# An Early Illness Recognition Framework Using a Temporal Smith Waterman Algorithm and NLP

**Zahra Hajihashemi, MS, Mihail Popescu, PhD**
**University of Missouri, Columbia, MO**

## Abstract

*In this paper we propose a framework for detecting health patterns based on non-wearable sensor sequence similarity and natural language processing (NLP). In TigerPlace, an aging in place facility from Columbia, MO, we deployed 47 sensor networks together with a nursing electronic health record (EHR) system to provide early illness recognition. The proposed framework utilizes sensor sequence similarity and NLP on EHR nursing comments to automatically notify the physician when health problems are detected. The reported methodology is inspired by genomic sequence annotation using similarity algorithms such as Smith Waterman (SW). Similarly, for each sensor sequence, we associate health concepts extracted from the nursing notes using Metamap, a NLP tool provided by Unified Medical Language System (UMLS). Since sensor sequences, unlike genomics ones, have an associated time dimension we propose a temporal variant of SW (TSW) to account for time. The main challenges presented by our framework are finding the most suitable time sequence similarity and aggregation of the retrieved UMLS concepts. On a pilot dataset from three Tiger Place residents, with a total of 1685 sensor days and 626 nursing records, we obtained an average precision of 0.64 and a recall of 0.37.*

## Introduction

Rapid aging of the population in developed countries requires increased attention from health care providers and from the entire society as a whole. While the elderly population (aged over 65) will increase by 8% until 2050 in the developed countries, the working-age population (age between 15 and 64 years) will decrease and its ratio to the elderly population will decline from 4.3 to 2.3 [1]. While older adults prefer to live independently, many of the health conditions associated with old age, such as frailty, dementia, and risk of falling require increased attention and monitoring. However, independent living may lead to infrequent health assessments due to lack of continuous monitoring or fear of being institutionalized. Late health assessments may miss unreported complications, which in turn lead to poor long-term prognosis and quality of life [21]. A possible solution to prevent unreported health problems in independently living older adults is through automatic health monitoring systems.

One efficient approach to health monitoring is to use sensor networks to collect information about the older adult's activity. In the last decade, many of such health monitoring systems have been piloted. For example, MIT's PlaceLab, Georgia Tech's aware House, Honeywell's Independent Lifestyle Assistant, and University of Missouri' s TigerPlace have demonstrated possible approaches to activity monitoring [3,4,5,10]. A variety of methodologies for detecting activity and assess medication compliance have been reported in the literature [6,7,8,9,10]. Even though these and many other systems are successful examples of applying sensor networks to monitor activity patterns, the major unsolved challenge is to consider the health context of the monitored older adults. In this paper, we propose a health pattern recognition framework that also accounts for the health context provided by a nursing EHR. In Tiger Place, our "living laboratory", we have installed since the fall of 2005 sensor networks in the apartments of 47 residents. In the previous work [2], we described a version of our framework system for early illness recognition. In this paper, we extend our previous work by introducing two new features: a sensor similarity algorithm based on a modified Smith-Waterman [16] algorithm and an improved medical term extraction method based on Unified Medical Language System (UMLS), which have been tested on multiple patient datasets.

To predict early illness from combining the non-wearable sensor data with medical concepts extracted from nursing notes, we employ a Natural Language Processing (NLP) methodology. There are several challenges in this research that we intend to address. First challenge is to identify the most suitable similarity measure for comparing multi-dimensional time series. Second, we want to improve the illness prediction using sensor series similarity by utilizing sets of UMLS concepts. Finally, we employ aggregation methods to infer the most probable health concepts associated to an unknown sensor sequence. This paper is organized as follows. In the next section, we review and discus related work. In the "Monitoring System Architecture" section, we describe the system architecture and available sensor data. In the "Methods" we present our illness prediction methodology based on UMLS medical concepts and sensor data similarity. Next, we describe our experiments and show results. Finally, in the last section, we provide conclusions and future work.

## Related Work

Sensor networks have been used in last decade as a promising solution to monitor the health of older adults. One possible approach for the automatic recognition of health problems is to assume that similar medical conditions result in similar behavior, hence in similar sensor activity patterns. While this assumption may not be generally true for a younger, more mobile population, it has a certain degree of validity for the elderly [22, 23, 24]. To better understand an older adult's behavior, elderly monitoring systems may use multiple sensors, such as motion, Kinect, radar, sound, etc. [10]. Typically, in eldercare applications the sensors are not wearable, but rather placed in the living environment to collect data about the resident's behavior [10].

