Prediction of Length of Hospital Stay in Preterm Infants A Case-Based Reasoning View

Ana Coimbra[1], Henrique Vicente[2, 3], António Abelha[3], M. Filipe Santos[3], José Machado[3], João Neves[4], and José Neves[3*]

[1] Departamento de Informática, Universidade do Minho, Braga, Portugal
cecilia.coimbra@hotmail.com
[2] Departamento de Química, Escola de Ciências e Tecnologia, Universidade de Évora, Évora, Portugal
hvicente@uevora.pt
[3] Centro Algoritmi, Universidade do Minho, Braga, Portugal
{abelha; jmac; jneves}@di.uminho.pt; mfs@dsi.uminho.pt
[4] Drs. Nicolas & Asp, Dubai, United Arab Emirates
joaocpneves@gmail.com
* Corresponding author: phone: +351-934201337; fax: +351-253604471;
e-mail: jneves@di.uminho.pt

Abstract. The length of stay of preterm infants in a neonatology service has become an issue of a growing concern, namely considering, on the one hand, the mothers and infants health conditions and, on the other hand, the scarce healthcare facilities own resources. Thus, a pro-active strategy for problem solving has to be put in place, either to improve the quality-of-service provided or to reduce the inherent financial costs. Therefore, this work will focus on the development of a diagnosis decision support system in terms of a formal agenda built on a Logic Programming approach to knowledge representation and reasoning, complemented with a case-based problem solving methodology to computing, that caters for the handling of incomplete, unknown, or even contradictory information. The proposed model has been quite accurate in predicting the length of stay (overall accuracy of 84.9%) and by reducing the computational time with values around 21.3%.

Keywords: Preterm Infants · Length of Stay · Neonatology · Knowledge Representation and Reasoning · Logic Programming · Case-Based Reasoning.

1 Introduction

In the current century hospital deliveries increased significantly and, consequently, issues related with the Length of Stay (LoS) of subsequent postpartum hospitalization became of the utmost importance. As a result, indicators that may assess the effectiveness of organizations and promote the improvement in hospital quality-of-care, long-term resource planning and family counseling were enforced [1-4].

adfa, p. 1, 2011.

There are different opinions about the time considered appropriate to hospitalization, ranging from drive-through deliveries (with only a few hours of LoS) to 14 (fourteen) day lying-in periods [1]. With the early discharge the family ties are promoted, the patient satisfaction is improved, the risk of iatrogenic infection is diminished and the hospitalization care and patients' costs are reduced [5, 6]. However, some aspects should not be ruled out, namely the fact that until two or more days after delivery some adverse events are not noticeable, which may lead to an increased readmission rate. On the other hand these readmission episodes may cause anxiety in parents and induce the premature cessation of breastfeeding [5, 6]. In an ideal scenario the LoS should be reduced to the strictly necessary time, without endangering the mother's and infant's health, allowing identification of early problems and ensuring that the family is able and prepared to care for the infant at home. Thus, the LoS should be customized for each mother/infant dyad, taking into consideration the health of the mother, the health of the infant and the conditions that they will have outside the hospital [7, 8].

One of the main causes of 4 million neonatal deaths that annually occur all over the world stands for premature birth [9]. Premature or preterm labor may be defined as one in which pregnancy ends between the 20th and the 37th week. The LoS are more than six times greater for extremely preterm infants than for late ones and, therefore, the hospitalization costs are significantly higher under the former case [2]. In fact, the prediction of the LoS of preterm infants in a neonatology service can contribute to the improvement of the long-term planning, leading to a cost reduction, and it also can identify modifiable risk factors that can support quality improvement initiatives. Thus, this paper addresses the length of hospital stay in the preterm infants theme and describes an attempt to predict such period, using a Case-Based Reasoning (CBR) approach to problem solving [10, 11]. CBR allows handling new problems by reusing knowledge acquired from past experiences, namely when similar cases have similar terms and solutions [10 11]. Actually, its use may be found in different arenas, namely in online dispute resolution [12, 13], medicine [14, 15] or education [16], among others. An approach that allows one to deal with incomplete, contradictory or even unknown information.

