NEW SYSTEM FORECAST E-HEALTH BY USING HYBRID OF ASSOCIATION RULES AND DECISION TREE

Ahmed EL HADDIOUI, Soumaya EL MAMOUNE, Loubna CHERRAT, Mostafa EZZIYYANI
Abdelmalek essaâdi university
Faculty of Sciences
Tetuan, Morocco,
ahmed.elhaddioui@gmail.com
soumaya.elmamoune@gmail.com
cherrat81@gmail.com
ezziyyani@gmail.com

Tayeb SADIKI and Mohammed BOULMALF
Electronic, Logistic,informatics,and Telecom Laboratory
Technopolis Rabat-Shore, Int. Univ. of Rabat, Sala el Jadida, Morocco
tayeb.sadiki@uir.ac.ma,
mohammed.boulmalf@uir.ac.ma

***Abstract*—E-health technologies can play a major role in improving the lives of patients and especially with chronic diseases, these technologies can provide quality care at a distance without consulting their doctor regularly. These technologies provide many benefits for all stakeholders: patients and health providers to patients, the main advantage of e-health is remote access of this information, and communications services as well as communication. Directly with professionals and health services without having to travel to doctors. For providers of health, the main advantages of e-health is improving access to real patient data remotely and improve the quality of their decision and their services through systems help in the decision.**

We propose in this paper to compare the two techniques of data mining association rules and decision trees in discussing their advantages and limitations depending on the efficiency, accuracy and timeliness in real time. Our discussion leads us to propose a new solution of a system of monitoring and suggestion based on the hybrid use of association rules and decision trees where the association rules in first phase is used to generate the relevant relationships for a homogeneous group of patients who share the same profile and the second phase involves generating a decision tree for each group of patients, from medical relationships that are generated in the first phase and the elimination of transitivity and regenerating the canonical decision tree.

Keywords—e-health; risk prevention; chronic disease; association rule; decision tree.

I. Introduction

One of the biggest challenges facing health systems in the 21st century is the need to try to lighten the heavy burden of chronic diseases, which are complex and require a multi-level response in the long term. In Morocco, the results of the investigation into the 2011 National Population and Family Health (NPFH) show that 18.2% of the Moroccan population is suffering from a chronic illness in 2011 (against 13.8% in 2004). [8]. an example of chronic disease: diabetes is a real public health phenomenon, which affects ≈ 2.5 million people in Morocco in 2010, half of which do not even conscious and requires monitoring every day and Lifetime health professionals.

The exponential increase in the number of people suffering from chronic diseases has led to an increasing burden for health care facilities that becomes a real problem, to solve this problem is to change the current model of health system and use the equipment mobile technology coupled with wireless networks to monitor patients remotely and in real time to reduce the costs of hospitalization. The current e-health systems have been designed primarily to collect and convey patient data remotely via sensors and Smartphone in real time to health care providers who can treat them. The data is not intended to be processed in real time, but stored for later examined by doctors and specialists to make decisions and recommend lifestyle. So our objective is to complete the system health monitoring e-health system data mining from medical data which are collared in a given patient and the problem of non-availability of physicians.

This article is organized as follows, section 2 presents a state of the art research work that investigated the challenges of building information remote monitoring systems and discusses their strengths and limitations in section 3 presents the requirements necessary to complete the system of health monitoring system by exploring medical data, Section 4 presents a comparative study of data mining techniques, Section 5 presents our proposal to complete the e-health system. Section 6 presents the conclusions and directions for future work.

II. STATE OF THE ART

Several research works have investigated the challenges of building information remote monitoring E-Heath systems and solutions were differentiated by how researchers approach the issue of integration given the heterogeneity of systems architectures used, security, confidentiality, communication, the characteristics of detection sensors, managing groups of assistants.

