34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Dynamic Self-adaptive Remote Health Monitoring System for Diabetics

Myung-kyung Suh[1], Tannaz Moin[2], Jonathan Woodbridge[1], Mars Lan[1], Hassan Ghasemzadeh[1], Alex Bui[3], Sheila Ahmadi[4], Majid Sarrafzadeh[1]

Abstract— **Diabetes is the seventh leading cause of death in the United States. In 2010, about 1.9 million new cases of diabetes were diagnosed in people aged 20 years or older. Remote health monitoring systems can help diabetics and their healthcare professionals monitor health-related measurements by providing real-time feedback. However, data-driven methods to dynamically prioritize and generate tasks are not well investigated in the remote health monitoring. This paper presents a task optimization technique used in WANDA (Weight and Activity with Blood Pressure and Other Vital Signs); a wireless health project that leverages sensor technology and wireless communication to monitor the health status of patients with diabetes. WANDA applies data analytics in real-time to improving the quality of care. The developed algorithm minimizes the number of daily tasks required by diabetic patients using association rules that satisfies a minimum support threshold. Each of these tasks maximizes information gain, thereby improving the overall level of care. Experimental results show that the developed algorithm can reduce the number of tasks up to 28.6% with minimum support 0.95, minimum confidence 0.97 and high efficiency.**

I. Introduction

Diabetes is a chronic disease caused by a shortage of insulin or an inability to properly use insulin in the body. In the United States, 25.8 million people have diabetes with $174 billion spent annually on the disease [1]. One in three Americans is expected to have diabetes in their lifetime, losing 10-15 years from their lifespan. Diabetes can cause complications including blindness, hypoglycemia, renal failures, cardiovascular disease, etc. [2]. Recent studies show that lifestyle changes with careful symptom and health-related monitoring can help to delay or prevent diabetes. Studies have shown that regulating and monitoring various comorbid conditions including blood glucose, symptom, blood pressure, weight, activities, etc. can have a significant positive impact [3][4][5][6]. Remote health monitoring systems can help diabetics and their healthcare professionals monitor health-related measurements and provide real-time feedback. IDEATel is one of the most successful remote health monitoring studies demonstrated between 2000 and 2008 [7]. Diabetic patients participated in taking and uploading blood pressure and blood glucose values, videoconferencing, electronic messaging, and accessing study web pages. The clinical trial included 1665 diabetics and showed improved HbA1c, LDL-cholesterol and blood pressure levels over 5 years compared to control.

In the remote health monitoring domain, patients are required to perform daily tasks provided by their health care professionals. However, the amount of tasks that can be asked of a patient is limited. Patients in Chaudhry's work answered 16 questions and entered their weight using telephone keypads [8]. This study had a high missing data rate due to system non-use. Similar missing data problems were shown in [9]. This study required patients to measure weight, blood pressure, and 12 symptom questionnaires on a daily basis. System non-use, such as those in [8] [9] can severely degrade the effectiveness of remote monitoring systems. Results in [10][11] suggest that a system's ease of use and perceived usefulness is strongly correltated with a patient's behavioral intention.

Task analysis is a process of analyzing how a task is completed including its sequence, frequency, environmental factors, types, etc. In remote health monitoring, task analysis can help schedule sequence of tasks, avoid unnecessary tasks, increase usability and effectiveness of the system, and even predict future emergency events. Most remote health monitoring systems use medical domain experts' knowledge to determine and assign priorities and sequences of tasks. Tang applied a heuristic evaluation method using expert knowledge to identify usability problems in the early stage of software development [12]. Dabbs utilized expert knowledge and survey feedback from patients for designing a health monitoring system [13]. However, most remote health monitoring systems do not apply data-driven dynamic design process for assigning tasks. Therefore, tasks asked of a patient often yield minimal or redundant information gain.

This paper combines domain expert's knowledge and data mining techniques to a mobilehealth monitoring system. This technique was verified on 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 [9]. The WANDA diabetes study was designed in collaboration between the UCLA Computer Science Department and the Ronald Regan UCLA Medical Center. In this analysis, we applied association rule mining techniques to 14 of the subjects with type 2 diabetes enrolled

[1] M. Suh (dmksuh@ucla.edu), J. Woodbridge (jwoodbridge@ucla.edu), M. Lan (marslan@cs.ucla.edu), H. Ghasemzadeh (hassan@cs.ucla.edu) and M. Sarrafzadeh (majid@cs.ucla.edu) are with the Computer Science Department, University of California, Los Angeles, CA, 90095, USA.

