A Case-Base Approach to Workforces' Satisfaction Assessment

Ana Fernandes, Henrique Vicente, Margarida Figueiredo, Nuno Maia, Goreti Marreiros, Mariana Neves, José Neves

A Case-base Approach to Workforces' Satisfaction Assessment

Ana Fernandes[1], Henrique Vicente[2,3], Margarida Figueiredo[2,4], Nuno Maia[5], Goreti Marreiros[6], Mariana Neves[7] and José Neves[3*]

[1] Organização Multinacional de Formação, Lisboa, Portugal
anavilafernandes@gmail.com
[2] Departamento de Química, Escola de Ciências e Tecnologia, Universidade de Évora, Évora, Portugal
{hvicente, mtf}@uevora.pt
[3] Centro Algoritmi, Universidade do Minho, Braga, Portugal
jneves@di.uminho.pt
[4] Centro de Investigação em Educação e Psicologia, Universidade de Évora, Évora, Portugal
[5] Departamento de Informática, Universidade do Minho, Braga, Portugal
nuno.maia@mundiservicos.pt
[6] Departamento de Engenharia Informática, GECAD – Grupo de Engenharia do Conhecimento e Apoio à Decisão, Instituto Superior de Engenharia do Porto, Porto, Portugal
goreti@dei.isep.ipp.pt
[7] Deloitte, London, United Kingdom
maneves@deloitte.co.uk

* Corresponding author: phone: +351-934201337; fax: +351-253604471;
e-mail: jneves@di.uminho.pt

Abstract. It is well known that human resources play a valuable role in a sustainable organizational development. Indeed, this work will focus on the development of a decision support system to assess workers' satisfaction based on factors related to human resources management practices. The framework is built on top of a *Logic Programming* approach to *Knowledge Representation* and *Reasoning*, complemented with a *Case Based* approach to computing. The proposed solution is unique in itself, once it caters for the explicit treatment of incomplete, unknown, or even self-contradictory information, either in terms of a qualitative or quantitative setting. Furthermore, clustering methods based on similarity analysis among cases were used to distinguish and aggregate collections of historical data or knowledge in order to reduce the search space, therefore enhancing the cases retrieval and the overall computational process.

Keywords: Human Resources Management · Logic Programming · Case-Based Reasoning · Knowledge Representation and Reasoning · Decision Support Systems

1 Introduction

In a global and competitive world either organization is under a constant state of worry and urgency, and to survive may need to adapt to new economic, organizational and technological sceneries. Undeniably, the organizations should create innovative strategies to promote their own competitive advantages. A company's staffs are a key asset and play an important role in order to undertake its objectives [1, 2]. Definitely, a company's productively is toughly related to its people and their strategies or, in other words, workers' satisfaction stands for a significant instrument in human resources management, leading to [1, 2]:

- Enhanced quality of the offered products and services;
- Positive attitude towards the company;
- Better observance of deadlines;
- Low personnel fluctuation;
- Small absenteeism rates; and
- Creativity and assuming of responsibilities.

The management of workers' satisfaction encompass factors like training (aiming the development of the workers' skills), and the creation of work environments that encourage the productivity, commitment and motivation. In this way, the organizations reveal major concerns in promoting practices which give some support to its employees, seeking the possible balance between professional and private lives [3, 4].

The present study addresses the theme of the *Human Resources Management*, in particular with regard to *Workers' Satisfaction*. However, the assessment of workers' satisfaction is a complex phenomenon that involves a large number of factors, some of which depend on the worker in itself, and even on the organisation [3-5]. Because of this, it is difficult to assess the *Workers' Satisfaction* since it is necessary to consider different conditions with complex relations among them, where the available data may be incomplete/unknown (e.g., absence of answers to some questions presented in the questionnaire), or even contradictory (e.g., questions relating to the same issue with incongruous answers). In order to overcome these difficulties, the present work reports on the founding of a not common approach to *Knowledge Representation and Reasoning* [6], complemented with a *Case Based* attitude to computing [7, 8].

Undeniably, *Case Based* (*CB)* provides the ability to solve new problems by reusing knowledge acquired from past experiences [7, 8], i.e., *CB* is used especially when similar cases have similar terms and solutions, even when they have different backgrounds [9]. Its use may be found in many different arenas, namely in *Online Dispute Resolution* [9] or *Medicine* [10, 11].

This paper involves five sections. In the former one a brief introduction to the problem is made. Then the proposed approach to Knowledge Representation and a *CB* view to computing are introduced. In the fourth and fifth sections it is assumed a case study and presented an answer to the problem. Finally, in the last section the most relevant conclusions are described and possible directions for future work are outlined.

2 Background

Many approaches to Knowledge Representation and Reasoning have been proposed using the *Logic Programming* (*LP*) epitome, namely in the area of *Model Theory* [12, 13] and *Proof Theory* [6], [14]. In the present work the *Proof Theoretical* approach in terms of an extension to the *LP* language is followed. An *Extended Logic Program* is a finite set of clauses, given in the form:

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

where the first clause stand for predicate's closure, "*,*" denotes "*logical and*", while "*?*" is a domain atom denoting falsity, the p_i, q_j, and p are classical ground literals, i.e., either positive atoms or atoms preceded by the classical negation sign $\neg$ [6]. Indeed, $\neg$ stands for a strong declaration that speaks for itself, and *not* denotes *negation-by-failure*, or in other words, a flop in proving a given statement, once it was not declared explicitly. Under this formalism, every program is associated with a set of abducibles [12, 13], given here in the form of exceptions to the extensions of the predicates that make the program, i.e., clauses of the form:

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

that stand for data, 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*, allows one to set the context under which the universe of discourse has to be understood. The term $scoring_{value}$ stands for the relative weight of the extension of a specific predicate with respect to the extensions of peers ones that make the inclusive or global program.