Among the proposed solutions, there is a system of electronic health SOCBeS [1], a system for monitoring invasive health PHMS (System on Pervasive Healthcare Monitoring) [2], the RHM (Remote Health Monitoring) system [3] monitoring system intelligent mobile health (IMHMS) [5]. The first solution has several advantages; it facilitates integration with other systems of care current or future health and develops new services, integrating different detection technologies through the creation of interfaces that allow the adoption of different protocols, including, for example, Bluetooth, 3G, and Wi-Fi, allow synchronization of data between different storage depots and cloud data centers, using a portable device for the detection and collection of data while optimizing (minimization) power consumption of the device and processing capabilities, rely on the use of cloud services while ensuring high quality of service (QoS), data privacy and security and gestion.la second solution ensures the confidentiality and integrity of data with authentication in an environment of pervasive health. The third solution proposes an architecture of three-tier system of health monitoring in real-time area of the body which consists of sensors and a gateway: body sensors to monitor the physiological state of the patient and ambient sensors to monitor conditions of the environment around the patient as air temperature, humidity and a gateway that collects information from the sensor and send them to the field of online service, network and communication field, is to connect the two areas via wireless communication technologies, service area: who is responsible for receiving, recording and analyzing information related to the user health, the server takes an accurate and complete picture of the effects (including the healing effect and side effects) medical treatment taken by the patients followed. So the server is able to provide valuable information for new patients taking the same medication or treatment; fourth solution allows the collection of patient data by biosensors, data mining, intelligent guesses the state of health of the patient, provide information to patients via their mobile devices.

All these solutions have the common drawbacks of these drawbacks is found that: Data are analyzed manually and there is no usage of data mining explore useful knowledge, there are only the comparison of the measured values relative to the previously set threshold values in the system.

III. PREPARE REQUIREMENTS OF THE PROPOSED SOLUTION

In this section, we present the requirements for the complete system health monitoring system by data mining from medical data: the system of health surveillance should be appropriate to read critical health data collected to differentiate the signals of a life-threatening situation from those of normal variation due to the ever-changing activities. Otherwise, such a search service first proactive local emergency triggered by incorrect decisions will result in a large amount of manpower and medical resources [3], as there are always constraints to the disconnection Internet, or patients who travel in areas not covered by the Internet zones, the system must have a offline mode that allows the treatment of minimal data in the event of unavailability of server processing, and in the case of an urgent response is required, the mobile application must continue the process of data collection and dissemination of emergency the people closest support immediate first aid, the collection of health data and the frequency call which these data are collected should be smart to take into account the various contexts [1], as a former physical location, gestures, health, medical, and temperature of the environment, the context information should be obtained by infrastructure sensors formed on the basis of needs and seniors pathologies. Accordingly, the context information suggests a high degree of heterogeneity, which makes it necessary to carefully determine the mechanisms and tools to transform the raw data to information about context [6], is a medical major area of application of data mining. Through data mining, we can extract knowledge and interesting patterns, the extracted information can then be applied in the field corresponding to increase work efficiency and improve the quality of decision making [7].

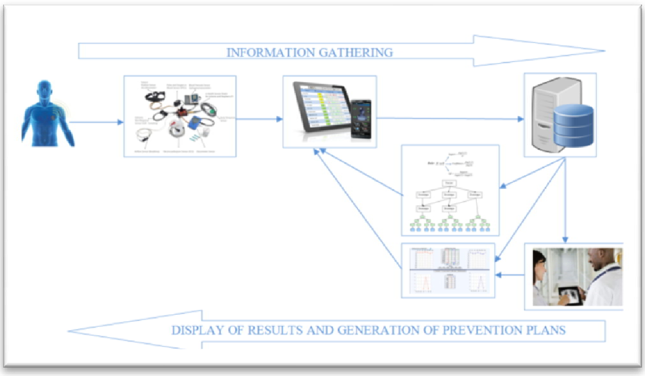

Fig. 1. SystemArchitecture for e-health

The e-health patient Smartphone system must work in two modes, online mode: each connection between the patient and Smartphone central processing server, the system must collect medical data of the patient are collared via body sensors in real time, send the data to the staff Smartphone patient, send the data to the central processing server to treat the basis of decision trees that are generated according to the patient profile, upload the new decision tree and update its basis of decision which are already stored, dynamically generate results and prevention plans based on the patient profile and the measurement context, download the results of treatment and displayed as a summary file or a graph and update based on historical results, validate the interpretation of results by the physician and patient to confirm plans via his Smartphone and detect the risk of worsening of the disease where the central system of e-health to send alerts as SMS or Mail or phone call to the doctor who follows the state of the patient or to a support group. In offline mode, the e-health patient Smartphone system must be able to work continuously despite it is not connected to a central server that is to say the system must process incoming data the patient based decision tree that are already downloaded from central processing server last connection, so the system should react and behave as the central e-health system to collect medical data of the patient are collared via sensors body in real time, send the data to the staff Smartphone patient, treat the patient medical data which are collared via body sensors in real time based on decision tree that are already downloaded, view the results of treatment in the form of a summary file or a graph and update based on historical results and dynamically generates prevention plans based on the patient profile and context, if the system detects a risk of worsening of the disease, must send alerts as SMS or Mail or phone call to the doctor who follows the state of the patient or to a support group.

