Real-Time Decision Support in Intensive Medicine: An intelligent approach for monitoring Data Quality

Filipe Portela and Manuel Santos
Algoritmi Center, University of Minho, Azurém, Guimarães, Portugal
Email: {cfp, mfs}@dsi.uminho.pt

José Machado and António Abelha
CCTC, University of Minho, Braga, Portugal,
Email: {jmac, abelha}@di.uminho.pt

Álvaro Silva and Fernando Rua
Serviço de Cuidados Intensivos, Centro Hospitalar do Porto, Hospital Santo António, Porto, Portugal
Email: moreirasilva@clix.pt, fernandoruasci@hgsa.min-saude.pt

Abstract—Intensive Medicine is an area where big amounts of data are generated every day. The process to obtain knowledge from these data is extremely difficult and sometimes dangerous. The main obstacles of this process are the number of data collected manually and the quality of the data collected automatically. Information quality is a major constrain to the success of Intelligent Decision Support Systems (IDSS). This is the case of INTCare an IDSS which operates in real-time. Data quality needs to be ensured in a continuous way. The quality must be assured essentially in the data acquisition process and in the evaluation of the results obtained from data mining models. To automate this process a set of intelligent agents have been developed to perform a set of data quality tasks. This paper explores the data quality issues in IDSS and presents an intelligent approach for monitoring the data quality in INTCare system.

Index Terms—intensive care; data quality, real-time; intelligent decision support; intcare.

I. INTRODUCTION

Nowadays and due the importance which the patient data have for the decision making process, the number of data present in Intensive Care Units (ICU) increased significantly [1], [2]. This situation motivates the use of Intelligent Decision Support System (IDSS) to obtain knowledge. However the implementation of this type of system is difficult due to the complexity of the environment; the number of data which are regularly collected in paper format [3] and the quality of information / data collected [4]. To overcome these limitations, a system called INTCare [5], [6] was developed with the objective to predict the organ failure and patient outcome for the next 24 hours. This system uses Data Mining (DM) techniques and some variables presented in ICU to predict the targets.

At the beginning most of those variables were collected manually, in paper format and in an hourly and offline base mode. In order to resolve this problem an electronic data acquisition system was developed. This new system gets values from the, patient sensors (vital signs, ventilation), electronic health record (EHR), laboratory and pharmacy system.

The data collected from the sensors attached to the patients require a more carefully analysis on the data quality provided. The sensors can be easily disconnected from the patients; measuring bad values and gives a wrong idea of the patient situation. At same time many of the values haven’t a correct patient identification. These situations can influence significantly the Data Mining (DM) models and the decision making process.

With the objective to validate the values collected automatically, was necessary to implement some procedures. When the data is collected it needs to be processed and transformed according to DM input variables.

Finally, the models are induced in real-time using online learning and some new knowledge will be available. At this point some other problems arise, like is the assessment of the quality of the results obtained by the models. All process reported before is executed automatically and in real-time. The difficult of validate each process automatically is evident.

The most difficult it is the process of evaluating the DM models in a continuous and real-time way. The quality of models developed by INTCare is evaluated using an ensemble and an agent analyses the quality measures defined to each result whenever a new prediction is done. This agent presents the best data mining results / models or refreshes the models according to the evaluation of the results obtained.
The main objective of this paper is to present an intelligent and integrated data quality system. This system increases the quality of information used in the decision process by the ICU professionals. This paper is divided in seven chapters. After an introduction to the subject a set of concepts are addressed in second chapter.

The third chapter presents the quality process and the agents designed. The fourth chapter evaluates the data acquisition process in terms of bedside monitored values, patient identification and process dematerialization. The DM process is evaluated in the fifth chapter. Finally, in the sixth chapter some results are analyzed and in the seventh chapter some conclusions and future work are outlined.

II. BACKGROUND

A. Intensive Medicine

Intensive medicine (IM) is a particular area of Medicine. It has specialists which apply their knowledge in Intensive Care Units (ICU). ICU is recognized as a critical environment because it has some complex health care situations and their patients are in very weak health conditions [7].

