Propaganda or Clickbait? Understanding and Classifying Types of Misinformation using Recurrent Neural Networks

UNC CHARLOTTE

Sashank Santhanam and Samira Shaikh

Department of Computer Science

INTRODUCTION

World Economic Forum reports that spread of misinformation on online social media has global consequences and poses significant risks to society[1]. Social media platforms such as Twitter, Facebook provide a means for information to reach to the masses in near real time. However, these platforms have increasingly been used to spread misinformation. Prior research has shown that misinformation in the form of propaganda, hoaxes, clickbait, rumors, satire and conspiracy theories spreads more rapidly than does accurate information. There is an urgent need to accurately identify and classify misinformation; a recent poll conducted by Pew research center found that nearly 66% of users get news from Facebook, and 59% from Twitter. Extant research to tackle this problem has relied on a network-based approaches while not adequately addressing the linguistic aspects of content. Research into linguistic aspects that distinguish reliable news from types of misinformation is still nascent.

OBJECTIVES

- In this work, we report on our research to combat the spread of misinformation from different news outlets on social media (e.g. Twitter).
- Use deep learning models to classify (binary and multi-class) tweets produced by a set of 83 news outlets on Twitter when compared to traditional machine learning approaches such as SVM and Naive Bayes.
- Identify linguistic features using LIWC that would help the model to distinguish between real news and other types of accounts and validate it with the help of ANOVA's

RELATED WORK

The problem of misinformation has been gained significant traction over recent years and has been well studied in the field of journalism and computer science [2]. Misinformation is a broad term and consists of different types of news categories. Some of the identified types of news categories include propaganda, clickbait, hoax and satire which produce news similar to the content similar to real news but not providing the accurate facts or biasing the content in favor or against a person or organization.

Some of the approaches from a computer science perspective towards making attempts to solve misinformation included work done by Controy et al. who used n-gram plus syntax model along with incorporating profile information to classifying fake news[5]. Recently, Volkova et al. create a predictive model using neural networks to classify between the four types of misinformation outlets and also identify some linguistic features that help improve the performance of the model[3]. Other work done in this area, include the work done by Horne et al on three publicly available datasets to exploit some of the traits exhibited by misinformation news outlets when compared to real news[2].

DATA EXPLORATION

To create our dataset, we used the Twitter Streaming API to extract tweets from a set of 83 accounts annotated as propaganda, clickbait, hoax, satire and real news by external sources. The data collection period was between May 23, 2017 to June 6, 2017. These 83 accounts were shortlisted based on the conditions:

1. Account creation date
2. Low friends to follower ratio

Table 1 - Distribution of different types of news outlets.

Account Type	Number
Real News	31
Propaganda	30
Clickbait	18
Hoax	2
Satire	2

METHOD

Method - 1: Classification

Baseline Models: We used Support Vector Machines with linear kernel and Naive Bayes models with 10 fold cross validation to classify (binary) tweets from the real news account or the misinformation accounts. To train these models we divided it into 85% for the training data and 15% for the testing data

RNN + LSTM: We built a Recurrent Neural Network (RNN) with LSTM to perform a binary classification between real and misinformation and also a used the same model to perform multi-class classification between the different types of news outlets. We initialized the embedding layer of the RNN with pre-trained GloVe embedding[7]. The input sequences were padded to a uniform length of 200. Table 2. represents the hyper parameters used in building the model.

Table 2. Hyperparameters

Hyperparameters	Values
Batch size	500
Learning Rate	0.001
Epochs	10
Dropout	0.5
Optimizer	Adam

Method - 2: Linguistic Analysis

To get a better understanding of the language used by real news when compared to misinformation outlets. We identify several features with the help of LIWC dictionary[8] and categorized into three categories:

1. Stylistic
2. Linguistic
3. Psychological

RESULTS

Method – 1: Classification

Table 3 represents the accuracy of the baseline and RNN + LSTM models for binary classification of tweets from real and misinformation accounts. RNN + LSTM model significantly outperforms the the baseline models.

