Weighted Fuzzy Similarity Relations in Case-Based Reasoning: a Case Study in Classification

Sandra Sandri*, Jonas Mendonça*, Flávia Martins-Bedê*, Ricardo Guimarães† and Omar Carvalho‡
* Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos - SP, Brazil
Email: sandri@lac.inpe.br, jonas.henrique01@gmail.com, flavinha@dpi.inpe.br
† Laboratório de Geoprocessamento/Inst. Evandro Chagas/SVS/MS, Belém - Pa, Brazil
Email: ricardojpsg@gmail.com
‡ Centro de Pesquisas René Rachou/FIOCRUZ, Belo Horizonte - MG, Brazil
Email: omar@cpqrr.fiocruz.br

***Abstract*—This paper describes a fuzzy similarity relation approach to Case-Based Reasoning. Residuated implication operators are used to create a fuzzy resemblance relation between cases, modeling the CBR basic principle "the more similar the problem descriptions are, the more similar the solution descriptions are" as a fuzzy gradual rule. We take the classification of Schistosomiasis prevalence estimation in a region of Brazil as case study, in order to investigate the effects in such a framework of weighting cases individually in classification tasks, considering a set of training strategies.**

I. Introduction

Case-Based Reasoning (CBR) [7], [1] proposes to solve a problem using a principle that can be stated as "similar problems have similar solutions" [1]. A case base consisting of solved problems, modeled as pairs (problem, solution), is used to determine the solution to a new problem. The first step of this procedure consists in retrieving problems in the base that are similar to the considered problem: it determines the cases in the base that are relevant to solving the problem at hand. The second step consists in reusing the solutions of these relevant problems, adapting them to the considered problem.

Weights can be attached to cases, so that cases considered more important for a given application have higher weights. Weight vectors can also be assigned to description variables: one can use either the same weight vector for all cases, or assign individual weight vectors to each case, so that more significant attributes inside a case receive higher weights.

In [2] and [8], fuzzy similarity relations associated to each description and solution variables spaces were used to derive individual vector weights through the learning algorithm proposed in [12]. It was shown that weighting the attributes in each case in the training set tends to lead to better results than the non weighted counterparts.

The problem with weighting is that the learning process is usually computationally expensive, which may make it impossible to be used in large case bases. An approach to allow learning weights in large case bases consists in extracting fragments of the case base and obtain weights for cases inside those fragments, thus considering each fragment as a case base itself. When a new problem is presented to the base, first a solution is calculated from each fragment, taking into account the weighted cases inside it. Then a final solution is either chosen from one of the fragments or aggregated from various ones.

An approach to create case base fragments was proposed in [6]). The proposed method is based on a binary similarity measure between cases, called Case Resemblance Relation, that takes into account their resemblance in the problem description space but also in the solution description space. This measure defines a binary relation between cases; the corresponding graph of cases is then exploited and decomposed to identify clusters of similar cases; the clusters group cases that have both similar problem descriptions and similar solutions. To compute a solution for a new problem, the authors propose to adopt the solution from the cluster with which the problem at had has the highest overall problem resemblance. Only the clusters that are maximal are considered in the process.

This approach was generalized to the case where the similarity measure between cases, defined as the aggregation of the similarities in the problem and the solution spaces, is not binary but takes values in $[0, 1]$, leading to a fuzzy relation FCRR instead of a crisp one [11]. The problem is then to extract clusters from this fuzzy relation. The authors propose to first extract the relevant level-cuts from the fuzzy relation, thus creating a set of crisp relations, for which the procedure devised in [6] can be applied in a straightforward manner.

The main goal of the present work is to study the effect of weighting cases individually in classification tasks, using case resemblance relations. The most relevant question we address here is what should be the training base for obtaining weights for a cluster. Two reasonable options are i) the cases inside the cluster itself and ii) a larger set containing the cases in the cluster. Note that, contrary to option ii), option i) means that negative information (i.e. cases that have similar problem descriptions but dissimilar solutions) is not taken into account to compute the weights in a cluster. In this paper we propose a set of training strategies that can be used in this framework and investigate their effectiveness on a real-world

classification task, that consists of estimating the prevalence of *Schistosomiasis mansoni*, a disease with social and behavioral characteristics, for a region of Brazil [10].

This paper is organized as follows. In Section II we give some basic definitions and the notation used throughout the text and in Section III we describe fuzzy similarity relations approach to CBR. In Section IV we propose training strategies for our approach and in Section V we describe the real-world experiment and compare our results to those from the literature. Section VI finally brings the conclusions.

II. Definitions and Notations

In this section, we recall some basic definitions that are used in the rest of the paper and provide some new definitions and notations, successively concerning fuzzy operators, similarity measures and hypergraphs. based reasoning. The definitions involving fuzzy concepts, such as residuated operators and similarity relations come from well established literature (see, e.g., [4] [5]). The definitions related to graph theory are the consecrated ones, apart from *imprecise partitions*, which is new (up to our knowledge).

