An Assessment of Pharmacological Properties of *Schinus* Essential Oils A Soft Computing Approach

José Neves
Algoritmi
Universidade do Minho
Braga, Portugal
jneves@di.uminho.pt

M. Rosário Martins
Departamento de Química
Escola de Ciências e Tecnologia
Laboratório HERCULES
Universidade de Évora, Évora, Portugal
mrm@uevora.pt

Fátima Candeias, Sílvia Arantes
Departamento de Química
Escola de Ciências e Tecnologia
Instituto de Ciências Agrárias e Ambientais Mediterrânicas
Universidade de Évora, Évora, Portugal
{mfbc, saa}@uevora.pt

Ana Piteira
Departamento de Química
Escola de Ciências e Tecnologia
Universidade de Évora, Évora, Portugal
anaisabelanaisabel14@hotmail.com

Henrique Vicente
Departamento de Química
Escola de Ciências e Tecnologia
Universidade de Évora, Évora, Portugal
Algoritmi, Universidade do Minho
hvicente@uevora.pt

KEYWORDS
Schinus spp., Essential Oils, Logic Programming, Case Base Reasoning, Knowledge Representation and Reasoning, Similarity Analysis.

ABSTRACT

Plants of genus *Schinus* are native South America and introduced in Mediterranean countries, a long time ago. Some *Schinus* species have been used in folk medicine, and *Essential Oils* of *Schinus* spp. (*EOs*) have been reported as having antimicrobial, anti-tumoural and anti-inflammatory properties. Such assets are related with the *EOs* chemical composition that depends largely on the species, the geographic and climatic region, and on the part of the plants used. Considering the difficulty to infer the pharmacological properties of *EOs* of *Schinus* species without a hard experimental setting, this work will focus on the development of an Artificial Intelligence grounded Decision Support System to predict pharmacological properties of *Schinus EOs*. The computational framework was built on top of a *Logic Programming Case Base* approach to knowledge representation and reasoning, which caters to the handling of incomplete, unknown, or even self-contradictory information. New clustering methods centered on an analysis of attribute's similarities were used to distinguish and aggregate historical data according to the context under which it was added to the Case Base, therefore enhancing the prediction process.

INTRODUCTION

Schinus L. species are trees from Anacardiaceae family characterized by pungent-smell of essential oils of their leaves and fruits. Plants of genus *Schinus* are native to South America, including approximately 29 species, and some of them have been introduced to southern Europe, including Portugal, as an ornamental plant (Bendaoud et al. 2010).

Some *Schinus* species, namely *S. molle* L., *S. terebinthifolius* Raddi and *S. longifolius* (Lindl.) Speg. are used in folk medicine to treat pathologies like rheumatism, high blood pressure, respiratory and urinary infections, or as digestive, diuretic and purgative (Duke, 2002; Atti dos Santos et al. 2010; Murray et. al 2012).

The chemical characterization of *EOs* of leaves and berries of *Schinus* spp. have been reported with the presence of different monoterpenes, sesquiterpenes and triterpenes, as secondary metabolites. However, the chemical composition of *EOs* is different according to the geographic and seasonal factors and the part of the plant used for extraction, fruit or leaves (Díaz et al. 2008; El-Massry et al. 2009; Gomes et al. 2013; Martins et al. 2014).

Some studies highlighted several biological properties of *EOs*, namely antimicrobial (El-Massry et al. 2009; Deveci et al. 2010; Martins et al. 2014), antioxidant (Díaz et al. 2008; Bendaoud et al. 2010; Martins et al. 2014), anti-tumoural (Díaz et al. 2008; Bendaoud et al.

Proceedings 30th European Conference on Modelling and Simulation ©ECMS Thorsten Claus, Frank Herrmann, Michael Manitz, Oliver Rose (Editors)
ISBN: 978-0-9932440-2-5 / ISBN: 978-0-9932440-3-2 (CD)

2010), analgesic and anti-inflammatory activities (Simionatto et al. 2011; Bigliani et al. 2012), and correlated them with their biochemical structure.

Taking into account the geographical and seasonal variability of chemical composition of *Schinus EOs* and the difficulty to infer their pharmacological properties without experimental assays for each *EO*, this paper describes an intelligent support system to predict pharmacological properties of *Schinus* essential oils using a *Case Based Reasoning* (*CBR*) approach to problem solving (Aamodt and Plaza 1994; Richter and Weber 2013). To set the structure of the information and the associate inference mechanisms, a computational framework centered on a *Logic Programming* (*LP*) based approach to knowledge representation and reasoning was used. It caters to the handling of unknown, incomplete, forbidden, or even self-contradictory data, information or knowledge.

