An assessment of environmental and toxicological risk to pesticide exposure based on a case-based approach to computing

Cristina Coelho[1], Henrique Vicente[1, 2], M. Rosário Martins[3], Nelson Lima[4], Mariana Neves[5] and José Neves[2, 6]

[1] Departamento de Química, Escola de Ciências e Tecnologia, Universidade de Évora, Évora, Portugal
[2] Centro Algoritmi, Universidade do Minho, Braga, Portugal
[3] Departamento de Química, Laboratório HERCULES, Escola de Ciências e Tecnologia, Universidade de Évora, Évora, Portugal
[4] Centro de Engenharia Biológica, Micoteca da Universidade do Minho, Universidade do Minho, Braga, Portugal
[5] Deloitte, London, United Kingdom

E-mail: jneves@di.uminho.pt

Abstract. Pesticide environmental fate and toxicity depends on its physical and chemical features, the soil composition, soil adsorption, as well as residues that may be found in different soil slots. Indeed, pesticide degradation in soil may be influenced by either biotic or abiotic factors. In addition, the toxicity of pesticides for living organisms depends on their adsorption, distribution, biotransformation, dissemination of metabolites together with interaction with cellular macromolecules and excretion. Biotransformation may result in the formation of less toxic and/or more toxic metabolites, while other processes determine the balance between toxic and a nontoxic upcoming. Aggregate exposure and risk assessment involve multiple pathways and routes, including the potential for pesticide residues in food and drinking water, in addition to residues from pesticide use in residential and non-occupational environments. Therefore, this work will focus on the development of a decision support system to assess the environmental and toxicological risk to pesticide exposure, built on top of a *Logic Programming* approach to *Knowledge Representation and Reasoning*, complemented with a *Case Based* attitude 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.

1. Introduction

The use of agricultural pesticides is mounting nowadays, aiming at the control of insect pests and diseases. They are a diverse group of inorganic and organic chemicals that include herbicides, insecticides, nematicides, fungicides and soil fumigants. Pesticide sprays may directly hit non-target vegetation, or may drift or volatilize from the treated area and contaminate air, surface, ground water, soil, and non-target plants. They can be moved from soil by runoff and leaching, thereby constituting a potential problem for the supply of drinking water to the inhabitants [1]. Pesticide environmental fate and toxicity depends on the physical and chemical characteristics of pesticide, the soil composition,

and soil adsorption. Pesticide degradation in soil may be influenced by both biotic and abiotic factors. Microbial activity has been considered to be the most influential and significant cause of degradation of organic pesticides in soil and surface waters. Nevertheless, other physic-chemical factors such as light, volatility, pH, temperature, soil moisture and organic carbon content, adsorption to sediment particles and lixiviation can influence the degradation of a pesticide and the exposure pathway [2].

In addition, pesticides may be toxic to non-target organisms that may range from beneficial soil bacteria, to insects, plants, fish, or birds. Insecticides are generally the most acutely toxic class of pesticides, but herbicides may also pose risks to non-target organisms. The excessive and frequent application of pesticides may also result in high level of pesticides residues accumulated on vegetables, which poses a potential health risk to consumers [3]. On the one hand, the human hazard is determined by the pesticide properties, exposure time and the individual's susceptibility, affecting the magnitude of these processes and the final fate of pesticide. On the other hand, the toxicity of pesticides for living organisms depends on their adsorption, distribution, biotransformation, distribution of metabolites or the interaction with cellular macromolecules and excretion [4]. Indeed, biotransformation may result in the formation of less toxic and/or more toxic metabolites, while the various other processes determine the balance between toxic and a nontoxic upcoming [5].

Human exposure may be dietary recreational and/or occupational, and toxicity might be acute or chronic [6], i.e., to ensure the safety of the food supply for human consumption, *Maximum Contaminant Levels* (*MCLs*) set the legal limits for the amount of pesticides allowed in food and drinking water, which is correlated with *Acceptable Daily Intake* (*ADI*), defined as the amount of a chemical that can be intake on a daily base [7-10]. More recently some works established pesticide impact and toxicity based on chemical properties, environmental fate and exposure considerations [8-10].

Therefore, this work will focus on the development of a decision support system to assess the environmental and toxicological risk to pesticide exposure, built on top of a Logic Programming approach to Knowledge Representation and Reasoning, complemented with a Case Based attitude 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.