An operator $\top : [0,1]^2 \rightarrow [0,1]$ is called a *t-norm* if it is commutative, associative, monotonic and has 1 as neutral element. An operator $\perp : [0,1]^2 \rightarrow [0,1]$ is called a *t-conorm* if it is commutative, associative, monotonic and has 0 as neutral element. Examples of t-norms are the minimum and the product; examples of t-conorms are the maximum and the bounded sum.

Given a left-continuous t-norm $\top$, a *residuated implication operator* $\rightarrow_\top$ is defined as $\forall x, y \in [0,1], x \rightarrow_\top y = \sup_{z \in [0,1]} \top(x,z) \leq y$. Some well-known operators are

- the Gödel implication, residuum of $\top = \min$, defined as $x \rightarrow_{\top G} y = 1$, if $x \leq y$, and $x \rightarrow_{\top G} y = y$, otherwise;
- the Goguen implication, defined as $x \rightarrow_{\top \Pi} y = 1$, if $x \leq y$, and $x \rightarrow_{\top \Pi} y = y/x$, otherwise.

Also noteworthy is the Rescher-Gaines implication operator, defined as $x \rightarrow_{\top RG} y = 1$, if $x \leq y$, and $x \rightarrow_{\top RG} y = 0$, otherwise. $x \rightarrow_{\top RG}$ is not a residuated operator itself but is the point-wise infimum of all residuated implications.

A *similarity relation* S on a domain U is a binary fuzzy relation, i.e. a mapping $S : U \times U \rightarrow [0,1]$ that is both reflexive ($\forall x \in U, S(x,x) = 1$) and symmetric ($\forall x, y \in U, S(x,y) = S(y,x)$). Some authors require it to also satisfy the t-norm transitivity property ($\forall x, y, z \in U, \top(S(x,y), S(y,z)) \leq S(x,z)$ for some t-norm $\top$), but we do not take it into consideration here as it does not play a role in our framework.

The set of similarity relations on a given domain U forms a lattice (not linearly ordered) with respect to the point-wise ordering (or fuzzy-set inclusion) relationship. The top of the lattice is the similarity S_{top} which makes all the elements in the domain maximally similar: $S_{top}(x,y) = 1$ for all $x, y \in U$. The bottom of the lattice S_{bot} is the classical equality relation: $S_{bot}(x,y) = 1$, if $x = y$, and $S_{bot}(x,y) = 0$, otherwise.

Particularly useful are families of parametric similarity relations $\mathcal{S} = \{S_0, S_{+\infty}\} \cup \{S_\beta\}_{\beta \in I \subseteq (0,+\infty)}$ that are such that: (i) $S_0 = S_{bot}$, (ii) $S_{+\infty} = S_{top}$, and (iii) $\beta < \beta'$, then $S_\beta \prec S_{\beta'}$, where $S \prec S'$ means $\forall x, y \in U, S(x,y) \leq S'(x,y)$ and $\exists x_0, y_0 \in U, S(x_0,y_0) < S'(x_0,y_0)$.

A *hypergraph* is a generalization of a non-directed graph, where edges can connect any number of vertices. Formally, it can be represented as a pair, $H = (N, E)$, where N is a set containing the nodes (or vertices) and E is a set of non-empty subsets of N, called hyperedges. The set of hyperedges E is thus a subset of $2^N \backslash \emptyset$, where 2^N is the power set of N. An "ordinary graph" is then a hypergraph in which all hyperedges have at most two elements.

Each graph can be associated to a corresponding hypergraph, whose hyperedges are the cliques of the initial graph. Given a hypergraph $H = (N, E)$, a hyperedge $A \in E$ is said to be maximal when $\nexists B \in E$, such that $A \subseteq B$ and $A \neq B$. Each hyperedge in E is a clique, therefore, the set of maximal hyperedges is the set of maximal cliques of E.

Let B be a subset of a domain U. A set $B' = \{B_1, ..., B_z\}, B_i \subseteq B$, $B_i \neq \emptyset$, is an *imprecise partition* of B when $\bigcup_{i=1,z} B_i = B$ and $\forall i, j \in 1, z, i \neq j, \nexists B_i, B_j \in B$, such that $B_i \subseteq B_j$. Each $B_i \in B$ is called an *imprecise class*. An imprecise partition does not allow one class to be contained inside another one but, contrary to precise partitions, it does allow non-empty intersections between classes. Let B' and B'' be two imprecise partitions of B. B' is said to be *finer* than B'' (denoted by $B' \preccurlyeq B''$) when $\forall h' \in B', \exists h'' \in B'', h' \subseteq h''$. Reciprocally, B'' is said to be *coarser* than B'.

III. Fuzzy Similarity Relations CBR Approach

In the following we describe a fuzzy similarity relations approach to Case-Based Reasoning, based on the use of residuated implications operators. Most of the definitions and notations are the same as those in [6] and [11].

Let a case c be defined as an ordered pair $c = (p, o) \in P \times O$ where p is the description of the case problem and o the description of its solution. $P = \{P_1 \times ... \times P_n\}$ and O are respectively the (n-ary) problem description and the (unary) solution spaces.

