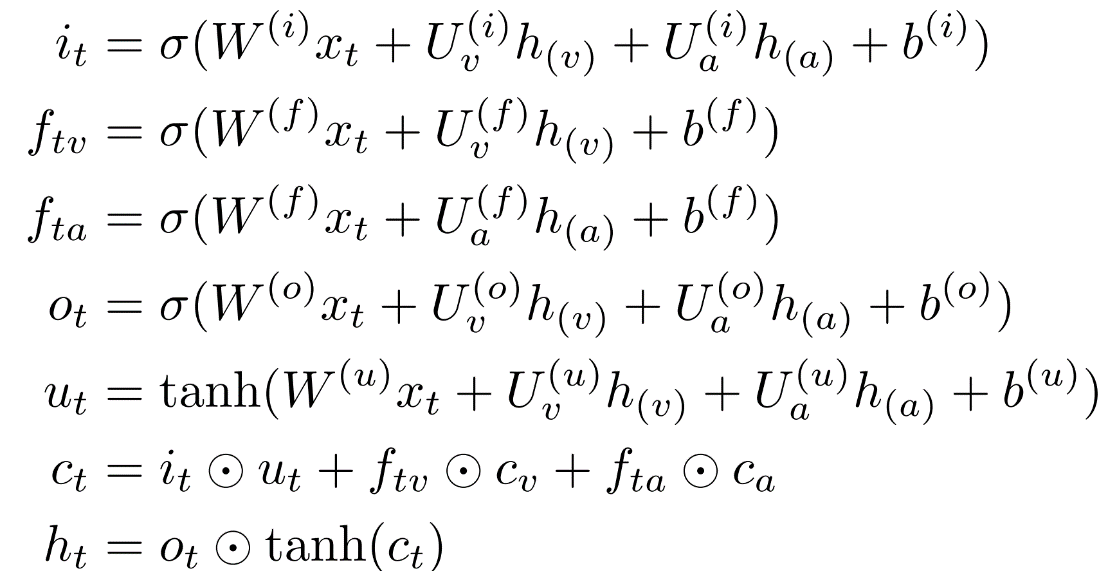
*Multimodality data*

Inspired by Tree-LSTM, we plan to implement a graph-based LSTM to model multimodality time series data. In this case, each mode is a node in Tree-LSTM model, and their hidden state and memory cell state are inter-exchanged during time series sequencing. For instance, to compute the next hidden state of text data, a combination of hidden states and memory cell state of video and audio data will be the current hidden state and memory cell state of text data node. We will also have multiple forget gates for those hidden states. The mathematical formulation of computing the next hidden state and memory cell state of text data is shown below. Subscripts indicate text, visual and audio respectively.



We use the most recent MOSI data released by CMU. As used in the Hackathon, the audio, visual and text features are extracted from COVAREP [1], FACET [2] and Glove [3] respectively. However, each data sequence now may have different lengths and they are aligned among different modes according to the spoken words. Our results are averaged over ten random experiments and are shown in the following table. We achieve results that are comparable to the results originally reported by [4].

|  |  |
| --- | --- |
| Tensor Fusion [4] | Graph LSTM |
| 42.0% | 42.19% |

*Lesson*

For the Duke data, our goal is to predict infection within 24-hour after the inoculation based on RNA-seq and physiological sensor data (electro-dermal activity, heart rate, skin temperature). We took both a spectral-analysis approach and a random-forest approach. The spectral-analysis approach consists of two steps: first take short-window Fourier transform (each of 1-hour length) with no overlap between each consecutive window and then use the dynamic time warping to measure the distances. These distances are used as features for classification. The random-forest approach is using data from the 48-hour before inoculation to train a random forest model for predicting the 24-hour data after inoculation; the distances between the true values and predicted values are used as features for classification. The random-forest based approach provides with about 80% accuracy and is stable with respect to different initializations of the algorithm.

There are two main lessons. First is that the amount of data in this project is relatively small and this makes the training of the random-forest a challenging task. So we have to make use of the periodicity to increase the training set. The second lesson is that there is no control group in the data, so we have to assume that the one does not show symptoms are not infected. This lead to some subjective rather than objective way to make decisions.

After participating a series of Deep-Purple hackathon, we realize that it is not straightforward to utilize deep learning models into other academic fields like biology. We believe that the main advantage of deep learning models compared to machine learning models are their trainable feature extraction modules. We have CNN for visual and word embedding for text data. These feature extraction methods are not only adaptable to a particular dataset but are also transferrable to other datasets in the same field. In biology, we usually have less amount of data and the feature extraction model for biological data remain unclear. It would be better to start from a large biological dataset first and then explore proper feature extraction methods which can be transferred to other biological datasets.

We also have some difficulties with the computation platform Xsede. Despite of the fact that we are issued GPU service units, the platform does not allow the users to install packages. For example, our lab uses Pytorch as our deep learning frameworks, but only Bridges support a very old version of it. We have to submit tickets to ask for them to install it. It would be better that the users can have their own Virtual Machine Image or at least grant some freedom on software installation.

*Reference*

[1] G. Degottex, J. Kane, T. Drugman, T. Raitio and S. Scherer, "COVAREP - A collaborative voice analysis repository for speech technologies", In Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy 2014.

[2] Ekman, Paul, and Erika L. Rosenberg, eds. What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA, 1997

[3] Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation

[4] Zadeh, Amir, et al. "Tensor fusion network for multimodal sentiment analysis." arXiv preprint arXiv:1707.07250 (2017).