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* [11, 12] and *Proof Theory* [13, 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:

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

where the first clause stand for predicate's closure, "," denotes "*logical and*", "*?*" is a domain atom denoting "*falsity*", "::" stands for "*where*", the p_i, q_j, and p (e.g., consider *tera* $(((0.02, 0.02)(1, 1)), \ldots, ((1, 1)(1, 1)))$ :: 1 :: 0.85 as it is depicted at the end of section 4) are "*classical ground literals*", i.e., either positive atoms or atoms preceded by the classical negation sign "¬" [13]. Indeed, "¬" 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 symbols' theory, every program is associated with a set of "*abducibles*" [11, 12], given here in the form of exceptions to the extensions of the predicates that make the program, i.e., clauses of the form:

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

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

In order to set one's approach to knowledge representation and reasoning, two metrics will be set, namely the Quality-of-Information (*QoI*) of a logic program that will be understood as a mathematical function that will return a truth-value ranging between 0 and 1 [15, 16], once it is fed with the extension of a given predicate. 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 $QoI_i \in]0, 1[$, in the form:

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

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 Figure 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 Figure 1(b), as a pie chart.

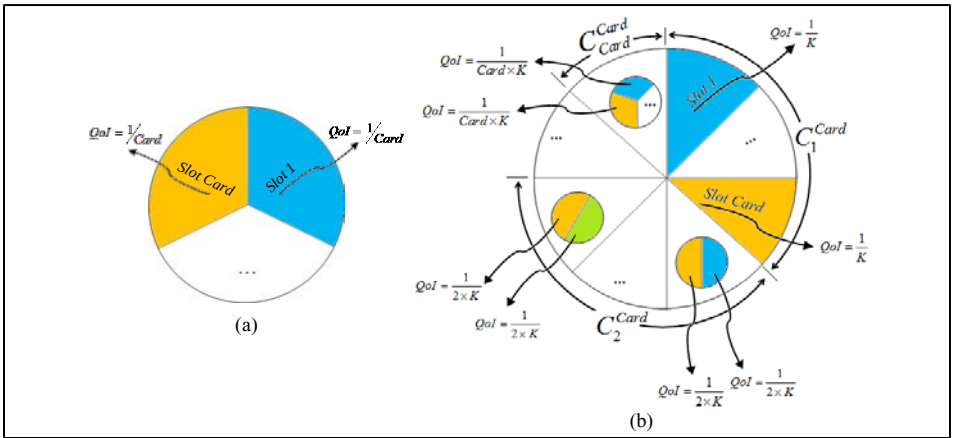

Figure 1. *QoI's* values for the abducible set for $predicate_i$ given in terms of an hatched area (a), i.e., considering only the abducible clauses, or with a circle (b), i.e., considering the possible combinations among the abducible clauses.

However, a term's *QoI* also depends on their attribute's QoI_s. In order to evaluate this metric, look to Figure 2, where the segment with bounds 0 and 1 stands for every attribute domain, i.e., all the attributes range in the interval [0, 1]. [*A*, *B*] denotes the range where the unknown attributes values for a given predicate may occur (Figure 2):

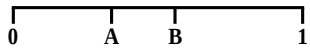

Figure 2. Setting the QoI_s for each attribute's clause.

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

where ||A–B|| stands for the modulus of the arithmetic difference between *A* and *B*, i.e., taking the absolute value. It must be also stated that unsharp (e.g., *fuzzy* or *probabilistic*) or linguistic attribute values (e.g., *good*, *bad*, …) may be transferred into an arithmetic difference as it is shown below (Figure 3 and subsection 2.2). Indeed, this generalized conception of observable enables a consistent notion of *unsharp reality* and with it an adequate concept of *joint properties*.

Under this setting, another 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 (which were set to the interval [0, 1], are in a given interval [17]. Therefore, the *DoC* is figured as $DoC = \sqrt{1 - \Delta l^2}$, where Δl stands for $\|A - B\|$ (Figure 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:

$$predicate_i \leftarrow \bigcup_{1 \le j \le m} clause_j\left(\left((A_{x_1}, B_{x_1})(QoI_{x_1}, DoC_{x_1})\right), \cdots, \left((A_{x_l}, B_{x_l})(QoI_{x_l}, DoC_{x_l})\right)\right)$$

$$:: QoI_j :: DoC_j \quad (4)$$

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]. On the other hand, either the subscripts of the QoI_s and the DoC_s, or those of the pairs (A_s, B_s), i.e., x_1, …, x_l, stand for the attributes' clauses values ranges.