During the analysis, the task of assessing behavior similarity can be cast as computing the similarity of multi attribute sensor time series (MATS). MATS similarity is a challenging task and is still an open research question. Here we mention two most popular approaches to this task. First approach consists of using the Euclidian distance as a measure of similarity between two sequences [11]. Euclidian distance is sensitive to outliers, cannot capture time and cannot be computed for sequences of variable length. The main limitation of using the Euclidean distance as a similarity measure is due to its weak performance in the presence of noise. The second commonly used approach to MATS similarity is based on the non-Euclidean metrics, such as Dynamic Time Warping (DTW) [12]. In [12], authors rephrased the problem of finding similar time series instead as a problem of sequence alignment. They presented an algorithm for DTW on multi-attribute time series and compared it to the ordinary DTW, where a single attribute is used for sequence alignment. The DTW algorithm for two sequences calculates the distance between each possible pair of data points. Then, the algorithm uses these distances to calculate a cumulative distance matrix, and it finds the best path through this matrix.

The second distance measure in the family of non-Euclidean measures is a modified Longest Common Subsequence (LCSS) for continuous domain [13]. LCSS offers more robustness in the presence of noise compare to DTW. Authors in [13] showed that the computational time required by this algorithm is quadratic to the sequence (in our case, the time series) length. To address this limitation, they applied upper boundary functions to limit the computational time. They utilized a fast pre-filtering schema that returns upper bound estimations for the LCSS similarity. In a similar way, authors in [14] proposed a Multi-Dimensional Dynamic Time Warping (MD-DTW) technique, which modifies the distance matrix by using the vector norm to calculate the distance between a pair of points. Although both measures presented a significant improvement compared to the Euclidean distance, some of their drawbacks are: quadratic time complexity, computational overload of the boundary detections, and filtering of data, which may remove valuable information.

In applications where the number of training samples is sufficient, Hidden Markov Models (HMM) typically outperform DTW. In [15], authors compared ordinary DTW, a probability-based DTW and a HMM for the sequence comparison. Using a Gaussian-based probabilistic model for a gesture recognition application, the probability-based DTW shows better results. In [15] a soft distance, based on posterior probability of the Gaussian Mixture Model (GMM), was defined. Despite of HMM's performance, there are several problems with this approach. They include the need for a large amount of data to train HMM, making strict assumptions about data, and setting a large number of parameters.

Our novel approach to sequence similarity is to consider MATS as one-dimensional sequence and to use bioinformatics techniques to find the best alignment. This approach is possible when sensor data is already quantized by the hardware system (see next section), which results in a symbolic sequence of sensor firings and the related firing time. Smith Waterman (SW) algorithm is a well-known algorithm to find an optimal local alignment [16]. Researchers have used SW not only in bioinformatics, but also in natural language processing [17,18]. In [18], the authors aligned words by using SW without a pre-existing ontology. In [19], authors implemented a search tool on protein structures by integrated SW with fuzzy logic to match protein energy profiles.

The idea behind the SW algorithm is to use dynamic programming to find the optimal local alignment with respect to a pre-defined scoring system. The scoring system includes a substitution matrix and a gap-scoring scheme. SW algorithm calculates a local similarity score in linear space, if we just look for a local alignment. In bioinformatics, one motivation for the local alignment is in a situation where biological sequences are highly related to each other, but the similarities in local regions are very low. In this case, it is hard to find the correct alignment. To tackle this problem, a local alignment discards such regions and concentrates on those with high similarity. Another motivation for local alignment is to use a reliable statistical model for optimal local alignment to produce expectation values [20].

SW can be used in time series similarity to find best local matches. In fact, finding local similarity in sensor time series is desirable, since it shows similar health patterns in the local sequence regions. However, for health sensor data, the time stamp is an important factor that needs to be considered when finding optimal local time series alignments. Unfortunately, the classical SW algorithm [16] does not consider time in alignment.

In this paper, we propose a new version of the SW algorithm that considers time in the scoring schema. This approach allows finding the similarity of two MATS and their optimal alignment with respect to time, an important factor in interpreting sensor data. In the next section, we described our proposed method in more details and illustrate it with examples.

## Monitoring System Architecture

We deployed our integrated monitoring system in 47 TigerPlace apartments with the University of Missouri IRB approval. After focus groups with TigerPlace residents early on in our research [9,10], we decided to use only non-wearable sensors for monitoring since they are unobtrusive and more acceptable by older adults. The monitoring has started in fall 2005. On average, we have two years of data for each resident. Figure 1 shows the architecture of our data acquisition system.