2 Knowledge Representation and Reasoning

The Logic Programming (LP) paradigm has been used for knowledge representation and reasoning in different areas, like Model Theory [17, 18], and Proof Theory [19, 20]. In the present work the proof theoretical approach is followed in terms of an extension to LP. An Extended Logic Program is a finite set of clauses in the form:

$$
\begin{aligned}
&\{ \;\; p \leftarrow p_1, \cdots, p_n, not\ q_1, \cdots, not\ q_m \\
&\quad ?\,(p_1, \cdots, p_n, not\ q_1, \cdots, not\ q_m) \;\; (n, m \geq 0) \\
&\quad exception_{p_1} \quad \ldots \quad exception_{p_j} \;\; (0 \leq j \leq k), \;\; being\ k\ an\ integer\ number \\
&\} :: scoring_{value}
\end{aligned}
$$

where “?” is a domain atom denoting falsity, the pi, qj, and p are classical ground literals, i.e., either positive atoms or atoms preceded by the classical negation sign ¬ [19]. Indeed, under this formalism, every program is associated with a set of abducibles [17, 18], given here in the form of exceptions to the extensions of the predicates that make the program, i.e., clauses of the form:

$$exception_{p_1} \quad \cdots \quad exception_{p_j} \ (0 \leq j \leq k), \ being\ k\ an\ integer\ number$$

that stand for information or knowledge that cannot be ruled out. On the other hand, clauses of the type:

$$?\,(p_1, \cdots, p_n, not\ q_1, \cdots, not\ q_m) \ \ (n, m \geq 0)$$

also named invariants or restrictions to complain with the universe of discourse, set the context under which it may be understood. The term *scoring_{value}* stands for the relative weight of the extension of a specific *predicate* with respect to the extensions of the peers ones that make the overall program.

In order to evaluate the knowledge that can be associated to a logic program, an assessment of the *Quality-of-Information* (*QoI*), given by a truth-value in the interval [0, 1], that stems from the extensions of the predicates that make a program, inclusive in dynamic environments, is set [21, 22]. Thus, $QoI_i = 1$ when the information is *known* (*positive*) or *false* (*negative*) and $QoI_i = 0$ if the information is unknown. Finally for situations where the extension of $predicate_i$ is unknown but can be taken from a set of terms, $QoI_i \in\ ]0, 1[$. Thus, for those situations, the *QoI* is given by:

$$QoI_i = 1/_{Card} \tag{1}$$

where *Card* denotes the cardinality of the *abducibles* set for *i*, if the *abducibles* set is disjoint. If the *abducibles* set is not disjoint, the clause’s set is given by $C_1^{Card} + \cdots + C_{Card}^{Card}$, under which the *QoI* evaluation takes the form:

$$QoI_{i_{1 \leq i \leq Card}} = 1/_{C_1^{Card}}, \cdots, 1/_{C_{\mathrm{Card}}^{Card}} \tag{2}$$

where C_{Card}^{Card} is a card-combination subset, with *Card* elements. The objective is to build a quantification process of *QoI* and *DoC* (Degree of Confidence), being the later a measure of one’s confidence that the argument values or attributes of the terms that make the extension of a given predicate, with relation to their domains, fit into a given interval [23]. The *DoC* is evaluated as depicted in [23] and computed using $DoC = \sqrt{1 - \Delta l^2}$, where Δl stands for the argument interval length, which was set in the interval [0, 1]. Thus, the universe of discourse is engendered according to the information presented in the extensions of such predicates, according to productions of the type:

$$predicate_i - \bigcup_{1 \leq i \leq m} clause_j\left(\left(QoI_{x_1}, DoC_{x_1}\right), \cdots, \left(QoI_{x_m}, DoC_{x_m}\right)\right) :: QoI_i :: DoC_i \tag{3}$$

where $\bigcup$ and m stand, respectively, for *set union* and the *cardinality* of the extension of *predicate$_i$*. QoI_i and DoC_i stand for themselves [23].

As an example, let us consider the logic program given by:

$$
\begin{aligned}
&\{\\
&\quad \neg\, f_1\left((QoI_{x_1}, DoC_{x_1}),\ (QoI_{y_1}, DoC_{y_1}),\ (QoI_{z_1}, DoC_{z_1})\right)\\
&\qquad\qquad\qquad \leftarrow not\left((QoI_{x_1}, DoC_{x_1}),\ (QoI_{y_1}, DoC_{y_1}),\ (QoI_{z_1}, DoC_{z_1})\right)\\
&\quad f_1\underbrace{\left((QoI_{[7,10]}, DoC_{[7,10]}),\ (QoI_{\perp}, DoC_{\perp}),\ (QoI_{1.5}, DoC_{1.5}),\right)}_{attribute`s\ values} :: QoI :: DoC\\
&\qquad\qquad \underbrace{[5, 35] \qquad\qquad [0, 10] \qquad\qquad [0.5, 2.5]}_{attribute`s\ domains}\\
&\quad exception_{f_{1,1}}\left((QoI_{18}, Doc_{18}), (QoI_{[1,2]}, DoC_{[1,2]}), (QoI_{\perp}, DoC_{\perp})\right) :: QoI :: DoC\\
&\quad \ldots\\
&\quad exception_{f_{1,k}}\left((QoI_{\perp}, DoC_{\perp}), (QoI_5, DoC_5), (QoI_{[0.6,\,1]}, DoC_{[0.6,\,1]})\right) :: QoI\\
&\qquad\qquad :: DoC
\end{aligned}
$$