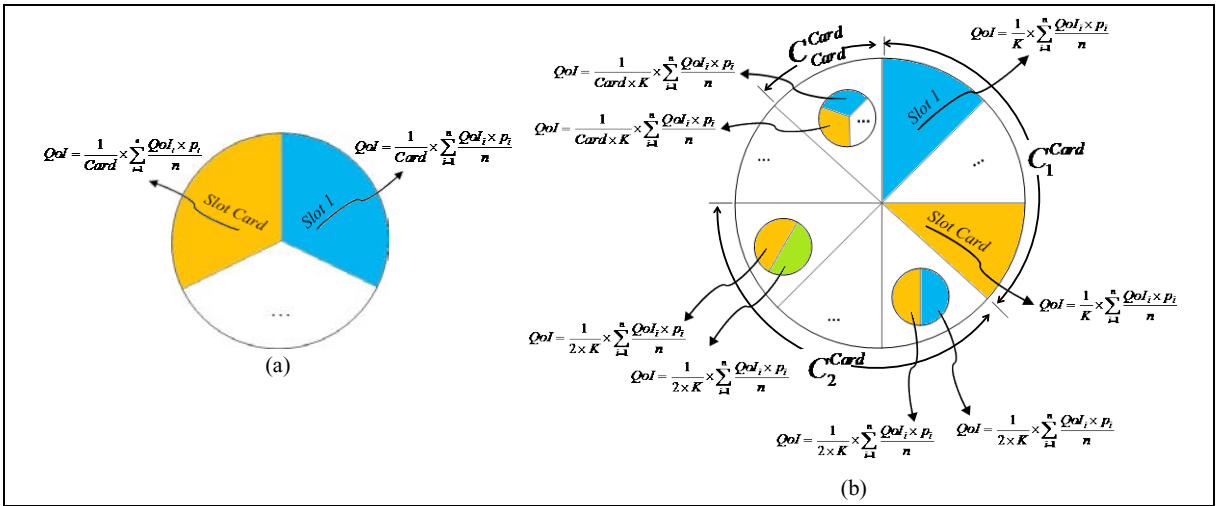

Figure 3. *QoI's* values for the abducible set for $predicate_i$ given in terms of an hatched area (a) considering only the abducible clauses, or with a circle (b), considering the possible combinations among the abducible clauses. $\sum_{i=1}^{n}(QoI_i \times p_i)/n$ denotes the *QoI's* average of the attributes of each clause 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 ($\sum_{i=1}^{n} p_i = 1$).

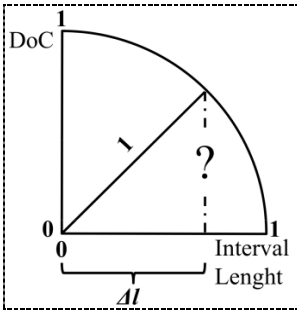

Figure 4. Evaluation of the Degree of Confidence.

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 will be presented in a graphical form. Taking as an example, consider a set of n issues regarding a particular subject, where the values of the k criteria, understood as linguistic ones, are *none*, *low*, …, *high* and *very high*. Now, enumerating a unitary area circle split into n slices (Figure 5), the marks in the axis resemble each of the possible criteria` values. If the answer to issue 1 is *high* the corresponding area is $\pi \times \left(\sqrt{(k-1)/k \times \pi}\right)^2/n$, i.e., $(k-1)/(k \times n)$ (Figure 5(a)). Assuming that in the issue 2 are chosen the alternatives *high* and *very high*, the resultant area ranges in the interval $\left[\pi \times \left(\sqrt{(k-1)/k \times \pi}\right)^2/n,\ \pi \times \left(\sqrt{k/k \times \pi}\right)^2/n\right]$, i.e., $\left[(k-1)/(k \times n),\ k/(k \times n)\right]$ (Figure 5(b)). Finally, in issue n if no alternative is ticked, all the hypotheses should be considered and the area varies in the interval $\left[0,\ \pi \times \left(\sqrt{k/k \times \pi}\right)^2/n\right]$, i.e., $\left[0,\ k/(k \times n)\right]$ (Figure 5(c)). Thus, the total area is the sum of the partial ones (Figure 5(d)), i.e., $\left[(2k-2)/(k \times n),\ (3k-1)/(k \times n)\right]$.

