Gathering Linguistic Information in Distributed Intelligent Agent on the Internet

Zheng Pei[1] Da Ruan[2] Yang Xu[3]

[1]School of Computers & Mathematical-Physical Science, Xihua University, Chengdu, Sichuan, 610039, China
[2]Belgian Nuclear Research Centre (SCK•CEN), Boeretang 200, B-2400 Mol, Belgium
[3]Department of Mathematics, Southwest Jiaotong University, Chengdu, Sichuan 610031, China

Abstract

Much information is expressed by natural languages. The management of linguistic information implies the use of operators of comparison and aggregation. Based on the Ordered Weighted Averaging (*OWA*) operator by Yager and modifying indexes of linguistic terms (their indexes are fuzzy numbers on $[0,T]\subseteq R^+$), a new linguistic weighted averaging operator (F_{lwa}) is proposed. Moreover from the distributed intelligent agent architecture on the Internet, information flows are discussed. By using the operator F_{lwa}, a new method of gathering linguistic weighted information on the Internet is presented.

Keywords: *OWA* operator, linguistic weighted averaging operator, intelligent agent, Internet.

1. Introduction

The fuzzy linguistic approach is an approximate technique that represents qualitative aspects as linguistic values by means of linguistic variables, *i.e.*, variables whose values are not numbers but words or sentences in a natural or artificial language [1]. In representation and manipulation of human knowledge automatically, many methods based on fuzzy set theory have already been successful developed [2]. However, at present, there is no uniformity in the construction of membership function and the interpretation of what a membership grade means, and many researchers have studied these problems [3], [4]. On the other hand, in some real world practice, it is very difficult to select suitable membership functions of fuzzy linguistic values, in many cases, human knowledge is expressed by directly using natural languages.

The management of linguistic information implies the use of operators of comparison and aggregation. Many researchers have studied operators of comparison and aggregation [5]-[11]. The fuzzy linguistic approach has been successfully applied to many aspects, such as decision, information retrieval, medicine, and education etc. In the framework of information retrieval, one of the most current problems for which a linguistic approach is very useful is the retrieval, handling, and identification of relevant information on Internet [5].

Intelligent agents (or Intelligent software agents), deal with the information gathering process assisting Internet users to locate the information that best fits their needs. In intelligent agents, it is a problem that the lack of connection, communication, and consensus among them has led to a decrease in the quality and suitability of the information retrieved in addition to the efficiency of the system in the recovery and filtering task. The great variety of representations and evaluations of information on Internet is the main obstacle to this communication, and the problem becomes more noticeable when a user takes part in the process. The complexity of this process reveals the need for more flexibility in communication among agents and between agents and the user. For this purpose, several approaches related to mechanisms to introduce and handle flexible information through linguistic labels have been proposed at both the agent and user levels [5], [12].

This paper is structured as follows: In Section 2, fuzzy linguistic values are briefly listed, modifying the index of linguistic label as a fuzzy number on $[0,T]$ is proposed, and based on fuzzy number index, the operator F_{lwa} is proposed. In Section 3, a distributed intelligent agent model on the Internet and information gathering are illustrated. In Section 4, information gathering through the operator F_{lwa} is discussed. Conclusion is Section 5.

2. Linguistic values

From a formal point of view, a linguistic variable is defined by a quintuple $\{L, H(L), U, G, M\}$ [1]. L is the name of the linguistic variable, and $H(L)$ (or simply H) is its linguistic term set, *i.e.*, the collection of linguistic descriptors used to assess. For example, a linguistic term set for the linguistic variable **Importance**, denoted by *H(Importance)*, can be defined as: H(Importance)= {high, very high, medium, low, more or less low, not very low, …, }. A numerical variable *u* called the base variable, which takes values in the universe of discourse *U* (usually [0,1]), is associated with each linguistic term of *H*. *G* is a syntactic rule, which usually takes the form of a context free grammar and generates the linguistic descriptors of *H*, and *M* is a semantic rule which associates the meaning *M(l)* with a fuzzy subset of *U* to each linguistic value $l\in H$. Each *l* generated by *G* is the label for the fuzzy restriction (membership function) *M(l)* on the values of the base variable, *i.e.*, $M(l)= \{(u, \mu_l(u)) \mid u\in U, \mu_l(u) \in[0,1]\}$. Hence, each linguistic value is characterized by a **syntactic value** (or **label**) and a **semantic value** (or **meaning**). The **label** is a word or sentence belonging to a linguistic term set, and the **meaning** is a fuzzy subset in a universe of discourse.

From the information granule (IG) point of view [13], we know that each linguistic label is a fuzzy IG that is a clump of physical or mental objects (points) drawn together by indistinguishability, similarity, proximity or functionality.

