

Hybrid Subjective Decision Support System based on Computational Semiotics and Computational Intelligence Techniques

Denis Martins and Fernando Buarque de Lima Neto

Computing Engineering Program, School of Engineering,
University of Pernambuco, Recife,
Brazil

{dmlm, fbln}@ecom.poli.br

Abstract. Subjective decision problems involve personal feelings and opinions, adding substantial complexity to evaluate different candidate options. In order to deal with this kind of problem, in which individual experience is considered and impacts directly in decision making process, many computational methods have been applied. However, the traditional approaches are often not flexible to consider uncertainties, imprecise situation contexts and idiosyncrasies. In this sense, we present a Hybrid Subjective Decision Support System based on Computational Semiotics and Computational Intelligence techniques. Our approach relies on Case-based Reasoning as the problem solving main methodology and Self-organizing Maps, which acts as pattern recognition tool, in order to organize more appropriately retrieval of similar past cases. Furthermore, a semiotic model handles a prior knowledge (i.e. knowledge acquired from a specialist) and domain specific restrictions to guide the search process towards an appropriate problem solution. In the paper we illustrate how the proposed approach can deal graciously with subjective concepts providing a more intuitive and evident decision making.

Keywords: Subjective decision, case-based reasoning, self-organizing maps, semiotics, decision support system.

1 Introduction

In dynamic environments, decision-making is usually an arduous task, especially when subjective decisions are necessary. In such situations, individual experience or personal opinions have to be considered for selecting an appropriate alternative from a set of possible options (*e.g.* decide which car to buy or make a leisure travel plan). This could take a large amount of time, not only because the concept of appropriateness change frequently, but also because environmental changes (than may occur rapidly).

In this context, Computational intelligence (CI) aims to provide adaptive mechanisms inspired by nature, possibilitating the construction of computer systems with intelligent behavior for tackling with such complex and dynamic kind of problems. The use of intelligent techniques such as Artificial Neural Networks and Case-Based Reasoning affords, for example, to these systems the ability to learn and adapt to new situations[1][2]. Thus, we argue that adaptive methods combined with Semiotic concepts of sign deconstruction and consequent individualized reasoning may assist well in subjective decision-making process, providing the needed individualization of decisions while improving efficiency and precision along time (due to the adaptive nature of CI mechanisms).

2 Semiotics

Semiotics is a field of human science which studies signs [3]. It involves the investigation of cognitive process such as communication and interpretation, as well as the study of how meaning are made and represented [4]. Charles Sanders Peirce (1857-1913), defined sign as something that, in some way, means something to someone in some respect or capacity [5]. Peirce model of sign can be seen as a triadic structure, as shown in Fig. 1. The object consists in something that is referred by the sign, the *interpretant* is the effect caused by a sign in someone's mind (*i.e.* meaning of the sign) and, finally, the sign or *representamen* is related to the form that sign takes.

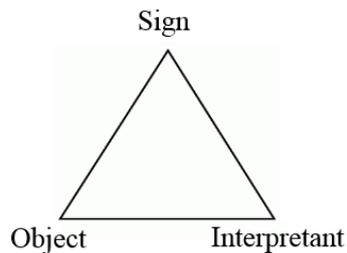


Fig. 1. Triadic structure of a sign as defined by Peirce.

Computational Semiotics is a relatively recent research field that aims to combine a computational approach with some important concepts grounded by Semiotics in order to develop a methodology to artificially create intelligent systems [3]. Because a sign can be deconstructed continuously and its constituents parts became grounded in the mind of the subject, a semiotic-based system is suitable for modeling dynamic processes in which, for example, decision parameters and options change frequently in relation to a particular individual [6].

In this context Computational Semiotics approaches can also apply CI methods, such as Artificial Neural Networks, Case-Based Reasoning, Evolutionary or Swarm computation, to provide the ability to learn and adapt themselves, this time, to problem contextual changes, new situations and, consequently, to idiosyncrasies as

well [1]. Indeed, those methods are already been used in function optimization problems, time series prediction, adaptive control, classification tasks and clustering data [7].

