Deep Recurrent Neural Networks for Predicting Intraoperative and Postoperative Outcomes and Trends

Shruthi Gopalswamy[1] and Patrick J. Tighe[2] and Parisa Rashidi[3]

Abstract— **Intraoperative vital signals convey a wealth of complex temporal information that can provide significant insights into a patient's physiological status during the surgery, as well as outcomes after the surgery. Our study involves the use of a deep recurrent neural network architecture to predict patient's outcomes after the surgery, as well as to predict the immediate changes in the intraoperative signals during the surgery. More specifically, we will use a Long Short-Term Memory (LSTM) model which is a gated deep recurrent neural network architecture. We have performed two experiments on a large intraoperative dataset of 12,036 surgeries containing information on 7 intraoperative signals including body temperature, respiratory rate, heart rate, diastolic blood pressure, systolic blood pressure, fraction of inspired O_2 and end-tidal CO_2. We first evaluated the capability of LSTM in predicting the immediate changes in intraoperative signals, and then we evaluated its performance on predicting each patient's length of stay outcome. Our experiments show the effectiveness of LSTM with promising results on both tasks compared to the traditional models.**

I. Introduction

A wide variety of physiological variables are routinely recorded during each standard surgery, such as body temperature, respiratory rate, blood pressure and heart rate. These variables are critical indicators of the physiological status of a patient. While anesthesiologists routinely monitor these signals, the demanding environment of the operating room calls for intelligent systems to reduce the cognitive load on anesthesiologists [1]. One of the major functions of such an intelligent decision support system is the capability to predict the immediate future trends. This can play an important role in timely interventions and can reduce the risk of post-operative complications. Besides predicting the immediate changes, one can use the intraoperative signals to gain important insights about patient's outcomes after the surgery. For example, predicting the length-of-stay (LOS) can help in optimized management and allocation of the resources in the hospital.

While analyzing intraoperative signals can reveal important information about patient's status and outcomes, these complex multivariate time series still pose many challenges such as varied time series length or missing values. Different surgeries can vary in length by hours and some signals might not be continuously recorded during the surgery. Finally, different intraoperative signals are measured at different frequencies, resulting in varied signal resolution.

In the past, various conventional machine learning methods have been used to analyze time-series. Examples include using k-Nearest Neighbors (kNN) along with Dynamic Time Warping (DTW) [2], conventional neural networks [3], and Markov models [1]. These techniques either require extracting hand-crafted features from the time series [3], or face challenges in the presence of many signals [2], or typically work in discrete spaces [1]. Recently, Recurrent Neural Networks (RNN) as an alternative approach have been successfully applied in a large number of domains, including machine translation [4], language understanding [5], speech recognition [6] and image captioning [7]. The gated RNNs such as the Long short-term memory (LSTM) [8] have furthermore improved upon regular RNNs by addressing problems such as the vanishing gradient problem to allow modeling of long-term dependencies [9].

In this paper, we use an LSTM architecture to predict short-term trend of the intraoperative signals during surgery, as well as to predict patient outcomes after the surgery. We have used a large intraoperative dataset of 12,036 surgeries containing information on 7 intraoperative signals including body temperature, respiratory rate, heart rate, diastolic blood pressure, systolic blood pressure, fraction of inspired O_2 and end-tidal CO_2. We use an LSTM model to predict the future values of blood pressure during surgery. The blood pressure needs to be continuously monitored during surgery by the attending anesthesiologist as it plays an important role in determining the cardiovascular status of a patient. [1]. Blood pressure is particularly important to measure given that it can be drastically altered by surgical stress, blood loss, fluid shifts, and anesthetics. Also, it has been postulated that significant intraoperative hemodynamic instability may reflect surgical complexity, stress, and latent cardiovascular factors which may influence the rate of postoperative recovery. Hence, we will also predict the LOS as an outcome after the surgery. One of the major factors that help gauge the competence of health care services is LOS, and numerous efforts have been designed to shorten LOS, such as the enhanced recovery programs [10]. Our proposed approach is generic and it can be also used to predict the changes in other physiological signals or outcomes of interest. To the best of our understanding, this is the pioneering effort to use an LSTM model to predict both short-term and long-term

