SISAIH: a Case-Based Reasoning Tool for Hospital Admission Authorization Management

Fabiana Lorenzi[1], Mara Abel[2], and Francesco Ricci[3]

[1] Universidade Luterana do Brasil
Rua Miguel Tostes, 101
Canoas, Brasil
lorenzi@ulbra.tche.br

[2] Universidade Federal do Rio Grande do Sul
Porto Alegre, Brasil
marabel@inf.ufrgs.br

[3] eCommerce and Tourism Research Laboratory
ITC-irst
via Sommarive 18
38050 Povo, Italy
ricci@itc.it

Abstract. In health care public programs, huge amounts of data are stored in computer systems: patient record, medical knowledge, or financial data. This paper describes the development process of a support tool for health care named SISAIH. This system reuses knowledge contained in hospital databases, and is aimed at detecting irregularities in the Brazilian health public system. The system was developed using an innovative approach to the knowledge acquisition process. Moreover a special case model is proposed, which aggregates two or more real patient admission cases to build a unique system case. This was motivated by the special requirements of the problem to solve, i.e., to support the decision making process in the control and evaluation of the Hospital admission authorizations.

1 Introduction

Quality care and cost savings in a health system are very important but could be attained only when information management achieve good quality standards. In fact information must be accurate, complete, reliable and verifiable, as well as error free and must contain all the critical facts needed to support the decision making process.

In Brazilian health care, where the resources are scarse and they should be used in the most efficient way, managing information systems becomes a delicate issue. A serious problem in the Brazilian health care system is the control and evaluation of the Hospital Admission Authorization procedures (HAA), which deals with the patient bill in and all his medical procedures and medicines.

The HAAs' management is manually handled by the City Health Secretary (CHS) and it consists in evaluating each HAA to decide if it should be paid or not to the hospital. The Brazilian Public Health Service has defined some rules that the hospitals should follow. For instance, a hospital can admit the patient more than once in a month only if different medical procedures are provided.

The main problem is that the CHS receives bills from almost one hundred hospitals and it must check these HAAs in just two days. Each HAA is checked to decide whether to pay or block the corresponding bill. If it is blocked, the HAA is returned to the hospital that must explain in details each charged item. This guarantees that only the right costs will be charged. The CHS experts use their know how to take these authorization decisions. This expert knowledge is therefore implicitly stored in the CHS databases. Conventional systems are not able to reuse this knowledge and this has motivated us to use a more sophisticated AI-based approach.

This paper presents the Hospital Admission Authorization System (SISAIH), a Case-Based Reasoning tool that supports the Hospital Admission Authorization Management. Applying the CBR methodology [5, 12], SISAIH exploits a memory of HAAs evaluation (stored in the case base) to support the decision process to block or not a new HAA.

Even though a long history of successful CBR systems is available, even in this application area [3, 10], creating a usable operational CBR system still poses major problems mostly related to the domain understanding and modelling. This basically relates to case representation and data management (structuring, cleaning, etc.). Patient records (cases) typically contain lot of heterogeneous data, and therefore it is difficult to structure them to make them self contained and clearly organized. Sometime, these difficulties in knowledge acquisition and representation jeopardize the application solution or even make it impossible. As a consequence, in the design and development of SISAIH we have analyzed problems about collecting, cleaning and structuring cases in a CBR system that imports and integrates data already structured in relational databases.

The paper is organized as follow. A brief explanation of the domain is presented in Section 2. Section 3 describes the whole SISAIH development process describing the stages we have followed. Section 4 illustrates the system validation methodology and section 5 provides the conclusion.

2 Checking the Hospital Admission Authorizations

An Hospital Admission Authorization (HAA) is basically the permission to hospitalize a patient. It is issued by the CHS and it generates a bill to be paid by the Brazilian Public Health. At the end of each month, every hospital of the Brazilian Public Health Care Service sends an invoice to the CHS that lists all the required payments, and their corresponding HAAs.

These HAAs must be checked for irregularities (e.g., a male patient interned for a delivery or a patient interned twice in the same month for the same medical procedure).

