

Health Care Monitoring System using Efficient Smith Waterman approach

Anaswara Jayan¹, Divya²

¹(Computer Science, Adishankara College Of Engg And Technology, Kalady, India)

²(Computer Science, Adishankara College Of Engg And Technology, Kalady, India)

Abstract: Data mining techniques have been applied to sensor data in a wide range of application domains, such as healthcare monitoring systems, manufacturing processes, intrusion detection, database management, and others. Many data mining techniques are based on computing the similarity between two sensor data patterns. A variety of representations and similarity measures for multi attribute time series have been proposed in the literature. In this paper, we describe a novel method for computing the similarity of two multi attribute time series based on a temporal version of Smith–Waterman (SW), a well-known bioinformatics algorithm. We then apply our method to sensor data from an eldercare application for early illness detection GATSW. Finally, we demonstrated how TSW can be used in a various frameworks for detecting health patterns. Since WTSW could be potentially slow for eldercare applications, we proposed a genetic version of it.

Keywords: bioinformatics algorithm, bioinformatics algorithm

I. Introduction

In the developed countries, the rapid aging of the population garnered attention from health care providers. In the meantime, older adults are eager to live independently, regardless of complex chronic conditions such as frailty, dementia, and risk of falling. However, independent living may lead to delayed health assessments due to lack of monitoring, which is associated with poor long-term health outcomes. Late health assessment is an aggravating risk factor that usually occurs because of the fear of institutionalization and the failure of a physician's assessments. A possible solution to unreported health problems is to utilize an automatic health monitoring system that is able to detect and report signs of early illness. In the last decades, a growing number of projects have deployed ubiquitous sensor networks to monitor health of older adults.

MIT Place Lab, Georgia Tech's aware House, Honeywell's Independent Lifestyle Assistant, and The University of Missouri's Tiger Place are such examples. A variety of activity recognition (AR) methodologies that assess individual stability to complete

II. Related Works

D. J. Cook [1] presented Sliding window based approach, which is used to perform activity recognition in an online or streaming fashion. Sensor events are recorded and different activities are characterized by different window lengths of sensor events. Additional information in the form of the previous activity and the activity of the previous window is also appended to the feature describing sensor window. A. Mahnot [2] focuses to predict early signs of illness in older adults by using the data generated by a continuous nursing home monitoring system. It describes the possibility of employing a multiple instance learning suitable for training classifiers when the available data presents temporal or location uncertainties. Presence or absence of illness is manually assessed based on the nursing visit reports. Z. Haji hashemi [3] proposed a method based on non-wearable sensor sequence similarity and NLP. EHR nursing comments are provided to automatically notify the physician when health problems are detected. For each sensor sequence, health concepts are extracted from the nursing notes using Meta map, a NLP tool provided by UML (Unified Medical Language) system. Another method is to identify deviations in patterns of day-to-day activities of older adults to generate alerts to the healthcare providers for timely interventions [4]. Daily routines can be monitored by automated in-home sensor systems. The distribution of duration, number, and average time between sensor hits to define the confidence level for routine visit extraction. A hierarchical clustering is applied to extract periodic pattern. Compare and Count algorithm is used to quantify and track changes in activities of daily living and in the mobility of a smart home resident over time using longitudinal smart home sensor data [5]. Automated activity recognition algorithm is used to recognize a smart home resident's activities of daily living from the generated sensor data. Activity of Daily Livings such as eating, grooming, cooking, drinking, and taking medicine has been reported in the literature. Primarily differences in there ported AR methodologies are due to underlying sensing technology, machine learning models, and the realism of experimental setup. Regardless of these differences, the majority of

