Employing time-series forecasting to historical medical data: an application towards early prognosis within elderly health monitoring environments

Antonis S. Billis[1] and Panagiotis D. Bamidis[1]

[1]Medical Physics Laboratory, Medical School, Faculty of Health Sciences, Aristotle University of Thessaloniki

Abstract. This work describes a first attempt to apply time-series forecasting analysis to health historical data in order to perform prediction of early pathological signs within telehealth applications, such as the Ambient Assisted Living environments for the elderly. A benchmark of state-of-the-art learning methods were applied to a set of artificial time-series data, simulating hypertensive patient profiles, based on blood pressure measurements. Results provided a fair proof of our initial hypothesis. Based on this first experimentation, our plans are to further investigate these findings in real –life or lab settings with seniors, thus proving the usefulness of time-series forecasting as a monitoring tool and an early prognosis mechanism in telehealth systems.

Keywords: telehealth, smart home, ambient assisted living, support vector machine, ARIMA models, time series forecasting, neural networks

1   Introduction

Senior citizens suffer nowadays from a wide variety of chronic conditions that reduce their independency level and deteriorate their health status [1]. This often happens due to the fact that they do not receive timely health assessment. In most of the cases this happens because any change is considered as part of the natural ageing process; seniors also tend to omit admitting they have a problem, or even of their fear of being institutionalized [2]. Early prediction of abnormal variation of health parameters may lead to early diagnosis of chronic diseases and subsequently to better medical decision-making and planning.

In this respect technology has much to offer the elderly. Recent technological advances resulted in the equipping of home environments with a plethora of sensors that aim to improve the seniors' quality of life and increase their independency by providing alerts in case of emergencies while increasing their socialization [3]. These approaches provided significant results in cases of fall detection, inappropriate use of electricity devices, estimation of participant's functionality, etc. [4]. However, they have mainly focused on the detection of life-threatening acute events and they neglected the significance of slow-varying trends that may influence the health status of senior citizens [5] at a later time.

Time series analysis is a methodology that provides: i) pattern recognition of historical data (detection of frequent types of sequences) and .ii) forecasting of future values based on historical trends. There are several time-series forecasting methods known to the literature, such as exponential smoothing [6], Box-Jenkins seasonal ARIMA models [7] and neural networks [8].

In this paper, we aim to showcase the use of several state-of-the-art machine learning algorithms, such as Gaussian Processes, Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and Box-Jenkins ARIMA models, with a set of artificially developed scenarios, relevant to patient models with high risk of hypertension. Our hypothesis is that one can take advantage of existing time-series forecasting methodologies to identify health trends over time and predict early signs of health deterioration, based on historical sets of health measurements and events.

The ARIMA model has advantages in its well-known statistical properties and effective (linear) modeling process. However, it may not work well in the presence of nonlinear relationships. In contrast, support vector machines and artificial neural networks time series models can capture the historical information by nonlinear functions and thus could prove to be efficient time series forecasting methods because of their flexible nonlinear mapping ability and tolerance to complexity in forecasting data.

2   Materials and Methods

In order to benchmark the forecasting methods a number of artificial scenarios were created. These concerned the simulation of day-to-day variation of diastolic blood pressure. Blood pressure is a prerequisite to monitoring seniors' physical health status progress or deterioration. This clinical parameter is very important since by monitoring it over a period of time it is possible to identify future risky health situations, such as hypertension. These situations may subsequently lead to a plethora of health problems, such as a sudden stroke episode or cardiovascular disease. Based on norms, an individual is hypertensive if he or she experiences repeatedly an elevated blood pressure exceeding 140 (systolic) and over 90 mmHg (diastolic).

The test data set consists of ten cases that reflect high risk profiles of seniors. The scenarios are produced randomly using a normal distribution with a mean diastolic blood pressure value nearly at the physiological margin of 80 mmHg and standard variation varying from 1 to 3.

The synthetic data were formed based on a priori knowledge derived from international norms, thus corresponding to hypertensive patient profiles. Specifically, the instances were modeled as a Gaussian process with a standard mean value and variability. The larger the variability the nosier the time-series is. The algorithms were tested under small, medium and high variability rates, so to provide cases of gradual forecast difficulty.

