Using the Big Picture – A Deep and Distant Approach to Classification of Suicidal Ideation in Autism Spectrum Disorder.

Xingyi Song[1], Johnny Downs[2,3], Sumithra Velupillai[2,4], Rachel Holden[2,5], Maxim Kikoler[2,5], Rina Dutta[2,3], Angus Roberts[2]

[1]Department of Computer Science, The University of Sheffield, Sheffield, UK
[2]NIHR Biomedical Research Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[3]South London and Maudsley NHS Foundation Trust, London, UK
[4]School of Electrical Engineering and Computer Science, KTH, Sweden
[5]University of Canterbury, Southborough, UK

Abstract

In order to understand complex language in the text of the Electronic Health Record (EHR), it is sometimes necessary to consider contextual information from other parts of the text. We introduce a novel neural network architecture to classify sentences from EHRs, taking in to account such distant contextual information. We test our method on detection of suicide-related information in documents from Autism Spectrum Disorder patients. Our approach improves accuracy over the state of the art.

Introduction

Classification of sentences (and more generally of any chunk or mention) is an important component of electronic health record (EHR) text analytics, and therefore for the reuse of EHRs for research. For example, classification of suicide-related mentions in psychiatric disorders may be used to create retrospective clinical cohorts for epidemiological research, and potentially to aid risk prediction in clinical settings [5]. Most previous classification approaches [1,2,3,4,5] consider only the sentence itself, and context from distantly surrounding text is ignored. However, this surrounding context may supply additional information to aid classification, with preceding and following sentences adding weight or nuance to a point. For example, consider these imaginary consecutive sentences:

1. *Her parents thought she felt life not worth living, although in clinic she denied ever having been suicidal and said she had never self-harmed.*

2. *She denied any current plans or thoughts of self-harm.*

Sentence 1 contains several negated statements, conjunctions, and discusses people other than the patient, making it difficult for a classifier to reach the correct decision. The classifier could use sentence 2 to provide additional context to correctly classify sentence 1 as negative for suicidal thoughts.

In this paper, we address this problem with a novel neural network architecture – Context LSTM-CNN (Figure 1) that is not only based on the sentence, but also takes the full context from the rest of the document into consideration, without significantly increasing the computational cost. Context LSTM-CNN first applies a bi-directional Long short-term memory (LSTM) [6] over sentence word embeddings to capture the sentence level relationship between the words, and passes these embeddings into a CNN [1] network for classification. At the same time, the left and right contexts of the sentence are encoded into fixed-size codes with a Fixed-Size Ordinally-Forgetting Encoding (FOFE) [7]. Finally we feed the left and right context FOFE codes together with the CNN output into a final output layer. The computational cost of adding the small, fixed size FOFE code to the network is negligible.
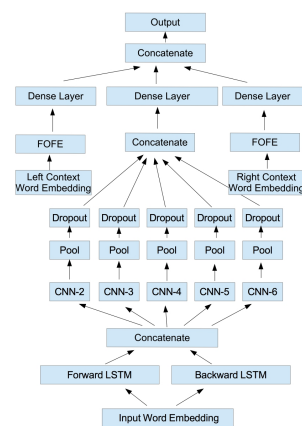
Figure 1. Architecture of Context LSTM-CNN

Preliminary Experiment

We conducted a preliminary experiment on the corpus reported in [5]. This corpus of 4916 documents from Autism Spectrum Disorder patients' records, labels all mentions of relevance to suicidal behaviour, in one of three categories: 'positive for suicidality', 'negated for suicidality' and 'suicidality uncertain'. The labeled mentions are mostly of sentence size or close to sentence size, so we treat this as a sentence classification task, with all other unlabeled sentences as 'not relevant', giving four classes. We compared our approach against two state of the art neural net baselines: LSTM and CNN. We randomly held out 20% of the document as a test set, and used the remainder for training. We repeated this holdout process five times, and report the average F-measure for each class in Table 1.

Results and Conclusions

The results show that Context LSTM-CNN improves F1 by up to 10 to 20% when compared to the baselines. In addition, the Context LSTM-CNN training and application time is almost identical to LSTM.

Table 1. F-1 score of classifiers over ASD suicidal ideation sentences

	Not relevant	Suicidality Pos	Suicidality Neg	Uncertain
CNN	98.27	58.66	66.70	35.23
LSTM	98.31	59.31	65.32	29.26
Context LSTM-CNN	**99.22**	**74.00**	**78.15**	**48.43**

References

1. Kim Y. Convolutional neural networks for sentence classification. In Proceedings of the 2014 Empirical Methods in Natural Language Processing, 2014.
2. Collobert R, Weston R, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. Journal of Machine Learning Research, 2011.
3. Zhou C, Sun C, Liu Z, Lau F. A c-lstm neural network for text classification. arXiv:151108630Z, 2015.
4. Huynh T, He Y, Willis A, Rueger S. Adverse drug reaction classification with deep neural networks. In Proceedings of the 26th International Conference on Computational Linguistics, 2016.
5. Downs J, Velupillai S, Gkotsis G, Holden R, Kikoler M, Dean H, Fernandes A, Dutta R. Detection of suicidality in adolescents with autism spectrum disorders: Developing a natural language processing approach for use in electronic health records. AMIA Annual Symposium Proceedings, 2016.
6. Hochreiter S and Schmidhuber J. Long short-term memory. Neural Computation, 1997.
7. Zhang S, Jiang H, Xu M, Hou J, Dai L. The fixed-size ordinally forgetting encoding method for neural network language models. In proceedings of 2015 annual meeting of computational linguistics, 2015