Extracting Linguistic Rules from Database using Linguistic Aggregation Operator

Zheng Pei Yingchao Shao

School of Mathematics & Computer Science, Xihua University, Chengdu 610039, China

Abstract

In real-world database, most attribute values of objects are numerical, numeral is too detail to obtaining good information or decision. Hence, linguistic rules of a set of data would be very desirable and human consistent. Based on a new aggregation operator for aggregating linguistic terms, extracting linguistic rules are presented. Due to obtain a linguistic rule with higher truth degree, genetic algorithm is used to optimize the number and membership functions of linguistic terms.

Keywords: Linguistic rules, Linguistic aggregation operator, Genetic algorithm

1. Introduction

In our real life, on one hand, we usually face an abundance of data that is beyond human cognitive and comprehension skills. On the other hand, natural language is used to communicate information by human beings. Hence, linguistic rules of a set of data would be very desirable and human consistent, *e.g.*, for a data set on employees, a statement (a linguistic rule) like "almost all younger and well qualified employees are well paid" would be useful and human consistent [1]-[5]. For a database, discovering linguistic rules from a database needs intelligent methods. The methodology of computing with words (CWW) [6] may be viewed as an attempt to harness the highly expressive power of natural languages by developing ways of CWW or propositions drawn from a natural language [7], [8]. Based on difference background, CWW have been studied in [9]-[21].

Formally, an approach to the linguistic data summary of a database could be expressed as following [1]:

1) V is a quality (attribute) of interest, *e.g.*, age, salary, etc, in a database of employees;

2) $Y = \{y_1, y_2, \cdots, y_n\}$ is a set of objects (records) that manifest quality V, *e.g.*, the set of workers. Hence, $V(y_i)$ $(i = 1, \cdots, n)$ are values of quality V for each object y_i;

3) $D = \{V(y_i) | i = 1, \cdots, n\}$ is a set of data (the "database" on question).

Accordingly, a simple linguistic rule ccould be expressed, *e.g.*, *most of employees are young is T*. It can be formalized by Qys *are* S *is* T, in which Q is a fuzzy linguistic quantifier, $Y = \{y_i | i = 1, \cdots, n\}$ is a set of objects, S is a summarizer (linguistic values), and T is a truth degree, *e.g.*, "most (Q) of employees (Y) are young (S) is T".

For linguistic value $s_{l'} \in S$, it's membership function is $\mu_{s_{l'}} : D_{s_{l'}} \longrightarrow [0, 1]$. membership functions of $q_{m'} \in Q$ and $t_{k'} \in T$ can be defined as follows [18]:

1) Let $P(Y) = \{A | A \subseteq Y\}$ be the power set of Y. Define a binary relation on $P(Y)$: $A \sim B$ if and only if $| A |=| B |$, where $| A |$ is power of set A. Obviously, "$\sim$" is an equivalence relation on $P(Y)$. The factor set of $P(Y)$ by $\sim$ is denoted by $\overline{P}(Y) = P(Y)/\sim$.

2) For each fuzzy linguistic quantifier $q_{m'} \in Q$, its fuzzy set is defined by

$$\mu_{q_{m'}} : \overline{P}(Y) \longrightarrow [0, 1]. \quad (1)$$

3) For each fuzzy linguistic truth degree $t_{k'} \in T$, its fuzzy set is defined by

$$\mu_{t_{k'}} : [0, 1] \longrightarrow [0, 1]. \quad (2)$$

For fixed Q, S and T, linguistic rules can be extracted automatically as follows [18]:

(1) Fixing a linguistic value $s_{l'} \in S$ (it can be one or several) and a level (threshold) θ, this can be done by experts or deciders. Let

$$D^{\theta}_{s_{l'}} = \mu^{-1}_{s_{l'}}(V(y_i)) = \{V(y_i) | \mu_{s_{l'}}(V(y_i)) \geq \theta\} \quad (3)$$

(2) Selecting $q_{m'} \in Q$. According to Eq.(3), $q_{m'}$ can be selected such that

$$\mu_{q_{m'}}(A) = max\{\mu_{q_1}(A), \cdots, \mu_{q_m}(A)\}, \quad (4)$$

in which $A = \{y_i | V(y_i) \in D^{\theta}_{s_{l'}}\}$.

(3) Selecting $t_{k'} \in T$. From the logical point of view, $t_{k'}$ can be selected as $\mu_{t_{k'}}(\mu_{q_{m'}}(A)) = max\{\mu_{t_1}(\mu_{q_{m'}}(A)), \cdots, \mu_{t_k}(\mu_{q_{m'}}(A))\}$.

