Irony Detection in English Tweets

Group 15:

M S Nishanth CS18MTECH11005
Nikhil P Kumar CS18MTECH11017
Sharath R CS18MTECH11026

Gretchen Grundler @Ebony_n_Irony · 2m
Replying to @OMG_AnnGee
Wait, am I late? Everybody knew but me?!?!?!? Lol

B Dubs @dmartino5000 · 7m
Replying to @JoeSaluzzi
The **irony** there........

1

tim mullaney @timmullaney · 13m
Replying to @rickjnewman
but i mean, could you imagine? it would bust all known **irony** meters. to little bits

Retarded Pingas @MonsieurPingas · 7m
Replying to @Corbelle13 @marquisliberal @BortoneA
That **irony**

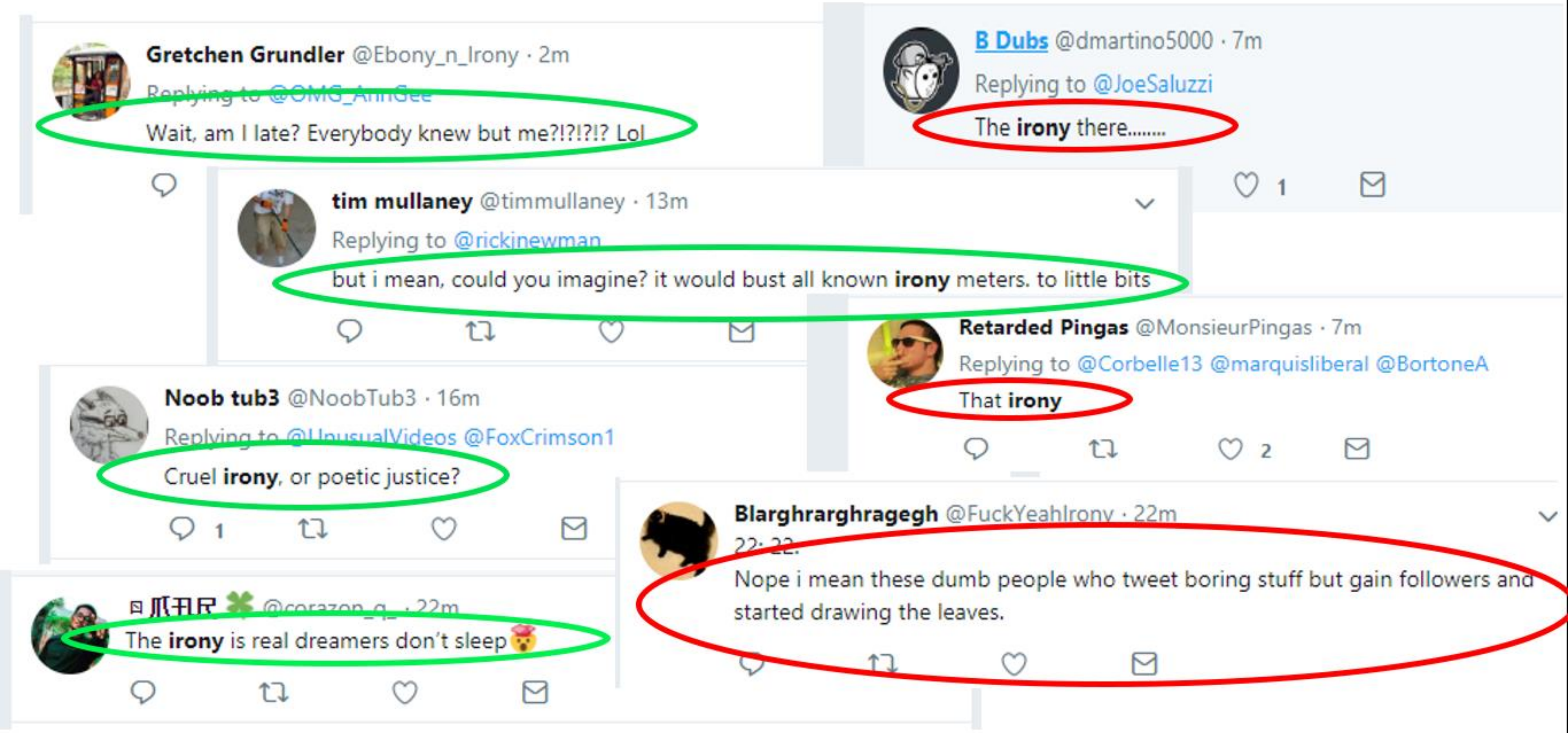
2

Noob tub3 @NoobTub3 · 16m
Replying to @UnusualVideos @FoxCrimson1
Cruel **irony**, or poetic justice?