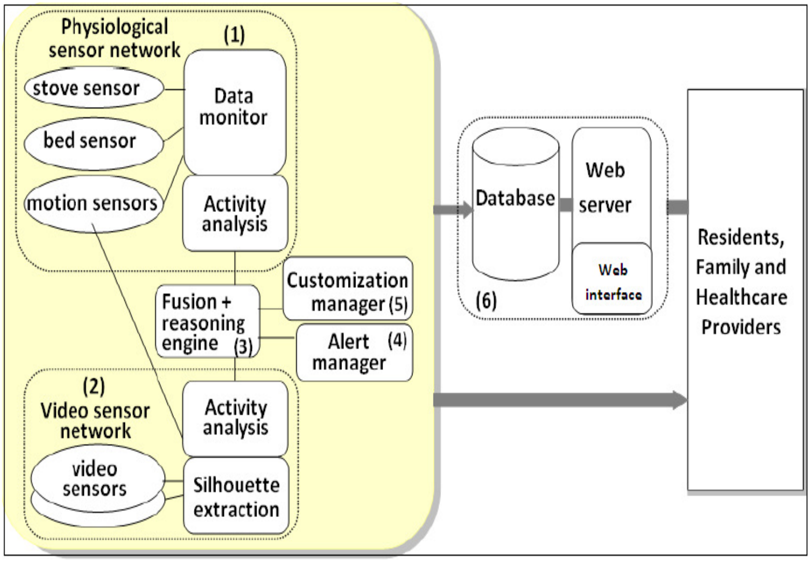


**Figure 1. TigerPlace sensor network architecture.**

Various sensors have been deployed in each apartment: motion, radar, Microsoft Kinect, and bed. The bed sensor is able to measure bed motion, pulse and breathing [6]. Each motion sensor sends an X10 signal (firing) indicating its ID that is logged with a time stamp in our sensor database (see more details at http://eldertech.missouri.edu).

In previous work [2], we used aggregated data from motion, pulse, breathing and restlessness sensors. Aggregating sensor data with respect to location and time can severely reduce the performance of the activity recognition algorithms because of the information loss in the aggregation process. In this study, we extend our work and use 23 different sensors firings, which capture a person's behavior in bed (restlessness, breathing and pulse), kitchen, and the overall apartment motion. Table 1 lists our sensor firings. Note that a physical sensor can have multiple firing types, because a physical sensor can be complex (such as the bed sensor or the vision sensors) and have multiple firings. For example, firings with ID 1-10 are all provided by the bed sensor: "BedMovement1-4" are four levels of motion in bed (less than 3 seconds (s), 3 to 7s, 7 to 15s, greater than 15s), "Breathing1-3" are three levels of breathing (low, normal, high), and "Pulse1-3" are three levels of pulse (low, normal, high). Firings with ID 11 to 15 and 18 to 23 are all produced by motion sensors, while 16 and 17 are given by a temperature sensor.

**Table 1.** ID of the 23 sensor firings used in this project.

| **ID \| Sensors firings** | 6. Breathing2 | 12. LivingMotion | 18. Drawer |
|---|---|---|---|
| 1. BedMovement1 | 7. Breathing3 | 13. BathRoom | 19. Cabinet |
| 2. BedMovement2 | 8. Pulse1 | 14. Off Chair | 20. Cup Cabinet |
| 3. BedMovement3 | 9. Pulse2 | 15. On Chair | 21. Refrigerator |
| 4. BedMovement4 | 10. Pulse3 | 16. TempHigh | 22. Plate Cabinet |
| 5. Breathing1 | 11. BedroomMotion | 17. TempLow | 23. Silverware Drawer |

Table 2 shows a snippet of the sensor firing data recorded in the log file for the resident with ID 3 (On October 5, 2005, around 12:30am the person had several episodes of bed restlessness: 8 of 3-7 s, 2 of less than 3 s and one 7-15 s, and two episodes of low breathing ). Note that the data in Table 2 is a sensor sequence of length 12.

**Table 2**. Sensor sequence snippet for TigerPlace resident ID 3

| UserID | SensorID | Year | Month | Day | Hour | Minute | Second |
|---|---|---|---|---|---|---|---|
| **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** | 2 | 2005 | 10 | 5 | 12 | 37 | 32 |
| **3** | 2 | 2005 | 10 | 5 | 12 | 41 | 18 |
| **3** | 2 | 2005 | 10 | 5 | 12 | 41 | 11 |
| **3** | 2 | 2005 | 10 | 5 | 12 | 41 | 4 |
| **3** | 5 | 2005 | 10 | 5 | 12 | 42 | 40 |
| **3** | 5 | 2005 | 10 | 5 | 12 | 42 | 58 |

Table 3 shows the pilot sensor data from the apartments of three TigerPlace residents in this study. For each resident, we also retrieved visit notes about physical, emotional and other health complaints, recorded by the nurses in the TigerPlace nursing EHR [25]. In our dataset, there are fewer notes than sensor data (automatically logged for each day per resident), as some days didn't have any nursing comments. However, some residents (for example, #3) have days with multiple comments. In addition, each day was manually annotated based on the EHR data (nursing comments, blood pressure, etc.) as normal or abnormal. The abnormal days (column 4 in Table 3) were subjectively determined by retrospectively inspecting the nursing notes. We only use them to tune our sequence similarity algorithm.

