Towards a Probabilistic Method for Longitudinal Monitoring in Health Care

Ning Xiong(✉) and Peter Funk

School of Innovation, Design and Engineering, Mälardalen University, 72123 Västerås, Sweden
{ning.xiong, peter.funk}@mdh.se

Abstract. The advances in IoT and wearable sensors enable long term monitoring, which promotes earlier and more reliable diagnosis in health care. This position paper proposes a probabilistic method to address the challenges in handling longitudinal sensor signals that are subject to stochastic uncertainty in health monitoring. We first explain how a longitudinal signal can be transformed into a Markov model represented as a matrix of conditional probabilities. Further, discussions are made on how the derived models of signals can be utilized for anomaly detection and classification for medical diagnosis.

Keywords: Health monitoring · Longitudinal signal · Symbolic time series · Markov model · Case-based reasoning

1 Introduction

In recent years there has been rising interest in wearable sensors for personal health care [1]. Integrating these devices with wireless communication provided by IoT [2] is hopeful to create new technology that would have significant impact on the way clinical monitoring is performed nowadays. Particularly IoT provides a convenient means of transmitting and recording long term biological signals, which convey much richer information than conventional lab-test based static measurements. Utilizing dynamic longitudinal data is beneficial to promote earlier and more accurate diagnosis results for future health monitoring systems.

However, longitudinal monitoring in health care faces two major challenges. The first lies in the big data volume that is collected continuously. There is a gap between the rate at which data become available and our ability to interpret and handle them. It is crucial to develop novel data analysis and mining tools to extract concise information and identify abnormality in real-time during the monitoring of the subject.

The second challenge arises from the inherently stochastic nature of data evolution in health monitoring. It is important to characterize the truly dynamic property of temporal patterns while ignoring random triviality in data analysis. How to represent uncertain characteristics residing in data and how to utilize such uncertain information in reasoning/learning is a key issue for reliable (anomaly) detection and diagnosis.

This position paper aims to suggests a probabilistic method to address the above two challenges. The proposed roadmap consists of three consecutive stages. The first is to convert original, real-valued signals into shorter symbolic time series. In the second

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2016
M.U. Ahmed et al. (Eds.): HealthyIoT 2016, LNICST 187, pp. 30–35, 2016.
DOI: 10.1007/978-3-319-51234-1_5

stage, the converted time series data are further transformed into a concise matrix to capture the stochastic and dynamic property of the underlying process. Finally, in the third stage, different matrices (derived from original sensor signals) are compared with each other in order to detect significant change of data transition patterns as well as to identify possible health problems. We hope that the presented work would offer an initial step towards the development of a useful framework to tackle uncertain and longitudinal data profiles in health monitoring.

The remainder of the paper is as follows. Section 2 presents the ways in which a longitudinal signal can be modeled into a concise matrix. Section 3 discusses how such matrices can be utilized for anomaly detection and diagnosis in health monitoring. Finally, the paper is concluded in Sect. 4.

2 Concise Modeling of Longitudinal Signals

This section explains how a longitudinal signal can be compressed by a concise model. It is accomplished by the following two steps: (1) converting the original signal into a symbolic series; (2) modeling the symbolic series with a matrix of pattern transition probabilities.

2.1 Converting Signal into Symbolic Time Series

The first step in our solution is to convert the sampling-point based representation of the signal into an interval-based representation. An interval consists of a set of consecutive sampling points and thus it encompasses multiple sampling periods in the time dimension. Subsequently, data within an identical interval have to be generalized into one symbolic value; the symbolization is conducted via discretization of the range of possible values of the signal. Next we shall outline three approaches that can be used in practice to convert a primary numerical signal into a shorter time series profile.

Symbolic Approximation was proposed in [3], in which the whole duration of the signal is divided into equally sized intervals, i.e., each interval encompasses the same amount of sampling periods. The data in each interval is averaged into a mean value, thereby creating an intermediate sequence of real numbers summarizing signal behaviors in the consecutive time intervals. This sequence is termed as PAA (Piecewise Aggregate Approximation) of the original signal. Then the PAA sequence is further transformed into a symbolic form by mapping the real numbers in it into corresponding symbols.

Temporal Abstraction [4, 5] was proposed to derive high level generalization of data from time-stamped representations towards interval-based interpretations. Basically this is achieved by aggregating adjacent entities falling in the same region into a cluster and summarizing behaviors in this cluster with a concept (symbol) corresponding to the region. Thereafter, arranging concepts of clusters according to the order of their appearances produces a required symbolic time series.

More specifically, the tasks of temporal abstraction can be performed on state abstraction or trend abstraction. The former focuses on the measured values themselves to extract intervals associated with qualitative concepts such as low, normal, and high, while the latter focuses differences between two neighboring records to discover patterns of changes such as increase, decrease, and stationarity in the series. Obviously trend abstraction is equivalent to applying state abstraction to the secondary series of differences derived from the primary signal of measurements.