IV. COMPARATIVE STUDY OF DATA MINING TECHNIQUES

The medical field is an important application of data mining field. Through data mining, we can extract knowledge and regularities hidden, interesting, and relevant from a very large amount of raw data. The extracted information can then be applied to increase the work efficiency and improve the quality of decision making, but the problem of making decisions for chronic diseases need at the same time high accuracy and speed of execution. There are several data mining techniques applied to medical field, among them mining techniques there: Association rules and decision tree, we discussed the advantages and limitations of each technology depending on the efficiency, accuracy and speed of real-time execution.

We distinguish the supervised data mining that is used primarily for data classification and unsupervised data mining that is used in research associations or groups of individuals. The scope of data mining extends loading and cleaning of data ETL (Extract, Transform and Load) in databases, the formatting of the results through the most important: the classification and linking of different data.

A. Decision trees

A decision tree is, as its name suggests, a tool for decision support that divides a population of individuals into homogeneous groups according discriminating attributes based on a fixed and known goal. It allows you to make predictions based on the known problem reduction, level by level, in the field of data solutions.

Each internal node of a decision tree provides a specifying attribute element to classify which distribute these evenly between children of this node. Branches linking a node to represent his children discriminate values of the attribute node. And finally, the leaves of a decision tree are his predictions for the data to classify.

It is a method that has the advantage of being readable for analysts and identifies couples < Attribute, Value> discriminating from a very large number of attributes and values.

B. The association rules

The association rules are rules that are extracted from a transactional database (item-set) that describe associations between certain elements. This technique allows to highlight the associations between commodities (essential goods, for which the client moves) and additional products, which can implement business strategies to increase profits by promoting, for example, the additional sales.

C. Discussion

Decision trees have several advantages. Mention may be made, the simultaneous consideration of qualitative and quantitative variables, the selection of the most informative variables, and producing logical classification rules easily interpretable, ease of implantation, the readability of decision rules, and speed execution, without loss of optimality.

These shafts have a number of limitations: the digital processing variables, the choice of thresholds takes into account not only the properties of density (area) values. They are susceptible to changes, which causes variations in the trees and performance products.

Association rules also have several advantages: simple and easy to understand techniques and the results are clear and easy to interpret when there are not too many rules generated but there are limitations include: the volume of calculations increases square or cube of volume data, operating results difficult when there are many rules and useless when highly connected data without significant support is sought.

In the medical field, association rules are a promising technique for improving the prediction of disease. Unfortunately, when association rules are applied to a set of medical data they produce a large number of rules, most of associations rules applied to a set of medical data are clinically relevant and the time required for finding may be impractical.

According to the comparative study of association rules and decision trees for disease prediction made by author [6] approved that decision tree is less efficient than forced association rule: decision trees are not as powerful as the association rules use a set of numeric attributes binned manually and categorical attributes and target multiple attributes related, decision trees do not work well with combinations of several target variables (arteries), this you must define a class attribute for each combination of values, decision trees fail to identify many clinically relevant combinations of independent variables spans digital and nominal values (ie perfusion measurements and risk factors), the decision trees to find some predictive rules of reasonable size (> 1%) sets of patients, which is well known as a data set fragmentation [10] problem, a decision tree is a predictive model of the dataset while association rules are disconnected them, a decision tree is guaranteed to have at least 50% forecast accuracy and generally above 80% accuracy for binary target variables, while rules Association specifically require trial and error works to find a good or acceptable threshold.

TABLE I. summary of the comparative study

	Decision Trees	**Association rules**
Timeliness	+	-
Efficiency	-	+
Quality results	+	-
Visibility	+	+
Easy to implementation	+	-
Performance	-	+

Cost calculation	+	-
Medical field	-	+

Where: "+" is the best technique
And : "-" is Bad technique.

V. PROPOSED SOLUTION

The problem of decision making for chronic disease surveillance system remotely and in real time at the same time need a high accuracy of the decision, and speed of execution in real time, view the benefits and limitations of each mining techniques, and to meet all the requirements of future systems (section 3) we propose a new solution of a system of monitoring and suggestion based on the hybrid use of association rules and decision trees.

