1. Hi everyone, we are group 20 and I am the presenter today, my name is TAKEHIRO MATSUNAGA. My team members include EDUARDO WANG ZHENG, and KAR CHUN TEONG who is currently in Malaysia joining us online.
2. We attempted 3 different model architectures for this project, they are listed on this slide.
3. First, let’s look at DMRNets. This architecture is analysed by KAR CHUN TEONG and implemented by KAR CHUN TEONG and EDUARDO WANG ZHENG. DMRNets use CNN originally, but we use FCN instead.
4. To explain what is DMRNets, we need to start from Residual Net. ResNet originated from CNN. Researchers found out that stacking layers leads to a degradation problem, and this is not caused by overfitting. The more layers they stack, the higher the training error. Some proposed a solution which is residual mapping, as shown on the image. And This is how residual network works, just concatenate processed data and unprocessed data.
5. Next, we move on to the variation based on ResNet, the DMRNets(CNN). Basically what it does is to assemble the residual branches in parallel through a merge-and run mapping.
6. At last, we come to the model in our project, the DMRNets(with FCN).
7. Here’s part of the model structure. Actually, the only variation we did is to change the convolutional layers to fully connected layers.
8. For unknown reasons, the training of the DMRNets model did not converge, we tried modifying the model, but in the end it couldn't work.
9. Now we look at the second model we use, the LSTM model, this model is analysed and implemented by TAKEHIRO MATSUNAGA.
10. LSTM is derived from recurrent neural network . As our brain can understand articles based on knowledge that we learned before, RNN works in a similar way. It can derive some parameters from its forward node, this is what traditional neural networks could not achieve. However, traditional RNN could not store what it learned for a dragged time period because of its algorithm. Then LSTM was invented to solve this problem.
11. LSTM is a type of RNN. A common LSTM unit is composed of a cell which contains 3 gates. The cell stores values over time intervals as cell states, and the 3 gates’ main job is to regulate the flow of information into and out of the cell. To put it in an abstract high-level view, LSTM is using past context as reference to predict the future output.
12. This picture shows a cell of LSTM, the forget gate will selectively forget some information getting from forward cell. And the formula at right side determine which data should keep and which one should forget.
13. Performance.
14. The third model we use is bidirectional LSTM model (biLSTM). this model is Analysed and implemented by TAKEHIRO MATSUNAGA.
15. biLSTM is a variant of LSTM. LSTM uses past context only, while biLSTM uses both past and future context. biLSTM generally provided more context, which leads to better performance.The picture shows that there are two LSTM in two different direction and each LSTM will provide the future prediction for the other model. With both forward and backward infomation we could predict with more confidence.
16. This picture shows our model structure, actually its quite similar to the LSTM model, just change the layer from lstm to bilstm. Beside layer changing, I inspired from merging method in our first model. and contribute the model to bilstm + merging method.
17. Performance
18. Experiment and analysis, this part is Tested by TAKEHIRO MATSUNAGA and Analysed by KAR CHUN TEONG
19. We wanted to use cloud platforms to train our model, but most of them are expensive. Google Cloud Platform offers free trial, but to get it, we need to register with credit card, which unfortunately we don't possess, at the end of the day we need to train the model using our local machines. Limited by the performance of our local machines, we could only use a small subset of the datasets for training (the first 200 training shards and the first 100 validation shards).
20. This figure shows the graph of training curve of the LSTM model.
21. This figure shows the graph of training curve of the biLSTM model. As shown in both figures, both models converged at around 0.75, with biLSTM having a slightly better performance.
22. The performances of 2 models are listed in the table. we use the 300 data shards as training datasets which leads to possible overfitting. However, we consider that the LSTM and biLSTM's performance of > 0.75 for the whole datasets as satisfactory, since we only use an extremely small subset to train them. We hypothesize that if we use larger datasets for training, the model would lead to more satisfactory results, too bad that we don't have the computing power to prove this hypothesis.
23. Contributions. (read the ppt or just let them read by saying “the contributions of each member are listed here”)
24. conclusion, Based on the experiments, we conclude that the biLSTM model has a superior performance over the other models we used. Therefore we use the biLSTM model in our final submission.
25. Thank you!