Stance Detection in the Context of Fake News

Izzat Alsmadi[1], Iyad Alazzam[2], and Mohammad Al-Ramahi[3]

[1,3] Texas A&M-San Antonio, San Antonio, USA

[2]Yarmouk University, Jordan

Izzat.Alsmadi@tamusa.edu[1], eyadh@yu.edu.jo[2], Mohammad.Abdel@tamusa.edu[3]

Abstract. Online Social Networks (OSNs) are overwhelmed by the daily volume of news reported by humans from all around the world. The diffusion of information can start from any OSNs user and spread rapidly. The process of news fact-checking is very labor and resource intensive. Researchers are looking for machine-learning approaches to automate the detection of fake news. Toward such goal, this paper focused on stance detection of content producers, whether they are in favor or against the content subject. In this study, our goal is to develop and evaluate advanced text-mining models that are enhanced with the cosine similarity between the headline and body of news articles to predict users' stance from the articles' content. We specifically aim to explore if the cosine distance feature will enhance the models' prediction performance.

Keywords: Stance detection, fake news, data analytics, cosine similarity.

1 Introduction

The power of OSNs enables all users around the world to be content producers rather than consumers in the early Internet. Users can create content in response to news events, personal lives, evaluating products, etc. The content generated by users may include a combination of events, facts, and opinions. Users' opinions reflect their personal perspectives, sentiments, cultures, and backgrounds. But how can we extract from the tremendous volumes of contents generated by humans daily, facts versus opinions? Do people intend to change facts or create misinformation or just look for and spread what they want to hear, assuming it is true with or without verification? This is particularly true when users receive such information from their networks, close friends, or people that share similar opinions or backgrounds. Each user may forward messages from their friends where the original source may not be known or verified by anyone.

Fact-checking websites such as Snopes.com and FactCheck.org track and check popular claims. Users evaluate claims and try to label them based on manual investigations. Researchers in machine-learning evaluate methods to label such labels using machine learning classifiers and learners as an alternative. Those classifiers start with building their learning algorithms based on previously known claims. As contents created by users will embed their personal preferences and opinions, researchers evaluate correlations between users' sentiments or stances and misinformation. Sentiments and stances can be confused with each other where both reflect connections between user emotions or opinions and the subject of the content. Users' stances can be broader in context than their sentiments. Several factors can impact their opinions toward some news or information in addition to sentiments.

Stance detection is profound and has been researched extremely in Natural Language Processing (NPL).

It is the process of deciding and concluding the audience and listener polarity from the text whether the audience preference is positive, negative, or none about the object. Stance classification, stance prediction, stance identification, debate stance classification, and debate-side classification are interconnected with stance detection and fall all on the same research area (Walker et al. 2012; Zhang et al. 2018; Qiu et al. 2015; Anand et al. 2011; Hasan and Ng 2013).

Stance detection has been applied in many different areas; fake news identification is one of these areas. Many researchers have worked on fake news identification concentrating on certain topic stance prediction wherein the text stance is related to a specific target or entity (Mohammad et al. 2016). Prior research results show that traditional methods of text processing proved effective. However, the key seemed to be in comparing the features from the headline to the features from the body. Intuitively, this makes sense as the target (stance) is a result of how the body relates to the headline. As a result, in this research, our primary objective is to develop and evaluate text-mining models that are enhanced with the cosine similarity between the headline and body of news articles to predict user stance. We aim to explore if the cosine distance feature will enhance the models' prediction performance.

2 Method

2.1 Dataset

Training and test datasets are provided on the FNC website. An overview of the data is provided on the FNC website and shown in Figures 1 and 2.

> **Input**
>
> A headline and a body text - either from the same news article or from two different articles.
>
> **Output**
>
> Classify the stance of the body text relative to the claim made in the headline into one of four categories:
>
> 1. **Agrees**: The body text agrees with the headline.
> 2. **Disagrees**: The body text disagrees with the headline.
> 3. **Discusses**: The body text discuss the same topic as the headline, but does not take a position
> 4. **Unrelated**: The body text discusses a different topic than the headline