$\}$ :: 1 (*once the universe of discourse is set in terms of the extension of only one predicate*)

where $\perp$ denotes a null value of the type unknown. It is now possible to split the abducible or exception set into the admissible clauses or terms and evaluate their QoI_i. A pictorial view of this process is given in Fig. 1, as a pie chart.

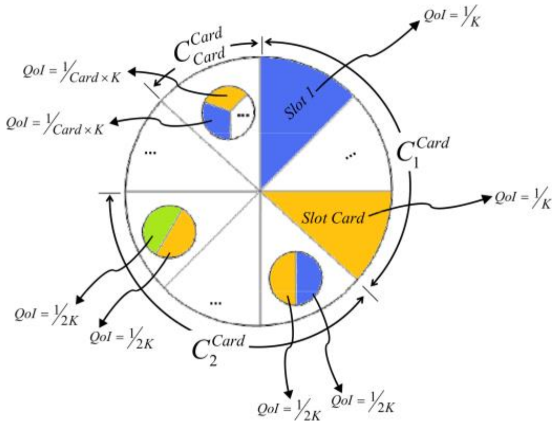

Fig. 1. *QoI's* values for the abducible set of clauses referred to above, where the clauses cardinality set, *K*, is given by the expression $C_1^{Card} + C_2^{Card} + \cdots + C_{Card}^{Card}$.

3 Methods

Aiming to develop a predictive model to estimate the length of hospital stay in preterm infants, a database was set with a total of 284 cases. The data was taken from the health records of patients at a major health care institution in the north of Portugal. This section demonstrates briefly the process of extraction, transformation and loading. Moreover, shows how the information comes together and how it is processed.

3.1 Case Study

Nowadays information systems are often confused with computer systems, however, they are two distinct things, because information systems are as old as the institutions themselves, which has varied is the technology which supports them [24]. The main purpose of Hospital Information Systems is to store the patients' treatment information, and other associated medical needs [25]. Nonetheless, there was a need to organize the information, get only the necessary in order to facilitate its access. Hence, the information was organized as a *star schema* (Fig. 2), which consists of a collection of tables that are logically related to each other. To obtain a star schema it was essential to follow a few steps. In the former one it was necessary to understand the problem under study and gather the parameters that have influence in the final outcome. The following stage was related with the dimensions that would be needed to set these parameters on the facts table. Finally, information from several sources was collected, transformed according the fact and dimension table and loaded into the fact table [26].

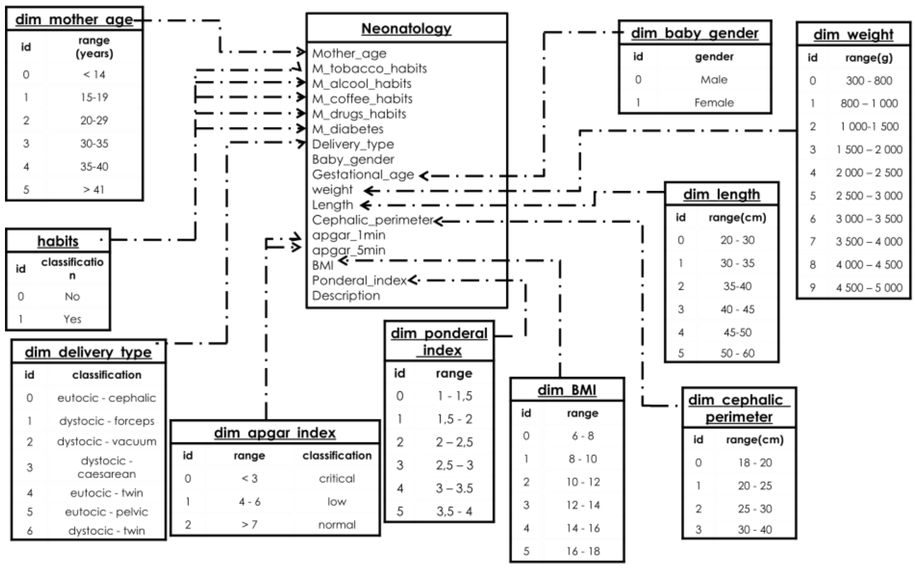

Fig. 2. An overview of the rational model used to gather the information.