*This work was supported by NIH NIGMS K23102697 and NIGMS 1R01GM114290

[1]S. Gopalswamy is with the Department of Computer & Information Science, University of Florida, Gainesville, Florida, USA. shruthigopal16@ufl.edu

[1]P. Tighe MD MS is with the Department of Anesthesiology, College of Medicine, University of Florida, Gainesville, Florida, USA. ptighe@anest.ufl.edu

[3]P. Rashidi is with the J. Crayton Pruitt Department of Biomedical Engineering, University of Florida, Gainesville, Florida, USA. parisa.rashidi@ufl.edu

978-1-5090-4179-4/17/$31.00 ©2017 IEEE

dependencies using a multitude of intraoperative signals. Our results show the effectiveness of LSTM with promising results on both tasks compared to the traditional models.

In the remaining part of the paper, first in section II, we will describe related work, then we will describe our LSTM model in more detail in section III, finally we will describe the experiment results in section IV, followed by the conclusion.

II. RELATED WORK

Time-series prediction has been extensively studied in the past. Various classical methods such as the Autoregressive integrated moving average model (ARIMA) [11] have been used in numerous applications, e.g. to predict stock prices, cyclone formation, and epidemic disease propagation [11]. Generally, two different approaches are taken: (1) feature extraction, and (2) direct time series manipulation. In the former approach, various frequency/time domain and statistical features are obtained from a time series. This is then used as input to a conventional statistical or machine learning model such as a support vector machine (SVM) [12]. Feature extraction methods include discrete wavelet transform or discrete Fourier transform [13]. While this former approach is convenient, since conventional methods can be reused, nonetheless it relies on a set of hand-crafted features which can be quite different for each new problem. The second approach avoids the feature extraction step by comparing the similarities of the entire time series using a similarity measure such as DTW [2], or making predictions based on a subset of the most recent values as in the family of ARIMA models [11]. While the conventional approaches have shown promising results in the past, utilizing both short-term and long-term dependencies have been difficult.

Recently, RNN models have become popular in many domains as an alternative approach for processing temporal or sequential data, outperforming the classical models [4]–[7]. The gated RNNs such as LSTMs [8] or Gated Recurrent Units (GRU) [14], further improve upon regular RNNs by enabling them to also model long-term dependencies in an efficient manner [9].

In this paper, we use an LSTM stacked network. An LSTM basic unit is composed of a memory cell which has a *forget gate*, an *input gate*, and an *output gate*, as can be seen in Fig. 1. These memory units can be stacked to form hidden layers. The presence of the multiplicative gates in the memory cells in LSTM ensures that the network retains a weighted average of those patterns that are relevant, while disallowing the patterns that do not contribute to the final outcome. LSTMs have the added advantage of obviating the need for human-engineered features, and thus can work directly on the raw multivariate time series data.

LSTMs have been used in a few recent clinical and medical tasks with promising results. Chauhan and Vig use LSTMs for arrhythmia detection from electrocardiogram (ECG) signals [15]. Other deep learning models besides LSTMs also have been used in similar tasks, for example Lehman et al. use a switching vector autoregressive-

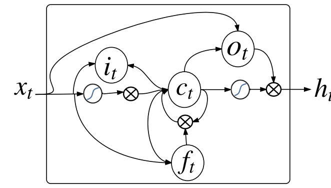

Fig. 1. A basic LSTM unit with forget gate f ,input gate i and output gate o. c_t is the cell state and h_t is the final output

convolutional neural network (SVAR-CNN) to predict hospital mortality from blood pressure time series data [16]. Choi et al. use GRUs to predict the diagnosis and medication categories for a subsequent visit [17]. Similarly, Lipton et al. [18] use an LSTM model to predict the primary diagnosis code based on Pediatric intensive care unit (PICU) multivariate time series data. Their work is the most similar to our work, with the difference that we are using intraoperative signals to predict short-term changes in the intraoperative signals, as well as to predict outcomes after the surgery. This will represent the first effort towards using an LSTM model for predicting both short-term and long-term dependencies using a multitude of intraoperative signals.