1

Blarghrarghragegh @FuckYeahIrony · 22m
22: 22.
Nope i mean these dumb people who tweet boring stuff but gain followers and started drawing the leaves.

爪ㄒ卄尺 🍀 @corazon_q_ · 22m
The **irony** is real dreamers don't sleep 🤯


Gretchen Grundler @Ebony_n_Irony · 2m
Replying to @OMG_AnnGee
Wait, am I late? Everybody knew but me?!?!?!? Lol
B Dubs @dmartino5000 · 7m
Replying to @JoeSaluzzi
The irony there........
1
tim mullaney @timmullaney · 13m
Replying to @rickinewman
but i mean, could you imagine? it would bust all known irony meters. to little bits
Retarded Pingas @MonsieurPingas · 7m
Replying to @Corbelle13 @marquisliberal @BortoneA
That irony
2
Noob tub3 @NoobTub3 · 16m
Replying to @UnusualVideos @FoxCrimson1
Cruel irony, or poetic justice?
1
Blarghrarghragegh @FuckYeahIrony · 22m
Nope i mean these dumb people who tweet boring stuff but gain followers and started drawing the leaves.
@corazon_q_ · 22m
The irony is real dreamers don't sleep

Introduction

The expression of one's meaning by using language that normally signifies the opposite, typically for humorous or emphatic effect.

Types:

- Situational irony
- Dramatic irony
- Socratic irony
- Verbal irony

Motivation

>> To build a stronger partnership between humans and machines. Moving to a world where the machines understand us, answering (and behaving) in the way a human would do.

Challenges

- Irony introduce new challenges to many text-related tasks including information retrieval (IR), classification, and clustering.
- Mere identification of a word as polar (positive or negative) word is inadequate for fine-grained sentiment analysis which results beyond positive or negative.
- There are many hidden properties of words other than being positive or negative which can lead to enrichment of existing sentiment analysis.
- This work focuses on finding these properties in polar words in light of different applications of sentiment analysis.

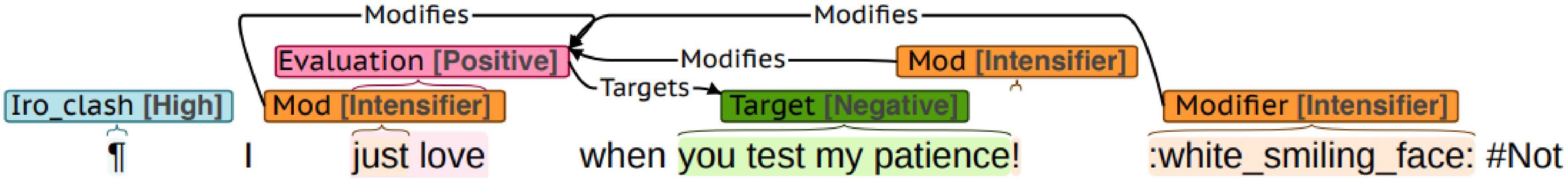
Problem Definition

P1: To predict if a given text is ironic or not.

P2: We next try to find the category of irony

- verbal irony realized through a polarity contrast,
- verbal irony without such a polarity contrast
- descriptions of situational irony
- non-irony