In order to perform time-series forecasting we have used well known machine learning methods and employed their existing implementations within publicly available AI suites such as WEKA [10] and Phicast [9]. More specifically, from the WEKA suite the following learning methods have been applied: GaussianProcesses (kernel: RBFKernel with gamma parameter set to default value 0.01), SMOreg (ker-

nel: RBFKernel with gamma parameter set to default value 0.01) and finally MultilayerPerceptron (all parameters set to the default values). The parameters of the ARIMA model were selected as follows: p=1, d=1, q=1, resulting to a ARIMA(1,1,1) model.

3 Experimentation and results

The experiments compare predicted values with test data provided by the random artificial scenarios. Initial scenarios were split to two sets: one training set used for learning, which consisted of 50 instances and a second one which served as the test set for the evaluation of the forecasting methods. All instances represent measurements taken ideally on a daily basis.

Two experiments were conducted. Firstly, the ARIMA method alone was employed to forecast future value ranges and then the rest methods were applied to the same data sets to predict single future values. In the first case the metric used for the evaluation of the method is the accuracy of the prediction calculated as percentage of the correctly predicted ranges that the actual value was included and the total amount of the test set. In the second case, the following standard metrics are calculated: the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE).

An indicative result of the ARIMA prediction algorithm is depicted below (**Fig. 1**). Average prediction accuracy resulted to 91.2%.
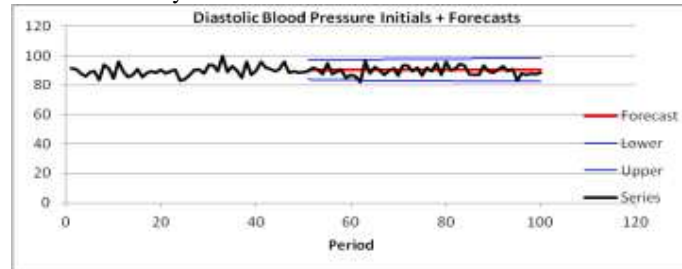

Fig. 1. Forecast Range vs Actual Values of Diastolic Data (95% prediction intervals)

Table 1 summarizes the evaluation results of the forecasting learning methods As shown GaussianProcesses learning algorithm provides the least error-prone analysis among the three. Neural Nets provided the worst forecasting results.

Table 1. Benchmark results of the tested forecasting algorithms

Learning Method	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
GaussianProcess	1.798633	2.292558
SVM	2.005342	2.523508
Neural Network	6.061283	7.126667

4 Discussion

In this paper, time-series forecasting was employed to identify early pathological symptoms based on historical measurements. The clinical utility of forecasting (e.g.

vital signs) is of major importance in order to avoid hospitalization and alleviate the socioeconomical burden for caring, while it may improve significantly life quality of senior citizens.

Box-Jenkins ARIMA forecasting provides relatively accurate short-term forecasts regarding future ranges of the monitored variable (blood pressure). SVMs and GaussianProcessess can provide relatively fair near-future predictions. Preliminary results on artificial data partially support our initial hypothesis. However, there are several issues that set limits to the aforementioned results.

AI techniques were applied just in univariate time-series, whereas in most real-life health problems that elderly people suffer from, tend to be multifactorial. This means that forecasting a single parameter may not provide sufficient indications of early disease, since it needs to be investigated under its temporal relation with associative factors/parameters. Another limitation when it comes to real-life applications is that these methods require a larger amount of data than may be available within smart home environments; especially where a small set of unobtrusive sensors will be available. Also sensor failure may lead to missing data. In order to overcome, such problems several data imputation techniques may be applied [11].

Future plans of this research include among others: the setup of pilots, where actual sensors will be placed within a lab setting [13] and several seniors will be recruited to perform daily life activities for a long enough period so to gather enough historical data.

The forecasting output will be adaptive in terms of the size of the temporal window in order to facilitate the expert to estimate the possibility of both short-term events and long-term future trends. Furthermore, each actual/predicted instance will be mapped to normal or abnormal class and the classification accuracy will be calculated based on whether forecasting methods can predict the onset of a pathological condition at an early stage.