**Table 3.** TigerPlace pilot dataset

| Resident Code | Number of sensor days | Number of comments | Number of abnormal days |
|---|---|---|---|
| #1 | 440 | 83 | 81 |
| #2 | 745 | 44 | 35 |
| #3 | 500 | 499 | 335 |

While we acknowledge that our dataset is small, we mention that it has to be seen as three separate experiments with more than 400 samples each, rather than one experiment with only three samples. The reason for this perspective is that mathematical models (such as classifiers, algorithm parameters) for early illness recognition are not usually transferrable from one patient to another, because the disease-behavior associations could be vastly different between people. In addition, in this work we merely intend to open a research direction rather than claiming that our framework is ready for clinical trials.

## Methods

### *I. Temporal Smith Waterman Sequence Similarity Measure*

Sensor networks have previously been used in health care to predict health patterns [22,23]. Defining a suitable distance/similarity measure to find similar patterns in sensor networks data is still an open question. The multi-dimensional nature of the sensor network complicates the task of measuring similarity. In [2], we used aggregated sensor data time-wise (in one hour time intervals) and sensor type-wise (aggregated all motion sensor data together) obtaining a four-dimensional sequence vector. We used a root mean square error (RMS) approach, a widely used techniques in signal processing to find the distance between time series. That is, for two sensor sequences $X = \{x_{ij}\}$, $Y = \{y_{ij}\} \in R^4 \times R^{24}$ we can compute the RMS distance as:

$$\mathrm{RMS}(X, Y) = \sqrt{\frac{1}{24}\left((X_1 - Y_1)^2 + \cdots + (X_{24} - Y_{24})^2\right)} \ ,$$

where $X_i = \sqrt{\frac{1}{4}\left(x_{1i}^2 + x_{2i}^2 + x_{3i}^2 + x_{4i}^2\right)}$ $\qquad Y_i = \sqrt{\frac{1}{4}\left(y_{1i}^2 + y_{2i}^2 + y_{3i}^2 + y_{4i}^2\right)}$.

In this work, we propose a new approach where we don't aggregate the data and we maintain the order of the sensor firings as recorded in the log file. To find similar sensor firing sequences, we employ the SW algorithm [16]. However, the drawback of SW algorithm is that it does not consider time in local alignment. Time is an important factor in sensor firing sequence, because various time intervals can be associated with different behaviors (for example fast or slow walking). In this paper, we propose a methodology based on the SW algorithm that accounts for time in sequence similarity.

The SW algorithm was first proposed in [16] to align two molecular sequences, $T_1=\{C_{11}, C_{12}, \ldots, C_{1m}\}$ and $T_2=\{C_{21}, C_{22}, \ldots, C_{2n}\}$ where $m, n \in N$ and $C_{ij}$ belongs to alphabet $\Sigma$ (nucleotides or amino acids). First, SW builds a similarity score matrix H where H(i,j) is equal to the similarity score between a suffix of $T_1[1\ldots i]$ and a suffix of $T_2[1\ldots j]$ such that

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

$$H_{ij}=\max\{0,\ H_{i-1,j-1}+S(C_{1i}, C_{2j}),\ \max_{k\geq 1}\{H_{i-k,j} - W_k\},\ \max_{k\geq 1}\{H_{i,j-k} - W_k\}\} \qquad (2)$$

$$W_k = g + ck \qquad (3)$$

$$\mathrm{FinalScore} = \frac{\max\{H_{ij}\}}{\min\{n,m\}} \qquad (4)$$

**Figure 2.** Smith Waterman algorithm

In equation 2, $S(C_{1i}, C_{2j})$ is the similarity between characters $C_{1i}, C_{2j} \in \Sigma$ denoted henceforth as $S_{ij}$. The penalty of opening a gap will be calculated by equation 3 where g and c are two constants and $k \in N$ is the length of the gap. The final alignment score is derived from equation 4. Based on these calculations, for each character-to-character comparison SW calculates a score that is positive for exact matches or substitution, and negative for insertion or deletions. In weight matrices, scores are added together and the highest scoring alignment is reported.

To consider time as a factor in calculating the similarity score, we propose a temporal variant of the SW algorithm (TSW) by considering time as a gap. In order for this approach to work, we assume that each character in the sequence has an associated time stamp. The time stamp shows the time when a character (firing) is emitted by the related sensor network. Consider two time-stamped sequences of characters as $T_1=\{(C_{11}\, t_{11}), (C_{12}\, t_{12}), \ldots, (C_{1m}\, t_{1m})\}$ and $T_2=\{(C_{21}\, t_{21}), (C_{22}\, t_{22}), \ldots, (C_{2n}\, t_{2n})\}$ where $m, n \in N$ and $C_{ij}$ belongs to alphabet $\Sigma$ and $t_{ij}$ are the time of the day of the firing $C_{ij}$. Figure 3 shows the temporal SW algorithm.