The professionals of ICU spend more time in the patient direct care than in data documentation [8]. Due to the reduced number of automatic acquisition system, this attitude can cripple a system like it is INTCare. Being the direct care, correctly always the first concern the solution is automating as possible the acquisition and data quality processes [6]-[9].

This solution reduces the human efforts with data documentation and makes it feasible to introduce an IDSS in intensive medicine.

B. INTCare

INTCare is a research project and has as main objective develop a Pervasive Intelligent Decision Support System to automatic and in real-time predict the organ failure (renal, coagulation, hepatic, cardiovascular, neurologic and respiratory) and patient outcome.

During the development of the project a set of change in the ICU were done. A gateway was implemented to obtain the vital signs values from bedside monitors. Some changes in protocols of obtain the data from Laboratory and pharmacy systems were done.

The objective of these changes was increase the number of data available and getting the patient results and therapeutic plans in an open format.

C. Data Mining variables

The idea of designing this system arose after obtaining good results using offline data. These results [10] were obtained during a studied using EURICUS database which showed be possible using some patient variables predict the organ failure.

The variables used were: Age, Critical Events, Admission Variables, Outcome, and SOFA. During the INTCare some other variables were added to the Data Mining Input:

a) SOFA Cardio, Respiratory, Renal, Liver, Coagulation, neurologic;
b) Case Mix = {Age, Admission type and admission from};
c) Critical Events Accumulated (ACE) = {ACE of Blood Pressure, ACE of Oxygen Saturation, ACE of Heart Rate, and ACE of Urine Output}
d) Ratios;
e) Outcome;

To ensure the quality of the values used, which are now collected in real-time it is necessary a constant and automatic validation of those variables. After the induction of the models and having in account that INTCare also is an adaptive system it is too important have the possibility to automatically and in real-time assess the quality of models and induce new models when necessary.

D. Oracle Data Mining

One of the most difficulty to implement INTCare was the possibility of perform all KDD tasks automatically and in real-time. Most of the solutions explored don’t allow a data transformation, model induction or apply the results using always a database. After do a depth research a solution was found.

Oracle Data Mining (ODM) provides a comprehensive collection of Data Mining analytics as part of the Oracle database environment which supports the development, integration and deployment of Data Mining applications [11]. All KDD tasks are now performed using ODM.
The process has two starting points: having some new data or there is an INTCare prevision request. In the first case the validity of data collected are analysed.

In the second case the Data Mining models are executed and the quality of the scenarios created is evaluated.

The Data quality process is ensured by three agents. However there are other four agents which their tasks are dependent from the data quality process. The agents in use or directly conditioned by this process are:

- **Pre-processing Agent** is integrated in Data Acquisition sub-system and verifies all data collected automatically. For each value collected it analyses if the value is possible or not.

- **Data Mining Agent** executes the data mining engine always some new request to improve data mining models is done. By default it is executed once a day or always the INTCare system is used by first time.

- **Model Initialization Agent** is responsible by answer to the INTCare request. After receive the message the agent processes the models according the request.

- **Prediction Agent** processes the results presented in knowledge base using the data warehouse. This agent only is executed when scenarios evaluation agents gives some indication about quality of scenarios.

- **Scenarios Evaluation Agent** is the other agent which has data quality tasks. This agent has the responsibility of evaluate the results provided from prediction agents. According the quality of the models it sends a message to the data mining agent to induce new models or send a message to the prediction agent with the results of the scenario chosen.

- **Data Retrieval Agent** is responsible to ensure that the data in use by prediction agent are correct and are the most recently collected to each patient. This agent is used to cases where there are subsequent changes, as is the patient outcome – patient is discharged in live condition and days after he die, or cases where since models are executed until the results are obtained some new data arrive. This agent is also responsible to delete rows with null values or values not recognized.

- **Interface Agent** is executed by indication of prediction agents and only has the task of put the results available in the INTCare interface.

The agents are integrated in decision making process and allow having data with high level of quality for automatically and in real-time attain:

- a) **Intensive Care Scores**[14];
- b) **Patient Critical Events**[15];
- c) **Predict organ failure and patient outcome**[16].

