

Closed loop Knowledge Discovery for Decision Support in Intensive Care Medicine

PEDRO GAGO¹ AND MANUEL FILIPE SANTOS²

¹Escola Superior de Tecnologia e Gestão

Instituto Politécnico de Leiria

Leiria

PORTUGAL

pgago@estg.ipleiria.pt

²Departamento de Sistemas de Informação

Universidade do Minho

Guimarães

PORTUGAL

mfs@dsi.uminho.pt

Abstract: - Clinical Decision Support Systems (CDSS) are becoming commonplace. They are used to alert doctors about drug interactions, to suggest possible diagnostics and in several other clinical situations. One of the approaches to building CDSS is by using techniques from the Knowledge Discovery from Databases (KDD) area. However using KDD for the construction of the knowledge base used in such systems, while reducing the maintenance work still demands repeated human intervention. In this work we present a KDD based architecture for CDSS for intensive care medicine. By resorting to automated data acquisition our architecture allows for the evaluation of the predictions made and subsequent action aiming at improving the predictive performance thus closing the KDD loop.

Key-Words: - Clinical Decision Support, Intelligent Decision Support Systems, Knowledge Discovery, Intensive care, Ensembles, Multi-Agent Systems

1 Introduction

Clinical Decision Support Systems (CDSS) are computer systems designed to impact decision making about individual patients at the point in time that these decisions are made [1]. Such systems may help physicians perform a large number of tasks related to patient care. Some of the most important factors associated with their success [2] are:

- providing alerts/reminders automatically as part of the workflow;
- providing the suggestions at a time and location where these decisions are being made;
- providing actionable recommendations; and
- computerizing the entire process.

Thus, integration into both the culture and process of care is going to be necessary for these systems to be optimally used. In fact, there are several issues that must be addressed in order to promote the use of CDSS. Some are issues of how the data are entered. Others are related to the development and maintenance of the knowledge base. Finally, there are also physician motivation issues.

An increasing number of Intensive Care Units (ICUs) are using computer applications in their daily routine

[3]. Those applications usually encompass data integration, where data from several sources is combined in a central repository. Moreover, the availability of medical data in digital support allows for the progressive replacement of the traditional paper based records with the associated gains in quality of care ([3-6]). In addition, the availability of data may ease the development of CDSS through the use of Knowledge Discovery from Databases (KDD) techniques. However, there are issues with such DSSs as there are several problems regarding data availability and quality and also there is the need for frequent system updates in order to maintain a good performance.

In this paper we present the work we have been developing in building one such CDSS. The architecture we propose allows the development of (semi) autonomous CDSSs as data acquisition, pre-processing and model updates can be performed without human intervention.

Our system (INTCare) is being tested in a real environment, the Intensive Care Unit (ICU) of Hospital Geral de Santo António (HGSA), Oporto, north of Portugal.

The following sections will introduce the problem definition and the state of the art, followed by the

INTCare system formal description in terms of the Agent-Based paradigm. Then, the obtained results will be presented by means of their usability and efficiency. Next, a discussion will be set over the critical aspects of the INTCare system. Finally, the paper will be concluded, showing also perspectives of future work.

2 Background

2.1 Clinical Decision Support Systems

One can now find CDSS being used with very diverse objectives, including [7]:

- alerts and reminders;
- diagnostic assistance;
- therapy critiquing and planning;
- prescribing decision support;
- information retrieval;
- image recognition and interpretation.

Nevertheless, CDSS are not as widespread as would be expected given the potential benefits [8]. Some insights into the reasons behind some CDSS failures are reported in [9] and include several ideas with one of the most striking being that “the critical impediment to the development of decision programs useful in medicine lays in the impossibility of developing an adequate database and an effective set of decision rules”. Also, CDSS that require significant data entry are not likely to be used on a regular basis [10]. The architecture we propose attempts to circumvent these difficulties includes an agent for automatic data gathering from the available sources such as bedside monitors or the patient’s Electronic Health Record (EHR). Moreover, by using Knowledge Discovery in Databases (KDD) techniques INTCare is able to use the available data in order to automatically build and/or maintain the knowledge base.

