2017 16th IEEE International Conference on Machine Learning and Applications

Monitoring Health Changes in Congestive Heart Failure Patients using Wearables and Clinical Data

Robert Fisher
Carnegie Mellon University
Pittsburgh, PA
rwfisher@cs.cmu.edu

Asim Smailagic
Carnegie Mellon University
Pittsburgh, PA
asim@cs.cmu.edu

George Sokos
West Virginia University School of Medicine
Morgantown, WV
george.sokos@wvumedicine.org

***Abstract*—In this work we present systems to monitor the health of a wide variety of high risk patients living with Congestive Heart Failure. For critical hospitalized patients, we introduce a deep learning framework for hospital records that uses a Word2Vec vector space representation to learn from a combination of structured data and unstructured text. The deep learning framework is able to assess patient risk, and accurately predict medical outcomes into the future. For less critical patients living at home, we also present algorithms for remote monitoring that can track a patient's changing health indicators using a wearable heart-rate sensors. The pool of individuals living with congestive heart failure is very diverse, which can make multi-faceted approaches to health monitoring, such as those presented in this work, attractive for observing large pools of high-risk patients. Collectively the methods presented in this paper allow us to continuously monitor a person living with CHF through hospitalization, discharge, and into their home.**

***Index Terms*—Intelligent monitoring, congestive heart failure, deep neural networks, Word2Vec, latent variable autoregression**

I. Introduction

More than a quarter of a million Americans die from complications resulting from Congestive Heart Failure (CHF) every year[1]. Previous research has shown that clinical intervention during initial presentation of symptoms can significantly reduce the risks of a cardiac event [1]. Early intervention requires careful monitoring of the patient's vital signs using in-home telehealth monitoring. However, by the time symptoms manifest, it may already be too late for effective intervention. In this work, we present methods to detect changes in health for CHF patients in a variety of situations. In particular, for high risk patients admitted to a hospital, we introduce a deep learning model with linguistic understanding to model clinical records and physician notes in order to predict risk factors for the patient. For healthier patients living at home with CHF, we show activity recognition and heart rate modeling algorithms to track a subject cardiac health using wearable heart rate monitors.

By combining intelligent monitoring systems in the home with clinical data we are able to construct a more complete understanding of a subject's health profile. This allows us to detect small changes over time, that could otherwise lead to life threatening complications.

[1]http://www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_heart_failure.htm

II. Related Work

There has been significant clinical research aimed at targeted intervention of Congestive Heart Failure patients to reduce the risks of readmission. One study suggested that continual monitoring of a patient in home can reduce the risk of hospital readmission by as much as 86% [2]. However, this same study concluded that telehealth systems, while extremely effective, did not lead to better outcomes than daily phone calls between a nurse and a patient. This evidence could suggest that by the time medical complications become apparent in telehealth data, the subject is already sufficiently uncomfortable to seek intervention when it is available.

Data driven methods offer us the possibility to detect more subtle, multimodal changes in a patient's physiology that may not be immediately apparent to a clinician. It has become more common for CHF patients to be evaluated by a machine learning model upon hospital discharge to determine the risk of readmission. These methods can achieve above 80% predictive accuracy, exceeding that of a physician in some instances [3]. In particular, these models are very useful for detecting the highest risk patients, that can then be monitored more frequently in order to prevent readmission. Machine learning methods have also been designed to diagnose CHF [4].

III. Approach

In this section we will present activity recognition and time series analysis algorithms to track patients in the home using wearable heart-rate monitors. We will then describe a deep learning framework using Word2Vec, that can predict medical diagnoses, prescriptions, and health outcomes using hospital records and physician notes. By building tools for the home as well as the hospital, we hope to reach the broadest population of CHF patients possible.

A. Datasets

This work is based on three datasets. In the first, heart Rate data was collected from several subjects using Apple Watch wearable heart rate monitors during 6 minute walks. The subjects' cardiac responses before and after the walk were also recorded.

In addition to the Apple Watch data collected with healthy subjects, we have also acquired data from the Physical Activity Monitoring for Aging People (PAMAP) project

0-7695-6321-X/17/31.00 ©2017 IEEE
DOI 10.1109/ICMLA.2017.000-9

includes heart rate and Inertial Measurement Unit (IMU) data from 30 elderly subjects (aged 55 to 86) performing a set of 13 physical activities such as bicep curls and leg extensions. The IMU units are placed on the subject's wrist, ankle, and torso. Collectively, the IMUs allow us to identify what type of physical activity the patient underwent, and the heart-rate monitor allows us the measure the cardiac response to that physical activity. Because heart dynamics can differ greatly depending on the subject's activity, it is important to consider the physical activity a subject is undergoing when analyzing heart-rate data. With the Apple watch data, we have collected data using a prescribed and codified physical regimen, which simplifies analysis and allows us to focus on measuring the cardiac response. However, in general we would prefer a system that can identify different types of physical activity seamlessly and autonomously. The PAMAP dataset allows us to develop methods to automatically perform activity recognition, and modify our heart dynamics analysis conditioned on the detected physical activity.