Table 3 – Binary classification results

Method	Accuracy (%)
Naïve Bayes	41.5
SVM	65
RNN + LSTM	85.4

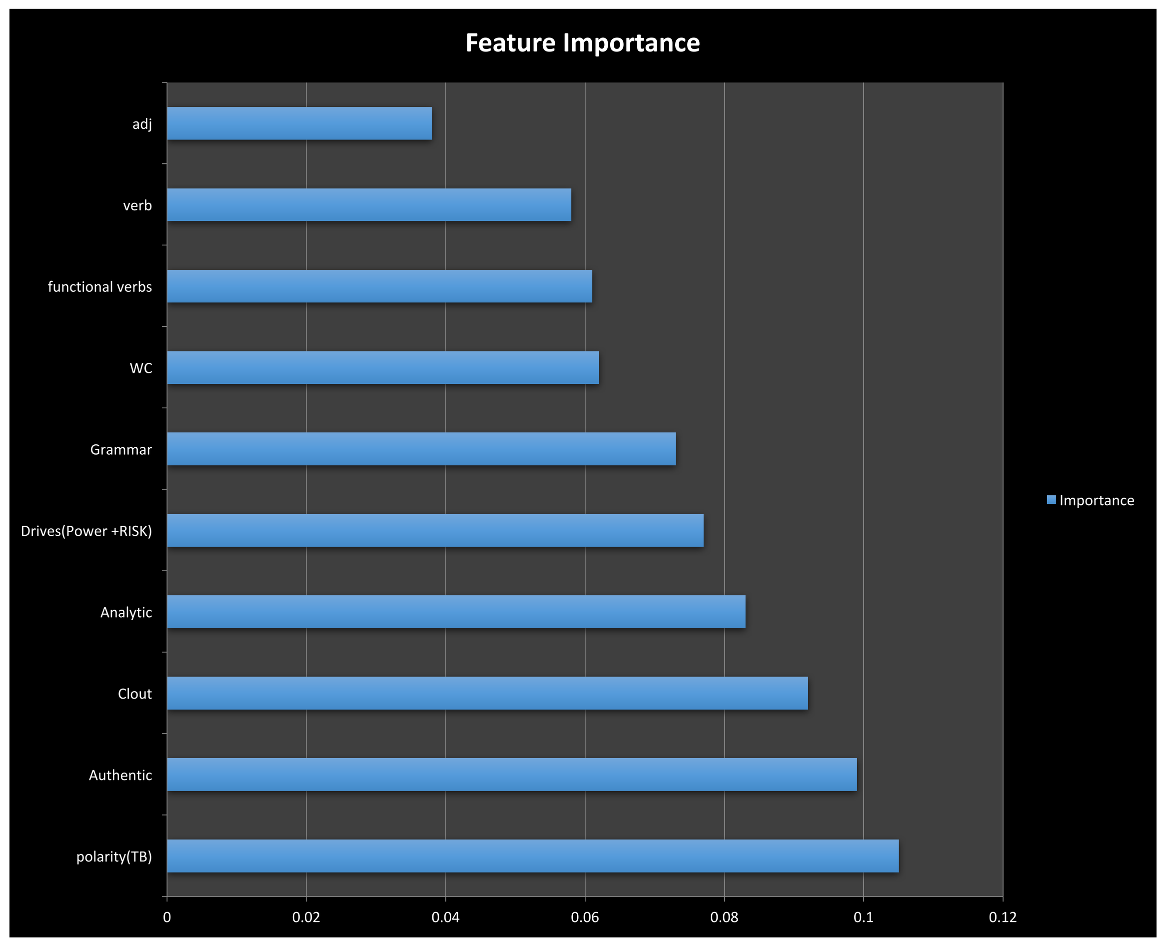
Table 4 represents the performance of the model on multi-class classification and compares the performance to an existing model.

Table 4- Multi-class classification results

Method	Accuracy (%)
Existing Method	63
RNN + LSTM	75.8

Method – 2: Linguistic Analysis

Feature	Significance
Polarity	P<0.0001 ***
Authentic	P<0.0001 ***
Clout	P<0.0001 ***
Analytic	P=0.085 *
Drive (Power + Risk)	P<0.0001 ***
Grammar	P<0.0001 ***
WC	P<0.0001 ***
Functional verbs	P<0.0001 ***
Verb	P<0.0001 ***
Adj	P<0.0001 ***



CONCLUSION & FUTURE WORK

- This preliminary work shows that misinformation can be detected on social media.
- Deep Learning performs better compared to traditional machine learning algorithms.
- This work facilitates understanding of the linguistic style present in tweets from misinformation accounts.
- In future work, we will incorporate linguistic features into the deep learning model and evaluate the classification performance. The long term goal of this project is to include features from images and combine it with the RNN model to maximize performance