III. MODEL ARCHITECTURE

The LSTM units form the basis of our model. Each LSTM basic unit is composed of 3 gates which serve different purposes. The input gate is a sigmoid gate that decides whether the input value in the cell needs to be passed. The forget gate outputs a value between 0 and 1 which would indicate whether the value in the cell needs to be discarded or remembered. The output gate is a sigmoid gate which decides what values will be provided as output. The value of the cell state is the sum of the product of the old cell state with the forget gate, and the product of the input gate with an activation function value. The cell state is then transformed by an activation function and multiplied with the output gate. This value is fed back into the LSTM unit. More specifically, at each step, we apply the updates outlined in Eq. 1 [6], [8].

$$\begin{aligned}
i_t &= \sigma(W_i.[h_{t\text{-}1}, x_t] + b_i) \\
f_t &= \sigma(W_f.[h_{t\text{-}1}, x_t] + b_f) \\
o_t &= \sigma(W_o.[h_{t\text{-}1}, x_t] + b_o) \\
\hat{c}_t &= \phi(W_c.[h_{t\text{-}1}, x_t] + b_c) \\
c_t &= f_t.c_{t\text{-}1} + i_t.\hat{c}_t \\
h_t &= o_t.\phi(c_t)
\end{aligned} \tag{1}$$

In the above equations,W_i,W_f,W_o stand for the weight matrices for the forget gate (f_t), input gate (i_t), output gate (o_t). Similarly, b_i ,b_f,b_o are the corresponding bias vectors. $\hat{c}_t$ is a temporary variable, c_t and $c_{t\text{-}1}$ stand for the cell state at time t and $t\text{-}1$ respectively, ϕ is the activation function and h_t is the final output.

Our model is composed of 2 hidden layers containing LSTM units and one fully connected dense layer containing regular neural units, as shown in Fig. 2.

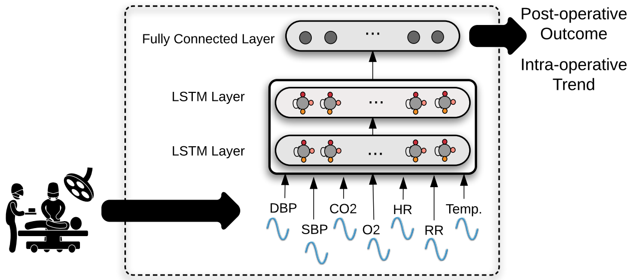

Fig. 2. Our model is composed of 2 hidden layers containing LSTM units and a fully connected layer to analyze intraoperative signals.

We follow a target replication strategy in order to achieve better generalization [18]. Missing values are handled by using zero imputation and forward-filling [19].

IV. EXPERIMENT

Our data has been collected in a study approved by the University of Florida's Institutional Review Board (IRB). The data consists of multivariate intraoperative signals obtained from 12,036 subjects aged 21 or older who received non-ambulatory and non-obstetric surgery in years 2014-2015. It contains information about body temperature, respiratory rate, heart rate, diastolic blood pressure, systolic blood pressure, fraction of inspired O_2 and end-tidal CO_2. The number of samples per patient episode varied between 38-312. Sample signals (Heart rate and blood pressure) for 3 different patients with a range of LOS values are shown in Fig. 3.

Each signal is measured irregularly and contains missing data. The data contains a combination of features that are measured often and those that are measured infrequently. For example, heart rate is recorded at the rate of 1 reading per minute, blood pressure is recorded every 1-5 minutes on an average, and temperature is recorded every 15 minutes-30 minutes. To overcome the irregularities in length of each patient's time series, as well as to standardize the time gaps between each measurement, the data has been re-sampled to every 5 minutes by smoothing over 5 minutes.

