Medical Examination Data Prediction with Missing Information Using Long Short-Term Memory

Han-Gyu Kim[1], Gil-Jin Jang[2], Ho-Jin Choi[1],
Minho Kim[3], Young-Won Kim[3], Jaehun Choi[3]
[1] School of Computing, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea
[2] School of Electronics Engineering, Kyungpook National University, Daegu 41566, South Korea
[3] Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea

kimhangyu@kaist.ac.kr
gjang@knu.ac.kr
hojinc@kaist.ac.kr
kimmh@etri.re.kr
everywkim@etri.re.kr
jhchoi@etri.re.kr

In this work, we use recurrent neural network (RNN) to predict the medical examination data with missing parts. There often exist missing parts in medical examination data due to various human factors, for instance, because human subjects occasionally miss their annual examinations. Such missing parts make it hard to predict the future examination data by machines. Thus, imputation of the missing information is needed for accurate prediction of medical examination data. Among various types of RNNs, we choose long short-term memory (LSTM) to predict the missing information as well as the future medical examination data, as LSTM shows good performance in many relevant applications. In our proposed method, the temporal trajectories of the medical examination measurements are modelled by LSTM with the missed measurements compensated, which is then used to predict the future measurements to be used as diagnosing the diseases of the subjects in advance. We have carried out experiments using a medical examination database of Korean people for 12 consecutive years with 13 medical fields. In this database, 11500 people took the medical check-up every year, and 7400 people missed their examination occasionally. We use complete data to train LSTM, and the data with missing parts are used to evaluate the imputation and future measurement prediction performance. In terms of root mean squared error (RMSE) between the prediction and the actual measurements, the experimental results show that the proposed LSTM network predicts medical examination data much better than the conventional linear regression in most of the examination items.