IV. EVALUATION OF DATA ACQUISITION VALUES

The process of validation values in real-time and in areas like is IM is too important and isn’t always correct.

A. Bedside Monitors Values

The values validated by this process are collected by the bedside monitor (BM). BM is used to collect vital signs data and through patient sensors. These sensors are connected to the patient and can be easily moved by them or the nurses during the patient care. Proving this fact they are for example the temperature and diastolic sensors. For the temperature if the sensor is connected to the patient is measuring the body temperature (~36º) otherwise is measuring room temperature (~22º). In case of diastolic can be measuring the diastolic or, if moved, systolic values. To resolve these errors an automatic procedure was implemented. This procedure is started by a trigger always some new value is inserted in database.

The validation process has two stages. First, the values collected (Spo2 and Heart Rate) are validated according the values present in Table I. Second, for the Blood Pressure (BP) the validation of Mean Arterial Pressure (MAP) is done. It is defined that MAP is \((1/3(\text{systolic blood pressure}) + 2/3 (\text{diastolic blood pressure}))\) [17]. The procedure compares at the same time if the BP values collected (systolic and diastolic) are between the possible ranges and if the MAP of these values are upper than 40 (minimum value possible)[18].

Table I presents the range of the values that normally can be accepted in ICU (ICUMIN and ICU MAX); it also presents a second range, which contains some abnormal values that can be verified (Amin and Amax). In the automatic pre-processing phase is been used Amin and Amax. These ranges are defined in accordance with the ICU doctors and with the patient data reality.

<table>
<thead>
<tr>
<th>Vital Sign</th>
<th>ICUMin</th>
<th>ICUMax</th>
<th>AMin</th>
<th>AMax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood Pressure (BP)</td>
<td>50</td>
<td>180</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td>SPO2</td>
<td>80</td>
<td>100</td>
<td>40</td>
<td>100</td>
</tr>
<tr>
<td>Temperature (Temp)</td>
<td>34</td>
<td>42</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>Respiratory Rate (RR)</td>
<td>0</td>
<td>40</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>Heart Rate (HR)</td>
<td>5</td>
<td>40</td>
<td>0</td>
<td>250</td>
</tr>
</tbody>
</table>

In this process there are other data quality procedures. At same time it validates the data inserted, i.e., verify if there isn’t a negative value, in case of scales the value collected need to be between the max and min number possible. In other case if only numbers are allowed, this agent ensures that data collected are numbers. If some value doesn’t achieve these requirements the respective row is deleted. All of the data can be manual validated through Electronic Nursing Record (ENR). This situation allows having values out of range when valid. If some value is out of range it is deleted automatically. In the ENR only appear the values automatic validated. The humans can delete the automatic value and put manually the correct. When the value is manually inserted the automatic validation didn’t do nothing, because only the professionals can know if some value is true or not, independent it is or not out of range. The values manually inserted are considered valid and are used by the DM Models. In case of be verified some error the humans can correct this values.

B. Patient Identification

One of the most commons problems in ICUs is the patient identification (PID). The PID is put manually in the bedside monitors; normally the nurses are concerned with the patient and not with putting the correct
identification. This situation leads to an automatic acquisition of a high set of values without identification.

In order to overcome this situation two solutions were tested. One is verify the PID through the Electronic Health Record (EHR), crossing the Bed Number and the information of the patient admitted in this bed. The procedure verifies the bed and consults in the EHR the respective PID. Then always some new value is collected the procedure update the data automatically putting the correct PID. The second solution still is in tests and uses RFID technologies [19], [20] to identify the patient. When the patient comes to ICU a tag is put in the arm. Then the nurses associate in the EHR: the PID, the patient bed and TAG ID. All beds have one antenna to read the tag and identify the patient. When the patient is in the bed, the RFID system reads the TAG and identify the patient [21]. When the patient goes out, the RFID can’t detect any patient and stops the vital signs registry. Both processes use intelligent agents to perform the tasks described above.