Fig. 1: Provided data description

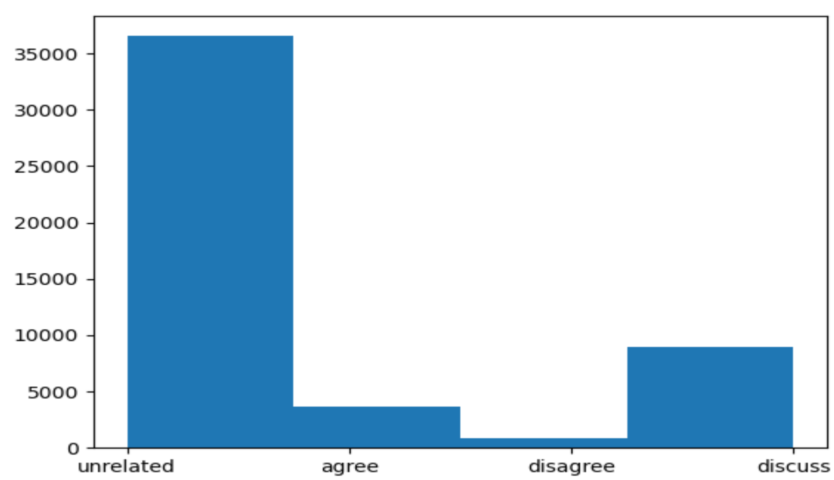

Fig. 2: distribution of the outcome variable

This data is available on Github in the following form:

- train_bodies.csv
- train_stances.csv
- compeition_test_bodies.csv
- compeition_test_stances.csv

The bodies and stances are associated with each other using a foreign key called body ID included in the body and stance files. The training data includes 49,972 news articles, 36,545 of which are unrelated. Figure 3 shows a breakdown of the training data by stance.

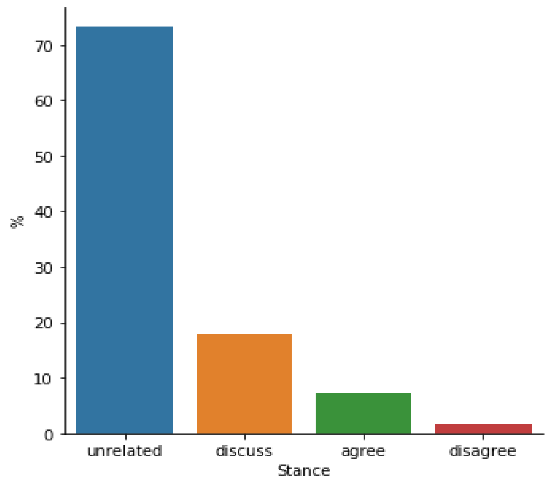

Fig. 3: Training data by the stance

All the training data may be used to train models since separate test data with 25,414 articles is provided.

2.2 Feature selection & engineering

Text processing

The only two features in the provided dataset are text features. As such, text processing is an essential component of the task. Research of previous analyses shows that traditional methods of text processing proved effective. However, the key seemed to be in comparing the features from the headline to the features from the body. Intuitively, this makes sense as the target (stance) is a result of how the body relates to the headline.

To keep the scope of the project focused, only one method of text processing was used instead of many. The top two teams used a plethora of text processing methods such as word counts, TF-IDF frequencies, word embeddings, and sentiment analysis. This project focuses on using TF-IDF frequencies to extract features from the article headlines and bodies. The TfidfVectorizer from Scikit-learn was used to obtain the TF-IDF frequencies. The vectorizer is initialized with a corpus comprised of all the headline and body text. Then it identifies the best 1000 1-grams, 2-grams, 3-grams, and 4-grams in the corpus and vectorizes the headline and body text.

Headline/body similarity

As the headline and body's stance that are similar are more likely to agree or discuss, Cosine distance was used to assess the similarity between the headline and body TF-IDF frequencies. Scikit-learn provides a function called pairwise_distances, which calculates each headline/ body row pair's value in the dataset.

2.3 Pipeline

Preprocessing the data results in a dataset with 2001 features; the 1000 TF-IDFs from the headline, the 1000 TF-IDFs from the body, and the cosine similarity between those two vectors. Figure 4 illustrates the preprocessing pipeline.

