Group 4

Shubham Agrawal , Ayush Agarwal, Janak Kapuriya, Akshara Nair

Problem Definition

Given a tweet, predict whether it is worth fact-checking by professional fact-checkers or not.This is a binary classification task.

A claim is an assertion that can have a truth value. So, a personal like/dislike isn’t even a claim. Further, of the valid claims, those which can negatively impact a large no. of people is of our interest, hence is a check-worthy claim. Our task is to classify the fact check-worthy claims.

Related Works

The initial works made use of classifiers such as SVMs, Decision Trees, Naive Bayes like by [1]. For feature extracion, the first attempts were made by using count based methods like TF-IDF, POS tags, sentiment scores etc. Additionally, works such as [2] added features such as average embedding vectors of a sentence, linguistic features, position awareness amongst sentences. Classifiers such as Deep Feed Forward Neural Networks were used.

[3] used RNN based network where tokens were represented from a mix of embeddings, POS tags and one hot encoding of syntactic dependencies. [4] used n-gram features and used a KNN classifier.

Recently, there has been a trend of using transformer models, like [5][6][7] . Since its inception, Transformer models[8] have been the goto technique for Natural Language tasks such as summarization, sequence classification, named entity recognition, text generation, language modeling, extractive question answering etc.

Further, a large number of models were derived from the transformer architecture like the BERT[9] which produces bidirectional encodings by eliminating the decoder part of the transformer.

Methodology

Ref: Check square at CheckThat! 2020: Claim Detection in Social Media via Fusion of Transformer and Syntactic Features Gullal S. Cheema 1 , Sherzod Hakimov 1 , and Ralph Ewerth 1,2

1. **Preprocessing**
   We use publicly available preprocessing tools from nltk library and some regex expressions. We apply the following normalization steps: tokenization, lower-casing, removal of punctuation, hashtags, all-caps, and URLs. We also removed stopwords since past experiments (Gullal et.al) showed reduction in score, probably because stop-words can result in less meaningful sentence representation by dominating in the average.As the transformer networks have their own tokenization process, we use raw text to extract word embeddings from them.

2. **Syntactic Features**
   We use the following syntactic features for English: Parts-of-Speech (POS) tags and dependency parse tree relations. We didn’t take named entity (NE) extracted embeddings as they showed reduction in score based on past experiments.(Gullal et.al) We use the pre-processed text and run ready-made tools (spaCy) to extract syntactic information of tweets and then convert each group of information to feature sets.

2a. Part-of-Speech

We extracted 16 POS tags in total generated using spacy out of which we keep the following eight tags as our most useful features: NOUN, VERB, PROPN, ADJ, ADV, NUM, ADP, PRON.This chosen set of POS tags is used to encode the syntactic information of tweets.

2b.Syntactic Dependencies

Dependency relations between tokens in a given tweet are used to construct features. We use the dependency relation between two nodes in the parsed tree if the child and parent nodes’ POS tags are one of the following ADJ, ADV, NOUN, PROPN, VERB or NUM. All dependency relations that match the defined constraint are converted into the pairs such as (child node-POS, dependency-relation) . Also, for every tweet we consider only the most frequent 15 dependencies. Now, to encode a feature, we create a frequency vector which contains the number of type of tag, and syntactic relation pair.

This feature encoding process is shown in Figure 1.