$$H_{i0}=H_{0j},\ i\epsilon[1,n] \text{ and } j\epsilon[1,m] \qquad (5)$$

$$H_{ij}=\max\{0,\ H_{i-1,j-1}+S(C_{1i}, C_{2j}),\ \max_{k\geq 1}\{H_{i-k,j} - W_{\Delta t}),\ \max_{k\geq 1}\{H_{i,j-k} - W_{\Delta t}\}\} \qquad (6)$$

$$W_{\Delta t} = g + c|t_{1i} - t_{2j}| \qquad (7)$$

$$\mathrm{FinalScore} = \frac{\max\{H_{ij}\}}{\min\{n,m\}} \qquad (8)$$

**Figure3.** Temporal Smith Waterman (TSW) algorithm.

Essentially, TSW considers the time of the day between the sensor firings as a gap and computes the gap penalty $W_{\Delta t}$ by using time stamps (as shown in equation 7 above). We use the "time of the day" metric since we would like to find similar behaviors across different days that happen at roughly the same time. Note that the type of time used in (7) is critically important and is dependent on the application. For example, if one would like to compare patients represented by the sequences of ICD-9 diagnoses, the time would more likely include the year of each diagnosis. An interesting aspect of our algorithm is one doesn't need to know in advance the time unit required by equation 7. Instead, the search for best constant *c* should be able to discover the scale of the events as we will see in the Experimental results section.

### *II. Sensor sequence similarity evaluation*

To evaluate our sensor sequence similarity measure, we used the data from Table 3, column 1 and 4 (sequences with normal/abnormal labels). Then, we employed a k-nearest-neighbor (KNN) method with leave-one-out approach to classify each sequence as normal or abnormal. To test our TSW similarity measure, we found the k nearest sequences (k=5) to test one and get the majority vote for the class label (normal or abnormal day). The results are reported in terms of Sensitivity and Specificity, that is:

$$\text{Specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives+number of false positives}} \qquad (9)$$

$$\text{Sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives+number of false negatives}} \qquad (10)$$

Finally, we compared the classification results obtained using TSW to the ones provided by RMS similarity used in [2].

### *III. Sensor sequence annotation using TSW Algorithm*

Computing sequence similarity allows us to proceed toward our goal of predicting patient's health status using sensor data and EHR medical concepts. We obtained medical concepts by parsing the text data from our nursing EHR (nursing notes) using the Metamap NLP tool provided by UMLS (http://metamap.nlm.nih.gov/). Metamap associates each medical concept found in the nursing notes with a Concept Unique Identifier (CUI). The prediction algorithm used a k nearest neighbor with a leave-one-out cross-validation approach. For each unknown sensor sequence $X_i$ ,we compute the distances (using both TSW and RMS for comparison) to all other daily sequences. Then we select k (k=5) most similar sequences and their associated CUIs, $T_j$, j=1,k. Finally, we assign to $X_i$ the medical terms with CUI's that are computed as the intersection of all $T_j$ that is we take the CUIs common to all retrieved sequences. The annotation algorithm is given in Figure 4 below:

```
Illness recognition algorithm using Temporal Smith Waterman
    1. Initialize k, g, c;
        a. k = number of neighbors
        b. g = gap opening
        c. c = gap extend
    2. For each sensor sequence $X_i$, compute its distance (TSW or RMS) to all N sequences
    3. Find k most similar sequences $X_j$, j=1,k
    4. Extract UMLS terms $T_j$ ($X_j$)= { CUI's associated with $X_j$ }, j=1,k
    5. Annotate sequence $X_i$, with the UMLS terms $T_i = \bigcap_{j=1}^{k} T_j$
```

**Figure 4.** Illness recognition algorithm

We used precision, recall, and F-measures for the evaluation of the algorithm, defined as:

$$\text{Precision} = \frac{\text{True\_Positive}}{\text{True\_Positive+ False\_Positive}} \qquad (11)$$

$$\text{Recall} = \frac{\text{True\_Positive}}{\text{True\_Positive +False\_Negative}} \qquad (12)$$

$$\text{F\_Measure} = 2\frac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \qquad (13)$$

During these experiments, we only used the days from our data set that have associated nursing notes (see column 3 in table 3). The results are given in the next section.

## Experimental Results

To investigate the performance of the proposed illness prediction methodology based on TSW, we compare the results to our previous approach [2] that used multi-dimensional sensor time series and RMS distance. Our evaluation has two steps. In the first set of experiments, we investigate the properties of the TSW similarity and then compare it to RMS. In the second set of experiments, we demonstrate the results of illness prediction based on TSW sensor sequence similarity.