2.1 Knowledge Representation and Reasoning – Quantitative Knowledge

On the one hand, the Quality-of-Information (*QoI*) of a logic program will be understood as a metric that will be given by a truth-value ranging between 0 and 1 [15, 16]. Indeed, $QoI_i = 1$ when the information is *known* (*positive*) or *false* (*negative*) and $QoI_i = 0$ if the information is *unknown*. For situations where the extensions of the predicates that make the program also include *abducible* sets, its terms (or clauses) present a $QoI_i \in]0, 1[$, which will be given by:

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

if the *abducible* set for *predicates* *i* and *j* satisfy the *invariant*:

$$?\left(\left(exception_{p_i};\ exception_{p_j}\right), \neg\ \left(exception_{p_i};\ exception_{p_j}\right)\right)$$

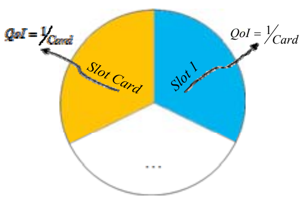
where "*;*" denotes " *logical or*" and "*Card"* stands for set cardinality, being $i \neq j$ and $i, j \geq 1$ (a pictorial view of this process is given in Fig. 1(a), as a pie chart).

On the other hand, the clauses cardinality (*K*) will be given by $C_1^{Card} + \cdots + C_{Card}^{Card}$, if there is no constraint on the possible combinations among the abducible clauses, being the *QoI* acknowledged as:

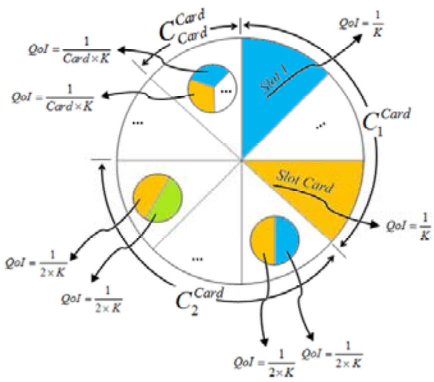
$$QoI_{i_{1\leq i\leq Card}} = {}^{1}/_{C_1^{Card}}, \cdots, {}^{1}/_{C_{Card}^{Card}} \tag{2}$$

where C_{Card}^{Card} is a card-combination subset, with *Card* elements. A pictorial view of this process is given in Fig. 1(b), as a pie chart.

However, a term's *QoI* also depends on their attribute's *QoI*. In order to evaluate this metric let us look to the Fig. 2, where the segment with limits 0 and 1 stands for every attribute domain, i.e., all the attributes range in the interval [0, 1]. [*A*, *B*] denotes the scope where the unknown attributes values for a given predicate may occur (Fig. 2). Therefore, the *QoI* of each attribute's clause is calculated using:



(a)



(b)

Fig. 1. *QoI's* values for the abducible set for $predicate_i$ with (a) and without (b) constraints on the possible combinations among the abducible clauses.

Fig. 2. Setting the *QoIs* of each attribute's clause.

$$QoI_{attribute_i} = 1 - \|A - B\| \quad (3)$$

where ‖A–B‖ stands for the modulus of the arithmetic difference between A and B. Thus, in Fig. 3 is showed the *QoI's* values for the abducible set for *predicate*$_i$.

Under this setting, a new metric has to be considered, which will be denoted as *DoC* (*Degree-of-Confidence*), that stands for one's confidence that the argument values or attributes of the terms that make the extension of a given predicate, having into consideration their domains, are in a given interval [17]. The *DoC* is figured using $DoC = \sqrt{1 - \Delta l^2}$, where Δl stands for the argument interval length, which was set to the interval [0, 1] (Fig. 4).

Thus, the universe of discourse is engendered according to the information presented in the extensions of such predicates, according to productions of the type:

$$predicate_i - \bigcup_{1 \le j \le m} clause_j \left((QoI_{x_1}, DoC_{x_1}), \cdots, (QoI_{x_l}, DoC_{x_l}) \right) :: QoI_j :: DoC_j \quad (4)$$

where $\bigcup$, *m* and *l* stand, respectively, for *set union*, the *cardinality* of the extension of *predicate*$_i$ and the number of attributes of each clause [17]. The subscripts of *QoIs* and *DoCs*, x_1, ..., x_l, stand for the attributes values ranges.

