Dynamic Task Optimization in Remote Diabetes Monitoring Systems

Myung-kyung Suh[1], Jonathan Woodbridge[1], Tannaz Moin[2,4], Mars Lan[1], Nabil Alshurafa[1], Lauren Samy[1], Bobak Mortazavi[1], Hassan Ghasemzadeh[1], Alex Bui[3], Sheila Ahmadi[2], Majid Sarrafzadeh[1]
Computer Science Department[1]
Department of Medicine, Division of Endocrinology[2]
Medical Imaging Informatics Group, Department of Radiological Sciences[3]
University of California, Los Angeles
HSR&D Center of Excellence for the Study of Healthcare Provider Behavior, VA Greater Los Angeles[4]
{dmksuh, jwoodbridge, marslan, nabil, lsamy, bobakm, hassan, majid}@cs.ucla.edu
{tmoin, sahmadi}@mednet.ucla.edu
buia@mii.ucla.edu

Abstract— **Diabetes is the seventh leading cause of death in the United States, but careful symptom monitoring can prevent adverse events. A real-time patient monitoring and feedback system is one of the solutions to help patients with diabetes and their healthcare professionals monitor health-related measurements and provide dynamic feedback. However, data-driven methods to dynamically prioritize and generate tasks are not well investigated in the domain of remote health monitoring. This paper presents a wireless health project (WANDA) that leverages sensor technology and wireless communication to monitor the health status of patients with diabetes. The WANDA dynamic task management function applies data analytics in real-time to discretize continuous features, applying data clustering and association rule mining techniques to manage a sliding window size dynamically and to prioritize required user tasks. The developed algorithm minimizes the number of daily action items required by patients with diabetes using association rules that satisfy a minimum support, confidence and conditional probability thresholds. Each of these tasks maximizes information gain, thereby improving the overall level of patient adherence and satisfaction. Experimental results from applying EM-based clustering and Apriori algorithms show that the developed algorithm can predict further events with higher confidence levels and reduce the number of user tasks by up to 76.19 %.**

Keywords- ***remote health monitoring, wireless health, telemedicine, diabetes, real-time feedback, task optimization, association rule mining, Apriori algorithm, expectation maximization algorithm***

I. INTRODUCTION

In the United States, 8.3% of the population has diabetes with $174 billion spent annually on the disease. If this trend continues, one in three Americans is expected to have diabetes in their lifetime in the next forty years [1]. Diabetes is the seventh leading cause of death in the United States and accompanied by significant complications including blindness, hypoglycemia, renal failure, cardiovascular disease [2]. Various studies have shown that regulating and monitoring comorbid conditions including blood glucose, symptoms, blood pressure, weight and activities can have a significant impact in helping delay or prevent complications [3][4][5][6]. With recent advances in technology, people supplement or even tailor this process using sensors and data processing units in their home or in the hospital using remote health monitoring systems. Remote health monitoring systems can help patients with diabetes and their healthcare professionals monitor health-related measurements and provide real-time feedback.

In remote health monitoring, patients are required to perform a series of daily tasks requested by their healthcare professionals. For instance, congestive heart failure patients in Chaudhry's work were required to answer 16 questions and measure and enter their weight using telephone keypads [7] and the study results showed a high missing data rate. Authors in [8] required patients to measure weight, blood pressure, and a 12-symptom questionnaire on a daily basis and showed frequent system non-use. Such system non-use in remote patient monitoring can severely degrade the patient participation rate and the effectiveness of designed systems. As missing data can lead to biased and dangerous conclusions, it is important to reduce the missing data rate and adequately handle missing data [9]. As task complexity is one of the main factors that highly affect user participation and satisfaction [10][11][12][13], reducing the number of required tasks should decrease missing data rate.

For designing a human-centered system, it is critical to distinguish which tasks should be handled by users or automatically processed by computers [13]. In remote health monitoring, analyzing the output of user tasks in real time can help schedule sequences of tasks, avoid unnecessary tasks, and increase usability and effectiveness of the system. Most remote health monitoring systems utilize medical domain experts' knowledge to determine and assign priorities and task sequences. For example, Tang applied a heuristic evaluation method using expert knowledge [14] and Dabbs utilized expert knowledge and patients' survey feedback for designing a health monitoring system [15]. As such, most remote health monitoring systems do not apply a data-driven dynamic process for designing human-centered units and yield redundant information gains.

This paper describes a human-centered task optimizing process combining data discretization methods and first-order logic to reduce the burden in remote diabetes patient monitoring systems. This technique was verified using WANDA (Weight and Activity with Blood Pressure and Other Vital Signs), a remote monitoring system leveraging wireless sensor and communication technologies to monitor the health status of patients with diabetes [8]. The WANDA diabetes study was designed in collaboration between the UCLA Computer Science Department and the UCLA Ronald Regan Medical Center. In this study, we applied data clustering and association rule mining techniques to 21 of the subjects with Type-2 Diabetes enrolled in the intervention group. The experimental results show that the developed algorithm can reduce the required time windows by 80% compared to the case of using experts' knowledge to reach the maximum conditional probability. In addition, the developed algorithm reduces the number of tasks by up to 76.19% with a minimum confidence of 0.95.

