Satire vs Fake News: You Can Tell by the Way They Say It

Dipto Das and **Anthony J Clark**

Computer Science Department

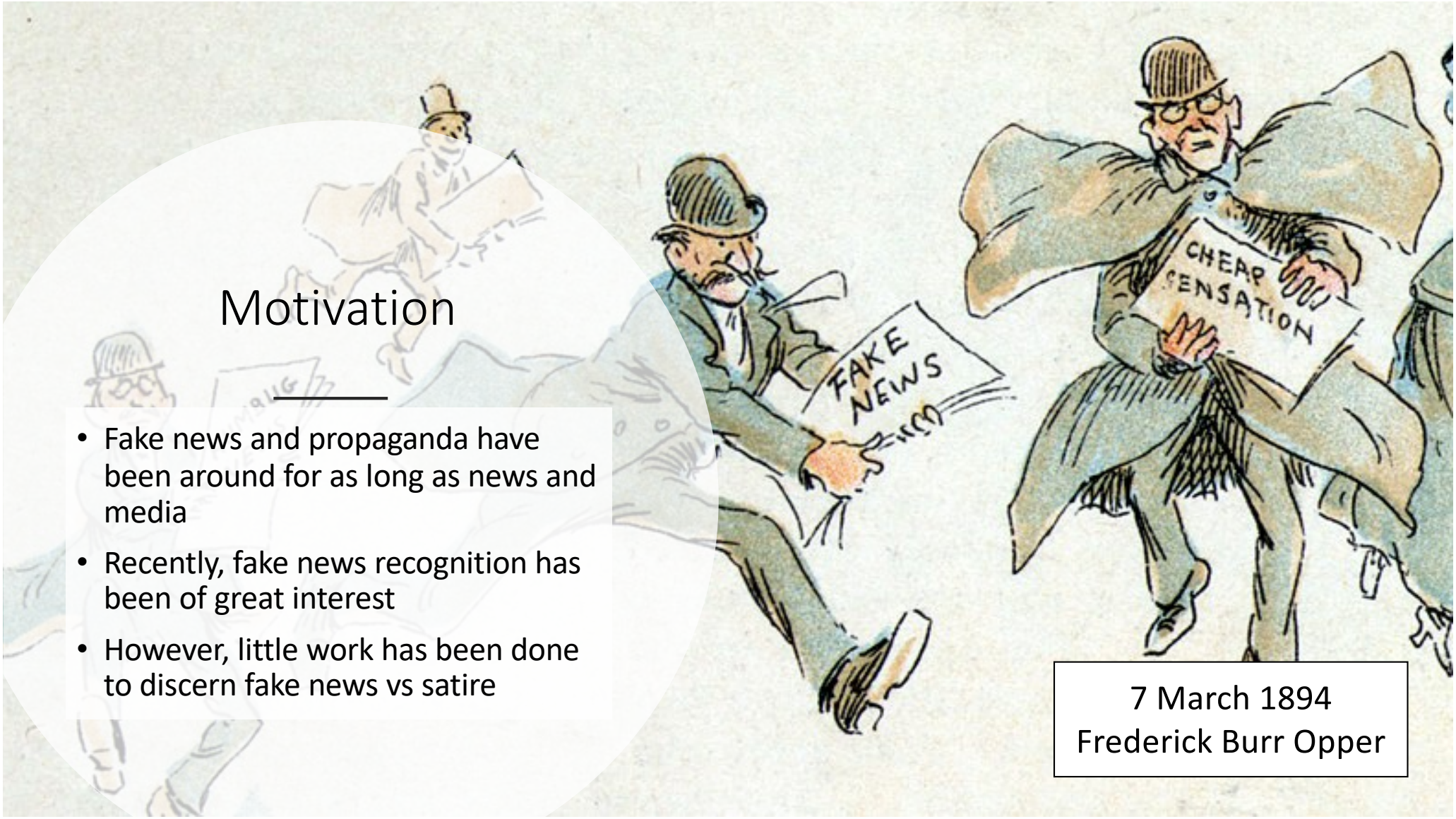
Missouri State University

Detecting Satire and Sarcasm

Sample sarcastic review:	Thank you for your WONDERFUL service! 
Sample auto-reply:	Thank you for your kind review.