Example 1 *[18] Given a part of a database. Let* $S_{\textbf{age}} = \{young(y),\ middle\ age(ma)\}$, $S_{\textbf{salary}} = \{low(l),\ high(h)\}$, $Q = \{several(s),\ about\ half(ah),\ most(m)\}$, $T = \{approximately\ true(at),\ true(t),\ very\ true(vt)\}$. *Membership functions are given, e.g.,*

$$\mu_y(x) = \begin{cases} 1, & if x \in [25,30], \\ 4 - \frac{1}{10}x, & if x \in (30,40], \\ 0, & if x > 40. \end{cases}$$

$$\cdots$$

$$\mu_{vt}(x) = \begin{cases} 5x - 4, & if x \in [0.8,1], \\ 0, & if x \in [0,0.8). \end{cases}$$

(1) Fixing linguistic values $s' = young \in S_{\textbf{age}}$ *and* $s'' = high \in S_{\textbf{salary}}$. *Let threshold* $\theta = 0.5$, *then* $D^{0.5}_{s'} = \{V(y_i)|\mu_{s'}(V(y_i)) \geq 0.5\} = \{25,31,35,28,34,27\}$, $D^{0.5}_{s''} = \{V(y_i)|\mu_{s''}(V(y_i)) \geq 0.5\} = \{2.8,3.0,3.5,2.9,3.1\}$, $A_{s'} = \{y_i|V(y_i) \in D^{0.5}_{s'}\} = \{y_1,y_3,y_4,y_5,y_9,y_{10}\}$ *and* $A_{s''} = \{y_i|V(y_i) \in D^{0.5}_{s''}\} = \{y_3,y_4,y_5,y_6,y_9,\ y_{10},y_{11},y_{12}\}$ *(2) According to* μ_s, μ_{ah}, μ_m *and* $A_{s'}$, *obtain* $\mu_s(A_{s'}) = 0, \mu_{ah}(A_{s'}) = 1$, *and* $\mu_m(A_{s'}) = 0$, *i.e.,* $max\{\mu_s(A_{s'}),\mu_{ah}(A_{s'}),\mu_m(A_{s'})\} = \mu_{ah}(A_{s'})$, *and* $\mu_{at}(\mu_{ah}(A_{s'})) = \mu_t(\mu_{ah}(A_{s'})) = \mu_{vt}(\mu_{ah}(A_{s'})) = 1$. *The linguistic rule is "about half of employees are young is very true." (3) Similar (2), obtaining the linguistic rule is "most of employees have high salary is approximately true".*

2. Aggregation of fuzzy linguistic values

The management of linguistic information implies the use of operators of comparison and aggregation. In [11], the *linguistic weighted averaging (LWA)* operator was presented as a tool to aggregate linguistic weighted values. Based on information granule (IG) [7], modifying the index of linguistic label and OWA operator [22], new linguistic ordered weighted averaging operators F_{lowa} and F_{i-lowa} have been proposed as follows [16]:

- Let $X \subseteq R$ be a universal set, all granules of X denotes $P(x)$, and $R^+ = \{x|x \geq 0\}$, define a linear function $F : X \longrightarrow R^+$;
- Let μ_A be a membership function of a granule $A \in P(X)$. Let $\widetilde{F}(X)$ be the collection of membership functions on X such that $\forall \mu_A(x) \in \widetilde{F}(X)$, $\mu_A(x)$ be a membership function of a granule $A \in P(X)$. Define $G : \widetilde{F}(X) \longrightarrow P(X)$.
- An equivalence relation "$\simeq$" on $\widetilde{F}(X)$ can be obtained: $\mu_A(x) \simeq \mu_B(x) \leftrightarrow G(\mu_A(x)) = G(\mu_B(x))$, *i.e.,* $\mu_A(x)$ and $\mu_B(x)$ are membership functions of the same granule A, and each equivalence class is denoted by $[\mu_A(x)] \in \widetilde{F}(X)/\simeq$.
- Select a representative element $\mu_A(x)$ of $[\mu_A(x)]$, based on F and the extension principle of fuzzy set, one-to-one mapping can be obtained:

$$E : \widetilde{F}(X)/\simeq \longrightarrow D = \{\chi|\chi : R^+ \rightarrow [0,1],$$
$$E([\mu_A(x)]) = E(\mu_(A)(x)) = \chi\} \qquad (5)$$

Above contents can be explanted by Fig. 1.

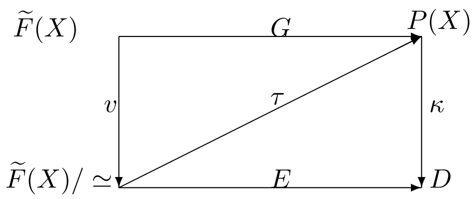

Fig. 1: Commutativity of a diagram of maps between $P(x)$ and D.

