Using Latent Variable Autoregression to Monitor the Health of Individuals with Congestive Heart Failure

Robert Fisher
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213
Email: rwfisher@cs.cmu.edu

Asim Smailagic
Institute for Complex Engineered Systems
Carnegie Mellon University
Pittsburgh, PA 15213
Email: asim@cs.cmu.edu

Reid Simmons
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
Email: reids@cs.cmu.edu

Kimitake Mizobe
Seiko Epson Corp
Tokyo, Japan 160-8801
Email: kmizobe@erd.epson.com

***Abstract*—Sudden weight gain in patients living with Congestive Heart Failure (CHF) is often an indication that the individual is retaining fluid, which often means that patient's heart has weakened leading to increased risk of kidney or cardiac failure. Clinical interventions can be made at this stage, leading to better outcomes, however it is essential that the interventions take place before the patient's health declines too drastically. In this work, we present a latent variable autoregression model that tracks patient weight and blood pressure over time, allowing us to predict weight values into the future. We are also able to model continuous heart-rate signals and evaluate a subject's response to physical activity. This allows us to detect signs of health decline days earlier than existing rule-based systems, leading to the possibility of earlier clinical interventions, potentially preventing deadly medical emergencies.**

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 a model to predict sudden weight changes in CHF patients, which is an indication that the subject has begun retaining fluid. Integrated Autoregression is a popular class of algorithms for predicting future time series values, however with Congestive Heart Failure patients, we may see sudden changes in the time series distribution that traditional autoregressive models cannot anticipate. To account for this, we introduce Markovian latent variables to monitor the patient's global health trends. This allows us to quickly introduce new autoregression parameters to account for rapidly changing conditions in the patient's health.

Using a latent variable autoregression model, we are able to predict patient weight two days in advance with a mean residual error of 0.92 pounds. This allows us to detect sudden weight changes with a sensitivity of .8313. This framework could allow for clinical interventions to take place days before current systems, potentially saving the lives of many individuals living with Congestive Heart Failure.

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 statistically significant improvements when compared with daily phone calls between a nurse and a patient. This evidence could suggest that by the time medical

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

complications become apparent in telehealth data, the subject is already sufficiently uncomfortable to accept intervention when it is offered.

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 [5].

Latent variables have been used in conjunction with autoregression models in the past to allow for more rapid and dynamic responses to changing conditions. However, these efforts have mostly been in the field of economics [7], with much less effort being given to medical applications.

III. Dataset

The data for this project was collected by Meadville Medical Center[2] using Authentidate telehealth systems. The data includes 9 volunteers living with Congestive Heart Failure over an average period of 6 months. Each subject provided daily or twice daily measurements that include systolic blood pressure, diastolic blood pressure, heart rate, and current weight. This dataset includes 11 instances of sudden weight gain, which are likely the product of fluid retention resulting from poor kidney health as a result of CHF.

IV. Methods

When monitoring large populations of patients, the cost of clinical supervision becomes a prohibitive factor. Telehealth systems can provide useful means for monitoring high-risk patients, but we cannot depend on annotations supplied by clinical experts. Instead, we must infer what we can from the signals themselves. Autoregression is a very popular school of methods that allows us to model the dynamics of a time series and predict future values, without requiring supervised training. In this class of algorithms, future values of a time series are estimated using a function that takes the most recent time series values as inputs. Multivariate autoregression, which operates over multiple time series simultaneously, has been successfully used in medical applications in the past [9]. Unfortunately, many autoregression models treat the time series as stable, meaning that the statistical moments of the distribution do not change over time. With congestive heart failure data, a patient with declining health will produce a time series with changing moments in the distribution, so we do not meet the stability requirement of many approaches. Autoregressive Integrated Moving Average (ARIMA) models are built to address just such a problem, allowing the first moment of the time series distribution to change over time.

An ARIMA model is denoted ARIMA(p, d, q), with p representing the degree of the autoregression, d is the degree of difference required for stationarity, and q represents 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.

This type of moving average model works well for non-stationary time series, however it often requires that the moments of the time series distribution change slowly over time. For congestive heart failure patients, a sudden medical emergency can lead to very sudden changes in the time series distribution. As such, we model these sudden changes using Markovian hidden states, similar to a Hidden Markov Model. If our model has hidden states $h_i \in h_1 ... h_N$, each hidden state will have its own ARIMA parameters θ and ψ (though p, d, and q remain the same). We then 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. By modeling the latent states of the patient's health, we are able to make rapid transitions between ARIMA models when it appears that the subject's condition is beginning to change.

We use a small parameter validation set from the patient data, and train the latent variable model and a multivariate ARIMA model on the weight time series, using diastolic blood pressure as a side input. This validation set is used to select the hyper-parameters of the model, resulting in $N = 3$ hidden states, with each

[2] http://meadvillemedicalcenter.com/

Model	Residual Mean (lbs)
BL (Current Value)	2.1
BL (Mean)	1.61
ARIMA	1.08
Latent Variable ARIMA	0.92

TABLE I
MEAN RESIDUAL ERROR COMPARISON (WEIGHT)

Model	Residual Mean (BPM)
BL (Current Value)	16.2
BL (Mean)	21.2
ARIMA	18.2
Latent Variable ARIMA	10.8

TABLE II
MEAN RESIDUAL ERROR COMPARISON (HEART-RATE)

state representing a $ARIMA(5, 2, 3)$ model. The hidden state parameters are trained using gradient descent, while the ARIMA parameters are estimated using the Box-Jenkins model.