Let $S_{in} : P^2 \rightarrow [0,1]$ and $S_{out} : O^2 \rightarrow [0,1]$ respectively denote the similarity relations used on the problem and solution spaces. S_{in} may obtained by using a suitable aggregation function applied on the set of similarity relations $\{S_1, ..., S_n\}$, each of which corresponding to a description variable. For example, using the arithmetic means, we have: $S_{means}(p_i, p_j) = \frac{1}{n} \sum_{k=1,n} S_k(p_{ik}, p_{jk})$.

A. Fuzzy Case Resemblance Relations

We here model a gradual formalization of the basic CBR principle "the more similar the problem descriptions are, the more similar the solution descriptions are".

We define a *Fuzzy Case Resemblance Relation* (FCRR) as the mapping $F_\phi : C^2 \rightarrow [0,1]$, defined as

$$F_\phi(c_a, c_b) = \begin{cases} 0, \text{if } S_{in}(p_a, p_b) = 0 \\ \phi(S_{in}(p_a, p_b), S_{out}(o_a, o_b)), \text{otherwise,} \end{cases} \quad (1)$$

where $\phi = \rightarrow_\top$ is a residuated implication operator obtained from a T-norm $\top$ (see Section 2). Note that two cases are considered dissimilar whenever their problem descriptions and/or their solutions are completely dissimilar.

B. Obtaining Crisp Case Resemblance Relations

A fuzzy case resemblance relation F_ϕ is not necessarily crisp. To be able to obtain clusters from the case base, we propose to first take a relevant level-cut from F_ϕ, and then extract clusters from the resulting crisp relation.

The proposed methodology relies on the α level cut decomposition of the fuzzy case resemblance relation: we propose to derive hypergraphs from the crisp relations

$$\forall \alpha \in (0,1], F_{\phi,\alpha}(c_i, c_j) = \begin{cases} 1, \text{if } F_\phi(c_i, c_j) \geq \alpha \\ 0, \text{otherwise} \end{cases} \quad (2)$$

Each $F_{\phi,\alpha}$ is called a Crisp Case Resemblance Relation (CCRR). The number of cases is finite and we thus obtain a finite number of CCRRs from a FCRR F_ϕ, one for each distinct value greater than 0 in F_ϕ.

C. Obtaining Clusters from a Case Base

Let $c_a = (p_a, o_a)$ and $c_b = (p_b, o_b)$ denote two cases in C. Let R be a CCRR obtained as an α level cut from a given F_ϕ.

Based on CCRR R, the case set can be organized through a decomposition in clusters based on this crisp resemblance relation. Several (possibly intersecting) clusters of cases can be obtained from R. The latter can in turn be represented as a hypergraph. More precisely, a hypergraph $H = (C, E), E \subseteq C^2$ is said to *be compatible with* CCRR R iff it obeys the following conditions:

- $\forall c_a, c_b \in C$, if $R(c_a, c_b) = 1$, then $\exists h \in E$, such that $\{c_a, c_b\} \subseteq h$.
- $\forall c_a, c_b \in C$, if $R(c_a, c_b) = 0$, then $\nexists h \in E$, such that $\{c_a, c_b\} \subseteq h$.

A notable hypergraph compatible with R is the one containing the maximal cliques of E.

D. Computing a Solution to a New Problem according to a Cluster

Given a case base C, similarity measures S_j for each description variable v_j, global similarity measures S_{in} and S_{out}, and a hypergraph $H = (C, E)$ compatible with $R = F_{\phi,\alpha}$, for a given a residuated operator ϕ and a value $\alpha \in (0, 1]$, we have to derive an appropriate solution o^* for a new problem description, denoted p^* in the following.

We propose to derive solution o^* from the clusters containing cases whose problem descriptions are somewhat similar to p^*, denoted $E^* = \{h \in E \mid \forall c_i = (p_i, o_i) \in h, S_{in}(p_i, p^*) > 0\}$. For each $h = \{c_1, ..., c_r\} \in E^*$, we compute its corresponding solution for p^*, denoted by o^*_h, using a suitable aggregation function that takes into account the set of solutions o_i as well as the similarity between each p_i and p^*, considering the cases (p_i, o_i) in h. For example, using the weighted means as aggregation function we obtain $o^*_h = \sum_{i=1,r} \frac{S_{in}(p_i,p^*) \times o_i}{\sum_{i=1,r} S_{in}(p_i,p^*)}$.

S_{in} is substituted by its counterpart in weighted frameworks (see below). Also, in case non-numerical description variables occur, the aggregation method above cannot be applied but can be replaced with a weighted voting method.

E. Determining Cluster Strength in Relation to a New Problem

Let O^* be the set of solutions for p^* from the clusters in E^*. To select a final solution o^* from O^*, one can aggregate the solutions produced by the clusters, or simply choose a cluster $h \in E^*$ and make $o^* = o^*_h$.