Definition 1 *[16] Let* $S = \{s_{\widetilde{i}}|i = 0,\cdots,T\}$ *be a finite set,* $A = \{a_{\widetilde{j_1}}, a_{\widetilde{j_2}}, \cdots,\ a_{\widetilde{j_n}}\} \subseteq S$ *be a set of labels to be aggregated* $(n \leq T)$. $W = \{w_1,w_2,\cdots,w_n\}$ *be a weighting vector such that* $\forall s \in \{1,\cdots,n\}, w_s \in [0,1]$, *and* $\sum_{s=1}^n w_s = 1$. *Let* $B = (j_1,j_2,\cdots,j_n)$, *where* j_s *is the center of* $\widetilde{j_s}$, $C = \sigma(B) = (j_{\sigma(1)},\cdots,j_{\sigma(n)})$ *such that* $j_{\sigma(s')} \geq j_{\sigma(s)}, \forall s' \leq s$ *(*C' *such that* $j_{\sigma(s')} \leq j_{\sigma(s)}, \forall s' \leq s$*), denote* $w = f_{owa}(B) = WC^T = \sum_{s=1}^n w_s j_{\sigma(s)}$, *and* $w' = f_{i-owa}(B) = WC'^T = \sum_{s=1}^n w_s j_{\sigma(s)}$, *then the new (inverse-)linguistic ordered weighted averaging operator* F_{lowa} *(*F_{i-lowa}*)is defined by*

$$F_{lowa}((a_{\widetilde{j_1}}, a_{\widetilde{j_2}}, \cdots, a_{\widetilde{j_n}})) = a_{\widetilde{j_k}}, \qquad (6)$$
$$F_{i-lowa}((a_{\widetilde{j_1}}, a_{\widetilde{j_2}}, \cdots, a_{\widetilde{j_n}})) = a_{\widetilde{j'_k}} \qquad (7)$$

where $a_{\widetilde{j_k}}$ *(*$a_{\widetilde{j'_k}}$*) such that* $\widetilde{j_k}(w) = max\{\widetilde{j_1}(w),\widetilde{j_2}(w),\cdots,\widetilde{j_T}(w)\}(|\ S\ |= T)$*(* $\widetilde{j'_k}(w') = max\{\widetilde{j_1}(w'),\widetilde{j_2}(w'),\cdots,\widetilde{j_T}(w')\}(|\ S\ |= T)$*).*

3. Extracting complex linguistic rules based on operator F_{lowa} and F_{i-lowa}

The so-called complex linguistic rules have the form "*Qy's are* $S_{\widetilde{j_1}}$ *and* $\cdots$ *and* $S_{\widetilde{j_r}}$ *is T*." According to Fig. 1, it is easy to obtain "$Q_{\tilde{j_1}} y's$ *are* $S_{\tilde{j_1}}$ *is* $T_{\tilde{j_1}}$," $\cdots$, and "$Q_{\tilde{j_r}} y's$ *are* $S_{\tilde{j_r}}$ *is* $T_{\tilde{j_r}}$." Hence, for a complex linguistic rule, the problem is how to combine $Q_{\tilde{j_1}}, \cdots, Q_{\tilde{j_r}}$ to obtain Q, and $T_{\tilde{j_1}}, \cdots, T_{\tilde{j_r}}$ to obtain T. Based on operator F_{lowa} and F_{i-lowa}, Q and T can be obtained.

$$Q = F_{lowa}(Q_{\tilde{j_1}}, \cdots, Q_{\tilde{j_r}}), \quad (8)$$

$$T = F_{lowa}(T_{\tilde{j_1}}, \cdots, T_{\tilde{j_r}}). \quad (9)$$

In Eq.(8) and Eq.(9), weighting vectors W and W' can be computed by Yager's linguistic quantifier $Q(r, a, b)$ [22],

$$Q(x, a, b) = \begin{cases} 0, & \text{if} \quad x < a, \\ \frac{x-a}{b-a}, & \text{if} \quad a \le x < b, \\ 1, & \text{if} \quad x \ge b. \end{cases}$$

where $x, a, b \in [0, 1]$. Some examples of $Q(r, a, b)$ are *most*, *at least half* and *as many as possible*, their parameters (a, b) are $(0.3, 0.8)$, $(0, 0.5)$, and $(0.5, 1)$, respectively. By using $Q(r, a, b)$, weighting vectors W and W' are

$$w_i = Q(\frac{i}{r}, a, b) - Q(\frac{i-1}{r}, a, b), i = 1, \cdots, r, \quad (10)$$

$$w'_i = Q(\frac{i}{r}, a', b') - Q(\frac{i-1}{r}, a', b'), i = 1, \cdots, r. \quad (11)$$

Example 2 *Continue Example 1, combining fuzzy linguistic quantifiers {about half, most} and fuzzy linguistic truth degree {approximately true, very true} are needed. select Yager's linguistic quantifier* most, *then weighting vectors* $W = W' = (0.4, 0.6)$, *and* $F_{lowa}(most_{\tilde{30}}, about\ half_{\tilde{18}}) = about\ half_{\tilde{18}}$, $F_{lowa}(very\ true_{\tilde{3}}, approximately\ true_{\tilde{0}}) = approximately\ true_{\tilde{0}}$.