From the algebra point of view, let *U* be a universal set. All granules of *U* denote *P*(*U*)(= *H(L)*), the set $R^+=\{x|x\geq 0\}$ denotes indexes of *H(L)*, then the linear function F can be defined as follows [14]:

$$F: U \to R^+ \qquad (1)$$

Let $\mu(A)$ be a membership function of a granule $A\in P(U)$. Generally, the granule *A* has many membership functions that are defined by different researchers. Let $\tilde{F}(U)$ be the collection of membership functions on *U*. Natural mapping is

$$\tau: \tilde{F}(U) \to P(U) \qquad (2)$$

An equivalence relation $\cong$ on $\tilde{F}(U)$ can be obtained,

$$\mu(u) \cong \mu'(u) \Leftrightarrow \tau(\mu(u)) = \tau(\mu'(u)) \qquad (3)$$

i.e., $\mu(u)$ and $\mu'(u)$ are membership functions of the same granule of *P*(*U*). Let an equivalence class $[\mu_A(u)] \in \tilde{F}(U)/\cong$, for a representative element $\mu_A(u)$, according to *F* and the extension principle, one-to-one mapping can be obtained

$$E: \tilde{F}(U)/\cong \to D=\{\chi \mid \chi = \mu_A F^{-1} : R^+ \to [0,1]\},$$

$$E([\mu_A(u)]) = E(\mu_A(u)) = \chi. \qquad (4)$$

χ is a fuzzy number on R^+, in this paper, the index of granule $A\in P(U)$ is identified by χ. If the granule *A* is crisp, then $\mu_A(u)$ is the characteristic function of crisp set *A*, where χ is also a characteristic function of a crisp set on R^+. Using the natural numbers theory, the indexes of crisp granules are the natural numbers that can be obtained. If the granule *A* is fuzzy, where χ is a fuzzy number on R^+, *i.e.*, the indexes of fuzzy granules ought to be a fuzzy number, this coincides with intuition understanding fuzzy granule. Commutativity of a diagram of maps between *P*(*U*) and *D* is shown in Fig.1, and $\kappa = E\varphi^{-1}$.

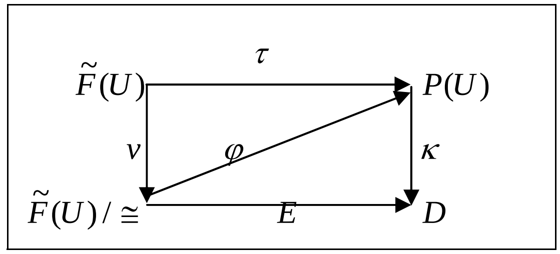

Fig. 1: Diagram of maps between *P*(*U*) and *D*.

Example 1 [8] Let linguistic labels be *P* (perfect), *VH* (very-high), *H* (High), *M* (Medium), *L* (Low), *VL* (Very-low), their membership functions are triangular fuzzy sets on [0,1], *i.e.*, (1, 0.25, 0), (0.75, 0.15, 0.25), (0.6, 0.1, 0.25), (0.5, 0.1, 0.1), (0.4, 0.15, 0.1) and (0.25, 0.25, 0.15), respectively. Where the first component is the center of the triangular fuzzy set, the second component is left width, the third component is right width. Let linear function F be

$$y = F(u) = 6u \qquad (5)$$

Correspondingly, indexes of linguistic labels are $P_{\tilde{6}} = P_{(6,1.5,0)}$, $VH_{\tilde{4.5}} = VH_{(4.5,0.9,1.5)}$, $H_{\tilde{3.6}} = H_{(3.6,0.6,1.5)}$, $M_{\tilde{3}} = M_{(3,0.6,0.6)}$, $L_{\tilde{2.4}} = L_{(2.4,0.9,0.6)}$, $VL_{\tilde{1.5}} = VL_{(1.5,1.5,0.9)}$.

Definition 1 [14] Let

$$P(U) = H(L) = \{s_{\tilde{i}} \mid i = 0,\cdots,T \in R^+\},$$

be a finite label sets, $A=\{a_{\tilde{j}_1},\cdots,a_{\tilde{j}_n}\} \subseteq P(U)$ a set of labels to be aggregated, $W=\{w_1,\cdots,w_n\}$ a weighting vector such that

- $\forall w_i \in W, i = 1,\cdots,n, w_i \in [0,1]$,
- $\sum_{i=1}^{n} w_i = 1$.

