.26	53.79	58.16
GI-Full	68.00	70.50	64.21	72.33	68.02
GI-Basic	62.50	59.00	65.68	63.87	64.23
SentiWordNet-Full	66.50	66.50	61.89	67.00	65.02
SentiWordNet-Basic	59.25	62.50	62.89	59.92	61.47
Subjectivity-Full	72.75	71.75	65.42	77.21	72.04
Subjectivity-Basic	64.75	63.50	68.63	64.83	66.51
SO-CAL-Full	80.25	80.00	76.37	80.16	**78.74**
SO-CAL-Basic	65.50	65.25	68.05	64.70	66.04

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. **Lexicon-Based Methods for Sentiment Analysis**. *Computational Linguistics*: 37(2), 267-307. 2011.

More on Lexicon Based Methods

- Negation
  - Not good vs. bad
- Intensification
  - Pretty good, very good, vs. good
- Composition

ADJ	NOUN	→	NP	Example
NEG	POS	→	NEG	disappointed hope
NEG	NEG	→	NEG	a horrible liar
POS	POS	→	POS	a good friend
POS	NEG	→	NEG	a perfect misery
POS	NEU	→	POS	a perfect meal
NEG	NEU	→	NEG	a horrible meal

Fig. 1: *NP composition*

Machine Learning Based Sentiment Classification

- Basic idea
  - Treat sentiment classification simply as a special case of topic-based categorization
    - With the two “topics” being positive sentiment and negative sentiment

- **Data + Feature + Model**

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. **Thumbs up? Sentiment Classification using Machine Learning Techniques**. *EMNLP*: 79-86, 2002.

Method

- Feature engineering
  - Each document d is represented by a feature vector $\tilde{}d := (n_1(d), n_2(d), \ldots, n_m(d))$
  - $n_i(d)$ could indicate **presence**, term frequency

- Classification models
  - Naive Bayes, Maximum Entropy, SVM

Data

- Movie reviews
  - Internet Movie Database (IMDb)
    - http://www.cs.cornell.edu/people/pabo/movie-review-data/
    - http://reviews.imdb.com/Reviews/
  - 700 positive / 700 negative
- Experiment setting for ML classifiers
  - 3-fold cross validation
  - Treating punctuation as separate lexical items
  - No stemming or stoplists were used

Experimental Results

- Baseline: use a few words written by human to classify

	Proposed word lists	Accuracy
Human 1	positive: *dazzling, brilliant, phenomenal, excellent, fantastic* negative: *suck, terrible, awful, unwatchable, hideous*	58%
Human 2	positive: *gripping, mesmerizing, riveting, spectacular, cool, awesome, thrilling, badass, excellent, moving, exciting* negative: *bad, cliched, sucks, boring, stupid, slow*	64%

- ML-based methods

	Features	# of features	frequency or presence?	NB	ME	SVM
(1)	unigrams	16165	freq.	**78.7**	N/A	72.8
(2)	unigrams	"	pres.	81.0	80.4	**82.9**
(3)	unigrams+bigrams	32330	pres.	80.6	80.8	**82.7**
(4)	bigrams	16165	pres.	77.3	**77.4**	77.1
(5)	unigrams+POS	16695	pres.	81.5	80.4	**81.9**
(6)	adjectives	2633	pres.	77.0	**77.7**	75.1
(7)	top 2633 unigrams	2633	pres.	80.3	81.0	**81.4**
(8)	unigrams+position	22430	pres.	81.0	80.1	**81.6**

More on Learning Based Methods

- Negation features
  - NOT_good, NOT_bad
- N-gram features
- POS tags features
- POS based pattern features

Which method is better

- No conclusion yet
  - Both are answers in different research papers
  - Depend on details, data, domains, scenarios, etc.
- Observations
  - Lexicon: simple, intuitive, interpretable
  - Learning: scale, data-driven, uninterruptable
- Maybe (?)
  - Sentence: lexicon is better
  - Document: learning is better

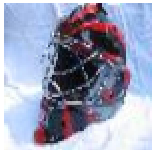
Do we really need sentiment analysis

The Task might be Over-simplified

- Document-level
- Review corpus
  - Movie reviews, product reviews, hotel reviews, etc.
- Holder := the author of the review
- Target := the movie/product being commented

- Two class, balanced classification in research papers

Why not use the intelligence from users



JetTech33
SW, ONT

Contributor

15 reviews

Reviews in 11 cities

68 helpful votes

"Wonderful lunch, incredible staff"

Reviewed September 23, 2012 NEW

My wife, 2yr old son and I had a fabulous lunch here on Fri Sept 21, 2012. The staff were incredibly kind, warm and accomodating to our 2yr old son. He was treated like a regular customer. Can't say enough about the cuisine staff and the waiting staff. True professionals.

My wife and I ordered the Chef's tasters' menu ($50EC) to share plus the chicken tikka masala and naan. The opening palate cleanser was a sweet mint/lemon "shooter" with an ice cube, much like a virgin mojito. The 5 dishes were incredibley well prepared and very flavourful. As a meal for one the Chef's menu is plenty of food. Also comes with a generous portion of basmati, bread and a dessert. All portions were hot, fresh and well seasoned. The dessert was such a surprise and left us wanting more. A vermicelli pasta with warm coconut milk, cardamon and a mellow sweetness to it that we both loved. Well done Chef. The staff also brought a bowl of ice cream for our son which was a very kind gesture.

Our meal came to $120EC with the tikka masala, naan and large bottle of water. We were very please with the value of this meal. Superb flavours, beautiful atmosphere, air conditioned and amazing staff.

Visited September 2012

Value	Service
Atmosphere	Food

Was this review helpful? Yes

Problem with this review?

See 2 more reviews by JetTech33 for Gros Islet

Ask JetTech33 about Spice Of India

Reviews from TripAdvisor (2012-09-30)

Question ?

- Why Taobao does not apply sentiment classification?

Sentiment Analysis on Social Media

- Core technology for social media analysis
  - A key function of social media is for people to express views & opinions
- We do not have ratings!
- The target is not explicit
- Sentence-level sentiment analysis

Twitter Sentiment Analysis

Twitter sentiment classification

- The objective
  - From a cluster of tweets, find positive and negative tweets on a given topic
  - Extended work: opinion summary
- Sentiment classification task
  - Holder: the author who publishes the tweet
  - Target: normally it is the given topic (query)
  - Polarity: to be decided
- Example
  - For a target: Windows7 (as a query)
  - Get a tweet "*Windows 7 is much better than Vista!*"
  - Output: positive