Time-series forecasting could be ideally employed within the context of an ambient assisted living environment, providing answers as whether we could detect transition patterns indicative of future health deterioration or timely estimate chronic alterations in the presence of outliers that may be either due to system (sensors) failure or to acute events. Combining seniors' health profile and time-series forecasting analysis, an intelligent Ambient Assisted Living environment would be able to propose the adoption of optimal lifestyle patterns according to the user's needs (acting proactively). This strategy could combine various non-pharmacological interventions (affective/cognitive/exer-games), alterations to the diet and/or some simple advice regarding a healthier lifestyle.

5   Conclusion

The reason that motivated us to use time-series forecasting methodologies is that telehealth monitoring should be able to detect apart from rapid deterioration states or life-threatening situations also slow-varying chronic conditions and provide early prognosis. More specifically, cases such as the geriatric depression or cognitive decline are characterized by a gradual impairment that may last for years and will be hardly noticed by the seniors or their carers, except for late stages where the outcomes

of the disease remain irreversible [12]. Initial experiments with artificial data provided encouraging results. However, real-life experimentation will give us the chance to further fine-tune existing prediction algorithms and evaluate time series forecasting on the basis of its accuracy in the health monitoring field of use.

References

1. W. Lutz, B.C. O'Neill, and S. Scherbov, "Europe's population at a turning point", Science, vo. 28, no. 299, pp. 1991-1992, 2003.
2. Hayes TL, Pavel M, Kaye JA. An unobtrusive in-home monitoring system for detection of key motor changes preceding cognitive decline. *Proc. of the 26th Annual Intl. Conf. of the IEEE EMBS*, pp. 2480-2483, San Francisco, CA. (2004)
3. M. Popescu, G. Chronis, R. Ohol, M. Skubic, and M. Rantz: An eldercare electronic health record system for predictive health assessment. Paper presented at: Proceedings of the IEEE 13th International Conference on e-Health Networking, Applications and Services; Columbia, MO; June 13-15, pp. 193-196, 2011.
4. Rantz MJ, Scott SD, Miller SJ, Skubic M, Phillips L, Alexander G, et al. Evaluation of health alerts from an early illness warning system in independent living. Computers Informatics Nursing. 31(6):274–280, 2013.
5. T. Hayes, M. Pavel, and J. Kaye, An approach for deriving continuous health assessment indicators from in-home sensor data. In Technology and aging: Selected papers from the 2007 international conference on technology and aging, vol. 21, pp. 130–137, 2008
6. Winters, P. R. "Forecasting sales by exponentially weighted moving averages." Management Science, 6:324–342., 1960
7. Box GEP, Jenkins GM. Time Series Analysis: Forecasting and Control, 2nd edition. San Francisco, CA: Holden Day, 1976
8. Cortez, P., Rocha, M., and Neves, J. "Time Series Forecasting by Evolutionary Neural Networks." chapter III, Artificial Neural Networks in Real-Life Applications, Idea Group Publishing, USA, pages 47–70, 2005
9. Rob J. Hyndman, Anne B. Koehler, Ralph D. Snyder and Simone Grose. A state space framework for automatic forecasting using exponential smoothing methods http://www.elsevier.com/locate/ijforecast. (2002).
10. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten: The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1, (2009).
11. Shuai Zhang, Sally I. McClean, Member, IEEE, and Bryan W. Scotney Probabilistic Learning From Incomplete Data for Recognition of Activities of Daily Living in Smart Homes, IEEE Transactions On Information Technology In Biomedicine, Vol. 16, No. 3, 2012
12. Abellan van Kan, G., Rolland, Y., Nourhashémi, F., Coley, N., Andrieu, S., Vellas, B.,. Cardiovascular disease risk factors and progression of Alzheimer's disease. Dementia and geriatric cognitive disorders 27, 240–6 (2009)
13. Artikis, A., Bamidis, P.D., Billis, A., Bratsas, C., Frantzidis, C., Karkaletsis, V., Klados, M., Konstantinidis, E., Konstantopoulos, S., Kosmopoulos, D., Papadopoulos, H., Perantonis, S., Petridis, S., Spyropoulos, C.S.: Supporting tele-health and AI-based clinical decision making with sensor data fusion and semantic interpretation: The USEFIL case study. International Workshop on Artificial Intelligence and NetMedicine. p. 21 (2012).