We have conducted 2 separate sets of experiments to evaluate the performance of LSTM for predicting both

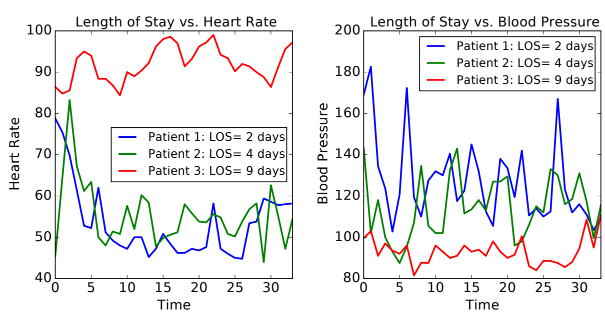

Fig. 3. Heart rate and blood pressure of 3 patients with varied length of Stay (2, 4, and 9 days).

postoperative outcomes and intraoperative trends. In order to select the most suitable model for this study, we conducted a series of tests, to determine the best parameters. The weights have been initialized using Glorot initialization [20]. By optimizing the parameters using 3-fold cross validation, we concluded that using adadelta [20] with a learning rate of 0.3 gave rise to the best results. Parameter tuning was done with regards to the activation function and the dropout rate. Some of the results have been shown in Table. II.

For predicting the LOS postoperative outcomes, the duration of hospitalization was divided into two different groups: patients with LOS less than or equal to 5 days and patients with LOS greater than 5 days. The average number of samples per patient episode was 62. Hence the data has been fed into the LSTM in 62 time steps. The data was split by patient into a 75: 25 ratio for training and testing having equally balanced or stratified classes. After selecting the most optimal set of parameters using 3-fold cross validation on a validation set, the model was used to predict each patient's LOS in the test set. Using metrics such as recall, precision and F1 score, the performance of the model was evaluated. The results of the evaluation has been compared against baseline methods such as KNN and SVM, as shown in Table. I. For the KNN and SVM model, we used hand-engineered features such as standard statistical features (e.g. mean, median, standard deviation, max/min, etc.). From the results shown in Table. I, we can see that the LSTM demonstrates an improved performance. The accuracy, precision, and recall values show an increase of 20% over the traditional baseline methods.

TABLE I. Comparison of k-nearest neighbor, Support Vector Machine, and LSTM on predicting LOS from intraoperative signals

Model	Accuracy	Precision	Recall	F1 score
KNN	49.25	49.17	49.25	49.20
SVM	52.84	54.64	52.84	36.76
LSTM	73.10	73.73	73.10	72.78

Next, we predicted the range in which the systolic blood pressure of a patient would fall into, based on the range of the previous systolic blood pressure values. The optimal window size was decided based on results from an ARIMA model. The blood pressure signal is first checked for stationary characteristics using Dickey-Fuller Test of stationarity [11]. It is made stationary by using the first difference.

The above results have been used in LSTM to predict the range of future blood pressure value. The blood pressure has been divided into different levels in order to discretize the time series. For example, values of blood pressure ranging from 0-90 are Class 1 (Low Blood pressure), values of blood pressure ranging from 90-120 are classified as Class 2 (Normal Blood Pressure) and the values greater than 120 (High Blood Pressure) are classified as Class 3. A window sampling size of 6 has been used to predict the value of the blood pressure at the 7th time step. The results of the experiment are shown in Table. III.

As shown in the Table. III, we can see the performance

TABLE II. Comparison of different initialization, learning rate, dropout and activation function on LOS prediction

Intialization		*Learning Rate*		*Dropout*		*Optimizer*	
Model	*Accuracy*	*Value*	*Accuracy*	*Value*	*Accuracy*	*Function*	*Accuracy*
Glorot Uniform	72.31	0.3	72.50	0.2	72.54	Adadelta	72.28
He Normal	71.66	0.01	64.22	0.3	72.47	Adamax	71.99
He Uniform	71.85	0.1	71.46	0.4	72.46	Adagrad	72.27
Uniform	72.10	0.001	48.80	0.5	72.41	RMSProp	72.16
Glorot Normal	71.60	0.2	72	0.6	72.19	Adam	69.59

TABLE III. Performance of LSTM on: a) Predicting the range of blood pressure, b) Predicting the length of stay (LOS)

Blood Pressure				*Length of Stay*			
Target	*Recall*	*Precision*	*F1 score*	*Target*	*Recall*	*Precision*	*F1 score*
Low Blood Pressure(Class 1)	29	57	38	Length of stay < 5 days	83	70	76
Normal Blood Pressure(Class 2)	91	76	83	Length of stay > 5 days	63	77	69
High Blood Pressure(Class 3)	61	82	70				

of the LSTM on both the tasks of predicting the length of stay as well as predicting the range of the blood pressure. The LSTM shows a high value of recall, precision and F1 score. These results point to the usefulness of deep recurrent models such as LSTM in predicting intraoperative and postoperative outcomes and trends, which subsequently can help in optimizing patient care and timely interventions.