Let $B=\{j_1,\cdots,j_n\}$, j_i be the center of $\tilde{j}_i$, $C=\sigma(B)=\{j_{\sigma(1)},\cdots,j_{\sigma(n)}\}$ such that $\forall i' \leq i$, $j_{\sigma(i')} \geq j_{\sigma(i)}$, denote

$$w = f_{owa}(B) = WC^T = \sum_{i=1}^{n} w_i j_{\sigma(i)},$$

then the linguistic ordered weighted averaging operator F_{lwa} is

$$F_{lwa}((a_{\tilde{j}_1},\cdots,a_{\tilde{j}_n})) = a_{\tilde{j}}$$

and $a_{\tilde{j}}$ such that

$$\tilde{j}(w) = \max\{\tilde{j}_1(w),\cdots,\tilde{j}_n(w)\}$$

Remark 1 There exists a bijective between $a_{\tilde{j}}$ and $\tilde{j}$, hence, aggregating linguistic labels A is transformed by aggregating the indexes of the linguistic labels. After obtaining w by using the *OWA* operator, similar to defuzzification methods in fuzzy control theory, $a_{\tilde{j}}$ can be obtained. The other aggregating methods based on fuzzy number indexes and their properties will be discussed elsewhere.

3. Distributed agents structure and communication among agents

Intelligent agents on the Internet not only retrieve and filter information (in the sense of Web documents), but also handle electronic mail, news lists, FAQ lists, and so on. Distributed agents structure is shown in Fig.2. In this paper, a single-user is considered.

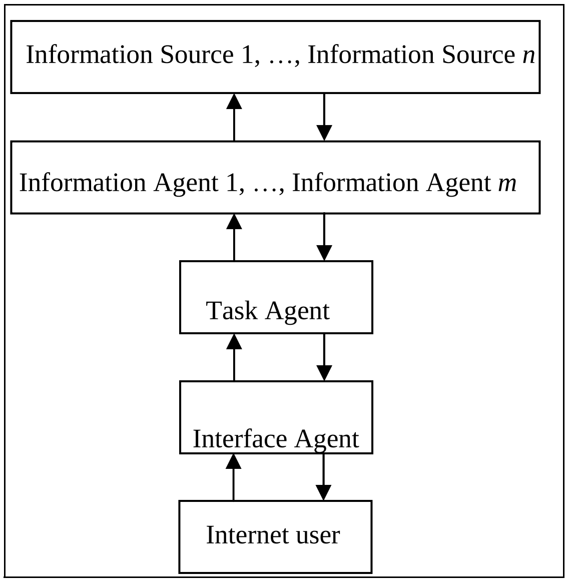

Fig. 2: Distributed agents' structure.

Actions of these levels are showed as follows [5]:

- Internet Users, who look for Web documents on the Internet by means of a weighted query where a set of terms related to the documents desired is specified.
- Interface Agents (one for the user, generally), which communicate the user's weighted query to the Task Agents and filter the documents retrieved from Task Agents to give users those that best satisfy their needs.
- Task Agents (one for the Interface Agent, generally), which communicate the user's query to the Information Agents and get those documents from every Information Agent that best fulfills the query, fusing them and resolving the possible conflicts among Information Agents.
- Information Agents, which receive the weighted query from the Task Agents, look for the information in the data sources, and give the retrieved documents back to the preceding level.
- Information Sources, consisting of all data sources within the Internet, such as databases and information repositories.

In the process of information gathering, as a response to a user's query on the model Fig. 2, there are two parts,

- On one hand, there is a communication between agents at information source and information agent, or information agent and task agent, which is far from the user's participation and where the question that the Task Agent must answer concerns which Information Agents would better satisfy the user's needs.
- On the other hand, there is a communication between agents at levels interface agent and the user, where the information element is specifically the set of retrieved documents that will be analyzed and filtered by the Interface Agents.

4. Information gathering through the operator F_{lwa}

Assume that linguistic labels set S has been obtained, according to Fig. 2. The process of information gathering is based on the weighted queries that are given by an internet user and communication among levels in Fig. 2, the process of information gathering can be described as follows:

- Internet User makes a query to look for documents related to the terms $\{t_{\tilde{1}}, t_{\tilde{2}}, \cdots, t_{\tilde{n}}\}$, which are weighted by $W_1 = \{w_1, \cdots, w_n\}$. These values are communicated to the Interface Agent.
- Interface Agent gives the terms and their importance weights to the Task Agent.
- Task Agent makes the query to all the Information Agents, the query is connected to the terms $\{t_{\tilde{1}}, t_{\tilde{2}}, \cdots, t_{\tilde{n}}\}$ and W_1.
- All Information Agents that have received the query look for the information that best satisfies it in the Information Sources, and retrieve from them the documents.
- Task Agent receives a set of documents from Information Agent h ($h = 1, \cdots, m$). Every document h has an associated term $\{c^h_{\tilde{1}}, c^h_{\tilde{2}}, \cdots, c^h_{\tilde{r}_h}\}$ which is weighted by $W_h = \{w_{h1}, \cdots, w_{hr_h}\}$
- Task Agent aggregates information obtained from every Information Agent h by the operator F_{lwa}, *i.e.*, for $W_h = \{w_{h1}, \cdots, w_{hr_h}\}$,
  $$F_{lwa}((c^h_{\tilde{1}}, \cdots, c^h_{\tilde{r}_h})) = c^h_{\tilde{k}_h},$$
  where, $\tilde{k}_h$ such that
  $$w_h = f_{owa}((1, 2, \cdots, r_h))$$
  $$\tilde{k}_h(w_h) = \max\{1(w_h), \cdots, \tilde{r}_h(w_h)\}$$
- According to weights W_1, Interface Agent aggregates the terms $\{t_{\tilde{i}_1}, t_{\tilde{i}_2}, \cdots, t_{\tilde{i}_n}\}$, *i.e.*,
  $$F_{lwa}((t_{\tilde{1}}, \cdots, t_{\tilde{n}})) = t_{\tilde{l}}$$
- Once Task Agent has calculated the overall performances $c^h_{\tilde{k}_h}$ of Information Agents

through the operator F_{lwa}, Task Agent must be decided which Information Agent best fulfills the user's query. Generally, the performances $c^h_{\tilde{k}_h}$ ($h = 1, \cdots, m$) are ordered decreasingly by a criterion. In this paper, the criterion is selected as follows:

$$d_h = \left| t_{\tilde{l}} - c^h_{\tilde{k}_h} \right| \equiv \left| \tilde{l} - \tilde{k}_h \right|.$$

In which, d_h is the measure between the fuzzy number $\tilde{l}$ and $\tilde{k}_h$.

- Interface Agent receives an ordered list of documents from Task Agent, and sends them to Internet User.

5. Conclusion

In this paper, on one hand, based on the *OWA* operator and modifying indexes of linguistic terms, a new linguistic weighted averaging operator (F_{lwa}) is presented. On the other hand, based on the distributed intelligent agent architecture on the Internet, information flows are discussed. By using the operator F_{lwa}, a new method of gathering linguistic information on the Internet is presented. The other aggregating methods based on the fuzzy number indexes, their properties and information filtering based on linguistic aggregating methods will be discussed elsewhere.

6. Acknowledgements

The work is supported by the National Natural Science Foundation of China (Grant no. 60474022).

7. References

[1] L. A. Zadeh, "The concept of a linguistic variable and its applications to approximate reasoning Part I, II, III," *Information Sciences*, 8: pp. 199-249, 1975; *Information Sciences*, 8: pp. 301-357, 1975; *Information Sciences*, 9: pp. 43-80, 1975.

[2] G. J. Klir, B. Yuan, "Fuzzy Sets and Fuzzy Logic: Theory and Applications", *Prentice-Hall PTR, Upper Saddle River, NJ*, 1995.

[3] G. Resconi, G. J. Klir, D. Harmanec, U. S. Clair, "Interpretations of various uncertainty theories using models of modal logic: a summary," *Fuzzy Sets and Systems*, 80: pp. 7-14, 1996.

[4] D. Dubois, H. Prade, "The three senmantics of fuzzy sets," *Fuzzy Sets and Systems*, 90: pp. 141-150, 1997.

[5] P.P. Wang, "Computing with Words," *John Wiley and Sons, Inc*, 2001.

[6] Zheng Pei and Yang Xu, "Lattice implication algebra model of linguistic variable Truth and its inference," edited by Da Ruan et al., *Applied Computational Intelligence, World Scientific*, pp. 93--98, 2004.

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

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

[9] V. Novak, "Antonyms and linguistic quantifiers in fuzzy logic," *Fuzzy Sets and Systems*, 124: pp. 335-351, 2001.

[10] A. Dvorak and V. Novak, "Fromal theories and linguistic descriptions," *Fuzzy Sets and Systems*, 143: pp. 169-188, 2004.

[11] 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: pp. 229-254, 2002.

[12] K. Sycara, A. Pannu, M. Williamson, D. Zeng, "Distributed intelligent agents," *IEEE Expert*, 11 (6): pp. 36-46, 1996.

[13] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets and Systems*, 90: pp. 111-127, 1997.

[14] Zheng Pei, Yajun Du, Liangzhong Yi, Yang Xu, Obtaining a complex linguistic data summaries from database based on a new linguistic aggregation operator, IWANN2005, LNCS-Springer proceedings, Spain, June 8-10, 2005. (accepted)