A New Method for Fuzzy-number Query Operators of Document Retrieval

Shi-Jay Chen & You-Yi Shih

Department of Information Management

sjchen@nuu.eud.tw, m0333001@smail.nuu.edu.tw

Abstract. In this paper, we firstly point out some drawbacks of the existing fuzzy-number AND/OR query operators based on averaging operators. Then we propose a new query operators based on the extended quadratic-mean operators to overcome these drawbacks of the existing methods. The proposed novel generalized fuzzy-number AND/OR query operators are more useful than the existing methods for fuzzy document retrieval.

Keywords: AND/OR query operators, fuzzy-number, generalized fuzzy numbers

INTRODUCTION

In [1], [2], [7], we can understand that the AND/OR query operators based on averaging operations (e.g., GMA operators [27], Infinite-One operators [3], P-Norm operators [8], T-operators [25], Waller-Kraft operators [21] and WPMA operators [20]) are used to deal with fuzzy document retrieval. In [11], Miyamoto pointed out that linguistic queries processing is an important research topic in fuzzy information retrieval. Recently, some papers deal with linguistic queries processing problems by using fuzzy numbers [1,6,12-18] and it's more flexibility by using generalized fuzzy numbers to represent linguistic queries for dealing with linguistic queries processing problems.

In this research, we present novel generalized fuzzy-number AND/OR query operators based on extended quadratic-mean operators for handling linguistic queries by generalized fuzzy numbers. The proposed extended QMA operators can deal with fuzzy-number queries in a more flexible and more intelligent manner.

Finally, we briefly review the AND/OR query operators [19] and use some examples to illustrate how to use the AND/OR query operators for dealing with fuzzy document retrieval of the AND/OR operations.

PRELIMINARY

In this section, we briefly review the concepts of quadratic mean [1], generalized fuzzy numbers [23], [27]24, the ranking method for generalized trapezoidal fuzzy numbers [22], AND/OR query operators based on quadratic-mean averaging operators [30], and the existing AND/OR query operators based on fuzzy numbers for information retrieval [24], [29].

ISBN 978-967-0850-51-1

Quadratic Mean

The quadratic mean is the square root of the average of sum of square. The quadratic mean of positive numbers a_1, a_2, …, and an is defined as

$$\sqrt{\frac{\sum_{i=1}^{n} a_i^2}{n}}, \tag{1}$$

where 1 ≦ i ≦ n from [10].

The Generalized Fuzzy Numbers

Fig. 1 shows a generalized trapezoidal fuzzy number $\tilde{A}$ = (a1, a2, a3, a4; $\hat{w}_{\tilde{A}}$), where 0 < $\hat{w}_{\tilde{A}}$ ≦ 1 and a1, a2, a3, and a4 are positive real numbers that a1 ≦ a2 ≦ a3 ≦ a4 [23]. The generalized trapezoidal fuzzy number $\tilde{\tilde{A}}$ can be denoted a decision-maker's opinion. The value $\hat{w}_{\tilde{A}}$ means the degree of confidence of the decision-maker's opinion.

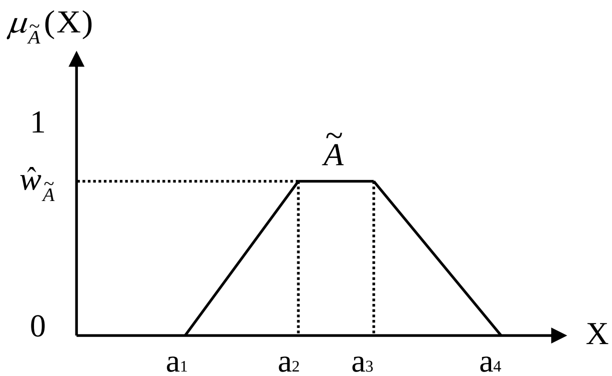

FIGURE 1. A trapezoidal generalized fuzzy number

The Ranking Method for Generalized Trapezoidal Fuzzy Numbers

In [28], the approach for ranking generalized trapezoidal fuzzy numbers and interval-valued fuzzy numbers. Assume that there is a generalized trapezoidal fuzzy numbers $\tilde{A}$ where $\tilde{A}$ = (a_1, a_2, a_3, a_4; $\hat{w}_{\tilde{A}}$), and 0 < $\hat{w}_{\tilde{A}}$ ≦ 1. The ranking value $R(\tilde{D})$ of the generalized trapezoidal fuzzy numbers $\tilde{A}$, shown as follows:

$$R(\tilde{D}) = \frac{1}{8}[a_1 + a_2 + a_3 + a_4 + 4a_1 + 2a_2 + 2a_3 + 4a_4 + 3(a_2 - a_1 - a_4 + a_3)]. \tag{2}$$

The larger the value $R(\tilde{D})$, the better the ranking of $\tilde{A}$.