Motivation

- Fake news and propaganda have been around for as long as news and media
- Recently, fake news recognition has been of great interest
- However, little work has been done to discern fake news vs satire



7 March 1894
Frederick Burr Opper

Our Goals

1. Short term: classify articles as either fakes new or satire
   - We will not consider other classes
   - Start with a pre-existing dataset
   - Consider only recent articles in English

2. Long term: develop social media tools for tagging content
   - Classify posts as fake news, satire, serious, funny, etc.
   - Help new users that are not familiar with newer forms of communication (e.g., memes)
   - Transfer tools to other languages and domains

Related Work

- Most recent studies consider satire, news parody, manipulation, fabrication, and large-scale hoaxes as different kinds of fake news
  - Rubin et al, Tandoc et al, etc.
  - These studies do **not** consider the motivation of content creators

- Other studies do not consider satire, but they define fake news as misinformation that is presented to deceive
  - Golbeck et al.
  - Did not provide any definition of satire

Fake News or Satire

For this study, we consider

Fake News is misinformation meant to deceive

And

Satire is misinformation meant to entertain and criticize

The key difference between Fake News and Satire is the motivation

How We Read and Write Sarcastic Content

Finding from qualitative study

- Unusual expression of sentiment in text, i.e., storytelling approach of satire should be different.

- Narrative Trajectory of satire and fake news should be different.

Our Key Idea

- Rather than use raw text, we propose to use narrative trajectories

- Narrative trajectory based on sentiment is an important indicator of the storytelling patterns of text articles
  - Gao et al., Reagan et al., Samothrakis et al.

- Idea: use filtered sentence-wise sentiment scores of an article to indicate the motivation and thereby the classification

This Study

Text-tone-based approach to classify fake news and satire

Background

Investigating an Existing System

Tone to Differentiate Satire and Fake News

Existing System

- Dataset from Golbeck et al.
- 203 satires, 283 fake news
- Relate to the 2016 US presidential election
  - Minimal variation in the theme of the articles

Existing System

Multinomial naïve Bayes

- 79.1% accuracy
- 0.88 ROC area

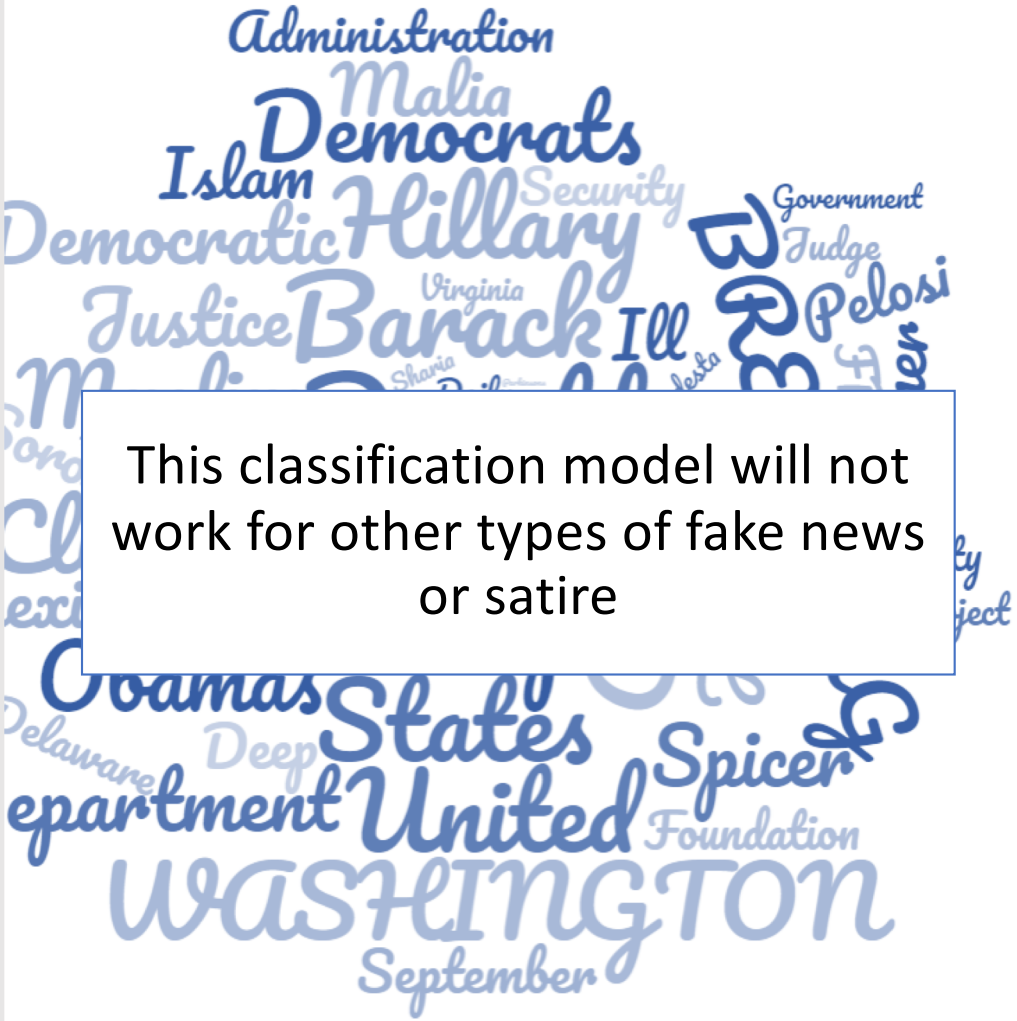
- High dependence on proper nouns in the articles
- Shannon Information Gain is used to get most occurring words