The new system solution is divided into two treatment phase, the first phase includes a generating relations medically necessary by the use of association rules for each group of patients who are cataloged according to their profile (age, sex, disease, activities, habits, ...) between all medical history medical measures, risk factor, environmental factor (former physical location, gestures, health, environmental temperature), the context of levy , trust (accuracy) and the support to ensure the accuracy of decision.

Trust and support are the parameters used to evaluate the medical importance and reliability of association rules generates, these parameters must be equal to or above the thresholds defined by the expert according to its objectives and the type of disease being treated.

The second phase involves generating a decision tree for each group of patients, from medical relationships that are generated in the first phase (the size of database rules is reduced).

To ensure the process execution speed, the system must treat the patient connected medical measures using the decision tree based on their profile, such treatment should dynamically generate results and prevention plans.

VI. MATHEMATICAL MODEL DATA

In this section, we present a mathematical model of data used to develop an algorithm for automatically generating decisions for a given patient.

Pi = patient characterized with set of attribute.

Pi = {Age, Gender, weight, size, diseases, disease extent, the result of examinations, laboratory tests, activities, habits, geographical position ...}.

Let GP(ProfilPatient) i = GP1, GP2, GP3, ... GPn a whole group of patients classified by profile.

GP (ProfilPatient) i = {Interval old, weight range, range size, list of activities, list of symptoms, geography ...}.

D (ProfilPatient) = set of decisions generated by the use of hybrid association rules and decision tree for each group of patients

Step 1: Generate the association rules for each group of patients:

$$Support\ (Symptom\ Pi \rightarrow Disease\ Pi) = (Symptom\ Pi \wedge Disease\ Pi) / D$$

Symptom Pi ∧ Disease Pi: Number of lines where Pi gives disease symptom is true
D : the total number of rows (transactions)
Pi = patient characterized by a set of attribute
Pi = {Age, Gender, weight, height, disease, disease extent, the result of examinations, laboratory tests, activities, habits, geographical position ...}

The confidence of a rule is defined by:

$$Confidence\ (Symptom\ Pi \rightarrow Disease\ Pi) = Support\ (Symptom\ Pi \rightarrow Disease\ Pi) / Support\ (Symptom\ Pi)$$

The association rules generated are included if they have a higher minSupp support and confidence greater than MinConf. These two constants are defined by the experts.

Step 2:

Building the decision tree based association rules generated in the first step by using C4.5 to remove unnecessary rules.

Step 3:

Transform the decision tree in the canonical form consists mainly two specific tasks:

- It describes the links between the nodes of the tree which will bring all the unambiguous treatment, which helps to distinguish them and order them.
- It refers to an appropriate methodology for treating a member selected from the shaft conventionally a set of elements with similar properties.

In order to transform the decision tree AD in the canonical form and respect the two rules of the canonical form of a decision tree, we apply the algorithm (under development) which remove weak arcs, eliminating the farm tours and dispose of transitivity in the decision tree AD.

Step 4: To find the right decisions for a given patient must seek first what patient group GP (ProfilPatient) i belongs to that patient Pi:

$$Pi \in GP(ProfilPatient)i \rightarrow Decision(ProfilPatient)i$$

To calculate the degree of membership between the Pi and the GP (ProfilPatient) i:

$$Degree_{membership(Pi,GPi)} = weight(Pi)$$

Weight (Pi) is the occurrence of Pi similar number in the transactional database of patient group Pi.

For Pi ∈ GP(Profil_Patient)i if and only if the Degree_membership (Pi, GPi) is max compared with the other.

Step 5: If the patient connects treatment causes changes in association rules, this update involves the automatic update at decision tree generates.

A. *Proposed algorithm*

- Initialize (base_rules_association B_RA)
- Initialize (Tree).
- Initialize List membership degree
- Classify_patient_profile (patients Pi)
- Generate_Rule_Association (GPi)
- Generate_Decision_Tree (B_RA)
- CONONIQUE (Tree AD) {.
  - WHILE nombre_lien (Node N)> 1 DO
    - Removes_transitivity (AD, N)
  - END WHILE }
- FOR each patient connects Pi $\in \{S^i\}$

  DO

  - Calculate_degre_membership (Pi, GP (ProfilPatient) i)
  - Store_list_ degree _membership (list_DA, DA)
  - MAX _ DA (list_DA)
  - Choose_GP (max_DA, Pi)
- Tester attributes patient Pi with tree generated for this group GP(ProfilPatient)i
  - IF Attri_1 < α AND Attri_2< β THEN
    - Attri_ 3 < γ
  - IF Attri_1 < α AND Attri_4 < β THEN
    - Attri_8 < δ
  - IF α < Attri_1 < γ AND Attri_5 = λ AND Attri_6 < ε AND Attri_7 < β THEN
    - Attri_8 < ζ
  - IF α < Attri_1 < γ AND Attri_5 = λ AND Attri_9 < ε AND Attri_10 < β THEN
    - Attri_11 < η
- Construction of the decision tree AD
- Displays the final decision