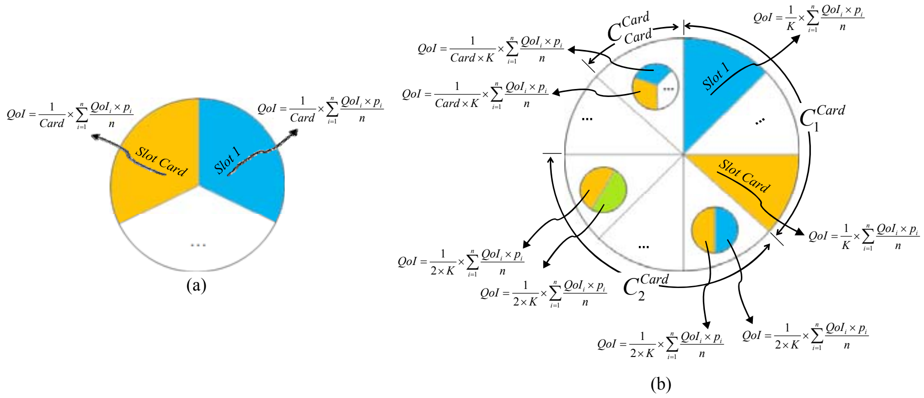


Fig. 3. *QoI's* values for the abducible set for *predicate*$_i$ with (a) and without (b) constraints on the possible combinations among the abducible clauses. $\sum_{i=1}^{n} (QoI_i \times p_i)/n$ denotes the *QoI's* average of the attributes of each clause (or term) that sets the extension of the predicate under analysis. *n* and p_i stand for, respectively, for the attribute's cardinality and the relative weight of attribute p_i with respect to its peers ($\sum_{i=1}^{n} p_i = 1$).

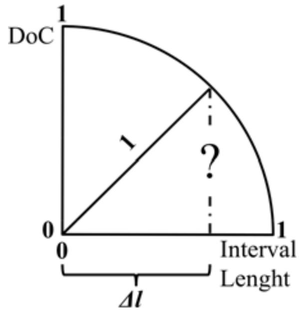


Fig. 4. Evaluation of the Degree of Confidence.

2.2 Knowledge Representation and Reasoning – Qualitative Knowledge

In present study both qualitative and quantitative data/knowledge are present. Aiming at the quantification of the qualitative part and in order to make easy the understanding of the process, it was decided to put it in a graphical form. Taking as an example a set of n issues regarding a particular subject (where the possible alternatives are *none*, *low*, *moderate*, *high* and *very high*), a unitary area circle split into n slices is itemized (Fig. 5). The marks in the axis correspond to each of the possible choices. If the answer to issue 1 is *high* the area correspondent is $\pi \times \left(\sqrt{\frac{0.75}{\pi}}\right)^2 / n$, i.e., $0.75/n$ (Fig. 5(a)). Assuming that in the issue 2 are chosen the alternatives *high* and *very high*, the correspondent area ranges between $\left[\pi \times \left(\sqrt{\frac{0.75}{\pi}}\right)^2 / n,\ \pi \times \left(\sqrt{\frac{1}{\pi}}\right)^2 / n\right]$, i.e., $[0.75/n,\ 1/n]$ (Fig. 5(b)). Finally, in issue n if no alternative is ticked, all the hypotheses should be considered and the area varies in the interval $\left[0,\ \ \pi \times \left(\sqrt{\frac{1}{\pi}}\right)^2 / n\right]$, i.e., $[0\ ,1/n]$ (Fig. 5(c)). The total area is the sum of the partial ones (Fig. 5(d)).

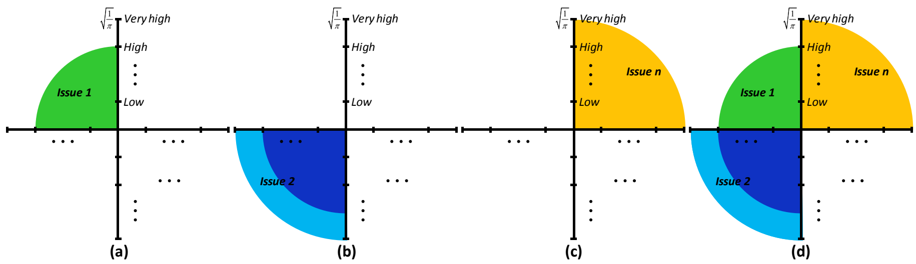


Fig. 5. A view of the qualitative data/information/knowledge processing.

3 A Case Based Methodology for Problem Solving

The *CB* methodology for problem solving stands for an act of finding and justifying a solution to a given problem based on the consideration of similar past ones, by reprocessing and/or adapting their data/knowledge [7, 8]. In *CB* – 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 *CB* cycle presents the mechanism that should be followed, where the former stage entails an initial description of the problem. The new case is defined and it is used to retrieve one or more cases from the *Case Base*.

Despite promising results, the current *CB* systems are neither complete nor adaptable enough for all domains. In some cases, the user cannot choose the similarity(ies) method(s) and is required to follow the system defined one(s), even if they do not meet their needs. Moreover, in real problems, the access to all necessary information is not always possible, since existent *CB* systems have limitations related to the capability of dealing, explicitly, with unknown, incomplete, and even self-contradictory information. To make a change, a different *CB* cycle was induced (Fig. 6). It takes into consideration the case's *QoI* and *DoC* metrics. It also contemplates a cases optimization process present in the *Case Base*, whenever they do not comply with the terms under which a given problem as to be addressed (e.g., the expected *DoC* on a prediction was not attained). This process that uses either *Artificial Neural Networks* [18, 19], *Particle Swarm Optimization* [20] or *Genetic Algorithms* [21], just to name a few, generates a set of new cases which must be in conformity with the invariant:

$$\bigcap_{i=1}^{n} (\boldsymbol{B_i}, \boldsymbol{E_i}) \neq \emptyset \tag{5}$$

i.e., it denotes that the intersection of the attribute's values ranges for cases'set that make the *Case Base* or their optimized counterparts ($\boldsymbol{B_i}$) (being n its cardinality), and the ones that were object of a process of optimization ($\boldsymbol{E_i}$), cannot be empty (Fig. 6).