reported AR techniques were performed on scripted or pre segmented sequences of activities. One of the strengths of our paper is that it uses data collected in a real living environment for a relatively long period of time. Moreover, since we do not have a given time interval for experiments, our data set is continuously expanding. A key feature of our health-monitoring systems is the ability to continuously and unobtrusively collect information about daily activities of older adults. The system processes the acquired sensor data, infers activities such as “bathroom visit” or “out of the apartment,” and tries to detect changes in behavior of the monitored residents. Early signs of an impending illness or an exacerbation of an existing chronic condition may produce detectable behavior changes. The link between behavior changes and health patterns is based on the assumption that, for a given person, similar medical conditions result in abnormal behaviors, hence in similar sensor patterns. If a sensor pattern is not similar to previous ones observed in a similar context (time of the day, location), then we assume that it is produced by some health condition. While these assumptions might not in general be true for a younger, more mobile population, they are in general valid in elderly. To capture a person’s behavior, researchers employed various sensors, such as motion, Kinect, radar, sound, etchant produced multi attribute time series data sets (MATS). The mathematical basis of assessing similarity of behavior patterns based on sensor data is computing the similarity of multi attribute sensor time series. There are many approaches to computing MATS similarity depending on the application and the attribute type (discrete or continuous). Euclidian-based distance functions are reported for two continuous MATS sequences of equal lengths, whereas algorithms based on dynamic time warping (DTW) or longest common subsequence (LCSS) are applied on length mismatched sequences. Symbolic aggregate approximation, SAX, is a type of MATS representation that allows dimensionality reduction and lower bounds distance measures, based on DTW, on the original time series. For discrete sequences (symbolic data), such as protein sequences in bioinformatics, similarity measures such as Smith– Waterman (SW) have been proposed. Measuring the similarity of mixed sequences (some dimensions are continuous, such as heart rate values, and others are discrete, such as motion detector hits) in applications such as eldercare monitoring systems is a challenging task. Although similarity measures such as LCSS or S Ware originally designed for discrete data, they can be applied to continuous data by either modifying the original algorithm or converting the sequence from discrete to continuous format. There are two factors that distinguish our work from other related works: the dataset and the machine learning model.

Dataset: our dataset consists of real world living environments where there is no prescript scenario. Elderly residents live in their personal environment and conduct their daily routines with no instructions from researchers. Our dataset reflects the complexity of unconstrained real-world and it allows for testing the robustness of our AR approach.

Machine Learning Model: we describe a similarity measure for discrete time series and its applications to eldercare. Our approach does not require series data conversion to continuous format. Instead, we employ the natural arrangement of all different sensor hits together with their timestamps into a one dimensional sequence. We briefly introduced the idea of a modified SW framework for time stamped discrete sequences, Temporal Smith–Waterman (TSW), and performed some limited testing for three Tiger Place residents. In this paper, we extend our previous work by providing a scalable TSW version, showing other possible applications and testing TSW on larger datasets.

III. System Architecture

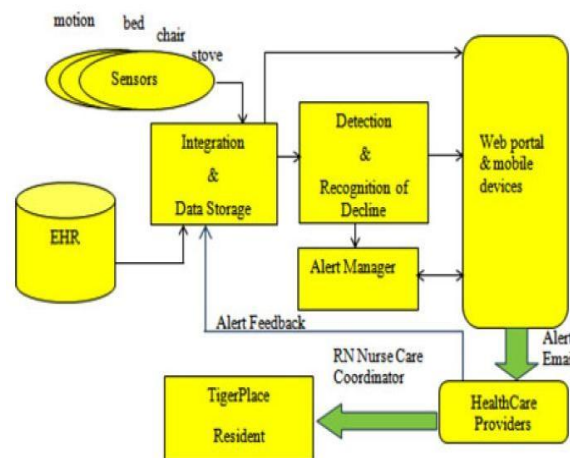


Fig.1

Fig. 1 shows the architecture of our data acquisition system. Various motion sensors have been deployed in different parts of the Tiger Place apartments: bedroom, kitchen, living room, and bathroom. For activity monitoring, we used wireless passive infrared (PIR) and pneumatic bed sensors networked using the X10 protocol. Each sensor sends an X10 signal that is logged with a time stamp in our sensor database. When triggered, the X10 motion sensor sends an ON signal and waits for at least seven seconds before motion will trigger the sensor to send another ON signal. After 60 seconds of no motion after an ON signal, the sensor sends an OFF signal. The bed sensor differs from the X10 motion sensors in which they do not check for motion using PIR sensing but instead use a pneumatic pressure strip to monitor various levels of three kinds of signals. The pneumatic strip captures vibrations from movement, breathing, and the resident's ball is to cardiogram, which is the mechanical effect of the heartbeat. The discretization of the bed sensor is performed in hardware by the bed sensor system. We are currently in the process of deploying a new bed sensor that provides continuous signals for all three previously mentioned quantities without the artificial quantization. The pilot sensor data is used with a total of 1902 sensor days and 2.1 million sensor hits, obtained from nine residents of Tiger Place. For each resident, we also retrieved visit notes about various physical and/or mental health complaints added by the nurses on site to the resident's EHR. In our dataset, there are fewer notes than sensor data (automatically logged for each resident), as for some days, there were no nursing comments. However, some residents may have multiple comments per day. In addition, each day was manually coded as normal or abnormal for each individual resident. Multiple clinical experts retrospectively reviewed the Tiger Place EHR data (nursing comments, vital signs, etc.) to flag days (denoted as "abnormal") with resident illness, but no inter rater agreement was computed. We use the abnormal days to test our early illness recognition (EIR) algorithm.