The star schema conceived for this study (Fig. 2) takes into account the variables that allow one to estimate the LoS in preterm infants (Neonatology Table), where Dim Tables show how the data regarding those variables were pre-processed (e.g., to babies with apgar index < 3 (critical), ranging between [4, 6] (low), and > 7 (normal) were assigned the values 0 (zero), 1 (one) or 2 (two), respectively).

3.2 Data Processing

Based on the star schema presented in Fig. 2, it is possible to build up a knowledge database given in terms of the tables depicted in Fig. 3, which stand for a situation where one has to manage information aiming to estimate the LoS. Under this scenario some incomplete and/or unknown data is also available. For instance, in case 1, the *Apgar Index 1 min* is unknown, which is depicted by the symbol $\perp$, while the *Cephalic Perimeter* ranges in the interval [2, 3].

The *General Information* and *Mother/Infant Related Factors* tables are filled according to Fig. 2. The values presented in the *Mother Habits* column of *Length of Stay* table are the sum of the columns *Coffee/Tobacco/Alcohol/Drugs Consumption*, ranging between [0, 4]. The *Descriptions* column stands for free text fields that allow for the registration of relevant case's features.

Applying the reduction algorithm presented in [23] to the fields that make the knowledge base for LoS estimation (Fig. 3), excluding of such a process the *Description* ones, and looking to the DoC_s values obtained as described in [23], it is possible to set the arguments of the predicate ***l**ength_**o**f_**s**tay* (*los*) referred to below, that also denotes the objective function with respect to the problem under analyze:

Infant Related Factors

#	Gestational Age	Gender	Weight	Length	Cephalic Perimeter	Apgar 1 min	Apgar 5 min	BMI	Ponderal Index
1	35	0	5	4	[2, 3]	$\perp$	2	2	2
2	32	0	2	3	2	2	2	1	2
…	…	…	…	…	…	…	…	…	…
n	34	1	4	4	3	1	2	2	2

Mother Related Factors

#	Age	Diabetes	Coffee Consumption	Tobacco Consumption	Alcohol Consumption	Drugs Consumption
1	3	0	$\perp$	$\perp$	$\perp$	$\perp$
2	2	0	1	1	0	0
…	…	…	…	…	…	…
n	1	1	1	0	0	0

General Information

#	Date	Delivery Type	Description
1	2015/03/14	1	*Description 1*
2	2015/02/02	3	*Description 2*
…	…	…	…
n	2015/01/03	0	*Description n*

Length of Stay

#	Delivery Type	Gestational Age	Gender	Weight	Length	Cephalic Perimeter	Apgar 1 min	Apgar 5 min	BMI	Ponderal Index	Mother Age	Mother Diabetes	Mother Habits	Description
1	1	35	0	5	4	[2, 3]	$\perp$	2	2	2	3	0	$\perp$	*Description 1*
2	3	32	0	2	3	2	2	2	1	2	2	0	2	*Description 2*
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
n	0	34	1	4	4	3	1	2	2	2	1	1	1	*Description n*

Fig. 3. A fragment of the knowledge base to predict the length of stay in preterm infants.

los: $D_{elivery}T_{ype}$, $G_{estational}A_{ge}$, Gen_{der}, W_{eight}, L_{ength},

$C_{ephalic}P_{erimeter}$, $A_{pgar}1_{min}$, $A_{pgar}5_{min}$, BMI, $P_{onderal}I_{ndex}$, $M_{other}A_{ge}$,

$M_{other}D_{iabetes}$, $M_{other}H_{abits} \rightarrow \{0,1\}$

where 0 (zero) and 1 (one) denote, respectively, the truth values *false* and *true*.

Exemplifying the application of the reduction algorithm presented in [23], to a term (case) that presents feature vector ($D_{elivery}T_{ype}$ = 1, $G_{estational}A_{ge}$ = 33, Gen_{der} = 0, W_{eight} = 3, L_{ength} = 3, $C_{ephalic}P_{erimeter}$ = [2, 3], $A_{pgar}1_{min}$ = 1, $A_{pgar}5_{min}$ = 2, BMI = 2, $P_{onderal}I_{ndex}$ = 2, $M_{other}A_{ge}$ = 5, $M_{other}D_{iabetes}$ = 1, $M_{other}H_{abits}$ = $\perp$), one may have:

Begin (DoCs evaluation),

The predicate's extension that maps the Universe-of-Discourse for the term under observation is set ←