II. RELATED WORKS

A. Remote Health Monitoring for Diabetics

According to Desai [16], an effective remote health monitoring system must contain the necessary elements that together complete the circle of disease management. Some of the important circle elements are the reliable measurement of physiological variables that can help in the early detection of adverse events, the efficient transmission of data to enable a timely response, the direct reception of data by personnel qualified to recommend an effective intervention, and patient adherence.

Many studies [17][18][19] have shown the effectiveness of remote health monitoring for patients with diabetes. Criteria for evaluating the effectiveness of remote health monitoring include the accuracy of the collected data, automated feedback and decision support and improvements in clinical outcomes, inlcuding HbA1c or glycemic variability [20]. The studies have demonstrated the effectiveness of well-designed remote monitoring systems that collect accurate data using sensors and reduce the time interval between blood glucose measurements and feedback from healthcare professionals. Well-designed remote health monitoring can also help patients manage their medications and daily behavioral routines including their dietary intakes and activity patterns.

IDEATel is one of the most successful remote health monitoring studies between 2000 and 2008 [18][21]. IDEATel utilized devices and techniques from American Telecare [22]. In this study, patients with diabetes monitored and uploaded blood pressure and blood glucose values through a serial port connected to a computer, and also participated in videoconferencing, electronic messaging, and accessing study web pages. The IDEATel study resulted in improved HbA1c, LDL-cholesterol and blood pressure levels over 5 years compared to the control group who did not use IDEATel components.

In Stone's study [19], patients with diabetes measured blood glucose, blood pressure, and weight using the Viterion 100 Telemonitor [23] connected to a telephone line. If the transmitted readings were in an abnormal range, nurse practitioners adjusted medication for blood glucose, blood pressure and lipid control based on American Diabetes Association target values. The intervention resulted significant reductions in HbA1C levels, which is relevant to long-term blood glucose level.

Montori's study [24] also demonstrated improvement in HbA1c levels after 6 months use of a telemonitoring system. The patients in the intervention arm measured their blood glucose using the ACCU-CHEK Complete glucometer [25] and a phone-line connection. Nurses spent an average of 50 more minutes per patient providing feedback to patients over the phone, demonstrating the increased collaboration between patients and their healthcare providers with the use of these technologies. As diabetes control is so closely linked to variations in daily activities such as dietary intake and physical activity, the exchange of information regarding daily blood glucose readings, meal and/or activity planning are critical for patient education and support [26].

B. Patient-orientedRemote Health Monitoring

One way to quantify patient satisfaction with health monitoring systems is to evaluate the amount of system use [20]. Huang [27] designed a neural network-based remote health monitoring system adoption model to predict the behavior intention toward using the system, using survey answers as inputs to the model (Figure 1).

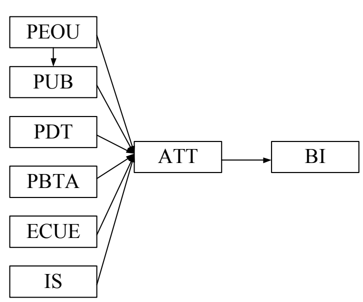


Figure 1. Huang's research structure[27]. Note: perceived ease of use (PEOU), perceived usefulness and benefits (PUB), perceived disease threat (PDT), perceived barriers of taking action (PBTA), external cues to action (ECUE), internal signs(IS), attitude toward using (ATT), and behavioral intention to use (BI).

Wu's study [28] showed that acceptance and satisfaction with a mobile healthcare system is related to compatibility, ease of use and perceived usefulness. In addition, this study showed that the perceived usefulness and ease of use are highly related to self-efficacy, which is a belief that the person has an ability to execute a series of required tasks. As self-efficacy is related to task complexity [29], it is important to make the task procedure simple.

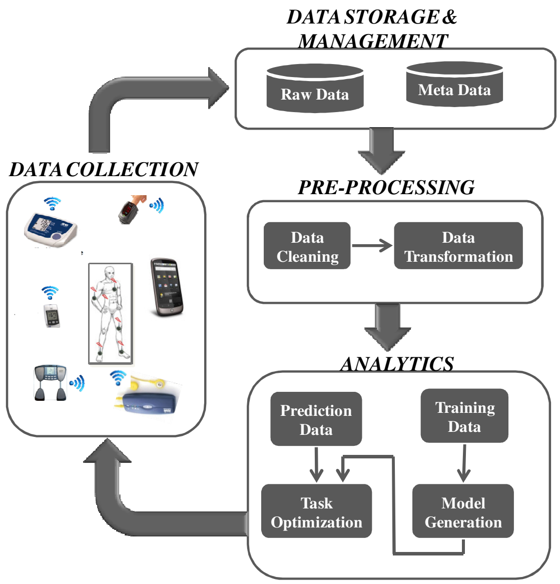


Figure 2.System Architecture of WANDA.

Time and cost-effectiveness is also an important criterion related to patient satisfaction of remote health monitoring systems. Time-effectiveness is a factor of perceived ease of use, while cost-effectiveness is one of perceived usefulness [30]. Estimations in the degree of cost reduction with use of remote health monitoring studies compared to traditional care have varied between 1.6% to 68.3% [31]. Reduced hospitalization and nursing home visits are major contributors of the cost lowering effects in remote health monitoring. However, initial costs of device and service purchase can be an obstacle to user satisfaction despite long-term savings [18]. Therefore, reducing the number of required devices and daily tasks can reduce further costs of equipments and physician workload.