KNOWLEDGE REPRESENTATION AND REASONING

At decision times the available information is not always exact in the sense that it can be estimated values, probabilistic measures, or degrees of uncertainty. Furthermore, knowledge and belief are generally incomplete, self-contradictory, or even error sensitive, being desirable to use formal tools to deal with the problems that arise from the use of those types of data, information, or knowledge (Neves 1984; Neves et al. 2007). *Logic Programming* (*LP*) has been used for knowledge representation and reasoning in different areas, like Model Theory (Kakas et al. 1998; Pereira and Anh 2009), and Proof Theory (Neves 1984; Neves et al. 2007). In the present work the proof theoretical approach is followed in terms of an extension to *LP*. An *Extended Logic Program* is a finite set of clauses in the form:

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

where "*?*" 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 $\neg$ (Neves 1984). Under this formalism, every program is associated with a set of abducibles (Kakas et al. 1998; Pereira and Anh 2009) given here in the form of exceptions to the extensions of the predicates that make the program. The term $scoring_{value}$ stands for the relative weight of the extension of a specific *predicate* with respect to the extensions of the peers ones that make the overall program.

In order to evaluate the knowledge that can be associated to a logic program, an assessment of the *Quality-of-Information* (*QoI*), given by a truth-value in the interval [0, 1], that stems from the extensions of the predicates that make a program, inclusive in dynamic environments (Lucas 2003; Machado et al. 2008), was set. Indeed, the objective is to build a quantification process of *QoI* and *DoC* (*Degree of Confidence*), the latter being a measure of one's confidence that the argument values or attributes of the terms that make the extension of a given predicate, with relation to their domains, fit into a given interval (Fernandes et al. 2015). The *DoC* is evaluated as depicted in Figure 1 and computed using $DoC = \sqrt{1 - \Delta l^2}$, where Δl stands for the argument interval (set to the interval [0, 1]). Thus, the universe of discourse is engendered according to the information presented in the extensions of such predicates, according to productions of the type:

$$
\begin{aligned}
predicate_i - \bigcup_{1 \leq i \leq m} clause_j\big((QoI_{x_m}, DoC_{x_m}), \cdots\\
\cdots, (QoI_{x_m}, DoC_{x_m})\big) :: QoI_i :: DoC_i
\end{aligned}
$$

where $\cup$ and m stand, respectively, for set union and the cardinality of the extension of $predicate_i$. QoI_i and DoC_i stands for themselves.

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

$$
\begin{aligned}
&\{\\
&\quad \neg f_1\big((QoI_{x_1}, DoC_{x_1}), (QoI_{y_1}, DoC_{y_1}), (QoI_{z_1}, DoC_{z_1})\big)\\
&\qquad \leftarrow not\big((QoI_{x_1}, DoC_{x_1}), (QoI_{y_1}, DoC_{y_1}), (QoI_{z_1}, DoC_{z_1})\big)\\
&\quad f_1\underbrace{\big((QoI_{[10,\,15]}, DoC_{[10,\,15]}), (QoI_{\perp}, DoC_{\perp}), (QoI_{2.5}, DoC_{2.5})\big)}_{attribute`s\ values}\\
&\qquad :: QoI :: DoC\\
&\qquad \underbrace{[0, 20] \qquad [0, 9] \qquad [5, 20]}_{attribute`s\ domains}\\
&\quad exception_{f_1}\big((QoI_3, DoC_3), (QoI_{[4,\,6]}, DoC_{[4,\,6]}), (QoI_{\perp}, DoC_{\perp})\big)\\
&\qquad :: QoI :: DoC\\
&\quad \ldots\\
&\quad exception_{f_k}\big((QoI_{\perp}, DoC_{\perp}), (QoI_5, DoC_5), (QoI_{[6,\,9]}, DoC_{[6,\,9]})\big)\\
&\qquad :: QoI :: DoC\\
&\}:: 1 \qquad (once\ the\ universe\ of\ discourse\ is\ set\ in\ terms\ of\ the\\
&\qquad\qquad extension\ of\ only\ one\ predicate)
\end{aligned}
$$

where $\perp$ denotes a null value of the type unknown. It is now possible to split the abducible or exception set into the admissible clauses or terms and evaluate their QoI_i. A pictorial view of this process is given in Figure 2, as a pie chart.