Brat annotation

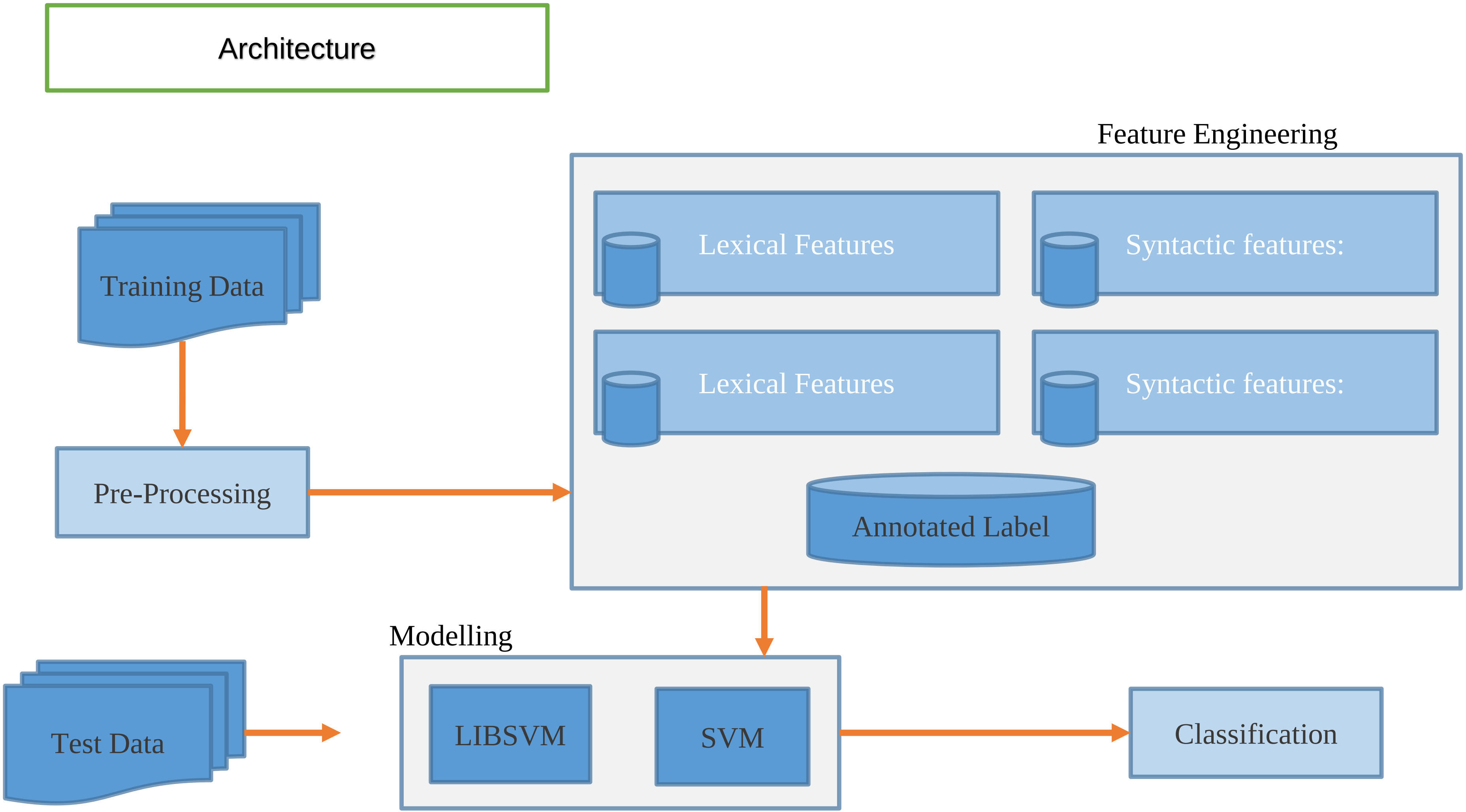


Approaches

Rule Based Approach : It attempts to identify sarcasm through specific evidences captured in terms of rules that rely on indicators of sarcasm.

Statistical Approach : Statistical approaches to irony detection vary in terms of features and learning algorithms. Most of the approaches use bag-of-words as features.

Deep Learning based Approach : The use of architectures based on deep learning can be used to get an improvement in performance.


Architecture
Feature Engineering
Training Data
Lexical Features
Syntactic features:
Lexical Features
Syntactic features:
Pre-Processing
Annotated Label
Modelling
Test Data
LIBSVM
SVM
Classification

Feature Engineering

We extracted the a number of features from the dataset. These features would help us explain the tweets to machine in a better way.

1. Lexical features
2. Syntactic features
3. Sentiment features
4. Semantic features

Lexical features:

- Bags-of-words (BoW) features.
- token bigrams
- Character flooding
- Punctuation flooding
- Capitalization
- Hashtag frequency
- Hashtag-to-word ratio
- Emoticon frequency
- Tweet length.

Syntactic features :

For each PoS tag:

- whether it occurs in the tweet or not.
- whether the tag occurs 0, 1, or > 2 times.
- The frequency of the tag .
- The number of interjections.

For each Named entity:

- named entity present or not.
- the number of named entities in the text.
- the number of tokens part of named entity.
- frequency of tokens that are part of a named entity.