In our study, we focus on the perceived ease of use and usefulness of the system for enhancing patient satisfaction and adherence rates by decreasing the number of required sensors and tasks.

III. SYSTEM ARCHITECTURE

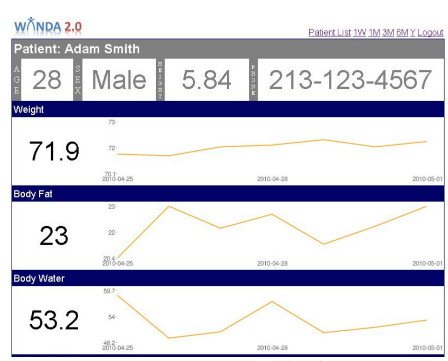
WANDA is a three-tier end-to-end remote monitoring system with extensive hardware and software components designed to cover the broad spectrum of the telehealth and remote monitoring paradigm. The overall architecture is summarized in Figure 2 and further details are available in [8].

The first tier of the architecture consists of a data collection framework, which is formed from a heterogeneous set of sensing devices that measure various bodily statistics such as blood glucose, weight, body fat, body water, blood pressure, heart rate, blood oxygen saturation and body movements. Considering the variability in age and preferences with regard to network options [1][32] , we offer several communication options. The data from these sensors are collected, processed, and transmitted via a phone-line, Ethernet or smartphone-based gateway to the cloud—the second tier of the WANDA architecture.

Data are stored and indexed using a scalable database and can be easily accessed. Data collected from the first tier are sent to web servers to store data and provide monitoring applications such as those in Figure 3. Through the monitoring applications, healthcare providers can leave comments and annotation of collected data, as well as export data. Additionally, the WANDA web application includes a basic statistical analysis tool to verify the test result and the effectiveness of the clinical trial. This function includes Wilcoxon rank test, log-rank test, t-test, etc. which are widely used in many randomized trials [33].

The last tier of the WANDA architecture is a backend analytics engine capable of continuously generating statistical models and predicting outcomes using various machine learning and data mining algorithms. Once data are transmitted to the server, basic preprocessing and dimensionality reduction algorithms are executed prior to data analytics. Data cleaning and signal transformations are the main goals of this pre-processing step. The analytics process normally consists of two stages. First, the data are downloaded and analyzed offline based on various hypotheses. Once a strong model has been generated and validated, it can then be uploaded to the server to perform real-time analytics. One of the challenges is to optimize the algorithm so that it can be executed in a real-time fashion. Finally, when the algorithm detects a pattern that is strongly associated with a predictable user action, a predicted outcome of task optimization, missing data or an undesirable outcome, real-time feedback is provided to remote health monitoring systems, patients and healthcare professionals.

(a) iPhone and Android Smartphone Applications



(b) Web-based Application

Figure 3.WANDA Monitoring Applications.

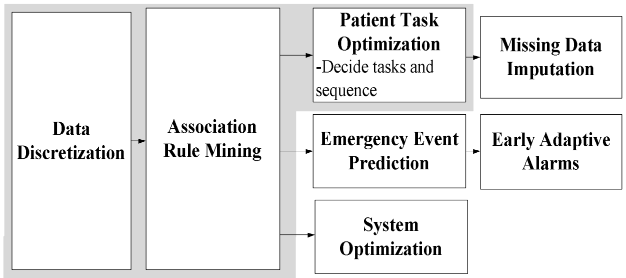


Figure 4. Association Rule Learning Workflow in WANDA.

IV. TASK OPTIMIZATION

As one of the data analytics in the third tier, WANDA performs data transformation for quantizing sensory data readings and executes data association rule mining. Instead of using experts' knowledge, WANDA finds data clusters and their ranges in order to discretize timestamps and blood glucose readings. Association rule analysis is a method to find interesting relations among attributes in large data sets. Rules are derived using previously collected data to help predict the current or future behaviors of a patient.

First order logic from association rule mining can be used for patient task optimization, emergency event prediction and system optimization, as shown in Figure 4. Association rule mining and its feedback are used for reducing the number of tasks required by patients while increasing information gain. One of the advantages of the task optimization step is improving patients' adherence to remote health monitoring and enhancing missing data imputation results by obtaining more data or finding underlying data relationships. After task optimization for improving patient participation, missing data imputation techniques can be applied to WANDA in order to predict missing values and provide alarms to healthcare professionals when the predicted missing values are out of acceptable range [35]. In addition, association rules related to emergency events can be used to generate early adaptive alarms and guidance to prevent emergency events. For example, analyzing past vital signal data can help find trends of readings, which cause emergent events such as hospitalization. Also, the association rule mining finds users' tendency of remote health monitoring system usage and provides dynamic feedback to the system for optimizing battery life.

A. *Data Discretization*

Sensor readings and corresponding timestamps in remote health monitoring are continuous signals. To reduce the dimensionality and complexity of processing continuous numeric and timestamp data, it is necessary to discretize the data. Under supervised settings, patients might have regular schedules of meal and blood glucose measurement such as 6:30 am, 11:30 am and 5:00 pm [38]. However, in unsupervised environments such as remote health monitoring, patients can have more flexible schedules, so it is hard to categorize time frames as morning, afternoon and evening. In addition, as patients with diabetes generally have higher blood sugar, their readings can be more biased and may not follow normal distribution or standards [3]. Moreover, results of Jarrett's study show that blood glucose levels in the afternoon or evening are generally higher than in the morning [39]. Since blood glucose levels vary depending on the time of day, discretizing blood glucose data equivalently for different time intervals can result in biased results.