2.2 Knowledge Discovery from Databases and Data Mining

The interest in Knowledge Discovery from Databases (KDD) and Data Mining (DM) arose due to the rapid emergence of electronic data management methods. In 1997, the Gartner Group suggested that DM is one of the top five key technologies that will have a major impact in the industry within the next years [11]. In effect, these techniques are now widespread and applications can be found in diverse areas, such as: marketing [12], banking [13], manufacturing and production [14], brokerage and securities trading [15], and health care [16]. References for other

applications can be found in various publications, including [17-19].

Within the Medicine arena, huge databases, with large, complex and multi-source information (e.g. text, images or numerical data), are commonplace. However, human experts are limited and may overlook important details. Furthermore, the classical data analysis (e.g. logistic regression) breaks down when such vast amounts of data are present. Hence, an alternative is to use automated discovery tools to analyze the raw data and extract high level information for the decision-maker [20].

More recently, there has been an emergent DM research area that involves the use of ensembles for supervised learning, where a set of classification/regression models are combined in some way to produce an answer [21].

3 CDSS Architecture

In this section we present our architecture for decision support in the intensive care unit.

Most CDSS have three main components: the knowledge base, the inference or reasoning engine and a mechanism to communicate with the user [22]. However, such architecture usually demands manual updates to the knowledge base and may require extensive data entry by the users of the system. Moreover, the need for manual updates to the knowledge base is one of the factors for CDSS failure [8]. Also, doctors tend not to use an application that demands them to repeat data entry [23].

In this architecture we opted to include a component that is responsible for data acquisition and pre-processing. This component will connect to whatever data sources are available and populate a data repository that can be used for automatic updates to the knowledge base. It is presented as an autonomous component as it is fundamental to autonomous operation. The architecture is presented in Figure 1.

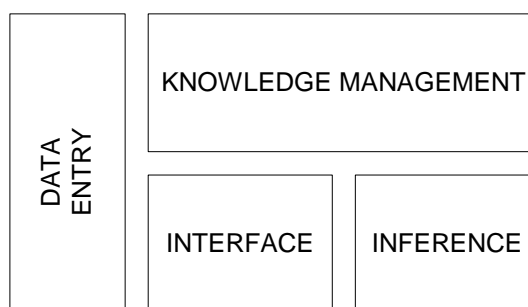


Figure 1 – The CDSS architecture

Data entry is of utmost importance for the correct operation of INTCare.

The Knowledge Management component maintains the prediction models and also includes allowing the evaluation of the predictions made. Moreover, a performance evaluation sub-component monitors the accuracy of the predictions made and when necessary will trigger model updates. We advocate the use of ensembles of models as they have been shown to perform better than single models [21] and allow for smooth updates. In fact, replacing a model in an ensemble is much less disruptive than having a single predictive model and replacing it for another.

The inference component responds to requests from the user (via the interface component). It retrieves the necessary data from the available sources and uses the adequate predictive models from the knowledge management component in order to present the user with an answer. The interface component allows the physicians to ask for predictions for specific patients.

A system based in this architecture is capable of functioning correctly when there is no information regarding the quality of the predictions made. However, if such information is made available, there are mechanisms that may force an automatic update of the predictive models being used thus allowing for an improved performance.

4 INTCare System

The proposed architecture was implemented in a real setting in order to evaluate its usefulness. The chosen location was an ICU of the Hospital de Santo António in Oporto, Portugal. It is a good location as there is a wealth of available data that can be collected either from the bed-side monitors or from the EHR. Our initial goal was that of predicting hospital mortality for ICU patients, given some data from the first 24 hours after ICU admission. As discharge data is entered into the EHR it is possible to (automatically) evaluate the predictions made.