To select a cluster in the latter option, we propose to take the cluster whose overall cases problem descriptions resembles most p^*. We define the *cluster strength* of each cluster $h = \{c_1, ..., c_r\} \in E^*$, $c_i = (p_i, o_i)$, in relation to a problem p^* as

$$str_f(h, p^*) = f(S_{in}(p_1, p^*), ..., S_{in}(p_n, p^*)).$$

where f is a suitable aggregation function, such as the arithmetic means, a t-norm, a t-conorm, an OWA operator, etc.

F. Using Attribute Weights

The solution to a cluster and the cluster strength can both be determined using weighted operators. Weights can be attached to cases, so that cases considered more important for a given application are assigned greater weights.

Weight vectors can also be assigned to description variables, so that the attributes considered more relevant receive greater weights. A single weight vector can be used for all cases in the base, or an individual attribute weight vector can be assigned to each case (see [12] for an individual weight vector learning algorithm). Note that the first method can be obtained from the second method, when the same weight is assigned to an attribute in all individual vectors.

Many aggregation functions have weighted counterparts, as for instance the means, t-norms and t-conorms operators, that can be used to derive a weighted S_{in} relation. For example, using the individual attribute weighting approach, the weighted means as aggregation operator yields $S^w_{means}(p_i, p_j) = \sum_k w_k \times S_k(p_{ik}, p_{jk})$, where w is a weight vector such that $\forall k,\ w_k \in [0, 1]$ and $\sum_k w_k = 1$.

A weighted version of S_{in} can also be used to compute the clusters themselves. If individual weight vectors are used, the resulting relation is possibly asymmetric and one has to make it symmetric before applying the rest of the formalism (see [6]).

G. Learning individual attribute weight vectors for cases

In order to learn attribute weight vectors for each case in a case base C, we first of all divide C in two parts, a training set C_T and a validation (or test) set $C_V = C - C_T$.

Furthermore, for any given case $c_0 = (p_0, o_0) \in C_T$, we need to select a subset LS_0 of C_T from which an individual weight vector w_0 for c_0 is learned.

In [12] and [13], a weight assignment $\mathbf{w}_0$ for c_0 is calculated in a leave-one-out fashion, where the learning base is obtained with the removal of c_0 from C_T, i.e., $\forall c_0 \in C_T$,

$LS_0 = C_T - \{c_0\}$. The algorithm uses similarity relations for both description and solution variables, by iteratively building description variable weight vectors that approximate the overall similarity between the descriptions of cases in LS_0 and p_0 and between the solution of cases in LS_0 and o_0.

H. Some relevant properties of FCRR's and CCRR's

Given a case base, let CCRRs $F_{\phi\uparrow}$ and $F_{\phi_\downarrow}$ be respectively calculated as

$$F_{\phi\uparrow}(c_i, c_j) = \begin{cases} 1, \text{if } 0 < S_{in}(p_a, p_b) \leq S_{out}(o_a, o_b) \\ 0, \text{otherwise;} \end{cases} \quad (3)$$

and

$$F_{\phi_\downarrow}(c_i, c_j) = \begin{cases} 1, \text{if } \min(S_{in}(p_a, p_b), S_{out}(o_a, o_b)) > 0 \\ 0, \text{otherwise.} \end{cases} \quad (4)$$

Using 1 to 4 it is straightforward to prove that for any residuated implication operator ϕ', $F_{\phi',1} = F_{\phi^*}$ and $F_{\phi',\epsilon} = F_{\phi_*}$, with $\epsilon = \inf\{x \in \mathbb{R}^+\}$, such that x tends to 0.

Using the maximal cliques strategy, the hypergraphs $H_\alpha = (C, E_\alpha)$ generated from the level cuts F_α of a given FCRR F_ϕ are nested:

$$\text{if } \alpha \geq \alpha' \text{ then } E_\alpha \preccurlyeq E_{\alpha'}$$

(see Section 2 for the definition of $\preccurlyeq$). E_α is thus an imprecise partition finer than $E_{\alpha'}$.

It follows from 3 (respec. 4) that the finest (respec. the coarsest) imprecise partition that can be obtained from a case base using a residuated implication operator ϕ' is the one derived from $F_{\phi\uparrow}$ (respec. $F_{\phi_\downarrow}$)[1].

Therefore both that finest and coarsest partitions that can be derived from a case base are unique, no matter which residuated operator is used to create a FCRR.

Let $CL = \{cl_1, ..., cl_k\}$ be a set of labels and let $class : O \rightarrow CL$ be a function that assigns labels (classes) to instances of the case base solution space. Let relation S_{class} be given by:

$$\forall c_i, c_j \in C, S_{class}(o_i, o_j) = \begin{cases} 1, \text{if } class(c_i) = class(c_j) \\ 0, \text{otherwise} \end{cases} \quad (5)$$

Taking $S_{out} = S_{class}$, we can prove that for any residuated implication operator ϕ, the derived FCRR F_ϕ is crisp, given by CCRR $F_{\phi_\downarrow}$. In other words, similarity relation S_{class} induces a precise partition of the output space. As a consequence, the case base is also precisely partitioned, and is such that at least one equivalence class is associated to each label in CL appearing as a solution to a case in C. Thus, two cases will be related through F_ϕ iff their solutions belong to the same class and their problem descriptions are somewhat related.

