Journal of Biomedical Informatics 69 (2017) 218–229

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

# Predicting healthcare trajectories from medical records: A deep learning approach

Trang Pham *, Truyen Tran, Dinh Phung, Svetha Venkatesh

*Center for Pattern Recognition and Data Analytics, Deakin University Geelong, Australia*

ARTICLE INFO

*Article history:*
Received 10 November 2016
Revised 23 March 2017
Accepted 1 April 2017
Available online 12 April 2017

*Keywords:*
Electronic medical records
Predictive medicine
Long-Short Term Memory
Irregular timing
Healthcare processes

ABSTRACT

Personalized predictive medicine necessitates the modeling of patient illness and care processes, which inherently have long-term temporal dependencies. Healthcare observations, stored in electronic medical records are episodic and irregular in time. We introduce DeepCare, an *end-to-end* deep dynamic neural network that reads medical records, stores previous illness history, infers current illness states and predicts future medical outcomes. At the data level, DeepCare represents care episodes as vectors and models patient health state trajectories by the memory of historical records. Built on Long Short-Term Memory (LSTM), DeepCare introduces methods to handle irregularly timed events by moderating the forgetting and consolidation of memory. DeepCare also explicitly models medical interventions that change the course of illness and shape future medical risk. Moving up to the health state level, historical and present health states are then aggregated through multiscale temporal pooling, before passing through a neural network that estimates future outcomes. We demonstrate the efficacy of DeepCare for disease progression modeling, intervention recommendation, and future risk prediction. On two important cohorts with heavy social and economic burden – diabetes and mental health – the results show improved prediction accuracy.



## 1. Introduction

When a patient is admitted to a hospital, there are two commonly asked questions: "what is happening?" and "what happens next?" The first question is about illness diagnosis, the second is about predicting future medical risk [43]. Whilst there are a wide array of diagnostic tools to answer the first question, fewer technologies address the second [41]. Traditionally, the prognostic question may be answered by experienced clinicians who have seen many patients or by clinical prediction models with well-defined risk factors. But both methods are expensive and restricted in availability. Modern electronic medical records (EMRs) promise a fast and cheap alternative. An EMR contains the history of hospital encounters, diagnoses, interventions, lab tests and clinical narratives. The wide adoption of EMRs has led to recent research to build predictive models from this rich data source [26,45,48,49].

Answering prognostic inquiries necessitates modeling patient-level temporal healthcare processes. Effective modeling must address four open challenges: (i) *Long-term dependencies in healthcare*: the future illness and care may depend critically on historical illness and interventions. For example, the onset of diabetes in middle age remains a risk factor for a person's remaining life; cancers may recur after years; and a previous surgery may prevent certain future interventions. (ii) *Representation of admission information*: an admission episode consists of a variable-size discrete set containing diagnoses and interventions. (iii) *Episodic recording and irregular timing*: medical records vary greatly in length, are inherently episodic in nature and irregular in time [47]. The data is episodic because it is only recorded when the patient visits the hospital and undergoes an episode of care. The episode is often tightly packed in a short period, typically ranging from a day to two weeks. The timing of arrivals is largely random. (iv) *Confounding interactions between disease progression and interventions*.

We address the four challenges to construct a predictive system that is both *end-to-end* and *generic* so that it can be deployed on different hospital implementations of EMRs. An end-to-end system requires minimal or no feature engineering, reads medical records, infers present illness states and predicts future outcomes.

Existing methods are poor in handling such complexity. They inadequately model variable length [45] and ignore the long-term dependencies [24,31,51]. Temporal models based on the Markovian assumption are limited to model temporal irregularity and have no memory, and thus they may completely forget previous major illness given an irrelevant episode [1]. Deep learning,

* Corresponding author.
*E-mail address:* phtra@deakin.edu.au (T. Pham).

http://dx.doi.org/10.1016/j.jbi.2017.04.001
1532-0464/

which has recently revolutionized cognitive fields such as speech recognition, vision and computational linguistics, holds a great potential in constructing end-to-end systems [27]. However, little work has been done using deep learning for healthcare [8,13,28,46]. While work in deep learning has been done to address the challenge of long-term dependencies [5,7,29], the three other challenges remain unsolved.