In which, $y = F(x) = 3x$. *Complex linguistic rules based on* F_{lowa} *and* F_{i-lowa} *can be obtained respectively, "about half of employees are young and have high salary is approximately true" and "most of employees are young and have high salary is true."*

4. Optimizing a complex linguistic rule based on genetic algorithms

Genetic algorithms (GA) are search algorithms that use operations found in natural genetics to guide the trek through a search space [23], [24]. A number of papers have been devoted to the automatic generation of the knowledge base of a fuzzy rule-based system (FRBS) using GA [25]-[31].

1) *Optimizing the number and membership functions of linguistic terms based on GA.*

From information systems point of view, a database is information system [32]. During obtaining linguistic rules from a database, how to select the number and membership functions of linguistic terms is a problem. Let there exist L attribute values in V, and each domain of attribute is denoted by $D_l \subset R^+, l = 1, \cdots, L$, then each object $y_i \in Y$ is understood as a point on space $D_1 \times D_2 \times \cdots \times D_L$ [18], *i.e.*, $y_i = (d_{i1}, d_{i2}, \cdots, d_{iL}), d_{il} \in D_l$. Let n' $(n' < n)$ pattern $y_i = (d_{i1}, d_{i2}, \cdots, d_{iL}), i = 1, \cdots, n'$ are given as training patterns from M classes: class 1 (C_1), $\cdots$, class M (C_M). The problem is to generate the number and membership functions of linguistic terms on D_l $(l = 1, \cdots, L)$ that divide the pattern space into M disjoint decision areas. Let each axis D_l of the space $D_1 \times D_2 \times \cdots \times D_L$ be partitioned into K_l fuzzy subsets $\{A^l_{k_l} | k_l = 1, \cdots, K_l\}$, then $D_1 \times D_2 \times \cdots \times D_L$ is divided into $K_1 \times K_2 \times \cdots \times K_L$ fuzzy subspaces, and each fuzzy subspace can be expressed by *If–Then* rule: $R_{k_1 \times \cdots \times k_L}$: If d_{i1} *is* $A^1_{k_1}$ *and* $\cdots$ *and* d_{iL} *is* $A^L_{k_L}$, *then* y_i *belongs to class* C_m *with* $CF = CF_{k_1} \times \cdots \times k_L$. In which $R_{k_1 \times \cdots \times k_L}$ is label of *If–Then* rule, $A^l_{k_l}$ $(l = 1, \cdots, L)$ is fuzzy subset on D_l, C_m $(m = 1, \cdots, M)$ is the consequent, and $CF_{k_1 \times \cdots \times k_L}$ is the grade of certainty of the *If–Then* rule and determined by following procedure,

1. For each class C_m and rule $R_{k_1 \times \cdots \times k_L}$, we have

$$\alpha_{C_m} = \sum_{y_i \in C_m} A^1_{k_1}(d_{i1}) \times \cdots \times A^L_{k_L}(d_{iL}), \quad (12)$$

2. Selecting

$$CF_{k_1 \times \cdots \times k_L} = max\{\alpha_{C_1}, \cdots, \alpha_{C_M}\}. \quad (13)$$

Remark 1 *If* $CF_{k_1 \times \cdots \times k_L} = 0$, *in this case, rule* $R_{k_1 \times \cdots \times k_L}$ *is useless to classify* y_i, *and the consequent of rule* $R_{k_1 \times \cdots \times k_L}$ *is modified by* $C_m = \emptyset$. *If two or more* α_{C_m} *are equal to* $CF_{k_1 \times \cdots \times k_L}$, *then the rule is not good to classify* y_i, *or dividing fuzzy subspaces are not suitable, and the consequent of rule* $R_{k_1 \times \cdots \times k_L}$ *is also modified by* $C_m = \emptyset$.

When a rule set R is given, a new pattern $y' = (d'_{i1}, \cdots, d'_{iL})$ is classified by the follows procedure based on R,

1. Calculate $\gamma_{k_1\times\cdots\times k_L}$ for each rule $R_{k_1\times\cdots\times k_L}$,

$$\gamma_{k_1\times\cdots\times k_L} = A^1_{k_1}(d'_{i1})\times\cdots\times A^L_{k_L}(d'_{iL}) \times CF_{k_1\times\cdots\times k_L}, \quad (14)$$

2. Find class $C_{m'}$ such that

$$\gamma_{C_{m'}} = max\{\gamma_{k_1\times\cdots\times k_L}|R_{k_1\times\cdots\times k_L}\in R\}. \quad (15)$$

If $\gamma_{C_{m'}} = 0$ or $C_{m'} = \emptyset$ of rule, then the classification of y' is rejected, *i.e.*, y' is left as an unclassified pattern, else assign y' to class $C_{m'}$ determined by Eq.(15).