Sentiment features :

- AFINN (Nielsen, 2011)
- MPQA (Wilson et al., 2005)
- General Inquirer (GI) (Stone et al., 1966)
- NRC Emotion Lexicon (Mohammad and Turney, 2013)
- Bing Liu (Minqing Hu and Bing Liu. 2004)

For each lexicon

- number of positive, negative and neutral lexicons.
- the overall tweet polarity.
- the absolute difference.
- a binary feature indicating whether there is a polarity .

The sentiment lexicon features were extracted in two ways:

- by considering all tokens in the instance.
- by considering only hashtag tokens.

Entry	Source	Positiv	Negativ	Pstv	Affil	Ngtv	Hostile
A	H4Lvd						
ABANDON	H4Lvd		Negativ			Ngtv	
ABANDON	H4		Negativ				
ABATE	H4Lvd		Negativ				
ABATEMEI	Lvd						
ABDICATE	H4		Negativ				
ABHOR	H4		Negativ				Hostile
ABIDE	H4	Positiv			Affil		
ABILITY	H4Lvd	Positiv					
ABJECT	H4		Negativ				
ABLE	H4Lvd	Positiv		Pstv			
ABNORM/	H4Lvd		Negativ			Ngtv	
ABOARD	H4Lvd						

MPQA_lexicon.csv

```
abandoned,-1
abandonment,-1
abandon,-1
abase,-1
abasement,-1
abash,-1
abate,-1
abdicate,-1
aberration,-1
aberration,-1
abhor,-1
abhor,-1
abhorred,-1
abhorrence,-1
abhorrent,-1
abhorrently,-1
abhors,-1
abhors,-1
abidance,1
abidance,1
```

```
aback,anger,0

aback,anticipation,0

aback,disgust,0

aback,fear,0

aback,joy,0

aback,negative,0

aback,positive,0

aback,sadness,0

aback,surprise,0

aback,trust,0

abacus,anger,0

abacus,anticipation,0

abacus,disgust,0

abacus,fear,0

abacus,joy,0

abacus,negative,0

abacus,positive,0

abacus,sadness,0

abacus,surprise,0
```

“I just love when you test my patience!”

What sort of emotion does it convey?

Semantic feature:

- used word embedding cluster features generated with Word2Vec (Mikolov et al., 2013).
- In this word embedding's, are the distributed vector representation of words that capture the syntactic and semantic relationships among words.
- Maps Love and Hate together.

Corpus Collection for Word2Vec

- from Twitter: API key, API secret, Access token and Access token secret.
- Collect the API keys and create your access token via twitter acc.
- A python library called Tweepy to download the corpus wrt hastag and time stamp.
- Hash tags:

#irony, #sarcasm, #not, #sarcastic, #ironic.

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW TEAM ms nishanth

A1 (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49)

	A
1	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49)
2	[0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
3	[1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
4	[0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
5	[0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
6	[0, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
7	[0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
8	[0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
9	[0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0]
10	[1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
11	[0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
12	[1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
13	[0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
14	[0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
15	[1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
16	[1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
17	[0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
18	[1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
19	[1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
20	[0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
21	[0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
22	[1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
23	[0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
24	[0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
25	[0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
26	[1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
27	[1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
28	[1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
29	[0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

word2vec

READY 100%

DataSet

SemEval2018-T3-train-taskA

SemEval2018-T3-train-taskA_emoji

SemEval2018-T3-train-taskA_emoji_ironyHashtags

SemEval2018-T3_input_test_taskA

SemEval2018-T3_input_test_taskA_emoji

Tweet index	Label	Tweet text
1	1	Sweet United Nations video. Just in time for Christmas. #imagine #NoReligion http://t.co/fej2v3OU
2	1	@mrdahl87 We are rumored to have talked to Erv's agent... and the Angels asked about Ed Escobar...
3	1	Hey there! Nice to see you Minnesota/ND Winter Weather
4	0	3 episodes left I'm dying over here
5	1	"I can't breathe!" was chosen as the most notable quote of the year in an annual list released by
6	0	You're never too old for Footie Pajamas. http://t.co/ElzGqsX2yQ
7	1	Nothing makes me happier then getting on the highway and seeing break lights light up like a Chris
8	0	4:30 an opening my first beer now gonna be a long night/day
9	0	@Adam_Klug do you think you would support a guy who knocked out your daughter? Rice doesn't deserv
10	0	@samcguigan544 You are not allowed to open that until Christmas day!

