DeepCare: A Deep Dynamic Memory Model for Predictive Medicine

Abstract

The field of predictive medicine studies the probability of diseases and appropriate preventative measures. Given a model of disease progression professionals can gauge when their patients will be diagnosed with a disease or if their patient will have an unplannned readmission. Since predictive diagnosis depends on understanding a patients past medical history, it makes sense to employ contemporary Deep Learning techniques to build such a model. In this regard, DeepCare is the next step for Deep Learning techniques in predictive medicine.

1 Introduction

DeepCare[5] is one of several works tackling the problem of prediction from Electronic Medical Record (EMR) data. The authors employ the popular Long Short-Term Memory (LSTM) recurrent neural net and deploy the network to predict disease progression and unplanned readmission for diabetic patients. The results are compared to the baseline of Markov models for the former and to the baselines of Support Vector Machines (SVMs) and random forests for the latter. Our review will critique the design of their network, feature vectors, benchmarks, and performance metrics. We will also discuss related works in predictive health care.

2 Background & Models

Five models are discussed in the paper: RNNs, LSTMs, Markov chains, SVMs, and random forests. A vanilla recurrent neural net takes identical neurons unrolls them into a chain dependency. The output of any given neuron can be given from a fully connected weighting of the previous neuron and the current input, followed by some nonlinear activation function. The LSTM is a type of recurrent neural net that also shares the same parameters between all of the nodes, but extends the vanilla RNN by holding some internal state that is responsible for influencing the output and influencing subsequent nodes. In the paper's own words: "Each LSTM unit has a memory cell that has state $c_t \in R^K$ at time t. The memory is updated through reading a new input $x_t \in R^M$ and the previous output $h_{t-1} \in R^K$. Then an output states h_t is written based on the memory c_t. There are 3 sigmoid gates that control the reading, writing and memory updating: input gate i_t, output gate o_t and forget gates f_t, respectively."
Markov chains are a type of probabilistic graphical model that assumes that the current state of the system is independent of any history, meaning that the system is memoryless (the 'Markovian' assumption). The nodes of the model, in our case, are disease states and the edges leaving any given node are probabilistic weights of transitioning to another disease state. The paper uses this simple model as a benchmark to compare with the vanilla RNN and the LSTM for disease progression prediction.
SVMs and random forests are both types of binary classifiers. Each input vector exists as a datapoint in some corresponding vector space of the same dimensionality, the SVM essentially finds hyperplanes to divide this space and classify the points. The random forest is a type of ensemble learning similar in

29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

nature to the SVM. The random forest scheme is resistant to overfitting by creating multiple decision trees that are fit to randomly drawn subsets of the data, and averaging is used afterwards across all the random trees.

3 Results

3.1 Modeling Disease Progression

The author's compare diagnoses prediction using DeepCare with Markov Models and plain Recurrent Neural Networks. Table 1 below shows the results of their study. Each column in the table corresponds to the precision of predictions after each discharge. DeepCare is shown to have better prediction rates than Markov Model with memoryless disease transition probabilities and plain Recurrent Neural Networks. Using memory dependent models clearly shows a significant gain in accuracy.

Table 1. Precision@n_{pred} diagnoses prediction.

Model	$n_{pred} = 1$	$n_{pred} = 2$	$n_{pred} = 3$
Markov	55.1	34.1	24.3
Plain RNN	63.9	58.0	52.0
DeepCare (interven. + param. time)	**66.0**	**59.7**	**54.1**

3.2 Predicting Unplanned Readmission

The author's also evaluate DeepCare for risk prediction. They compare risk prediction using DeepCare against SVM and Random Forests. Both max-pooling and sum-pooling strategies are used to compare the baseline models. The experiment involves taking a random discharge for each patient as a random prediction point. From this point, the model predicts whether an unplanned readmission will happen. Since the dataset is skewed the authors decided to evaluate the prediction using F-score, which takes into account both precision and recall. Table 2, shows the results of this experiments indicating that DeepCare has the highest F-Score.

Table 2. Results of unplanned readmission prediction within 12 months.

	Model	F-score (%)
1	SVM (*max-pooling*)	64.0
2	SVM (*sum-pooling*)	66.7
3	Random Forests (*max-pooling*)	68.3
4	Random Forests (*sum-pooling*)	71.4
5	LSTM (*mean-pooling + logit. regress.*)	75.9
6	DeepCare (*mean-pooling + nnets*)	76.5
7	DeepCare ([*interven. + time decay*]+ *recent.multi.pool. + nnets*)	77.1
8	**DeepCare ([*interven. + param. time*]+ *recent.multi.pool. + nnets*)**	**79.1**

4 Criticism

4.1 Choice of Dataset

The input to this model is diagnosis and interventions. While the authors reference personalized healthcare, they describe a prediction task that does not seem to incorporate any biological data or molecular data. In addition, they do not mention standard clinical data such as demographics, lifestyle or family history. This data is usually available for all patients in the case of diabetes and has high predictive power. In addition, standard lab tests are not mentioned. The authors evaluated DeepCare for both disease progression and unplanned readmission for diabetic patients. It would have been interesting to do these evaluations for a variety of diseases rather than just diabetes. The paper could have made a stronger argument for DeepCare had they showed improvements in predicting several

diseases which are known to be challenging to diagnose. For example, there is plenty of work in modeling the progression of Alzheimer's [1-3], however it remains a challenging problem.