$\{$

$$\neg\, los\left((QoI_{DT}, DoC_{DT}),\ \cdots,(QoI_{CP}, DoC_{CP}),\ \cdots,(QoI_{MH}, DoC_{MH})\right)$$

$$\leftarrow not\ los\left((QoI_{DT}, DoC_{DT}),\ \cdots,(QoI_{CP}, DoC_{CP}),\ \cdots,(QoI_{MH}, DoC_{MH})\right)$$

$$los\ \underbrace{\left((1_1, DoC_1),\ \cdots,(1_{[2,\,3]}, DoC_{[2,\,3]}),\ \cdots,(1_{\perp}, DoC_{\perp})\right)}_{attribute`s\ values} :: 1 :: DoC$$

$$\underbrace{[0,6] \qquad \cdots \qquad [0,3] \qquad \cdots \qquad [0,4]}_{attribute`s\ domains}$$

$\} :: 1$

The attribute's values ranges are rewritten ←

$\{$

$$\neg\, los\left((QoI_{DT}, DoC_{DT}),\ \cdots,(QoI_{CP}, DoC_{CP}),\ \cdots,(QoI_{MH}, DoC_{MH})\right)$$

$$\leftarrow not\ los\left((QoI_{DT}, DoC_{DT}),\ \cdots,(QoI_{CP}, DoC_{CP}),\ \cdots,(QoI_{MH}, DoC_{MH})\right)$$

$$los\ \underbrace{\left((1_{[1,1]}, DoC_{[1,1]}),\ \cdots,(1_{[2,\,3]}, DoC_{[2,\,3]}),\ \cdots,(1_{[0,4]}, DoC_{[0,4]})\right)}_{attribute`s\ values\ ranges} :: 1 :: DoC$$

$$\underbrace{[0,6] \qquad \cdots \qquad [0,3] \qquad \cdots \qquad [0,4]}_{attribute`s\ domains}$$

$\} :: 1$

The attribute's boundaries are set to the interval [0, 1] ←

$\{$

$$\neg\, los\left((QoI_{DT}, DoC_{DT}),\ \cdots, (QoI_{CP}, DoC_{CP}),\ \cdots, (QoI_{MH}, DoC_{MH})\right)$$

$$\leftarrow not\ los\left((QoI_{DT}, DoC_{DT}),\ \cdots, (QoI_{CP}, DoC_{CP}),\ \cdots, (QoI_{MH}, DoC_{MH})\right)$$

$$los\underbrace{\left(\left(1_{[0.2,0.2]}, DoC_{0.2,0.2]}\right), \cdots, \left(1_{[0.7,1]}, DoC_{[0.7,1]}\right), \cdots, \left(1_{[0,1]}, DoC_{[0,1]}\right)\right)}_{attribute`s\ values\ ranges\ once\ normalized} :: 1 :: DoC$$

$$\underbrace{[0, 1] \qquad \cdots \qquad [0, 1] \qquad \cdots \qquad [0, 1]}_{attribute`s\ domains\ once\ normalized}$$

$\} :: 1$

The DoC's values are evaluated ←

$\{$

$$\neg\, los\left((QoI_{DT}, DoC_{DT}),\ \cdots, (QoI_{CP}, DoC_{CP}),\ \cdots, (QoI_{MH}, DoC_{MH})\right)$$

$$\leftarrow not\ los\left((QoI_{DT}, DoC_{DT}),\ \cdots, (QoI_{CP}, DoC_{CP}),\ \cdots, (QoI_{MH}, DoC_{MH})\right)$$

$$los\underbrace{\left((1,\ 1),\ \cdots,\ (1,\ 0.94),\ \cdots,\ (1,\ 0)\right)}_{\substack{attribute`s\ quality-of-information \\ and\ respective\ confidence\ values}} :: 1 :: 0.92$$

$$\underbrace{[0.2, 0.2] \quad \cdots \quad [0.7, 1] \quad \cdots \quad [0, 1]}_{attribute`s\ values\ ranges\ once\ normalized}$$

$$\underbrace{[0, 1] \quad \cdots \quad [0, 1] \quad \cdots \quad [0, 1]}_{attribute`s\ domains\ once\ normalized}$$

$\} :: 1$

End.

This approach allows the representation of the case repository in a graphic form, showing each case in the Cartesian plane in terms of its *QoI* and *DoC*. Thus, the data can be presented in two different forms, i.e., one that is comprehensible to the user and the normalized one that speeds up the retrieval process.