C. Process Dematerialisation

Other change was in the record process through the processes dematerialization. Now, instead of the professional records the patient values manually in the paper, they have an electronically, online and touch-screen platform: Electronic Nursing Record (ENR) to registry the clinical data. This platform doesn’t allow text writing. All the data possible are predefined or number. This situation increases the data quality because all of them record the data in the same way, i.e., the text records aren’t allowed because they have to choose a predefined value or a number. The ENR is not only used to register but also to validate and confirm the values that is automatically collected. ENR make available all data than before were recorded manually in the paper.

V. EVALUATION OF DATA MINING MODELS

In order to ensure the quality of the models developed some evaluation procedures were designed. Like presented in figure 1 always some request is done to the Model Initialization Agent the knowledge base is used to predict the targets. For each scenario a set of evaluations are done. At the end of comparing all models, a DM ensemble is induced. INTCare System uses the models are done. At the end of comparing all models, a DM ensemble is induced. INTCare System uses the models presented in figure 1 always some request is done to the RFID system reads the TAG and identify the patient values manually in the professional records. Like presented in this paper, they have an electronically, online and touch-screen platform: Electronic Nursing Record (ENR) to registry the clinical data. This platform doesn’t allow text writing. All the data possible are predefined or number. This situation increases the data quality because all of them record the data in the same way, i.e., the text records aren’t allowed because they have to choose a predefined value or a number. The ENR is not only used to register but also to validate and confirm the values that is automatically collected. ENR make available all data than before were recorded manually in the paper.

The ensemble is organized in terms of six independent components (target) considers seven different scenario and applies three distinct DM techniques: Decision Trees (DT), Support Vector Machine (SVM) and Naive Bayes (NB). The ensemble process is divided in two steps:

- Predictive Models – 126 models are induced combining seven scenarios (S1 to S7), six targets and three different techniques (SVM, DT and NB);
- Ensemble – all the models are assessed in terms of the sensibility, accuracy and total error.

The models are induced in real-time and using online-learning by the DM agent. This agent runs whenever a request is sent or when the performance of the models decreases. In order to choose the best predictive model for each target, a set of tasks are performed automatically and in real-time:

1) Create the confusion matrix for each scenario;
2) Obtain the assessment measures;
3) Apply the quality measure;

For each model 10 runs are performed. Then the confusion matrix (CMX) is automatically obtained for each models. Through the CMX is possible obtain: Number of false positives (FP), Number of true positives (TP), Number of false Negatives (FN) and Number of true negatives (TN). Then, a procedure is executed recursively by scenarios agents for each model (M) in order to obtain the measures:

\[
\text{SENSIBILITY}(M) = \frac{TP(M)}{TP(M) + FN(M)}
\]

\[
\text{ACCURACY}(M) = \frac{TP(M) + TN(M)}{TP(M) + TN(M) + FP(M) + FN(M)}
\]

\[
\text{TERROR}(M) = \frac{FP(M) + FN(M)}{TP(M) + TN(M) + FP(M) + FN(M)}
\]

The main goal of the ensemble is to select the most suited model from a set of candidates. In order to assess the models quality, a measure was defined. This measure is based in the results obtained by the models during the 10 runs in terms of sensibility, accuracy and total error. The selected models are used only if they satisfy the following conditions: Total Error <= 40%, Sensibility >= 85% and Accuracy >= 60%

VI. RESULTS

The quality process is divided in a set of different continuous and integrated tasks. This process are performed automatically and in real-time.

<table>
<thead>
<tr>
<th>TABLE II. ICU VITAL SIGNS DATA ANALYSIS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>HR</td>
</tr>
<tr>
<td>SPO2</td>
</tr>
<tr>
<td>MAP</td>
</tr>
<tr>
<td>Systolic</td>
</tr>
<tr>
<td>Diastol</td>
</tr>
<tr>
<td>CVP</td>
</tr>
<tr>
<td>TEMP</td>
</tr>
<tr>
<td>RR</td>
</tr>
</tbody>
</table>
At first level are the data collected automatically by bedside monitors. The Table II presents an overview of the data collected in ICU. In 2009 most of the data were out of range [4]. Now and after talk with nurses about the importance of maintain the sensors connected, only Central Venous Pressure (CVP) (~40%) and temperature (~15%) present bad values. The quality system validates the values collected. The hours of values out of range are associated to the tasks of hygienic (10h-12h) or patient visits (15h-17h).