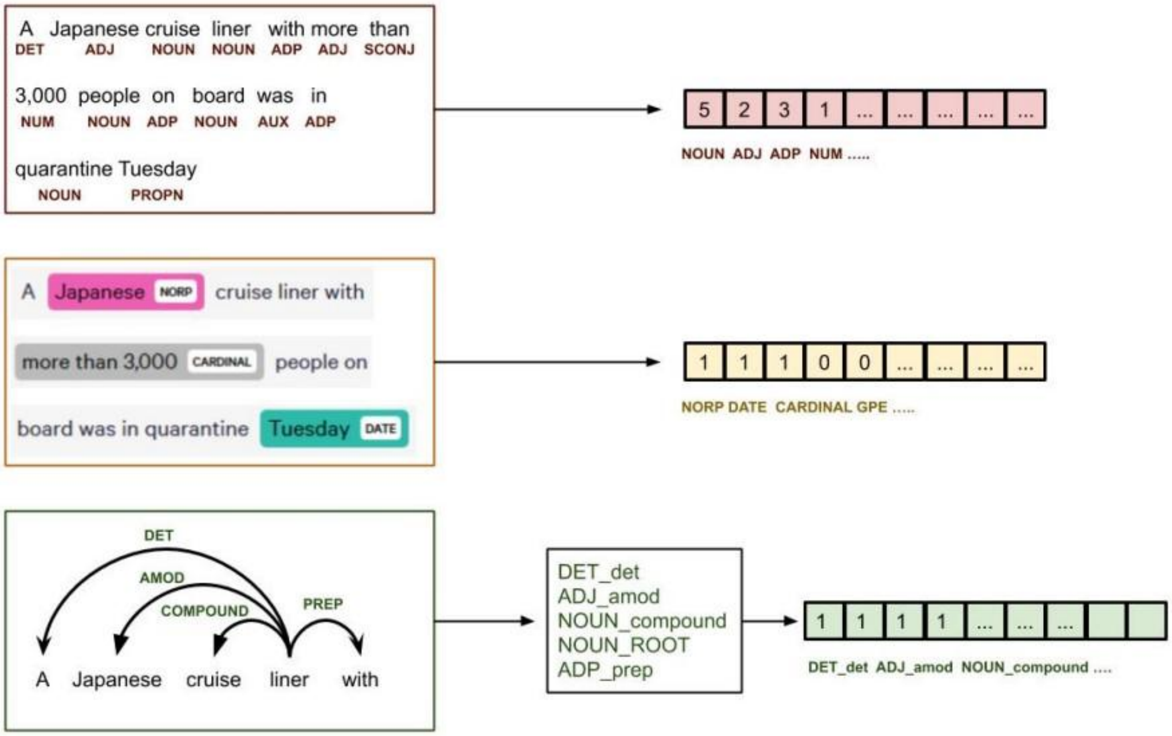

A Japanese cruise liner with more than
DET ADJ NOUN NOUN ADP ADJ SCONJ
3,000 people on board was in
NUM NOUN ADP NOUN AUX ADP
quarantine Tuesday
NOUN PROPN
5 2 3 1 ... ...
NOUN ADJ ADP NUM .....
A Japanese NORP cruise liner with
more than 3,000 CARDINAL people on
board was in quarantine Tuesday DATE
1 1 1 0 0 ... ...
NORP DATE CARDINAL GPE .....
DET
AMOD
COMPOUND
PREP
A Japanese cruise liner with
DET_det
ADJ_amod
NOUN_compound
NOUN_ROOT
ADP_prep
1 1 1 1 ... ...
DET_det ADJ_amod NOUN_compound ....

Fig. 1: Syntactic feature extraction and encoding process. Feature vectors are based on the number of times it is seen in the given sentence.

3. Contextual Word Embeddings

To extract contextual features we use all-MiniLM-L6-v2 embeddings that are trained using the context of the word in a sentence. all-MiniLM-L6-v2 is a sentence transformer model which maps sentences and paragraphs to 384 dimensional vector-spaced dense embeddings.It’s is usually trained on a very large text corpus which makes it very useful for ready-made feature extraction and fine-tuning for downstream tasks in NLP.

4.Concatenation, Dimensionality Reduction and Classification

To get the overall representation of the tweet, we concatenate all the syntactic features together with transformer-generated contextual features and then apply PCA for dimensionality reduction. SVM classifier is trained on the feature vectors of tweets to output a binary decision (check worthy or not). The overall process is shown in Figure 2.