Figure 1 shows the control and evaluation process for data analysis and payment authorization. A manual procedure is carried out monthly on every report

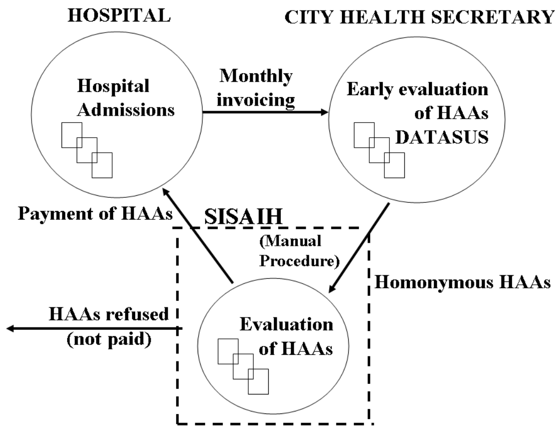

Fig. 1. Context model of the SISAIH system [8]

issued by hospitals belonging to the Brazilian Public Health Service. A CHS expert team double check each bill looking for irregularities.

The file is first evaluated by a system called *Datasus*, which finds out obvious problems such as the lack of essential information, inappropriate medical procedures and typing mistakes. In a second step, the CHS staff analyzes the HAAs. The most important issue is that the evaluation must be completed in less then two days after the arrival of the files. For this reason, the evaluation has focussed in the past on simple irregularities such as multiple admissions (the HAAs issued in the same month for the same patient), that are typically blocked even because of light suspects.

Therefore, we thought that the application of a more powerful and fast automatic approach, improving the evaluation of the HAAs, could speed up the hospital administration procedure. The SISAIH tool was then designed to reuses the knowledge already stored in the admission records to check new HAAs. We hypothesized that the use of appropriate knowledge engineering techniques, and CBR in particular, could allow the development of such a system. That would make possible to avoid unnecessary wastes and create routines which, according to the standard procedures defined by the CHS, will improve the general quality of the process.

3 The development of the SISAIH

As usual in CBR system development, we started with an extensive analysis of the application domain, the definition of the case representation format and the choice of the required indexes for case retrieval. We defined the similarity metric, as a method to evaluate the suitability of the previous solution in solving the current problem, the adaptation methodology, and the learning procedure based on case retention. We shall describe this issues in the next Sections.

3.1 Knowledge Acquisition

The knowledge acquisition phase is one of the most important steps in the development of a CBR system [6]. Knowledge acquisition refers to the task of acquiring knowledge from human experts, books, documents, sensors or computer files, and organizing these sources in a common and rational framework. Although there are many techniques described in the literature [13], the knowledge acquisition task is still poorly mechanized and the majority of the methods largely rely on subjective judgments. The better methodologies aims at speeding up the development process. They try to be enough general to be applicable to many domains and to model the acquired knowledge in such a way that can be easily converted into a computational approach [11].

The methods applied in our research are the open and structured interviews, coupled with the analysis of cases.

1. *Open or unstructured interviews:* the information was collected through direct interviews to the head of the CHS team (the expert). The interviews were tape recorded and then transcribed and analyzed. The interviews generally started with a few questions about the domain to stimulate the expert to freely talk about the topic. The questions were: "What is a HAA?"; "Is the authorization necessary to admit a patient to the hospital?"; "What is a hospital invoice?"; "How the public money spending is assessed in the heath department?". The collected information provided a general overview of the domain.
2. *Structured interviews:* At this stage, after the open interviews have provided basic knowledge of the problem specification, more structured interviews focused on the details of the checking and solution procedures. The answers were again transcribed and analyzed, and the main characteristics of the problem solving process was identified. Reviewing these characteristics, we understood an important issue, i.e., the expert must analyze and compare more than one HAA, belonging to the same patient, to detect irregularities. As we noted above, the Datasus system identifies only the simplest problems that can be detected double checking on single HAA per time. There is no system that can compare two or more HAAs. For instance, the expert showed us a case where the patient was admitted to amputate the left leg and one week after the same patient was admitted again for a surgery operation on the left foot. That is an obvious irregularity, but was not automatically detected. If a patient is admitted in the hospital three times in a short period, then it is necessary to compare in pairs all these hospitalizations. Hence, at least two HAAs belonging to the same patient must be included and modelled in one case. All the relevant attributes identified by the expert for one HAA were therefore "duplicated" as shown in table 1. The collected information was used to refine the case attributes and their values domains.

The last phase of the knowledge engineering preocess was dedicated to observe the CHS team during their job of screening the hospital invoices. They were invited to think aloud and we traced each step of their decision process, making clear how they tackle and accomplish the task. Their decisions were compared with the methods obtained in the previous interviews, revising and validating the acquired knowledge.