IV. Method

In this section, we describe the TSW similarity measure and its application.

A. TSW Similarity

In elderly population, decline of functional ability happens by slight changes in normal activities of elderly. We apply TSW measure on sensor sequences to track changes in sensor patterns and detect abnormal events. Abnormal events are defined by unusual sensor activity patterns that require a nurse's assessment of the resident. TSW measure is used to classify sensor sequences as "normal" or "abnormal" by employing the distribution of similar sensor sequences. TSW measure is applied to develop a novel approach to identify deviations in patterns of day-to-day activities of older adults to generate alerts to the healthcare providers for timely interventions. Daily routines, such as bathroom visits, can be monitored by sensor systems. We focus on bathroom visits, as an important daily routine. A hierarchical clustering approach using TSW is applied to cluster bathroom visits and extract frequent bathroom patterns.

Our bed motion sensor for which the bed motion was empirically divided [11] into four categories: less than 3 s, 3–7 s, 7–14 s, more than 14 s. The best value depends on the dataset. TSW considers the time difference between two sensor firings as a gap and, consequently, computes a gap using the time stamp associated to each symbol. The time is given in seconds and the constant c used in controls the time scale associated with the symbol firing. For example, in computing the similarity between patients represented by sequences of International Classification of Disease version 9 diagnoses, the time scale would probably be of the order of months where in our application, the time scale is of the order of minutes. The exact computation can be performed if a training set of known sequence similarities is available.

Given two discrete time series $S1 = \{(s_{11}, t_{11}), \dots, (s_{1n}, t_{1n})\}$ of length n and $S2 = \{(s_{21}, t_{21}), \dots, (s_{2m}, t_{2m})\}$ of length m , with $s_{1i}, s_{2j} \in \Sigma$, $1 \leq i \leq n$, $1 \leq j \leq m$, we can compute their TSW similarity, $TSW(S1, S2)$, using the following algorithm [3]:

$$H_{i0} = H_{0j}, \quad i \in [1, n] \text{ and } j \in [1, m] \quad (1)$$

$$H_{ij} = \max\{0, H_{i-1, j-1} + Sim(s_{1i}, s_{2j}),$$

$$\max_{k \geq 1} \{H_{i-k, j-W\Delta t}\},$$

$$\max_{k \geq 1} \{H_{i, j-k-W\Delta t}\} \quad (2)$$

$$k \geq 1$$

$$W\Delta t = g + c/|t_{1i} - t_{2j}| \quad (3)$$

$$TSW(S1, S2) = \max\{H_{ij}\} / \min\{n, m\} \quad (4)$$

H is a work matrix used in the dynamic programming which is trace back the alignment between $S1$ and $S2$ and Sim . It is a symbol similarity matrix it will reflect the compatibility between symbols. Example: *National Conference on "Emerging Research Trends in Electrical, Electronics & Instrumentation"* 31 | Page (ERTEEI'17)

$S_{im}(Bed\ Move\ ment1, BedMovement2) = 0.9$ sensor firings (see Table I), Then $S_{im}(BedMovement1, Cabinet) = 0$ to different type of sensors. G constant is a penalty for opening a gap, c is a penalty for extending a gap. The gap open penalty is the score taken away when a gap is created. The approximate value depends on the dataset. We set $g = 0$ it will limit our Search space. This would not have any effect on the performance of TSW on Tiger Placed at a set.

Similar to the traditional SW used in bioinformatics[6], TSW considers the time difference between two sensor firings as a gap and, consequently, It computes a gap penalty $W\Delta t$ using the time stamp associated to each symbol(3). The time is given in seconds and the constant c used in (3) controls the time scale associated with the symbol firing. While the TSW algorithm can be used for any symbolic time series, the time scale used in (3) is application dependent. The example, in computing the similarity between patients represented by sequences of International Classification of Disease version 9 (ICD-9) diagnoses [7], the time scale would probably be of the order of months where in our application, the time scale is of the order of minutes. The exact computation of c can be performed if a training set of known sequence similarities is available.

