1. Install:

numpy, pandas, nltk, tqdm

2. In this work, we choose to build a Naive Bayes classifier to determine if an article is fake or real.

The Naive Bayes classifier is based on Bayes theorem (1). In real application, we usually drop off the denominator p(E) as for each candidate Hi, the denominator is the same.

(1)

Thus we only need to compute the prior probability P(C=i) and the condition probability P(Wj|C=i)=and P(NM|C=i)=. we smooth the model by adding a small value deta=0.5.

3. Experiments designed:

(1) Load the training and testing data set with pandas.read\_csv().

(2) Build the vocabulary for training text+title and author's name.

(3) Calculate the prior probability P(C=i),i {0,1} and conditional probability P(NM|C=i) and P(Wj|C=i).

(4) Using the function score[i]=P(C=i)\*P(NM|C=i)\*P(Wj|C=i) to evaluate each test article and calculate the score for each class.

(5) Assign the class with the highest score.

4.Evaluation metrics:

In order to evaluate the performance of our model, we take accuracy, precision, recall and F1-measure as evaluation metrics.

Accuracy is the percentage of instances in the test set that the model has correctly classified.

Recall is the proportion of the instances in a special class C are labelled correctly.

Precision is the proportion of instances predicted to be class C are actually correct.

F-measure is a weighted combination of precision and recall.

5. Methods for model improving:

(1) Stop-words filtering (Used): remove the commonly used stop-words that don't convey real meaning.

(2) Word length filtering (Used): remove words with length (<= 2 or >=9).

(3) infrequent and high-frequency word filtering (In the future):

First, gradually remove words with frequency=1, <=5,<=10,<=15 and <=20. get a new vocabulary;

Then, sort the vocabulary by word frequency, gradually remove top 5%, 10%, 15%, 20% and 25% most frequent words of the vocabulary.

(4) The impact of different smoothing value changing from 0 to 1 with a step of 0.1 (In the future).

6. Experiment Results:

The test results of the Naïve Bayes classifier are shown in Table 1.

**Table 1**. Performance of Naive Bayes Classifier with stop-words filtering and word length filtering

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Predicted result | The True label | | Precision | Recall | F1\_measure | Accuracy |
| real | fake |
| real | 1522 | 1233 | 0.552 | 0.651 | 0.598 | 0.606 |
| fake | 817 | 1628 | 0.666 | 0.569 | 0.614 |