4 Case-Based Reasoning

CBR methodology for problem solving stands for an act of finding and justifying the solution to a given problem based on the consideration of similar past ones, by reprocessing and/or adapting their data or knowledge [11, 27]. In *CBR – the cases* – are stored in a *Case Base*, and those cases that are similar (or close) to a new one are used in the problem solving process. The typical CBR cycle presents the mechanism that should be followed to have a consistent model. In fact, it is an iterative process since the solution must be tested and adapted while the result of applying that solution is inconclusive. In the final stage the case is learned and the knowledge base is updated with the new case [10, 11]. Despite promising results, the current CBR systems are neither complete nor adaptable enough for all domains. In some cases, the user is required to follow the similarity method defined by the system, even if it does not fit into their needs [27]. Moreover, other problems may be highlighted. On the one hand, the existent CBR systems have limitations related to the capability of dealing with unknown, incomplete and contradictory information. On the other hand, an important feature that often is discarded is the ability to compare strings. In some domains strings are important to describe a situation, a problem or even an event [11, 27].

Contrasting with other problem solving methodologies (e.g., those that use *Decision Trees* or *Artificial Neural Networks*), relatively little work is done offline. Undeniably, in almost all the situations, the work is performed at query time. The main difference between this new approach and the typical CBR one relies on the fact that not only all the cases have their arguments set in the interval [0, 1] but it also allows for the handling of incomplete, unknown, or even contradictory data or knowledge [27]. The classic CBR cycle was changed in order to include a normalization phase aiming to enhance the retrieve process (Fig. 4). The *Case Base* will be given in terms of triples that follow the pattern:

$$Case = \{< Raw_{case}, Normalized_{case}, Description_{case} >\}$$

where Raw_{case} and $Normalized_{case}$ stand for themselves, and $Description_{case}$ is made on a set of strings or even in free text, which may be analyzed with string similarity algorithms.

When confronted with a new case, the system is able to retrieve all cases that meet such a structure and optimize such a population, i.e., it considers the attributes *DoC*'s value of each case or of their optimized counterparts when analysing similarities among them. Thus, under the occurrence of a new case, the goal is to find similar cases in the *Case Base*. Having this in mind, the reductive algorithm given in [23] is applied to the new case that presents feature vector ($D_{elivery}T_{ype} = 0$, $G_{estational}A_{ge} = [34, 35]$, $Gen_{der} = 1$, $W_{eight} = [2, 3]$ $L_{ength} = 3$, $C_{ephalic}P_{erimeter} = \perp$, $A_{pgar}1_{min} = \perp$, $A_{pgar}5_{min} = 2$, $BMI = 2$, $P_{onderal}I_{ndex} = 2$, $M_{other}A_{ge} = 3$, $M_{other}D_{iabetes} = 0$, $M_{other}H_{abits} = 2$, *Description* = *Description new*), with the results:

$$\underbrace{los_{new}\big((1,1),(1,0.99),(1,1), \ \cdots, \ (1,0),(1,0),\cdots, \ (1,1),(1,1)\big) \ :: 1 :: 0.84}_{new\ case}$$

Thus, the *new case* can be depicted on the Cartesian plane in terms of its *QoI* and *DoC*, and through clustering techniques, it is feasible to identify the clusters that intermingle with the new one (symbolized as a star in Fig. 5). In present work the technique used to induce clusters was the k-means clustering method [28].

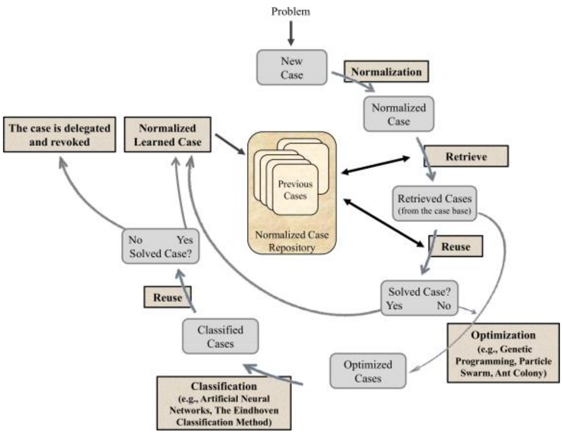

Fig. 4. An extended view of the CBR cycle [27].