Therefore, it is necessary to categorize and quantize data using data-driven methods instead of experts' knowledge or human intuition. In this paper, we assume that the collected sensor data follows a mixture of Gaussian distribution and apply an expectation maximization (EM) algorithm to cluster data into the pre-defined numbers of bins. The EM algorithm [40][41] is an iterative method to optimize the estimation of an unknown parameter Θ, given measured variables U and unmeasured variables J. The objective of the EM algorithm is the maximization of the posterior probability (1) of the parameter Θ given U and J.

$$\Theta^* = argmax_{\Theta} \sum_J P(\Theta, J|U) \qquad (1)$$

The EM algorithm consists of two steps: The expectation step (E step) and the maximization step (M step). The E step finds a local lower-bound to the posterior distribution while the M step optimizes the bound obtained from the E step using iteration. In the E step, the algorithm calculates the expected value of the log likelihood function under the current estimate of the parameters Θ^t of the conditional distribution J given U. This step finds the best lower bound, $B(\Theta|\Theta^t)$

$$B(\Theta|\Theta^t) = \sum_J f^t(J) \log \frac{P(U,J,\Theta^t)}{f^t(J)} \qquad (2)$$

while $f^t(J) = \frac{P(U,J,\Theta^t)}{\sum_J P(U,J,\Theta^t)} = P(J|U,\Theta^t)$. The M step iterates and chooses Θ^{t+1} by maximizing the bound, $B(\Theta|\Theta^t)$ from the E step

$$\begin{aligned}\Theta^{t+1} &= argmax_{\Theta} B(\Theta|\Theta^t) \\ &= argmax_{\Theta}[Q^t(\Theta) + \log P(\Theta)]\end{aligned} \qquad (3)$$

while $Q^t(\Theta)$ is the expected complete log-likelihood, $\log P(U, J|\Theta)$ and $P(\Theta)$ is the prior on the parameters Θ.

The developed algorithm quantizes timestamps of sensor readings and sensor readings in each time range are discretized. Based on mean and standard deviation values of each Gaussian curve from the EM algorithm, the developed algorithm finds intersection points and these points are used for discretizing time and blood glucose ranges.

B. Data Association Rule Mining

Association rule analysis is a method to find interesting and strong association among attributes in large data sets. One example is affinity analysis to find the purchase behavior of different groups of consumers and their market baskets [42]. The results of affinity analysis can be used for arranging items in the store, planning store promotions, etc.

In remote health monitoring for patients with chronic diseases, patients' health status changes dynamically, but health-related readings are correlated [43][44]. Therefore, finding trends and associations of patients' data can help to reduce the number of tasks, decide the order of tasks, and even enable to provide early adaptive alarms to prevent emergency situations.

In WANDA, rules are derived using previously collected data to help predict the status and behavior of a patient. The WANDA implementation uses data collected within a dynamic sliding window *w* determined by the algorithm in Figure 5 before the current or future measurement. Association rule mining and its feedback are used for reducing the number of tasks required by patients while increasing information gain. In the data preprocessing step, the developed algorithm performs data cleaning and discretization for removing erroneous data and discretizing timestamp and indexing data (see section above). The system also indexes blood glucose and questionnaire response data as multiple measurements and system non-use. Additionally, information on whether a caregiver contacted the patient for each day is used.

The developed algorithm applies the Apriori algorithm [45] to derive first order logic rules, after preprocessing the data (Figure 5). The discretized and categorized data are used in the algorithm as inputs and rules are derived by looking back a variable number of days (time window). The time window is increased by one day for each iteration. The algorithm calculates the support and confidence of each implication and chooses implications qualifying threshold limits. In each subsequent pass, the large item sets found in the previous step are used to generate the candidate sets (the largest item sets). The results of each step are large item sets of qualifying minimum support and confidence in the given time window.

Let $I = \{i_1, i_2,\ldots i_m\}$ be a superset of all possible task outputs. Let D be a set of events such that $D \subset I$. An association rule is an implication of $A \Rightarrow B$ where $A \subset I$, $B \subset I$ and $A \cap B = \emptyset$. Confidence c means that c% of events in D contain A and B. Support s indicates s% of events in D contain A or B. Conditional probability p indicates p% of events in D contain B when A happens. The developed algorithm requires generating association rules that have support, confidence and conditional probability greater than the user-defined thresholds, minimum support (s_{min}), confidence (c_{min}) and conditional probability (p_{min}) .

MAIN LOOP:
Rule:= ø;
w := 1;
while *Aw**≠ ø **do**
  Aw :=*Result* from **APRIORI** with minimum conditional probability p_{min};
  Aw' := Contrapositive of *Aw*;
  *Aw** := Subsets of *Aw* U *Aw'* which antecedents' maximum timestamp is smaller than consequents' minimum timestamp and not in *Rule* with smaller conditional probability;
  Rule :=*Rule* U *Aw**;
  w:= *w*+1;
end

APRIORI:
Result:=ø;
k:=1;
while C_k**≠ ø do**
  create a counter for each itemset in C_k;
  foreach events in database **do**
    Increase the counter of itemsets in C_k which appears in the events;
  L_k := All candidates in C_k with support s_{min}
  Resuit:=*Result* U L_k;
  C_{k+1}:=k+1-itemsets which have all subsets of L_k.
  k:=*k*+1;
end

Figure 5.Association Rule Learning Algorithm in WANDA.