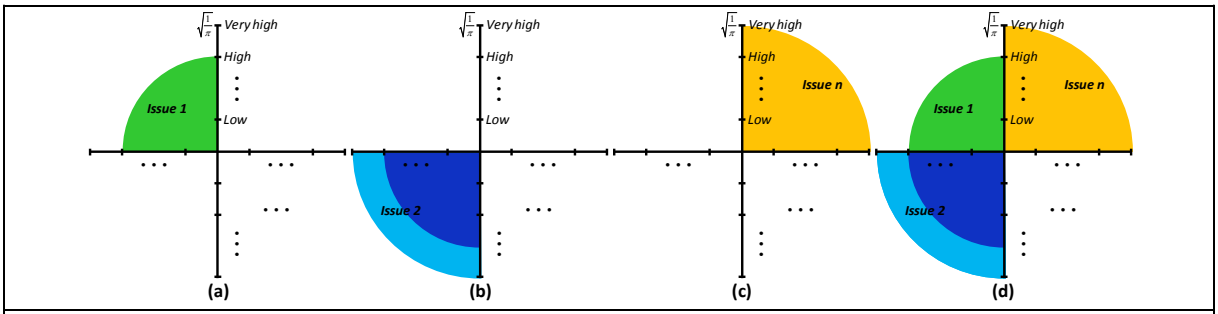

Figure 5. A view of a qualitative data/knowledge processing.

3. Case Based Reasoning

The *Case Based* (*CB*) approach to computing stands for an act of finding and justifying a solution to a given problem based on solutions of past ones [18, 19]. 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 (see section 5). The typical *CB* cycle presents the mechanism that must be followed, where the former stage entails an initial description of the problem. The new case is used to retrieve one or more cases from the *Case Base* [17], a process that will be explained below.

Undeniably, despite promising results, the current *CB* systems are neither complete nor adaptable for all domains. Moreover, in real problems, the access to the 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 (Figure 6). It takes into consideration the case's *QoI* and *DoC* metrics. It also contemplates a cases optimization process in the *Case Base*, whenever the retrieved cases 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* [20, 21], *Particle Swarm Optimization* [22] or *Genetic Algorithms* [14], just to name a few, generates a set of new cases from the retrieved ones which must be in conformity with the invariant:

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

i.e., it denotes that the intersection of the attribute's values ranges for the 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 (Figure 6), where the cases in the *Case Base* follow the pattern $Case = \{< Raw_{data}, Normalized_{data} >\}$.

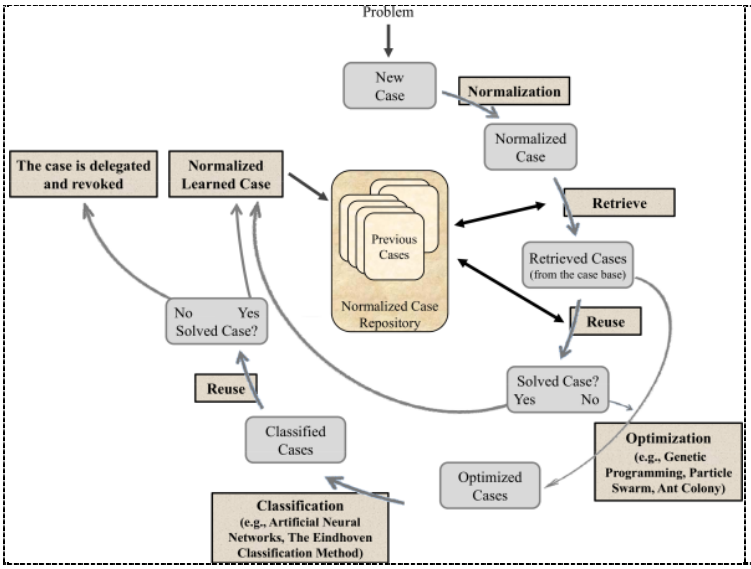

Figure 6. The CB cycle updated's view.

4. Methods

Aiming to develop a predictive model to assess the pesticides environmental and toxicological risk a knowledge database was set, and built around the pesticides records of the *National Pesticide Information Center* [23]. For each pesticide it was considered information regarding environmental fate, human exposure and toxicity (i.e., acute and chronic) both in qualitative and quantitative terms.

