Retrieval of Actionable Information during Mass Emergency: A Classification Approach

Aayush Chowdhury[1], Aishwarya Roy[2] and Aman Choudhary[3]

[1]University of Engineering and Management,Kolkata University Area, Plot No. III-B/5, Main Arterial Road (East-West), New town, Action Area-III, Kolkata, West Bengal - 700156

[2]University of Engineering and Management,Kolkata University Area, Plot No. III-B/5, Main Arterial Road (East-West), New town, Action Area-III, Kolkata, West Bengal - 700156

[3]University of Engineering and Management,Kolkata University Area, Plot No. III-B/5, Main Arterial Road (East-West), New town, Action Area-III, Kolkata, West Bengal - 700156

Abstract

This paper discusses our work submitted to FIRE 2021 IRMiDis Track. The goal was identification of claim or fact-checkable tweets, i.e., tweets that report some verifiable fact or claim.The two tasks addressed in this work are, first, Identifying claims or fact-checkable tweets and second,COVID Vaccine Stance Classification.The evaluation scores of the submitted runs are reported in terms of Precision@100, Recall@1000 and MAP@100. The average MAP score is 0.1587. The score for Vaccine Stance Classification is reported in terms of accuracy and macro-F1-score which came out to be 0.472 and 0.461 respectively.

Keywords

Fact-checkable, Non-fact-checkable, Information Retrieval, Machine Learning, Microblogs, ProVax, Antivax, Neutral, Stemming, Vectorization, automatic, classification.

1. Introduction

In the present digital world, the increasing use and popularity of various social media platforms has helped people to connect worldwide quickly and efficiently. Microblogging sites like Twitter are increasingly being used for aiding relief operations during various disasters. A lot of critical situational information is posted on microblogging sites during disaster events. Messages posted on microblogging sites, however, usually contain rumors and exaggerated information. It is extremely important for effective coordination of post-disaster relief operations to identify claim or fact-checkable tweets, i.e., tweets that report some verifiable fact or claim (rather than sympathy, prayers, or opinions) [1].

COVID-19 can spread quickly and widely. It has resulted in the deaths of over 1.9 million people worldwide.In the current scenario where the covid pandemic is so much prevalant, vaccines are only a solution if we want to normalize things.When enough people in the community are vaccinated, it slows down the spread of disease.So it is of utter importance to find out the

CEUR-WS.org/Vol-3159/T8-6.pdf

Forum for Information Retrieval Evaluation, December 13-17, 2021, India

aayushchowdhury.official@gmail.com (A. Chowdhury); royaishwarya693@gmail.com (A. Roy); amanchoudhary0109@gmail.com (A. Choudhary)

© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

natural sentiments of people towards the vaccine and try to eradicate and defy all misleading facts against the vaccine. [2] [3]

2. Tasks

- `Task 1:` **Identifying claims or fact-checkable tweets**
  This work primarily discusses our automated approach to develop automatic methodologies for identifying claim or fact-checkable tweets. This is mainly a classification problem, where tweets are classified into two classes – fact-checkable (claim) tweets and non-fact-checkable tweets.
  Any tweet that specifies the need or requirement of resources is termed as a need-tweet. This category also comprises the tweets which do not directly specify the need, but point non-availability of some resources. Below are the samples of a need-tweet and an availability-tweet identified from the provided dataset.

- Students of Himalayan Komang Hostel are praying for all beings who lost their life after earthquake!!!]|:
- ibnlive:Nepal earthquake: Tribhuvan International Airport bans landing of big aircraft [url]|:
  In this case, Tweet 1 is a non-fact-checkable tweet while tweet 2 is a fact-checkable tweet.

- `Task 2:` **COVID Vaccine Stance Classification**
  Here,we build an effective classifier for 3-class classification on tweets regarding their stance towards COVID-19 vaccines. The 3 classes are described below:
  `1. AntiVax` - the tweet is against the use of vaccines.
  `2. ProVax`- the tweet supports / promotes the use of vaccines.
  `3. Neutral` - the tweet does not have any discernible sentiment expressed towards vaccines or is not related to vaccines

