

CNN Hyperparameters. We set the number of feature maps for each convolution width to 100 and used dropout with a probability of 0.5 to prevent overfitting [1]. All parameters were initialized using a uniform distribution from -0.05 to 0.05 . The model was trained with adadelata [2]. After every parameter update, the parameters of the feature maps were normalized to a norm of 3.

Word Embedding Hyperparameters. We used word2vec to train the word embeddings on all the notes of MIMIC III v3. We used the continuous bag of words (CBOW) model, set the window size to 10, sampled 10 negative examples for each positive example, discarded words that appear less than 5 times, and performed 15 training iterations.

cTAKES Hyperparameters. We used cTAKES 3.2.2 with the default FilesInDirectoryCollectionReader.xml.

References

1. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*. 2014;15(1):1929–1958.
2. Zeiler MD. ADADELTA: an adaptive learning rate method. *arXiv preprint arXiv:12125701*. 2012;.