It is now possible to build up a knowledge database given in terms of the extensions of the relations (or tables) depicted in Figure 7, which denote a situation where one has to manage information in order to evaluate the *Pesticide Environmental and Toxicological Risk*. Under this scenario some incomplete and/or default data is present. For instance, in the former case the *ADI* is unknown (depicted by the symbol $\perp$), while the *Acute Toxicity* for *Mice/Rats* is not conclusive (*Slightly/Moderate*). The *Human Exposure* table is populated with 0 (zero) that stands for absence, 1 (one) that denotes only food or drinking water (in the *Dietary Exposure* (*DE*) column), and only dermal or inhalation exposure (in the *Occupational Exposure* (*OE*) column), and 2 (two) that stands for simultaneous exposition. The issues presented in *Environmental Fate* (*EF*) table are populated with *absence*, *low, medium*, *high* and *very high*, while the issues present in columns of *Acute Toxicity* (*AT*) and *Chronic Toxicity* (*CT*) tables are qualified as *absence*, *slightly*, *medium*, *high* and *very high*. In order to measure the information present in these tables the procedures already described in section 2.2 were followed.

Applying the algorithm presented in [17] to the table or relation's fields that make the knowledge base for *Pesticide Environmental and Toxicological Risk Assessment* (Figure 7), and looking to the DoC_s values obtained as described in [17], it is possible to set the arguments of the predicate ***t**oxicological and **e**nvironmental **r**isk **a**ssessment* (***tera***) referred to below, whose extension also denotes the objective function with respect to the problem under analyze:

$$tera\colon A_{cceptable}D_{aily}I_{ntake},\ M_{aximum}C_{ontamination}L_{evel},E_{nvironmental}F_{ate},A_{cute}$$
$$T_{oxicity},C_{hronic}T_{oxicity},\ D_{ietary}E_{xposure},\ O_{ccupational}E_{xposure} \rightarrow \{0,1\}$$

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

Indeed, in this work it is presented a construction method for interval-valued incomplete preference relations (or predicates) from an incomplete preference relation (or predicate) and the representation of the lack of knowledge or ignorance that experts suffer when they define the membership values of the elements of such incomplete preference relation (or predicate). We also prove that, with this construction method, we obtain membership intervals for an element which domain's length is equal to the ignorance associated with that element. Undeniably, the algorithm presented in [17] encompasses these 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.