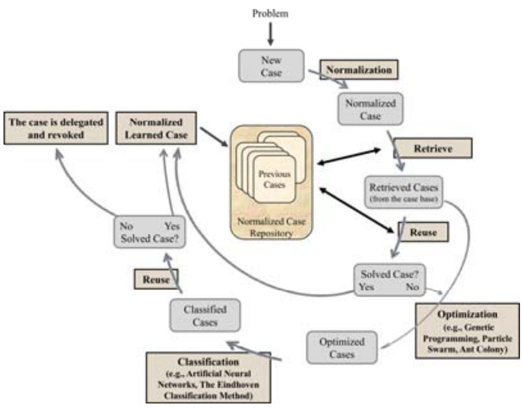


Fig. 6. The updated view of the *CB* cycle.

4 Methods

Aiming to develop a predictive model to assess workers' satisfaction a questionnaire was designed specifically for this study and applied to a cohort of 236 employees of training companies. This section describes briefly the data collection tool and how the information is processed.

4.1 Questionnaire

The questions included in the questionnaire aimed to evaluate the degree of worker's satisfaction. The respondents participated in the study voluntarily and the questionnaires were anonymous to ensure the confidentiality of information provided. The questions included in the questionnaire were organized into five sections, where the former one includes the general questions related with workers' age, gender, length of service and functional area. The second one comprises questions related with the workers' opinions about the received training (Table *Training Related Factors* in Fig. 7), while the third section is about occupational medicine service (Table *Occupational Medicine Related Factors* in Fig. 7). Finally, the fourth and fifth sections comprise issues related with the workers' opinions about the resources (Table *Resources Related Factors* in Fig. 7) and organizational climate (Table *Organizational Climate Related Factors* in Fig. 7), respectively.

4.2 Workforces' Satisfaction Knowledge Base

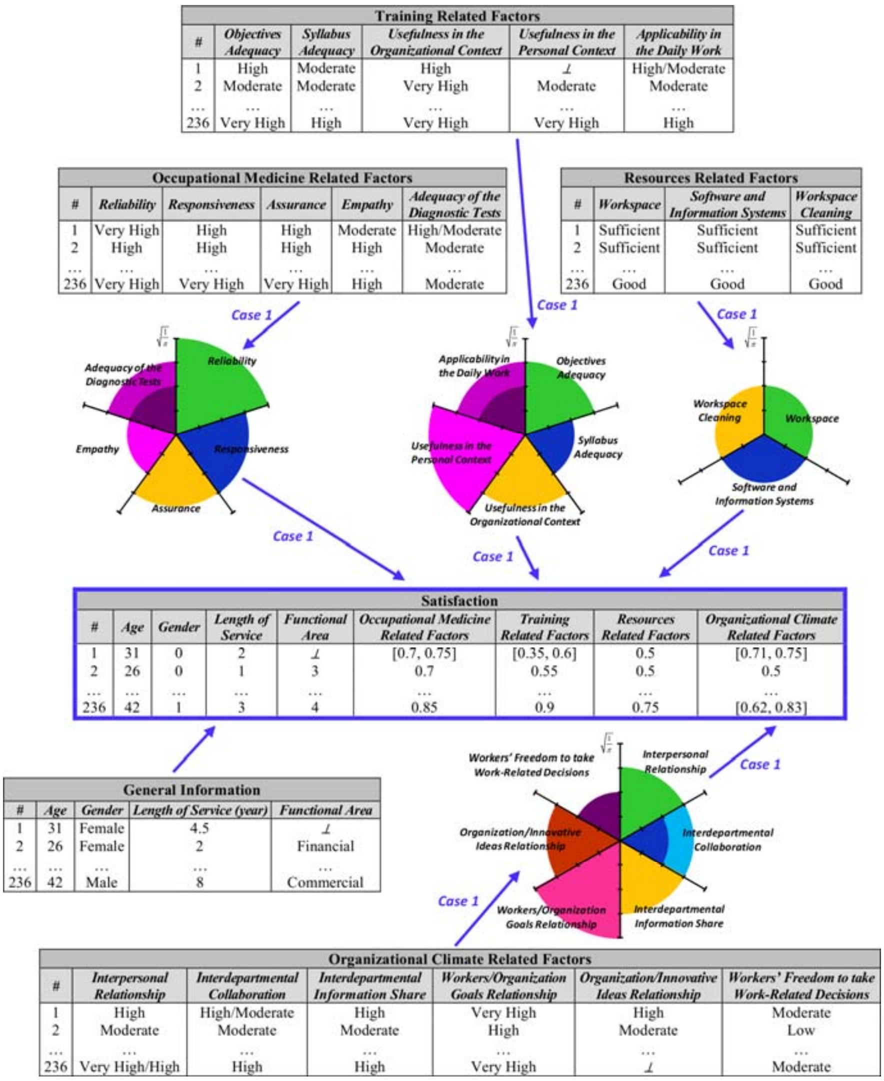
It is now possible to build up a knowledge database given in terms of the extensions of the relations (or tables) depicted in Fig. 7, which stand for a situation where one has to manage information aiming to estimate the workers' satisfaction. Thus, the *General Information*, *Training*, *Occupational Medicine*, *Resources*, and *Organizational Climate Related Factors* tables are populated with the responses to the issues presented in the questionnaire, where some incomplete, default and/or unknown data is present. For instance, in the former case the *Functional Area* is unknown (depicted by the symbol $\perp$), while the opinion about the *Applicability of the Training Received in the Daily Work* is not conclusive (*High/Moderate*).