4.2 Interpretation of Results

The authors present results on one diabetes cohort. No validation is done on an external cohort (even a diabetic external cohort would make the paper stronger). Also, no validation is done with non-overlapping data-collection periods, since they select data points for training and testing randomly from the same cohort. In addition, evaluation on a public data set could also be more useful.

The author's claim that DeepCare has a "realistic model of disease progression," without describing this term. The authors thus far have used machine learning to solve a regression problem on a sequence of clinical events and outcomes, but this "model" is not a scientific model, in the traditional sense, where some sort of mechanics of causality is described. In other words, correlation is not causation. They may have purposely chosen not to, because each disease has a different model of progression. However, a general description would have been useful. It should also be noted that they detailed the progression of diagnoses from pre-diabetes to post-diabetes. Furthermore, they reference Wang et al. [4] , which describes disease progression models in detail for a variety of diseases such as Alzheimer's, Diabetes, and Chronic Obstructive Pulmonary Disease. Given the scope of the paper, having a such a reference seems sufficient.

4.3 Evaluation of Unplanned Readmission

When evaluating risk prediction, the authors decide to select a single random discharge from each patient rather than multiple discharges. It is not clear why they do this. While it makes sense to select the discharge point randomly, they could have selected a pool of random discharge points to evaluate on. Selecting multiple discharge points could prove to be more effective for testing purposes.

4.4 Network Design

The authors arbitrarily chose to make emergency admissions to have an input feature weight of 1 and non-emergency admissions to have an input feature weight of 0.5. There is no clear reason why the authors chose to do this. This information could just be left in the feature vector where the neural net could decide whether or not these features are important. It may be that longer training and convergence times forced this decision, but the way to correct for that is with smarter weight initialization.

4.5 Baseline Comparison Models

The authors use vanilla Markov models where the could have used more advanced sequence modeling like Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs). Both of these models have a chain dependency that makes them much more amenable to labeling disease progression as opposed to the memoryless vanilla Markov model.

5 Related Work

This paper is the first to apply neural networks to this type of data (diagnoses and interventions) to model disease progression. Related work on EMR data used other approaches. Neural networks have also been applied to clinical time series data for the classification task of clinical diagnosis. Tran et al [7] proposed an automatic feature extraction method from EMR and compared it to manually extracted features, using a logistic regression risk prediction model. Lipton et al [6] applied LSTMs to multivariate time series data: diastolic and systolic blood pressure, peripheral capillary refill rate, end-tidal CO2, fraction of inspired O2, Glascow coma scale, blood glucose, heart rate, pH, respiratory rate, blood oxygen saturation, body temperature, and urine output in the form of episodes which vary from hours to months. They use sequential target replication to allow the network to pass information across many sequence steps in learning the target function. They use auxiliary target training as a way to avoid overfitting because of their small data set size. Auxiliary training helps because loss is minimized over actual and auxiliary targets combined. There are significant differences between this

paper and our original paper e.g. the task and the data type. This speaks to heterogeneity of EMR data and the complexities inherent in analysis and prediction from EMR.

6 Conclusion

This paper applies neural networks to the problem of disease progression. The authors focus on the episodic, time dependent properties in EMR to motivate their choice of architecture. The paper describes LSTM as a the basis of the architecture and uses time parameterization and multiscale temporal pooling to deal with heterogeneous temporal nature of the input data. This model is evaluated on the tasks of predicting disease progression and emergency (unplanned) admission using a single diabetes cohort with no validation on another diabetes cohort from a different institution or other diseases. Comparison of model performance is done with respect to baseline models chosen by the authors rather than existing literature.

References

[1] Modeling Alzheimer's disease progression using the disease system analysis approach. Roberto Gomeni, Monica Simeoni, Marina Zvartau-Hind, Michael C. Irizarry, Daren Austin, Michael Gold Alzheimers Dement. 2012 Jan; 8(1): 39–50. Published online 2011 Jul 22. doi: 10.1016/j.jalz.2010.12.012

[2] H.M.J. Fonteijn, M.J. Clarkson, M. Modat, J. Barnes, M. Lehmann, S. Ourselin, N.C. Fox, D.C. Alexander An event-based disease progression model and its application to familial Alzheimer's disease Proc. Information Processing in Medical Imaging (IPMI), Vol. 6801 of Lecture Notes in Computer Science, Springer (2011), pp. 748–759

[3] Predicting the progression of Alzheimer's disease dementia: A multidomain health policy model Green, Colin et al. Alzheimer's and Dementia: The Journal of the Alzheimer's Association , Volume 12 , Issue 7 , 776 - 785

[4] Wang, X., Sontag, D., Wang, F.: Unsupervised learning of disease progression models. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 85–94. ACM (2014)

[5] Trang Pham, Truyen Tran, Dinh Phung, and Svetha Venkatesh. DeepCare: A Deep Dynamic Memory Model for Predictive Medicine. arXiv preprint arXiv:1602.00357, 2016.

[6] Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzell. Learning to diagnose with lstm recurrent neural networks. arXiv:1511.03677, 2015.

[7] T. Tran, W. Luo, D. Phung, S. Gupta, S. Rana, R. L. Kennedy, A. Larkins, and S. Venkatesh, "A framework for feature extraction from hospital medical data with applications in risk prediction," BMC bioinformatics, vol. 15, no. 1, p. 6596, 2014.