CS273B Paper Review 2

DeepCare: A deep dynamic memory model for predictive medicine

Abhimanyu Banerjee, Raunaq Rewari, Wen Torng, Junhong Choi, Joy Xiang

Background

Electronic medical records (EMRs) is a compilation of patient illness diagnosis and treatment intervention processes with irregular and varying time intervals. Developing better informatics tools to analyze EMR data could significantly improve healthcare for patients such as providing better understanding of patient disease progression and treatment outcome, reducing preventable errors, improving communication among health care providers and facilities, and controlling the cost of medical care. However, existing dynamic models to predict patient disease trajectories do not sufficiently take into consideration the irregular timings between diagnosis and intervention. Hence, the authors of DeepCare aim to address these shortcomings of current models to improve predictive medicine.

Summary

Pham et al has developed a modified Long Short-Term Memory recurrent neural network to predict prognosis from electronic medical records of diabetic patients, which was shown to outperform support vector machines and random forests baselines. To achieve this, DeepCare 1) incorporated intervention events as part of the output and forget gates of the LSTM to model its targeted effect on disease progression, and 2) modeled the time irregularity between events with both a time decay function and a parametric time model. 3) Further, to focus the attention of the model to the more recent records, the model does multiscale pooling of the illness states from the LSTM using three look-backs of 12 months, 24 months and all available history. By incorporating invention and time interval information, DeepCare was shown to be better than an SVM model by F-score 12.4% and a Random forest model by F-score 7.7%. It also outperforms a simple LSTM by F-score of 3.6%, which suggests modest improvement but highlights the importance of developing better models to integrate such time based information in predicting unplanned readmission.

Critiques

One of authors’ main hypotheses for developing DeepCare is that intervention at a patient’s admission affects disease progression and is included in the output and forget gates of the LSTM. Intuitively, this implies that certain intervention measures could significantly impact patient outcome in the future, whereas an inappropriate intervention following a diagnosis would have a different effect. However, it is unclear what the independent benefit of including this variable is on the performance of the model, as the authors only report the performance of DeepCare including, alongside intervention, either of the two time models and multiscale pooling. When the model switched from a time decay to a parametric time model, performance improved, which shows evidence that time difference matters, but does not shed any light on the importance of intervention. One could guess that if this parameter is not important on its own, the insight is either that for diabetes treatment, none of the interventions affected outcome, or that there should be a better way to integration this information in the model. The former conclusion, if true, would invite a reexamination of the impact of the treatment measures on patient outcome. Assuming evidence-based medicine had justified the appropriate interventions, the latter conclusion may be more likely and it is unclear if the authors could better incorporate intervention information, perhaps also in the input gate or with a different function than sigmoid. Another point to be noted is that the previous intervention supposedly only affects the forget gate, but we argue that past interventions should also show up in the output gate, as the previous intervention very closely affects what we prescribe as intervention in the current time point and also the overall outcome.

Similarly, the independent advantage of the multiscale pooling layer is unknown and would require further evidence from the authors that it substantially improves the model. The authors could also demonstrate how varying the type of multiscale pools might result in better or worse predictions e.g. whether 12 month, 24 month and all history (12 years in the case of the data that the paper uses) is the best way to partition the different pools.

The authors have converted the medical codes into vectors, but fail to provide the logic they used to do this. The point of using vectors is both to be able to do back propagation on them as well as derive some semantic meaning i.e. codes that are related to each other should occur close to each other in vector space.

The authors also use a measure, m, to indicate planned versus unplanned admissions, giving a score of up to 1 or 1/2. This value impacts the input gate in the LSTM and seems to signify that unplanned admissions are twice as important than planned ones. It would be better if a few choices of m are described and an effect of choosing various different m is shown.

The authors have used multi-scale pooling, and the logic behind this decision does make sense. However, an alternate way that should have been tried is to set a threshold (which would be a hyper-parameter), above which all hidden states are average pooled. The rationale may be that certain points in the entire history are important for predicting the next outcome.

DeepCare explored two different mechanisms to capture time irregularity, resulting in limited improvement over LSTMs that do not capture this information. The first is a direct scaling of the forget gate by an exponential decay function of time difference between two recorded inputs i.e. admissions. This model captures the shape of the 40 channels of forgetting as shown in Fig 4 (left). The other model is a parametric model of the first, second and third power of the time difference between admissions. However, it is unclear why parameterizing up to the power of 3 of the time difference improved the F-score by 2% as the authors rightfully pointed out that it seems to be overparameterizing.

For predicting diabetic patient disease progression, DeepCare considered only EMR data from one Australian hospital and simplified the nature of the data as illness or intervention. The strength of predictions could be improved by incorporating Electronic Health Record (EHR) information, which contains more comprehensive information about a patient history from different providers, in addition to EMR. Conceivably, the dataset would be richer in information and charts a more complete history of each patient, which could be highly relevant to chronic diseases such as diabetes. However, such datasets could be much greater in complexity and DeepCare may be limited in its simple delineation of data as illness and intervention.

Conclusion

To conclude, Pham et al. has modified an existing LSTM architecture to fit the problem of predicting the disease development from the personal medical records. This is a necessary and timely attempt to utilize a large amount of existing data set to aid healthcare system and improve a disease prognosis process via deep learning. However, the improvement is modest at best compared to simple LSTM, and the mechanism in which an aspect of DeepCare contributed best to prediction remains unclear. Specifically, effects of medical intervention, multi-scale pooling and different prioritization of admission type needs to be explored further, if these are indeed the source of improvements for the DeepCare. Given the task of predicting patient's' health is complex with many variables, more complex model encompassing all aspect of disease – patients' individual health record, effectiveness of each medical intervention, genetic propensity – may need to be considered in order to capture the full disease progression.