



THE SIMILARITY QUANTIFICATION OF MULTIDIMENSIONAL TIME SERIES DATA SETS IN SURVEILLANCE APPLICATIONS FOR HEALTH MONITORING SYSTEM

Dr. Muhammad Ashfaqul Hasan

Department of Computer science & Engineering
Malla Reddy Institute of Engineering & Technology
Dhulapally, Secundrabad - 500100.

Dr.P.Kiran kumar Reddy

Department of Computer science & Engineering
Malla Reddy Institute of Engineering & Technology
Dhulapally, Secundrabad - 500100.

Abstract: In the last decade, Data mining techniques have been applied for sensor data in a wide range of applications. Like health care monitoring systems, manufacturing process. Intruder detection, database management and other. A lot of data mining engineering is based on the calculation of the similarity between two models of sensor data. A number of representations and Equality measures for multi - assign time series was suggested in the literature. In this paper, we describe a new way of calculating whether two similarities in the series of multiple series are based on the temporal version of Smith-Waterman (SW), a known information algorithm. Next, we apply our method to detect data on the demand for care of the elderly to early detection of the disease. Our procedure absorber is difficulties linked to the data uncertainty and aggregation that often occurs during treatment sensor data. The trials will take place one aging-in-place installation, Tiger Place placed in Columbia, MO. To validate our method we used data on nine no-portable a sensor for one-p located in TigerPlace apartments, combined with information of one Electronic Health Record (EHR). We deliver a set of experiments studying the temporal version of SW properties, with experiences on TigerPlace dataset.

Keywords: multi-attribute time series, time series data mining, Smith-Waterman algorithm, genetic algorithm, EHR.

1. INTRODUCTION

The American Academy of Nursing (AAN), asked Researchers to propose new solutions for change and Improving standards of care for the elderly [1]. Developed countries have achieved rapid population aging Attention of health providers. Published statistics show (65 years and over) from 13 per cent 2010 to 19% in 2030, while the relationship between working age (Ages 15 to 64) and older people 4.3 to 2.3 [2]. At the same time, older people are Anxious to live independently, regardless of chronic complex Conditions such as weakness, dementia and fall hazards. However, independent lifestyles may lead to delays and health assessments due to lack of monitoring, which is Associated with long-term poor health problems [3]. Late Health assessment is a risk factor that is stressed as normal this happens because of the fear of institutionalization Lack of medical evaluation [4]. Possible solution to unreported health problems are the benefits of automatic health a monitoring system that can detect and report characters early illness.

In recent decades, there has been an increasing number of projects Network sensor implemented everywhere to monitor health Elderly. MIT Place Lab, House, Independent lifestyle assistant in Honeywell, WA Are examples of such [5, 6, 7, 8]. A number of activity identification Methods of assessing the ability of individuals to complete them The activities of everyday life such as eating, Cooking, drinking and medicine have been reported in Literature [9,10,11,12]. Key differences in reports The Activity Recognition methodologies are attributable to basic sensor technology, Automated and realistic learning models for experimental groups On [13, 42]. Regardless of these differences, the majority Text-based techniques have been implemented or predicted Sequence of activities. One of our strengths of this

paper which uses data collected in a realistic environment over a relatively long period. Incidentally, because we did not have a specific test period, our datasets constantly expanding.

An important part of our health surveillance systems is the ability to collect confidential information about daily of older activities. The system processes acquired sensor data, accompanying activities such as "bathroom visit" or "apartment", and "attempts to make changes to the behavior of accompanying residents". The first signs of an imminent illness or the worsening of the current chronic condition can lead to behavioral changes that can be detected. The relationship between behavioral changes and health patterns is based on similar (abnormal) behavior in comparable sensor models.

If the sensor model does not match, so that such similar terms can be seen earlier (from time to time), it is assumed that it is a certain state of health (unknown). Is prepared, although these beliefs are not common to a small population, more exercise, they generally apply to the elderly [14, 15, and 16]. For the behavior of a person, researchers have used different sensors, such as movement, range, radar, sound, etc. [8] a series of multiple datasets (MATS) has been designed. The mathematical basis calculates the equality of multifunctional sensor series for a comparable evaluation of sensitivity-based behavioral models.

There are several ways to calculate equality MATS depending on the application and attribute type. The remote Euclidean information function [17] for two joined matrices together sequences of equal lengths, while using dynamic timing algorithms (DTL) as a function of sequence lengths are only not identical [18, 19, 20, 21, 22]. It is a type of representation mat that reduces on the basis of limit values DTL and dimensions to the original time series [41]. For

discrete sequences, such as protein sequences in bioinformatics, measurements have been proposed for the equality between Smith-Waterman (SW) [23].