Challenges of tweet sentiment classification

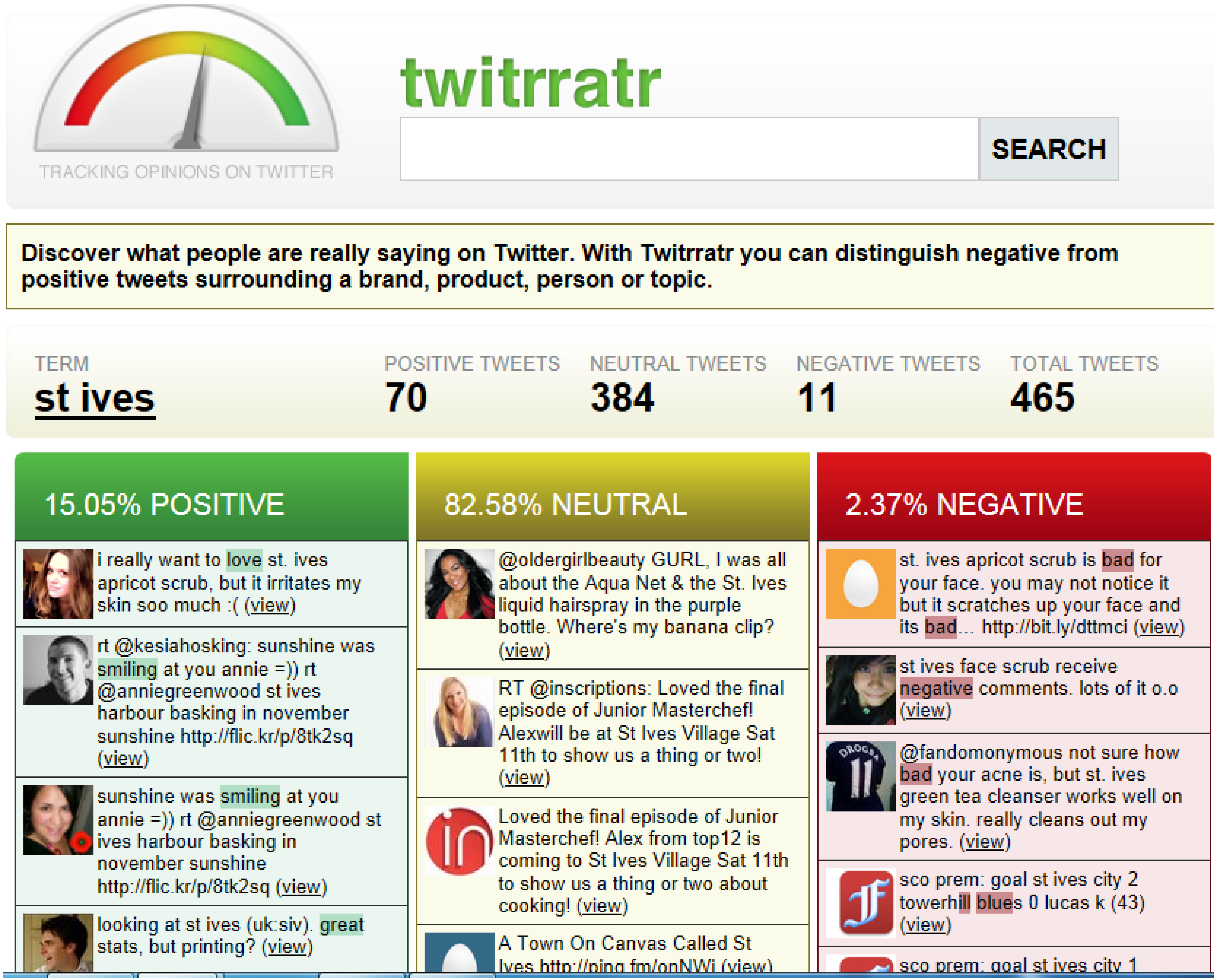
- Sentence level rather than document level
- Short and incomplete sentences
- Full of ambiguous words, abbreviation words
- Informal and unedited texts
  - *"another part of me by Micheal Jackson is soo nicee! Loooveeeeee itttttttttt!"*
- Not enough annotated data readily available

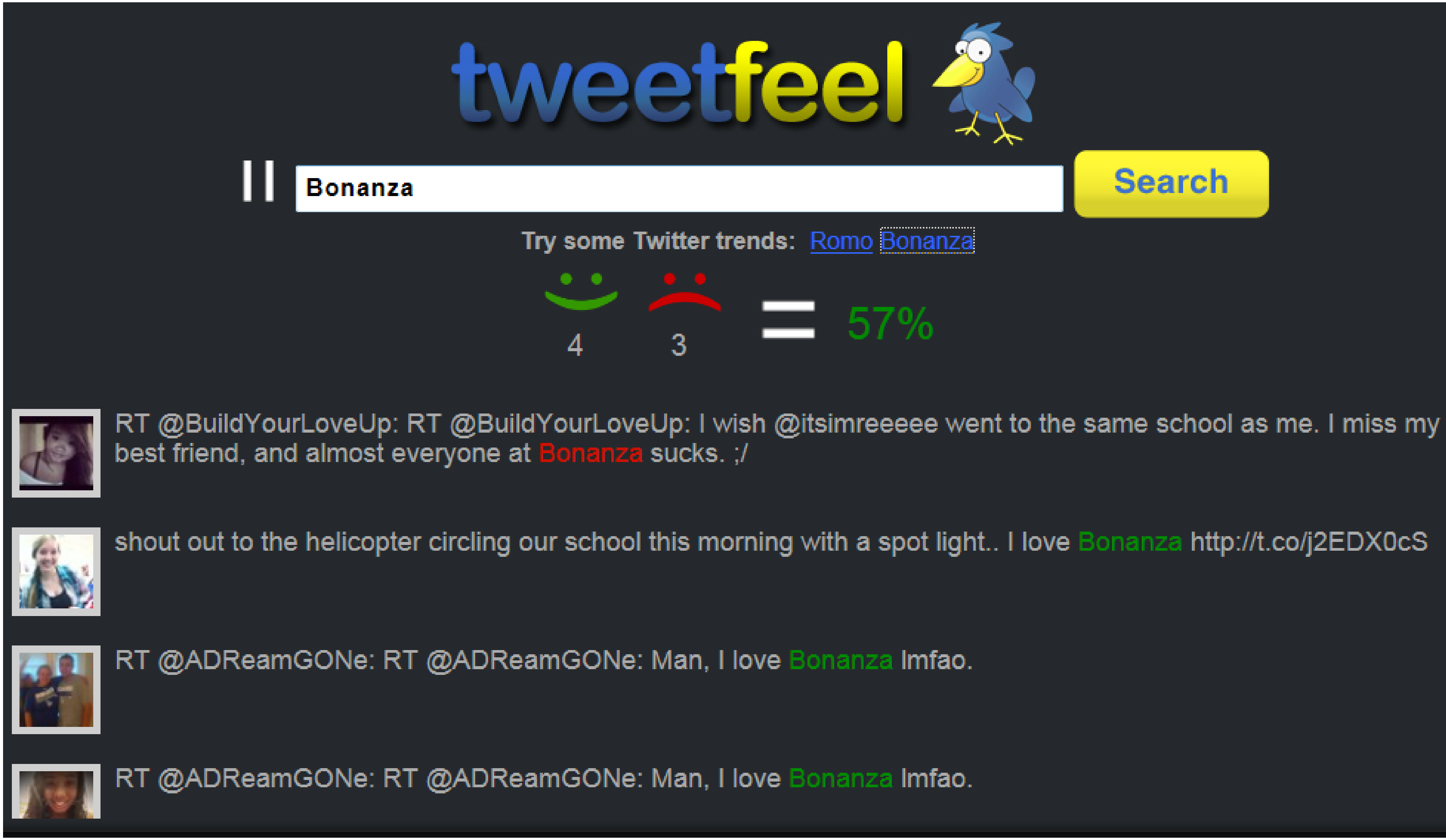
Progress of the existing research

- Two step approach
  - Barbosa and Feng, 2010: Two-step approach to classify the sentiments of tweets using SVM classifiers
- Using hashtags, smileys, emoticons to collect training data
  - Davidiv et al., 2010 : Classify tweets into multiple sentiment types using hashtags and smileys as labels
  - Go et al., 2009: SVM classifier + collect training data using emoticons

Existing Systems

- Lexicon-based methods
  - Based on the counting of # of positive words and # of negative words
  - Twittratr
- Rule-based methods
  - Based on syntactic rules, e.g., [query] is pos-adj
  - Tweetfeel
- Machine learning based methods
  - Based on the classifier built on a training data
  - Twitter sentiment