Experiment

- Preprocessing
  - We normalized hyperlinks and @-replies/mentions to http://someurl and @someuser, respectively.
  - Other preprocessing steps involve tokenization and lemmatization.
- Extracted Features
  1. Lexical features - 7
  2. Syntactic features - 82
  3. Sentiment features -16*5
  4. Semantic features - 50

Model Training

For Binary Classification:

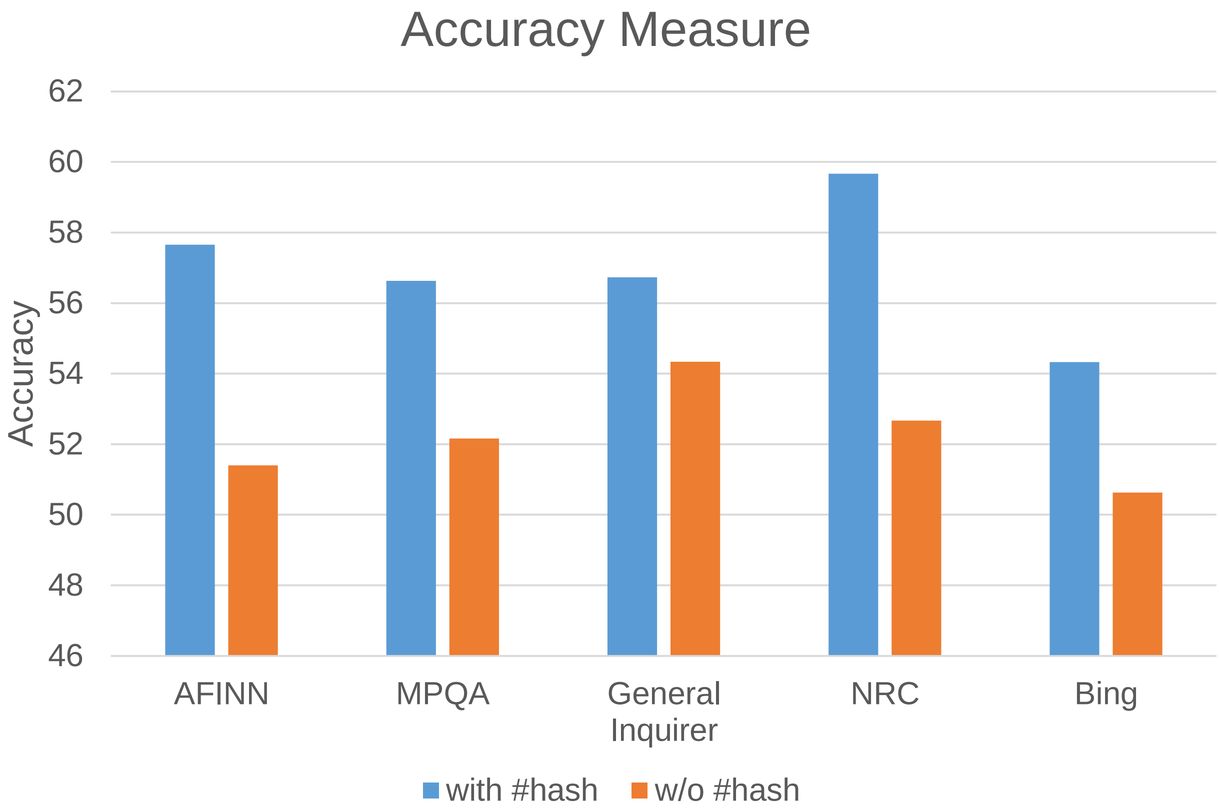
- SVM
- Naïve Bayes

For Multi-Class

- LIBSVM
- Naïve Bayes
- Neural Network

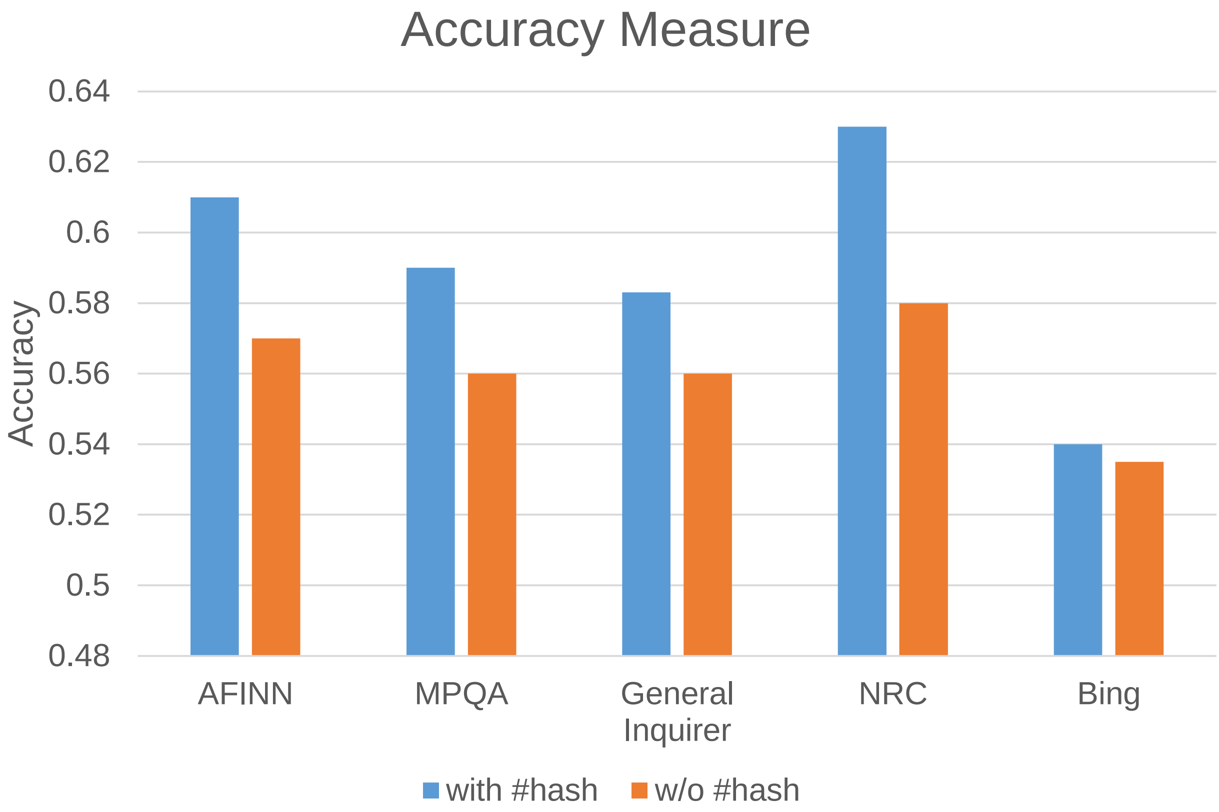
Results

- Sentiment Analysis (TEST A)