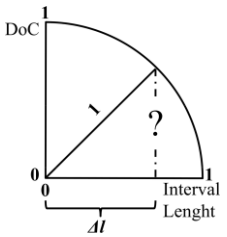

Figure 1: Evaluation of the Degree of Confidence

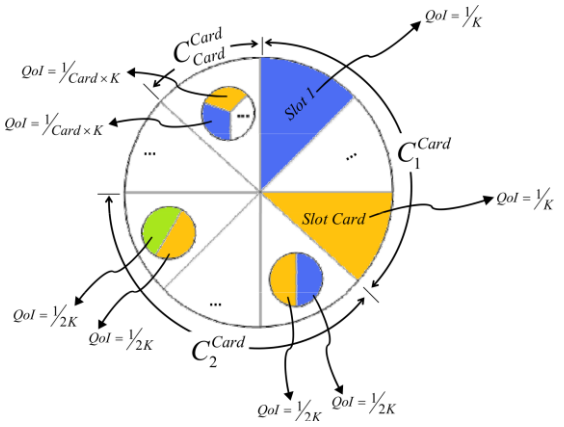

Figure 2: *QoI's* values for the abducible set of clauses referred to above, where the clauses cardinality set, *K*, is given by the expression $C_1^{Card} + C_2^{Card} + \cdots + C_{Card}^{Card}$

CASE BASED REASONING

The *CBR* methodology for problem solving stands for an act of finding and justifying the solution to a given problem based on the consideration of similar past ones, by reprocessing and/or adapting their data or knowledge (Aamodt and Plaza 1994; Richter and Weber 2013). In *CBR* – the cases – are stored in a *Case Base*, and those cases that are similar (or close) to a new one are used in the problem solving process. There are examples of its use in *The Law* with respect to *Dispute Resolution* (Carneiro et al. 2013), in *Medicine* (Janssen et al. 2014; Ying et al. 2015), among others. The typical *CBR* cycle presents the mechanism that should be followed to have a consistent model. The first stage consists of the initial description of the problem. The new case is defined and it is used to retrieve one or more cases from the repository. At this point it is important to identify the characteristics of the new problem and retrieve cases with a higher degree of similarity to it. Thereafter, a solution for the problem emerges, on the *Reuse* phase, based on the blend of the new case with the retrieved ones. The suggested solution is reused (i.e., adapted to the new case), and a solution is provided (Aamodt and Plaza 1994; Richter and Weber 2013). However, when adapting the solution it is crucial to have feedback from the user, since automatic adaptation in existing systems is almost impossible. This is the *Revise* stage, in which the suggested solution is tested by the user, allowing for its correction, adaptation and/or modification, originating the test-repaired case that sets the solution to the new problem. The test-repaired case must be correctly tested to ensure that the solution is indeed correct. Thus, one is faced with an iterative process since the solution must be tested and adapted while the result of applying that solution is inconclusive. During the *Retain* (or *Learning*) stage the case is learned and the knowledge base is updated with the new case (Aamodt and Plaza 1994; Richter and Weber 2013).

Despite promising results, the current *CBR* systems do not cover all areas, and 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 (Richter and Weber 2013; Neves and Vicente n.d.). But, worse than that, in real problems, access to all necessary information is not always possible, since existent *CBR* systems have limitations related to the capability of dealing, explicitly, with unknown, incomplete, and even contradictory information. To make a change, a different *CBR* cycle was induced (Figure 3). It takes into consideration the case's *QoI* and *DoC* (Neves and Vicente n.d.). It deals not only with unknown, incomplete, forbidden, and even self-contradictory data, information or knowledge, in an explicit way, but also contemplates the cases optimization 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 degree of confidence on the diagnostic was not attained), either using particle swarm optimization procedures (Mendes et al 2003), or genetic algorithms (Neves et al 2007), just to name a few.

METHODS

The data set was obtained based on experimental researches with EOs of leaf and fruit of *Schinus molle* collected in Alentejo (Martins et al. 2014) and *S. molle*, *S. terebinthifolius* and *S. longifolius* collected in Brazil and Argentina (Atti dos Santos et al. 2010; Gomes et al. 2013; Murray et. al 2012).

The knowledge database is specified in terms of the extensions of the relations depicted in Figure 4, which denotes a situation where one has to manage information aiming to evaluate the pharmacological properties of *Schinus* essential oils. Under this scenario some incomplete and/or unknown data is also available. For instance, in the former case, the data regarding antioxidant tests are unknown, as depicted by the symbol $\perp$, while the percentage of monoterpenes hydrocarbons ranges in the interval [68, 72]. The *Plant Part* column ranges in the interval [0, 1], wherein 0 (zero), and 1 (one) denote, respectively, *leaves* and *fruit*.