[2] T. Moin (tmoin@mednet.ucla.edu) is with the VA Greater Los Angeles Center of Excellence for the Study of Healthcare Provider Behavior and Department of Medicine, Division of Endocrine, Los Angeles CA. 90073

[3] A. Bui (buia@mii.ucla.edu) is with the Medical Imaging Informatics Group, Department of Radiological Sciences, University of California, Los Angeles, CA, 90024, USA.

[4] S. Ahmadi (sahmadi@mednet.ucla.edu) is with the Division of Endocrinology, University of California, Los Angeles, CA, 90024, USA.

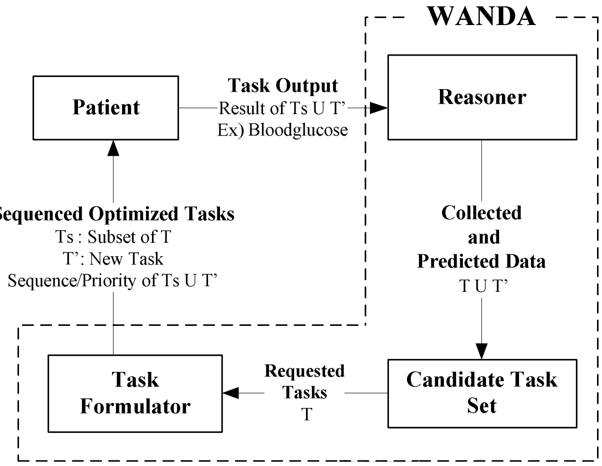

Figure 1 WANDA Dynamic Task Manager Architecture

in the intervention arm for this study. The experimental results show that the developed algorithm can reduce the number of tasks up to 28.6 % with minimum support 0.95 and confidence 0.97.

II. System Design

A. System Architecture

WANDA is built using a three-tier architecture [9]. The first tier comprises of sensors that measure patients' health-related measurements and transmits data to the second tier. The second tier consists of web servers that receive data from the first tier, archive and maintain data integrity; provide patient monitoring applications; and plan assignment tasks. Additionally, data in the back-end databases is used for system backup and recovery, data analysis, missing data prediction [14] , signal search, user task optimization and early adaptive alarms.

WANDA's association rule learning is performed by the third tier. This tier finds data-driven rules and delivers optimized task sets to the user (Figure 1). Tasks are initially defined by domain experts and are accomplished by the patients. Tasks consist of data measurements (such as measuring a patient's blood glucose) and questionnaires. The algorithm derives rules from data obtained from patient's task outputs. Using its derived association rules, the algorithm defines priorities and optimal sequences of tasks and minimizes the number of tasks. If necessary, the system can replace predictable tasks (tasks with minimal information gain) with more informative tasks (higher information gain). The developed algorithm sends a formulated task list to the second tier and the second tier delivers the task to the respective user. The user's sensor readings are delivered to the third tier as an input to predict unmeasured output using generated implications. Therefore, the final output of the system will request a minimal number of tasks while maximizing information gain. If system non-use or data delivery failure happens, WANDA performs missing data imputation algorithms [15] for predicting binomial and non-binomial readings and improving the quality of care in remote patient monitoring.

B. Data Association Rule Learning Algorithms

The back-end database system performs data association rule analysis and missing data imputation [14]. 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 status of a patient. Association rule mining and its feedback are used for reducing the number of tasks required by patients while increasing information gain. Association rule analysis is unlike missing data imputation in that missing data imputation predicts missing values using observed data from the same, previous, or future days. In addition, data predicted by missing data imputation does not affect user task sets. In this paper, we will focus on data association rule learning.

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. The developed algorithm requires generating association rules that have support and confidence greater than the user-defined thresholds, minimum support (s_{min}) and confidence (C_{min}) .

In the data preprocessing step, the developed algorithm performs data cleaning and discretization for removing erroneous data and discretizing timestamp and indexing data. The algorithm discretizes timestamp data as morning, afternoon, and evening. The system also indexes blood glucose and questionnaire response data as multiple measurements, system non-use, and data out of acceptance range. Additionally, information on whether a caregiver contacted the patient for each data is used.