Existing System

Multinomial naïve Bayes

- 79.1% accuracy
- 0.88 ROC area

- High dependence on proper nouns in the articles
- Shannon Information Gain is used to get most occurring words



This classification model will not work for other types of fake news or satire

Improving the Existing System

- Word Stemming
  - Reduce words to their root/base forms; e.g.: working → work
  - Lovins Stemmer algorithm
- Discarding stop-words
  - As defined by McCallum et al. ("the", "of", "is“)
- Minor accuracy improvement

Metric	Golbeck et al.	Our improvement
Accuracy	79.10%	80.30%
ROC area	0.88	0.87

Tone Analysis

- Next we want to look at using sentiment to discover motivation
- Motivation is the difference between fake news and satire

- We use the IBM Tone Analyzer to calculate scores for each sentence in an article

- The IBM Tone Analyzer produces 13 values for each sentence

IMB Tone Analyzer Output Per Sentence

Language Scores

1. Analytical
2. Confidence
3. Tentative

All scores are between 0 and 1

Emotion Scores

4. Anger
5. Joy
6. Fear
7. Disgust
8. Sadness

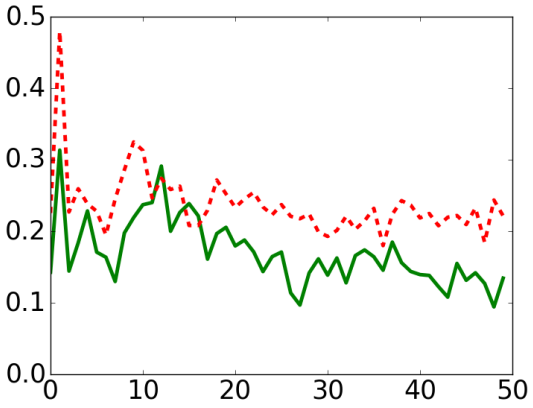
Social Scores

9. Agreeableness
10. Conscientiousness
11. Emotion
12. Extraversion
13. Openness

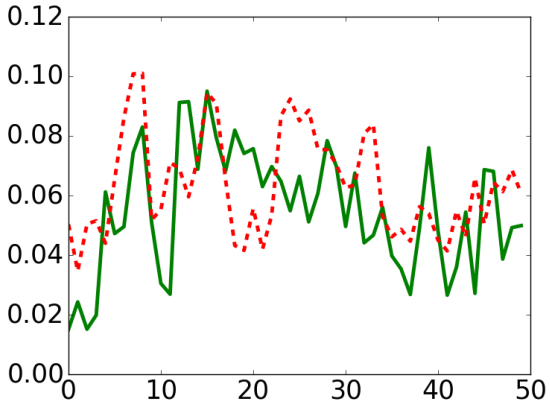
Narrative Trajectories

Satire
Fake news

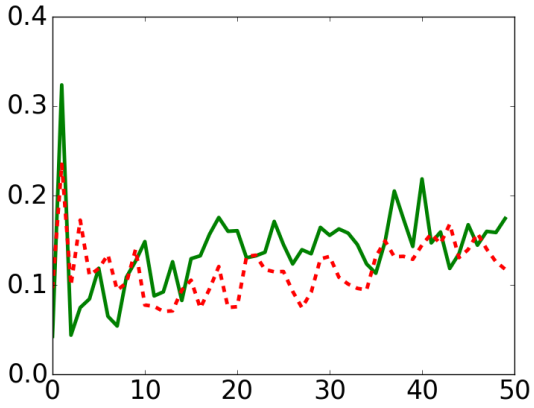
- Hanning smoothing (window size = 3)
- Cropped to remove boundary effects from filtering
- Interpolated to have a canonical length of 50 samples