Furthermore, a semiotic model handles a prior knowledge (i.e. knowledge acquired from a specialist) and domain specific restrictions to guide the search process towards an appropriate problem solution. As this type of information is, in general, subjective and imprecise, this system may provide a useful computational modeling framework which can encompass some problem-domain concepts that are not clear or contain some uncertainty [8].

The application of Semiotics concepts in intelligent systems can also bring advantages in problems that involve situation or context awareness [9]. As semiotic analysis intends primarily to understand situational dependent concepts, we argue that these analytical processes are also able to model objects or contextual information [3]. In this work, we argue that Semiotics can therefore provide a framework to recognize situation patterns, identifying relevant not obvious information and be helpful in the decision-making process.

3 Case-Based Reasoning

Differently from other Artificial Intelligence methods, Case-based Reasoning (CBR) is a problem solving technique which allows a system to reason and to learn from previous experienced (contextual) situations [10]. These experiences are represented as cases, which contain both problem description and its solution. Applying CBR, whenever a new problem appears, the system searches into a Case Base (CB) in order to retrieve the most similar past cases based on a similarity measure. Then, the retrieved cases are combined and adapted for building a solution that can be used for the new problem.

The Case Base is a collection of past cases used CBR cycle. In general, CB can contains both successful and non-successful cases (*i.e.* cases with a solution that did not respond well to a given problem situation). The first ones guide the search process towards an appropriate problem solution while failed experiences allow system to avoid some states that are unwanted, like occur in human mind reasoning process [2].

In CBR, the reasoning process is incremental, similar to natural human decision making, and can be typically summarized in four steps [11] below, also shown in Fig. 2:

1. Retrieve: given a new problem situation, recall the most similar cases;
2. Reuse: use or combine retrieved cases in order to solve the problem;
3. Revise: evaluate the proposed solution, by an expert or in a real world application, for example. If the solution is not appropriated, some adaptations are made in order to better fit the current problem or to satisfy a specific constraint;
4. Retain: store the proposed solution in Case Base as a new solved case to be utilized in the next CBR cycle.

In this work, CBR approach is used as a problem solving methodology to support intelligent decision-making process which learns from the user interaction. As

presented in following sections, we applied a Self-organizing Maps technique, to improve the case-based process, especially to make it more flexible – as a semiotic system would require being.

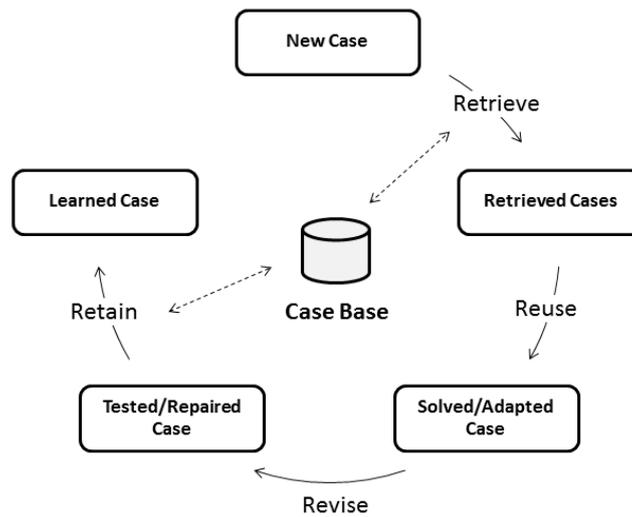


Fig. 2. CBR Cycle. Reasoning process organized in four steps: Retrieve, Reuse, Revise and Retain.

4 Self-Organizing Maps

The Self-organizing Map (SOM) introduced by Teuvo Kohonen in the early 80's [12] is an unsupervised Neural Network algorithm. Inspired by the human brain cortex, SOM creates a topologically map in which similar data are grouped together by specialized neurons. Hence, this technique is a non-parametric model usually applied as a clustering or a pattern recognition algorithm [13].

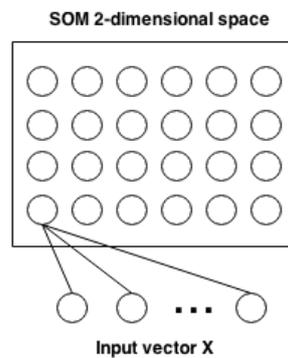


Fig. 3. Basic representation of a two-dimensional SOM.