When Apriori returns first order logic rules, $A \Rightarrow B$, the algorithm calculates the contrapositive rules, $\neg B \Rightarrow \neg A$. If the timestamp of the consequent in either implication (original rule or contrapositive rule) is larger than the timestamp of the antecedent and the implication is not a subset of any existing rules, the generated rule is added to the rule set. However, if a subset of the existing rule has a higher conditional probability, the algorithm updates the existing rule with a new conditional probability value. The process stops when there is no new rule and the algorithm returns the final rule set (Figure 5).

The generated rules in *Rule* are prioritized based on the conditional probability values and applied to the remote health monitoring system. Using the implication rules, the system can reduce user tasks by monitoring necessary tasks and predicting unperformed tasks. As the Apriori algorithm has excellent scale-up properties, the developed algorithm can be applied to the system for dynamically arranging daily patient tasks depending on the size of dataset [45].

TABLE I. PATIENT POPULATION INFORMATION

Group	Total	Male	Female	Avg. Age
Intervention	21	18	3	48.13
Control	24	16	8	65.25

V. RESULT

A. Subjects and Data Sets

This study was approved by the UCLA Institutional Review Board (IRB) and patients were randomized to either

TABLE II. WANDA MONITORED ITEMS AND ACCEPTABLE RANGES

Items	Values
Blood Glucose	80 - 200 mg/dl
Q1. Have you had any blood sugar readings < 80 or > 200?	No
Q2. Have you missed doses of your medication?	No
Q3. Today, is your health, good, fair or poor?	Good, Fair
Q4. Compared to yesterday, are you feeling better, about same, or worse?	Better, About same

intervention or control groups starting June 1st, 2011. Participants eligible for recruitment were adults with Type 2 Diabetes, HbA1c≥7.5 who were recently hospitalized. Patients with active malignancy or those unable to provide informed consent were excluded.

In this analysis, we used data from 21 study participants assigned to the intervention arm (see Table I) and the average participation duration is 52.23 days. Patients in the intervention arm are required to measure their blood sugar up to three times a day (morning, afternoon and evening) and answer four questions per day (see Table II). The defined acceptable ranges for blood glucose are between 80 and 200 mg/dL [3]. Acceptable ranges of questionnaire values are denoted in Table II.

The timestamps are discretized into three different categories and blood glucose data in each time category were discretized into three levels. Collected data are also indexed if there are any missing data or multiple measurements. Existence of call logs between a patient and caregivers are labeled. The pre-defined threshold for s_{min} and c_{min} in **APRIORI** are both 0.95 and p_{min} is 0.85 in **MAIN LOOP** in Figure 5. The total number of instances used in this study is 1117 and each data has 54 attributes per day including patients' profile.

B. Data Discretization Methods and Result

To find the best discretization method, we applied different discretization approaches using: 1) the experts' knowledge utilized in Hanefeld and Malherbe's studies

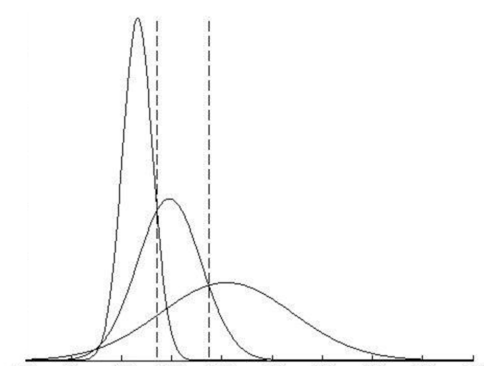


Figure 6.Gaussian Mixture of Blood Glucose in *EMBD*

TABLE III. TIMESTAMP AND BLOOD GLUCOSE RANGES IN *EMTBD*

	Blood Glucose (mg/dl)					
	Level 1		Level 2		Level 3	
Time	Mean	STD	Mean	STD	Mean	STD
00:00:00-08:29:59 (T1)	124.9	28.90	213.2	53.35	426.4	16.24
08:30:00-15:09:45 (T2)	133.6	27.52	187.1	60.65	273.8	138.6
15:09:46-23:59:59 (T3)	139.9	38.69	215.5	68.41	438.3	161.1

([3][38]), on blood glucose and timestamp accordingly; 2) the EM algorithm ; and 3) the combination of experts' knowledge and the EM algorithm.

For experts' knowledge-based discretzation (*EKTBD*), timestamps are categorized into three different time periods (T_i) and each timestamp period isT_1 : 7:30:00-12:00:00, T_2: 12:00:00-16:30:00 and T_3: 16:30:00-21:00:00, and blood glucose readings are categorized into three different level (B_i) and each blood glucose level is B_1: <80 mg/dl, B_2: 80-200 mg/dl and B_3:> 200 mg/dl.

For discretizing timestamps only (*EMTD*), we applied the EM algorithm on collected timestamps of blood glucose measurements and applied experts' knowledge on blood glucose readings. Timestamps are discretized as T_1: 0:00:00 - 8:29:59, T_2: 8:29:59-15:09:45 and T_3: 15:09:45- 23:59:59 and blood glucose readings are discretized as B_1: < 80 mg/dl, B_2: 80-200 mg/dl andB_3: > 200 mg/dl.