3.2 Case Representation

"A case is a piece of knowledge in a particular context representing an experience that teaches an essential lesson to reach the goal of the reasoner" [5]. A case represents knowledge at an operational level, i.e., it makes explicit how a HAA evaluation was carried out or how a piece of knowledge (for instance a set of admission features) was applied to decide whether to accept or reject a specific HAA.

Case representation deals with the information and knowledge to be represented in a case, and its representation formalism [12]. In SISAIH a case basically represents an event of HAA evaluation. The knowledge acquisition phase showed that simultaneously two HAAs are needed to model an evaluation case. The expert indicated the attributes of a HAA that are more relevant for the evaluation, i.e., those that are always checked when comparing two or more HAAs. These attributes are:

- The medical specialty required for the patient hospitalization (e.g., pediatrics or cardiology);
- The medical procedures performed during the hospitalization;
- The time period of each HAA, i.e., the number of days the patient was hospitalized;
- The interval, in days, between the end of a previous hospitalization and the beginning of the next one;
- The motivations for admitting the patient into the hospital, hence the reason for the claimed expenses.

In addition, each case is augmented with the solution for the corresponding HAA together with an explanation that provide a rationale for the decision. To model this knowledge the following three new attributes were added to each case:

- Blocked 1: a boolean field indicating that the first HAA was rejected;
- Blocked 2: a boolean field indicating that the second HAA was rejected;
- Explanation: the reason wht some or both of the referred HAAs were rejected.

3.3 Populating the Case Base

Each case in the case base contains the operational knowledge applied by the expert in solving one problem. SISAIH builds a case by joining two invoice records for patients with multiple hospitalization. This provides the description of the problem, whereas the solution corresponds to the decision of rejecting or not one HAA, together with the justification (explanation) of the solution.

Table 1. Case structure in SISAIH

First HAA	Second HAA
Medical specialty	Medical specialty
Medical procedure	Medical procedure
Period of interment	Period of interment
Interval	Interval
Motive of admission	Motive of admission
Blocked 1	Blocked 1
Blocked 2	Blocked 2
Explanation	Explanation

The case base was generated from the invoice files of the Cardiology Institute Foundation for the period May-August 2003. The case-generation module of the system reads the invoice files month by month, looking for group of records that refer to multiple admissions. For each pair of multiple admission records, the module creates a new case. The interval and admission period were calculated when the case is generated. Moreover, the expert's motivations to block the HAAs (one or both) are also stored in the explanation field. The case base that was generated in this way contains 60 cases, that are physically stored in a relational database.

The whole process of generating the case base has to be done each time new management procedures are applied and we want te system to integrate them in its knowledge repository. In this situation, the system needs to process the monthly invoices that are representative of the new procedure that must be learned. Obviously, in addition to the invoice files, it is necessary to collect the expert evaluations and explanations.

3.4 Indexing

CBR systems derive their effectiveness from the ability of efficiently retrieve relevant cases from the case base [7]. This is assured by the application of the right indexes. These identify which attributes of a case must be exploited in the comparison with the other cases, how the case base must be structured and the best retrieval algorithms.

In SISAIH the indexes were built following two methodologies: explanation-based and mathematical analysis [9]. The explanation-based method was applied during the knowledge acquisition stage, where the expert pointed out what characteristics are used to identify the solution. These characteristics were chosen, in principle, to build the indexing (as we already showed in table 1).

In the indexing process feature weights were also assigned [4]. These weights are used in the nearest neighbor retrieval algorithm (see Equation 1). These indexes were refined through mathematical analyzes, which pointed out how much the attributes influence the solution. The more important attributes were chosen as the primary indexes. The first step was to assess the possible ranges for each characteristic, based on the historical data. The values that pointed to the

Table 2. Characteristics and their possible values (ranges)

Characteristics	Possible values (ranges)
Medical specialty 1	1..9
Medical specialty 2	1..9
Medical procedure 1	47801024..91904013; 31000002..47801018
Medical procedure 2	47801024..91904013; 31000002..47801018
Period of interment 1	1..100; 101..200
Period of interment 2	1..100; 101..200
Interval	1..5; 6..100
Motive of admission 1	11..19; 21..25; 31..39; 41..43; 51..53
Motive of admission 2	11..19; 21..25; 31..39; 41..43; 51..53

same decision were grouped in subsets. These subsets were defined previously by the expert (see table 2. For instance, the medical procedure has two possible ranges: the first range (7801024..91904013) means that it is a clinical procedure and the second one (31000002..47801018) means that it is a surgical procedure.