A common representation of a SOM is a two dimensional grid of neurons, as shown in Fig. 3. Each output neuron is fully connected with an input neuron by a weight link.

The mapping process realized by a SOM arranges the input data preserving topological information [14]. Thus, the arrangement reflects the underlying neighborhood relationships between data samples creating groups called clusters. Moreover, this arrangement often gives a quite relevant overview about what is not known, obvious or about the expected patterns; incidentally, information already present in data set.

During the training process, a given input data vector is presented to SOM and the most similar (closest) output neuron is select as a winner neuron, or best match unit (BMU). In Kohonen’s work, similarity criterion used is the Euclidean distance [12] and the winner neuron is the one which has the minimum distance value. This process is unsupervised because no labeled data is required.

SOM employs a self-organizing training that aims to minimize the distance adjusting the BMU weights and its neighborhood toward the input vector using the following rule:

$$w_{ij}(t + 1) = w_{ij}(t) + \alpha(t)h_{ci}(t)[x(t) - w_{ij}(t)], \quad (1)$$

where, w_{ij} is the weight j of neuron i , $\alpha(t)$ is the learning rate, and $h_{ci}(t)$ is the neighborhood radius function centered on winner neuron c with respect to neuron i . Typically, both leaning rate and neighborhood radius are functions which decrease with time, in order to facilitate system convergence.

Moreover, a trained SOM can be viewed as a data recognition system in which a set of specialized neurons respond to certain stimuli (i.e. a specific type or class of inputs), similar to the process that occurs in brain in which different neurons (sensory areas) respond to a different kind of sensorial stimulus [15].

Other SOM applications involve dimensionality reduction, knowledge discovery in databases (KDD) and intelligent decision support [16]. In our proposed model, a SOM is used as data visualization tool which allows users to identify similar cases during the Case-based Reasoning process.

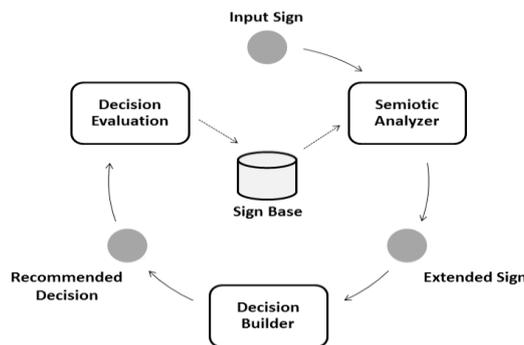


Fig. 4. Proposed subjective decision support process, inspired on CBR cycle.

5 Overview of the Proposed System

We aim to elaborate an intelligent system based on Semiotics and Computational intelligence techniques in order to support subjective decision-making. The conceptual construction followed a blend of CI and Semiotics ethos, derived hybridization of CI techniques, and inspiration on the CBR cycle. The Hybrid Subjective Decision Support Cycle is illustrated in Fig 4.

We use the concept of sign, proposed by Peirce, as the knowledge representation model. Similarly in CBR, a sign is a special type of case that represents a specific situation perceived by the system. In other words, signs can model user interest on a product (*e.g.* on a product for purchasing support or e-commerce applications or patient data in clinical domains, for example). A sign consists of three basic parts:

- Problem description: a feature vector describing both contextual information and situational description, corresponding to *representamen* in Peirce definition;
- Solution: contains subjective information about how the sign was interpreted by someone (*e.g.* a domain specialist or system users), as *interpretant* concept;
- Historical data: describes association among the sign and other signs previously perceived by the system.

Whenever a new problem situation appears, represented by an input sign, the Semiotic Analyzer module (SA) retrieves similar signs from the Sign Base. SA promotes a semiotic deconstruction of retrieved signs extracting contextual data in order to extend input sign. Thus, the extended sign (containing historical information) are adapted by the Decision Builder, combining data from retrieved signs, outputting a recommended decision. Hence, a user can test the decision and store it in the Sign Base for future application.

5.1 Retrieval of Similar Signs and Semiotic Deconstruction

To realize the retrieve phase, SA employs a SOM to search for similar signs. Each sign stored in Sign Base is indexed by a BMU in the SOM neuron grid during the system training phase, as proposed by Wang et al. [17]. Thus, similarity assessment is realized by calculating Euclidean Distance between the index of the input sign and all the indexes of previous signs. The top n indexes with smallest distance are selected and retrieved as similar signs.