Measure equality between mixed settings (some problems are consistently consistent - such as heart rate values and other dynamic animation moves). A major care monitoring system is very difficult in applications. Although equal measures such as LCSS [33] or SW was originally designed for discrete data, They can be used for current data by changing original algorithm [18] or convert the sequence of discrete to continuous format [15, 22, 24, 25]. LCSS [33] or SW was originally designed for discrete data, They can be used for current data by changing original algorithm [18] or the conversion of discrete to continuous format [15, 22, 24, 25].

There are two factors that characterize our proposed work from other related work: data set and automated learning model.

Data Set: Data sets are made up of real environments where there is no script. The elderly live in their own environment and carry out their daily routine without instructions from researchers. Our data set reflects the complexity of the real world without restrictions and allows us to test the durability of our R approaches.

Automatic learning Model: We describe a measure of equality for a discrete time series and its applications in elderly care. Our approach requires serial data conversion in continuous format. Instead we use the natural setting of all the different sensor images with their time stamps in a one-dimensional sequence. Our article is structured as follows. In section II we describe our system architecture and data sets. In section III we explain our method and its application to AR in the care of the elderly. In the episode IV, we provide in-depth experience to investigate our approach. Finally, we discuss in section V consequences of our approach and conclusion.

2. ARCHITECTURE OF THE SYSTEM

We set up our integrated monitoring system TigerPlace, an aging onsite facility from Columbia, MO, with the approval of the IRB from the University of Missouri. We decided to Use only non-portable sensors for surveillance purposes after the management, several focus groups in 2004 with seniors residents of TigerPlace [8, 37, 38]. The surveillance started in autumn 2005. Currently we have 50 apartments online.

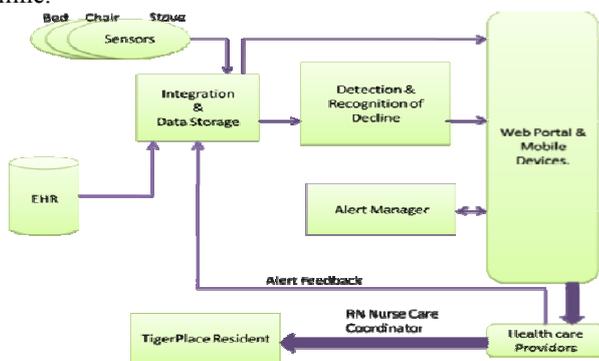


Fig. 1. Sensor network architecture for health care system.

Different motion sensors are used in different parts of the apartments: bedroom, kitchen, living room. In order to

control the activity, we used passive infrared and infrared sensors using the X10 protocol.

Each sensor sends an X10 signal recorded in our sensor database with a timestamp. When the X10 motion sensor is activated, an activation signal is transmitted and waits at least seven seconds before the movement leads the sensor to send another ON signal. After 60 seconds no after an ON signal, the sensor sends an OFF signal. Table I shows the sensors and their identifiers used here. The bed sensor can measure the movements of the bed (ID1-4), breathing (ID 5-7) and pulse (ID 8-10). The bed sensor differs from X10 motion sensors in which they are not check the movement using PIR detection, but use a pneumatic pressure band to control different levels of three types of signals [39]. Pneumatic tape attracts vibrations of movement, breathing and inhabitant ballistocardiogram, which is the mechanical effect of the heart rate the discretion of the bed sensor is performed in hardware by the bed sensor system as described in [11]. We are currently implementing a new bed sensor [9] which offers for all three continuous signals said quantities without artificial quantification.

TABLE I. SENSOR IDENTIFIERS USED IN THIS PAPER

ID	Type: Sensors	8. Pulse1 (low) [*]	16. Stove TempHigh ^{**}
1. BedMovement1 [*]		9. Pulse2 (ok) [*]	17. Stove TempLow ^{**}
2. BedMovement2 [*]		10. Pulse3 (high) [*]	18. Drawer ^{**}
3. BedMovement3 [*]		11. BedroomMotion ^{**}	19. Cabinet ^{**}
4. BedMovement4 [*]		12. Living Motion ^{**}	20. Cup Cabinet ^{**}
5. Breathing1 (low) [*]		13. Bath Room ^{**}	21. Refrigerator ^{**}
6. Breathing2 (ok) [*]		14. Off Chair ^{**}	22. Plate Cabinet ^{**}
7. Breathing3 (high) [*]		15. On Chair ^{**}	23. Silverware Drawer ^{**}

^{**} indicates X10 sensors and ^{*} indicates bed sensors.