The main components of optimizing the number and membership functions of linguistic terms based on GA is describe as follows [18]:

1) *Encoding the solution*: The two components of the solution to be encoded are the number of linguistic terms and the membership functions of linguistic terms.

1. Number of labels (S_1). In this paper, there are L variables (qualities), the number of labels per variable is stored into an integer array of length L. In this contribution, the possible values considered are the set $\{3,\cdots,9\}$.
2. Membership functions (S_2). In this paper, we deal with triangular functions, a real number array of $L\times 9\times 3$ positions is used to store the membership functions. Of course, if a chromosome does not have the maximum number of labels in one variable, the space reserved for the values of these labels is ignored in the evaluation process.

If s_l is the granularity of variable l $(l=1,\cdots,L)$, $s_l \in \{3,\cdots,9\}$, $P^1_{lj}, P^2_{lj}, P^3_{lj}$ are the definition points of the label j of the variable l, and S_{2l} is the information about the fuzzy partition of variable l in S_2, then a graphical representation of the chromosome is shown as follows:

$$S_1 = (s_1, \quad s_2, \quad \cdots \quad, \quad s_L),$$

$$S_{2l} = (P^1_{l1}, P^2_{l1}, P^3_{l1}, \cdots, P^1_{ls_l}, P^2_{ls_l}, P^3_{ls_l}),$$

$$S_2 = (S_{21}, S_{22}, \cdots, S_{2L}), \quad S = S_1 S_2.$$

Uniform fuzzy partitions are denoted by $(V^1_{lj}, V^2_{lj}, V^3_{lj})$ for each variable. For general fuzzy partition, a variation interval is defined for each one of the membership function definition points [29], *i.e.*, $P^1_{lj}\in[L^1_{lj}, R^1_{lj}] = [V^1_{lj} - \frac{V^2_{lj}-V^1_{lj}}{2}, V^1_{lj} + \frac{V^2_{lj}-V^1_{lj}}{2}]$, $P^2_{lj}\in[L^2_{lj}, R^2_{lj}] = [V^2_{lj} - \frac{V^2_{lj}-V^1_{lj}}{2}, V^2_{lj} + \frac{V^3_{lj}-V^2_{lj}}{2}]$, $P^3_{lj}\in[L^3_{lj}, R^3_{lj}] = [V^3_{lj} - \frac{V^3_{lj}-V^2_{lj}}{2}, V^3_{lj} + \frac{V^3_{lj}-V^2_{lj}}{2}]$.

2) *Initial gene pool*: The initial population is composed of four groups:

1. In the first group, each chromosome will have the same number of labels in all its variables and the membership functions are uniformly distributed across the domain of variable.
2. In the second group, each chromosome can have a different granularity per variable (different values in S_1) and the membership functions are uniformly distributed as in the first part.
3. In the third group, each chromosome will have the same number of labels in all its variables. Then a uniform fuzzy partition is built for each variable as in the first group and the variation intervals of all the definition points are calculated. Finally, a value for all the definition points is randomly chosen from the correspondent variation interval.
4. In the last group, each chromosome can have different number of labels per variable as in second group and the membership functions are calculated in the same way as in the third group, a random value is in the variation interval.

3) *Evaluating the chromosome*: Each chromosome represents a kind of fuzzy classification on $D_1\times D_2\times\cdots\times D_L$. Our problem is to obtain optimal solution which is to maximize the number of correctly classified pattern and to minimize the number of *If–Then* rule. This problem can be formulated as following two-objective combinatorial optimization problem,

$$\text{Minimize}: f(s) = \omega_1 DCP(s) + \omega_2 \mid s \mid, \quad (16)$$

Where s is a chromosome, $DCP(s)$ is the number of unclassified patterns by s, $\mid s \mid$ is the number of *If–Then* rules in s. In general, the classification power of s is more important then its compactness, therefor the weights in Eq.(16) should be specified as $0 < \omega_2 \ll \omega_1$ [27]. The objective function $f(s)$ is treated as the fitness function in GA.

4) *Genetic operators*: Since there is a strong relationship among the two chromosome parts, operators working cooperatively in S_1 and S_2 are required in order to make best use of the representation used.

a) *Selection*: Let current population Ψ. The selection probability $P(s)$ of chromosome s is

$$P(s) = \frac{(f_{max}(\Psi) - f(s))}{\sum_{s'\in\Psi}(f_{max}(\Psi) - f(s'))}, \quad (17)$$

in which $f_{max}(\Psi) = max\{f(s)|s\in\Psi\}$.