The INTCare (Figure 2) is an agent based system, composed by several semi-autonomous agents in charge for the functionalities inherent to the system. Formally, the INTCare system is defined as a tuple $\Xi \equiv \langle C_{INTCare}, \Delta_{INTCare}, a_{pp}, a_{cde}, a_{dm}, a_{pf}, a_{mi}, a_{dr}, a_{pd}, a_{sc}, a_{int}, a_{ic} \rangle$, where:

$C_{INTCare}$ is the context and corresponds to a logical theory, represented as a triple $\langle Lg, Ax, Dñ \rangle$, where Lg stands for an extension to the language of programming logic, Ax is a set of axioms over Lg, and D is a set of inference rules;

$\Delta_{INTCare}$ is the set of bridge rules defining the interaction among the systems' components (the agents);

a_{pp}, \dots, a_{ic} are the system's agents.

4.1. Description of the Agents

The INTCare agents follow the subordinated architecture [24], they are social computational entities, semi-autonomous, reactive, with internal-state, and pre-defined goals incited by the system creators, whose activity contributes to the goal of the overall system. In a technical perspective the INTCare agents are high granularity objects aggregating a great number of capacities.

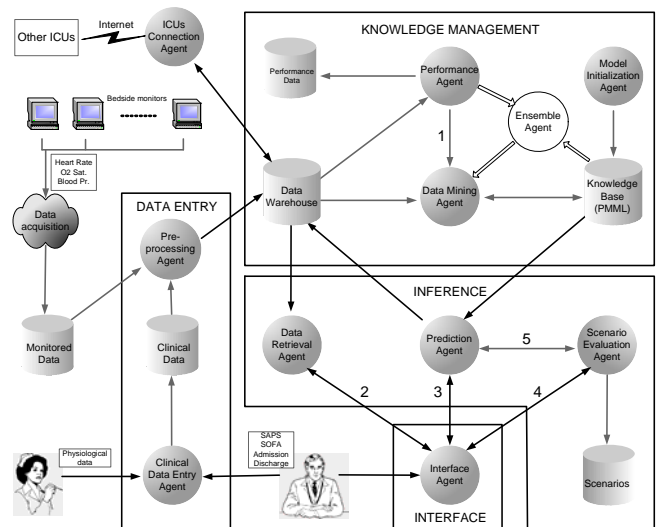


Figure 2. The INTCare system

The social model is static, pre-defined, and incorporated into the system by the developers. The intelligent behavior, the accuracy, the robustness, the flexibility and efficiency of this kind of system emerges from the agents and their interaction. The overall behavior is the combined result of its purpose and its interaction with the environment [25]. The next lines describe the agents' functionalities:

Clinical Data Entry (a_{cde}) agent is responsible for the capture of clinical data from the medical and nursing staff;

Pre-Processing (a_{pp}) agent is responsible for the correct linking of all the values in order to create a valid (even if limited in scope) medical record for the patient. It proceeds with the copy of the values entered by the medical and nursing staff (or recorded via the bedside monitors), examines them and derives new fields, if necessary;

Data Mining (a_{dm}) is responsible for the retrieval of the relevant data in order to make possible the application of AI algorithms to train new models (train), whenever requested by the Performance (a_{pr}) agent, storing them into the Knowledge Base. After

training, the models are stored in the Knowledge Base;

Ensemble agent intended to enhance predictive performance by combining several models (e.g. using a majority voting scheme) in order to produce an efficient answer [21]. ;

Performance (a_{pf}) agent continually consults, in a proactive way, the the Data Warehouse for updates that allow statistics collection (e.g. discharge data that may or may not confirm a prediction made), as a base to calculate a set of assessment parameters maintained in the Performance Database. The evaluation metrics include classification accuracy, sensitivity and specificity values [26]. These statistics are updated every time new relevant information is collected. Whenever the collected statistics show that the performance has fallen below a predefined threshold (a configuration parameter) a new model is requested in order to allow the replacement of the poor performing one;