The *Length of Service* column of the *Satisfaction* table is populated with 0 (zero), 1 (one), 2 (two) or 3 (three) that stands, respectively, for a length of service lesser than a year, comprised in the range [1, 3[, ranging between [3, 5[, and with more than 5 years. The *Functional Area* column, in turns, is filled with 0 (zero), 1 (one), 2 (two), 3 (three) or 4 (four) that denotes human resources, quality, marketing, financial and commercial issues, respectively. In the *Gender* column 0 (zero) and 1 (one) stand, respectively, for *female* and *male*.

In order to quantify the information present in the *Training*, *Occupational Medicine*, *Resources*, and *Organizational Climate Related Factors* tables the procedures already described above were followed. Applying the algorithm presented in [17] to the table or relation's fields that make the knowledge base for workers' satisfaction assessment (Fig. 7), and looking to the *DoCs* values obtained as described before, it is possible to set the arguments of the predicate ***satis**faction* (*satis*) referred to below, whose extensions denote the objective function regarding the problem under analyze:

$$satis\colon Age\,,\ G_{ender}\,,L_{ength}{o_f}S_{ervice}\,,F_{unctional}A_{rea}\,,T_{raining}\ R_{elated}F_{actors}\,,O_{ccupational}M_{edicine}R_{elated}F_{actors}\,,R_{esources}\ R_{elated}F_{actors}\,,O_{rganizational}C_{limate}R_{elated}F_{actors} \rightarrow \{0,\,1\}$$

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



Training Related Factors					
#	*Objectives Adequacy*	*Syllabus Adequacy*	*Usefulness in the Organizational Context*	*Usefulness in the Personal Context*	*Applicability in the Daily Work*
1	High	Moderate	High	⊥	High/Moderate
2	Moderate	Moderate	Very High	Moderate	Moderate
…	…	…	…	…	…
236	Very High	High	Very High	Very High	High

Occupational Medicine Related Factors					
#	*Reliability*	*Responsiveness*	*Assurance*	*Empathy*	*Adequacy of the Diagnostic Tests*
1	Very High	High	High	Moderate	High/Moderate
2	High	High	High	High	Moderate
…	…	…	…	…	…
236	Very High	Very High	Very High	High	Moderate

Resources Related Factors			
#	*Workspace*	*Software and Information Systems*	*Workspace Cleaning*
1	Sufficient	Sufficient	Sufficient
2	Sufficient	Sufficient	Sufficient
…	…	…	…
236	Good	Good	Good

Satisfaction								
#	*Age*	*Gender*	*Length of Service*	*Functional Area*	*Occupational Medicine Related Factors*	*Training Related Factors*	*Resources Related Factors*	*Organizational Climate Related Factors*
1	31	0	2	⊥	[0.7, 0.75]	[0.35, 0.6]	0.5	[0.71, 0.75]
2	26	0	1	3	0.7	0.55	0.5	0.5
…	…	…	…	…	…	…	…	…
236	42	1	3	4	0.85	0.9	0.75	[0.62, 0.83]

General Information				
#	*Age*	*Gender*	*Length of Service (year)*	*Functional Area*
1	31	Female	4.5	⊥
2	26	Female	2	Financial
…	…	…	…	…
236	42	Male	8	Commercial

Organizational Climate Related Factors						
#	*Interpersonal Relationship*	*Interdepartmental Collaboration*	*Interdepartmental Information Share*	*Workers/Organization Goals Relationship*	*Organization/Innovative Ideas Relationship*	*Workers' Freedom to take Work-Related Decisions*
1	High	High/Moderate	High	Very High	High	Moderate
2	Moderate	Moderate	Moderate	High	Moderate	Low
…	…	…	…	…	…	…
236	Very High/High	High	High	Very High	⊥	Moderate

Fig. 7. A fragment of the knowledge base for workers' satisfaction evaluation.

The algorithm presented in [17] encompasses different phases. In the former one the clauses or terms that make extension of the predicate under study are established. In the subsequent stage the arguments of each clause are set as continuous intervals. In a third step the boundaries of the attributes intervals are set in the interval [0, 1] according to a normalization process given by the expression $(Y - Y_{min})/(Y_{max} - Y_{min})$, where the Y_s stand for themselves. Finally, the *DoC* is evaluated as described in section 2.1.