Table II shows a small sequence (with a length of 8) of the sensor store the data included in the system log file Resident # 3 (the person is probably out of bed and went in his living room). Table III describes the pilot sensor data used in this study with a total of 1902 sensor days and 2.1 millions of soundings, obtained from nine inhabitants of TigerPlace. For each resident, we also assessed the comments on various complaints that have been added to physical and / or mental health care by nurses on location to the occupant of the Ministry of Economic and Social Affairs. In our dataset there are fewer observations of sensor data (automatically connected to each evaluator) because it did not breastfeed for a few days, but some assessors (eg resident 3) can have multiple reactions per day.

In addition, each day is coded manually as normal or abnormal for each resident. Many clinical experts have retrospective data (nursing comments, biomarkers, etc.) scientific days (so-called "abnormal") with episodes of the disease listed (column 3 of Table III), but no agreement has been calculated between the router [46]. We use abnormal days to test our early recognition test algorithm for diseases.

TABLE II. SENSOR SEQUENCE SNIPPET FOR RESIDENT#3

ResidentID	SensorID	Year	Month	Day	Time
3	3	2005	10	5	12:34:38
3	2	2005	10	5	12:36:52
3	2	2005	10	5	12:37:04
3	2	2005	10	5	12:37:11
3	1	2005	10	5	12:37:26
3	1	2005	10	5	12:37:28
3	12	2005	10	5	12:38:32
3	12	2005	10	5	12:41:18

TABLE III. PILOT DATASET

Resident Code	Number of days	Abnormal days
1	440	83
2	463	44
3	499	335
4	225	130
5	81	60
6	72	55
7	45	15
8	39	10
9	38	15

3. METHOD

A. Temporal similarity of Smith-Waterman

We define a time series S as a set of n couples (si ti), that is to say S = {(s1 t1)... (sn tn)}. Each pair (si ti), 1 ≤ i ≤ n has two components: a sensor signal if belonging to an array of symbols Σ and a timestamp ti which represents the moment when he was included in the database. The alphabet Σ is a set of identifiers we use a multidimensional sensor to represent the time series. In our case, Σ includes the symbols shown in Table I.

Although some sensors can be like motion detectors naturally described by symbolic data, theirs, such as the bed sensor, can be quantified. This was our case bed motion sensor for which the movement of the bed was empirical divided [11] into 4 categories: less than 3s, 3-7s, 7-14s, plus only 14 s (sensor ID 1-4 in Table I).

Given two discrete time series S1 = {(s11 t11)... (s1n t1n)} of length n and s2 = {(s21 t21), ..., (s2m t2m)} of length m, with s1i, s2j ∈ Σ, we can calculate their TSW similarity, TSW(S1, S2), using the following algorithm [26]:

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

$$H_{ij} = \max\{0, H_{i-1, j-1} + \text{Sim}(s_{1i}, s_{2j}), \max_{k \geq 1} \{H_{i, j-k} - W_{\Delta t}\}, \max_{k \geq 1} \{H_{i-k, j} - W_{\Delta t}\}\} \tag{2}$$

$$W_{\Delta t} = g + c|t_{1i} - t_{2j}| \tag{3}$$

$$\text{TSW}(S_1, S_2) = \max\{H_{ij}\} / \min\{n, m\} \tag{4}$$

Where H is a working matrix used in dynamic programming plot the best alignment between S1 and S2 and Sim is a symbol resemblance matrix that reflects compatibility between the symbols. For example, Sim (BedMovement1, BedMovement2) = 0.9, since these are two bed motion sensors fireworks (see Table I), while Sim (BedMovement1, Cabinet) = 0 because they belong to a different type of

sensors. The constant g is a punishment for opening a gap while it's a punishment widens a gap. The opening penalty has been removed When an interval is created the best value depends on the dataset. We define g = 0 to limit our search space. It would not have effect on the implementation of TSW on the TigerPlace dataset. As with traditional SW used in bioinformatics [23] TSW considers the time difference between two shots as a gap and calculates a gap penalty WΔt using the timestamp attached to each symbol (equation 3). Time is given in seconds and the constant c used in Equation 3 is controlled by the time scale associated with the symbol lights. Although the TSW algorithm can be used in any symbolic time series, the schedule used in Equation 3 depends on the application. For example, in a correspondence account between patients representing the classification of diagnostic sequence 9 (ISD-9) [27], the time of orders can be months when the application is a measure. From time to minutes. The exact calculation of c can be performed if there is a set of known game sequence training is available.

B. Sequential search using TSW

Our main motivation for developing a series of equivalent measurement sensors is to compare human behavior and activities. A simple method [26] is to divide the entire sensor sequence for fixed, flat periods, such as days (sets of 24-hour sensor sequences) and to use TSW to calculate the matches between them. However, since a certain activity can be performed at different times of the day, week, or year, you must be able to search the entire database for a specific reason. To solve this problem, we use a window based on TSW (WTSW) in this article. This algorithm uses a policy cursor window to search for submenus that resemble a specific user (query) based on the equality metric. In our case, examples of user-defined subdivisions are bath visits or food preparation activities.