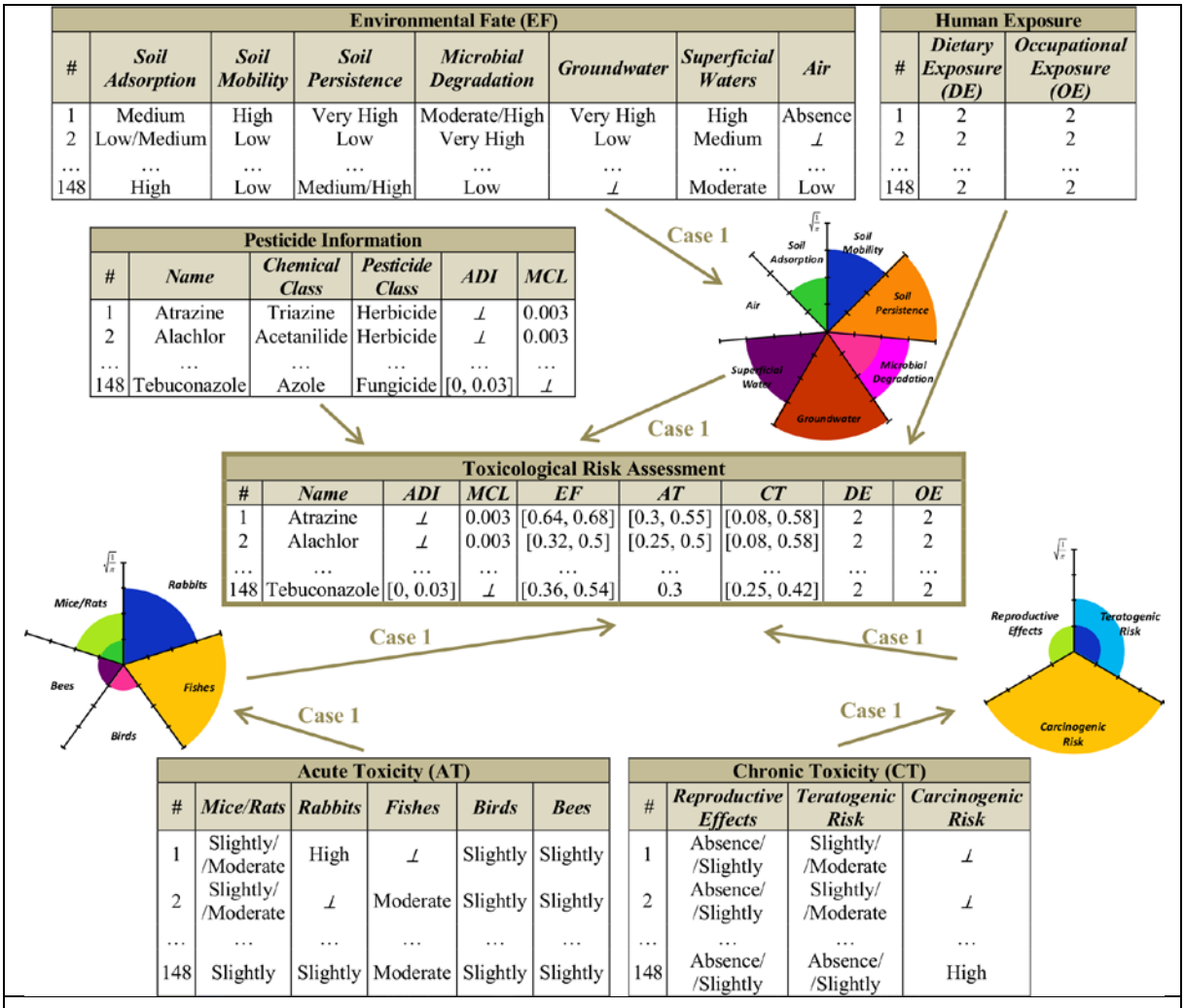

Figure 7. A fragment of the knowledge base for *Environmental and Toxicological Risk Assessment*.

Exemplifying the application of the algorithm presented in [17] in relation to a fictitious term (once a real pesticide assessment is per se not easy) that presents feature vector *ADI* = 0.01, *MCL* = ⊥, *EF* = [0.39, 0.68], *AT*= [0.7, 0.85], *CT* = 0.17, *DE* = 2, *OE* = 2, one may have:

$$\{ \quad \neg\, tera\ \Big(\big((A_{ADI}, B_{ADI})(QoI_{ADI}, DoC_{ADI})\big),\ \cdots,\ \big((A_{OE}, B_{OE})(QoI_{OE}, DoC_{OE})\big)\Big)$$

$$\leftarrow not\ tera\ \Big(\big((A_{ADI}, B_{ADI})(QoI_{ADI}, DoC_{ADI})\big),\ \cdots,\ \big((A_{OE}, B_{OE})(QoI_{OE}, DoC_{OE})\big)\Big)$$

$$tera\underbrace{\Big(\big((0.02,\ 0.02)(1,\ 1)\big),\ \cdots,\ \big((1,\ 1)(1,\ 1)\big)\Big)}_{\substack{attribute`s\ values\ ranges\ once\ normalized\ and \\ respective\ QoI\ and\ DoC\ values}} :: 1 :: 0.85$$

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

5. A Case Based approach to Computing

In this section is set the model of the universe of discourse, where the computational part is based on a *CBR* 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 [24]. Undeniably, in almost all the situations the work is performed at query time. The main difference between this 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], a situation that is complemented with the prospect of handling incomplete, unknown, or even self-contradictory data or knowledge (Figure 6).

When confronted with a new case, the system is able to retrieve all cases that meet such a structure and optimize, when necessary, such a population, i.e., it considers the attributes *DoC*'s value of each case or of their optimized counterparts when analyzing 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 algorithm described above is applied to a new case, that presents the feature vector (*ADI* = ⊥, *MCL* = 0.003, *EF* = 0.71, *AT*= [0.65, 0.8], *CT* = [0.42, 0.58], *DE* = 1, *OE* = 2), having in consideration that the cases retrieved from the *Case Base* satisfy the invariant:

$$\bigcap_{i=1}^{n}\left(\boldsymbol{B_i}\,,\ \boldsymbol{E_i}\right) \neq \emptyset \tag{6}$$

denoting that the intersection of the attributes range in the cases that make the *Case Base* repository or their optimized counterparts ($\boldsymbol{B_i}$), and the equals in the new case ($\boldsymbol{E_i}$), cannot be empty. Then, the computational process may be continued, with the outcome (once applying the algorithm presented in [17]):

$$tera_{new}\underbrace{\Big(\big((0,\ 1)(1,\ 0)\big),\ \cdots,\ \big((1,\ 1)(1,\ 1)\big)\Big)}_{new\ case} :: 1 :: 0.85$$