Alternatively, symbolic time series can also be obtained via **Phase-Based Pattern Identification**, as suggested in [6]. It is motivated by the fact that sometimes a lengthy sensor signal from health monitoring may comprise a series of phases and every phase has its importance to identify its property (pattern) alone. In such cases, we need to separate the profile of the signal into a set of sub-signals with each of which corresponding to a phase inside the whole duration. As sub-signals are shorter and simpler, it would be relatively easy to classify their patterns using traditional signal processing and machine learning approaches. The final symbolic series is constructed by combining the patterns of sub-signals in terms of the order of appearance, which provides a compact and abstract representation of the evolution of data in the whole signal profile.

2.2 Characterization as Markov Model

After conversion of the primary signal as stated in Subsect. 2.1, we acquire a symbolic time series $x(1)$, $x(2)$, …, $x(t)$, …, $x(n)$, in which an element $x(t) = S_i$ reflects the fact that the process under monitoring is in state (symbol) S_i at time step t. We have to focus on transitions of states between adjacent time steps rather than single symbolic values for characterizing the evolution of data in the time series.

Since this time series originates from a stochastic process in health monitoring, we suggest using the Markov model to depict the uncertain transitions in it. According to the Markov property, the probability for the state at time step $t + 1$ is only dependent on the state at time step t, regardless of the states in the previous time steps, i.e.,

$$P\{x(t+1) = S_j|x(t) = S_i, x(t-1) = S_k, \cdots\} = P\{x(t+1) = S_j|x(t) = S_i\} \quad (1)$$

Equation (1) implies that only transitions between two successive time steps are required in the model of the symbolic time series.

Let $\{S_1, S_2, \ldots, S_M\}$ be the set of possible states (symbols) of the process monitored for health care. We use a_{ij} ($i, j = 1, 2, \ldots, M$) to denote the probability for the process to move from state S_i to state S_j in two consecutive time steps. Hence a_{ij} is defined as a conditional probability:

$$a_{ij} \equiv P\{x(t+1) = S_j|x(t) = S_i\} \quad \forall t \quad (2)$$

This conditional probability in Eq. (2) can simply be calculated as the ratio of the number of transitions from state S_i to S_j to the number of transitions starting from S_i in the series.

Finally, the stochastic Markov model of the symbolic time series can be formulated as a concise matrix as follows:

$$G = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ \cdots & \cdots & \cdots & \cdots \\ a_{M1} & a_{M2} & \cdots & a_{MM} \end{pmatrix} \quad (3)$$

with $a_{ij} \geq 0$ and $\sum_{j=1}^{M} a_{ij} = 1 \quad \forall i$

where the elements in row i reveal the probability distribution for the next state after state S_i. Note that the size of the matrix is merely determined by the number of states (or symbols), which is independent of the length of the time series. This offers an attractive opportunity of strong data reduction to benefit data storage and handling in the health monitoring system.

3 Anomaly Detection and Diagnosis

This section addresses how the model of the symbolic time series can be utilized for anomaly detection and diagnosis in health monitoring. First we shall explain the the ways of calculating the distance between matrices of time series in Subsect. 3.1. Then, in Subsect. 3.2, we discuss how the developed distance metric can be employed to support detection and classification of abnormal situations.

3.1 Measuring the Distance Between Two Models

Our goal is to evaluate the distance between two symbolic time series cases that are represented by matrices G and G' respectively. As each row in these matrices represents a distribution of probabilities of state transition, we first calculate the distances for pairs of probability distributions from the two matrices. Then the distances between probability distributions for various starting states are aggregated to achieve an overall dissimilarity between the two models of time series.

The matching of two probability distributions can be performed in terms of relative entropy or information gain. Hence we apply Jeffreys divergence (J-divergence) [7] to quantitatively distinguish two probability distributions in comparison. Suppose that $TB(i) = [a_{i1}, a_{i2}, \cdots, a_{iM}]$ and $TB'(i) = \left[a'_{i1}, a'_{i2}, \cdots, a'_{iM}\right]$ are two probability distributions described in the ith rows of G and G' respectively, the J-divergence between $TB(i)$ and $TB'(i)$ is formulated as follows:

$$\begin{aligned} J(TB(i), TB'(i)) &= \sum_{j=1}^{M} a_{ij} \cdot log\left(\frac{a_{ij}}{a'_{ij}}\right) + \sum_{j=1}^{M} a'_{ij} \cdot log\left(\frac{a'_{ij}}{a_{ij}}\right) \\ &= \sum_{j=1}^{M} \left(a_{ij} - a'_{ij}\right) \cdot log\left(\frac{a_{ij}}{a'_{ij}}\right) \end{aligned} \quad (4)$$

For acquiring the overall distance between matrices (representing the time series), the values of J-divergence on different probability distributions have to be combined.

If we deem probability distributions for various starting states are equally important, we can simply average the J-divergence values derived from every pair of probability distribution in G and its counterpart in G'. Otherwise, we can define the weighted average of the J-divergence values as the final distance metric, where the weights reflect the importance of different probability distributions.