Sequence Search Using TSW Our main motivation for developing a sensor sequence similarity measure is to be able to compare human behaviors and activities. One simple approach is to divide the entire sensor sequence acquired from a Tiger Place apartment into fixed intervals such as days and use TSW to compute the similarity between them. However, since a given activity can be performed at different times of the day, week, or year, we should be able to search the entire database for a given pattern. To address this question, in this paper we employ a window-based TSW (WTSW) method. WTSW algorithm uses a sliding window approach to find the similar subsequences to a user defined (query) subsequence based on TSW similarity measure. In our case, examples of user defined sub sequences are bathroom visits or food preparation activities.

C. Speeding of WTSW Using a Genetic Algorithm Approach

Similarity search is a subtask of most of time-series data mining algorithms. The time complexity of search algorithm has always been the bottleneck for large time series. Depending on the properties of the similarity measure used in the search, different techniques have been proposed in the literature to speed up searching process. In this paper, we use genetic algorithm as a first try to speed up WTSW and leave other possible speeding up methods for future work. We acknowledge that using genetic algorithm to speed up the searching process is not the best solution. We applied genetic algorithm as a solution that mimics the sliding window feature of WTSW method and produces satisfactory results in a reasonable time. Lower boundary searches and early abandoning may speed up TSW. For computing the similarity of two subsequences, early abandoning keeps track of the minimal distance found so far when computing the distance of two new elements of the subsequences. If it approximates that the current minimum will be exceeded, it terminates. Although the early abandoning speeds up computing the similarity of two subsequences, it still needs to perform all pair wise comparisons. Lower bounding techniques perform well on uniform scaling distances.

Chromosome: A chromosome is a sequence of sensor firings within a specified time interval, namely a set of genes.

Population: A population of size p is a set of p different chromosomes. In any genetic algorithm, we use a fitness function to determine which chromosomes survive from one generation to another.

Fitness function: In GATSW, the fitness of a chromosome is calculated by assessing the similarity of the chromosome to the given user defined subsequence.

$Saj(position) = [Sa - 1j(position) + Fitness - threshold]$.

Saj(position): It is the position of chromosome

Sj: iteration a, and mutation rate and threshold are parameters of mutation process.

Mutation: The amount of mutation depends on the fitness of the parent chromosome. The fittest parent chromosome changes its position slightly, whereas the parent chromosome with the lower fitness changes its position drastically.

The parameter *threshold* demonstrates the best similarity score that is equal to the fitness of a perfect match. It is to control the rate of selection of children chromosomes for the next generation. The parameter *mutation rate* controls the influence of its fitness on the chromosome's position in the next generation. The more dissimilar is a chromosome from the user defined subsequence, the higher the mutation patterns.

Our goal is to detect events that a resident needs to be assessed. Abnormal day is a subjective matter that differs for each elderly resident. That means an event that is considered abnormal for a resident might be a normal event for another resident. For example, some residents have more sound sleeps during the night, for these residents multiple bathroom visits during midnight is considered as an abnormal event, whereas for another group of residents, several midnight bathroom visits are considered normal. The labeling of

normal/abnormal events depends on the functional ability of residents, their health status, taken medications, etc.

```
1 . Initialize
   Initialize the mutation rate and the generation size of population.
   Initialize population
2 . while(condition for termination)
   {
     1.using TSW a evaluate the fitness
     2.select the parent chromosome and then mutate it with them
     3.select the new population
3  select the fittest chromosome
```

V. Conclusion

In this paper, we described a similarity measure, TSW, for MATs data. To scale up TSW for real data, we introduced a window-based algorithm, WTSW, which uses TSW to search for the best match in long MATS. The proposed method provides a natural segmentation of the sequences that provides finer grained illness recognition solution. Since WTSW could be potentially slow for eldercare applications, we proposed a genetic version of it, GATSW. Finally, we demonstrated how TSW can be used in a various frameworks for detecting health patterns. For future directions, we suggest researchers to use TSW with multiple methods together with a fusion methodology to reduce the false alarms rate.

References

Journal Papers:

- [1] C. K. Narayanan and D. J. Cook, "Activity recognition on streaming sensor data," *Pervasive Mobile Comput.*, vol. 10, pp. 138–154, 2014
- [2] M. Popescu and A. Mahnot, "Early illness recognition in older adults using in-home monitoring sensors and multiple instance learning," *Methods Inf. Med.*, vol. 51, no. 4, pp. 359–367, 2012.