Now, the *new case* may be portrayed on the *Cartesian* plane in terms of its *QoI* and *DoC*, and by using clustering methods [25] it is feasible to identify the cluster(s) that intermingle with the *new one* (epitomized as a square in Figure 8). The *new case* is compared with every retrieved case from the clusters 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,* which is crucial when different clustering methods are examined. Thus, one may have:

$$\underbrace{\begin{array}{l} tera_1\left(((0.04,\ 0.04)(1,\ 1)),\ \cdots,\ ((1,\ 1)(1,\ 1))\right) :: 1 :: 0.99 \\ tera_2\left(((0.06,\ 0.06)(1,\ 1)),\ \cdots,\ ((0.5,\ 0.5)(1,\ 1))\right) :: 1 :: 0.84 \\ \vdots \\ tera_j\left(((0.03,\ 0.03)(1,\ 1)),\ \cdots,\ ((1,\ 1)(1,\ 1))\right) :: 1 :: 0.85 \end{array}}_{normalized\ cases\ from\ retrieved\ cluster}$$

Assuming that every attribute has equal weight, for the sake of presentation, the dissimilarity between $tera_{new}$ and the $tera_1$, i.e., $tera_{new \to 1}$, may be computed as follows:

$$tera_{new\to1}^{DoC} = \frac{\|0-1\| + \cdots + \|0.99-0.95\| + \|0.99-1\| + \|1-1\| + \|1-1\|}{7} = 0.15$$

Thus, the similarity for $tera_{new\to1}^{DoC}$ is set as *1 – 0.15 = 0.85.* Regarding *QoI* the procedure is similar, returning $tera_{new\to1}^{QoI} = 1$. Thus, one may have:

$$tera_{new\to1}^{QoI,\,DoC} = 1 \times 0.85 = 0.85$$

These procedures should be applied to the remaining cases of the retrieved clusters 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 [21]. In the implementation of the respective dividing procedures, ten accomplishments 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 30 (thirty) experiments. A perusal to Table 1 shows that the model accuracy was 87.8% (i.e., 130 instances correctly classified in 148 cases). Thus, the predictions made by the *CB* model are satisfactory, attaining accuracy close to 90%. The sensitivity and specificity of the model were 91.7% and 80.8%, while *Positive* and *Negative Predictive Values* were 89.8% and 84.0%, respectively. Thus, the proposed model correctly identified 89.8% of the positive cases, i.e., pesticides with high *potential environmental* or *toxicological risk.* Moreover, it also classified appropriately 84.0% of the negative cases, i.e., pesticides with *low environmental and toxicological risk.*

The present model, beyond to consider the pesticide chemical properties, enables the integration of *Acute* and *Chronic Toxicity* data with other factors such as *Environmental Fate* and *Human Exposure*, being therefore assertive in the prediction of *Pesticide Toxicological Risk.* Thus, it is our claim that the proposed model is able to evaluate the *Environmental and Toxicological Risk* of each pesticide properly, and may be a major contribution to achieve high standards concerning environmental sustainability and public health protection.

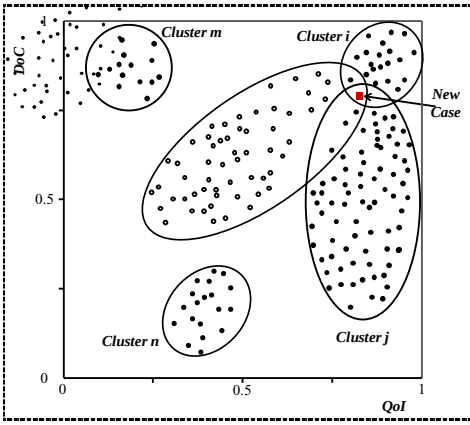

Figure 8. A case's set clusters.