For discretizing blood glucose only (*EMBD*), we utilize the EM algorithm on collected blood glucose readings and applied experts' knowledge on timestamps. Each timestamp period is T_1 : 7:30:00-12:00:00, T_2: 12:00:00-16:30:00 and T_3: 16:30:00-21:00:00, and blood glucose readings are categorized as B_1: < 170.8 mg/dl, B_2: 170.8-274.0 mg/dl and B_3: > 274.0 mg/dl (Figure 6).

For quantizing both timestamps and blood glucose readings (*EMTBD*), we utilized EM algorithm on collected timestamps of blood glucose measurements to discretize data into three bins and readings collected in each time period is also discretized into three levels. In other words, each time interval has different standards of categorizing blood glucose data in three different levels. Each discretized timestamp period and mean and standard deviation values of its three different blood glucose levels are in Table III. We assume that the readings are a mixture of Gaussian distribution and find intersection of Gaussian curves. The obtained B_1, B_2, B_3 ranges of T_1 (B_{11}, B_{12}, B_{13}) are <166 , 166-372, and >372and B_{21}, B_{22}, $B_{23,}$ B_{31}, B_{32}, $B_{33}$are < 169 , 169-265, > 265, < 185 , 185-318, > 318 mg/dl accordingly.

The experimental results show that the maximum sliding window size to make the conditional probabilities of ten best Apriori first order logic rules 1.00 is 4 days in *EMTBD*, while other methods require 5 days (Figure 7). Therefore, the developed algorithm can reduce the required time windows to 4 days to reach the maximum conditional probability, while utilizing experts' knowledge requires 5 days. Furthermore, discretizing timestamp and quantizing blood glucose data of each time frame using the EM algorithm yields less computational power and maximizes information

TABLE IV RESULTS OF DATA ASSOCIATION RULE MINING

IF	THEN	Conditional Probability
Answer of Q3 on Day 1 is Good/Fair AND Answer of Q4 on Day 2 is Better/About Same	Answer of Q3 on Day 2 is Good/Fair	0.9970
Answer of Q3 on Day 1 is Good/Fair AND Answer of Q4 on Day 2 is Better/About Same AND Answer of Q4 on Day 3 is Good/Fair	Answer of Q3 on Day 3 is Good/Fair	0.9990
Answer of Q3 on Day 1 is Poor	Answer of Q3 on Day 2 is Poor Answer of Q3 on Day 3 is Poor Answer of Q4 on Day 1 is Worse Answer of Q4 on Day 2 is Worse Blood glucose is above 372 before 8:30 am on Day 1 Blood glucose is above 372 before 8:30 am on Day 2 Blood glucose is above 372 before 8:30 am on Day 3	0.8889
Answer of Q4 on Day 1 is Better/About Same	Answer of Q3 on Day 1 is Good/Fair	0.9961
Answer of Q4 on Day 1 is Better/About same AND Answer of Q4 on Day 2 is Better/About same	Answer of Q3 on Day 2 is Good/Fair	0.9980
Blood glucose is less than 372 before 8:30 am on Day 1 AND Answer of Q4 on Day 2 is Better/About Same AND Answer of Q4 on Day 3 is Better/About Same	Answer of Q3 on Day 3 is Good/Fair	0.9990
Blood glucose is not below 166 AND not above 372 before 8:30 am on Day 1	No Multiple measurement before 8:30 am on Day 1	0.9964
Multiple measurement before 8:30 am on Day 1	Blood glucose is above 372 before 8:30 am on Day 1 Blood glucose is below 166 before 8:30 am on Day 1 Blood glucose is above 265 between 8:29:59am and 15:09:45 pm on Day 1 Multiple measurement between 8:29:59 am and 15:09:45 pm on Day 1 Blood glucose is above 318 after 15:09:45 pm on Day 1 Multiple measurement after 15:09:45 pm on Day 1	0.9333

gain in a shorter period of time compared to using experts' knowledge-based methods

C. Data AsoociationRule Mining Result

The proposed algorithm had optimal results with a look-back window of 5 days. The minimum confidence peaks at

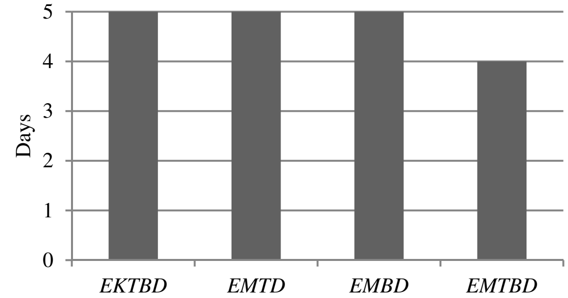


Figure 7. Required Window Size (Days) to Reach Maximum Conditional Probability, 1.

1.00 at 2 to 5 days. Compared to our earlier study [37] which only utilizes experts' knowledge, the combination of EM algorithm-based discretization and Apriori algorithm shows improvement in 1.292% of minimum confidence.

Figure 8 shows the number of new rules added or updated with increasing window size. A total of 7 rules were added with a window size of one day and a total of 19 rules in Table IV were updated with a window size of 5 days. No rules were added or updated with a look-back window of more than 5 days. Compared with earlier study results in [37], a larger amount of data (546 data in [37] and 1117 data in this study after data cleaning) and the EM algorithm-based discretization yields more rules with larger size of sliding window from the Apriori algorithm.