Suppose we point out order of sensor shots by D and the defined user Query by Q. To find the most comparable subsequence to Q in D, we drag a window of magnitude tΔ to D.

The size of a window, tΔ, must be greater than the difference times in which S_{q1} and S_{qn} were observed in Q, it is tΔ > (t_{qn} - t_{q1}). The consecutive windows overlap with an interval of time (tΔ - wΔ). To exclude trivial agreements (subsequences of Q), we do not consider subsequences with time stamps row with Q. After extracting all non-trivial extractors Subsequences give us a similarity score to all subsequences using the TSW equality measure. We then choose the subsequence with the highest score as the most comparable sub-order to the user defined subsequence Q. pseudo code of the WTSW algorithm is shown in Figure 2. The WTSW has faced two challenges. The first challenge is speed. The TSW algorithm takes a lot of time. The second Challenge launches the initialization of the WΔ parameter. If WΔ is small then the results are more accurate, but the calculation time is high. The maturity can be reduced by reducing the overlap; however, some relevant subsequences cannot be retrieved.

```

Step 1: Initialization
Initialize parameter  $W_A$ 
Calculate parameter  $t_A$  using user defined subsequence  $Q$ ;
Form a window of size bigger than  $t_A$  from the start point of
data set sequence;
Initialize subsequence set  $W = \Phi$ ;
Step 2: while (termination condition)
{
Form  $W_i$  = the subsequence (i) of size bigger than  $t_A$ 
Calculate the similarity score of  $W_i$  using TSW algorithm
 $Score(W_i) = TSW(W_i, Q)$ ;
Add subsequence  $W_i$  to the set  $W$ ;
Slide the window at the rate of  $W_A$  to get the next
subsequence;
}
Step 3:
Pick the highest score  $W_i$  for the set  $W$ ;
    
```

Fig. 2.Pseudo code for WTSW method

The TSW similarity of two parts of lengths m and n, O (m n) has complexity. If the length of D r is then we can needed between r/n (no-overlap) and r-n (overlap n-1) TSW assessments. If $r \gg n$, then the upper bound of complexity rn^2 , which can be high for a large r. For example, "bathroom visit" behavior usually has 200 symbols (n = 200) and the 5 Annual behavioral sequences can be 2 million symbols long (R = 2,000,000). To meet these two challenges, we use the genetic algorithmic approach described in the next section.

C. WTSW using a genetic based algorithm approach

Comparison is a sub-task of most mining algorithms in time series. Quality search algorithm search time has always been bottleneck for large time chains. Target uses the searches; different techniques have been suggested in the literature to speed up the search process [19, 20, 21, 22, and 40]. In the paper of Gaze we have genetic algorithm as a Speed up Explorer WTSW in the possible speed methods for future work. We acknowledge the use Genetic algorithm to accelerate the search process is not the best solution we have used the genetic algorithm as a solution mimics the slider window of the WTSW method and delivers satisfactory results within a reasonable period of time. Lower border investigation and early abandonment [22] TSW. To calculate the similarity of two sub-sequences early giving up to the minimum distance found so far when calculating the distance of two new elements of the subsequences. If it is close to the current minimum will be exceeded, it ends. Although it started rate the calculation of the similarity of two Subsequent, he still has to complete all pairs comparison. Lowest limit techniques, on the other hand, perform well on uniform scale sizes. For larger values of n, however, the algorithm is still unbearable.

Note that the purpose of this article is to study the effectiveness measure of TSW on the data collected from TigerPlace and are practical utility to solve the problem of elderly care in the real world like detecting daily routines and abnormal events detection In what follows, we define the most important GA parameters.

Gene: Given a subsequence S of length m, where $S = \{(s_1 t_1) \dots (s_m t_m)\}$, a gene is a pair $(s_i t_i)$, $1 \leq i \leq m$. Identify the sensor as who shot down at the moment t_i . In other words, a gene in GATSW is a window that indicates the triggering of a sensor and time patch.

Chromosome: A chromosome is a series of sensor fires in a certain time interval, namely a set of genes. Given a time

series T of length n, a subsequence S_i of T is ordered the sampling of the length m,

$$S_i = \{(s_i t_i), \dots, (s_{i+m-1} t_{i+m-1})\} \quad \text{Where}$$

$$m \leq n, 1 \leq i \leq n - m + 1..$$

Inhabitants: An inhabitant of size p is a series of unlike p-chromosomes. In each genetic algorithm we use a fitness function: conclude which chromosomes live on from one generation (iteration) to another.