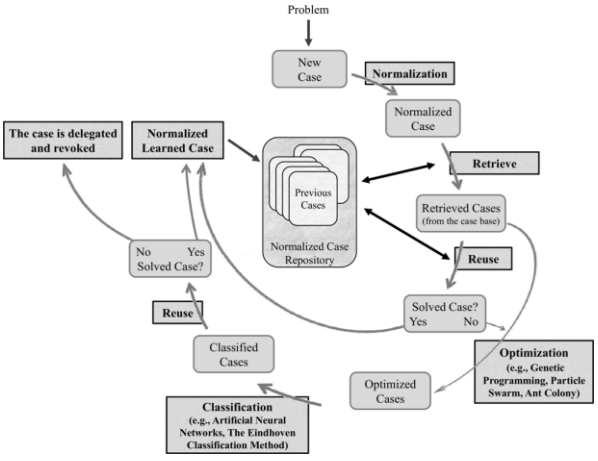

Figure 3: An extended view of the CBR cycle

	Pharmacological Activity Screening Assessment										
Attributes of the Feature Vector:	#	Plant Part	Monoterpenes Hydrocarbons	Oxygenated Monoterpenes	Sesquiterpenes Hydrocarbons	Oxygenated Sesquiterpenes	CL_{50}	DL_{50}	Hippocratic Screening	Antioxidant Activity	Description
Feature Vector Attributes:	1	0	[68, 72]	[0.5, 2]	[2, 4]	[13, 15]	48	2000	1	⊿	Description 1
	2	0	[63, 65]	[0.6, 0.8]	[30, 32]	[2, 3]	⊿	1500	2	2	Description 2
	…	…	…	…	…	…	…	…	…	…	…
	73	1	[91, 98]	[0, 1]	[0, 1]	[1, 3]	67	2500	3	5	Description 73
Feature Vector Domains:		[0, 1]	[0, 100]	[0, 100]	[0, 100]	[0, 100]	[25, 3000]	[100, 5000]	[0, 6]	[0, 12]	

Figure 4: A fragment of the knowledge base aiming to predict pharmacological activity of essential oils of *Schinus* species

Applying the algorithm presented in Fernandes (2015) to the fields that make the knowledge base for pharmacological activity screening assessment (Figure 4), excluding of such a process the *Description* ones, and looking to the DoC_s values obtained, it is possible to set the arguments of the predicate ***pharm**acological **act**ivity* ($pharm_{act}$) referred to below, that also denotes the objective function with respect to the problem under analysis:

$$pharm_{act} : P_{lant}P_{art}, M_{onoterpenes}H_{ydrocarbons}, M_{onoterpenes}O_{xigenated}, S_{esquiterpenes}H_{ydrocarbons}, S_{esquiterpenes}O_{xigenated}, CL_{50}, DL_{50}, H_{ippocratic} S_{creening}, A_{ntioxidant}A_{ctivity} \rightarrow \{0, 1\}$$

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

Exemplifying the application of the algorithm presented in Fernandes (2015), to a term (case) that presents feature vector (P_{lant} $P_{art} = 1$, $M_{onoterpenes}$ $H_{ydrocarbons} = [85, 93]$, $M_{onoterpenes}$ $O_{xigenated} = [1.2, 2.4]$, $S_{esquiterpenes}$ $H_{ydrocarbons} = [0.8, 3.2]$, $S_{esquiterpenes}$ $O_{xigenated} = [1.2, 4.3]$, $CL_{50} = 63$, $DL_{50} = 2400$, $H_{ippocratic}$ $S_{creening} = 2$, $A_{ntioxidant}$ $A_{ctivity} = ⊿$), and applying the procedure referred to above, one may get:

Begin %DoCs evaluation%,

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

$$\{$$
$$\neg\, pharm_{act}\left((QoI_{PP}, DoC_{PP}), (QoI_{MH}, DoC_{MH}), \cdots, (QoI_{AA}, DoC_{AA})\right) \leftarrow not\ pharm_{act}\left((QoI_{PP}, DoC_{PP}), (QoI_{MH}, DoC_{MH}), \cdots, (QoI_{AA}, DoC_{AA})\right)$$
$$pharm_{act}\ \underbrace{\left((1_1, DoC_1), (1_{[85,93]}, DoC_{[85,93]}), \cdots, (1_{\perp}, DoC_{\perp})\right)}_{attribute`s\ values} :: 1 :: DoC$$
$$\underbrace{[0,1] \quad [0,100] \quad \cdots, \quad [0,12]}_{attribute`s\ domains}$$
$$\} :: 1$$