[1] $F_{\phi\uparrow}$ is named S_{res} in [6].

IV. Training strategies

Clusters can be used to calculate weight vectors and/or to calculate results. We are now interested in what should be the training base to derive a weight vector for a case inside a cluster. Two reasonable options are i) the cases inside the cluster itself and ii) a larger set containing the cases in the cluster.

Contrary to option ii), option i) means that negative information (i.e. cases that have similar problem descriptions but dissimilar solutions) is not taken into account to compute the weights in a cluster, which intuitively does not lead to optimal vector weights.

We have devised a series of types of experiments to test the most effective combination of sets of cases used for training and for the calculation of results. We use the letters "W" and "R" in order to specify the type of training and the type of calculation of results, respectively. We may either not use weights (W-) in an experiment, or obtain them considering the whole base as a single cluster (W), or yet obtain them considering several clusters (W+). The results may be calculated using either a single cluster (R) or several ones (R+). Moreover, W++ denotes the training strategy in which when a set of cases larger than the cases in a cluster itself are used for obtaining weights for the cases in the cluster[2].

Here we address six types of experiments (see Figure IV):

- W-R (without weights, results calculated for single cluster),
- W-R+ (without weights, results calculated for several clusters),
- WR (with weights learned for a single cluster, results calculated for single cluster),
- WR+ (with weights learned for a single cluster, results calculated for several clusters),
- W+R+ (with weights learned for several clusters using the cluster itself as training base, results calculated for several clusters), and
- W++R+ (with weights learned for several clusters using enlarged cluster as training base, results calculated for several clusters).

The missing combinations W+R and W++R would require a further step to be implemented. They both mean, on the one hand, that as many weight vectors are obtained for a given case as the number of clusters to which it belongs. On the other hand, a single cluster is used for the calculation of results. So either a unique vector would have to be derived for each case (using the means of the weight vectors for example) or the base would have to have as many instances of a given case, each of which associated to a weight vector, as the number of clusters to which that case belongs.

In the weighted strategies, we need to construct a training base for each cluster $h \in E$ derived from case base C, denoted as CT_h. For W+ we simply make $CT_h = h$.

[2] Note that the use of and R++, in which enlarged clusters would be used to calculate the result for a new case, could also be envisaged.

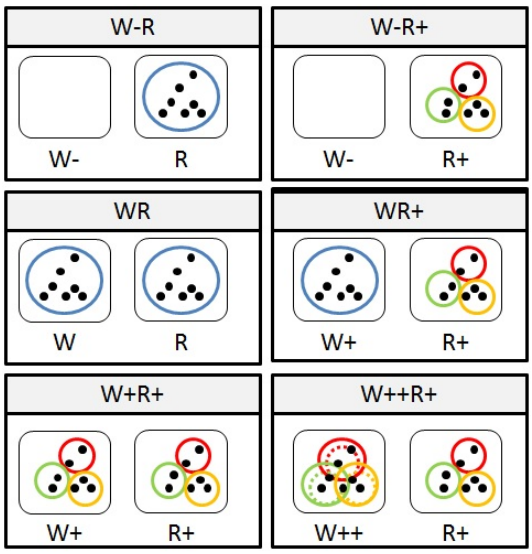

Fig. 1. Illustration of experiment typology.

For W++, an enlarged cluster serves as the training base for the calculation of weight vectors to be associated to the cases in that cluster. Let $c' \in C$ be a case, such that $c' \notin h$. Two reasonable methods to allow case c' to belong to CT_h are:

- c' is somewhat related to all cases in h:
  $c' \in CT_h$ iff $\inf_{c \in h} S_{in}(c, c') > 0$.
- c' is somewhat related to some cases in h:
  $c' \in CT_h$ iff $\sup_{c \in h} S_{in}(c, c') > 0$;

Approach i) is more restrictive and produces training bases smaller than ii). In our experiments, we have used approach i), since approach ii) tends to make the enlargement of the cluster CT_h become the whole base C itself.

Other functions can be devised. An interesting possibility is to relax the conditions above, setting another value than 0 as threshold for the similarities between a case and a cluster.

V. Classification of Schistosomiasis prevalence

Schistosomiasis mansoni is a disease with social and behavioral characteristics. Snails of the *Biomphalaria* species, the disease intermediate host, uses water as a vehicle to infect man, the disease main host. In Brazil, six million people are infected by it, mainly in poor regions of the country [14]. According to the data presented at the Brazilian Information System for Notifiable Diseases (SINAN) of the Ministry of Health, from 1995 to 2005, more than a million positive cases were reported, 27% of them in the State of Minas Gerais. In [10], the authors present a classification Schistosomiasis prevalence for the State of Minas Gerais, using remote sensing, climate, socioeconomic and neighborhood related variables. Two approaches were used, a global and a regional one. In the first approach, a unique regression model was generated and used to estimate the disease risk for the entire state. In the second approach, the state was divided in four regions, and a model was generated for each one of them. Imprecise classifications were also generated for both approaches, using the estimated standard deviation and several reliability levels as basis. In [8] the authors compare those experiments with a similarity based approach, weighting the cases individually.