Strength function: In GATSW, the capability of a S_i chromosome is designed by evaluating the chromosome conformity with the user-defined Q subsequence. The appropriateness of the S_i chromosome is strength $(S_i) = TSW (S_i, Q)$.

```

Step 1. Initialization
I. Initialize the size of population and generation, and
mutation rate;
II. Generate initial population;
Step 2: While (termination condition){
I. Evaluate the fitness of individuals using TSW;
II. Select parent chromosomes;
III. Mutate parent chromosome;
VI. Select new population;
}
Step 3: Pick the fittest chromosome as the final solution;
    
```

Fig. 3.The GATSW algorithm

Mutation: Given a chromosome S_j changes a mutation with respect to S_j the starting position according to its relevance.

$$S_j^a(position) = [S_j^{a-1}(position) + (mutation_{rate}) \times (|Fitness(S_j^{a-1}) - threshold|)] \quad (5)$$

In 5 equation, $S_j^a(position)$ is the location of chromosome in the repetition a, $mutation_{rate}$ and $threshold$ are parameter of change process. $Fitness(S_j^{a-1})$, is the fitness of chromosome in the previous repetition (a-1). The amount of change depends on the strength of the parent chromosome. The fittest parent chromosome changes its location a little, whereas the parent chromosome with the lower health changes its location significantly. Fig. 4 represents the simulated code of GATSW.

The stricture *threshold* demonstrates the best likeness score that is equal to the strength of a ideal match. We use this parameter in transformation to control the rate of variety of children chromosomes for the next generation. The parameter $mutation_{rate}$ controls the persuade of its strength on the chromosome's location in the next generation. The more different is a chromosome from the user defined subsequence, the higher the transformation.

D. Health model Prediction: TSW Applications:

An important purpose of our TigerPlace research is to predict the change in the health status of residents based on sensor data produced by the home surveillance system. The ability of sensor networks to predict health profiles using logistic regressions, multi-instance learning and temporal clustering has already been studied in aggregated sensor data [28, 29]. In [42, 43] the authors provided new methods

for using sensor network data to track changes in daily routines that can be used in assessing cognitive health. Multiple characteristics such as sensor count aggregation and time characteristics were extracted from sensor sequences to quantify the performed activities. An SVM student and search algorithm were used to detect activities and follow changes. Linear regression and the Gaussian process were used in [44] to name functions extracted from aggregated sensor count to resident functional capacity. With the same data aggregation approach [45], authors have proposed a statistical method for estimating the average time spent by a resident in each room of his apartment. Our proposed approach is different so that the number of sensors is not aggregated. In this article we investigate the ability of the TSW method to detect health patterns.

1) Detection of Anomalous Activities

In older patients, impairment of functional ability occurs due to slight changes in the normal activities of the elderly. We apply TSW measurement to the sequence of sensors to track changes in sensor models and detect abnormal events. Non-natural events are defined by the unusual activity patterns of sensors that require nurse assessment by the client. This measurement is used to classify sensor sequences as "normal" or "abnormal" using the distribution of similar sensor sequences.

We affirm that if a particular sensor sequence does not resemble the normal sensor sequences previously observed; this is an indication of a possible abnormal event [30]. We use TSW measurement because we want to find similar behaviors at different times at the same time [30].

Algorithm: Abnormal events detection using TSW similarity distribution.
Input: NS : a set of normal sensor sequences of size n ; S_x : an unknown sensor sequence.
Output: Label of S_x as normal or abnormal sequence;
Steps:

1. Using all sequences in set $NS=\{S_i\}_{i=1,n}$ build the similarity matrix $S=\{S_{ij}\}_{i,j=1,n}$ by calculating their pair-wise similarities using TSW method.
2. Assume a distribution function that best suites data points in matrix S .
3. For the sequence S_x , computing its similarities to all normal sequences $S_1, \dots, S_n, \{S_{ix}\}_{i=1,n}$, and find the maximum similarity of $\{S_{ix}\}, S_{x,max}$.
4. Calculate the confidence of the similarity, $C(S_x)$, using the similarity distribution function in step 2.
5. If $C(S_x) < \delta$ then classify the unknown sensor sequence as an abnormal event, otherwise as a normal event.

Fig. 5. Abnormal events detection using TSW similarity distribution

E. Assessment Metrics and Experiment Setup

Our goal is to detect events (days) that should evaluate an evaluator (called an abnormal day). The abnormal day is another personal question for every elderly resident. It means if an event that is considered abnormal for the resident can be a normal event for another resident. For example, some residents have more nap because they are considered different visits to the bathroom during the night as an unusual event.