REFERENCES

1. Howell, L., 2013. Digital wildfires in a hyperconnected world. *WEF Report*, *3*, pp.15-94.
2. Horne, B.D. and Adali, S., 2017. This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. *arXiv preprint arXiv:1703.09398*.
3. Volkova, S., Shaffer, K., Jang, J.Y. and Hodas, N., 2017. Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (Vol. 2, pp. 647-653).
4. Conroy, N.J., Rubin, V.L. and Chen, Y., 2015. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology, 52(1), pp. 1-4.*
6. Zhiguo Wang, Wael Hamza, and Radu Florian. Bilateral multi-perspective matching for natural language sentences. 2017.
7. Pennington, J., Socher, R. and Manning, C., 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*(pp. 1532-1543).
8. Pennebaker, J.W., Francis, M.E. and Booth, R.J., 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, *71*(2001), p.2001.

Propaganda or Clickbait? Understanding and Classifying Types of Misinformation using Recurrent Neural Networks

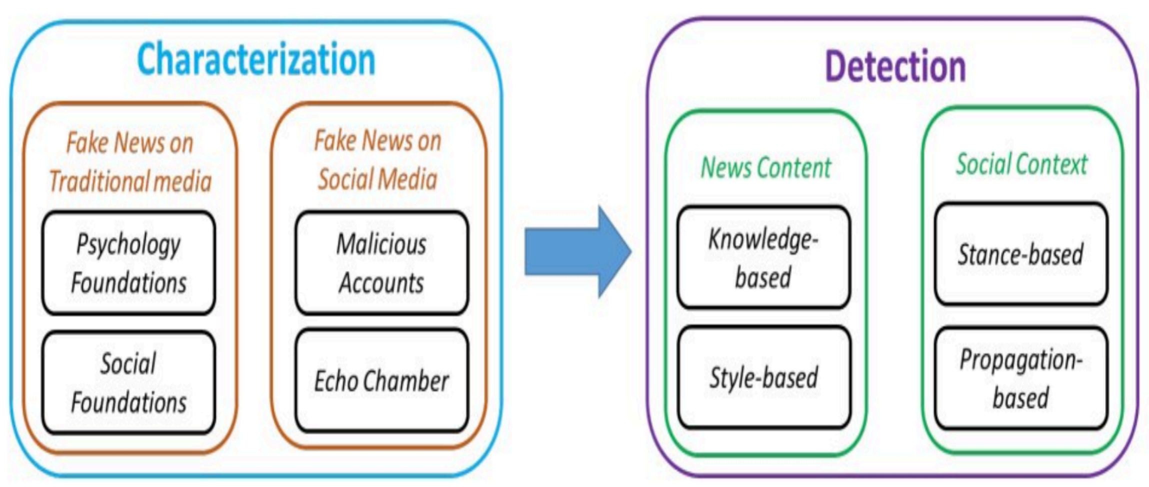
Sashank Santhanam and Samira Shaikh

Department of Computer Science

UNC CHARLOTTE

INTRODUCTION

- Social media platforms such as Twitter, Facebook provide a means for information to reach to the masses in near real time.
- Prior research has shown that misinformation in the form of propaganda, hoaxes, clickbait, rumors, satire and conspiracy theories spreads more rapidly than does accurate information.
- A recent poll conducted by Pew research center found that nearly 66% of users get news from Facebook, and 59% from Twitter.



OBJECTIVES

- In this work, we report on our research to combat the spread of misinformation from different news outlets on social media (e.g. Twitter).
- Use deep learning models to classify (binary and multi-class) tweets produced by a set of 83 news outlets on Twitter when compared to traditional machine learning approaches such as SVM and Naive Bayes.
- Identify linguistic features using LIWC that would help the model to distinguish between real news and other types of accounts and validate it with the help of ANOVA's.

RELATED WORK

The problem of misinformation has been gained significant traction over recent years and has been well studied in the field of journalism and computer science [2]. Misinformation is a broad term and consists of different types of news categories. Some of the identified types of news categories include propaganda, clickbait, hoax and satire which produce news similar to the content similar to real news but not providing the accurate facts or biasing the content in favor or against a person or organization.