The total number of patient tasks was reduced by up to 76.19% with negligible information loss. The reduction in patient tasks allows the system to generate additional tasks for patients to increase information gain. For example, as shown in Table IV, it was found that responses to Q3 and Q4 can be inferred from each other. This allows the system to generate a new unrelated question to replace Q3 to learn additional information about this patient (with no added work by the patient).

Compared with the algorithm in [46], the developed

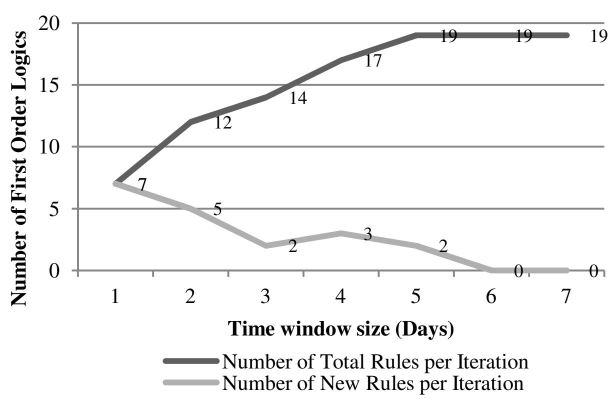


Figure 8.Number of First Order Logics Added or Updated per Iteration.

algorithm shows higher efficiency. However, since Flach's algorithm finds new rules that the developed algorithm doesn't generate, they can be used for generating helpful tips or reminders.

VI. CONCLUSION

The WANDA system was developed in conjunction with the University of California Los Angeles Computer Science and the UCLA Ronald Regan Medical Center. WANDA monitors health-related readings such as blood glucose, weight, blood pressure, etc. and analyzes sensor readings and patient profile data for improving the quality of care and preventing emergency situations.

In this study, we developed WANDA, a three tier remote health monitoring system and focused on increasing ease of use in order to improve patients' system adherence. The developed system applies EM-based data discretization and Apriori rule learning algorithms and finds association rules using collected sensor readings with dynamic sliding windows. We assumed that sensor readings from patients are Gaussian mixture and quantize continuous features and applied Apriori algorithm which efficiently finds related data using support values. The designed algorithm minimizes the number of action items and reorganizes series of tasks for maximizing information gain.

In this work, we applied the developed algorithms to 1117 data sets from 21 patients with diabetes enrolled in the intervention arm. Patients are required to measure their blood sugar up to three times a day and answer four questionnaires daily. The experimental results show that the developed algorithm can reduce the number of tasks by up to 76.19% with minimum support 0.95, minimum confidence 0.95, minimum conditional probability 0.85 and maximum time window size of 5 days. Compared to our earlier study [37], the EM-based discretization helps improve confidence levels of first order logics and predict further events. As the Apriori algorithm has excellent scale-up properties [45], the developed algorithm can be applied to the remote patient with low complexity.

Future studies will investigate and validate the significance of the obtained first order logic rules in this paper. To make the first-order logic richer to reduce required patient tasks dynamically, more data association rule mining techniques will be exploited and maximize the conditional probability and confirmation. In addition, patient survey data will be combined to predict adherence rate in advance, based on their perception and experience of remote health monitoring technologies.

ACKNOWLEDGMENT

This work was supported by NIH/National Library of Medicine Medical Informatics Training Program Grant T15 LM07356. The authors would like to thank Nancy Lee, Rose Healy, Florence-Joy Figueras, Jane Lee, and Toni Ann Riveros for their help in patient recruitment and advice. Tannaz Moin was an endocrinology fellow in the UCLA Department of Medicine, Division of Endocrinology, Diabetes & Hypertension and is now a health services research fellow supported by VA Office of Academic Affiliations, Health Services Research and Development through the Health Services Fellowship Training Program (TMP 65-020), VA Greater Los Angeles.

REFERENCES

[1] CDC. Centers for disease control and prevention. http://www.cdc.gov/. 2012.

[2] Diabetes Control and Complications Trial Research Group. Effect of intensive diabetes treatment on the development and progression of long-term complications in adolescents with insulin-dependent diabetes mellitus: Diabetes Control and Complications Trial. The journal of pediatrics, 125(2), 177-188. 1994

[3] Hanefeld, M. Risk factors for myocardial infarction and death in newly detected NIDDM: the Diabetes Intervention Study, 11-year follow-up. Diabetologia, 39(12), 1577-1583. 1996.

[4] Ciechanowski, P.The relationship of depressive symptoms to symptomreporting, self-care and glucose control in diabetes.General Hospital Psychiatry, 25(4):246–252. 2003.

[5] Vijan, S. Treatment of hypertension in type 2 diabetes mellitus: blood pressure goals, choice of agents, and setting priorities in diabetes care. Annals of internal medicine, 138(7):593–602. 2003.

[6] The Diabetes Prevention Program Research Group. The diabetes prevention program (DPP) : Description of lifestyle intervention, 2002.

[7] Chaudhry, S I. Randomized trial of telemonitoring to improve heart failure outcomes (Tele-HF): study design. Journal of cardiac failure, 13(9), 709-714. 2007.

[8] Suh, M. A Remote Patient Monitoring System for Congestive Heart Failure. Journal of medical systems, 35(5), 1165-1178. 2011.

[9] Wood, A M. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clinical trials, 1(4), 368-376. 2004.