In the original experiments presented in [10], the disease prevalence data was provided by the Health Secretary of the State of Minas Gerais state. The prevalence is known for 197 municipalities out of the 853 composing the state (see Figure 2.a). In the original experiments, 86 independent variables of various types were used to classify prevalence: Remote Sensing (22), climatic (6), socioeconomic (34) and neighborhood characterization (24). The Remote Sensing variables were derived from sensors MODIS (Moderate Resolution Imaging Spectroradiometer) and SRTM (Shuttle Radar Topography Mission), and are supposedly related to the snail habitat type. The climatic variables were obtained from the Weather Forecast and Climate Studies Center (CPTEC) from the National Institute for Space Research (INPE) and reflects the conditions of survival of the snail and the various forms of the *Schistosoma mansoni* larvae. The socioeconomic variables were obtained from SNIU (Brazilian National System of Urban Indicators) such as the water accessing means and sanitation condition aspects. The neighborhood characterization variables measure the disparity between neighboring municipalities with relation to variables of income, education, sewerage, water access and water accumulation.

From the original 86 variables, a smaller set was selected, according to tests using multiple linear regression [10]; the independent variables chosen were those that had high correlation with the dependent variable and low correlation with other independent variables. Two main approaches were used: i) a global one, in which all the municipalities with known disease prevalence were used, for either constructing or validating a linear regression model, and ii) a regional one, in which the state was divided in four homogeneous regions and a linear regression model was created for each one of them. The number of independent variables used in the experiment varied; in the global approach 5 variables were used, while in the regional approach 2 variables were used for region R1, 5 for region R2, 4 for region R3 and 3 for region R4 (see details in [10]). In both the global and regional approaches, approximately 2/3 of the samples were used as training set, and the remainder 1/3 as the test set. Algorithm SKATER [3] was used to obtain the homogeneous regions in the regional model; this algorithm creates areas such that neighboring areas with similar characteristics belong to the same region (see Figure 2.b).

The prevalences were classified as low, medium or high, respectively defined by intervals $[0, 5]\%$, $(5, 15]\%$ and $(15, 100]\%$. Table I reproduces the results from [10], with the accuracy of the results obtained using regression and decision trees for all regions, for both the training and validation data sets. Table II details the results presented in [8], using similarity based weighting, it is in fact an instance of experiment type WR. The sizes of the case bases used for training and test are indicated in the tables. Note that in [10], the total amount of samples from region R1 was used for training.

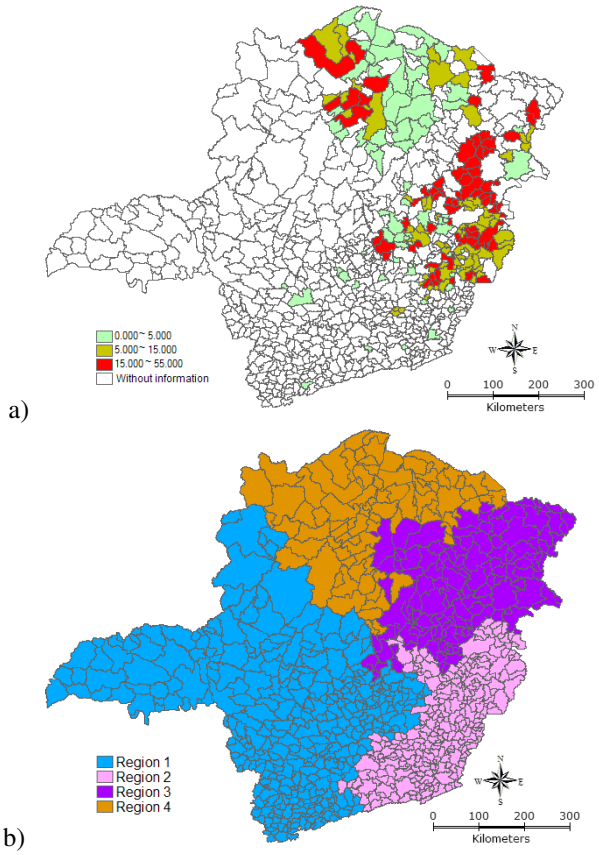

Fig. 2. State of Minas Gerais in Brazil: a) known prevalence of Schistosomiasis (source: Minas Gerais Secretary of Health) and b) regionalization obtained through algorithm SKATER (source: [10]).