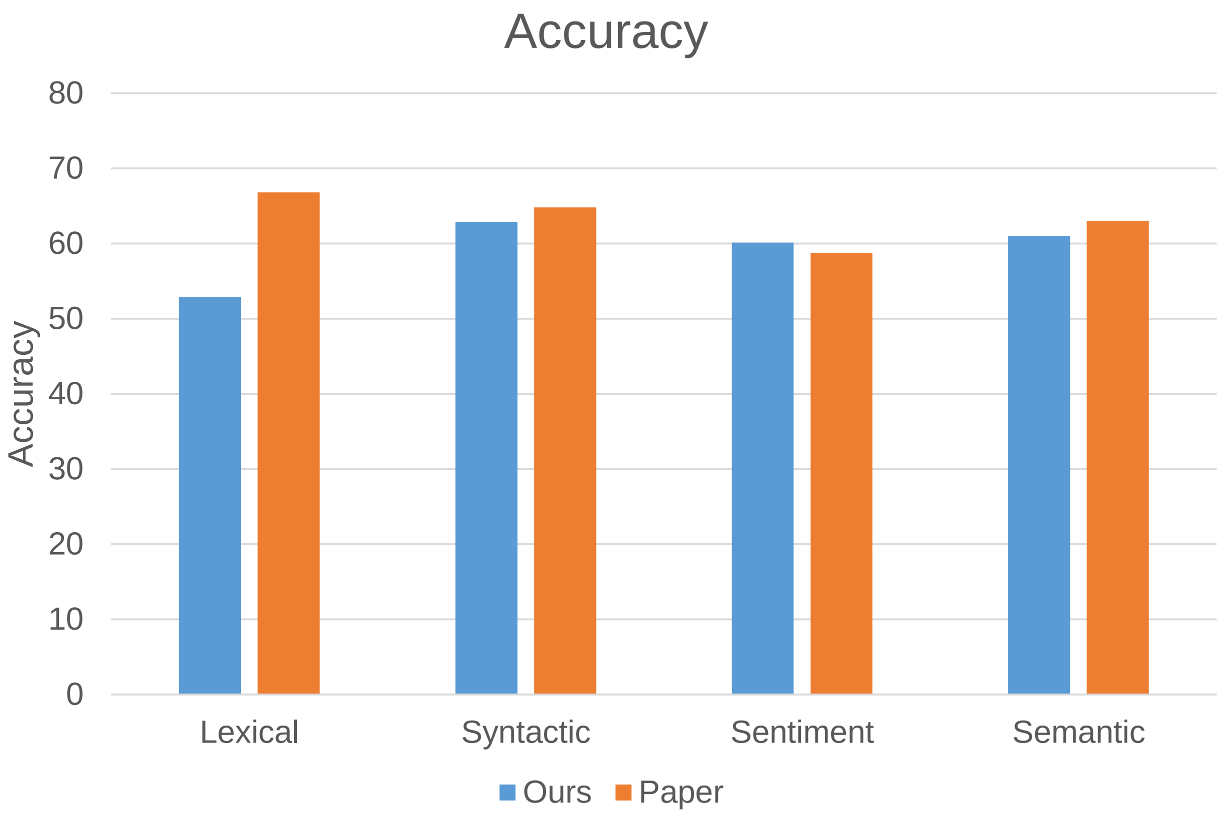


Results

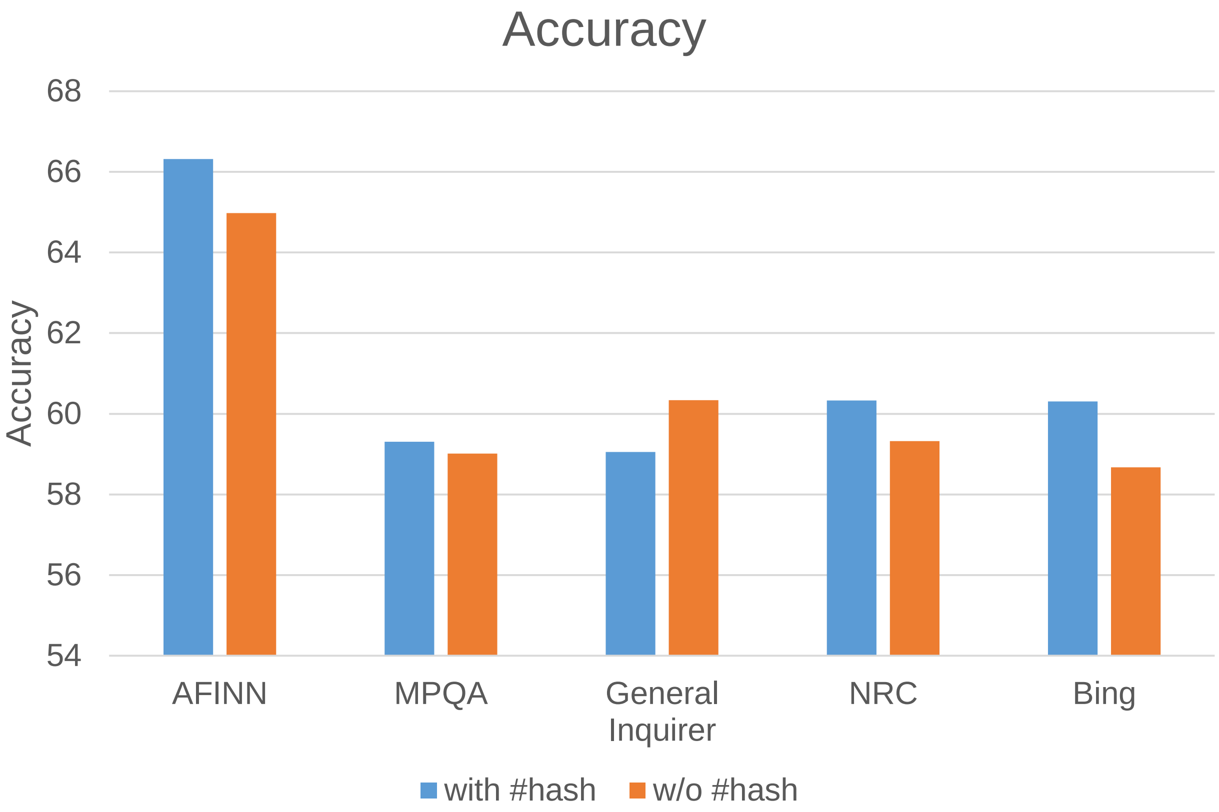
- F1-Score (TEST A)