For another group of people, different visits at midnight are normal. Marking normal / abnormal events depends on the functional ability of the population, state of health, medicine, etc. We report the Tsu's performance in the early identification of the disease using different approaches in section IV. We use the following statistics in our experiments.

$$Precision = \frac{tp}{tp+fp} \tag{6}$$

$$Recall = \frac{tp}{tp+fn} \tag{7}$$

$$F - measure = 2 * \frac{Precision+Recall}{Precision+Recall} \tag{8}$$

Here, t_p is the number of true positives, t_n is the number of true negatives, f_p is the number of false positives, and f_n is the number of false negatives.

4. RESULTS

In this section, we demonstrate the performances of proposed methods on TigerPlace dataset.

A. Abnormal Events Detection

Our goal is to build an early illness recognition (EIR) system for TigerPlace residents using sensor pattern similarity. For a given resident with available historical sensor data, the EIR system searches for similar patterns to the current recorded activity. The high similarity of the current sensor pattern to one in the past that was related to an abnormal health event indicates the possibility that the previous illness reoccurred. The health pattern prediction experiments are performed using a k -nearest neighbor and a leave-one-out cross-validation approach. For each unknown sensor sequence S_i , we compute the distances (using WTSW and GATSW for comparison) to all other past sequences. Then, we select k ($k \sim \sqrt{n}$, n is the number of samples in the training dataset for each resident) most similar sequences together with their "normal" or "abnormal" labels. Finally, the classifier predicts the label of S_i based on the label of its k -nearest neighbors using the following heuristic: if any of the labels of the k retrieved sequences is "abnormal" then we label S_i as abnormal; if all k labels are normal, then S_i is labeled as normal. This heuristic was motivated by the fact that in medical applications the cost of a missed detection far outweighs the one of a false alarm.

Here, for each resident, we analyze the performance of WTSW and GATSW in health pattern recognition in terms of retrieving "normal" and "abnormal" days. Table IV shows the parameter setting, and Table V compares the average performance of WTSW and GATSW on all residents of TigerPlace dataset. In this experiment, different parameter settings have been tested and the setting that provides the best performance has been reported in Table IV. Column "Time" presents the maximum execution time over all residents' data for each method in seconds.

TABLE IV. PARAMETER SETTINGS OF WTSW AND GATSW METHODS

WTSW Method	W_d (min)	t_d (hour)	Time(sec)
	20	2	14
GATSW Method	Population Size	Generation Size	Time(sec)
	50-60	50-60	7

TABLE V. PERFORMANCE COMPARISONS ON WTSW AND GATSW METHODS

WTSW Method	Precision	Recall	F-measure
	0.66	0.6	0.66
GATSW Method			
	0.77	0.71	0.75

Generally, by increasing $t\Delta$ or $W\Delta$ the *F-measure* decreases. The reason is that in a larger time interval it is less likely that residents do the exact same thing at the exact same time of every day. Moreover, larger $W\Delta$ results in fewer window candidates and decreases the performance of prediction. In GATSW experiments with 60 chromosomes and after 60 generation, it achieves 0.75 of F-measure in almost half a time that WTSW takes. Even though our dataset has about two millions of sensor hits and stretched over three years, this experiment shows that for larger data sets (billions of hits) GATSW can still be time demanding. This can happen if the search will go, for example, across multiple residents with similar health conditions.

We acknowledge that the imbalance data set may in some cases reduce the performance of the classifier. We avoided normal over / under sampling strategies for the following reasons. In part sampling, the majority class would decrease the F-size in some cases where we only have a few abnormal events; therefore, the classifier should be formed with little global samples. Oversampling of the minority class is difficult because we do not understand the problem enough to oversimplify (not sure how to adjust / generate abnormal events). To solve the problem, we used a classifier in a class (see section IV-B.1) that uses the majority class only to identify abnormal events.

5. CONCLUSION

In this paper, we described an agreement to increase TSW's real data; we introduced a Windows-based algorithm, TSW, used by TSW to find the best match in Mats for a long time. The proposed method provides a likely sequence segmentation that provides an acknowledgment solution for micro-grains. Since WTC may be slow for older care applications, we have proposed a genetically modified version of GATSW.

Finally, we have shown how TSW can be used in different frameworks for detection of health plans. We have tested our algorithms on different datasets: two synthetic, one obtained in TigerPlace and the other obtained from [32]. On the TigerPlace data set, we obtained the abnormal F-rating predictions of 0.75 for all residents in this study. For future clues, we suggest that researchers use TSW using different methods and a merger methodology to reduce false alarms. On the sensor data set obtained from [32], we obtained a measurement F of 0.82 on attitude recognition.