Model Initialization (a_{mi}) agent populates the Knowledge Base with the models obtained from off-line training. This agent is currently used only when first starting INTCare;

Data Retrieval (a_{dr}) agent is an information agent, whose only objective is that of retrieving, from the Data Warehouse, the information requested by the interface agent and returning it;

Prediction (a_{pd}) agent answers user questions by applying the adequate models contained in the Knowledge Base to the data stored in the Data Warehouse. Next, results are sent back to the calling agent and if the calling agent is the interface agent, the prediction made is recorded into the Data Warehouse;

Scenario Evaluation (a_{sc}) agent makes it possible to the doctor to create and evaluate what-if scenarios. After receiving the data from the interface agent, the scenario evaluation agent requests a forecast from the prediction agent, the scenario is then stored in the Scenarios Database and the result is sent back to the interface agent;

Interface (a_{int}) agent allows (web-based) interaction with the system by providing an easy way for doctors to request prognostics and evaluate scenarios. Whenever new data is needed, this agent messages the data retrieval agent;

Connection Agent (a_{ic}) is in charge for the knowledge interchange among the various instances of the INTCare system running in the ICUs.

5 Deployment and Experimentation

In this section we describe both the deployment process and some of the experiments we have conducted.

INTCare is an application programmed in Java that utilizes the WEKA library [27]. Its deployment was controlled, having started with data collection from a single bed and is now and has evolved to the current situation where data is collected from all the beds in the ICU unit. Data collection depends on proprietary software from the suppliers of the bed-side monitors. Also, the system connects to the hospital network to access the relevant clinical records, thus reducing the need for data entry by the physicians.

After some initial work with offline data [28, 29] we tested the INTCare system using data collected via the bed-side monitors [30]. It was a somewhat limited experience as it was performed on a small dataset as at the time data acquisition was available for a single ICU bed. Nevertheless, both the results and the physicians' comments were encouraging.

Other experiments regarded how show the model updates be performed so that the system is able to function without the need for human intervention.

Two different experiments [31, 32] showed that good performance can be achieved through the use of ensembles that are updated following a procedure based on the Dynamic Weighted Majority [33].

6 Contributions and limitations

In order to be successful CDSS must merge into the existing clinical environment and be as inconspicuous as possible to physicians and nursing staff. Manual data entry must be kept to an absolute minimum avoiding the need for duplicate data entry. Moreover, the knowledge base must be kept up to date. By connecting to the existing electronic records INTCare greatly reduces the need for manual data entry from the medical staff.

Moreover, INTCare works in a non-intrusive way as it provides predictions on request and does not burden physicians with unwanted alarms and recommendations.

Also, the automatic updates to the knowledge base lower the maintenance costs and allow the system to maintain a good predictive performance.

With these characteristics INTCare is a system that blends into the existing environment and operates without placing any extra requirements on the medical staff.

As a byproduct, the deployment of INTCare allows for data collection that may be used in future medical studies.

The architecture allows for the growth of the number of predictive objectives through simple changes in the agents.

A computer based system has other advantages such as not being subject to between-physician variability in prognostic skill or beliefs regarding the construct of “futility”. Furthermore model-based predictions may be more equitable because they do not incorporate value-based judgments regarding the “worth” of one life over another ([34]).

7 Conclusions and further work

Clinical Decision Support Systems are potentially very useful having being used in conjunction with CP Order Entry. Current uses include alerts for drug interaction, etc.

Reports on the effective use of such systems stress the need for integration into the existing environment and the difficulties found when there is a need for knowledge base updates.

The INTCare architecture allows for the development of CDSS that blend into the existing systems as data acquisition is done by automatically connecting to existing systems and knowledge base updates are automatically performed in response to self evaluation. The resulting system can be deployed as part of the informatics infrastructure with prediction results being shown to doctors when they require them. Regardless if this, the initial stages of deployment required great effort so that automatic data collection could work as intended. Indeed, nurses were asked to adhere more strictly to guidelines regarding the use of bed-side monitors and authorizations had to be obtained so that access to data over the hospital’s networks could be possible.

