**Phase I Report**

**Project Topic (Title)**

Variational Autoencoder-aided Multi-task Rumor Classifier with Sentiments Analysis

**Team member information**

Yuhang Lu yuhangl@usc.edu

Zechen Ha zechenha@usc.edu

Zehao Huang [zehaohua@usc.edu](mailto:zehaohua@usc.edu)

**Problem Description**

The widespread of rumors can cause various far-reaching consequences. In this way, we are trying to build an LSTM-based VRoC combined with sentiment analysis system to detect rumors. It is difficult for us to grasp the whole idea about the VRoC algorithm. And we are trying to add a new element into this architecture. Also, implementing this architecture in both Python and c++ is challenging.

**Project Timeline:**

For now, we have learnt basic ideas about LSTM, RNN, VRoC. Also, we have learnt how the rumor classifier works. And we have also designed a web spider to collect a dataset.

For phase 2, we will implement what the VRoC paper has achieved, experiment on the given

dataset and get similar accuracy as that of VRoC paper. We are trying to add sentiment analysis into the system. Then, we will write section 3 (our model description) as well as part of Section 4. Additionally, we will write a separate short progress report for phase II and add information the course staff and supervisors should be aware of (any change of plan, etc.)

For Phase III, our team will complete/revise the software implementation, complete Section 4

(experimental results), revise previous sections based on the results. We hope to improve the

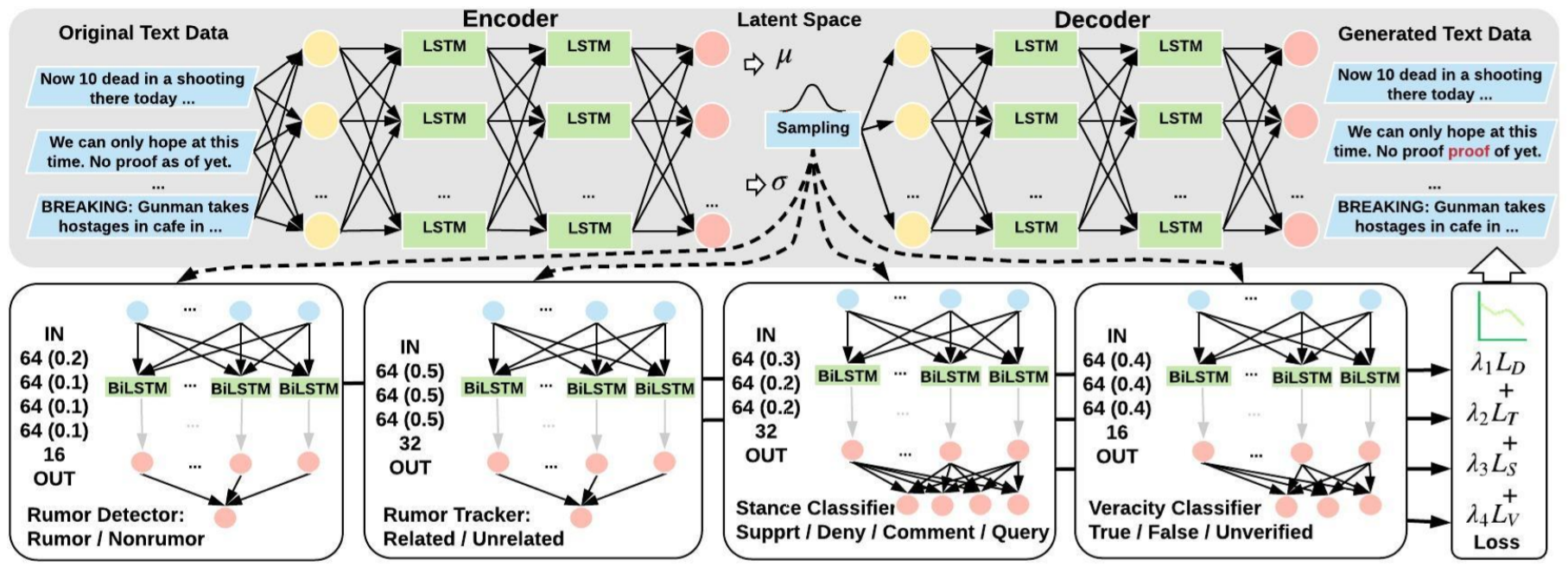
accuracy of the final rumor detection in this stage. We are also going to write a separate progress report for phase III and add information the course staff and supervisors should be aware of (any change of plan, etc.

**Analysis and Implementation**

In this task, we are going to add sentiment analysis on the basis of VRoC to improve the detection performance. The algorithms used in VRoC to implement rumor detection are mainly variational autoencoders (VAEs), which is based on recurrent neural networks(RNNs) with long short-term memory(LSTM).

The VAE in this system is LSTM-based, which plays the role of extracting latent representations of tweet-level text. The rumor classification system is made up of four independent parts: rumor detector, rumor tracker, stance classifier and veracity classifier.

Because Tensorflow is a package which integrates lots of practical functions and methods, it is a good choice to use it to implement the framework of our model.