### *I. Results for the TSW similarity measure*

Before comparing TSW to RMS we have to choose the best values for the temporal gap parameters *g* and *c* (see Formula 7). For this purpose, we compute the sensitivity of classifying each sequence as normal/abnormal using the dataset shown in Table 3 (columns 2 and 4) and the method described in section Methods.II. To limit our search space we set *g=0*. The variation of the Sensitivity with *c* for patient #1 is shown Figure 5. Note that times $t_{ij}$ in formula 7 are given in seconds.

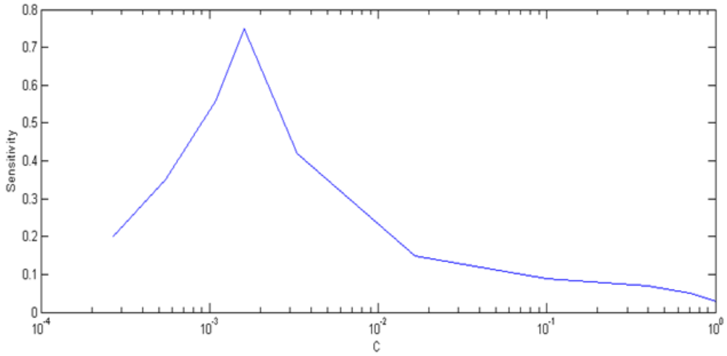


**Figure 5.** Variation of sequence classification sensitivity vs. temporal gap parameter c

We see that the sensitivity has a sharp maximum around c=0.002. Since the times in (7) are in seconds, it implies that our time scale is effectively about 480s/60~8 min. We obtained similar results for the other two residents.

The classification results obtained using TSW and RMS on the sensor sequences mentioned in Table 3 (column 2 and 4) as normal/abnormal are shown in Table 4 and 5, respectively.

**Table 4.** Sensor sequence classification using TSW similarity

| **Resident** | **Sensitivity** | **Specificity** |
|---|---|---|
| **#1** | 0.89 | 0.36 |
| **#2** | 0.96 | 0.53 |
| **#3** | 0.48 | 0.85 |

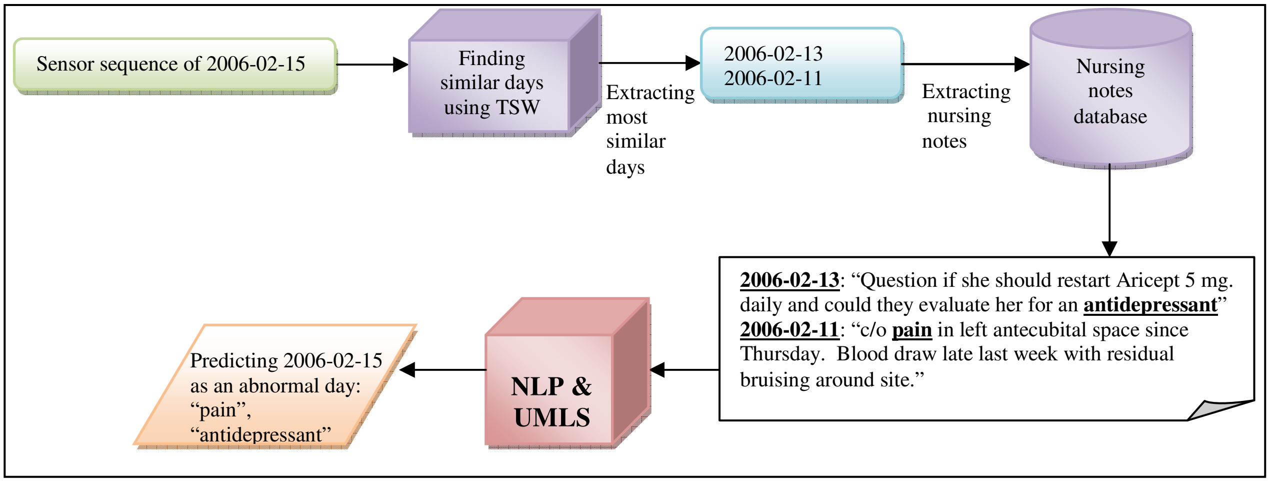
**Table 5**. Sensor sequence classification using RMS similarity

| **Resident** | **Sensitivity** | **Specificity** |
|---|---|---|
| **#1** | 0.66 | 0.27 |
| **#2** | 0.71 | 0.10 |
| **#3** | 0.16 | 0.67 |

As it can be seen from above tables, TSW clearly outperforms RMS by about 20-30% in sensitivity and 10-40% in specificity. Obviously, sensor aggregation (both temporal and by sensor type) is detrimental to the sequence classification. Interestingly, the sensitivity results for resident #3 is lower than for the other two elders, for both TSW and RMS algorithms. The reason for the lower sensitivity is that the classification of abnormal days is not as reliable as for the normal days. The resident #3 has the greatest proportion of abnormal days in the dataset (has more medical problems than the other two). This could be caused by both the diversity of abnormal days (multiple behaviors, possible very different) but also by the subjective labeling of the days as abnormal.