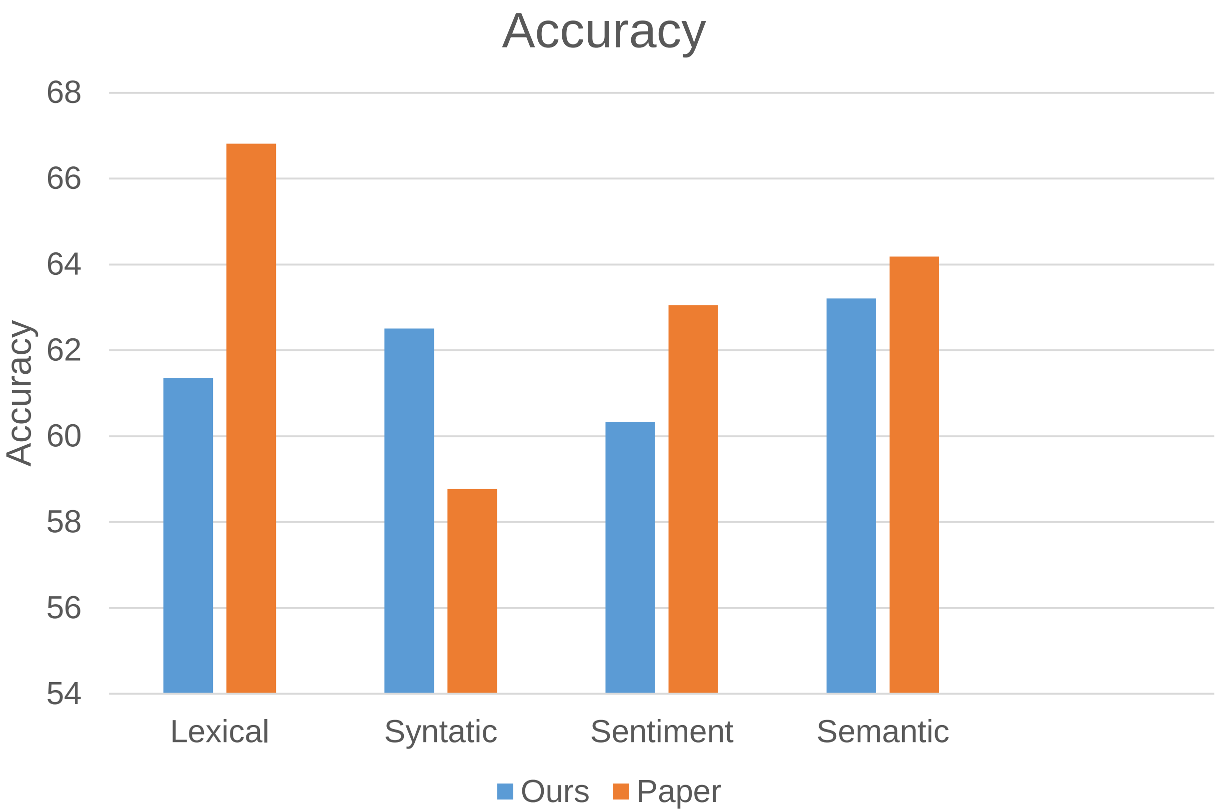
Result Comparison with Paper(Set A)



Sentiment Analysis (TEST B)



Result comparison with paper(Set B)



Future Study

- Using result of P1 to get better result for P2.
- GloVe Embedding for Semantic Feature
- LDA to group word Embedding.

References

- Van Hee,C., Lefever, E., and Hoste, V. (2016). Exploring the Automatic Recognition of Irony in English tweets
- Van Hee,C., Lefever, E., and Hoste, V. (2016). Exploring the Realization of Irony in Twitter Data
- Joshi, A., Bhattacharyya, P., Carman, M. (2016). Automatic Sarcasm Detection: A Survey

. THANK YOU