After preprocessing data, the developed algorithm applies the Apriori algorithm [15] to derive rule sets (Figure 2). The Apriori algorithm is efficient for finding all frequent itemsets. The discretized and categorized data are used in the algorithm as inputs. Current rules are derived by looking back a variable number of days (sliding time window). The time window is increased by one day for every iteration. The algorithm calculates the support and confidence of each implication and chooses implications qualifying threshold limits. In each

```
Result := ø;
k := 1;
   while Ck ≠ ø do
   create a counter for each itemset in Ck;
   foreach events in database D do
      Increase the counter of itemsets in Ck which appears in the
      events;
   Lk := All candidates in Ck with support Smin
   Resuit := Result U Lk;
   Ck+ 1 :=  k+1 -itemsets which have all subsets of Lk.
   k:=k+1;
end
```

Figure 2 Apriori algorithm [15]

TABLE 1 PATIENT POPULATION INFORMATION

Group	Total	Male	Female	Avg. Age
Intervention	14	11	3	51.07
Control	14	10	4	66.28

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 qualifying minimum support and confidence in the given time window.

When Apriori returns implication rules, $A \Rightarrow B$, the algorithm calculates their 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. The process stops when there is no new rule and the algorithm returns the final rule set (Figure 3).

The generated rules in *Result* are prioritized based on the confidence 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 other outputs. Since 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 [15].

III. RESULT

A. Subjects and Datasets

This study was approved by the UCLA Institutional Review Board (IRB) and patients were randomized to either intervention or control starting June 1st, 2011. Participants eligible for recruitment were adults with Type 2 Diabetes, HbA1c$\geq$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 14 study participants assigned to the intervention arm (TABLE 1) and the average participation duration is 59.67 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 2). The defined acceptable ranges for blood glucose are between 80 and 200 mg/dL. Acceptable ranges of questionnaire values are denoted in TABLE 2. The timestamp are discretized as morning, afternoon and evening. Collected data are discretized as

TABLE 2 WANDA DIABETES TESTBED QUESTIONNAIRES AND ACCEPTABLE RANGES

Questionnaire Items		Values
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

```
Rule:= ø;
w := 1;
  while Aw*≠ ø do
    Aw := Result from Apriori ;
    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 confidence;
    Rule := Rule U Aw*;
    w := w+1;
end
```

Figure 3 Association rule learning algorithm in WANDA

missing data, multiple measurements and out of normal ranges. Existence of call logs between a patient and caregivers are labeled. The total number of instances used in this study is 1688 and each data has 26 attributes.

B. Results of Data Association Rule Learning

The proposed algorithm had optimal results with a look back window of 2 days. Figure 4 shows the minimum confidence with varying window sizes. The minimum confidence peaks at 0.98 at two and three days. Figure 5 shows the number of new rules added or updated with increasing window size. A total of 12 rules were added with a window size of one day and a total of 2 rules were updated with a window size of 2 days. No rules were added or updated with a look back window of more than 2 days.

The total number of patient tasks was reduced by an average of 28.6% 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 Figure 6, it was found that outputs of Q3 and Q4 are highly correlated. Therefore, the system only needs to ask Q4 (since predicting Q3 with Q4 has a higher confidence than predicting Q4 with Q3). 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 algorithm in [16], the developed algorithm show higher confidence level and higher efficiency. However, since Flach's algorithm finds new rules that WANDA doesn't generate, they can be used for generating

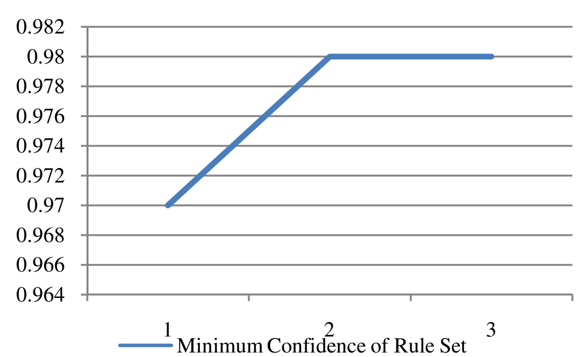

Figure 4 Minimum Confidence of the total rule set with varying window sizes

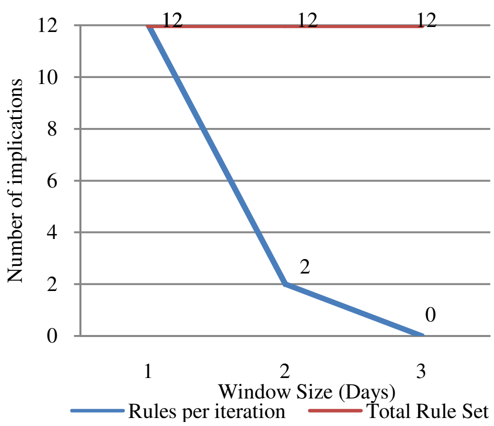

Figure 5 Number of rules added or updated per iteration

helpful tips or reminders.