The major contributions of this approach are:

- this architecture is being tested in a real environment that demands INTCare to be customized in order to answer physician needs;
- it allows for a seamless integration into the work processes in the ICU thus increasing the chances that it will be used in daily work;
- the use of ensembles of diverse predictive models allows for a better predictive performance;
- mechanisms for automatic update of the models in the knowledge base are included;
- the overall KDD process is implemented in an automatic manner.

At the present time there are still some INTCare agents to be fully implemented, as is the case with the scenario evaluation agent. Future work includes the implementation of these agents and the evaluation of other ensemble methods.

Acknowledgements

The INTCare project is financially supported by FTC (PTDC/EIA/72819/2006).

References:

- [1] Berner, E. S. and La Lande, T. J., Berner, E. S. (ed.) *Clinical Decision Support Systems - Theory and Practice, Overview of Clinical Decision Support Systems*, Springer, 2007, pp. 3-22.
- [2] Kawamoto, K., Houlihan, C.A., Bala, E.A., Lobach, D.F., *Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success*, BJM, 2005.
- [3] Manjoney, R., *Clinical information systems market - An insider’s view*, Journal of Critical Care, Vol. 19, No. 4, 2004, pp. 215-220.
- [4] Zanier, E., Ortolano, F., Ghisoni, L., Colombo, A., Losappio, S. and Stocchetti, N., *Intracranial pressure monitoring in intensive care: clinical advantages of a computerized system over manual recording*, Critical Care, 2007, Vol. 11, R7.
- [5] Bosman, R., Rood, E., Oudemans-van Straaten, H., Van der Spoel, J., Wester, J. and Zandstra, D., *Intensive care information system reduces documentation time of the nurses after cardiothoracic surgery*, Intensive Care Medicine, Vol 29, 2003, pp. 83-90.
- [6] Fraenkel, D. J., Cowie, M. and Daley, P. *Quality benefits of an intensive care clinical information system*, Critical Care Medicine, Vol. 31, No. 1, 2003, pp. 120-125.
- [7] Coiera, E. *Guide to Health Informatics – 2nd Ed.*, Hodder Education, 2003.
- [8] Carter, J. Berner, E. S. (ed.) *Clinical Decision Support Systems - Theory and Practice Design and Implementation Issues*, Springer, 2007, pp. 64-98.
- [9] Engle R.L. Jr., *Attempts to use computers as diagnostic aids in medical decision making: a thirty-year experience*. Perspect Biol Med, Vol 35, No. 2, 1992, pp.207-219
- [10] Clayton, P.D. and Hripcsak, G. *Decision Support in Healthcare*, Int. J. Biomed Comput, Vol. 39, 1995, pp. 59-66.
- [11] Lee S. and Siau K., *A review of data mining techniques*, Industrial Management & Data Systems, Vol. 101, No. 1, 2001, pp. 41-46.
- [12] Kim Y. and Street W.N., *An intelligent system for customer targeting: a data mining approach*, Decision Support Systems, Vol. 37, No. 2, 2004, pp. 215-228.
- [13] Hsieh N.C., *An integrated data mining and behavioral scoring model for analyzing bank*