TRACKING OPINIONS ON TWITTER
twitrratr
SEARCH
Discover what people are really saying on Twitter. With Twitrratr you can distinguish negative from positive tweets surrounding a brand, product, person or topic.
TERM
st ives
POSITIVE TWEETS
70
NEUTRAL TWEETS
384
NEGATIVE TWEETS
11
TOTAL TWEETS
465
15.05% POSITIVE
i really want to love st. ives apricot scrub, but it irritates my skin soo much :((view)
rt @kesiahosking: sunshine was smiling at you annie =)) rt @anniegreenwood st ives harbour basking in november sunshine http://flic.kr/p/8tk2sq (view)
sunshine was smiling at you annie =)) rt @anniegreenwood st ives harbour basking in november sunshine http://flic.kr/p/8tk2sq (view)
looking at st ives (uk:siv). great stats, but printing? (view)
82.58% NEUTRAL
@oldergirlbeauty GURL, I was all about the Aqua Net & the St. Ives liquid hairspray in the purple bottle. Where's my banana clip? (view)
RT @inscriptions: Loved the final episode of Junior Masterchef! Alexwill be at St Ives Village Sat 11th to show us a thing or two! (view)
Loved the final episode of Junior Masterchef! Alex from top12 is coming to St Ives Village Sat 11th to show us a thing or two about cooking! (view)
A Town On Canvas Called St Ives http://ping.fm/onNWi (view)
2.37% NEGATIVE
st. ives apricot scrub is bad for your face. you may not notice it but it scratches up your face and its bad... http://bit.ly/dttmci (view)
st ives face scrub receive negative comments. lots of it o.o (view)
@fandomonymous not sure how bad your acne is, but st. ives green tea cleanser works well on my skin. really cleans out my pores. (view)
sco prem: goal st ives city 2 towerhill blues 0 lucas k (43) (view)
sco prem: goal st ives city 1


tweetfeel
Bonanza
Search
Try some Twitter trends: Romo Bonanza
4
3
= 57%
RT @BuildYourLoveUp: RT @BuildYourLoveUp: I wish @itsimreeeee went to the same school as me. I miss my best friend, and almost everyone at Bonanza sucks. ;/
shout out to the helicopter circling our school this morning with a spot light.. I love Bonanza http://t.co/j2EDX0cS
RT @ADReamGONe: RT @ADReamGONe: Man, I love Bonanza lmfao.
RT @ADReamGONe: RT @ADReamGONe: Man, I love Bonanza lmfao.

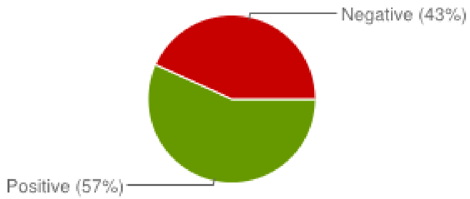
Sentiment140

Tweet 353 | Like 140 | +1 74

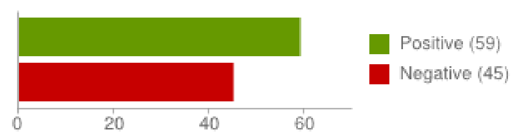
microsoft | English | Search | Save this search

Sentiment analysis for microsoft

Sentiment by Percent



Sentiment by Count



Tweets about: microsoft

Isaydumb: @Youporn, in my humble opinion you have nothing to do on the @Xbox Live. What the fuck is **@Microsoft** doing?!
Posted 46 seconds ago

Megan_Maracle: I hate this class. **#Microsoft** #die
Posted 2 minutes ago

dilwortha: @carasmith10 oh okay, you'll have to explain when i see you as i dont understand this disk haha. is it for **microsoft** project do you

Posted 5 minutes ago

jlebrech: @rsslldnphy it happens to be **microsoft** this time, but a superset is the next best thing from a compiled bytecode, as valid JS is also

http://www.sentiment140.com/

The results for this query are: Accurate

Issues (1)

- Most current research or systems do not consider the target when classifying the sentiment
- Example
  - Input a target “google”
  - *“Here's a great article about Monte Veronese cheese. It's in Italian so just put the url into Google translate and enjoy http://ow.ly/3oQ77”*
  - Sentiment classification: positive , however actually it is not towards the target that user inputs

Issues (2)

- Most current research and systems treat a tweet independently with its context
- However, there is severe shortage of the information in a single tweet as it is short and often uncompleted

Observation

- There is strong association between the polarities of connected tweets
  - An user tends to have same polarities towards some celebrities with a fixed length of time
  - The user who retweets a tweet normally has the same opinions with the original tweet

Target-dependent twitter sentiment analysis (ACL 2011)

Our approach on tweet sentiment classification (1)

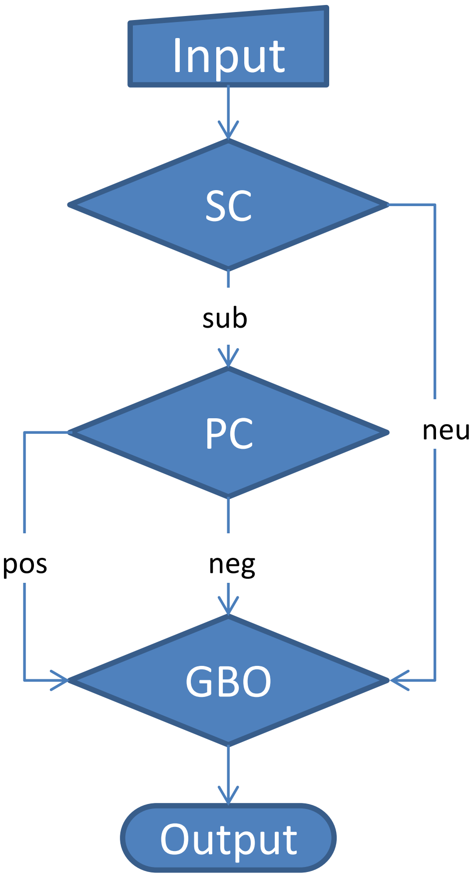
- Consider dependency relation between the target and the sentiment word
  - *E.g., Windows 7 is much better than Vista!*
    - *Output positive if target is Windows 7, and output negative if target is Vista*
  - *People everywhere love Windows & vista. Bill Gates*
    - *Output positive if target is Windows & vista, and output neutral if target is Bill Gates (If we don't consider the relation between Windows & Vista and Bill Gates)*

Our approach on tweet sentiment classification (2)

- Consider the context of the tweet
  - *E.g., "First game: Lakers!"*
  - *It is too short to decide the polarity, but when we consider the tweets in its context, we will make a better judgment*

Overview of our approach

- Task definition
  - Input
    - a collection of tweets containing the target (or query)
  - Output
    - labels assigned to each of the tweets
- Three steps
  - Subjectivity classification (SC)
  - Polarity classification (PC)
  - Graph-based optimization (GBO)



Preprocessing

- Tweet normalization
  - A simple rule-based model
  - “gooood” to “good”, “luve” to “love”
- POS tagging
  - OpenNLP POS tagger
- Word stemming
  - A word stem mapping table (about 20,000 entries)
- Syntactic parsing
  - A Maximum Spanning Tree dependency parser (McDonald et al., 2005)

Classification of subjectivity and polarity

- Binary SVM classifiers with linear kernel
  - Target-independent features
    - Content features
      - Words, punctuations, emoticons, and hashtags
    - Sentiment lexicon features
      - The number of positive or negative words in the tweet according to a sentiment lexicon (General Inquirer)
  - Target-dependent features(see next page)

Target-dependent features (1)

- Templates for generating target-dependent features
  - *Subject/object of* a transitive verb wi
    - wi _arg2, e.g., “I love iPhone”, => “love_arg2”
    - wi_arg1, e.g., “Obama reaffirms ..” => reaffirm_arg1
  - *Subject of* a intransitive verb
    - Wi_it_arg1
  - *Head of* an adjective or noun
    - Wi_arg1
  - Connected by a copula (verb “to be”) with an adjective or noun
    - Wi_cp_arg1
  - ….

