RESEARCH STAY WEEK 6, LSTM and GRUs

Daniel Cajas A01708637

CONTEXT

One problem with ANNs is the fixed input size and lack of temporal awareness. With RNNs we can solve that by adding a feedback signal into the network, allowing us to do multiple forward propagations keeping some state in between them (the feedback signal).

SEARCH METHODOLOGY

Last week i accidentally got to LSTM and RNNs when researching the application of ANNs in sentiment analysis. This week wanted to follow up last weeks comparison with one extra paper that i found interesting. I found it while searching more specifically for RNNs in sentiment analysis instead of just ANNS.

Preliminary terms

- ANN: Artificial Neural Network
- CNN: Convolutional Neural Network
- RNN: Recurrent Neural Network
- LSTM: Long Short Term Memory
- BiLSTM: Bidirectional Long Short Term Memory
- GRU: Gated Recurrent Unit

COMPARISON

Similarly to last week's papers. This week I found a model with CNNs and LSTM or BiLSTM for sentiment analysis. However in this case it isn't used in the classification task or as two parallel feature extraction layers. This model uses the CNNs output and feeds it to the LSTM (after a max pooling layer). This is a more straightforward approach and seems to give some interesting conclusions. I do need to note that the results are not directly comparable as the datasets and word embeddings used were not the same.

COMPARISON

Hybrid Features

Model	SST1 Random initialized word vectors	SST1 Pre-trained word vectors	SST2 Random initialized word vectors	SST2 Pre-trained word vectors
CNN-RNN+KCCA	47.62	49.42	86.05	88.73
CNN-RNN+SVM2K	46.54	48.3	84.35	87.45
CNN-RNN+MVMED	50.25	52.09	89.12	90.32
CNN-RNN+SMVMED	50.89	52.94	90.25	91.93

TABLE 8. Percentage accuracies of all variations of the proposed model

CNN fed to LSTM

Table 4. Model comparaison.

Word Embedding	Model	Accurary
Doc2vec	CNN	88.00%
	LSTM	85.87%
	BiLSTM	86.40%
	CNN-LSTM	90.13%
	CNN-BiLSTM	90.66%

RNN classifier

CNN-RNN	CNN-GRU-word2vec	**82.28**	50.68	**89.95**
	CNN-LSTM-word2vec	81.52	**51.50**	89.56
	AVG-GRU-word2vec	81.44	50.36	89.61
	CNN-GRU-rand	76.34	48.27	86.64
	CNN-LSTM-rand	77.04	49.50	86.80

As we can see, feeding the CNN output to the BiLSTM got similar results as the Hybrid features approach although the hybrid one did get an edge. We can conclude from these 3 papers that RNNs (specifically LSTM) are best used for feature extraction.

BIBLIOGRAFÍA

[1] Xingyou Wang, Weijie Jiang, and Zhiyong Luo. 2016. Combination of Convolutional and Recurrent Neural Network for Sentiment Analysis of Short Texts. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 2428–2437, Osaka, Japan. The COLING 2016 Organizing Committee.

[2] H. Sadr, M. M. Pedram and M. Teshnehlab, "Multi-View Deep Network: A Deep Model Based on Learning Features From Heterogeneous Neural Networks for Sentiment Analysis," in IEEE Access, vol. 8, pp. 86984-86997, 2020, doi: 10.1109/ACCESS.2020.2992063.
keywords: {Feature extraction;Sentiment analysis;Deep learning;Data mining;Recurrent neural networks;Convolutional neural networks;Deep learning;multi-view learning;convolutional neural network;recursive neural network;sentiment analysis}

[3] Rhanoui, M., Mikram, M., Yousfi, S., & Barzali, S. (2019). A CNN-BiLSTM Model for Document-Level Sentiment Analysis. Machine Learning and Knowledge Extraction, 1(3), 832–847. doi:10.3390/make1030048