*II. Results for illness prediction*

In the second set of experiments, we performed illness prediction by annotating sensor sequences with UMLS concepts from the nursing notes. We describe this process by an example shown in the figure 6.



**2006-02-13**: "Question if she should restart Aricept 5 mg. daily and could they evaluate her for an **antidepressant**"
**2006-02-11**: "c/o **pain** in left antecubital space since Thursday. Blood draw late last week with residual bruising around site."

**Figure 6.** Predicting abnormal days using nursing notes.

In Figure 6 we have a sensor sequence of resident #2 on February 15, 2006, and we want to classify this day as "normal" – no annotation available or "abnormal"- find all related UMLS terms. The system processes this sensor sequence with the TSW algorithm to detect the most similar sensor sequences to this pattern in the stored database of all past motion sensor sequences for this resident. Then, the system extracts nursing notes associated with the selected similar days detected by the TSW algorithm. We apply NLP and UMLS to convert raw text (nursing notes) to a set of CUI's (concepts unique identifiers of the medical terms). Finally, the system identifies the sensor sequence of 2006-02-15 as an abnormal day because the similar days, such as 2006-02-13 and 2006-02-11, are tagged as "abnormal", and the system also propose a set of CUI's as suggested abnormalities concerns ("pain" and "antidepressant").

While the final version of the system uses all available sensor history, in the following experiments we only used the abnormal days (column 4 in table 3), that is the search was conducted only among the "abnormal" days. The scaled down experiments were necessary to increase speed as our TSW implementation is not time-efficient at this time. Moreover, including the "normal" days requires solving supplementary problems such as what to do when out of three most similar sequences two are "normal" and one "abnormal". The comparison of the annotation results obtained with the TSW and the RMS measures are shown in Tables 6 and 7, respectively.

**Table 6.** Sequence annotation using TSW similarity

| Resident | Precision | Recall | F_Measure |
|---|---|---|---|
| **#1** | 0.62 | 0.37 | 0.46 |
| **#2** | 0.41 | 0.9 | 0.55 |
| **#3** | 0.44 | 0.57 | 0.47 |

**Table 7.** Sequence annotation using RMS similarity

| Resident | Precision | Recall | F_Measure |
|---|---|---|---|
| **#1** | 0.51 | 0.52 | 0.42 |
| **#2** | 0.41 | 0.34 | 0.35 |
| **#3** | 0.35 | 0.54 | 0.35 |

As it can be seen from above tables, the F-measures improved by 4-12% by using the TSW measure, but the improvement is not as large as the one observed in Tables 4, 5. The main reason of this outcome is that in this experiment we did not use the "normal" days, that is, the days without any annotation. This could be easily done by labeling each day without any nursing note with the term "normal" and perform the same experiments as before.

## Study limitations

We acknowledge that our study has the following limitations:

1. Our study is based on the fundamental assumption that there is a univocal relation between disease states and behavior. While we acknowledge that, for younger, active adults this assumption does not hold, we believe that this becomes increasingly true for older, less active adults.
2. Our patient dataset is small. However, it is not insignificant: it contained 1685 days of sensor data (85.8 million sensor firings) and 626 nursing notes. We have 47 sensor networks deployed at TigerPlace and we plan to extend our study to include our entire resident population.
3. We did not include all of the available sensors such as Kinect, radar and microphone array. Including vision based sensors, such as Kinect, can greatly increase behavior recognition capabilities. We plan to include Kinect data in the future studies.
4. Our sensor sequences represented one 24-hour day, from 0h to 23:59h, which would not allow us to capture the similarity of two days, say {d1, d2}, where d1 had one event 5 minute after midnight (hence at the beginning of the sequence) and d2 had the same event 5 minutes before midnight (hence at the end of the sequence). In future implementations of TSW we plan to search for the similar sequence using a sliding window over the entire time frame.

## Conclusions

In this paper we presented a framework for illness prediction using sensor networks and a temporal Smith Waterman (TSW) algorithm. The health patterns were described by NLM CUI terms that were extracted from nursing notes using MetaMap, a NLP tool released by NLM. The association between nursing notes and sensor sequences was made possible by our unique nursing EHR that captures both sources of information for the TigerPlace residents. We tested our framework on a pilot set that consists of 1685 days of sensor data and 626 nursing notes. We compared our results with an approach that uses sensor aggregation and a Euclidean-type (RMS) similarity measure. In this comparison, the proposed TSW framework outperformed the one based on RMS. In future research we plan to add more sensors and residents to our study and to refine the annotation procedure.

## Acknowledgement

This project has been funded by a NSF SHWB grant, award # IIS-1115956.