3. Datasets

In the first phase of the track, we got two training datasets, one containing the fact checkable tweets and the other containing non fact checkable tweets, to train our model. Both the datasets consist of a tweet ID column and a tweet column. The total number of tweets in the training dataset was 299 – 149 fact checkable and 150 non fact checkable tweets. The training dataset was a balanced one. In the second phase, we got an unlabeled test dataset containing 11195 total tweets.
We did not use any other data source apart from the ones mentioned above.
For our second task the entire dataset provided to us contains 4392 tweets posted on Twitter during Covid-19. The tweets in the provided collection were all written in English. The dataset was divided into two phases:
1) Training dataset: It has three columns id, tweet, and label. There is a total of 2792 tweets in the training dataset labeled as either Provax, Antivax, or Neutral. There are 1010 Neutral, 991

ProVav, and 791 AntiVax labeled tweets.
2) Test dataset: It has two columns id and tweet consisting of 1600 tweets that were not labeled.
We have not used any other data resources apart from the above-mentioned ones for classification

4. Methodology

We discuss the overall design and implementation of our approach in this section. We borrow the advantages and capabilities of machine learning algorithms to implement the two above-mentioned tasks which are properly discussed in the consequent subsections

4.1. Task 1

Identifying claims or fact-checkable tweets:
In this task, we are looking for claims or fact checks that can be done on microblogging sites like Twitter to eliminate rumors or exaggerated statements. The overall process can be divided into three parts: Preprocessing, Vectorization and Model Selection

4.1.1. Preprocessing:

This phase involves cleaning up of the provided tweets labeled as Fact Checkable or Non Fact Checkable in the training dataset. All the words starting with hashtags or containing multiple spaces, special characters, URLs, punctuations or stopwords are firstly trimmed from every tweet. The URL's such as https://t.co/TMB8kyNO4D, usernames such as @sagarikaghose, etc are removed since they appear in majority of tweets and are not of much help for training the classifier model. In the next step we tokenize the words and apply wordnet lemmatization on each tweet which converts words into its root form thus decreasing noise and increase the speed and accuracy.

4.1.2. Vectorization:

The lemmatized tweets are then transformed into a TF-IDF matrix (Term Frequency - Inverse Document Frequency) using TF-IDF Vectorizer to determine how many times an item appears in each tweet. The word that appears frequently in a tweet has greater relevance to that tweet, meaning that there is a higher probability that the tweet is about or concerning that specific word. Each word in our dataset is assigned a TF-IDF score. This score increases when the word reappears in a tweet but decreases when the word reappears in another tweet. Two parameters are passed during vectorization –
a) min-df (used to remove terms that appear infrequently) - Set to 5, it means that if a term appears in less than five tweets, it is ignored.
b) max-df (used to remove terms with excessive frequency) - This is set to 0.4, which indicates that terms that occur in more than 40 percent of the tweets will be ignored.
TF-IDF matrix is then converted to an array and stored in the independent variable. Meanwhile, the dependent variable contains an array of labels – Fact checkable and Non fact checkable

tweets encoded as 0 and 1 respectively using LabelEncoder (like [0, 0, 1, .,... , 1, 0, 1]).

4.1.3. Model selection:

After preprocessing and vectorization, we feed our results into the classifier which encodes the tweets of test data to 0 or 1 based on its predictions after learning from the training model. We experimented with three classifiers SVM, Naïve Bayes and Logistic Regression. The Gaussian Naïve Bayes algorithm predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class. SVM with RBF kernel has higher speed and better performance with a limited number of samples. Logistic regression helps us estimate a probability of falling into a certain level of the categorical response given a set of predictors. The SVM (RBF kernel) gave the best recall and precision value. Finally, we mapped the predicted classes to their meaningful labels (0 - Fact Checkable, 1 - Non Fact Checkable) in our dataset and saved the results. The predictions which were predicted as Fact Checkable were ranked according to their confidence scores.

4.2. Task 2: Covid vaccine stance classification

This task is concerned with building an effective classifier for 3-class classification on tweets regarding their sentiment towards COVID-19 vaccines, the three classes being: • AntiVax-The tweets that are against the use of vaccines. • ProVax -The tweets that supports/promotes the use of vaccines. • Neutral –The tweets that not have any discriminable sentiment expressed towards vaccines or are not related to vaccines. The overall process can be divided into three parts: Preprocessing, Vectorization and Model Selection

4.2.1. Preprocessing:

This phase involves cleaning up of the provided tweets labeled as AntiVax, Provax or Neutral in the training dataset.All the words starting with hashtags or containing multiple spaces, special characters,urls,punctuations or stopwords are firstly trimmed from every tweet.The URL's such as https://t.co/WJXen4MfTV, usernames such as @DYGalvezINQ etc are removed since they appear in majority of tweets and are not of much help for training the classifier model. In the next step we tokenize the words and apply wordnet lemmatization on each tweet which converts words into its root form thus decreasing noise and increase the speed and accuracy.