b) *Crossover*: Two different crossover operators are considered depending on the two parents' scope [26], Crossover when both parents have the same granularity level per variable, in this case, crossover operator in S_2 and obviously, by maintaining the parent S_1 values in the offspring. If $(S_2^v)^t = ((P_{11}^1)^v, \cdots, (P_{Ls_L}^3)^v)$ and $(S_2^w)^t = ((P_{11}^1)^w, \cdots, (P_{Ls_L}^3)^w)$ are to be crossed, the follows four offspring are generated, in which, $i = 1, 2, 3$,

$$\begin{aligned}(S_2^{vw})_1^{t+1} &= d(P_{1l}^1)^{vw}, \cdots, (p_{Ls_L}^3)^{vw}),\\ (P_{ls_l}^i)^{vw} &= d(P_{ls_l}^i)^v + (1-d)(P_{ls_l}^i)^w,\\ (S_2^{vw})_2^{t+1} &= d((P_{11}^1)^{vw}, \cdots, (P_{Ls_L}^3)^{vw}),\\ (P_{ls_l}^i)^{vw} &= (1-d)(P_{ls_l}^i)^v + d(P_{ls_l}^i)^w,\\ (S_2^{vw})_3^{t+1} &= d((P_{11}^1)^{vw}, \cdots, (P_{Ls_L}^3)^{vw}),\\ (P_{ls_l}^i)^{vw} &= max\{(P_{ls_l}^i)^v, (P_{ls_l}^i)^w\},\\ (S_2^{vw})_4^{t+1} &= d((P_{11}^1)^{vw}, \cdots, (P_{Ls_L}^3)^{vw}),\\ (P_{ls_l}^i)^{vw} &= min\{(P_{ls_l}^i)^v, (P_{ls_l}^i)^w\},\end{aligned}$$

This operator uses a parameter that is either a constant, or a variable whose value depends on the age of the population [26]. The resulting descendants are the two best of the four aforesaid offspring.

Crossover when the parents encode different granularity levels. Let

$$S^v = ((s_1)^v, \cdots, (s_l)^v, (s_{l+1})^v, \cdots, (s_L)^v, (S_{21})^v,$$

$$\cdots, (S_{2l})^v, (S_{2(l+1)})^v, \cdots, (S_{2L})^v),$$

$$S^w = ((s_1)^w, \cdots, (s_l)^w, (s_{l+1})^w, \cdots, (s_L)^w, (S_{21})^w,$$

$$\cdots, (S_{2l})^w, (S_{2(l+1)})^w, \cdots, (S_{2L})^w)$$

be crossed at point l, the two resulting offspring are,

$$S_1^{vw} = ((s_1)^v, \cdots, (s_l)^v, (s_{l+1})^w, \cdots, (s_L)^w, (S_{21})^v,$$

$$\cdots, (S_{2l})^v, (S_{2(l+1)})^w, \cdots, (S_{2L})^w),$$

$$S_2^{vw} = ((s_1)^w, \cdots, (s_l)^w, (s_{l+1})^v, (s_L)^v, (S_{21})^w,$$

$$\cdots, (S_{2l})^w, \cdots, (S_{2l})^w, (S_{2(l+1)})^v, \cdots, (S_{2L})^v).$$

c) *Mutation*: Two different operators are used,

1. Mutation on S_1, in this case, once a new value $s_l' \in \{3, \cdots, 9\}$ at point l of S_1 is selected, a uniform fuzzy partition for this variable is stored in its corresponding zone of S_2.
2. Mutation on S_2: Let $(S_2^v)^t = ((P_{11}^1)^v, \cdots, (P_{ls_l}^i)^v, \cdots, (P_{Ls_L}^3)^v)$ and the element $(P_{ls_l}^i)^v$ was selected for this mutation (the domain of $(P_{ls_l}^i)^v$ is $[(P_{ls_l}^i)_l^v, (P_{ls_l}^i)_r^v])$, the result is a vector $(S_2^v)^{t+1} = ((P_{11}^1)^v, \cdots, ((P_{ls_l}^i)^v)', \cdots, (P_{Ls_L}^3)^v)$, and

$$((P_{ls_l}^i)^v)' = \begin{cases} (P_{ls_l}^i)^v + \Delta(t, (P_{ls_l}^i)_r^v - (P_{ls_l}^i)^v), \\ \quad \text{if} \quad e = 0, \\ (P_{ls_l}^i)^v + \Delta(t, (P_{ls_l}^i)^v - (P_{ls_l}^i)_l^v), \\ \quad \text{if} \quad = 1. \end{cases}$$

with t being the current generation, e a random number that may have a value of zero or one, and the function $\Delta(t, y)$ [24] is

$$\Delta(t, y) = y(1 - r^{(1-\frac{t}{T})^b}),$$

with r being a random number in the interval $[0, 1]$, T the maximum number of generations and b a parameter chosen by the user.