Target-dependent features (2)

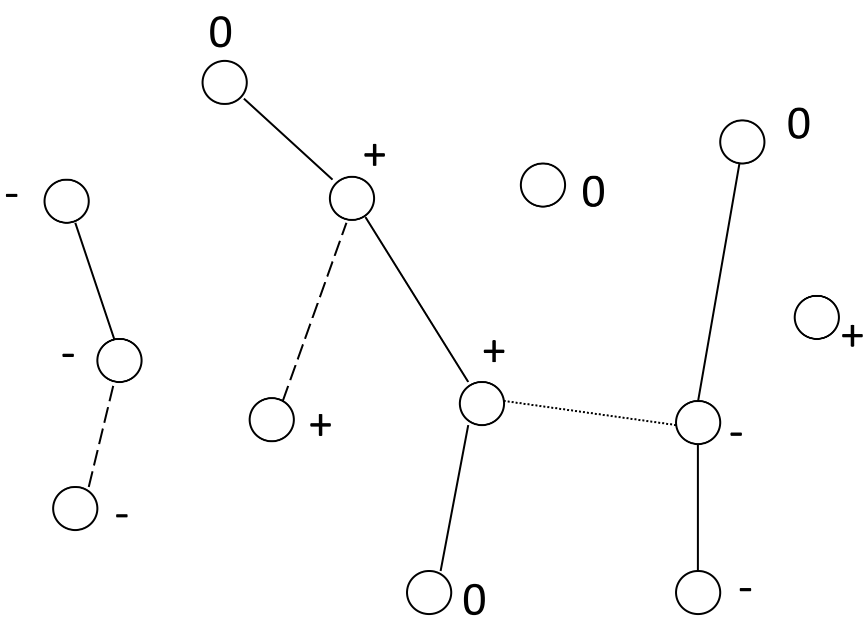
- Handle negations by adding "*neg-*"
  - "*iPhone does not work better with the CellBand*" => *neg-work_arg1*
- *Seven negations are used including not, n't, neither, seldom, hardly, etc.*

Target expansion

- Sometimes, sentiments are not expressed exactly towards the target
  - *"I am passionate about Microsoft technologies especially Silverlight."*
  - *Microsoft (input target) vs. Microsoft technologies(actual appearance)*
- Extended targets are viewed equally as target
  - All noun phrases including the target
  - Mentions co-referring to the target
  - Top K nouns with strong association with the target
- Note: we don't use ontology now

Graph-based sentiment optimization

- Relation types among the input tweets
  - Retweeting
  - Being published by the same person
  - Replying



Graph-based sentiment optimization

$$p(c \mid \tau) \longrightarrow \boxed{\text{Relaxation Labeling (Angelova and Weikum, 2006)}} \longrightarrow p(c \mid \tau, G)$$

The output score by SC and PC to calculate p(c| τ)

Iteratively optimize p(c| τ,G)

$$p(c \mid \tau, G) = p(c \mid \tau) \sum_{N(d)} p(c \mid N(d)) \, p(N(d))$$

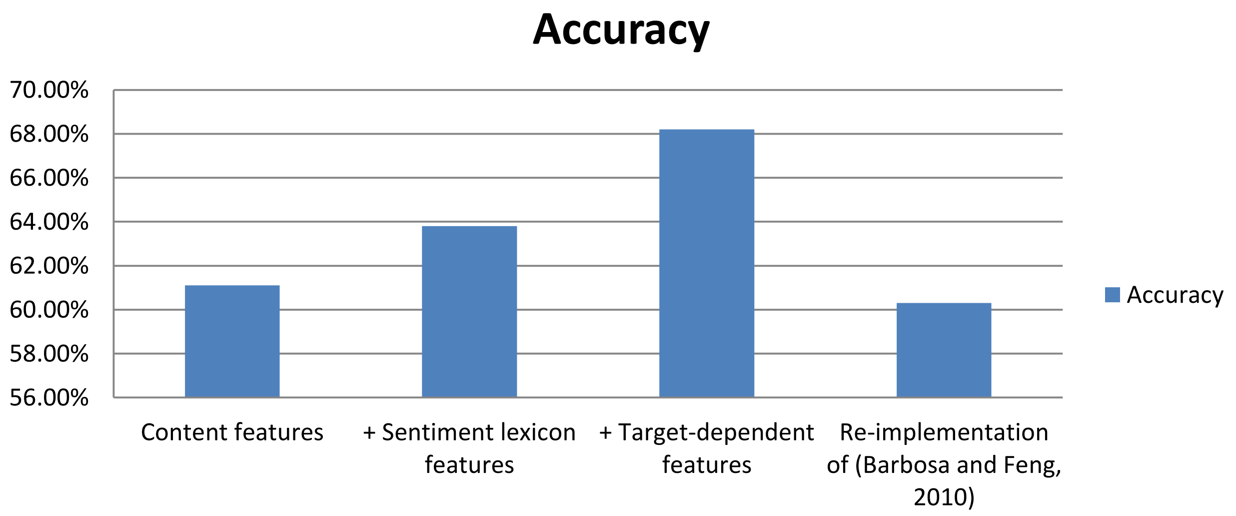
- *c* is the sentiment label of a tweet, which belongs to {positive, negative, neutral}
- *G* is the tweet graph
- *N(d)* is a specific assignment of sentiment labels to all immediate neighbors of the tweet
- τ is the content of the tweet

Experimental setting

- Raw data
  - 5 queries: *Obama, Google, iPad, Lakers, Lady Gaga*
  - 400 English tweets downloaded for each from Twitter
- Annotation
  - 2 human annotators
  - 3 labels: positive, negative or neutral
  - 459 positive, 268 negative and 1,212 neutral tweets
- Inter-annotator study
  - For 86% of tweets, two annotators give identical labels
  - For 13%, neutral-subjective disagreement
  - For 1%, positive-negative disagreement

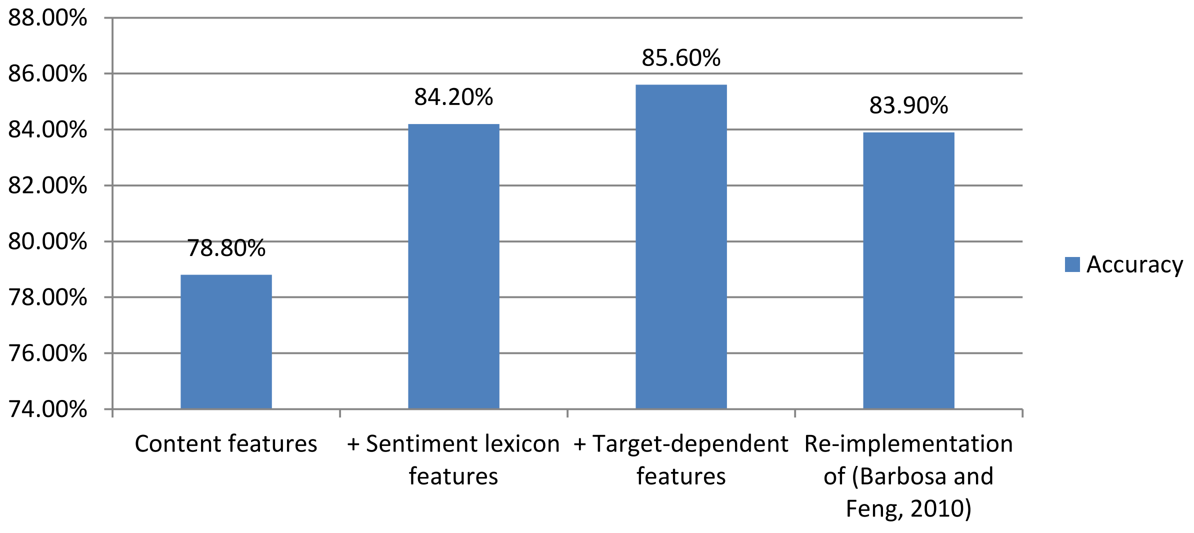
Subjectivity classification evaluation

- Data
  - 727 subjective (positive + negative) tweets and 1212 neutral tweets
  - 5 fold cross validation