Training	R1 (16)	R2 (59)	R3 (44)	R4 (28)
G-Regr	50.00% (8)	42.37% (25)	54.55% (24)	57.14% (16)
R-Regr	56.25% (9)	54.24% (32)	59.09% (26)	60.71% (17)
G-DT	62.50% (10)	64.41% (38)	77.27% (34)	78.57% (22)

Test	R1(0)	R2 (27)	R3 (14)	R4 (9)
G-Regr	-	37.04% (10)	28.57% (4)	77.78% (7)
R-Regr	-	29.63% (8)	42.86% (6)	100.00% (9)
G-DT	-	29.63% (8)	35.71% (5)	44.44% (4)

TABLE I
CLASSIFICATION ACCURACY, FOR TRAINING AND TEST, WITH LEARNING MADE ON: R-REGR (REGIONAL BASIS AND REGRESSION), G-REGR (GLOBAL BASIS AND REGRESSION) AND G-DT (GLOBAL BASIS AND DECISION TREES [9].

Training	R1 (11)	R2 (59)	R3 (44)	R4 (28)
G-Sim	18.18% (2)	50.84% (30)	54.54% (24)	25.00% (7)
R-Sim	63.63% (7)	37.28% (22)	56.81% (25)	35.71% (10)

Test	R1 (5)	R2 (27)	R3 (14)	R4 (9)
G-Sim	20% (1)	33.33% (9)	28.57% (4)	33.33% (3)
R-Sim	20% (1)	55.55% (15)	35.71% (5)	33.33% (3)

TABLE II
CLASSIFICATION ACCURACY, FOR TRAINING AND TEST, WITH LEARNING MADE ON: R-SIM (REGIONAL BASIS AND SIMILARITY APPROACH) AND G-SIM (REGIONAL BASIS AND SIMILARITY APPROACH) [10].

Test	R1 (5)	R2 (27)	R3 (14)	R4 (9)
W-R	20% (1)	18.51% (5)	14.28% (2)	55.56% (5)
W-R+	20% (1)	40.74% (11)	42.86% (6)	55.56% (5)
WR	20% (1)	29.63% (8)	57.14% (8)	33.33% (3)
WR+	20% (1)	44.44% (12)	64.29% (9)	55.56% (5)
W+R+	20% (1)	37.04% (10)	64.29% (9)	55.56% (5)
W++R+	20% (1)	33.33% (9)	57.14% (8)	33.33% (3)

TABLE III
CLASSIFICATION ACCURACY: TEST SET, GLOBAL APPROACH, FUZZY SIMILARITY BASED METHOD.

Test	R1 (5)	R2 (27)	R3 (14)	R4 (9)
W-R	40% (2)	51.85% (14)	42.86% (6)	33.33% (3)
W-R+	20% (1)	48.15% (13)	57.14% (8)	55.56% (5)
WR	40% (2)	37.04% (10)	50.00% (7)	33.33% (3)
WR+	20% (1)	48.15% (13)	71.42% (10)	66.67% (6)
W+R+	40% (2)	29.63% (8)	14.29% (2)	66.67% (6)
W++R+	40% (2)	51.85% (14)	57.14% (8)	66.67% (6)

TABLE IV
CLASSIFICATION ACCURACY: TEST SET, REGIONAL APPROACH, FUZZY SIMILARITY BASED METHOD.

VI. CLASSIFICATION OF SCHISTOSOMIASIS PREVALENCE USING WEIGHTED CLUSTER BASED SIMILARITY APPROACH

Below we address the problem of classification in our framework, using the *Schistossomiasis* estimated prevalence classification problem described in the previous section as case study. We have focused on the validation case bases only, since they are the most relevant ones and those with poorer results in the literature.

A. Methodology

In the experiments, we used the same variables of the experiments described in Section V. We have also used the same similarity relations for description variables that resulted in Table II, constructed by means of the parameterized similarity relation family described in Section II. The only exception is that we used S_{class} for S_{out} (see Section III.H), instead of the parameterization used in [8]. The arithmetic means was used to calculate the solution in the non-weighted framework, and the weighted means was used to both determine cluster strength and calculate the solution in the weighted frameworks.

In our experiments, the training and test case bases are the same as the ones in the literature referenced in the previous section, except for region R1, for which 11 samples were used for training and 5 for testing. Tables III and IV bring the results using the global and regional approaches, respectively.

B. Analysis

In the following we concentrate our discussion on regions R2, R3 and R4. Region R1 has a very small number of samples making it not very significative. We also focus exclusively on the test results.

We can see in Tables I to IV that regional basis methods fared usually better than the global ones, which indicates that

using more specialized sets of variables did indeed lead to better predictions. We can also see that, with the notable exception of region R4, the similarity based methods (see Tables II, III and IV) produced better results than regression or decision trees (see Table I).

The results from Table II are equivalent to using strategy WR. Comparing them to WR in Tables III and IV show that the use of S_{class} as the similarity relation for the solution space, and consequent precise partitioning of the case base did not lead to significantly better results than those using another parameterization.

Considering only the FCCR methods, Tables III and IV indicate that usually weighting strategies perform better than non-weighting ones. A possible explanation is that the weighting is capable of partially compensating poor parameterizations for the similarity relations. The notable exception is W+R+ in R3, which fares worse than all other methods, both using our framework and those from the literature.