Some of the approaches from a computer science perspective towards making attempts to solve misinformation included work done by Controy et al. who used n-gram plus syntax model along with incorporating profile information to classifying fake news[5]. Recently, Volkova et al. create a predictive model using neural networks to classify between the four types of misinformation outlets and also identify some linguistic features that help improve the performance of the model[3]. Other work done in this area, include the work done by Horne et al on three publicly available datasets to exploit some of the traits exhibited by misinformation news outlets when compared to real news[2].

DATA EXPLORATION

To create our dataset, we used the Twitter Streaming API to extract tweets from a set of 83 accounts annotated as propaganda, clickbait, hoax, satire and real news by external sources. The data collection period was between May 23, 2017 to June 6, 2017. These 83 accounts were shortlisted based on the conditions:

1. Account creation date
2. Low friends to follower ratio

Table 1 - Distribution of different types of news outlets.

Account Type	Number
Real News	31
Propaganda	30
Clickbait	18
Hoax	2
Satire	2

METHOD

Method - 1: Classification

Baseline Models: We used Support Vector Machines with linear kernel and Naive Bayes models with 10 fold cross validation to classify (binary) tweets from the real news account or the misinformation accounts. To train these models we divided it into 85% for the training data and 15% for the testing data

RNN + LSTM: We built a Recurrent Neural Network (RNN) with LSTM to perform a binary classification between real and misinformation and also a used the same model to perform multi-class classification between the different types of news outlets. We initialized the embedding layer of the RNN with pre-trained GloVe embedding[7]. The input sequences were padded to a uniform length of 200. Table 2. represents the hyper parameters used in building the model.

Table 2. Hyperparameters

Hyperparameters	Values
Batch size	500
Learning Rate	0.001
Epochs	10
Dropout	0.5
Optimizer	Adam

Method - 2: Linguistic Analysis

To get a better understanding of the language used by real news when compared to misinformation outlets. We identify several features with the help of LIWC dictionary[8] and categorized into three categories:

1. Stylistic
2. Linguistic
3. Psychological

RESULTS

Method – 1: Classification

Table 3 represents the accuracy of the baseline and RNN + LSTM models for binary classification of tweets from real and misinformation accounts. RNN + LSTM model significantly outperforms the the baseline models.

Table 3 – Binary classification results

Method	Accuracy (%)
Naïve Bayes	41.5
SVM	65
RNN + LSTM	85.4

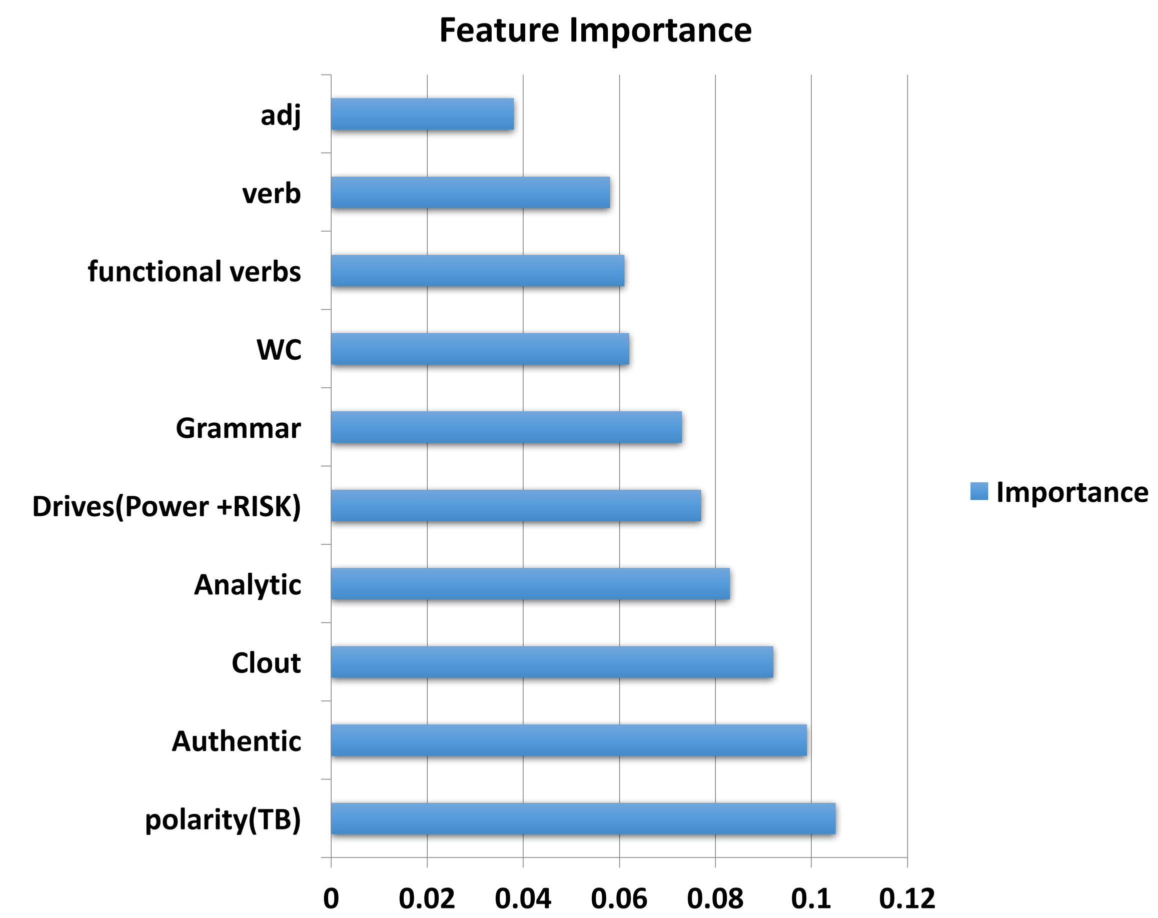
Table 4 represents the performance of the model on multi-class classification and compares the performance to an existing model.