Exemplifying with a term (worker) that presents the feature vector ($Age = 37$, $G_{ender} = 0$, $L_{ength}\,o_f S_{ervice} = \perp$, $F_{unctional}\,A_{rea} = 2$, $O_{ccupational}\,M_{edicine}\,R_{elated}\,F_{actors} = [0.45, 0.55]$, $T_{raining}\,R_{elated}\,F_{actors} = [0.65, 0.8]$, $R_{esources}\,R_{elated}\,F_{actors} = 0.67$, $O_{rganizational}\ C_{limate}\ R_{elated}\ F_{actors} = [0.58, 0.75]$), one may have:

Begin (DoCs evaluation)

The predicate's extension that sets the Universe-of-Discourse for the case (term) under observation is fixed

$\{$

$$\neg\, satis\left((QoI_{Age}, DoC_{Age}), \cdots, (QoI_{LoS}, DoC_{LoS}), \cdots, (QoI_{OCRF}, DoC_{OCRF})\right)$$

$$\leftarrow not\ satis\left((QoI_{Age}, DoC_{Age}), \cdots, (QoI_{LoS}, DoC_{LoS}), \cdots, (QoI_{OCRF}, DoC_{OCRF})\right)$$

$$satis\ \underbrace{\left((1_{37}, DoC_{37}), \cdots, (1_{\perp}, DoC_{\perp}), \cdots, \left(1_{[0.58,\,0.75]}, DoC_{[0.58,\,0.75]}\right)\right)}_{attribute`s\ values} :: 1 :: DoC$$

$$\underbrace{[19, 64] \quad \cdots \quad [0, 3] \quad \cdots \quad [0, 1]}_{attribute`s\ domains}$$

$\} :: 1$

The attribute's values ranges are rewritten

$\{$

$$\neg\, satis\left((QoI_{Age}, DoC_{Age}), \cdots, (QoI_{LoS}, DoC_{LoS}), \cdots, (QoI_{OCRF}, DoC_{OCRF})\right)$$

$$\leftarrow not\ satis\left((QoI_{Age}, DoC_{Age}), \cdots, (QoI_{LoS}, DoC_{LoS}), \cdots, (QoI_{OCRF}, DoC_{OCRF})\right)$$

$$satis\ \underbrace{\left(\left(1_{[37,\,37]}, DoC_{[37,\,37]}\right), \cdots, \left(1_{[0,\,3]}, DoC_{[0,\,3]}\right), \cdots, \left(1_{[0.58,\,0.75]}, DoC_{[0.58,\,0.75]}\right)\right)}_{attribute`s\ values\ ranges}$$

$$:: 1 :: DoC$$

$$\underbrace{[19, 64] \quad \cdots \quad [0, 3] \quad \cdots \quad [0, 1]}_{attribute`s\ domains}$$

$\} :: 1$

The attribute's boundaries are set to the interval [0, 1], according to a normalization process that uses the expression $(Y - Y_{min})/(Y_{max} - Y_{min})$

{

$$\neg\, satis\ \Big((QoI_{Age}, DoC_{Age}), \cdots, (QoI_{LoS}, DoC_{LoS}), \cdots, (QoI_{OCRF}, DoC_{OCRF})\Big)$$

$$\leftarrow not\ satis\ \Big((QoI_{Age}, DoC_{Age}), \cdots, (QoI_{LoS}, DoC_{LoS}), \cdots, (QoI_{OCRF}, DoC_{OCRF})\Big)$$

$$satis\ \underbrace{\Big((1_{[0.4,0.4]}, DoC_{[0.4,0.4]}), \cdots, (1_{[0,1]}, DoC_{[0,1]}), \cdots, (1_{[0.58,0.75]}, DoC_{[0.58,0.75]})\Big)}_{attribute`s\ values\ ranges\ once\ normalized} :: 1 :: DoC$$

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

} $:: 1$

{

$$\neg\, satis\ \Big((QoI_{Age}, DoC_{Age}), \cdots, (QoI_{LoS}, DoC_{LoS}), \cdots, (QoI_{OCRF}, DoC_{OCRF})\Big)$$

$$\leftarrow not\ satis\ \Big((QoI_{Age}, DoC_{Age}), \cdots, (QoI_{LoS}, DoC_{LoS}), \cdots, (QoI_{OCRF}, DoC_{OCRF})\Big)$$

$$satis\ \underbrace{((1,\ 1),\ \cdots,\ (1,\ 0),\ \cdots,\ (1,\ 0.98))}_{\substack{attribute`s\ quality-of-information\\ and\ respective\ confidence\ values}} :: 1 :: 0.87$$

$$\underbrace{[0.4,\ 0.4]\ \cdots\ [0,\ 1]\ \cdots\ [0.58,\ 0.75]}_{attribute`s\ values\ ranges\ once\ normalized}$$

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

} $:: 1$

End.