In the next step we analyzed the CHS database searching for the characteristics frequency in each subset. Weights ranging from 0.1 to 0.3 were assigned to the subsets of values of the characteristic. The higher the frequency of a subset of values in the case base, the larger the weight assigned to each subset (see Table 3. Hence, the frequency is a good indicator of the importance of that specific attribute and set of values in refusing a HAA. Since the medical specialty has just one group is not mentioned in Table 3.

Table 3. Frequency of subset values for attributes and their respective weights

Characteristics	1 subset	2 subset	3 subset	4 subset	5 subset
Medical procedure 1	50 (0.2)	50 (0.2)	-	-	-
Medical procedure 2	50 (0.2)	50 (0.2)	-	-	-
Period of admission 1	70 (0.3)	30 (0.1)	-	-	-
Period of admission 2	70 (0.3)	30 (0.1)	-	-	-
Interval	29 (0.1)	71 (0.3)	-	-	-
Motive of admission 1	70 (0.3)	15 (0.1)	15 (0.1)	0	0
Motive of admission 2	70 (0.3)	15 (0.1)	15 (0.1)	0	0

3.5 Case Retrieval

Case retrieval starts with a problem-description, as input, and terminates when the most similar cases are found. The nearest neighbor algorithm was applied in SISAIH to retrieve the closest cases, where distance is given by a weighted Euclidean metric, with a local weighting scheme [2]. A local distance function d measures the difference between the attribute value in the case and in the new problem. The local distance function d is given by the module of the difference between the two values, $|x_f - c_f|$.

$$Distance(q, c) = K\sqrt{\sum w(c_f) * d(q_f, c_f)^2} \tag{1}$$

Where q is the case in the case base and c is the current case that describes the new problem, $w(c_f)$ is the weight for the subset of values to whom c_f belongs, and K is a normalizing constant that make the distance ranging between 0 and 1. The algorithm assumes that each case $c = c_1, c_2..., c_n$ is defined by a set of n attributes (features) f. Considering a new problem c, the case base CB, and the weights $w = (w(c_1), \ldots, w(c_n))$, which were previously computed in the indexing process, the nearest neighbor algorithm retrieves the case with the smallest distance from the current problem. Two cases are considered "similar" if their distance is smaller than 0.5 (this is derived from expert evaluation). The five most similar cases are shown to the user suggesting that both the HAAs analyzed should be blocked or not and why.

It is important to note that the retrieval step has only the goal to present to the user the most "useful" cases for deciding about the current one, but it is only up to the clerk (Health Secretary) to evaluate the situation and decide whether to reject or not a HAA.

4 SISAIH Validation and Tests

SISAIH was developed to be used both in the hospitals and in the CHS. They have different operational systems and relational databases and for this reason it was used a portable visual programming tool during the development.

As figure 2 shows, two modules were developed in the SISAIH. The *input* module reads the invoice file and produces a new problem representation for each multiple admissions found in the file. They are stored in a second entity called problem-table.

The retrieval stage is done in the *compare* module, where the cases of the case base and the problem-table are matched. The solutions associated with the matched cases are written into the relevant record of the problem-table, producing the output of the system.

A data set was provided later on at the final stage of system development and applied to test the system. Using the invoice files of the Independence Hospital pertaining to the period September-November 2003, we carried out a first test. This test set contained 660 authorizations already evaluated by the CHS that generated a problem-table with 54 new problems (the multiple admissions). After the new problems' analysis, the system suggested that 15 HAAs should be blocked, i.e., all the five retrieved nearest neighbors agreed on a "block" evaluation for these 15 cases. The system was correct in finding exactly the same HAAs reject by the CHS expert team. But we are aware that if the data set of the test would be increased new cases would show up.

The explanation associated with each refused HAA was considered significant by the expert team. The most important feature is that the system provides

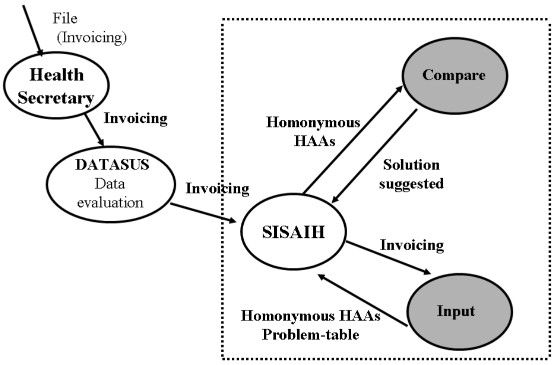

Fig. 2. SISAIH's modules

instantaneous access to information (cases) in a just-in-time manner. And an important CBR's feature, as the case base is increased more efficient are the results.