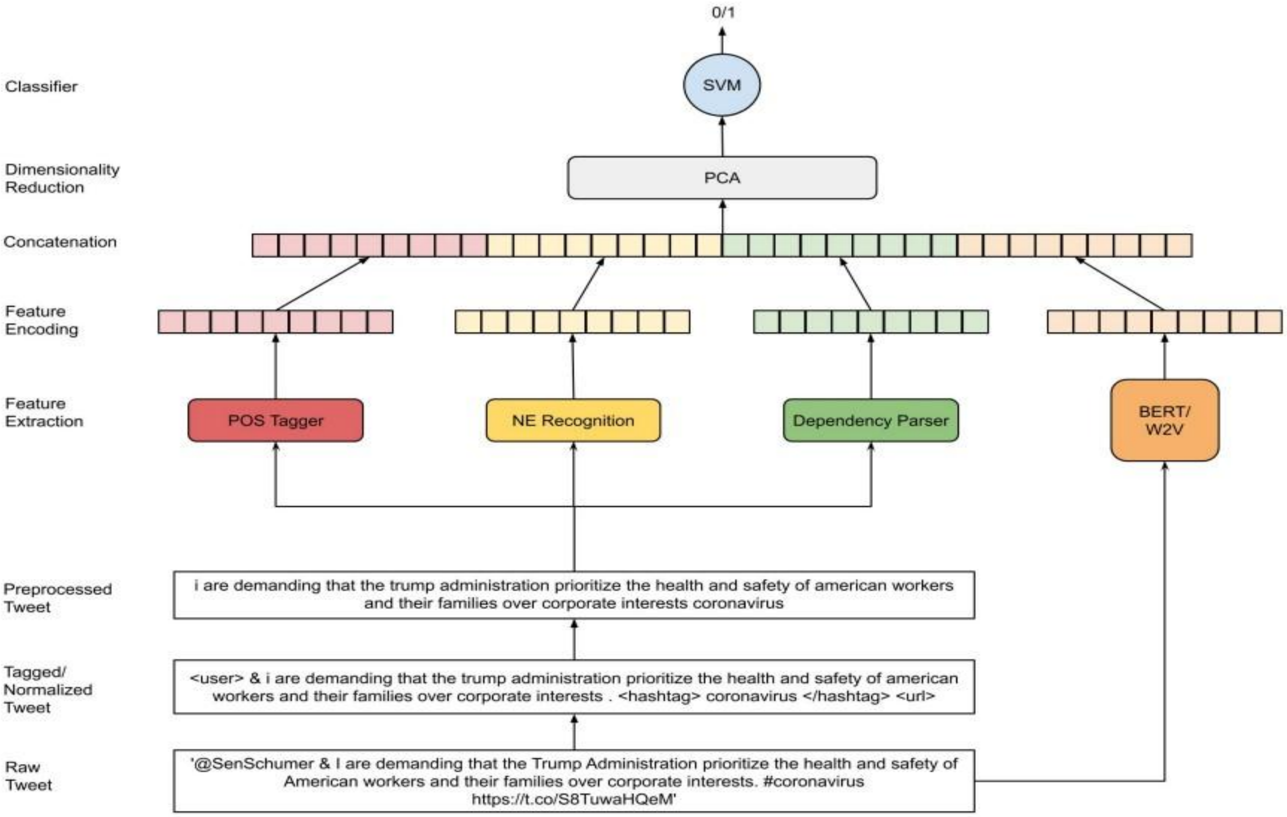
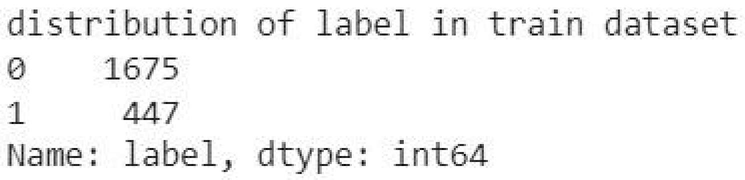

0/1
Classifier
SVM
Dimensionality Reduction
PCA
Concatenation
Feature Encoding
Feature Extraction
POS Tagger
NE Recognition
Dependency Parser
BERT/ W2V
Preprocessed Tweet
i are demanding that the trump administration prioritize the health and safety of american workers and their families over corporate interests coronavirus
Tagged/ Normalized Tweet
<user> & i are demanding that the trump administration prioritize the health and safety of american workers and their families over corporate interests . <hashtag> coronavirus </hashtag> <url>
Raw Tweet
'@SenSchumer & I are demanding that the Trump Administration prioritize the health and safety of American workers and their families over corporate interests. #coronavirus https://t.co/S8TuwaHQeM'

Fig. 2: Proposed Approach for Check-Worthiness Prediction

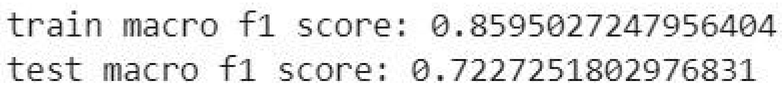
Experimental Results

Distributions of labels



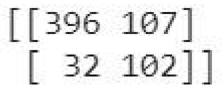
```
distribution of label in train dataset
0    1675
1     447
Name: label, dtype: int64
```