Hence, SA employs a semiotic deconstruction on retrieved signs comparing their data with the input sign. This operation is realized analyzing each feature value in sign problem description part. Whenever the input sign presents an unexpected value, then it is marked with a warning tag. Deconstruction is described as follows:

- Consider S a sign feature vector: $S = \{s_1, s_2, s_3, \dots, s_n\}$
- For each feature i in retrieved signs feature set, easy calculations of average μ_i and standard deviation σ_i are carried out.
- Mark feature s_i with corresponding tag using the following rule:

$$\text{tag}(s_i) = \begin{cases} \text{normal}, & \mu_i - \sigma_i \leq s_i \leq \mu_i + \sigma_i \\ \text{warning}, & \text{otherwise} \end{cases} \quad (2)$$

5.2 Recommended Decision

After extend the input sign with tags based on historical and similarity data, the Decision Builder module adapts the extended sign attaching a meaning to it, *i.e.* combine solution information from retrieved signs and associate it to input sign. This is simple, yet very relevant information for the aimed individualization process that tackles subjectivity. In this work, we employ most frequent adaptation to create the recommended decision. This adaptation process evaluates the solution part of all retrieved signs and uses the one with more occurrences to determine final solution decision configuration.

6 Experiments and Results

In order to evaluate our approach we apply our model to subjective decision making problems, where context of evaluation is central.

We select two benchmarking datasets provided by UCI Machine Learning Repository [18]: Pittsburgh Bridges and Auto Imports Database to try out our proposal: The former, in Semiotic Analyzer Module, and the latter in the Decision Builder Module. All data sets have been normalized using Eq. (3), nominal values converted into numeric values and entries with missing values were disregarded.

$$\text{Normalized}(x) = \frac{x - x_{MIN}}{x_{MAX} - x_{MIN}}, \quad (3)$$

where x is the current value, x_{MIN} and x_{MAX} are minimum and maximum values of x in dataset, respectively.

6.1 Selection of Parameters

Selection of parameters of computational models often impacts the obtained results quality. Especially in neural-based systems, the selection of appropriate parameters and training configurations can improve performance and avoid problems such as overfitting [1]. As this topics is widely discussed in the literature and out of the scope of the article, we assumed that parameters were selected using common heuristics for neural-system, *e.g.*, Kasaboy [22].

Related to SOM, some important parameters are neuron grid size, initial learning rate, neighborhood radius, and its updating methods. Although there are many approaches to set these parameters, some standards have been established by many authors [19][20]. In our experiments, we have used a 8×6 rectangular neuron map with initial learning rate as 0.1 (decreasing at each iteration to a minimum of 0.01) and performed 10 runs of 1000 iterations.

6.2 Evaluating Semiotic Analyzer Module

Pittsburgh Bridges dataset was used here as it contains information about several distinct bridges. Seven specification properties (problem and context information) and five design description properties were utilized. The evaluation task consists in enter an input bridge specification to the system for prediction of design properties. Results are summarized in Fig. 5 and Table 1.

To evaluate the quality of sign mapping provided by SA, based on SOM algorithm, we employ two measures: Quantization Error and Topological Error [12] [13]. Quantization Error (QE) evaluates the quality of SOM learning process computing how well the output neuron weights fits to data. It is calculated taking the average difference between N inputs and their correspondent BMU weights, as follows:

$$QE = \frac{1}{N} \sum_{i=1}^N |x_i - m_i| \tag{4}$$

where, x_i is the input vector and m_i is the BMU weight vector.

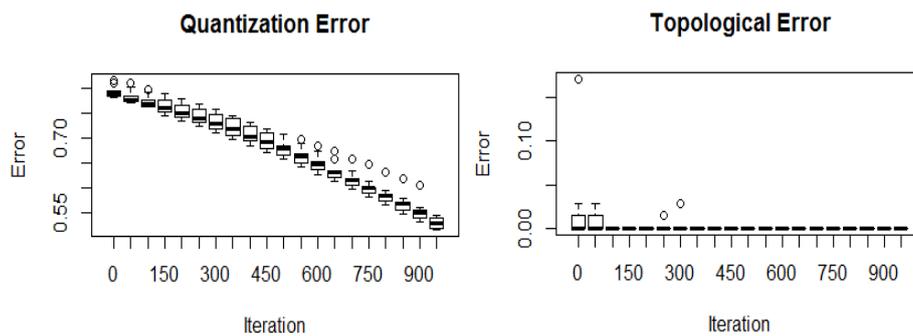


Fig. 5. Quantization and Topological errors for Pittsburgh Bridges dataset.