IV. CONCLUSION

In the United States, 11.8% of all men and 10.8% of all women aged 20 years or older have diabetes and diabetes contributed 231,404 deaths. Remote health monitoring helps patients and caregivers monitor health-related readings and provide feedbacks in real time. However, data-driven methods to minimize or prioritize user tasks are not investigated.

The WANDA system was developed in conjunction with the University of California Los Angeles Computer Science and the Ronald Regan UCLA Medical Center. WANDA monitors patients readings and analyze outputs of sensor readings for improving the quality of care using remote health monitoring.

In this study, we focused on increasing ease of use for system users for improving system adherence. The developed system applies Apriori-based association rule learning algorithms and finds association rules using collected sensor readings with different sizes of sliding windows. The algorithm minimizes the number of action items and reorganizes series of tasks for maximizing information gain.

The first pilot study has approved by the UCLA Institutional Review Board (IRB) which began enrolling patients in June of 2011. In this work, we applied the developed algorithms to 1688 data points from 14 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 up to 28.6% with minimum support 0.95, minimum confidence 0.97 and maximum time window size = 2 days. Since the Apriori algorithm has excellent scale-up properties [15], the developed algorithm can be applied to the remote patient with low complexity.

ACKNOWLEDGMENT

This work was supported by NIH/National Library of

Answer of Q4 = acceptable $\Rightarrow$ Answer of Q3 = acceptable (c = 1.0)
Answer of Q3 = acceptable $\Rightarrow$ Answer of Q4 = acceptable (c = 0.99)
Answer of Q3 = not acceptable $\Rightarrow$ Answer of Q4 = not acceptable (c = 1.0)
Answer of Q4 = not acceptable $\Rightarrow$ Answer of Q3 = not acceptable (c = 0.99)
Answer of Q3 = not acceptable $\Rightarrow$ Multiple measurements in morning (c = 0.97)
Multiple measurements in morning $\Rightarrow$ Answer of Q3 = not acceptable (c = 0.98)
Answer of Q3 = acceptable $\Rightarrow$ No multiple measurements in morning (c = 0.98)
No multiple measurements in morning $\Rightarrow$ Answer of Q3 = acceptable (c = 0.97)
Answer of Q4 = not acceptable $\Rightarrow$ Multiple measurements in morning (c = 0.97)
Multiple measurements in morning $\Rightarrow$ Answer of Q4 = not acceptable (c = 0.98)
Answer of Q4 = acceptable $\Rightarrow$ No multiple measurements in morning (c = 0.98)
No multiple measurements in morning $\Rightarrow$ Answer of Q4 = acceptable (c = 0.97)

Figure 6 Results from the collected data set

Medicine Medical Informatics Training Program Grant T15 LM07356. The authors would like to thank Florence-Joy Figueras, Jane Lee and Nancy Lee for their help in patient recruitment and advice.

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] P. Ciechanowski, W. Katon, J. Russo, and I. Hirsch. The relationship of depressive symptoms to symptom reporting, self-care and glucose control in diabetes. General Hospital Psychiatry, 25(4):246–252. 2003.
[5] S. Vijan and R. Hayward. 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] 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.
[8] 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.
[9] Suh, M. A Remote Patient Monitoring System for Congestive Heart Failure. Journal of medical systems, 35(5), 1165-1178. 2011.
[10] 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.
[11] 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.
[12] Tang, Z. Applying heuristic evaluation to improve the usability of a telemedicine system. Telemedicine and e-health, 12(1), 24-34. 2006.
[13] Dabbs, A D V. User-centered design and interactive health technologies for patients. Computers, informatics, nursing, 27(3), 175-183. 2009.
[14] Suh, M. Missing data imputation for remote CHF patient monitoring systems. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 3184-3187. 2011.
[15] Agrawal, R. Fast algorithms for mining association rules. VLDB, pp.487-489.1994.
[16] Flach, P A. Confirmation-guided discovery of first-order rules with Tertius. Machine learning, 42(1/2), 61-95. 2001.