The *new case* is compared with every retrieved case from the cluster using a similarity function *sim*, given in terms of the average of the modulus of the arithmetic difference between the arguments of each case of the selected cluster and those of the *new case* (once *Description* stands for free text, its analysis is excluded at this stage). Thus, one may have:

$$\underbrace{\begin{array}{l} los_1\big((1,1),(1,0.98),(1,0),\ \cdots,\ (1,1),(1,1),\cdots,\ (1,1),(1,0)\big) :: 1 :: 0.88 \\ los_2\big((1,1),(1,1),(1,1),\ \cdots,\ (1,0),(1,1),\cdots,\ (1,1),(1,0.95)\big) :: 1 :: 0.92 \\ \vdots \\ los_j\big((1,1),(1,0.92),(1,1),\ \cdots,\ (1,0),(1,0),\cdots,\ (1,1),(1,0)\big) :: 1 :: 0.78 \end{array}}_{normalized\ cases\ from\ retrieved\ cluster}$$

Assuming that every attribute has equal weight, the dissimilarity between los_{new} and the los_1, i.e., $los_{new\to 1}$, may be computed as follows:

$$los_{new\to 1} = \frac{\|1-1\| + \|0.99-0.98\| + \|1-0\| + \cdots + \|1-0\|}{13} = 0.24$$

Thus, the similarity for $los_{new\to 1}$ is 1 – 0.24 = 0.76.

Descriptions will be compared using String Similarity Algorithms, in order to compare the description of the new case with the descriptions of the cases belonging to the retrieved cluster (in this study the strategy used was the Dice Coefficient one [29]), with the results:

$$los_{new\to 1}^{Description} = 0.79$$

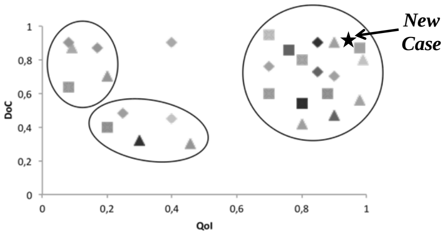

Fig. 5. A case's set separated into clusters.

With these similarity values it is possible to get a global similarity measure, i.e.,

$$los_{new \to 1} = \frac{0.76 + 0.79}{2} = 0.775$$

These procedures should be applied to the remaining cases of the retrieved cluster in order to obtain the most similar ones, which may stand for the possible solutions to the problem.

In order to evaluate the performance of the proposed model the dataset was divided in exclusive subsets through the ten-folds cross validation. In the implementation of the respective dividing procedures, ten executions were performed for each one of them. To ensure statistical significance of the attained results, 30 (thirty) experiments were applied in all tests. The model accuracy was 84.9% (i.e., 241 instances correctly classified in 284). Moreover, the computational time was shortened in 21.3%, when compared with classic CBR implementations.

5 Conclusions

The decision support system to estimate the length of hospital stay in preterm infants, presented in this work, is centred on a formal framework based on Logic Programming for Knowledge Representation and Reasoning, complemented with a *CBR* approach to problem solving that caters for the handling of incomplete, unknown, or even contradictory information. Under this approach the cases' retrieval and optimization phases were heightened and the time spent on those tasks shortened in 21.3%, when compared with existing systems. On the other hand the overall accuracy was around 84.9%. The proposed method allows also for the analysis of free text attributes using *String Similarities Algorithms*, which fulfils a gap that is present in almost all *CBR* software tools. Additionally, under this approach the users may define the weights of the cases' attributes on the fly, letting them to choose the most appropriate strategy to address the problem (i.e., it gives the user the possibility to narrow the search space for similar cases at runtime).

Acknowledgments. This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013.