Table 1. Semiotic Analyzer evaluation results.

Properties	Input bridge	Bridge 1	Bridge 2	Bridge 3	SA tag
River	M	M	A	A	Warning
Location	3	6	28	25	Warning
Period	Crafts	Crafts	Crafts	Crafts	Normal
Purpose	Highway	Highway	Aqueduct	Highway	Normal
Length	Medium	Medium	Medium	Medium	Normal
Lanes	2	2	1	2	Normal
Clear-G	N	N	N	N	Normal
T-or-D	-	Through	Through	Through	-
Material	-	Wood	Iron	Wood	-
Span	-	Medium	Short	Short	-
Span length	-	S	S	S	-
Type	-	Wood	Wood	Wood	-

Topological Error (TE) represents the quality of map topologic preservation. It indicates the proportion of data items for which the first and second-best match units are not close to each other [13]. Whenever those units are not next in map error value is 1, otherwise 0.

As shown in Table 1, SA analyzed bridge specification properties about period, purpose, length, number of lanes and clear-G and retrieve similar bridges, only using contextual information. For a given input say, river and location, SA alerts existence of some divergent items in retrieved data, this, importantly, directs the user attention towards those properties in order to evaluate their values more carefully, helping to reduce cognitive overload on the decision-making process at the same time it continues to produce sound recommendations.

By analyzing QT and TE measures we illustrate that SOM behaves as an effective mechanism of, in this example, bridge information assessment, indexing bridge signs in Sign Base and grouping them in a similarity map; precisely, as hypothesized.

6.3 Purchasing Decision

Purchasing decision is another common subjective decision example. When people look for a product or service to purchase, they consider a large number of set of interest variables: price, brand, presence of a virtual shop, quality and so on. However, those variables are often impacted by personal experience or opinion, increasing decision making complexity, particularly for decision support systems.

In this example, our proposed system helps users to find an appropriate car based on theirs interests. We use information about 197 cars from Auto Imports Database. This dataset describes a car model based on properties such as number of door, body-style, horsepower, fuel type and price.

User inputs a query containing his interest about a car and, based on this information, system shows a set of car options. To retrieve appropriate car models, we defined the concept of relevance. A relevant product is the one that is in accordance with user-defined query at a minimum of 80%, related to each feature of the desired item. Thus, we use an adapted Heterogeneous Euclidean-Overlap Metric (HEOM) as relevance function. This function defines a distance between two items, here the query and each car model in dataset, as shown in Eq (5) and (6):

$$Relevance = \sqrt{\sum_{i=1}^m d(x_i, y_i)^2} \quad (5)$$

$$d(x_i, y_i) = \begin{cases} 1, & \text{if } x_i \text{ or } y_i \text{ are unknown} \\ \frac{|x_i - y_i|}{max_i - min_i}, & \text{otherwise} \end{cases} \quad (6)$$

Next, we apply Recall and Precision measures [21], widely used to evaluate information retrieval systems, in order to assess cars relevance. Precision is defined as the proportion of retrieved documents that are relevant, while Recall is the proportion of relevant documents that are retrieved by the system. Tables 2 and 3 summarize the results.

Table 2. Results of the Recommender of cars.

Car properties	User query	Car option 1	Car option 2	Car option 2
Num. of doors	2	4	4	4
Fuel type	Gas	Gas	Gas	Gas
Body style	Sedan	Sedan	Sedan	Hatch
Fuel consumption	-	29	29	29,5
Price	\$10.000	\$9.960	\$9.995	\$9.988
Relevance	-	100%	100%	99%

The proposed system proved to be able to perform the recommendation of products based on user defined interests, even if it is provided with incomplete information. Thus, there is no need to fill all the characteristics of the product by the user: a small amount of information is enough to guide the process of searching for relevant items.