4.2.2. Vectorization:

The lemmatized tweets are then transformed into a TF-IDF matrix (Term Frequency - Inverse Document Frequency) using TF-IDF Vectorizer to determine how many times an item appears in each tweet. The word that appears frequently in a tweet has greater relevance to that tweet, meaning that there is a higher probability that the tweet is about or concerning that specific word. Each word in our dataset is assigned a TF-IDF score. This score increases when the word reappears in a tweet but decreases when the word reappears in another tweet. Two parameters

are passed during vectorization –
a) min-df (used to remove terms that appear infrequently) - Set to 8, it means that if a term appears in less than eight tweets, it is ignored.
b) max-df (used to remove terms with excessive frequency) - This is set to 0.4, which indicates that terms that occur in more than 40 percent of the tweets will be ignored.
TF-IDF matrix is then converted to an array and stored in the independent variable.
Meanwhile, the dependent variable contains an array of labels - AntiVax, Neutral, and Provax encoded as 0, 1, and 2 respectively using LabelEncoder (like [0, 0, 1, .,… , 2, 2, 0]).

4.2.3. MODEL SELECTION:

After preprocessing and vectorization, we feed our results into the classifier which encodes the tweets of test data to 0,1 or 2 based on its predictions after learning from the training model. We experimented with three classifiers SVM, Naïve Bayes and Random Forest Classifier. The Multinomial Naïve Bayes algorithm predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class. SVM with linear kernel has higher speed and better performance with a limited number of samples.This algorithm creates a line or a hyperplane which separates data into classes. The Random Forest Classifier (RFC) from sklearn. ensemble model with n-estimators=70 (HyperTuning parameter). The "forest" it builds, is an ensemble of decision trees, usually trained with the “bagging” method. The general idea of the bagging method is that a combination of learning models increases the overall result. The Naïve Bayes gave the best accuracy and macro F1 score Finally we mapped the predicted classes to their meaningful labels(0-ProVax,1-AntiVax,2-Neutral) in our dataset and saved the results.

5. Evaluation:

The initial gold-standard for the task was generated using human annotation. A pooling mechanism was applied, i.e., by manually checking the top-ranked tweets of every model(as commonly done in TREC tracks). Standard IR measures such as Precision, Recall, MAP, and F-score were used to evaluate the runs. Higher credit was given to runs that identified more number of claim / fact-checkable tweets For task 2,the run submissions are evaluated against the gold standards. Measures like accuracy and macro-F1-score are used for the evaluation of runs.

Table 1
Table for task 1

Model	Recall
SVM(RBF)	0.98
NB	0.9328
Logistic Reg	0.9131

Table 2
Table for task 2

Run Submission	Accuracy	macro-F1 Score
1	0.472	0.461
2	0.455	0.446
3	0.406	0.405
4	0.413	0.402

6. Conclusion

In this work submitted to IRMiDis Track, we used Natural Language Processing for the processing of the tweets. And used Machine Learning models to identify the tweets in the training dataset that were against or in support of the Covid-19 vaccine or if they were neutral. We used the NeatText package of NLP for cleaning our tweets. Then techniques like Tokenization and Lemmatization were used to pre-process the tweets. Later, we used a few Machine Learning models to train our dataset regarding their stance towards the Covid-19 vaccine among which the Naive Bayes model gave the highest F-score.

References

[1] D. Barnwal, S. Ghelani, R. Krishna, M. Basu, S. Ghosh, Identifying fact-checkable microblogs during disasters: a classification-ranking approach, in: Proceedings of the 20th International Conference on Distributed Computing and Networking, 2019, pp. 389–392.

[2] L.-A. Cotfas, C. Delcea, I. Roxin, C. Ioanăş, D. S. Gherai, F. Tajariol, The longest month: Analyzing covid-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement, IEEE Access 9 (2021) 33203–33223.

[3] M. M. Müller, M. Salathé, Crowdbreaks: tracking health trends using public social media data and crowdsourcing, Frontiers in public health 7 (2019) 81.