REFERENCES

- [1] Editor, "TigerPlace: An Innovative 'Aging in Place' Community," The American Journal of Nursing, 2013, 113(1), pp. 68-69.
- [2] KV Grayson, AV Victoria, "The next four decades: the older population in the United States: 2010 to 2050", U.S. Dep. of Commerce, Economics and Statistics Administration, U.S. Census Bureau, 2010, pp.1-14.
- [3] JS Goodwin, B Howrey, et. al. "Risk of Continued Institutionalization after Hospitalization in Older Adults". J Gerontol A Biol Sci Med Sci. 2011, pp.1321-1327.
- [4] TL Hayes, M Pavel, et al. "An unobtrusive in-home monitoring system for detection of key motor changes preceding cognitive decline", 26th Annual Intl. Conf. of the IEEE EMBS, 2004, pp. 2480-2483.
- [5] S Intille, K Larson, et al. "Using a live-in laboratory for ubiquitous computing research. Proc. of Pervasive Health. 2006.
- [6] CD Kidd, RJ Orr, et al. "The Aware Home: A living laboratory for ubiquitous computing research". 2nd International Workshop on Cooperative Buildings-CoBuild'9, 1999.
- [7] KZ Haigh, LM Kiff, G Ho. "Independent Lifestyle Assistant: Lessons learned". Assistive Technology. 2006; pp. 87-106.
- [8] M Skubic, G Alexander, et. al. "A Smart Home Application to Eldercare: Current Status and Lessons Learned." J Technology and Health Care. 2009, pp.183-201.
- [9] D Heise, M Skubic. "Monitoring pulse and respiration with a noninvasive hydraulic bed sensor." 32nd Annual Intl. Conf. of the IEEE EMBS. Buenos Aires, Argentina. 2010.
- [10] R Beckwith. "Designing for ubiquity: The perception of privacy." Pervasive Computing. 2003, pp. 40-6.
- [11] D Mack, M Alwan, et. al. "A passive and portable system for monitoring heart rate and detecting sleep apnea and arousals: Preliminary validation." Distributed Diagnosis and Home Healthcare (D2H2). 2006.
- [12] MJ Rantz, K D Marek, et al. "A technology and nursing collaboration to help older adults age in place." Nursing Outlook; 2005, pp.40-45.
- [13] CK Narayanan, DJ Cook. "Activity Recognition on Streaming Sensor Data", Pervasive and Mobile Computing, Vol. 10, 2014, pp. 138-154.
- [14] M Popescu, E Florea. "Linking Clinical Events in Elderly to In-home Monitoring Sensor Data: A Brief Review and a Pilot Study on Predicting Pulse Pressure." Computing Science & Engineering, 2008, pp.180-199.
- [15] M Popescu, A Mahnot. "Early Illness Recognition in Older Adults Using In-Home Monitoring Sensors and Multiple Instance Learning." Methods of Informatics in Medicine. 2012, 51(4), pp.359-67.
- [16] Y Lee, D Ho, M Popescu. "Microphone Array System for Automatic Fall Detection". IEEE Trans. on Biomedical Engin. 2012, pp.1291-1301.
- [17] T Kahveci, A Singh, A Gurel. "Similarity Searching for Multi-attribute Sequences." SSDBM. 2002, pp.175-184.
- [18] GA H. Ten, et al. "Multi-Dimensional Dynamic Time Warping for Gesture Recognition." 13th conf. of the Advanced School for Computing and Imaging. 2007.
- [19] M Vlachos, M Hadjieleftheriou, D Gunopulos, E Keogh. "Indexing Multi-Dimensional Time-Series with Support for Multi Distance Measures." SIGKDD. 2003, pp.216-225.
- [20] P. Sanguansat. "Multiple Multidimensional Sequence Alignment Using Generalized Dynamic Time Warping." WSEAS Transaction on Mathematics, Issue 8; Volume 11; 2012, pp. 684-694.
- [21] H Ding, G Trajcevski, et al. "Querying and mining of time series data: experimental comparison of representations and distance measures." PVLDB, 2008, 1(2), pp. 1542-1552.
- [22] T Rakthanmanon, B Campana, et al. "Searching and Mining Trillions of Time Series Subsequences under Dynamic Time Warping", SIGKDD, 2012, pp. 262-270.
- [23] TF Smith, MS Waterman. "Identification of common molecular subsequences." J of Molecular Biology. 1981, pp.195-197.
- [24] Z Hajjhashemi, M Popescu. "Predicting Health Patterns Using Sensor Sequence Similarity and NLP." IEEE International Conference on BIBM, 2012, pp.948-950.
- [25] F Duchene, C Garbay, V Riale. "Learning recurrent behaviors from heterogeneous multivariate time-series", Art. Intell. Med., vol 39, 2007, pp 25-47.
- [26] Z Hajjhashemi, M Popescu, "An Early Illness Recognition Framework Using a Temporal Smith Waterman Algorithm and NLP," Proc. Of AMIA Fall Symposium, 2013, pp. 548-557.