At the second level it is the patient with correct identification and the records with predefined data. Making a comparison between 2009 and 2012, at level of patient identification and data format, it is possible observe that in 2012 all data collected has a PID and are available electronically, online and in real-time. These changes result in an increase of the data quality.

Finally, it is assessed the quality of DM system. The data quality of the DM Models can be measured using the procedures earlier described. To a better comprehension of the quality results by the user a monitoring interface were implemented. The monitoring system is similar to a traffic light. Using this system the user can know, on moment when he is consulting the prediction, the quality level of results presented by the selected models. Each label corresponds to one of the target quality measure and it is filled with one colour (green, yellow and red). Table III shows the ranges, where first level is green (very good models), the second level is yellow (acceptable models) and the third level is red (excluded models).

![Table III. Evaluation Criteria](image)

Table IV displays the results achieved by the ensemble having in account the quality measures. At bold it is the results that achieve the quality measure and at red it is the rejected results. Looking to the table is possible observe that only 50% of the targets present satisfactory results. The INTCare system using the quality system only shows prevision to renal, cardiovascular and respiratory system. Due the fact of the data are in constant changes this values can be quickly modified and other models present best results.

![Table IV. Ensemble Performance](image)

**VII. Conclusions & Future work**

The introduction of the data quality component brings direct benefits to the IDSS. For a real-time system, is fundamental to assure the data quality in an autonomous and integrated way. The development of a system able to validate automatically and in real-time the data collected improves the quality of information used and the efficiency of the data mining models induced. Using the data provided by the quality system it is possible: to present better information to decision process and to create new knowledge. It is also possible to calculate ICU Scores: SAPS, SOFA, MEWS and TISS; calculate Critical Events (BP, Temperature, SPO2 and Urine Output) and predict organ failure or patient outcome with high level of confidence. All of this changes allow to have a bigger control of the data and a high number of data available automatically and in real-time. With this system the professionals have a complete control of the data (e.g. they can change the values whenever the values aren’t correct).

In the future, other techniques to validate the information and RFID will be explored in order to assure the correct identification of patients. Some improvements will be done in the quality system and in the data mining models.

**Acknowledgment**

This work is supported by FEDER through Operational Program for Competitiveness Factors – COMPETE and by national funds though FCT – Fundação para a Ciência e Tecnologia in the scope of the project: FCOMP-01-0124-FEDER-022674. The authors would like to thank FCT (Foundation of Science and Technology, Portugal) for the financial support through the contract PTDC/EA1/72819/2006. The work of Filipe Portela was supported by the grant SFRH/BD/70156/2010 from FCT.

**References**

Filipe Portela is a PhD Student in Information Systems in Minho University. He is inserted in the R&D Centre ALGORITMI and is developing his research work in the topic “Pervasive Intelligent Decision Support in the Critical Health Care”.

Manuel Filipe Santos holds a PhD in Computer Science – Artificial Intelligence. Actually he is associate professor in the department of information systems, school of engineering, University of Minho, Portugal. He is the Investigator Responsible in research projects (grid data mining and intelligent decision support systems for intensive care). He is the head of the Intelligent Data Systems group in the R&D Centre Algoritmi.

Álvaro Silva, holds a PhD and is guest professor in Abel Salazar Biomedical Sciences Institute and works as a Doctor in the Intensive Care Unit of Centro Hospital do Porto (CHP) in Portugal, where he has been doing research in the intensive medicine and data mining area.

Fernando Rua is a graduate doctor in Intensive Medicine. He is the Chief Officer of Intensive Care Unit of CHP in Portugal. He also develops research related with information systems and intensive medicine.

António Abelha is an Auxiliary Professor at University of Minho, Department of Informatics, in Braga, Portugal.

José Machado is an Associate Professor with habilitation in the same department. They got their PhD in Informatics, respectively in 2004 and 2002, and are researchers of Computer Science and Technology Center (in Portuguese CCTC). Their research interests span the domain of Electronic Health Records, Interoperability, Databases and Artificial Intelligence.