%The attribute's values ranges are rewritten%

$$\{$$
$$\neg\, pharm_{act}\left((QoI_{PP}, DoC_{PP}), (QoI_{MH}, DoC_{MH}), \cdots, (QoI_{AA}, DoC_{AA})\right) \leftarrow not\ pharm_{act}\left((QoI_{PP}, DoC_{PP}), (QoI_{MH}, DoC_{MH}), \cdots, (QoI_{AA}, DoC_{AA})\right)$$
$$pharm_{act}\ \underbrace{\left((1_{[1,1]}, DoC_{[1,1]}), (1_{[85,93]}, DoC_{[85,93]}), \cdots, (1_{[0,12]}, DoC_{[0,12]})\right)}_{attribute`s\ values\ ranges} :: 1 :: DoC$$
$$\underbrace{[0,1] \quad [0,100] \quad \cdots, \quad [0,12]}_{attribute`s\ domains}$$
$$\} :: 1$$

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

$$\{$$
$$\neg\, pharm_{act}\left((QoI_{PP}, DoC_{PP}), (QoI_{MH}, DoC_{MH}), \cdots, (QoI_{AA}, DoC_{AA})\right) \leftarrow not\ pharm_{act}\left((QoI_{PP}, DoC_{PP}), (QoI_{MH}, DoC_{MH}), \cdots, (QoI_{AA}, DoC_{AA})\right)$$
$$pharm_{act}\ \underbrace{\left((1_{[1,\ 1]}, DoC_{[1,\ 1]}), (1_{[0.85,\ 0.93]}, DoC_{[0.85,\ 0.93]}), \cdots, (1_{[0,\ 1]}, DoC_{[0,\ 1]})\right)}_{attribute`s\ values\ ranges\ once\ normalized} :: 1 :: DoC$$
$$\underbrace{[0,1] \quad [0,1] \quad \cdots, \quad [0,1]}_{attribute`s\ domains\ once\ normalized}$$
$$\} :: 1$$

%The DoC's values are evaluated%

$$
\begin{Bmatrix}
\neg\, pharm_{act}\left((QoI_{PP}, DoC_{PP}),\ (QoI_{MH}, DoC_{MH}),\ \cdots,\ (QoI_{AA}, DoC_{AA})\right) \\
\qquad\qquad \leftarrow not\ pharm_{act}\left((QoI_{PP}, DoC_{PP}),\ (QoI_{MH}, DoC_{MH}),\ \cdots,\ (QoI_{AA}, DoC_{AA})\right) \\
pharm_{act}\underbrace{\left((1,\ 0),\ (1,\ 0.997),\ \cdots, (1,\ 0)\right)}_{\substack{attribute`s\ quality-of-information \\ and\ respective\ confidence\ values}} :: 1 :: 0.89 \\
\qquad\qquad \underbrace{[1,\ 1]\ [0.85,\ 0.93]\ \cdots, [0,1]}_{attribute`s\ values\ ranges\ once\ normalized} \\
\qquad\qquad \underbrace{[0,1] \qquad [0,1] \qquad \cdots, [0,1]}_{attribute`s\ domains\ once\ normalized}
\end{Bmatrix} :: 1
$$

End.

SOFT COMPUTING APPROACH

A soft computing approach to model the universe of discourse based on *CBR* methodology for problem solving is now set. Indeed, contrasting with other problem solving methodologies (e.g., *Decision Trees* or *Artificial Neural Networks*), in a *CBR* based methodology relatively little work is done offline. Undeniably, in almost all the situations the work is performed at query time. The main difference between this new approach and the typical *CBR* one relies on the fact that not only all the cases have their arguments set in the interval [0, 1], but it also caters for the handling of incomplete, unknown, or even self-contradictory data or knowledge (Neves and Vicente n.d.). Thus, the classic *CBR* cycle was changed (Figure 3), being the *Case Base* given in terms of triples that follow the pattern:

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

where Raw_{data} and $Normalized_{case}$ stand for themselves, and $Description_{data}$ is made on a set of strings or even in free text, which may be analyzed with string similarity algorithms. When confronted with a new case, the system is able to retrieve all cases that meet such a structure and optimize such a population, i.e., it considers the attributes *DoC*'s value of each case or of their optimized counterparts when analysing similarities among them. Thus, under the occurrence of a new case, the goal is to find similar cases in the *Case Base*. Having this in mind, the algorithm given in Fernandes (2015) is applied to a new case that presents feature vector (P_{lant} $P_{art} = 0$, $M_{onoterpenes}$ $H_{ydrocarbons} = [68, 71]$, $M_{onoterpenes}$ $O_{xigenated} = [0.8, 2.1]$, $S_{esquiterpenes}$ $H_{ydrocarbons} = [2, 5.2]$, $S_{esquiterpenes}$ $O_{xigenated} = [14, 16]$, $CL_{50} = 45$, $DL_{50} = 2200$, $H_{ippocratic}$ $S_{creening} = \perp$, $A_{ntioxidant}$ $A_{ctivity} = 3$, *Description* = *Description_new*), with the results:

$$\underbrace{pharm_{act_{new}}\left((1,1),\ (1,0.99), \cdots,\ (1,1)\right) :: 1 :: 0.88}_{new\ case}$$

The *new case* can be depicted on the *Cartesian Plane* in terms of its *QoI* and *DoC*, and through clustering techniques, it is feasible to identify the clusters that intermingle with the new one (symbolized as a square in Figure 5). The *new case* is compared with every retrieved case from the cluster using a similarity function *sim*, given in terms of the average of the modulus of the arithmetic difference between the arguments of each case of the selected cluster and those of the *new case* (once *Description* stands for free text, its analysis is excluded at this stage). Thus, one may have:

$$
\underbrace{\begin{matrix}
pharm_{act_1}\left((1,\ 1), (1,\ 0.92), \cdots, (1,\ 0)\right) :: 1 :: 0.85 \\
pharm_{act_2}\left((1,\ 1), (1,\ 0.97), \cdots, (1,\ 0)\right) :: 1 :: 0.82 \\
\vdots \\
pharm_{act_j}\left((1,\ 1), (1,\ 0.99), \cdots, (1,\ 1)\right) :: 1 :: 0.91
\end{matrix}}_{normalized\ cases\ from\ retrieved\ cluster}
$$

Assuming that every attribute has equal weight, the dissimilarity between $pharm_{act_{new}}^{DoC}$ and the $pharm_{act_1}^{DoC}$, i.e., $dissim\_pharm_{act_{new\to 1}}^{DoC}$, may be computed as follows:

$$
dissim\_pharm_{act_{new\to 1}}^{DoC} = \\
= \frac{\|1-1\| + \|0.97-0.92\| + \cdots + \|1-0\|}{9} = 0.17
$$

Thus, the similarity between $pharm_{act_{new\to 1}}^{DoC}$ ($sim\_pharm_{act_{new\to 1}}^{DoC}$) is $1 - 0.17 = 0.83$. Regarding *QoI* the procedure is similar, returning $sim\_pharm_{act_{new\to 1}}^{QoI} = 1$.

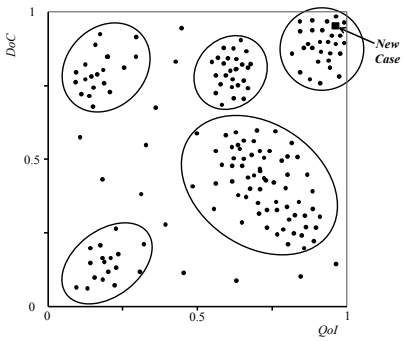

Figure 5: A case's set split into clusters

Descriptions will be compared using *String Similarity Algorithms*, in order to liken the description of the new case with the descriptions of the cases belonging to the retrieved cluster (in this study the strategy used was the *Dice Coefficient* one (Dice 1945)), with the results:

$$sim\_pharm_{act_{new \to 1}}^{Description} = 0.80$$

With these similarity values it is possible to get a global similarity measure:

$$sim\_pharm_{act_{new \to 1}} = \frac{0.83 + 1 + 0.80}{3} = 0.88$$

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

A common tool to evaluate the performance of the classification models is the coincidence matrix, i.e., a matrix of size $L \times L$, where L denotes the number of possible classes (two in the present case). This matrix is created by matching the predicted and target values. Table 1 presents the coincidence matrix (the values denote the average of the 30 experiments). It shows that the model accuracy was 87.7% (64 instances of 73 correctly classified). Based on coincidence matrix it is possible to compute the sensitivity and the specificity of the model:

$$sensitivity = TP/(TP + FN) \quad (1)$$

$$specificity = TN/(TN + FP) \quad (2)$$

where TP, FN, TN and FP stand, respectively, for true positive, false negative, true negative and false positive. Briefly, sensitivity and specificity are statistical measures of the performance of a binary classifier. Sensitivity measures the proportion of true positives that are correctly identified as such, while specificity measures the proportion of true negatives that are correctly identified. In this case both metrics show values higher than 85%, (i.e., 87.0% and 88.9% for sensitivity and specificity, respectively). In addition, the *Receiver Operating Characteristic* (*ROC*) curves were considered. An *ROC* curve displays the trade-off between sensitivity and specificity. The *Area Under the Curve* (*AUC*) quantifies the overall ability of the test to discriminate between the output classes. Figure 6 depicted the *ROC* curve for the proposed model. The area under *ROC* curve is 0.88 denoting that the model exhibits a good performance in the evaluation of pharmacological properties of *Schinus* essential oils.