W+R+ usually yields poorer results than WR+, indicating that learning weights taking the whole base into account is a better choice than learning them on clusters and taking the result calculated using the strongest cluster. However, enlarging the clusters (W++R+) seems to be somewhat compensate the gap, which is a very promising result.

For region R2, the simplest experiment, W-R, i.e. the cases are not weighted and results are calculated considering the whole base as a single cluster, leads with W++R+ to the best results. A careful analysis of the data, however, reveals that using W-R on R2, 26 out of 27 cases were classified as medium. The simple use of similarity relations, without clusters or weights, was not capable of distinguishing the classes in the test base. Also in R3, 10 out of 12 cases from W-R are placed on a single class. This means that the similarity relations employed should be improved, in order to decrease bias.

VII. Conclusions

We have presented an approach to Case-Based Reasoning grounded on fuzzy similarity relations and residuated implications operators. We propose to create clusters of cases based on fuzzy gradual rules, modeling the principle "the more similar the problem descriptions are, the more similar the solution descriptions are". The approach was first proposed having weighting cases individually for large case bases. The reduction of the training sets produced by the clustering may drastically reduce the computational cost of learning the weight vectors.

However, efficiency in computation is less important than performance and the approach can only be useful if it produces the same or better results than the use of the whole base to compute the weights.

Here we have studied this issue, focusing in a classification task. We have compared several strategies using a case base created to estimate prevalence of Schistosomiasis for a region of Brazil [10]. We have found that all our methods but one outperformed those used in the literature to deal with the same data.

The single exception is a method (W+R+) that clusterizes the case base, obtains weight vectors using the set of cases in the clusters themselves as training bases and then computes the results also using the clusters. It makes our best method (W++R+) more outstanding, since the only difference between them is that in the latter one, the training base for each cluster contains the cases in the cluster and also cases similar to those in the cluster, but that belong to other classes. We believe that these extra cases are able to provide negative examples in the training base, making the weighting more suitable to the application in hand.

As future work, we intend to deepen the study on how to enlarge the clusters training bases, so as to obtain both good performance in both what regards the quality of the results and the efficiency in obtaining them.

We also intend to apply the framework for other problems, building the application from scratch. Indeed, it is possible that variables chosen for an application to improve regression results are not the most suitable for our framework.

Acknowledgments

The authors would like to thank FAPESP for grant No 2012/02077-8 and both CNPq and CAPES for student scholarships. The authors are also indebted with Corina Freitas and the paper reviewers for useful comments and suggestions.

References

[1] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches", *AI Commun.*, 7(1):39–59, 1994.

[2] E. Armengol, F. Esteva, L. Godo, and V. Torra, "On learning similarity relations in fuzzy case-based reasoning", *Trans. on Rough Sets*, pp. 14–32, 2004.

[3] R.M. Assunção, M.C. Neves, G. Câmara, C.C. Freitas, "Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees", *IJGIS*, vol. 20, pp. 797–811, 2006.

[4] D. Dubois and H. Prade, *Possibility theory: an approach to computerized processing of uncertainty*. Plenum Press, 1988.

[5] D. Dubois and H. Prade, "Gradual inference rules in approximate reasoning" *Information Sciences* vol. 61/1-2,1992 pp. 103–122, Elsevier.

[6] T. Fanoiki, I. Drummond, and S. Sandri, "Case-based reasoning retrieval and reuse using case resemblance hypergraphs", *Proc. FuzzIEEE'10*, 2010.

[7] J. Kolodner, *Cased-based reasoning*. Morgan Kaufmann, 1993.

[8] F.T. Martins-Bedê, L. Godo, S. A. Sandri, L.V. Dutra, C.C. Freitas, O.S. Carvalho, R.J. Guimaraes, and R.S. Amaral, "Classification of schistosomiasis prevalence using fuzzy case-based reasoning" *Proc. IWANN'09, LNCS 5517*, pp. 1053–1060. Springer, 2009.

[9] F.T. Martins. *Mapeamento do risco da esquistossomose no estado de Minas Gerais, usando dados ambientais e sociais*, Computação Aplicada MSc Thesis (in Portuguese), INPE, SJCampos, 2008.

[10] F.T. Martins, C. Freitas, L. Dutra, S. Sandri, I. Drummond, F. Fonseca, R. Guimarães, R. Amaral, O. Carvalho, "Risk mapping of Schistosomiasis in the state of Minas Gerais, Brazil, using MODIS and socioeconomic spatial data", *IEEE TGRS*, vol. 47, n. 11, pp. 3899-3908, 2008.

[11] S. A. Sandri, M.-J. Lesot. "A Fuzzy Residuated Approach to Case-based Reasoning" submitted.

[12] V. Torra. "On the learning of weights in some aggregation operators: the weighted mean and owa operators", *Math. and Soft Comp.*, 6, 2000.

[13] V. Torra. "Learning weights for the quasi-weighted means", *IEEE Trans. on Fuzzy Systems*, vol. 10:5, pp. 653–666, 2002.

[14] World Health Organization, "The Control of Schistosomiasis. Second Report of the WHO Expert Committee". *Technical Report Series no. 830*, Geneva, 1993.