5 A Case Based approach to Computing

The framework presented previously shows how the information comes together and. In this section, a soft computing approach was set to model the universe of discourse, where the computational part is based on a *CB* approach to computing. Contrasting with other problem solving tools (e.g., those that use *Decision Trees* or *Artificial Neural Networks*), relatively little work is done offline [9]. Undeniably, in almost all the situations, the work is performed at query time. The main difference between this new approach and the typical *CB* one relies on the fact that not only all the cases have their arguments set in the interval [0, 1], but it also caters for the handling of incomplete, unknown, or even self-contradictory data or knowledge. Thus, the classic *CB* cycle was changed (Fig. 6), being the *Case Base* given in terms of the following pattern:

$$Case = \{< Raw_{data}, Normalized_{data} >\}$$

When confronted with a new case, the system is able to retrieve all cases that meet such a structure and optimize such a population, having in consideration that the cases retrieved from the *Case-base* must satisfy the invariant present in equation (5), in order to ensure that the intersection of the attributes range in the cases that make the *Case Base* repository or their optimized counterparts, and the equals in the new case cannot be empty. Having this in mind, the algorithm described above is applied to a *new case*, that in this study presents the feature vector ($Age = \bot$, $G_{ender} = 1$, $L_{ength}\, O_f S_{ervice} = 2$, $F_{unctional} A_{rea} = 1$, $O_{ccupational} M_{edicine} R_{elated} F_{actors} = 0.7$, $T_{raining}\ R_{elated}\ F_{actors} = [0.5, 0.7]$, $R_{esources} R_{elated} F_{actors} = [0.67, 0.75]$, $O_{rganizational} C_{limate} R_{elated} F_{actors} = 0.75$). Then, the computational process may be continued, with the outcome:

$$new_{case} \underbrace{\big((1,0), (1,1), (1,1), (1,1), (1,1), (1,0.98), (1,0.99), (1,1)\big)}_{new\ case} :: 1 :: 0.87$$

Now, the *new case* may be portrayed on the *Cartesian* plane in terms of its *QoI* and *DoC*, and by using clustering methods [22] it is feasible to identify the cluster(s) that intermingle with the *new one* (epitomized as a square in Fig. 8). 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*. Thus, one may have:

$$\underbrace{\begin{array}{c} retrieved_{case_1}\big((1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,0.95)\big) :: 1 :: 0.99 \\ retrieved_{case_2}\big((1,1), (1,1), (1,0), (1,0), (1,1), (1,1), (1,0.85), (1,1)\big) :: 1 :: 0.73 \\ \vdots \\ retrieved_{case_j}\big((1,1), (1,1), (1,0), (1,0), (1,0), (1,1), (1,1), (1,0.97)\big) :: 1 :: 0.62 \end{array}}_{normalized\ cases\ that\ make\ the\ retrieved\ cluster}$$

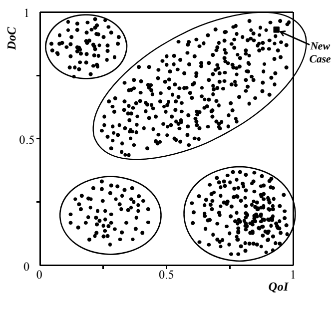


Fig. 8. A case's set divided into clusters.

Assuming that every attribute has equal weight, for the sake of presentation, the *dis(imilarity)* between new_{case} and the $retrieved_{case_1}$, i.e., $new_{case \to 1}$, may be computed as follows:

$$dis^{DoC}_{new\ case \to 1} = \frac{\|0-1\| + \cdots + \|0.98-1\| + \|0.99-1\|\|1-0.95\|}{8} = 0.14$$

Thus, the *sim(ilarity)* for $sim^{DoC}_{new\ case \to 1}$ is set as $1 - 0.14 = 0.86$. Regarding *QoI* the procedure is similar, returning $sim^{QoI}_{new\ case \to 1} = 1$. Thus, one may have:

$$sim^{QoI,\,DoC}_{new\ case \to 1} = 1 \times 0.86 = 0.86$$

a value that may now be object of a *reading* by part of experts. These procedures should be applied to the remaining cases of the retrieved cluster(s) in order to obtain the most similar ones, which may stand for the possible solutions to the problem. However, there is yet a problem, i.e., *when meeting the situation of multiple experts and multiple assessment methods, how to integrate them?*

In order to answer to this question, let us consider that one has p ($p \geq 2$) experts and the making $e_i/domain_i$, where e_i stands for $expert_i$, and $domain_i$ denotes the *metrics* or *methods* used by $expert_i$ to read the study outcome, i.e., the pair (QoI_i, DoC_i), here given in terms of $sim^{QoI,\,DoC}_{newcase \to j}$, where "$j$" stands for the $case_j$ in the cluster(s) of retrieved cases. $domain_i$ may be, for example, a set (e.g., *{low, moderate, high, very high}*), an interval (e.g., *[80,90]*), a number (e.g., *80*), or an unknown value (e.g., $\perp$). A pictorial view of the process is given by Fig. 9 (*that sets the relation experts/readings*), and Fig. 10 (*that sets the overall assessment*).

Experts

e_1/{low, moderate, high, very high}	e_2/[0, 100]	...	e_p/[20, 40]
low/moderate	[80, 90]	...	$\perp$

Readings

Fig. 9. The relation *experts/readings*.

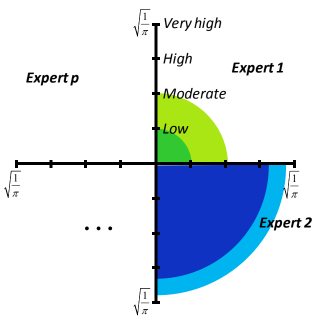


Fig. 10. The overall assessment that is given by the areas`sum.