Polarity classification evaluation

- Data
  - 268 negative and 459 positive tweets
  - 5 fold cross validation



Evaluation of graph-based optimization

- Data
  - 459 positive, 268 negative and 1,212 neutral tweets

System	Accuracy(%)	F1-score (%)		
		pos	neu	neg
Target-dependent sentiment classifier	66.0	57.5	70.1	66.1
+Graph-based optimization	68.3	63.5	71.0	68.5

Summary of our approaches

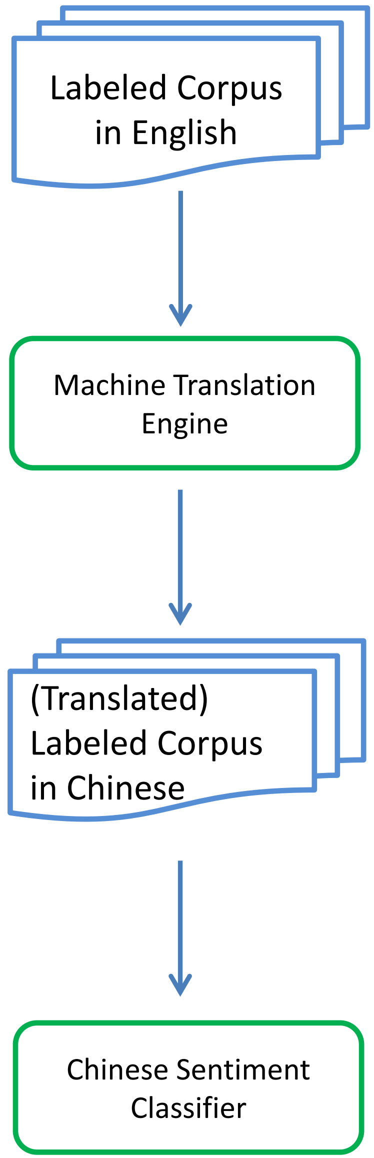
- Target-dependent features are used using dependency relation
- Targets are extended by various tricks to cover more appearances of targets
- A simple graph-model is used to take the context into consideration by relaxation labeling process

Multilingual Sentiment Analysis

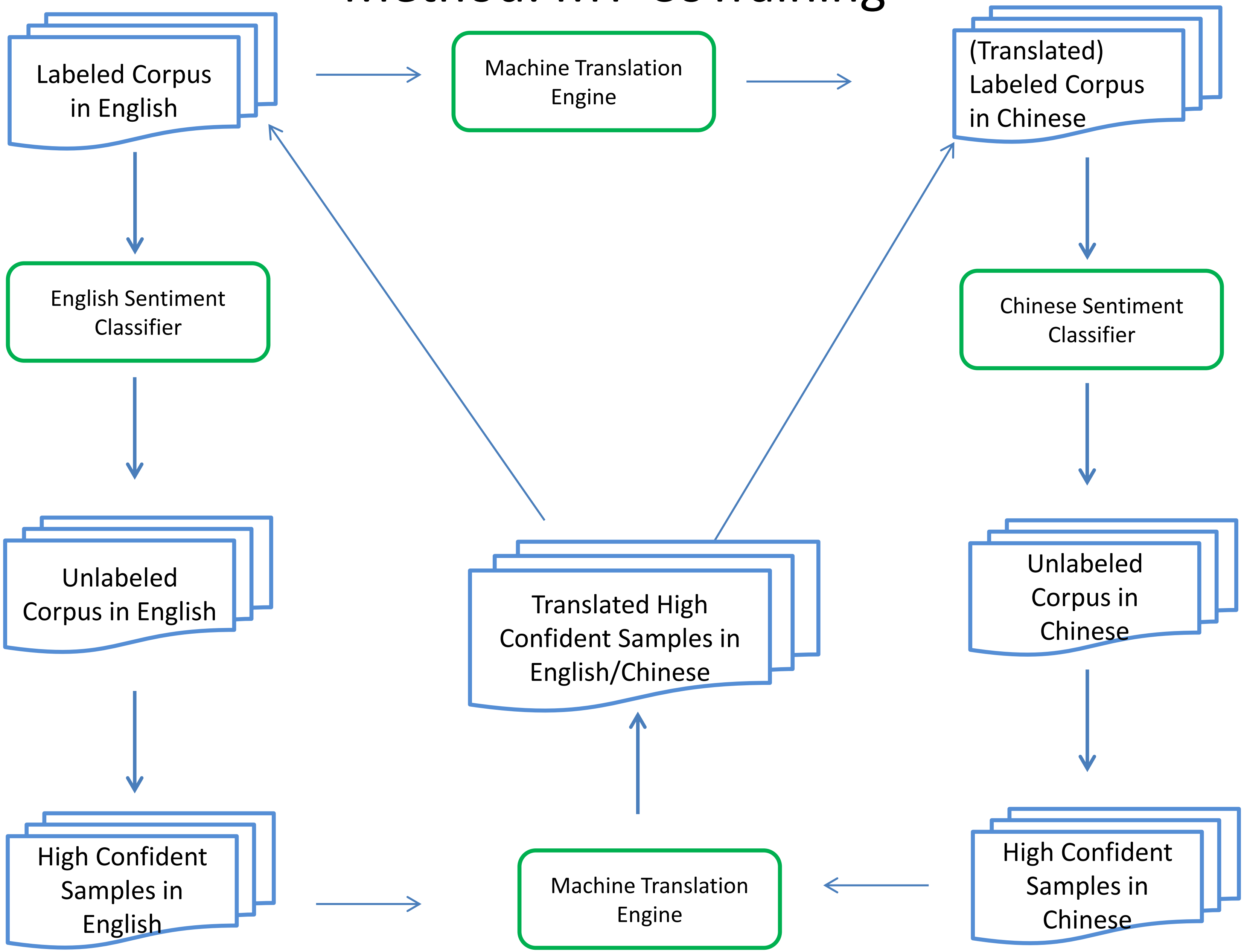
Task

- Train a sentiment classifier for a foreign language with labeled data on English and unlabeled parallel data on both languages
- Input
  - Labeled English data L_e
  - Unlabeled parallel data U_{ef}
  - Labeled foreign language data L_f (optional)
- Output
  - Sentiment classifier on the foreign language C_f