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

	Predictive	
Target	True (1)	False (0)
True (1)	88	8
False (0)	10	42

6. Conclusions

The proposed methodology for problem solving is able to give an adequate response to predict the environmental and toxicological risk of pesticide exposure. Nevertheless, it can be considered a hard task since it is necessary to consider different variables and/or conditions with complex relations entwined among them, where the data may be incomplete, self-contradictory, and even unknown. In order to overcome these difficulties the methodology followed in this work 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, susceptive of application in different arenas. Furthermore, 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). A possible limitation on its use is not on the model in itself, but on the unavailability of data, information or knowledge but, even in these situations, once it has the capacity to handle incomplete data, information or knowledge, either in its qualitative or quantitative form, its usefulness is assured. Future developments of the model should include the biotransformation pathways and routes of exposure, and consider the contact time and the individual's susceptibility.

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] Akatar W R, Sengupta D and Chowdhury A 2009 *Interdiscip. Toxicol.* **2** 1–12
[2] Popp J, Pető K and Nagy J 2013 *Agron. Sustain. Dev.* **33** 243–255
[3] Verma J P, Jaiswal D K and Sagar R 2014 *Rev. Environ. Sci Biotechnol.* **13** 429–466
[4] Hodgson E 2012 Introduction to Pesticide Biotransformation and Disposition *Pesticide Biotransformation and Disposition* ed E Hodgson (Amsterdam: Elsevier) chapter 1 pp 1–3
[5] Needham L L, Patterson D G, Barr D B, Grainger J. and Calafat A M 2005 *Anal. Bioanal. Chem.* **381** 397–404
[6] Environmental Protection Agency 2001 *General Principles for Performing Aggregate Exposure and Risk Assessments* (Washington: Office of Pesticide Programs)
[7] Renwick A G 2002 *Pest Manag. Sci.* **58** 1073–82
[8] Antón A, Castells F, Montero J I and Huijbregts M 2004 *Chemosphere* **54** 1225–35
[9] Alister C and Kogan M 2006 *Crop Prot.* **25** 202–211
[10] Juraske R, Antón A, Castells F and Huijbregts M A 2007 *Environ. Int.* **33** 886–893
[11] Kakas A, Kowalski R and Toni F 1998 *Handbook of Logic in Artificial Intelligence and Logic Programming* vol 5, ed D Gabbay *et al* (Oxford: Oxford University Press) pp 235–324
[12] Pereira L and Anh H 2009 *Studies in Computational Intelligence* vol 199, ed K Nakamatsu (Berlin: Springer) pp 51–64
[13] Neves J 1984 *Proc. 1984 Annual Conf. of the ACM on the 5th Generation Challeng* ed R Muller and J Pottmyer (New York: Association for Computing Machinery) pp 50–54

[14] Neves J, Machado J, Analide C, Abelha A and Brito L 2007 *Lecture Notes in Artificial Intelligence* vol 4874, ed J Neves *et al* (Berlin: Springer) pp 160–169
[15] Lucas P 2003 *Proc. of AI-2003 (Research and Developments in Intelligent Systems XX)* ed F Coenen *et al* (London: Springer) pp 309–321
[16] Machado J, Abelha A, Novais P, Neves J and Neves J 2008 *Proc. of the ESM 2008* ed C Bertelle and A Ayesh (Ghent: Eurosis) pp 291–298
[17] Fernandes F, Vicente H, Abelha A, Machado J, Novais P and Neves J 2015 *Proc. of the 2015 Science and Information Conf.* pp 362–370
[18] Aamodt A and Plaza E 1994 *AI Commun.* **7** 39–59
[19] Richter M M and Weber R O 2013 *Case-Based Reasoning: A Textbook* (Berlin: Springer)
[20] Vicente H, Dias S, Fernandes A, Abelha A, Machado J and Neves J 2012 *J. Water Supply Res. T.* **61** 446–459
[21] Haykin S 2009 *Neural Networks and Learning Machines* (New Jersey: Pearson Education)
[22] Mendes R, Kennedy J and Neves J 2003 *Proc. of the 2003 IEEE Swarm Intelligence Symp.* pp 88–94
[23] National Pesticide Information Center http://npic.orst.edu/index.html
[24] Carneiro D, Novais P, Andrade F, Zeleznikow J and Neves J 2013 *Knowl. Inf. Syst.* **36** 789–826
[25] Figueiredo M, Esteves L, Neves J and Vicente, H. 2016 *Chem. Educ. Res. Pract.* **17** 156–171