

# Fusion: Towards Automated ICD Coding via Feature Compression

## 1 Data: MIMIC-III

Medical Information Mart for Intensive Care (MIMIC III) (Johnson et al., 2016) is used for our experiment. The database includes many types of information like demographics, vital sign measurements, laboratory results, notes, reports, procedures, and medication codes. It's a free open data set. Following the setting of (Mullenbach et al., 2018), we used the discharge summaries as the text information. Following is a short peach of a discharge summary.

1. *In the ED, patient was saturating 96% on NRB. CXR did not reveal any consolidation. Per report EKG was unremarkable. Laboratory evaluation revealed a leukocytosis if 14 and lactate of 2.2. Patient received combivent nebs, solumedrol 125 mg IV x1, aspirin 325 mg po x1 ...*

Under the full codes setting, all codes that exist on MIMIC-III data set are required to predict. Under the 50 setting, the model is only required to predict the top 50 codes with the highest frequency. Totally there are 8,921 ICD-9 codes with 47,719, 1,631, and 3,372 discharge summaries for training, validation, and testing.

## 2 Experiment Environment Details

All experiments are run with a Tesla V100 GPU with Intel Xeon Gold 6242 CPU under Linux system, Pytorch framework.

## 3 Hyper Parameters

The training epoch is set to 30 with a learning rate of  $1e-4$ . Early stop procedure is used by checking Precision@8 score on the validation set. Pre-trained word embeddings are used and tuned during the training. The multi-filter sizes are set to

3,5,9,15,19,25. The filter dim is set to 50. Please refer to the submitted codes.

## References

- Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1):1–9.
- James Mullenbach, Sarah Wiegrefe, Jon Duke, Jimeng Sun, and Jacob Eisenstein. 2018. Explainable prediction of medical codes from clinical text. In *NAACL-HLT*.