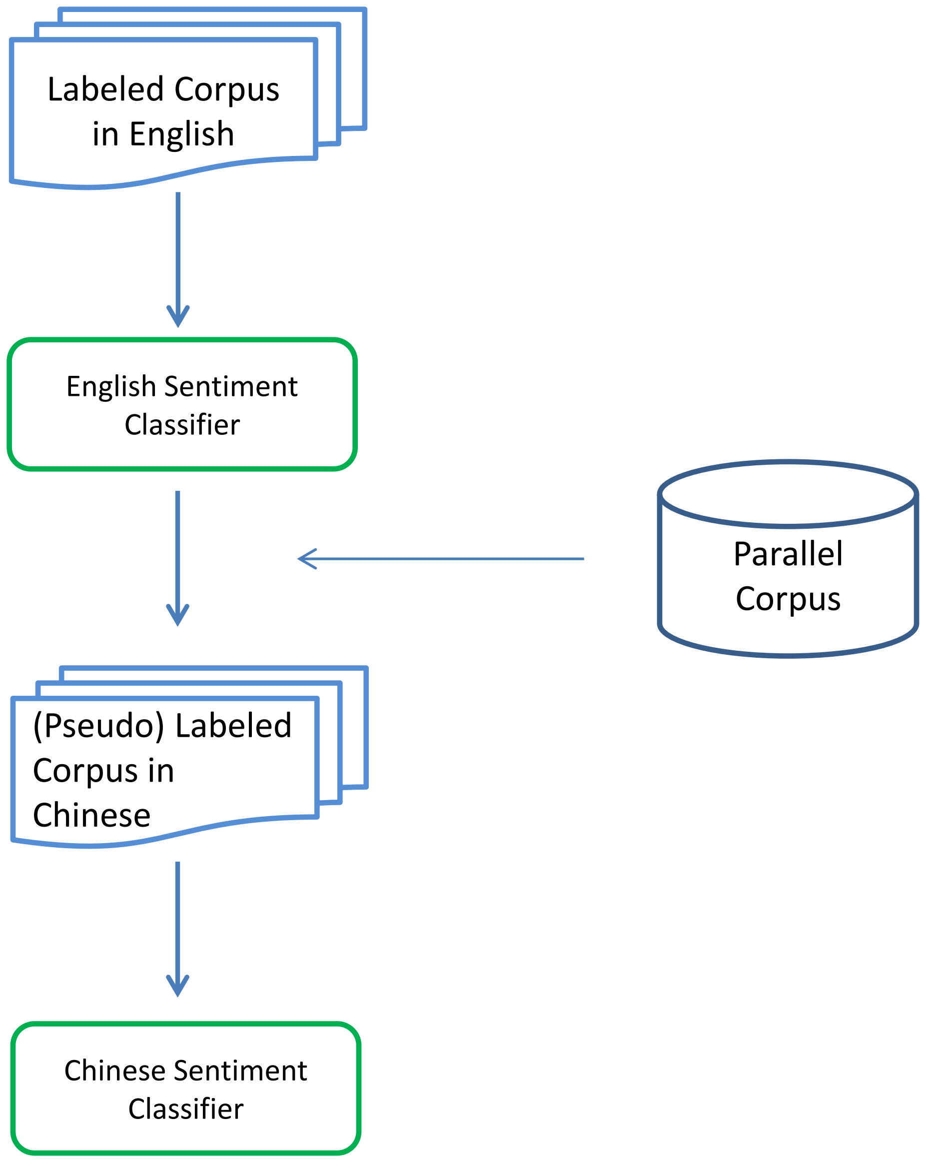
Method: MT-SVM



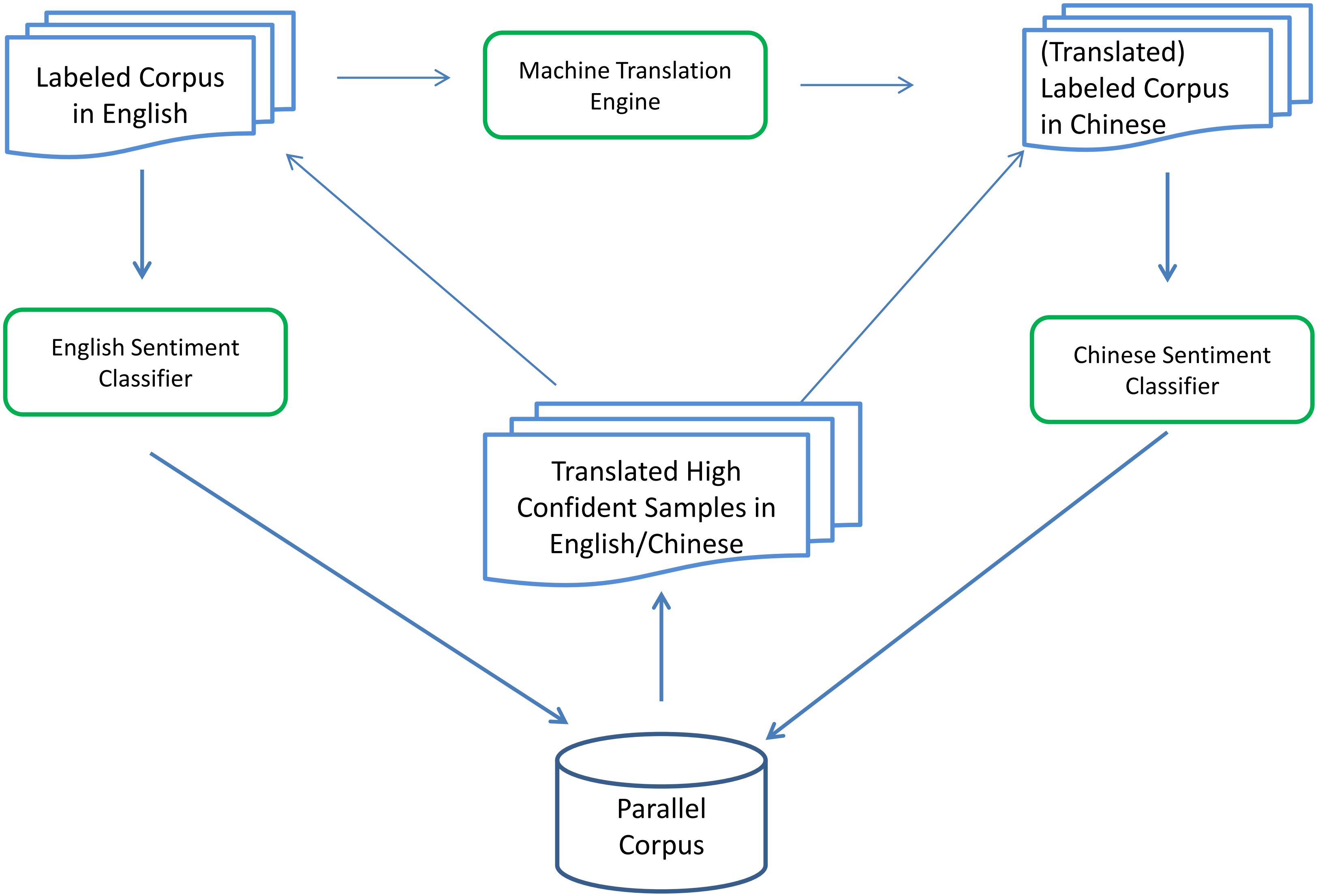
Method: MT-CoTraining



Method: Para-SVM



Method: Para-CoTraining



Cross-Lingual Mixture Model for Sentiment Classification (ACL 2012)

Existing Work

- (Wan 2009) uses machine translated text as training data
- (Prettenhofer 2010) projects both languages into the same space with bilingual words mapping and learn classifiers on this space
- (Lu 2011) improves bilingual sentiment classification by enforcing label consistency on parallel corpus

Challenges

- Problems and challenges
  - Machine translation engine
    - Feature/vocabulary coverage
    - Polysemy
  - Language projection with bilingual words mapping
    - Polysemy
  - Label consistency on parallel corpus
    - Label consistency is only determined by sentence alignment probability