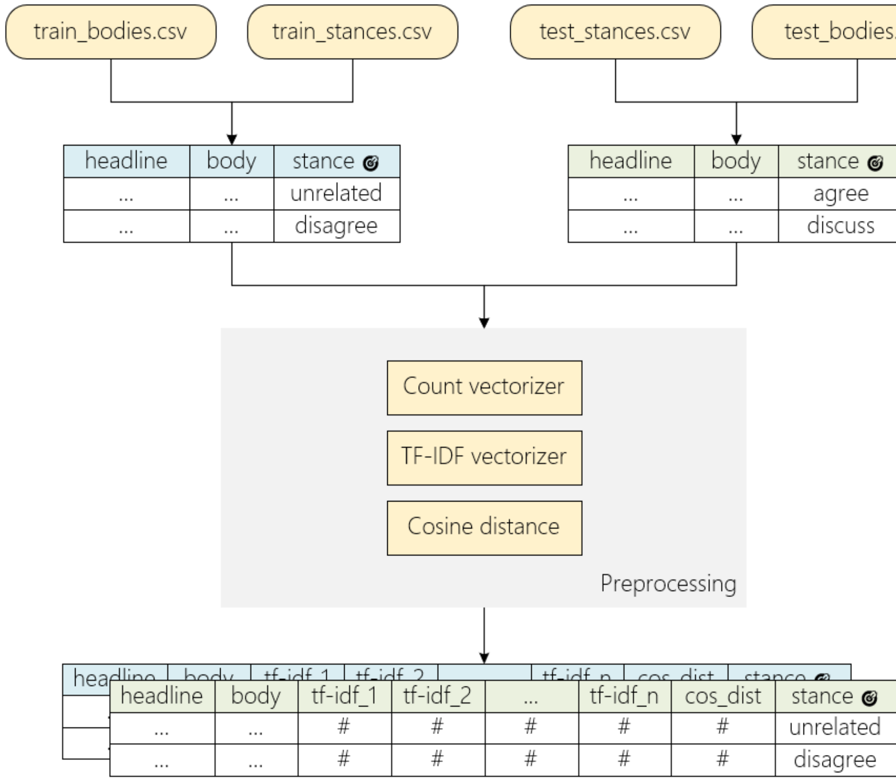

Fig. 4: Preprocessing pipeline

As expected, the cosine similarity proved to be the best predictor of an article's stance. Fig. 5 shows the feature correlation of each of the target stances. The TF-IDFs are named to indicate the origin of the feature ('h' for headline and 'b' for body) and the n-gram. The format of the TF-IDF names is as follows: <origin>_[<n-gram>]. The cosine similarity is the best predictor in all but the 'disagree' stance.

2.4 Training process

Training methods assessed include support vector machine classifiers, naïve Bayes classifiers, and ensemble methods such as random forest and gradient boosting classifiers. Scikit-learn makes it very easy to experiment with different models since they all implement the same methods. Fig. 6 shows the scores and training times of nine classifiers which were assessed for performance against the competition test data. The score shown is a simple percentage of correctly predicted targets.