Finally, the largest dataset, the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II) dataset, consists of data for 25,328 individual stays in the Intensive Care Unit. Of these ICU stays, 39.1% of them are hospitalized with cardiac related diagnoses, most often congestive heart failure. For each discrete hospital entry, the data includes waveform data consisting of continuous ECG data, blood pressure readings, respiration readings, and fingertip plethysmograph. Additionally the data includes complete clinical information, such as demographic information, healthcare provider type (RN, MD) admission and discharge notes, diagnoses, medications, lab tests, fluid balance, and unstructured clinician notes. This dataset allows us to test our multiple sensor prediction algorithms at scale with a dataset than includes many instances of severe cardiac health exasperations.

B. Latent Variable ARIMA Models

Much of the cardiac data we are concerned with in this work comes in the form of time-series data. Autoregression, a class of algorithms that predicts future observations in a time series based on previous observations, is a useful tool for modeling the dynamics of this sort of data. Unfortunately most autoregression assume stability in the data, $i.e.$ the moments of the distribution are constant. Even autoregression methods that support moving averages, such as Autoregressive Integrated Moving Average (ARIMA) models, respond to change very slowly over time. With CHF patients however, vital signs can quickly degrade over the course of days or even hours. As such, we introduce Markovian Latent Variables to a traditional ARIMA model to allow for rapid adaptation to sudden changes [8].

A traditional ARIMA model is represented as ARIMA(p, d, q), with p corresponding to the degree of the autoregression, d as the degree of difference required for stationarity, and q as the degree of the moving average model. This leads to a forecasting model as follows:

$$\hat{y}_t = \mu + \sum_{i=1}^{p} (\psi_i y_{t-i}) - \sum_{j=1}^{q} (\theta_i e_{t-j})$$

In this instance, e_j represent the moving average differences of the time series, which can be thought of as a discrete variant of a d^{th} degree derivative. When $d = 2$, these values can be thought of as the local acceleration of the time series.

With a latent variable ARIMA model, we introduce hidden states $h_i \in h_1 ... h_N$, and each hidden state will have its own ARIMA parameters θ and ψ (though p, d, and q remain the same). As with a traditional Hidden Markov Model, we have an initial state distribution π, and $N \times N$ transition matrix T denoting the probability of moving between states, and an observation matrix O denoting the probability of an observed value given a hidden state. Each timestep in the series will have a latent state assigned by a Hidden Markov Model, and the ARIMA parameter are changed accordingly. This results in a moving average autoregression system that is able to respond instantly to sudden changes in patient health.

C. Activity Recognition

In this section, we describe our prediction and modeling results using the Physical Activity Monitoring for Aging People (PAMAP) dataset [9]. This dataset consists of 30 elderly individuals performing 13 physical exercises while wearing a heart-rate monitor and 3 IMU devices. As outlined in section B, we would like to perform activity recognition, so that we can better understand the exercises that patients wearing wearable devices are undergoing. By conditioning our cardiac models on the specific type of exercise an individual is performing, we can gain more robust insights into their cardiac health.

We will use these small repetitive movements, which are often called activity motifs, to build an unsupervised classifier to identify which exercise is being performed. From the IMU signals, we compute standard signal processing features over one second frames and feed this sequence into our spectral Hidden Markov Model. From here, we interpret the sequence of latent states as a sequence of exercise motifs, such as raising one's arm, or extending one's leg. After generating these sequences using a total of 6 motifs, we then cluster these sequences into 13 clusters using the K-means algorithm. We can then interpret each cluster as a single exercise.

This dataset comes with labels telling us what activity is being performed in each trial. We use these labels to evaluate our method, however the model itself is fully unsupervised, and can therefore adapt to a wide range of unknown exercises in the field.

After clustering the exercises, we train a latent variable ARIMA model for the heart-rate data within each cluster. As we will see in the empirical results section, this approach significantly outperformed training a single ARIMA model for the entire dataset, or training a separate ARIMA for each individual trial. We also evaluated the activity detection method using only a single wrist-mounted IMU, which would simulate users wearing an Apple watch. For exercises that include full body movement, or only arm movement, this approach works nearly as well as the model trained using all three IMUs. Unsurprisingly, exercises that only use leg muscles, such as leg curls, cannot be reliable detected using only a single wrist mounted IMU.