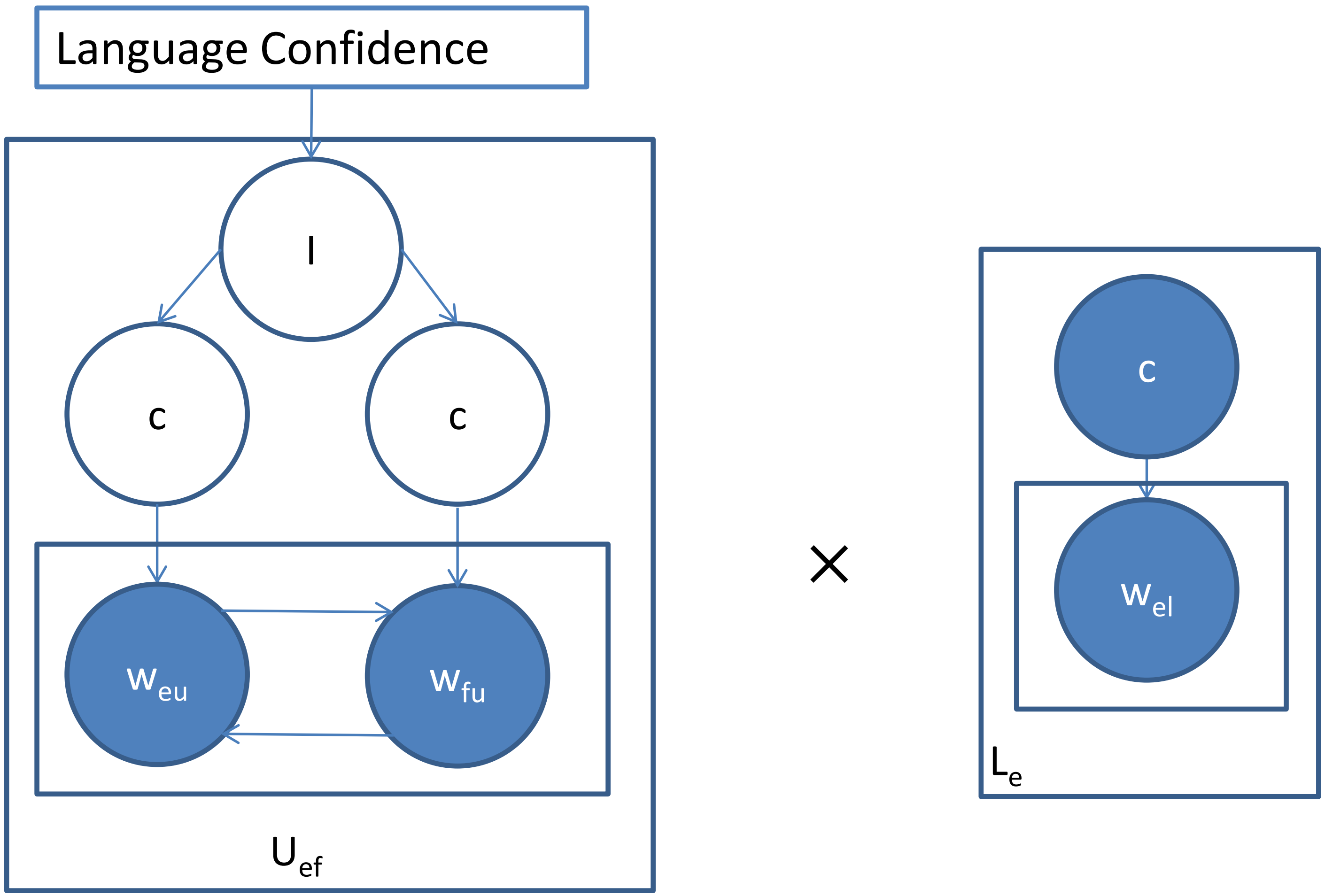
Answers to Challenges

- Improve feature/vocabulary coverage with parallel corpus
- Use word alignment in parallel corpus to choose mapping for polysemy
- Determine label consistency in parallel corpus with word alignment

Cross-Language Mixture Model

- Word level assumption in parallel sentences
  - The aligned words between English and Chinese have the same functions for determining the sentiment polarity for sentences
- Generative model
  - The process of sentence generation:
    - Select a polarity label wrt. prior distribution
    - Select words
      - Generate a Chinese word according to the polarity label
      - Generate a Chinese word by projecting an English word with the same polarity
  - Train the parameters by maximizing the likelihood of the large unlabeled parallel corpus and the labeled monolingual data (English with/without Chinese)

Cross-Language Mixture Model



$P(U_{ef}|\,\theta_1, \theta_2)$ $\times$ $P(L_{ef}|\,Ce, \theta_1, \theta_2)$

Unlabeled Parallel data

Labeled English data

Parameter Estimation

$$argmax_{\theta} \quad L(\theta|D_t, D_s, U) = L(\theta|D_s) + L(\theta|D_t) + L(\theta|U)$$

$$L(\theta|D_s) = \sum_{i=1}^{|D_s|} \sum_{j=1}^{|C|} \sum_{s=1}^{|V_s|} N_{si} \log P(w_s|c_j)\delta_{ij}$$

$$L(\theta|D_t) = \sum_{i=1}^{|D_t|} \sum_{j=1}^{|C|} \sum_{t=1}^{|V_t|} N_{ti} \log P(w_t|c_j)\delta_{ij}$$

$$L(\theta|U) = \sum_{i=1}^{|U_t|} \sum_{j=1}^{|C|} \sum_{t=1}^{|V_t|} [N_{ti} \log (\underline{P(w_t|c_j)} + \underline{P(w_t|w_s)P(w_s|c_j)})$$

$$+ \sum_{i=1}^{|U_s|} \sum_{j=1}^{|C|} \sum_{s=1}^{|V_s|} [N_{si} \log (P(w_s|c_j) + P(w_s|w_t)P(w_t|c_j))]$$

Generating a Chinese word according to a polarity

Generating a Chinese word by projectingan English word with same polarity

Parameter Estimation

$$argmax_{\theta} \quad L(\theta|D_t, D_s, U) = L(\theta|D_s) + L(\theta|D_t) + L(\theta|U)$$

$$L(\theta|D_s) = \sum_{i=1}^{|D_s|}\sum_{j=1}^{|C|}\sum_{s=1}^{|V_s|} N_{si} \log \underline{P(w_s|c_j)}\delta_{ij}$$

$$L(\theta|D_t) = \sum_{i=1}^{|D_t|}\sum_{j=1}^{|C|}\sum_{t=1}^{|V_t|} N_{ti} \log \underline{P(w_t|c_j)}\delta_{ij}$$

$$L(\theta|U) = \sum_{i=1}^{|U_t|}\sum_{j=1}^{|C|}\sum_{t=1}^{|V_t|} [N_{ti} \log (\underline{P(w_t|c_j)} + \boxed{P(w_t|w_s)}\underline{P(w_s|c_j)})$$

$$+ \sum_{i=1}^{|U_s|}\sum_{j=1}^{|C|}\sum_{s=1}^{|V_s|} [N_{si} \log (\underline{P(w_s|c_j)} + \boxed{P(w_s|w_t)}\underline{P(w_t|c_j)})]$$

Probability of generating a word *w* for polarity *c* *estimated by EM*

Projection probability estimated by word alignment probability

EM

$$P(c_j|u_{si}) = Z(c_{u_{si}} = c_j) = \frac{\prod_{w_s \in u_{si}} [P(w_s|c_j) + \sum_{P(w_s|w_t)>0} P(w_s|w_t)P(w_t|c_j)]}{\sum_{c_j} \prod_{w_s \in u_{si}} [P(w_s|c_j) + \sum_{P(w_s|w_t)>0} P(w_s|w_t)P(w_t|c_j)]} \quad (5)$$

E-Step

$$P(c_j|u_{ti}) = Z(c_{u_{ti}} = c_j) = \frac{\prod_{w_t \in u_{ti}} [P(w_t|c_j) + \sum_{P(w_t|w_s)>0} P(w_t|w_s)P(w_s|c_j)]}{\sum_{c_j} \prod_{w_t \in u_{ti}} [P(w_t|c_j) + \sum_{P(w_t|w_s)>0} P(w_t|w_s)P(w_s|c_j)]} \quad (6)$$