[10] McKeen, J D. The relationship between user participation and user satisfaction: an investigation of four contingency factors. MIS Quarterly, , 427-451. 1994.

[11] Or, C K L. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. Journal of the American Medical Informatics Association, 18(1), 51-59.2011.

[12] Lu, H P. An empirical study of perceived usefulness and perceived ease of use on computerized support system use over time. International journal of information management, 14(5), 317-3290. 1994.

[13] Maguire, M. Methods to support human-centred design. International journal of human-computer studies, 55(4), 587-634.2001.

[14] Tang, Z. Applying heuristic evaluation to improve the usability of a telemedicine system. Telemedicine and e-health, 12(1), 24-34. 2006.

[15] Dabbs, A D V. User-centered design and interactive health technologies for patients. Computers, informatics, nursing, 27(3), 175-183. 2009.

[16] Desai, A S. Connecting the circle from home to heart-failure disease management. The New England journal of medicine, 363(24), 2364-2367. 2010.

[17] Jaana, M. Home telemonitoring of patients with diabetes: a systematic assessment of observed effects. Journal of evaluation in clinical practice, 13(2), 242-253. 2007.

[18] Shea, S. A randomized trial comparing telemedicine case management with usual care in older, ethnically diverse, medically

underserved patients with diabetes mellitus: 5 year results of the IDEATel study. Journal of the American Medical Informatics Association, 16(4), 446-456. 2009.

[19] Stone, R A. Active care management supported by home telemonitoring in veterans with type 2 diabetes. Diabetes care, 33(3), 478-484. 2010.

[20] Klonoff, D C. Using Telemedicine to Improve Outcomes in Diabetes—An Emerging Technology. Journal of diabetes science and technology, 3(4), 624-628. 2009.

[21] J. Starren, G. Hripcsak, S. Sengupta, C. Abbruscato, P. Knudson, R. Weinstock, and S. Shea. Columbia university's informatics for diabetes education and telemedicine (ideatel)project. Journal of the American Medical Informatics Association, 9(1):25, 2002.

[22] American Telecare. http://www.americantelecare.com/. 2012.

[23] Viterion. http://www.viterion.com/. 2012.

[24] Montori, V M. Telecare for patients with type 1 diabetes and inadequate glycemic control. Diabetes care, 27(5), 1088-1094. 2004.

[25] ACCU-CHEK. https://www.accu-chek.com/. 2012.

[26] Williams, G C. Supporting autonomy to motivate patients with diabetes for glucose control. Diabetes care, 21(10), 1644-1651. 1998.

[27] Huang, J C. Remote health monitoring adoption model based on artificial neural networks. Expert systems with applications, 37(1), 307-314. 2010.

[28] Wu, J H. Mobile computing acceptance factors in the healthcare industry: A structural equation model. International journal of medical informatics, 76(1), 66-77. 2007.

[29] Speier, C. Generalized self efficacy as a mediator and moderator between control and complexity at work and personal initiative: A longitudinal field study in East Germany. Human performance, 10(2), 171-192. 1997.

[30] Helitzer, D. Assessing or predicting adoption of telehealth using the diffusion of innovations theory: a practical example from a rural program in New Mexico. Telemedicine and e-health, 9(2), 179-187. 2003.

[31] Seto, E. Cost comparison between telemonitoring and usual care of heart failure: a systematic review. Telemedicine and e-health, 14(7), 679-686. 2008.

[32] Census. United Staes Census Bureau. http://www.census.gov/. 2012.

[33] Peto, R. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. British Journal of Cancer, 35(1), 1-39. 1977.

[34] Mohan, C. An efficient and flexible method for archiving a data base. SIGMOD record, 22(2), 139-146. 1993.

[35] Suh, M. Missing data imputation for remote CHF patient monitoring systems. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society,3184-3187. 2011.

[36] Lan, M. WANDA: An end-to-end remote health monitoring and analytics system for heart gailure patients.2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. In Press. 2012.

[37] Suh, M. Dynamic self-adaptive remote health monitoring system for diabetics. 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. In Press. 2012.

[38] Malherbe, C. Circadian variations of blood sugar and plasma insulin levels in man. Diabetologia, 5(6), 397-404. 1969.

[39] Jarrett, R J. Diurnal variation in oral glucose tolerance: blood sugar and plasma insulin levels morning, afternoon, and evening. BMJ. British medical journal, 1(5794), 199-201.1972.

[40] Dellaert, F. The expectation maximization algorithm. Technical Report GIT-GVU-02-20, Georgia Institute of Technology. 2002.

[41] Hartley, H O. Maximum likelihood estimation from incomplete data. Biometrics, 14(2), 174-194. 1958.

[42] Menzies, T. Data mining for very busy people. Computer, 36(11), 22-29. 2003.

[43] Kenchaiah, S. Obesity and the Risk of Heart Failure. New England Journal of Medicine, 34, vol. 5, 305-313. 2002.

[44] Vasan RS. Impact of High-Normal Blood Pressure on the Risk of Cardiovascular Disease. New England Journal of Medicine, vol. 345, vol. 18, 1291-1297. 2001

[45] Agrawal, R. Fast algorithms for mining association rules. VLDB, pp.487-489.1994.

[46] Flach, P A. Confirmation-guided discovery of first-order rules with Tertius. Machine learning, 42(1/2), 61-95. 2001.