To collect a dataset of our rumor detection project, a web crawler is used to get rumors and the corresponding labels. The web crawler, which is also called web spider, simulates the behavior of browser users to automatically grab target information on the internet. In other words, a web crawler is a program or automated script which browses the World Wide Web in a methodical, automated manner.

The USC Melady Lab investigates twitter coronavirus topics a lot in the recent months, which makes it a good source of collecting our dataset. We use the URL of USC Melady Lab to track its website. Often are the cases that all the information of a website will be loaded once a browser visits it. However, the webpage of the USC Melady Lab dynamically loads the covid-19 rumors by Asynchronous Javascript And XML(AJAX). Therefore, we use the package “selenium” instead of the package “require”, which is broadly used in the websites without AJAX. After downloading the raw data from Melady Lab, we separate them into three parts: rumor(only content of a rumor), label(which kind of rumor it is) and source(the source of the rumor).

**Summary:**

**Detection and Resolution of Rumours in Social Media: A Survey**

This paper shows us the definition of rumors that a rumor is “an item of circulating information whose veracity status is yet to be verified at the time of posting". Also, it tells us what the different annotations of rumors are, rumor veracity, stance toward rumors and rumor relevance. What’s more, it shows us a brief introduction of the architecture of a rumor classification system. Usually, rumor handling research efforts cast four main elements: rumor detection, rumor tracking, rumor stance classification, and rumor veracity classification.

**MVAE: Multimodal Variational Autoencoder for Fake News Detection**

This paper proposes a multimodal variational autoencoder that learns shared (visual + textual) representations to aid fake news detection. And the result shows us with the help of visual analysis, the performance is improved. It also gives us the confidence that when sentiment analysis comes into picture the performance would be improved as well

**Information Credibility on Twitter**

This paper shows us automatic methods for assessing the credibility of a given set of tweets which focused on features of users’ posting and reposting behaviors. However, this method mostly utilizes hand-crafted features. And this paper uses credibility in the sense of believability and ignores the veracity of rumors directly. This kind of evaluation method is a little bit subjective.

**Prominent Features of Rumor Propagation in Online Social Media**

This paper identifies characteristics of rumors by examining the three aspects of diffusion: temporal, structural and linguistic. And those three classifiers are built based on decision trees, random forest and SVM. Moreover, this paper is one of the first to analyze the underlying process of rumor propagation which is similar to rumor tracking.

**All-in-one: Multi-task Learning for Rumour Verification**

An LSTM-based multi-task learning approach that allows joint training of the main and auxiliary tasks is introduced in this paper. Rumor detection and stance classification are combined to improve the performance.

**Tree LSTMs with Convolution Units to Predict Stance and Rumor Veracity in Social Media Conversations**

The paper proposes to use convolution units in Tree LSTMs that are better at learning patterns in features obtained from the source and reply posts. It gives two hints. One is that tree LSTMs can have a better effect in showing the connection between sources and reply, which will improve the effect. Another is that a useful signal (stance in this paper) will lead to a better performance in rumor classification.

**Reply-Aided Detection of Misinformation via Bayesian Deep Learning**

The paper uses a Bayesian deep learning model to represent the uncertainty of the prediction to rumor veracity. The paper firstly encodes a claim to be verified, and generates a prior belief distribution from which we sample a latent variable, then encodes all the people’s replies to using LSTM to update the prior belief generating posterior belief. The paper uses Stochastic Gradient Variational bayes to get the iterations.

**Determining the veracity of rumors on Twitter**

The paper extracts over 80 trustworthiness features including the authors’ profile and past behavior, the social network connections, and the content of tweets themselves, and uses classical machine-learning methods to classify rumors and get a good performance.

**Automatic Detection and Verification of Rumors on Twitter**

To predict the veracity of rumors, the paper identifies salient features of rumors by examining three aspects of information spread: linguistic style used to express rumors, characteristics of people involved in propagating information, and network propagation dynamics. Then the paper uses a hidden markov model to predict the veracity of tweets.

**Predicting Stances in Twitter Conversations for Detecting Veracity of Rumors: a Neural Approach**

The paper’s solution comprises two key steps. Firstly, detect the stance of each individual tweet, by considering the textual content of the tweet, its timestamp, as well as the sequential conversation structure leading up to the target tweet. Then use the predicted stances of all tweets in a conversation tree to determine the veracity of the original rumor.

**Writing:**

**Abstraction**

Social media platforms such as Twitter have become a plethora of rumors and the negative impact has shown in the Great East Japan Earthquake, President Election and recently COVID-19. This leads to attention of the research on developing efficient methods to detect and verify potential rumors. A typical rumor classifier may contain a rumor detector, rumor tracker, stance classifier and veracity classifier. The paper will focus on veracity classifier combining a sentiment signal.

**Introduction:**

Despite the increasing use of social media platforms for information, its unmoderated nature often leads to the emergence and spread of rumors (items of information that are unverified). The widespread of rumors can cause various far-reaching consequences. And sentiments, especially strongly positive or negative ones have a great impact on veracity identification. Hence, it is essential to explore how to automatically assess their veracity, using natural language processing and data mining techniques.

**Key Novelties:**

Based on the previous work, we planned to add sentiment analysis into VRoC architecture.

And we are trying to implement the system in C++.