To solve these problems, let’s come to the solution part. In this part, I’d like to talk about the related work and the proposed method.

Related work comes first.

The first section “Stock Forecasting Based on Deep Learning” illustrates that LSTM and CNN are the most typical deep learning models which have been commonly used in stock forecasting. For example, a hybrid model combined LSTM and Genetic Algorithm by Chuang, CEEMD-PCA-LSTM by Zhang’ s team, Hierachical Attention Networks by Zheng and so on.

The second section “Stock Forecasting with Social Media Sentiment Analysis” talks about enhancing stock movement prediction by public sentiment, such as news, articles, blog posts and other texts. It runs for 2 steps. Step 1, transform the text to an understandable form for machines. Step 2, utilize statistical models or machine learning models to make prediction.

Beyond the previous work, this article makes an innovation to utilize multiple news textual features for prediction.

Now let’s come to the proposed method. Firstly, let’s formulate the problem. Use 3 numbers to describe three-class classification, 1 for up, 0 for hold and -1 for down. Denote the stock prices sequences as XT and the multiple textual features as MT, then the stock price trend of the next day Y(T+1) can be written as this equation, in which F() is the mapping function from the input to the output, and K is the size of the sliding window.

Let multiple textual features MT consist of topic features MTT. Denote the sentiment features as MTS and the semantic features as MTW, we get this equation, in which this operation is to combine different textual features.

Now come to the proposed method which contains 3 stages. This is the diagram.

Stage 1 contains 2 steps. Step 1 is to use coherence score to evaluate the LDA model to determine the optimal number of topics. Step 2 is to represent each document as an n-dimension vector.

Stage 2 contains 3 steps. Step 1 is to use NLTK Valence Awareness Dictionary for Sentiment Reasoning to obtain compound score. Step 2 is to use pre-trained embeddings from a subset of the Google News, generate two 300-dimensional vectors. Step 3 is to decide the optimal size of sliding window K.

Stage 3 contains 2 steps. Step 1 is to compare the performance of different deep learning models. Step 2 is to adopt the most suitable model for proposed method.