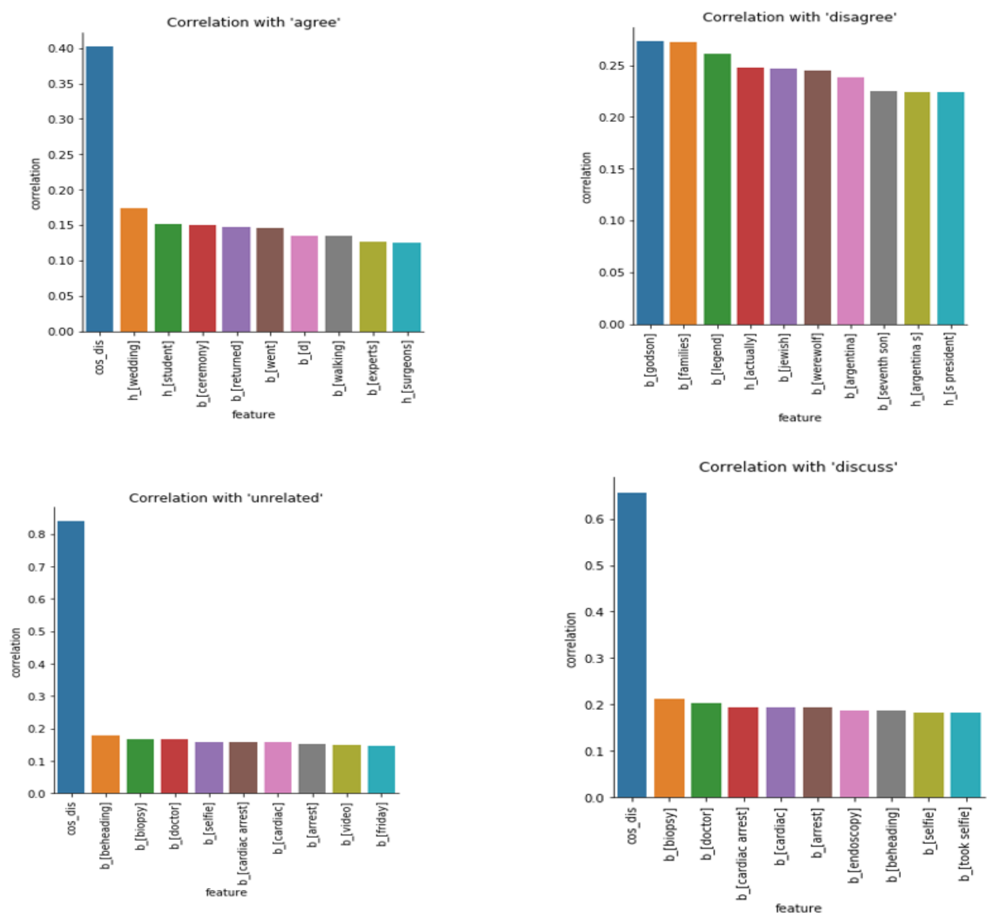

Fig. 5: Feature correlation

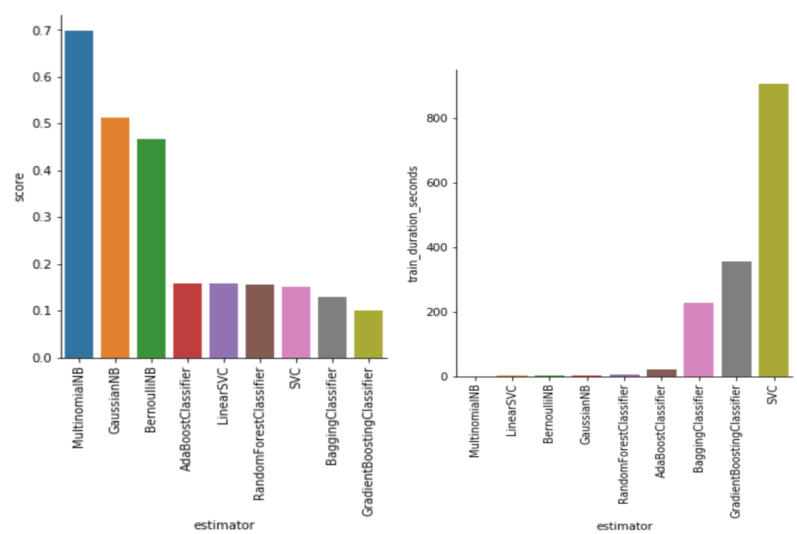

Fig. 6: Classifier scores and training durations

Multinomial naïve Bayes was chosen as the best classifier, although the scores were very poor in general. This is thought to include only one feature that compares the article headline to the article body. Taken on their own, features from the headline and the body do not seem to be good indicators of the article's stance.

The classifier chosen also happens to be the fastest to train, although this was not a factor in its selection. Figure 7 shows the training durations of the assessed classifiers. The long preprocessing step makes it difficult to experiment with different preprocessing methods because it makes the feedback loop very long. The impact of long preprocessing was mitigated somewhat by persisting preprocessed data to intermediate CSVs. With those files in place, it was possible to experiment quickly with different models and model parameters.

3 Conclusion and Future Work

In this paper, we examined the problem of detecting user stance in the context of news articles as a case study of detecting user stance in online social websites. We establish advanced text mining models using n-gram features and enhanced with the cosine distance between the headline and body of the news article. Our findings showed that the cosine similarity is the best predictor of an article's stance. A potential enhancement of this work is using other text feature extraction methods like word counts, word embeddings, and sentiment analysis. More advanced techniques like Singular Value Decomposition (SVD) and Latent Semantic Analysis (LSA) could be also employed. Another way to improve the performance of the analysis is to consider other measures of comparison between the article headlines and bodies. Using one feature, the cosine similarity seems to be insufficient. Some winning teams also used metrics like overlap to infer the relationship between the headlines and bodies.

REFERENCES

Anand, P., Walker, M., Abbott, R., Tree, J. E. F., Bowmani, R., & Minor, M. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011), 2011* (pp. 1-9)

Hasan, K. S., & Ng, V. Stance classification of ideological debates: Data, models, features, and constraints. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing, 2013* (pp. 1348-1356)

Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., & Cherry, C. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 2016* (pp. 31-41)

Qiu, M., Sim, Y., Smith, N. A., & Jiang, J. Modeling user arguments, interactions, and attributes for stance prediction in online debate forums. In *Proceedings of the 2015 SIAM international conference on data mining, 2015* (pp. 855-863): SIAM

Walker, M., Anand, P., Abbott, R., & Grant, R. Stance classification using dialogic properties of persuasion. In *Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies, 2012* (pp. 592-596)

Zhang, Q., Yilmaz, E., & Liang, S. Ranking-based method for news stance detection. In *Companion Proceedings of the The Web Conference 2018, 2018* (pp. 41-42)