In order to evaluate the performance of the proposed model the dataset was divided in exclusive subsets through the ten-folds cross validation [19]. In the implementation of the respective dividing procedures, ten executions were performed for each one of them. Table 1 presents the coincidence matrix of the *CB* model, where the values presented denote the average of 20 (twenty) experiments. A perusal to Table 1 shows that the model accuracy was 92.4% (i.e., 218 instances correctly classified in 236). Thus, the predictions made by the *CB* model are satisfactory, attaining accuracies higher than 90%. The sensitivity and specificity of the model were 95.0% and 87.0%, while *Positive* and *Negative Predictive Values* were 93.8% and 89.3%, respectively. The *ROC* curve is shown in Fig. 11. The area under *ROC* curve (0.91) denotes that the model exhibits a good performance in the assessment of workers' satisfaction.

Table 1. The coincidence matrix for *CB* model.

Target	**Predictive** True (1)	False (0)
True (1)	151	8
False (0)	10	67

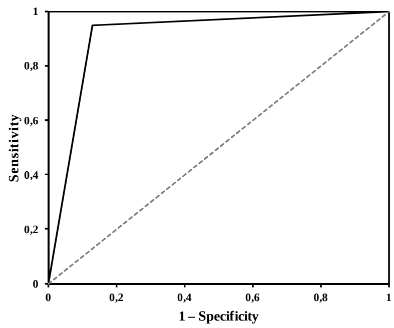


Fig. 11. The *ROC* curve regarding the proposed model.

6 Conclusion

The workers' satisfaction assessment is not only an inestimable practice, but something of utmost importance in the organization efficiency context. To meet this challenge it is necessary that the organizations optimize their efficiency in order to achieve excellence practices. However, it is difficult to assess the workers' satisfaction since it is necessary to consider different variables and/or conditions, with complex relations entwined them, where the data may be incomplete, contradictory, and even unknown. In order to overcome these difficulties this work presents a *Decision Support System* to estimate the workers' satisfaction. The methodology followed was centred on a formal framework based on *LP* for knowledge representation and reasoning, complemented with a *CB* approach to computing. It may set the basis for an overall approach to such systems in this arena. Indeed, it has also the potential to be disseminated across other prospective areas, therefore validating an universal attitude. Indeed, under this line of thinking the cases' retrieval and optimization phases were heightened when compared with existing tactics or methods. Additionally, under this approach the users may define the cases weights attributes on-the-fly, letting them to choose the appropriate strategies to address the problem (i.e., it gives the user the possibility to narrow the search space for similar cases at runtime). Finally, it was presented a solution to the question: *when meeting the situation of multiple experts and multiple assessment methods, how to integrate them?* This is quite important, for example, in-group(s) construction and in the assessment of their outcomes.

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. Elena, N.I.: Human Resources Motivation – An Important Factor in the Development of Business Performance. Annals of Faculty of Economics of University of Oradea 1(1), 1039–1045 (2012)
2. An, N., Liu, J., Wang L., Bai, Y.: Employee Satisfaction as an Important Tool in Human Resources Management. In Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing, pp. 1–4, IEEE Edition (2008)
3. Alfes, K., Shantz, A.D., Truss, C., Soane, E.C.: The link between perceived human resource management practices, engagement and employee behaviour: A moderated mediation model. The International Journal of Human Resource Management 24, 330–351, (2013)
4. Berman, E.M., Bowman, J.S., West J.P., Van Wart, M.R.: Human Resource Management in Public Service: Paradoxes, Processes, and Problems. SAGE Publications Inc., California (2015)
5. Petrescu, A.I., Simmons, R.: Human resource management practices and workers' job satisfaction. International Journal of Manpower 29, 651 – 667 (2008)
6. 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 5th Generation Challenge, pp. 50–54. Association for Computing Machinery, New York (1984)
7. Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Communications 7, 39–59 (1994)
8. Richter, M.M., Weber, R.O.: Case-Based Reasoning: A Textbook. Springer, Berlin (2013)
9. 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)
10. 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)
11. Quintas, A., Vicente, H., Novais, P., Abelha, A., Santos, M.F., Machado, J., Neves, J.: A Case Based approach to assess Waiting Time Prediction at an Intensive Care Unity., In:

Arezes, P. (ed.) Advances in Safety Management and Human Factors. Advances in Intelligent Systems and Computing, vol. 491, pp. 29–39. Springer International Publishing, Cham (2016)
12. 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)
13. 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)
14. 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)
15. 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)
16. 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)
17. 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)
18. Vicente, H., Dias, S., Fernandes, A., Abelha, A., Machado, J., Neves, J.: Prediction of the Quality of Public Water Supply using Artificial Neural Networks. Journal of Water Supply: Research and Technology – AQUA 61, 446–459 (2012)
19. Haykin, S.: Neural Networks and Learning Machines. Pearson Education, New Jersey (2009)
20. Mendes, R., Kennedy, J., Neves J.: Watch thy neighbor or how the swarm can learn from its environment. In Proceedings of the 2003 IEEE Swarm Intelligence Symposium (SIS'03), pp. 88–94, IEEE Edition (2003)
21. 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)
22. Figueiredo, M., Esteves, L., Neves, J., Vicente, H.: A data mining approach to study the impact of the methodology followed in chemistry lab classes on the weight attributed by the students to the lab work on learning and motivation. Chemistry Education Research and Practice 17, 156–171 (2016)