- customers, *Expert Systems with Applications*, Vol. 27, No. 4, 2004, pp. 623-633.
- [14]Musaev A., *Analytic information technologies in oil refinery*, *Expert Systems with Applications*, Vol. 26, No. 1, 2004, pp. 81-85.
- [15]Lam M., *Neural network techniques for financial performance prediction: integrating fundamental and technical analysis*, *Decision Support Systems*, Vol. 37, No. 4, 2004, pp. 567-581.
- [16]Li L., Tang H. , et al., *Data mining techniques for cancer detection using serum proteomic profiling*”, *Artificial Intelligence in Medicine*, Vol. 32, No. 2, 2004, pp. 71-83.
- [17]Turban E., Aronson J., Liang T., *Decision Support Systems and Intelligent Systems*, Prentice Hall, New Jersey, 2004.
- [18]Santos M.F. and Azevedo C., *Data Mining – Descoberta de Conhecimento em Bases de Dados*, FCA Editora, Lisbon (in Portuguese), 2005.
- [19]Liao S-h., *Knowledge management technologies and applications-literature review from 1995 to 2002*, *Expert Systems with Applications*, Vol. 25, No. 2, 2003, pp. 155-164.
- [20]Hand D., Mannila H., Smyth P., *Principles of Data Mining*, MIT Press, Cambridge, MA, 2001.
- [21]Dietterich T., *Ensemble methods in machine learning*, In: Kittler and Roli (eds), *Multiple Classifier Systems*, 2001, LNCS 1857, Springer, pp 1-15.
- [22]Tan J.K.H., and Sheps, S. *Health decision support systems*, Aspen Publishers, Inc., 1998.
- [23]Clayton PD, Hripcsak G. *Decision support in healthcare*, *Int J Biomed Comput*, Vol. 39, No.1, 1995 pp.59–66.
- [24]Weiss G., *Multiagent Systems A Modern Approach to Distributed Artificial Intelligence*, MIT Press, Cambridge MA, 1999.
- [25]Vahidov R., Kersten G., *Decision station: situating decision support systems*, *Decision Support Systems*, Vol. 38, No. 2, 2004, pp. 283-303.
- [26]Silva A., Cortez P., Santos M.F., Gomes L., Neves J., *Multiple organ failure diagnosis using adverse events and neural networks*, In: Seruca I. et al. (eds), *Enterprise Information Systems VI*, 2005, Springer.
- [27] Ian H. W. and Eibe F., *Data Mining: Practical machine learning tools and techniques, 2nd Edition*, Morgan Kaufmann, San Francisco, 2005.
- [28]Santos M.F., et al., *Augmented Data Mining over Clinical Databases: Using Learning Classifier Systems*, In: *Proceedings of the Fourth International Conference on Enterprise Information Systems (ICEIS 2002)*, INSTICC, pp. 512-516.
- [29]Santos M.F., Pereira J., Silva A., *A cluster framework for data mining models – an application to intensive medicine*, In: Chen C-S. et al. (eds) *Proceedings of the seventh International Conference on Enterprise Information Systems (ICEIS 2005)*, INSTICC, pp. 163-168.
- [30]Gago, P., Santos, M. F. , Silva, A., Cortez, P., Neves, J. and Gomes, L., *INTCare: A Knowledge Discovery based Intelligent Decision Support System for Intensive Care Medicine*, *Journal of Decision Systems*, Special issue on Design, Building and Evaluation of Intelligent DMSS, Lavoisier, Vol. 14 No. 3, 2005, pp. 241-259.
- [31]Gago, P., Silva, A. and Santos, M. F., *Adaptive Decision Support for Intensive Care*, *EPIA Workshops*, Springer, 2007, pp. 415-425.
- [32]Gago, P. and Santos, M. F., *Towards an Intelligent Decision Support System for Intensive Care Units*, *ECAI Workshop on Supervised and Unsupervised Ensemble Methods and their Applications*, Greece, 2008, pp. 21-25.
- [33]Kolter, J. Z. and Maloof, M. A., *Dynamic Weighted Majority: A New Ensemble Method for Tracking Concept Drift*, *Proceedings of the Third IEEE International Conference on Data Mining*, IEEE Press, 2003, pp. 123-130.
- [34]Barnato A. and Angus D., *Value and role of intensive care unit outcome prediction models in end-of-life decision making*. *Crit Care Clin*, Vol. 20, No. 3 , 2004, pp. 345-362.