Macro F1 score results



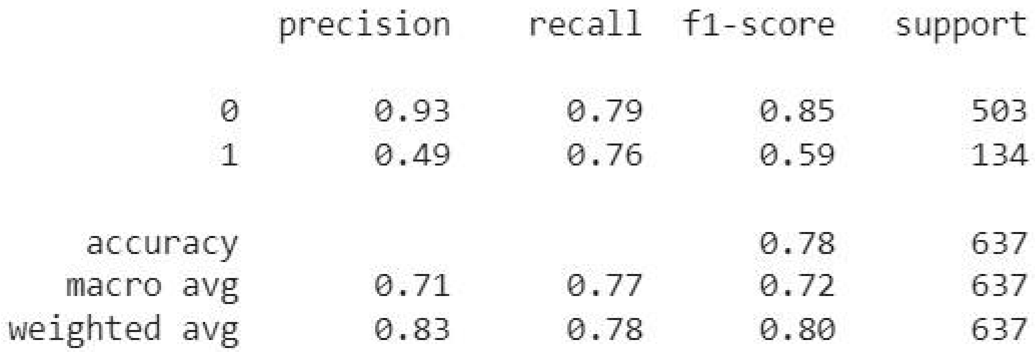
```
train macro f1 score: 0.8595027247956404
test macro f1 score: 0.7227251802976831
```

Confusion matrix of SVM



```
[[396 107]
 [ 32 102]]
```

Classification Report of Model



```
              precision    recall  f1-score   support

           0       0.93      0.79      0.85       503
           1       0.49      0.76      0.59       134

    accuracy                           0.78       637
   macro avg       0.71      0.77      0.72       637
weighted avg       0.83      0.78      0.80       637
```

Analysis

As we can see from the distributions of class labels that the dataset is imbalanced. So we have used weighted SVM on dataset which will give more weightage to the minority class during training.

Contribution

- Non contextual feature extraction(histogram vector storing counts of pos-tags and dependency parsers.- Akshara, Janak, Ayush.
- Contextual feature extraction(dense vector created using Sentence Transformer)- Shubham, Janak
- Research Methedology- Ayush, Akshara.
- PPT / Report: Akshara, Ayush, Janak, Shubham.

References

[1]N. Hassan, C. Li, M. Tremayne, Detecting check-worthy factual claims in presidential debates, in: Proceedings of the 24th acm international on conference on information and knowledge management, 2015, pp. 1835–1838.

[2]P. Gencheva, P. Nakov, L. Màrquez, A. Barrón-Cedeño, I. Koychev, A context-aware approach for detecting worth-checking claims in political debates, in: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, 2017, pp. 267–276.

[3]C. Hansen, C. Hansen, J. G. Simonsen, C. Lioma, The Copenhagen Team Participation in the Check-Worthiness Task of the Competition of Automatic Identification and Verification of Claims in Political Debates of the CLEF-2018 CheckThat! Lab (2018) 8.

[4]B. Ghanem, M. Montes-y Gomez, F. Rangel, P. Rosso, UPV-INAOE-Autoritas - Check That: Preliminary Approach for Checking Worthiness of Claims (2018) 6.

[5]E. Williams, P. Rodrigues, V. Novak, Accenture at CheckThat! 2020: If you say so: Post-hoc fact-checking of claims using transformer-based models, arXiv:2009.02431 [cs] (2020). URL: http://arxiv.org/abs/2009.02431, arXiv: 2009.02431.

[6]A. Nikolov, G. D. S. Martino, I. Koychev, P. Nakov, Team Alex at CLEF CheckThat! 2020: Identifying Check-Worthy Tweets With Transformer Models, arXiv:2009.02931 [cs] (2020). URL: http://arxiv.org/abs/2009.02931, arXiv: 2009.02931.

[7]G. S. Cheema, S. Hakimov, R. Ewerth, Check_square at CheckThat! 2020: Claim Detection in Social Media via Fusion of Transformer and Syntactic Features, arXiv:2007.10534 [cs] (2020). URL: http://arxiv.org/abs/2007.10534, arXiv: 2007.10534.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, \. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in neural information processing systems, 2017, pp. 5998–6008.

[9] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).