M-Step

$$P(w_s|c_j) = \frac{1 + \sum_{i=1}^{|D_s|} \Lambda_s(i) N_{si} P(c_j|d_i)}{|V| + \sum_{s=1}^{|V_s|} \Lambda(i) N_{si} P(c_j|d_i)}$$

$$P(w_t|c_j) = \frac{1 + \sum_{i=1}^{|D_t|} \Lambda_t(i) N_{ti} P(c_j|d_i)}{|V| + \sum_{t=1}^{|V_t|} \Lambda(i) N_{ti} P(c_j|d_i)}$$

Experiments

Methods	NTCIR-EN NTCIR-CH	MPQA-EN NTCIR-CH
MT-SVM	62.34	54.33
Para-SVM	N/A	N/A
MT-CoTrain	65.13	59.11
Para-CoTrain	67.21	60.71
CLMM	70.96	71.52

Classification Result using Only English Labeled Data

Note:

- # of parallel sentences: 20,000
- MT engine: Microsoft translator
- NTCIR-EN (English labeled corpus): 4,294
- NTCIT-CH (Chinese labeled corpus): 1,739
- MPQA-EN (English labeled corpus): 4,598

Experiments

Methods	NTCIR-EN NTCIR-CH	MPQA-EN NTCIR-CH
SVM	80.58	80.58
MT-CoTrain	82.28	80.93
Para-CoTrain	82.35	82.18
CLMM	82.73	83.02

Classification Result using English and Chinese Labeled Data

Lost in Translations? Building Sentiment Lexicons Using Context Based Machine Translation (COLING 2012)

Motivation

- Sentiment lexicons are very important for sentiment analysis
- Many lexicons (Ref. P14-P18) in English, but few or even unavailable for other languages

Task

- Use sentiment lexicons in English to automatically generate sentiment lexicons in other languages

Existing Work

- Straightforward translation
  - Suffer from low sentiment word coverage in the bilingual dictionaries
  - Two or more English sentiment words often are translated to the same foreign word
  - Smaller translated sentiment lexicons than the original ones
- English -> Romanian (Mihalcea et al., 2007)
  - 6,856 -> 4,983

Generate Sentiment Lexicon using Context-Aware Translation

- We put the English words into different contexts to effectively prompt the machine translation engine to query the large scale parallel corpora that it is trained on
- We can take advantage of the polysemy of words; one word can mean different things and it usually has various target language translations

Pipeline

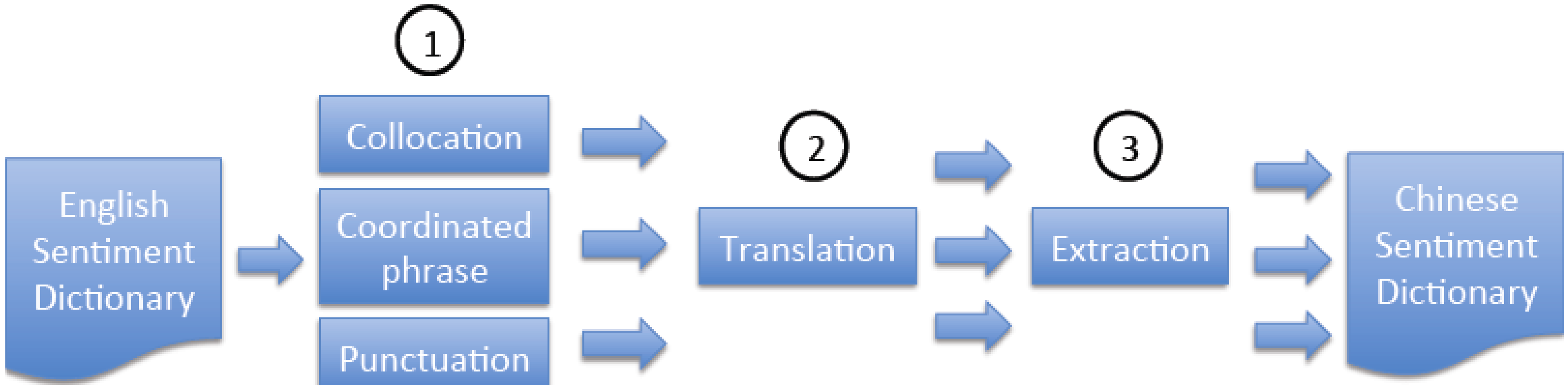


Figure 1: The Flow Chart of Our Approach

Context Generation

- **Collocation**: We obtain the most frequent bi-grams containing the English word. This technique effectively makes the word meaning more specific and concrete, which helps the translation engine to pick out more accurate and diverse translations. For example, we generate "graceful voice" and "graceful dance". Given the contexts, "voice" and "dance", two "graceful" are translated to "优美" and "曼妙", respectively, which are more natural Chinese translations.

- **Coordinated phrase**: We combine two English words that have the same Chinese translations. This makes the translation engine less likely to return the same translations for both words. For example, we create a coordinated phrase by joining "elegant" and "graceful" with the word "and". Joining together, the translations for both words are different from the original translation. More interestingly, putting the two English words in different orders lead to different translations.

- **Punctuation**: We place a punctuation mark, such as period or question mark, at the end of the English word. We use this simple rule to limit the possible parts-of-speech of the translations. For example, "effusive." is translated to "热情洋溢", while "effusive" is translated to "感情奔放的"; after adding punctuation context, "effusive" is translated to words that have different parts-of-speech. We can also combine this technique with the coordinated phrase technique.

Context	English	Translation
None	elegant graceful	优雅 优雅
Collocation	graceful voice graceful dance	优美的声音 曼妙的舞姿
Coordinated phrase	elegant and graceful graceful and elegant	典雅大方 雍容典雅
Punctuation	graceful. elegant. graceful and elegant.	优美。 优雅。 婉约和优雅。

Table 1: Chinese Translations of "graceful" and "elegant" in different contexts

Experimental Results

Lexicon	#POS	#NEG	#TOTAL
MPQA(EN)	1,481	3,080	4,561
DICT	742	1,139	1,881
DICT + Stem	814	1,230	2,044
DICT + Multiple	2,811	3,799	6,610
MT	1078	2,104	3,182
CONTEXT	**3,511**	**5,210**	**8,721**

Table 2: Vocabulary Size of Different Lexicons

Lexicon	Precision
DICT	93%
DICT + Stem	93%
DICT + Multiple	82%
MT	91%
CONTEXT	91.5%

Table 3: Precision of Different Lexicons

Experimental Results

Lexicon	NTCIR	Weibo
DICT	61.9%	57.6%
DICT + Stem	61.9%	57.5%
DICT + Multiple	64.7%	61.7%
MT	66.2%	64.6%
CONTEXT	**70.1%**	**73.5%**

Table 4: Classifier Accuracy Using Different Lexicons

Summary

- We present a very simple but effective approach to automatically generating sentiment lexicons for other languages with English sentiment lexicons
- The approach is language independent

Revisit Sentiment Analysis

Future Work (I)

- Sentiment vs. topic classification
  - Single word can determine the class
    - Subjective vs. objective
    - Positive vs. negative
  - Invertible (negation)
  - Contextual
    - Local and long distance context
  - Compositional
  - Target dependent
  - Topic (domain) dependent
  - Annotation and adaption (language, domain, topic)
  - Imbalance classification

Future Work (II)

- Implicit sentiment (semantic context of sentiment)
  - Sarcasm, Irony, Metaphor, Polysemous
- Sentiment insight mining (summarization)
- Spam detection
- Understand the user (sentiment holder)

Thanks