Table 1: The Coincidence Matrix for the ANN Model

Target	Predictive	
	True (1)	False (0)
True (1)	40	6
False (0)	3	24

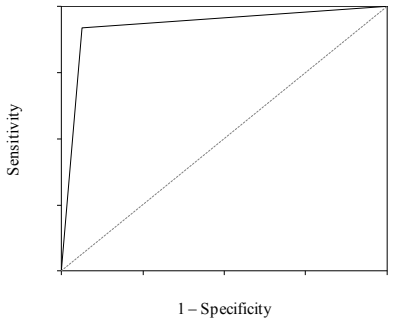

Figure 6: The *ROC* curve for the proposed model

CONCLUSIONS

This work presents an intelligent decision support system aiming to predict pharmacological activity of essential oils of *Schinus* species. It is centred on a formal framework based on *LP* for *Knowledge Representation and Reasoning*, complemented with a *CBR* approach to problem solving that caters for the handling of incomplete, unknown, or even contradictory information. Under this approach the cases' retrieval and optimization phases were heightened and the time spent on those tasks shortened in 11.3%, when compared with existing systems. The proposed approach is able to provide adequate responses since the overall accuracy was around 88% and the area under ROC curve is near 0.9. The proposed method allows for the analysis of free text attributes using *String Similarities Algorithms*, which fulfils a gap that is present in almost all *CBR* software tools. Additionally, under this approach the users may define the weights of the cases' attributes on-the-fly, letting them to choose the most appropriate strategy to address the problem (i.e., it gives the user the possibility to narrow the search space for similar cases at runtime).

ACKNOWLEDGMENTS

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

REFERENCES

Aamodt, A.; and E. Plaza. 1994. "Case-based reasoning: Foundational issues, methodological variations, and system approaches.", *AI Communications* 7, 39-59.

Atti dos Santos A. C.; M. Rossato; L. A. Serafini; M. Bueno; L. B. Crippa; V. C. Sartori; E. Dellacassa; and P. Moyna. 2010. "Antifungal effect of *Schinus* molle L., Anacardiaceae, and *Schinus terebinthifolius* Raddi, Anacardiaceae, essential oils of Rio Grande do Sul." *Brazilian Journal of Pharmacognosy* 20, 154-159.

Bendaoud, H.; M. Romdhane; J.P.Souchard; S. Cazaux; and J. Bouajila. 2010. "Chemical composition and anticancer and antioxidant activities of *Schinus molle* L. and *Schinus terebinthifolius* Raddi berries essential oils." *Journal of Food Science* 75, 466-472.

Bigliani, M.C.; Rossetti, V.; E. Grondona; S. Lo Presti; P.M. Paglini; V. Rivero; M.P. Zunino; and A.A. Ponce. 2012. "Chemical compositions and properties of *Schinus areira* L. essential oil on airway inflammation and cardiovascular system of mice and rabbits.*" Food and Chemical Toxicology* 50, 2282-2288.

Carneiro, D.; P. Novais; F. Andrade; J. Zeleznikow; and J. Neves. 2013. "Using Case-Based Reasoning and Principled Negotiation to provide decision support for dispute resolution." *Knowledge and Information Systems* 36, 789-826.

Deveci, O.; A. Sukan; N. Tuzun; and E.E.H. Kocabas. 2010. "Chemical composition, repellent and antimicrobial activity of *Schinus molle* L." *Journal of Medicinal Plants Research* 4, 2211-2216.

Díaz, C.; S. Quesada; O. Brenes; G. Aguilar; and J.F. Cicció. 2008. "Chemical composition of Schinus molle essential oil and its cytotoxic activity on tumor cell lines." *Natural Product Research* 22, 1521-1534.

Dice, L. 1945. "Measures of the Amount of Ecologic Association between Species." *Ecology* 26, 297-302.

Duke J. 2002. *Handbook of Medicinal Herbs*, CRC Press, Florida.