2) *Obtaining a complex linguistic rule with higher truth degree based on GA.*

In some cases, if all $Q_{\widetilde{j_{r'}}}$ and $T_{\widetilde{j_{r'}}} (r' = 1, \cdots, r)$ are used to obtain Q and T, respectively, then the truth degree of the complex linguistic rule is low. From the real-world practice point of view, a complex linguistic rule with lower truth degree is useless. To solve this problem, parts of $Q_{\widetilde{j_{r'}}}$ and $T_{\widetilde{j_{r'}}} (r' = 1, \cdots, r)$ are selected to obtain Q and T, respectively. Due to $Q_{\widetilde{j_{r'}}}$ and $T_{\widetilde{j_{r'}}} (r' = 1, \cdots, r)$ can be decided by each other, discussion is based on $T_{\widetilde{j_{r'}}} (r' = 1, \cdots, r)$.

1) *Encoding the solution*: The solution to be encoded is truth degree of each simple linguistic rule. The coding scheme generates fixed-length r chromosomes, a graphical representation of the chromosome is as follows: $\forall r' \in \{1, \cdots, r\}, t_{r'} \in \{0, 1\}$,

$$S = t_1 t_2 \cdots t_r, \tag{18}$$

in which if $t_{r'} = 0$, it means that the truth degree $T_{\widetilde{j_{r'}}}$ at point $t_{r'}$ does not take part in aggregation, otherwise, $T_{\widetilde{j_{r'}}}$ takes part in aggregation.

2) *Initial gene pool*: According to Eq.(18), there are 2^r solutions, as each chromosome is encoded as a binary coded GA, the initial population is randomly selected as usualness.

3) *Evaluating the chromosome*: Our aim is to obtain a complex linguistic rule with higher truth degree. For a solution $s = t_1 t_2 \cdots t_r$ and fixed Yager's linguistic quantifier $Q(r, a, b)$, denote

$$(T_{\widetilde{j}}, \alpha_j) = (F_{lowa}(T_{\widetilde{j_1}}, \cdots, T_{\widetilde{j_r}}), \widetilde{j}(w)), \tag{19}$$

in which, $T_{\widetilde{j_{r'}}}$ such that $t_{r'} = 1$, w is decided by Definition 1. According to Eq.(19), fitness function is obtained as follows:

$$\text{Maximize}: f(s) = \xi_1 j + \xi_2 \alpha_j. \quad (20)$$

In which, $\xi_1 + \xi_2 = 1$, ξ_1 and ξ_2 (decided by expert or user) express important degree of j and α_j, respectively. $f(s)$ means higher truth degree and it's membership degree.

4) *Genetic operators*: Let current population be Ψ. The selection probability $P(s)$ of chromosome s is

$$P(s) = \frac{(f(s) - f_{min}(\Psi))}{\sum_{s' \in \Psi}(f(s') - f_{min}(\Psi))}, \quad (21)$$

in which $f_{min}(\Psi) = min\{f(s)|s \in \Psi\}$, Let the index of $f_{min}(\Psi)$ be j', then

$$f(s) - f_{min}(\Psi) = \xi_1(j - j') + \xi_2(\alpha_j - \alpha_{j'}). \quad (22)$$

Based on these genetic operators and fitness function Eq.(20), the optimal solution can be obtained, *i.e.*, a complex linguistic rule with higher truth degree and it's membership degree is obtained.

5. Conclusion

From the formalization point of view, a linguistic data summary is equal to a fuzzy rule with fuzzy linguistic quantifier and truth degree. Based on linguistic ordered weighted averaging operator F_{lowa} and F_{i-lowa}, the method to extract Q, S and T of a complex linguistic rule is discussed. Based on GA, how to select the number and membership functions of linguistic terms are discussed during obtaining linguistic rules from a database.

Acknowledgement

This work is supported by the Excellent Young Foundation of Sichuan Province (Grant No.06ZQ026-037), the National Natural Science Foundation of China (60474022), the Special Research Funding to Doctoral Subject of Higher Education Institutes in China (Grant No.20060613007) and a Project Supported by Scientific Research Fund of Sichuan Provincial Education Department(Grant No.2005A121, 2006A084).

References

[1] R.R. Yager, A new approach to the summarization of data. *Information Sciences*, 28: 69-86, 1982.

[2] R.R. Yager, Database discovery using fuzzy sets. *International Journal of Intelligent Systems*, 11:691-712, 1996.

[3] R.R. Yager, J. Kacprzyk, Linguistic data summaries: A perspective. *in Proceedings of IFSA'99 Congress (Taipei, Taiwan R.O.C.)*, pp. 44-48, 1999 .

[4] G. Raschia, N. Mouaddib, SAINTETIQ: a fuzzy set-based approach to database summarization. *Fuzzy Sets and Systems*, 129:137-162, 2002.

[5] R. George, R. Srikanth, Data summarization using genetic algorithms and fuzzy logic. In F. Herrera and J.L. Verdegay (Eds.): *Genetic Algorithms and Soft Computing*, Physica-Verlag, Heidelberg, pp: 599-611, 1996.