## References

1. Pogorelc B, Gams M. Home-based health monitoring on the elderly through gait recognition. Journal of Ambient Intelligence and Smart Environments. 2012;Volume4, Number5:415-428.
2. Hajihashemi Z, Popescu M. Predicting Health Patterns Using Sensor Sequence Similarity and NLP. 2012 IEEE International Conference on BIBM 2012;948-950.
3. Intille S, Larson K, Munguia-Tapia E, Beaudin J, Kaushik P, Nawyn J, et al. Using a live-in laboratory for ubiquitous computing research. Proc. of Pervasive Health. 2006.
4. Kidd CD, Orr RJ, Abowd GD, Atkeson IA, Essa B, MacIntyre B, et al. The Aware Home: A living laboratory for ubiquitous computing research. Proc of the 2nd International Workshop on Cooperative Buildings-CoBuild'9, 1999.
5. Haigh KZ, Kiff LM, Ho G. Independent Lifestyle Assistant: Lessons learned. Assistive Technology. 2006; 87-106.
6. Heise D, Skubic M. Monitoring pulse and respiration with a non-invasive hydraulic bed sensor. Proc 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Buenos Aires, Argentina. 2010.
7. Beckwith R. Designing for ubiquity: The perception of privacy. Pervasive Computing. 2003; 40-6.
8. Mack D, Alwan M, Turner B, Suratt R, Felder R. A passive and portable system for monitoring heart rate and detecting sleep apnea and arousals: Preliminary validation. Arlington VA. Proceedings Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare (D2H2). 2006.
9. Rantz MJ, Dorman-Marek K, Aud M, Tyrer HW, Skubic M, Demiris G, et al. A technology and nursing collaboration to help older adults age in place. Nursing Outlook; 2005:40-45.
10. Skubic M, Alexander G, Popescu M, Rantz M, Keller J. A Smart Home Application to Eldercare: Current Status and Lessons Learned. *J Technology and Health Care*. 2009: 183-201.
11. Kahveci T, Singh A, Gurel A. Similarity Searching for Multi-attribute Sequences. SSDBM. 2002:175-184.
12. Ten Holt GA, Reinders MJT, Hendriks EA. Multi-Dimensional Dynamic Time Warping for Gesture Recognition. Thirteenth annual conference of the Advanced School for Computing and Imaging. 2007.
13. Vlachos M, Hadjieleftheriou M, Gunopulos D, Keogh E. Indexing Multi-Dimensional Time-Series with Support for Multi Distance Measures. SIGKDD. 2003:216-225.

14. Sanguansat P. Multiple Multidimensional Sequence Alignment Using Generalized Dynamic Time Warping. WSEAS Transaction on Mathematics. 2012;Issue 8; Volume 11; 684-694.
15. Bautista MA, Vela AH, Ponce V, Perez-Sala X, et Al. Probability-based Dynamic Time Warping for Gesture Recognition on RGB-D data. International Workshop on Depth Image Analysis. 2012.
16. Smith TF, Waterman MS. Identification of common molecular subsequences. Journal of Molecular Biology. 1981;147:195–197.
17. Karwarth A, Kersting K. Relational Sequence Alignments and Logos. In S. H. Muggleton and R. Otero, editors, Proceedings of the 16th International Conference on Inductive Logic Programming (ILP-06), Santiago, Spain. August 24-27, 2006.
18. Mrozek D, Maysiak B, S Kozielski . EAST: Energy Alignment Search Tool. L. Wang et al. (Eds.): FSKD 2006, LNAI 4223. 2006: 696-705.
19. Popescu M. An ontological fuzzy Smith-Waterman with application to patient retrieval in Electronic Medical Records. 2010 IEEE International Conference on Fuzzy Systems (FUZZ). 2010:1-6.
20. Karlin S, Altschul SF. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc Natl Acad Sc i. 1990; 87:2264-2268.
21. Goodwin JS, Howrey B, Zhang BD, Kuo Y-F. Risk of Continued Institutionalization after Hospitalization in Older Adults. J Gerontol A Biol Sci Med Sci. 2011:1-7.
22. Popescu M, Florea E. 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*. 2008; 2; 1:180-199.
23. Popescu M, Mahnot A. Early Illness Recognition in Older Adults Using In-Home Monitoring Sensors and Multiple Instance Learning. *Methods of Informatics in Medicine*. 2012 Aug 7; 51(4):359-67.
24. Lee Y, Ho D, Popescu M. Microphone Array System for Automatic Fall Detection. *IEEE Transactions on Biomedical Engineering*. May 2012; 59:1291-1301.
25. Popescu M, Chronis G, Ohol R, Skubic M, Rantz M. An Eldercare Electronic Health Record System for Predictive Health Assessment. IEEE 13th International Conference on e-Health Networking, Applications and Services. June 13-15, 2011: 194-196.