El-Massry, K.F.; A.H. Ghorab; H.A. Shaaban; and T. Shibamoto. 2009. "Chemical compositions and antioxidant/ antimicrobial activities of various samples prepared from *Schinus terebinthifolius* leaves cultivated in Egypt." *Journal of Agricultural and Food Chemistry* 57, 5265-5270.

Fernandes, F.; H. Vicente; A. Abelha; J. Machado; P. Novais; and J. Neves. 2015. "Artificial Neural Networks in Diabetes Control". In *Proceedings of the 2015 Science and Information Conference (SAI 2015)*, IEEE Edition, 362-370.

Gomes, V.; G. Agostini; F. Agostini; A. C. Atti dos Santos; and M. Rossato. 2013. "Variation in the essential oils composition in Brazilian populations of *Schinus molle* L. (Anacardiaceae)." *Biochemical Systematics and Ecology* 48, 222-227.

Janssen, R.; P. Spronck; and A. Arntz. 2014. "Case-based reasoning for predicting the success of therapy." *Expert Systems* 32, 165-177.

Kakas, A.; R. Kowalski; and F. Toni. 1998. "The role of abduction in logic programming". In *Handbook of Logic in Artificial Intelligence and Logic Programming*, D. Gabbay, C. Hogger and I. Robinson (Eds.). Vol. 5, Oxford University Press, Oxford, 235-324.

Lucas, P. 2003. "Quality checking of medical guidelines through logical abduction". In *Proceedings of AI-2003 (Research and Developments in Intelligent Systems XX)*, F. Coenen, A. Preece and A. Mackintosh (Eds.). Springer, London, 309-321.

Machado J.; A. Abelha; P. Novais; J. Neves; and J. Neves. 2008. "Quality of service in healthcare units". *In Proceedings of the ESM 2008*, C. Bertelle, and A. Ayesh (Eds.). Eurosis, Ghent, 291-298.

Martins, M.R.; S. Arantes; F. Candeias; M.T. Tinoco; and J. Cruz-Morais. 2014. "Antioxidant, antimicrobial and toxicological properties of *Schinus molle* L. essential oils." *Journal of Ethnopharmacology* 151, 485-492.

Mendes, R.; J. Kennedy; and J. Neves. 2003. "Watch thy neighbor or how the swarm can learn from its environment". In *Proceedings of the 2003 IEEE Swarm Intelligence Symposium (SIS'03)*, IEEE Edition, 88-94.

Murray A.P.; S.A. Rodriguez; and N.P. Alza. 2012. "Chemical Constituents and Biological Activities of Plants from the Genus *Schinus*". In *Recent Progress in Medicinal Plants*, J.N. Govil (Ed.). Ethnomedicine and Therapeutic Validation, Vol. 32, Studium Press LLC, Texas, 261-287.

Neves, J. 1984. "A logic interpreter to handle time and negation in logic databases". In *Proceedings of the 1984 annual conference of the ACM on The Fifth Generation Challenge*, R.L. Muller and J.J. Pottmyer (Eds.). ACM, New York, 50-54.

Neves, J.; and H. Vicente. (n. d.) "A Quantum approach to Case-Based Reasoning." (In preparation).

Neves, J.; J. Machado; C. Analide; A. Abelha; and L. Brito. 2007. "The halt condition in genetic programming". In *Progress in Artificial Intelligence*, J. Neves, M.F. Santos and J. Machado (Eds.). Lecture Notes in Artificial Intelligence, Vol. 4874, Springer, Berlin, 160-169.

Pereira, L.M. and H.T. Anh. 2009. "Evolution prospection". In *New Advances in Intelligent Decision Technologies – Results of the First KES International Symposium IDT 2009*, K. Nakamatsu (Ed.). Studies in Computational Intelligence, Vol. 199, Springer, Berlin, 51-64.

Richter, M.; and R. Weber. 2013. *Case-Based Reasoning: A Textbook*. Springer, Berlin.

Simionatto, E.; M. Chagas; M. Peres; S. Hess; C. Silva; N. Ré-Poppi; S. Gebara; J. Corsino; F.Morel; C. Stuker; M. Matos; and J. Carvalho. 2011. "Chemical composition and biological activities of leaves essential oil from *Schinus molle* (Anacardiaceae)." *Journal of Essential Oil Bearing Plants* 14, 590-599.

Ying, S.; C. Joël; J. Arnelle; and L. Kai. 2015. "Emerging medical informatics with case-based reasoning for aiding clinical decision in multi-agent system." *Journal of Biomedical Informatics* 56, 307-317.