[6] L. A. Zadeh, Fuzzy logic = computing with words. *IEEE Trans. Fuzzy Systems*, 4:103-111, 1996.

[7] L. A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in houman reasoning and fuzzy logic. *Fuzzy Sets and Systems*, 90:103-111, 1997.

[8] P. P. Wang, Computing with words, John Wiley and Sons, 2001.

[9] J. Kacprzyk, S. Zadrożny, Computing with words in intelligent database querying: standalone and Internet-based applications. *Information Sciences*, 34:71-109, 2001.

[10] J. Kacprzyk, R.R. Yager, Linguistic summaries of data using fuzzy logic. *International Journal of General Systems*, 30: 133-154, 2001.

[11] F. Herrera, E. Herrera-Viedma, Aggregation operators for linguistic weighted information. *IEEE Trans. System, Man, Cybernet. -Part A: Systems Humans*, 27:646-656, 1997.

[12] F. Herrera, E. Lopez, M.A. Rodriguez, A linguistic decision model for promotion mix management solved with genetic algorithms. *Fuzzy Sets and Systems*, 131:47-61, 2002.

[13] O. Cordón, F. Herrera, I. Zwir, Linguistic modeling by hierarchical systems of linguistic rules. *IEEE Transactions on Fuzzy Systems*, 10(1) :2-19, 2002.

[14] O. Cordón, F. Herrera, Hybridizing genetic algorithms with sharing scheme and evolution strategies for designing approximate fuzzy rule-based systems. *Fuzzy Sets and Systems*, 118:235-255, 2001.

[15] Z. Pei, K. Qin, Obtaining Decision Rules

and Combining Evidence based on modal logic. *Progress in Natural Science (in chinese)*, 14(5):501-508, 2004.

[16] Z. Pei, Y. J. Du, L. Z. Yi and Y. Xu, *Obtaining a complex linguistic data summaries from database based on a new linguistic aggregation operator, J. Cabestany, A. Prieto, and D.F. Sandoval (Eds.): IWANN 2005, LNCS 3512*, Springer-Verlag Berlin Heidelberg, pp. 771-778, 2005.

[17] Z. Pei, Y. Xu, Lattice implication algebra model of linguistic variable Truth and its inference, in Da Ruan et al (eds.). *Applied Computational Intelligence, World Scientific*, pp. 93-98, 2004.

[18] Z. Pei, Y. Xu, D. Ruan, K. Qin, Extracting a complex linguistic data summaries from perssonnel database via linguistic aggregation of simple ones, *Information Sciences* (to appear).

[19] V. Novák, Antonyms and linguistic quantifiers in fuzzy logic. *Fuzzy Sets and Systems*, 124:335-351, 2001.

[20] A. Dvořák and V. Novák, Formal theories and linguistic descriptions. *Fuzzy Sets and Systems*, 143:169-188, 2004.

[21] N. C. Ho, T. D. Khang, V. N. Huynh, An algebraic approach to linguistic hedges in Zadeh¡¯s fuzzy logic. *Fuzzy Sets and Systems*, 129:229-254, 2002.

[22] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Trans.Systems,Man Cybernet*, 18:183-190, 1988.

[23] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* , Addison-Wesley, New York, 1989.

[24] Z. Miehalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Vienna, Austria: Springer-Verlag, 1996.

[25] J. Casillas, O. Cordón, M. J. Del Jesus, F. Herrera, Genetic feature selection in a fuzzy rule-based classification system learning process for high-dimensional problems. *Information Sciences*, 136: 135-157, 2001.

[26] F. Herrera, M. Lozano, J. L. Verdegay, Fuzzy connectives based crossover operators to model genetic algorithms population diversity. *Fuzzy Sets and Systems*, 92:21-30, 1997.

[27] H. Ishibuchi, K. Nozaki, N. Yamamoto, H. Tanaka, Selecting fuzzy if-then rules for classification problems using genetic algorithms.*IEEE Transactions on Fuzzy Systems*, 3(3):260-270, 1995.

[28] O. Cordón, F. Herrera, L. Magdalena, P. Villar, A genetic learning process for the scaling factors, granularity and contexts of the fuzzy rule-based system data base. *Information Sciences*, 136:85-107, 2001.

[29] O. Cordón, F. Herrera, P. Villar, Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base.*IEEE Transactions on Fuzzy Systems*, 9(4):667-674, 2001.

[30] F. Herrera, M. Lozano, J. L. Verdegay, A learning process for fuzzy control rules using genetic algorithms. *Fuzzy Sets and Systems*, 100:143–158, 1998.

[31] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, L. Magdalena, Ten years of genetic fuzzy systems: current framework and new trends.*Fuzzy Sets and Systems*, 141:5-31, 2004.

[32] Z. Pei, K. Qin. Intuitionistic special set expression of rough set and its application in reduction of attributes. *Pattern Recognition and Artificial Intelligence(in chinese)*, 17(3): 262-266, 2004.