D. Deep Learning for Clinical Records Data

In this section, we describe a deep learning model with the network architecture shown in figure 1 trained with the much larger and higher dimensional Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) database. Of the 25,328 intensive care stays contained in the database, we restrict ourselves to roughly 3,000 subjects diagnosed with congestive heart failure that are hospitalized with complications related to cardiac failure. We wish to train the model to predict if a patient's health is likely to improve or worsen in the comings days. This is a difficult task as patients admitted to the ICU are already in critical condition, which results in more significant fluctuations in health metrics.

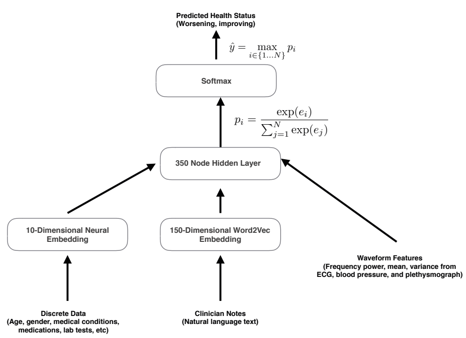

Fig. 1. Architecture of Deep Learning Model for MIMIC II Dataset

The architecture leverages a 150-dimensional semantic representation of natural language based on the Word2Vec model [10]. Word2Vec is a deep learning approach that learns the semantic meaning of words as a vector space representation. Intuitively, Word2Vec seeks to learn a representation of a word such that we can predict what the word will be given the representations of the words surrounding (the context) it in a piece of text. Similarly, we should be able to predict the context given a single word. This approach has proven successful in a wide range of real world applications. Our Word2Vec model is trained using articles from Wikipedia, which then allows us to compute a composed 150-dimensional representation of the clinician notes included in the MIMIC II Databse. Using Word2Vec in this way allows for some interesting arithmetic in the semantic space. If we begin with a concept such as 'king', subtract the concept of 'man', and add the concept of 'woman' we arrive at the concept corresponding to 'queen'. By representing the semantics of natural language in this way, we are able to leverage the rich information only present in unstructured clinical notes.

We also learn similar 10-dimensional neural embeddings for the discrete metadata, such as age, weight, gender, medical prescriptions etc. Finally, we compute standard time-series analysis (Fourier-frequency power, mean value, variance) for the ECG signal for the patient. The Word2Vec embeddings, the metadata embeddings, and the ECG features are then passed to a single 350 node hidden network layer. We use SoftMax for our core learning function, which seeks to minimize the entropy of the predicted class distributions such that the maximum likelihood class corresponds with the labels in the training data.

The predicted variable in this task is the reason for subject discharge from the ICU. If the subject is discharged normally with clinician approval, we conclude that the subject's health has stabilized and the health exacerbation has passed. On the other hand, a subject may be transfer to specialized facilities for more intensive care, or the subject may die while in the intensive care unit. In these two instances, we conclude that the subject's health has declined. We use all other features of the MIMIC II dataset described in the dataset section to predict these outcomes.

IV. Results

Table I shows the empirical results of our studies of the PAMAP dataset. This table shows prediction results for two tasks, activity recognition, and heart-rate time series prediction. For activity recognition, the naive baseline simply predicts the most common of the 13 known exercises, and we see that the motif clustering using the spectral HMM significantly outperforms this baseline. We also see that when we restrict ourselves to predicting upper body or full body exercises using only a single IMU, we can perform nearly as well as if we had access to all three IMUs.

For the heart-rate prediction task, we ask the model to predict heart-rates for 10 seconds into the future using previous time series data. We present two baselines, one which predicts the mean value seen in the time series thus far, and one that predicts the current time series value. We also train the latent variable ARIMA models in two ways. In the first approach, we train a single model using the entire dataset, while in the second approach we train 13 different ARIMA models, one for each clustered exercise. We see that while both ARIMA models outperform the baseline approaches, pre-conditioning training on the predicted exercise cluster significantly improves performance.

Task	**Model**	**Mean Accuracy**
Activity Recognition	Baseline	0.12
Activity Recognition	Spectral HMM	0.78
Activity Recognition (Upper body exercises)	Spectral HMM (Single IMU)	0.73
Task	**Model**	**Residual Error**
Heart-rate Prediction	BL (Mean)	18 BPM
Heart-rate Prediction	BL (Current Value)	15 BPM
Heart-rate Prediction	Single ARIMA Model	8 BPM
Heart-rate Prediction	Clustered ARIMA Model	4 BPM

TABLE I
PREDICTION RESULTS FOR PAMAP DATA

Model	Mean Accuracy
Baseline (Always predict improvement)	0.819
Support Vector Machine	0.862
Deep Learning without Word2Vec	0.886
Deep Learning with Word2Vec	0.910