V. Conclusion

In this paper, we have demonstrated that LSTMs are effective in extracting patterns from intraoperative signals and are effective in modeling long term dependencies in the data. We compared the performance of LSTM against baseline methods. While baseline methods required hand-engineered features, the LSTM could work on the raw time-series data. We also observed that LSTMs outperform conventional machine learning techniques such as SVM and KNN. Our future work involves using more sophisticated recurrent neural network structures to allow simultaneous processing of signals at various resolutions.

References

[1] P. Yang, G. A. Dumont, and J. M. Ansermino, "Online pattern recognition based on a generalized hidden markov model for intraoperative vital sign monitoring," *International Journal of Adaptive Control and Signal Processing*, vol. 24, no. 5, pp. 363–381, 2010.

[2] C. C. Hsu, P. H. Kung, M. Y. Yeh, S. D. Lin, and P. B. Gibbons, "Bandwidth-efficient distributed k-nearest-neighbor search with dynamic time warping," in *Big Data (Big Data), 2015 IEEE International Conference on*, pp. 551–560, Oct 2015.

[3] T. Koskela, M. Lehtokangas, J. Saarinen, and K. Kaski, "Time series prediction with multilayer perceptron, fir and elman neural networks," in *In Proceedings of the World Congress on Neural Networks*, pp. 491–496, Press, 1996.

[4] B. Zhao and Y. C. Tam, "Bilingual recurrent neural networks for improved statistical machine translation," in *Spoken Language Technology Workshop (SLT), 2014 IEEE*, pp. 66–70, Dec 2014.

[5] G. Mesnil, X. He, L. Deng, and Y. Bengio, "Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding," in *INTERSPEECH*, pp. 3771–3775, 2013.

[6] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*, pp. 6645–6649, IEEE, 2013.

[7] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: Lessons learned from the 2015 mscoco image captioning challenge," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2016.

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, pp. 1735–1780, Nov. 1997.

[9] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *Trans. Neur. Netw.*, vol. 5, pp. 157–166, Mar. 1994.

[10] M. Kelly, L. Sharp, F. Dwane, T. Kelleher, and H. Comber, "Factors predicting hospital length-of-stay and readmission after colorectal resection: a population-based study of elective and emergency admissions," *BMC health services research*, vol. 12, no. 1, p. 1, 2012.

[11] W. W.-S. Wei, *Time series analysis*. Addison-Wesley publ Reading, 1994.

[12] K.-j. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1, pp. 307–319, 2003.

[13] P. Chaovalit, A. Gangopadhyay, G. Karabatis, and Z. Chen, "Discrete wavelet transform-based time series analysis and mining," *ACM Comput. Surv.*, vol. 43, pp. 6:1–6:37, Feb. 2011.

[14] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[15] S. Chauhan and L. Vig, "Anomaly detection in ecg time signals via deep long short-term memory networks," in *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*, pp. 1–7, Oct 2015.

[16] L. w. Lehman, M. Ghassemi, J. Snoek, and S. Nemati, "Patient prognosis from vital sign time series: Combining convolutional neural networks with a dynamical systems approach," in *2015 Computing in Cardiology Conference (CinC)*, pp. 1069–1072, Sept 2015.

[17] E. Choi, M. T. Bahadori, and J. Sun, "Doctor AI: predicting clinical events via recurrent neural networks," *CoRR*, vol. abs/1511.05942, 2015.

[18] Z. C. Lipton, D. C. Kale, C. Elkan, and R. C. Wetzel, "Learning to diagnose with LSTM recurrent neural networks," *CoRR*, vol. abs/1511.03677, 2015.

[19] Z. C. Lipton, D. C. Kale, and R. C. Wetzel, "Directly modeling missing data in sequences with rnns: Improved classification of clinical time series," *CoRR*, vol. abs/1606.04130, 2016.

[20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The address: MIT Press, 1 ed., 2016.