B. *Description of they used functions*

- Classify_patient_profile (patients Pi): classification of patients by profile.
- Generate_Rule_Association (GPi): generation rules binding for each group of patients.
- Generate_Decision_Tree (B_RA): generation of decision trees.
- Removes_transitivity (AD, N) : eliminates the transitivity in the decision tree AD.
- Calculate_degre_membership (Pi, GP (ProfilPatient) i) : Calculate the degree of membership between the patient Pi and the GP (ProfilPatient) i
- Store_list_ degree _membership (list_DA, DA) : Store in the degree of membership list
- MAX_DA(liset_DA) : Compare the degrees of membership and return the max degree.
- Choose_GP (max_DA, Pi): select the group of patients who owns this patient.

VII. CONCLUSIONS AND PERSPECTIVES

The exponential increase in the number of people suffering from chronic disease is a health problem that automatically increases hospital expenditures and Ministry of Health to reduce these expenses must migrate the current health care system based on the monitoring of patients suffering from chronic hospitals to a system based on mobile monitoring these patients remotely and in real time by the use of new communication technologies and mobile.

In this paper, we presented a new solution for complete monitoring system for e-health system data mining by the hybrid use of association rules and decision trees.

Future work is to complete and test the algorithm based on the hybrid use of association rules and decision tree that ensures accuracy and decision speed of execution in real time and performs a case study to test well illustrate our approach: monitoring chronic illness gives by this monitoring system to evaluate this solution.

REFERENCES

[1] A.Benharref, M.Serhani, "Novel Cloud and SOA Based Framework for E-Health Monitoring Using Wireless Biosensors" Biomedical and Health Informatics, IEEE Journal of (Volume: PP, Issue: 99) Date of Publication: May 2013.

[2] N RukmaRakha, Prof.MS PrasadBabu "Secure Framework for Pervasive Healthcare Monitoring Systems", International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.2, No.2, April 2013.

[3] X. Liang, X. Li, M. Barua, L. Chen, R. Lu, X. Shen, and HY Luo, " Enable Pervasive Healthcare through Continuous Remote Health Monitoring ", IEEE Wireless Communications, vol. 19, no. 6, pp. 10-18, 2012.

[4] Khalid Elgazzar Muhammad Aboelfotoh, Patrick Martin, Hossam S. Hassanein, "Ubiquitous Mobile Health Monitoring Using Web Services", Procedia Computer ScienceVolume 10 , 2012, Pages 332-339 MobiWIS ANT 2012 and 2012.

[5] R.Shahriyar, Md.F.Bari, G.Kundu, SIAhamed, Md.M.Akbar Intelligent Mobile Health Monitoring System (IMHMS) International Journal of Control and Automation Vol.2, No.3, September 2009.

[6] C.Ordonez "Comparing Association rules and decision trees for disease prediction" Proc. ACM Workshop on Healthcare Information and Knowledge Management (HIKM, CIKM Conference Workshop) p. 17-24, 2006.

[7] Jing-Song Li, Hai-Yan Yu and Xiao-Guang Zhang (2011), "Data Mining in Hospital Information System, New Fundamental Technologies in Data Mining", Prof. KimitoFunatsu (Ed.), ISBN: 978-953-307-547-1, InTech, DOI: 10.5772/14052. Available from:

http://www.intechopen.com/books/new-fundamental-technologies-in-data-mining/data-mining-in-hospital-information-system.

[8] National Survey of Population and Family Health (enpsf-2011), Ministry of Health.

[9] National Survey on Population and Family Health (enpsf) 2011, Department of Planning and Financial Resources, Division of Planning and Studies, Office of Academic and health information.

[10] T. Hastie, R. Tibshirani, and JH Friedman.The Elements of Statistical Learning.Springer, New York, 1st edition, 2001.

[11] A. Abou El Kalam, Y. Deswarte, G. Trouessin, "Managing anonymized medical data: problems and solutions", the second Francophone Conference of the Hospital Management and Systems Engineering (GISEH'2004), Mons, Belgium, 9 -11 October2004.