Table 3. Information retrieval measures.

Measure	Average	Standard deviation
Precision	89%	0,17
Recall	75,2%	0,82

The obtained results show an interesting behavior of the system. Although user has preferred a sedan car, system returned a hatch car. This behavior is important, considering real purchase situations where a human seller could offer a product of different types, but which suits customer's interests. Ultimately, this reveals the system ability to produce results that evidences hidden information, necessary for the subject decision.

7 Conclusion

In this work, we proposed a hybrid decision support system based on Semiotics and Computational Intelligence (*e.g.* Neural Networks and Case-Based Reasoning) readily applicable to subjective decision problems. Semiotics concepts, such as signs, sign deconstruction and contextual information processing allowed the development of a meaning deconstruction process which, applied together with Computational Intelligence methods, provide a degree of individualization of decisions, while improving efficiency and precision.

Initial experimental results illustrate that our approach is easy to be used, yet effective to help users along the selection of decision alternatives, reducing their cognitive overload. Additionally the proposal may also support a more intuitive subjective and evident decision making.

More experiments have to be cared out, especially within larger decision scenarios, *i.e.* more cases and conflicting features. Other distance measures could also be tried out, as well as the assessment of appropriateness.

Acknowledgements. The authors would like to thanks CAPES and University of Pernambuco for the financial support of this research.

References

1. A. Engelbrecht: Computational intelligence: an introduction, 2nd ed. Chichester: Wiley (2007)
2. M. M. Richter and R. O. Weber: Case-Based Reasoning. Berlin, Heidelberg: Springer Berlin Heidelberg (2013)
3. R. Gudwin: From semiotics to computational semiotics. In Proceedings of the 9th International Congress of the German Society for Semiotic Studies/ 7th International Congress of the International Association for Semiotic Studies (IASS/AIS) (1999)
4. D. Chandler: Semiotics: the basics. Second ed. New York: Routledge (2007)
5. C. S. Peirce: Collected Papers of Charles Sanders Peirce. Harvard University Press (1958)
6. R. Gudwin and J. Queiroz: Semiotics and Intelligent Systems Development. Hershey: IGI Global (2006)
7. D. F. Pacheco: Hybrid Intelligent Suite for Decision Support. Vol. 55 (2007)
8. M. Kwiatkowska and K. Kielan: Fuzzy logic and semiotic methods in modeling of medical concepts. Fuzzy Sets Systems, Vol. 214 (2013)
9. J. Hugo: The semiotics of control room situation awareness. In Fourth International Cyberspace Conference. Ergon (2005)
10. A. Aamodt and E. Plaza: Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Community, Vol. 7 (1994)
11. J. L. Kolodner: An introduction to case-based reasoning. Artif. Intell. Rev., Vol. 6 (1992)
12. T. Kohonen: The self-organizing map. In Proc. IEEE (1990)
13. B. Hammer, A. Gisbrecht, and A. Schulz: Advances in Self-Organizing Maps. In 9th International Workshop, WSOM 2012 (2012)
14. S. Sklorz: Concept Formation by Combining Neural Networks and Machine Learning. (1996)
15. H. Ritter and T. Kohonen: Self-organizing semantic maps, Biol. Cybern., Vol. 254 (1989)
16. Z. Y. Zhuang, L. Churilov, F. Burstein, and K. Sikaris: Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners. Eur. J. Oper. Res., Vol. 195, no. 3, pp. 662–675 (2009)
17. H. Wang, J. Gao, and Z. Jiang: Rotating machinery fault case-based reasoning based on vector space model and SOM neural network. J. Comput. Inf. Syst., Vol. 19 (2012)
18. K. Lichman and M. Bache: UCI Machine Learning Repository. University of California, Irvine.
19. S. Haykin: Neural networks and learning machines. 3rd ed. New Jersey: Pearson Education, (2009)
20. T. Kohonen: Self-Organizing Maps. 3rd ed. Berlin: Springer (2001)
21. A. Drigas and J. Vrettaros: Using the Self-Organizing Map (SOM) Algorithm as a Prototype E-Content Retrieval Tool. Sci. Its Appl. (2008)
22. N. Kasbov: Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering (Computational Intelligence), MIT PRESS (1996)