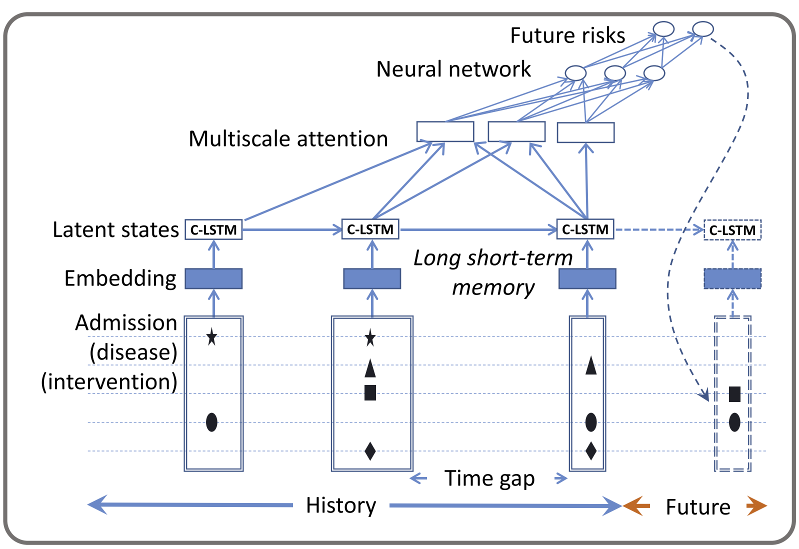
To this end, we introduce DeepCare, an *end-to-end* deep dynamic memory neural network that addresses the four aforementioned challenges [38]. DeepCare is built on Long Short-Term Memory (LSTM) [15,21], a recurrent neural network equipped with *memory cells* to store experiences. At each time-step, the LSTM reads an input, updates the memory cell, and returns an output. Memory is maintained through a *forget gate* that moderates the passing of memory from one time step to another, and is updated by seeing new input at each time step. The output is determined by the memory and moderated by an *output gate*. In DeepCare, the LSTM models the illness trajectory and healthcare processes of a patient encapsulated in a time-stamped sequence of admissions. The inputs to the LSTM are information extracted from admissions. The outputs represent illness states at the time of admission. *Memory maintenance enables capturing of long-term dependencies*, thus addressing the first challenge. In fact, this capacity has made LSTM an ideal model for a variety of sequential domains [15,17,44].

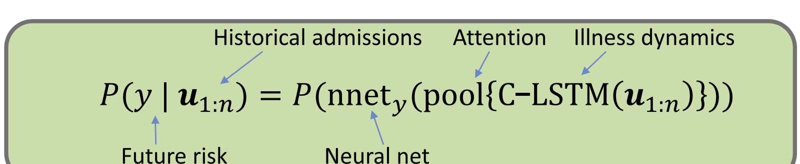
Addressing the other three drawbacks, DeepCare introduces C-LSTM as an extension of the standard LSTM unit (Fig. 1). For representing an admission, which is a set of discrete elements in different types such as diagnoses and interventions, the solution is to embed these elements into continuous vector spaces. Vectors of the same type are then pooled into a single vector. Type-specific pooled vectors are then concatenated to represent an admission. In that way, *variable-size admissions are embedded into continuous distributed vector space*. The admission vectors then serve as input features for the C-LSTM. As the embedding is learned from data, the model does not rely on manual feature engineering.

For *irregular timing, the forget gate is extended to be a function of irregular time gap* between consecutive time steps. We introduce two new forgetting mechanisms: monotonic decay and full time-parameterization. The decay mimics natural forgetting when learning a new concept in human. The parameterization accounts for more complex dynamics of different diseases over time. The resulting model is sparse in time and efficient to compute since only observed records are incorporated, regardless of the irregular time spacing. Finally, in DeepCare *the confounding interaction between disease progression and interventions is modeled* as follows: Interventions influence the output gate of current illness states and the forget gate which moderates memory carried into the future. As a result, the illness states (the output) are moderated by past and current interventions.

Once illness states are outputted by the C-LSTM layer, they are aggregated through a new time-decayed multiscale pooling strategy for future projection. This allows further handling of time-modulated memory. Finally at the top layer, pooled illness states are passed through a neural network for future prognosis. (See Fig. 1 for a graphical depict of DeepCare.) Overall, DeepCare is an end-to-end prediction model that relies on no manual feature engineering, is capable of reading generic medical records, memorizing a long history, inferring current illness states and predicting the future risk.

We demonstrate our DeepCare on answering a part of the question "what happens next?". In particular, we validate our model on *disease progression, intervention recommendation* and *future risk prediction*. Disease progression refers to the next disease occurrence given the medical history. Intervention recommendation is about predicting a subset of treatment procedures for the current diagnoses. Future risk may involve readmission or mortality within a predefined period after discharge. Our experiments are demonstrated on two datasets of very different nature – diabetes (a well-defined chronic condition) and mental health (a diverse mixture of many acute and chronic conditions). The cohorts were col-




$$P(y \mid \boldsymbol{u}_{1:n}) = P(\mathrm{nnet}_y(\mathrm{pool}\{\text{C-LSTM}(\boldsymbol{u}_{1:n})\}))$$

**Fig. 1.** DeepCare architecture. **Top**: The healthcare dynamics are modelled as a sequence of C-LSTM units which model irregular timing and interventions (see Fig. 4 for more detail). The symbols (e.g., stars, circles and triangles) in the bottom rectangles are diagnosis and interventions codes (see Section 2.2). **Bottom**: Predictive computation summarized in an equation.