5 Conclusion

CBR systems have shown to be an effective tool to improve the accuracy of decision making in organizations, and to support the definition of knowledge management policies. This is particularly important when the subject is public health, where the services must be fast and accurate and the public resources must be used in the correct way.

Developing low-cost decision support tools allows extracting, preserving, distributing and reusing effectively existing knowledge, and therefore better support the organizational decision processes.

The main contribution of this paper pertains to the knowledge acquisition methodology and to the case model. In our approach a case models a past HAA evaluation, i.e., the comparison of one HAA with a previous one and not the HAA itself. Hence, a case involves two HAAs, that share some common features, and must be jointly analyzed to judge the correctness of both HAAs.

The CBR systems, applied in supporting knowledge based decisions in health have shown to be profitable in many companies that have been testing the approach [1]. CBR is appropriated for this purpose, since it provides a fast and easy knowledge acquisition method when compared with model-based tools. Moreover, in our approach we kept the main aspect of the original problem description as unique information unit. This was motivated by the goal to be able to support additional needs that may arise and to rely on simple information processing tools, adapted from more conventional information system approaches. The resulting system can be therefore more easily integrated in corporate programs.

SISAIH provides a simple tool for supporting and justifying decisions in a manager's daily job. The system can be used by both CHS and the hospital. In this last case, the hospital can use its own invoice file to run the system and fix the problems before sending the invoice file to the CHS.

The distribution of knowledge allows a more effective stream of information and less rejected work. The system is also easily updated, by providing a new set of solved cases to be processed by the input module, i.e., those that generates a new case base. The decision is automatically learn by the system, although, at this stage of development, the Explanation field must be manually filled with the expert input.

References

1. K.-D. Althoff, R. Bergmann, S. Wess, M. Manago, E. Auriol, O. Larichev, Bolotov, Y. Zhuravlev, and A. Gurov. Case-based reasoning for medical decision support tasks: The inreca approach. *Artificial Intelligence in Medicine*, 12:25–41, 1998.
2. C. G. Atkenson, A. W. Moore, and S. Schaal. Locally weighted learning. *AI Review Journal*, 11:11–73, 1997.
3. I. Bichindaritz, E. Kansu, and K. Sullivan. Case-based reasoning in care-partner: Gathering evidence for evidence-based medical practice. *European Workshop on Case-Based Reasoning*, pages 334–345, 1998.
4. L. K. Branting. Techniques for the retrieval of structured cases. In *Working Notes of the AAAI Spring Symposium on Case-Based Reasoning*, Palo Alto, CA, 1990.
5. J. Kolodner. *Case-Based Reasoning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.
6. D. Leake and D. Wilson. Combining cbr with interactive knowledge acquisition, manipulation and reuse. *Third International Conference on Case-Based Reasoning - ICCBR-99*, 1999.
7. M. Lenz, B. Bartsch-Spörl, H.-D. Burkhard, and S. Wess, editors. *Case-Based Reasoning Technology, From Foundations to Applications*, volume 1400 of *Lecture Notes in Computer Science*. Springer, 1998.
8. F. Lorenzi. Uso da metodologia de raciocnio baseado em casos na investigao de irregularidades nas internaes hospitalares. Master dissertation, PGCC, 1998.
9. D. Michie, D. Spiegelhalter, and C. Taylor. *Machine learning, neural and statistical classification*. Ellis Horwood, 1994.
10. S. Montani, L. Portinale, G. Leonardi, and R. Bellazi. Applying case-based retrieval to hemodialysis treatment. *Workshop on CBR in the Health Sciences*, pages 53–62, 2003.
11. E. Turban. Knowledge acquisition and validation. *Expert Systems and Applied Artificial Intelligence*, pages 117–166, 1992.
12. I. Watson. *Applying Case-Based Reasoning: Techniques for Enterprise Systems*. Morgan Kaufmann, 1997.
13. G. Wright and P. Ayton. Eliciting and modeling expert knowledge. *Decision Support Systems*, 3:1–26, 1987.