References

1. Watt, S., Sword, W., Krueger, P.: Longer postpartum hospitalization options – who stays, who leaves, what changes? BMC Pregnancy and Childbirth 5, 1–10 (2005)
2. Hintz, S., Bann, C., Ambalavanan, N., Cotten, C., Das, A., Higgins, R.: Predicting time to hospital discharge for extremely preterm infants. Pediatrics 125, e146–e154 (2010)
3. Adebanji, A., Adeyemi, S., Gyamfi, M.: Empirical analysis of factors associated with neonatal length of stay in Sunyani, Ghana. Journal of Public Health and Epidemiology 7, 59–64 (2015)
4. Goyal, N., Zubizarreta, J., Small, D., Lorch, S.: Length of stay and readmission among late pre term infants: An instrumental variable approach. Hospital Pediatrics 3, 7–15 (2013)
5. Gupta, P., Malhotra, S., Singh, D., Dua, T.: Length of postnatal stay in healthy newborns and re-hospitalization following their early discharge. Indian Journal of Pediatrics 73, 897–900 (2006)
6. Farhat, R., Rajab, M.: Length of postnatal hospital stay in healthy newborns and re-hospitalization following early discharge. North American Journal of Medical Sciences 3, 146– 151 (2011)
7. American Academy of Pediatrics, Committee on Fetus and Newborn: Hospital stay for healthy term newborns. Pediatrics 113, 1434–1436 (2004)
8. American Academy of Pediatrics, Committee on Fetus and Newborn: Hospital stay for healthy term newborns. Pediatrics 125, 405–409 (2010)
9. Niknajad, A., Ghojazadeh, M., Sattarzadeh, N., Hashemi, F., Shahgholi, F.: Factors affecting the neonatal intensive care unit stay duration in very low birth weight premature infants. Journal of Caring Sciences 1, 85–92 (2012)
10. Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Communications 7, 39–59 (1994)
11. Richter, M.M., Weber, R.O.: Case-Based Reasoning: A Textbook. Springer, Berlin (2013)
12. Carneiro D., Novais P., Andrade F., Zeleznikow J., Neves J.: Using Case Based Reasoning to Support Alternative Dispute Resolution. In Carvalho, A.F., Rodríguez-González, S., Paz-Santana, J.F., Corchado-Rodríguez J.M. (eds.) Distributed Computing and Artificial Intelligence, Advances in Intelligent and Soft Computing, vol. 79, pp 123-130, Springer, Berlin (2010)
13. Carneiro, D., Novais, P., Andrade, F., Zeleznikow, J., Neves, J.: Using Case-Based Reasoning and Principled Negotiation to provide decision support for dispute resolution. Knowledge and Information Systems 36, 789–826 (2013)
14. Guessoum, S., Laskri, M.T., Lieber, J.: Respidiag: A Case-Based Reasoning System for the Diagnosis of Chronic Obstructive Pulmonary Disease. Expert Systems with Applications 41, 267–273 (2014)
15. Ping, X.-O., Tseng, Y.-J., Lin, Y.-P., Chiu H.-J., Feipei Lai, F., Liang J.-D., Huang, G.-T., Yang, P.-M.: A multiple measurements case-based reasoning method for predicting recurrent status of liver cancer patients. Computers in Industry 69, 12–21 (2015)
16. Tsinakos, A.: Asynchronous distance education: Teaching using Case Based Reasoning. Turkish Online Journal of Distance Education 4, 1–8, (2003)
17. Kakas, A., Kowalski, R., Toni, F.: The role of abduction in logic programming. In: Gabbay, D., Hogger, C., Robinson, I. (eds.) Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 5, pp. 235–324. Oxford University Press, Oxford (1998)
18. Pereira, L., Anh, H.: Evolution prospection. In: Nakamatsu, K. (ed.) New Advances in Intelligent Decision Technologies – Results of the First KES International Symposium IDT 2009, Studies in Computational Intelligence, vol. 199, pp. 51–64. Springer, Berlin (2009)

19. Neves, J.: A logic interpreter to handle time and negation in logic databases. In: Muller, R., Pottmyer, J. (eds.) Proceedings of the 1984 annual conference of the ACM on the 5^{th} Generation Challenge, pp. 50–54. Association for Computing Machinery, New York (1984)
20. Neves, J., Machado, J., Analide, C., Abelha, A., Brito, L.: The halt condition in genetic programming. In: Neves, J., Santos, M.F., Machado, J. (eds.) Progress in Artificial Intelligence. LNAI, vol. 4874, pp. 160–169. Springer, Berlin (2007)
21. Machado J., Abelha A., Novais P., Neves J., Neves J.: Quality of service in healthcare units. In Bertelle, C., Ayesh, A. (eds.) Proceedings of the ESM 2008, pp. 291–298. Eurosis – ETI Publication, Ghent (2008)
22. Lucas, P.: Quality checking of medical guidelines through logical abduction. In: Coenen, F., Preece, A., Mackintosh A. (eds) Proceedings of AI-2003 (Research and Developments in Intelligent Systems XX), pp. 309–321. Springer, London (2003)
23. Fernandes, F., Vicente, H., Abelha, A., Machado, J., Novais, P., Neves J.: Artificial Neural Networks in Diabetes Control. In Proceedings of the 2015 Science and Information Conference (SAI 2015), pp. 362–370, IEEE Edition (2015)
24. Almeida, A.: The management systems of archival information on Portuguese public hospitals. Master's thesis, University of Lisbon (2011)
25. Cardoso, M.: Auditing an hospital information system – SAM. Master's thesis, Polytechnic Institute of Bragança (2010)
26. O'Neil, P., O'Neil, B., Chen, X.: Star Schema Benchmark. Revision 3, June 5, 2009, http://www.cs.umb.edu/~poneil/StarSchemaB.pdf
27. Neves, J., Vicente, H.: Quantum approach to Case-Based Reasoning. (In preparation)
28. MacQueen J.B.: Some methods for classification and analysis of multivariate observations. In Proceedings of the 5^{th} Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281–297, University of California Press (1967)
29. Dice, L.R.: Measures of the Amount of Ecologic Association between Species. Ecology 26, 297–302 (1945)