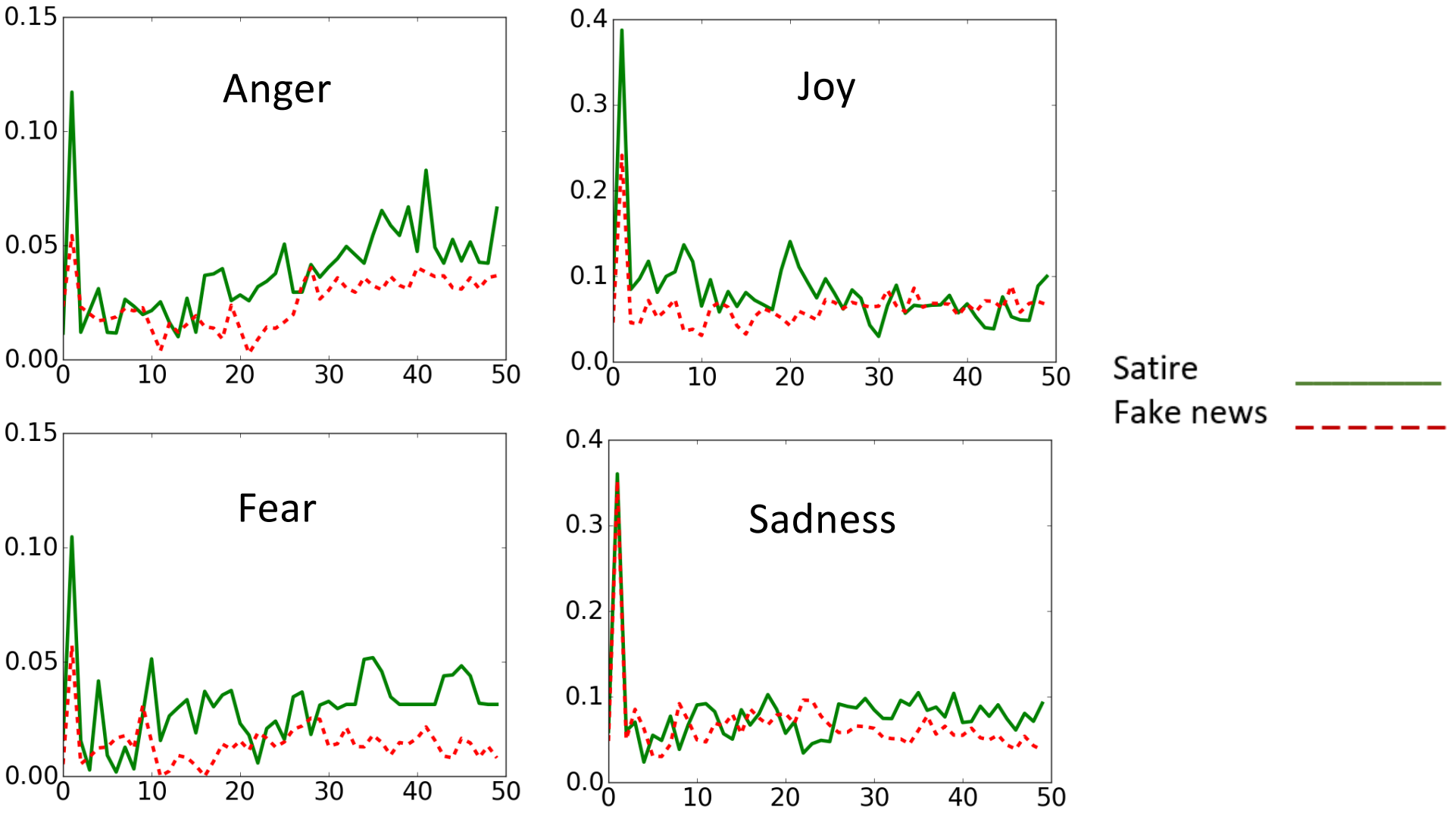
Analytical



Confident



Tentative


Anger
Joy
Fear
Sadness
Satire
Fake news
0.15
0.10
0.05
0.00
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
50