- [27] M Popescu. "An ontological fuzzy Smith-Waterman with application to patient retrieval in Electronic Medical Records." IEEE International Conference on Fuzzy Systems (FUZZ). 2010, pp. 1-6.
- [28] M Popescu, E Florea, "Linking Clinical Events in Elderly to In-home Monitoring Sensor Data: A Brief Review and a Pilot Study on Predicting Pulse Pressure", Journal of Computing Science and Engineering, March 2008, vol. 2(1), pp.180-199.
- [29] M. Popescu, A. Mahnot, "Acoustic Fall Detection Using One-Class Classifiers", 31th Int. Conf. of the IEEE EMBS, Minneapolis, MN, Sept 2-6, 2009, pp. 3505-3508.
- [30] Z Hajihashemi, M Popescu, "Detection of Abnormal Sensor Patterns in Eldercare", 4th Intl. conf. of the IEEE EHB, 2013, pp. 1-4.
- [31] Z Hajihashemi, M Popescu, "Detecting Daily Routines of Older Adults Using Sensor Time Series Clustering", 36th Intl. Conf. of the IEEE EMBS, Chicago, IL, August 26-30, 2014.
- [32] B Kaluza, V Mirchevska, et al, "An Agent-based Approach to Care in Independent Living", Ambient Intelligence, Lecture Notes in Computer Science, Volume 6439, 2010, pp. 177-186.
- [33] L. Bergroth, H. Hakonen, T. Raita, "A Survey of Longest Common Subsequence Algorithms". SPIRE, 2000, pp. 39-48.
- [34] S Waterman, "Some Biological Sequence Metrics", Advances in Mathematics, 1976, Vol.20, No.3, pp. 367-387.
- [35] M Skubic, RD Guevara, M Rantz, "Testing Classifiers for Embedded Health Assessment," International Conference on Smart Homes and Health Telematics, Artimino Italy, June 12-15, 2012, pp. 198-205. [36] L Wang, UTV Nguyen, JC Bezdek. et al, "iVAT and aVAT: Enhanced Visual Analysis for Cluster Tendency Assessment", Advance in Knowledge Discovery and Data Mining, 2010, Vol. 6118, pp. 16-27.
- [37] Demiris G, et al. "Older Adults' Attitudes Towards and Perceptions of 'Smarthome' Technologies: a Pilot Study," Medical Informatics and Internet in Medicine, June, 2004, vol. 29, no. 2, pp. 87-94.
- [38] Demiris G, Rantz MJ, Aud MA, Marek KD, Tyrer HW, Skubic M, and Hussam AA, "Seniors' Attitudes Towards Home-based Assistive Technologies," 29th Annual MNRS Research Conference, Cincinnati, Ohio, April 1-4, 2005.
- [39] D. C. Mack, J. T. Patrie, P. M. Suratt, R. A. Felder, and M. Alwan, "Development and preliminary validation of heart rate and breathing rate detection using a passive, ballistocardiography-based sleep monitoring system," Information Technology in Biomedicine, IEEE Transactions on, vol. 13, pp. 111-120, 2009.
- [40] L. Wei, E. Keogh, et al. "Atomic Wedgie: Efficient Query Filtering for Streaming Time Series." 5th IEEE International Conference on Data Mining (ICDM), pp. 490-497, Houston, Texas, Nov 27-30, 2005.
- [41] Lin, Jessica, et al. "Experiencing SAX: a novel symbolic representation of time series." Data Mining and Knowledge Discovery, Vol. 15, Issue 2, 2007, pp. 107-144.
- [42] Diane J Cook, et al. "Activity discovery and activity recognition: A new partnership," IEEE Transaction on Cybernetics, Vol. 43, Issue 3, 2013, pp. 820 – 828.
- [43] P. Dawadi, M. S. Edgecombe, et al. "Smart home-based longitudinal functional assessment." ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2014, pp. 1217-1224.
- [44] Robben, Saskia. "How Is Grandma Doing? Predicting Functional Health Status from Binary Ambient Sensor Data." AAAI Fall Symposium series, 2012, pp. 26-31.
- [45] Virone, G., et al. "Behavioral patterns of older adults in assisted living." Information Technology in Biomedicine, IEEE Transactions on 2(3), 2008, pp.387-398.
- [46] Rantz M, et al. "Evaluation of Health Alerts from an Early Illness Warning System in Independent Living," Computers, Informatics, Nursing, 2013, 31(6):274-280.