Further, the weights for the probability distributions can be determined automatically from a set of time series models (matrices) with known classes. Our idea, inspired from the work in [8], is that we retrieve the nearest models using a single J-divergence index and then we base the quality of retrieved models to assess the importance of the probability distribution, on which the J-divergence value is derived. More concretely, for every model in the collection, a set of nearest models are retrieved in terms of the J-divergence to yield a local alignment degree for that model. Secondly, the global alignment degree is calculated as the mean of the local alignment degrees for all models in the collection. Finally, the global alignment degree is assigned as the weight to the probability distribution in inspection.

3.2 Distance-Based Decisions

The distance metric developed for time series models can be used for two purposes. The first is to detect significant deviation of data evolution during the monitoring process (anomaly detection). The second is to further identify the class of the abnormal situation (if anomaly is detected) for medical diagnosis.

The anomaly detection can be made by comparing the model of time series in the latest time window with that of the preceding window. If the distance between them is sufficiently large, it indicates a potential abnormality since the probabilities of sate transitions have changed significantly in the new time window. Of course, the size of the window is an important parameter that affects the results of monitoring. One heuristic to find a proper value for that parameter would be gradually increasing the window size until the matrix of the time series becomes stable. Discovering optimal window sizes for different phases of the signal may improve anomaly detection.

For identification of the class of an abnormal situation, we advocate the application of case-based reasoning (CBR) which has been proved as a powerful methodology to solve new problems by learning from previous experiences [9]. CBR is based on the principle that similar problems have similar solutions. Therefore, given an abnormal time series in the latest window, we measure the distances of its matrix and the classified models (of time series) in the case library. The nearest models are thereby retrieved, and we resort to the classes of the retrieved models as the foundation to predict the class of the new abnormal situation. Feature selection [10] is sometimes needed here to identify the most important elements of the models for comparison, and fuzzy rule-based matching [11] can support more flexible criteria for assessment of the discrepancy between two time series models.

4 Conclusion

This paper puts forward a probabilistic method to deal with large data volumes in long term monitoring in health care. The key in our work lies in the conversion of longitudinal signals into shorter symbolic time series as well as depicting the stochastic property of the symbolic series with a Markov model. As the size of the Markov model only is related to the number of patterns rather than the length of the signal, it contributes with a big reduction of the data that needs to be stored and processed. We also illustrate that the Markov models derived from primary signals can be conveniently utilized to detect and classify abnormality in signals during the monitoring process.

However, it should be admitted that Markov models only consider the current state information. In the future work we are going to extend the current model to accommodate historical and contextual information to enable diagnosis and reasoning of higher accuracy.

Acknowledgement. This research is carried out within the research profile “Embedded Sensor Systems for Health”, funded by the Knowledge Foundation of Sweden.

References

1. Pantelopoulos, A., Bourbakis, N.: A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Trans. Sys. Man Cybern. Part C Appl. Rev. **40**, 1–12 (2010)
2. Milenkovi, A., Otto, C., Jovanov, E.: Wireless sensor networks for personal health monitoring: issues and an implementation. Comput. Commun. **29**, 2521–2533 (2006)
3. Lin, J., Keogh, E., Lonardi, S., Chiu, B.: A symbolic representation of time series, with implications for streaming algorithms. In: 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, pp. 2–11, San Diego, CA (2003)
4. Shahar, Y.: A framework for knowledge-based temporal abstractions. Artif. Intell. **90**, 79–133 (1997)
5. Bellazzi, R., Larizza, C., Riva, A.: Temporal abstractions for interpreting diabetic patients monitoring data. Intell. Data Anal. **2**, 97–122 (1998)
6. Funk, P., Xiong, N.: Extracting knowledge from sensor signals for case-based reasoning with longitudinal time series data. In: Perner, P. (ed.) Case-Based Reasoning in Signals and Images, pp. 247–284. Springer, Heidelberg (2008)
7. Kullback, S., Leibler, R.A.: On information and sufficiency. Ann. Math. Stat. **22**, 79–86 (1951)
8. Massie, S., Wiratunga, N., Craw, S., Donati, A., Vicari, E.: From anomaly reports to cases. In: Weber, R.O., Richter, M.M. (eds.) ICCBR 2007. LNCS (LNAI), vol. 4626, pp. 359–373. Springer, Heidelberg (2007). doi:10.1007/978-3-540-74141-1_25
9. Mantaras, R.L.D., et al.: Retrieval, reuse, revision and retention in case-based reasoning. Knowl. Eng. Rev. **20**, 215–240 (2005)
10. Xiong, N.: A hybrid approach to input selection for complex processes. IEEE Trans. Sys. Man Cybern. Part A Syst. Hum. **32**, 532–536 (2002)
11. Xiong, N.: Fuzzy rule-based similarity model enables learning from small case bases. Appl. Soft Comput. **13**, 2057–2064 (2013)