SMOTE Sampling

- We use synthetic minority over-sampling technique (SMOTE)

- The dataset includes 41.% and 58.3% satire and fake news articles, respectively

Classification

Using tone scores should result in less dependence on the actual text

- Less dependent upon a specific domain (e.g., politics)
- Less dependent upon a time (e.g., near an election)
- Less dependent upon the place
- Less dependent upon the language

Additional features

- Subjectivity of article titles
- Polarity of article titles
- Article themes

Classification Techniques

Classifiers

- Naïve Bayes
- Neural networks
- SVM
- Random forests

Approaches	Accuracy	ROC area
Naïve Bayes (Golbeck et al.)	79.10%	0.88
Improved naïve Bayes	80.30%	0.87
(Only) Tone-based classifier	75.80%	0.83
Text, Tone, Theme-based classifier	82.50%	0.91

Performance of classification task with tone data extracted from articles (text independent)								
Class	TP Rate	FP Rate	Precision	Recall	F1 Score	MCC	ROC Area	PRC Area
Satire	0.729	0.212	0.775	0.729	0.751	0.518	0.827	0.833
Fake news	0.788	0.271	0.743	0.788	0.765	0.518	0.827	0.788
Weighted Avg.	0.758	0.242	0.759	0.758	**0.758**	0.518	**0.827**	0.811

Performance of classifier model with text, tone, and theme data combined								
Class	TP Rate	FP Rate	Precision	Recall	F1 Score	MCC	ROC Area	PRC Area
Satire	0.905	0.254	0.782	0.905	0.839	0.660	0.911	0.894
Fake news	0.746	0.095	0.887	0.746	0.811	0.660	0.911	0.919
Weighted Avg.	0.826	0.174	0.834	0.826	**0.825**	0.660	**0.911**	0.907

Feature	Information Gain
Conspiracy (theme)	0.1035
Document Joy (tone)	0.0668
Document Analytical (tone)	0.0402
Sentences Analytical (tone)	0.0395
Sensationalist Crime/Violence (theme)	0.0390

Experiment on Non-English Dataset

Dataset Collection:

- 30 satire articles from Motikontho and Earki
- 30 fake news articles as identified by Jachai

- We tried training a classifier on both the native articles and using automatically translated versions

Experiment on Non-English Dataset

- Testing using our small Bengali Satire Dataset
- Trained improved naïve Bayes classifier and tone-based classifier
- Trained using English dataset from Golbeck et al.

Model	Accuracy
Improved Naïve Bayes	93.33%
Tone-based classifier	61.29%

Observations

- Tone-based approach < naïve Bayes approach: non-English dataset
- Tone-based approach > naïve Bayes approach: English dataset

The differences in tone between satire and fake news is enough

Or

Are the observations due to the particular features of the dataset

Effect Size of Features

Language/Emotion	t-value	p-value
Analytical	0.7816	0.44
Confident	0.2387	0.81
Tentative	0.9603	0.34
Anger	0.8443	0.4
Disgust	0.0	INF
Fear	0.3214	0.75
Joy	0.3044	0.76
Sadness	0.4674	0.64

Takeaways

- Some differences in narrative trajectories in sarcastic tones

- Tone information:
  - A useful feature
  - May not be enough to create a classifier

- Use of words in text is a better stand-alone predictor

References

- Jennifer Golbeck, Matthew Mauriello, Brooke Auxier, Keval H Bhanushali, Christopher Bonk, Mohamed Amine Bouzaghrane, Cody Buntain, Riya Chanduka, Paul Cheakalos, Jennine B Everett, et al. Fake news vs satire: A dataset and analysis. In Proceedings of the 10th ACM Conference on Web Science, pages 17–21. ACM, 2018.
- Mikhail Khodak, Nikunj Saunshi, and Kiran Vodrahalli. A large self-annotated corpus for sarcasm. arXiv preprint arXiv:1704.05579, 2017.
- Merriam-Webster Dictionary. Satire Definition. https://www.merriam-webster.com/dictionary/satire, n.a. Online; accessed 25 September 2018.
- Das, Dipto, "A Multimodal Approach to Sarcasm Detection on Social Media" (2019). *MSU Graduate Theses*. 3417.
- Mathieu Cliche. The sarcasm detector. http://www.thesarcasmdetector.com/, 2014. Accessed: May 19, 2018.

Thank you!

Questions?

Phrases + invenclo

Look @ tone of Bengali text