TABLE II
PREDICTION RESULTS FOR HEALTH EXACERBATIONS WITH MULTIPLE SENSORS USING MIMIC II DATA

Table II shows the results of applying the deep learning framework with Word2Vec to the multi-sensor MIMIC II dataset to predict health outcomes as described in section III-D. These results are obtained by training the network with 2,000 subjects diagnosed with congestive heart failure, and testing on another 1,000 subjects with congestive heart failure. We see that on average, 0.819 of the subjects were discharged normally from the ICU, which corresponds with our positive outcome label, therefore any predictive method must surpass this naive baseline. In addition to the deep neural network described in section III-D, we also trained a simple Support Vector Machine with a 3-dimensional polynomial kernel. While table II indicates that both methods outperform the simple baseline, the neural network shows the best performance with the cause of discharge being correctly predicted in 91% of instances. It is also worth noting here that the addition of the unstructured clinician notes with Word2Vec contributes roughly 2.5 percentage points to the predictive accuracy, indicating that there is some information contained in the clinical notes that cannot be inferred from the structured data alone.

Finally, table III compares various methods when predicted recovery time. In particular, the latent variable ARIMA model achieves a mean residual error of 6.4 seconds, with a ± 3.1 seconds .95 confidence bound. This indicates statistical significance when comparing again the baseline method of relying on an individual's average recovery time.

Method	Mean Residual Error
LV-ARIMA(15, 2, 10)	4.9 Seconds
ARIMA(15, 2, 10)	6.4 Seconds
Mean Recovery Time	10.4 Seconds

TABLE III
MEAN RESIDUAL ERROR FOR PREDICTED RECOVERY TIME

V. CONCLUSION

In this work, we have presented methods for tracking the health status of a wide spectrum of CHF patients.

By applying deep learning to clinical records for critical patients in hospital, we can identify the patients with the highest risk of poor medical outcomes. This system is able to represent the semantics of unstructured clinician notes, revealing information that is likely not included in the structured data considered by most intelligent record monitoring systems. The network can also be easily adapted to predict medical diagnoses, prescribed medicine, or other relevant clinical information.

For CHF patients living at home, we have introduced unobtrusive wearable monitoring systems to track changes in health over time. For CHF patients at risk of health exacerbations, remote monitoring can often save lives, but clinicians do not have the resources to track large populations of remote patients. Compared to other tele-health systems, wrist mounted wearables also have the advantage of allowing us to track a subjects response to physical activity, a vital indicator of cardiac health that is not tracked by the majority of remote monitoring systems.

Taken collectively, intelligent monitoring of high risk CHF patients offers us opportunities to reduce medical costs, improve outcomes, and avoid unnecessary health exacerbations. As electronic medical data becomes more and more prevalent, it will become increasingly necessary to supply clinicians with sophisticated tools to help them track the highest risk populations of patients.

ACKNOWLEDGMENTS

We gratefully acknowledge the support by Carnegie Mellons Disruptive Health Technology Institute through their grant and by the Pennsylvania Infrastructure Technology Alliance, a partnership of Carnegie Mellon, Lehigh University and the Commonwealth of Pennsylvania Department of Community and Economic Development (DCED).

REFERENCES

[1] S. Stewart, J. E. Marley, and J. D. Horowitz, "Effects of a multidisciplinary, home-based intervention on planned readmissions and survival among patients with chronic congestive heart failure: a randomised controlled study," *The Lancet*, vol. 354, no. 9184, pp. 1077–1083, 1999.

[2] A. F. Jerant, R. Azari, and T. S. Nesbitt, "Reducing the cost of frequent hospital admissions for congestive heart failure: a randomized trial of a home telecare intervention," *Medical care*, vol. 39, no. 11, pp. 1234–1245, 2001.

[3] K. Zolfaghar, N. Meadem, A. Teredesai, S. B. Roy, S.-C. Chin, and B. Muckian, "Big data solutions for predicting risk-of-readmission for congestive heart failure patients," in *Big Data, 2013 IEEE International Conference on*. IEEE, 2013, pp. 64–71.

[4] G. Liu, L. Wang, Q. Wang, G. Zhou, Y. Wang, and Q. Jiang, "A new approach to detect congestive heart failure using short-term heart rate variability measures," *PloS one*, vol. 9, no. 4, p. e93399, 2014.

[5] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[6] R. Fisher, A. Smailagic, R. Simmons, and K. Mizobe, "Using latent variable autoregression to monitor the health of individuals with congestive heart failure," in *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*. IEEE, 2016, pp. 1016–1019.

[7] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111–3119.