Once the model is trained, we can use it to instantiate alerts for clinicians about the possible decline of patient health. In particular, we evaluate the average slope of the predicted time series curve over the next two days (denoted ψ). We can then issue an alert any time this value surpasses a given threshold:

$$\psi \geq \delta$$

We can then control the sensitivity and specificity of the model according to a medical center's available resources. In particular, the cost of a false negative (a patient with declining health that is not detected by the system) is much higher than the cost of a false positive. As such, we would select the lowest value of δ possible, such that a medical center has the resources to personally monitor and intervene with the number of patient flagged by the system for that threshold value. In this way, we can achieve the highest sensitivity possible, at the cost of lower specificity.

V. Empirical Results

A. Modeling Patient Weight

Table I shows the mean residual error of a traditional ARIMA model, without latent variables, as well as two simple baselines. In the first baseline, the current weight of the patient is simply used as the predicted value, while the second baseline uses the average weight of the patient across the entire time series. A 95% confidence bound for this values is ± 0.15 pounds, Meaning that the improvement seen over the baseline by both ARIMA models is statistically significant.

While the mean residual error of the latent variable ARIMA model shows only a modest improvement over the traditional model, we see significantly better prediction for the latent variable model when dealing with sudden changes in weight. Of the 11 instances in the dataset in which a patient gained more than 2 pounds in one day, using a threshold value of $\delta = 1.0$ allows the latent variable ARIMA model to correctly predict the weight gain in 9 instances, resulting in a sensitivity value of 0.8313, along with 8 false positive examples.

B. Modeling Continuous Heart-rate

In addition to using tele-health to record daily readings, we can use wearable devices to continually monitor a subject's heart-rate, which can also allow us to observe a subject's response to physical activity. In particular, at the end of a period of physical activity, the length of time it takes for an individual to return to their average resting heart-rate is a significant indicator of heart health [12]. In this work, we have used the latent-variable ARIMA method to model data collected from 13 volunteers using wearable heart-rate monitors, as well as heart-rate data of hospitalized CHF patients monitoring using an electrocardiogram (ECG) [13].

Similar to the weight gain time series discussed in the previous section, the heart-rate time series are strongly multi-modal, with differing averages during exercise, rest, or recovery. This means to the latent-variable nature of the model is essential for adapting changing conditions. In particular, table II shows that a traditional ARIMA model underperforms a simple model that simply predicts the current heart-rate for future values. The latent variable ARIMA model, on the other hand, outperforms all other models evaluated.

VI. Conclusion

As we have seen in this work, we can quite accurately model the dynamics of a weight time series for patients with congestive heart failure, and we see some promising results for predicting sudden exacerbations of health. These preliminary results should be confirmed with a larger set of examples of sudden weight gain, potentially giving us the means to cut costs and save lives. It is interesting to note that for the 2 instances of fluid retention not detected by the system, using lower values of δ does not produce improved results. This suggests that daily weight and blood pressure readings alone may be insufficient to detect all causes of cardiac failure.

Similarly, we have shown that this model can accurately continuous heart-rate activity, which allows us to monitor a CHF patient's response to physical exercise. In future work, we should continue consolidating warning signs taken from many disparate indicators of heart-health into a unified framework for tracking patient health. Using these tools, we have the opportunity to allow clinicians to intervene with wavering patients days before current methods allow, which could potentially save the lives of many individuals living with chronic congestive heart failure.

Acknowledgment

We give thanks to the Epson company for their assistance and valuable feedback. We also gratefully acknowledge the National Science Foundation for their support through EAGER grant number IIS1450543, as well as funding from Carnegie Mellon's Disruptive Health Technology Institute. This material is also based upon work supported by the Quality of Life Technology Center and the National Science Foundation under Cooperative Agreement EEC-0540865.

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. I

[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. II

[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. II

[4] P. Melillo, N. De Luca, M. Bracale, and L. Pecchia, "Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability," *IEEE journal of biomedical and health informatics*, vol. 17, no. 3, pp. 727–733, 2013.

[5] 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. II

[6] J. Polisena, K. Tran, K. Cimon, B. Hutton, S. McGill, K. Palmer, and R. E. Scott, "Home telemonitoring for congestive heart failure: a systematic review and meta-analysis," *Journal of Telemedicine and Telecare*, vol. 16, no. 2, pp. 68–76, 2010.

[7] L. Bauwens and D. Veredas, "The stochastic conditional duration model: a latent variable model for the analysis of financial durations," *Journal of econometrics*, vol. 119, no. 2, pp. 381–412, 2004. II

[8] A. Ang and M. Piazzesi, "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables," *Journal of Monetary economics*, vol. 50, no. 4, pp. 745–787, 2003.

[9] A. E. Gelfand and P. Vounatsou, "Proper multivariate conditional autoregressive models for spatial data analysis," *Biostatistics*, vol. 4, no. 1, pp. 11–15, 2003. IV

[10] F. Z. N. Y. Z. Jinliang, "The impacts of monetary policy shock on stock market liquidity: An empirical study based on markov switching vector autoregression [j]," *Journal of Financial Research*, vol. 7, p. 006, 2011.

[11] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "Arima models to predict next-day electricity prices," *IEEE transactions on power systems*, vol. 18, no. 3, pp. 1014–1020, 2003.

[12] G. L. Pierpont, D. R. Stolpman, and C. C. Gornick, "Heart rate recovery post-exercise as an index of parasympathetic activity," *Journal of the autonomic nervous system*, vol. 80, no. 3, pp. 169–174, 2000. V-B

[13] D. S. Baim, W. S. Colucci, E. S. Monrad, H. S. Smith, R. F. Wright, A. Lanoue, D. F. Gauthier, B. J. Ransil, W. Grossman, and E. Braunwald, "Survival of patients with severe congestive heart failure treated with oral milrinone," *Journal of the American College of Cardiology*, vol. 7, no. 3, pp. 661–670, 1986. V-B