Table 4- Multi-class classification results

Method	Accuracy (%)
Existing Method	63
RNN + LSTM	75.8

Method – 2: Linguistic Analysis

Feature	Significance
Polarity	P<0.0001 ***
Authentic	P<0.0001 ***
Clout	P<0.0001 ***
Analytic	P=0.085 *
Drive (Power + Risk)	P<0.0001 ***
Grammar	P<0.0001 ***
WC	P<0.0001 ***
Functional verbs	P<0.0001 ***
Verb	P<0.0001 ***
Adj	P<0.0001 ***

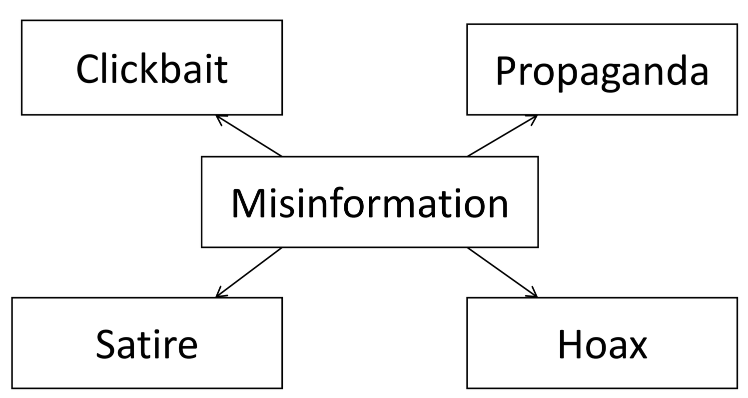


CONCLUSION & FUTURE WORK

- This preliminary work shows that misinformation can be detected on social media.
- Deep Learning performs better compared to traditional machine learning algorithms.
- This work facilitates understanding of the linguistic style present in tweets from misinformation accounts.
- In future work, we will incorporate linguistic features into the deep learning model and evaluate the classification performance. The long term goal of this project is to include features from images and combine it with the RNN model to maximize performance

REFERENCES

1. https://www.kdnuggets.com/2017/10/guide-fake-news-detection-social-media.html.
2. Horne, B.D. and Adali, S., 2017. This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. *arXiv preprint arXiv:1703.09398*.
3. Volkova, S., Shaffer, K., Jang, J.Y. and Hodas, N., 2017. Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (Vol. 2, pp. 647-653).
4. Conroy, N.J., Rubin, V.L. and Chen, Y., 2015. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology, 52(1), pp. 1-4.*
6. Zhiguo Wang, Wael Hamza, and Radu Florian. Bilateral multi-perspective matching for natural language sentences. 2017.
7. Pennington, J., Socher, R. and Manning, C., 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*(pp. 1532-1543).
8. Pennebaker, J.W., Francis, M.E. and Booth, R.J., 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, *71*(2001), p.2001.


Clickbait
Propaganda
Misinformation
Satire
Hoax