Review of Some Existing AND/OR Query Operators Based on Fuzzy Numbers

In the following, we briefly review existing averaging operators from [24], [29], as shown as follows:

ISBN 978-967-0850-51-1

Geometric-Mean Averaging Operators by[24]:

Assume there are n documents d_1, d_2, …, d_n, and there are m terms t_1, t_2, …, t_m. The value of $\tilde{e}_{ij}$ is the generalized fuzzy number means the degree of strength of terms t_j in document di, where $\tilde{e}_{ij}$ = (a_{ij}, b_{ij}, c_{ij}, d_{ij}; $\hat{w}_{ij}$), $0 \leqq a_{ij} \leqq b_{ij} \leqq c_{ij} \leqq d_{ij} \leqq 1$, $0 < \hat{w}_{ij} \leqq 1$, $1 \leqq i \leqq n$, $1 \leqq j \leqq m$, and $\hat{w}_{ij}$ means the degree of confidence of the value $\tilde{e}_{ij}$. GFNGMA operators shown as follows:

$\tilde{F}$ (d_i, qAND) = $\tilde{F}$ (d_i, t_1 AND t_2 AND … AND t_m)

$$= (\breve{\wedge}_{j=1}^{m} a_{ij}, \breve{\wedge}_{j=1}^{m} b_{ij}, \breve{\wedge}_{j=1}^{m} c_{ij}, \breve{\wedge}_{j=1}^{m} d_{ij}; [\prod_{j=1}^{m} \hat{w}_{ij}]^{\frac{1}{m}})$$

$$= ([\prod_{j=1}^{m}(\alpha + a_{ij})]^{\frac{1}{m}} - \alpha, [\prod_{j=1}^{m}(\alpha + b_{ij})]^{\frac{1}{m}} - \alpha,$$

$$[\prod_{j=1}^{m}(\alpha + c_{ij})]^{\frac{1}{m}} - \alpha, [\prod_{j=1}^{m}(\alpha + d_{ij})]^{\frac{1}{m}} - \alpha; [\prod_{j=1}^{m} \hat{w}_{ij}]^{\frac{1}{m}}), \quad (3)$$

$\tilde{F}$ (d_i, qOR) = $\tilde{F}$ (di, t_1 OR t_2 OR … OR t_m)

$$= (\breve{\vee}_{j=1}^{m} a_{ij}, \breve{\vee}_{j=1}^{m} b_{ij}, \breve{\vee}_{j=1}^{m} c_{ij}, \breve{\vee}_{j=1}^{m} d_{ij}; [\prod_{j=1}^{m} \hat{w}_{ij}]^{\frac{1}{m}})$$

$$= (\alpha + 1) - [\prod_{j=1}^{m}(\alpha + 1 - a_{ij})]^{\frac{1}{m}}, (\alpha + 1) - [\prod_{j=1}^{m}(\alpha + 1 - b_{ij})]^{\frac{1}{m}},$$

$$(\alpha + 1) - [\prod_{j=1}^{m}(\alpha + 1 - c_{ij})]^{\frac{1}{m}}, (\alpha + 1) - [\prod_{j=1}^{m}(\alpha + 1 - d_{ij})]^{\frac{1}{m}}; [\prod_{j=1}^{m} \hat{w}_{ij}]^{\frac{1}{m}}), \quad (4)$$

where $\alpha \in \{0, 1\}$ and $1 \leqq i \leqq n$ and $1 \leqq j \leqq m$, $\tilde{F}$ (di, qAND) $\in [0, 1]$ and $\tilde{F}$ (di, qOR) $\in [0, 1]$. The degree of satisfaction of document di relate to the query qAND and the degree of satisfaction $\tilde{F}$ (di, qOR) of document di relate to the query qOR are generalized fuzzy numbers.

Generalized Fuzzy-Number Quadratic-Mean Averaging Operators by [29]:

Assume there are n documents d_1, d_2, …, d_n, and there are m terms t_1, t_2, …, t_m. The value of $\tilde{e}_{ij}$ is the generalized fuzzy number means the degree of strength of terms t_j in document di, where $\tilde{e}_{ij}$ = (a_{ij}, b_{ij}, c_{ij}, d_{ij}; $\hat{w}_{ij}$), $0 \leqq a_{ij} \leqq b_{ij} \leqq c_{ij} \leqq d_{ij} \leqq 1$, $0 < \hat{w}_{ij} \leqq 1$, $1 \leqq i \leqq n$, $1 \leqq j \leqq m$, and $\hat{w}_{ij}$ means the degree of confidence of the value $\tilde{e}_{ij}$. GFNQMA operators shown as follows:

$\tilde{F}$ (d_i, qAND) = $\tilde{F}$ (d_i, t_1 AND t_2 AND … AND t_m)

$$= (\breve{\wedge}_{j=1}^{m} a_{ij}, \breve{\wedge}_{j=1}^{m} b_{ij}, \breve{\wedge}_{j=1}^{m} c_{ij}, \breve{\wedge}_{j=1}^{m} d_{ij}; [\prod_{j=1}^{m} \hat{w}_{ij}]^{\frac{1}{m}})$$

ISBN 978-967-0850-51-1

$$= ([2\alpha + Min(a_{ij})(1-\alpha) - \alpha \times \sqrt{\frac{\sum_{j=1}^{m}(2-a_{ij})^2}{m}}],$$

$$[2\alpha + Min(b_{ij})(1-\alpha) - \alpha \times \sqrt{\frac{\sum_{j=1}^{m}(2-b_{ij})^2}{m}}],$$

$$[2\alpha + Min(c_{ij})(1-\alpha) - \alpha \times \sqrt{\frac{\sum_{j=1}^{m}(2-c_{ij})^2}{m}}],$$

$$[2\alpha + Min(d_{ij})(1-\alpha) - \alpha \times \sqrt{\frac{\sum_{j=1}^{m}(2-d_{ij})^2}{m}}];[\prod_{j=1}^{m}\hat{w}_{ij}]^{\frac{1}{m}}),$$

(5)

$\tilde{F}$ (d_i, qOR) = $\tilde{F}$ (d_i, t_1 OR t_2 OR … OR t_m)

$$= (\breve{\vee}_{j=1}^{m} a_{ij}, \breve{\vee}_{j=1}^{m} b_{ij}, \breve{\vee}_{j=1}^{m} c_{ij}, \breve{\vee}_{j=1}^{m} d_{ij}; 1-[\prod_{j=1}^{m}(1-\hat{w}_{ij})]^{\frac{1}{m}})$$

$$= ([Max(a_{ij})(1-\alpha) + \alpha \times \sqrt{\frac{\sum_{j=1}^{m}(1+a_{ij})^2}{m}} - \alpha],$$

$$[Max(b_{ij})(1-\alpha) + \alpha \times \sqrt{\frac{\sum_{j=1}^{m}(1+b_{ij})^2}{m}} - \alpha],$$

$$[Max(c_{ij})(1-\alpha) + \alpha \times \sqrt{\frac{\sum_{j=1}^{m}(1+c_{ij})^2}{m}} - \alpha],$$

$$[Max(d_{ij})(1-\alpha) + \alpha \times \sqrt{\frac{\sum_{j=1}^{m}(1+d_{ij})^2}{m}} - \alpha]; 1-[\prod_{j=1}^{m}(1-\hat{w}_{ij})]^{\frac{1}{m}}),$$

(6)

where α ∈ {0, 1} and 1 ≦ i ≦ n and 1 ≦ j ≦ m, $\tilde{F}$ (d_i, qAND) ∈ [0, 1] and $\tilde{F}$ (d_i, qOR) ∈ [0, 1]. The degree of satisfaction $\tilde{F}$ (d_i, qAND) of document d_i relate to the query qAND and the degree of satisfaction $\tilde{F}$ (d_i, qOR) of document d_i relate to the query qOR are generalized fuzzy numbers.

ISBN 978-967-0850-51-1

AND/OR Query Operators Based on Extended Quadratic-Mean Operators

In [30], Chen and Shih proposed the AND/OR query operators based on extended quadratic-mean operators for dealing with fuzzy document retrieval of the AND and OR operators. The AND/OR query operators for dealing with the fuzzy document retrieval, shown as follows:

$$\begin{aligned} F(d_i, q_{AND}) &= F(d_j, t_1 \text{ AND } t_2 \text{ AND} \ldots \text{AND } t_m) \\ &= \mathrm{Min}(e_{ij})(1-\alpha) + \alpha \times \left(4 - \sqrt{\frac{\sum_{j=1}^{m} (4 - e_{ij})^2}{m}} \right), \end{aligned} \quad (7)$$

$$\begin{aligned} F(d_i, q_{OR}) &= F(d_j, t_1 \text{ OR } t_2 \text{ OR} \ldots \text{OR } t_m) \\ &= \mathrm{Max}(e_{ij})(1-\alpha) + \alpha \times \left(\sqrt{\frac{\sum_{j=1}^{m} (3 + e_{ij})^2}{m}} - 3 \right), \end{aligned} \quad (8)$$

where $\alpha \in \{0, 1\}$, $1 \leqq i \leqq n$ and $1 \leqq j \leqq m$, F(di, qAND) $\in$ [0, 1], and F(di, qOR) $\in$ [0, 1]. The values of F(di, qAND) and F(di, qOR) of the proposed AND/OR query operators are controlled by a parameter α, where α is either 0 or 1.

ANALYSIS OF THE EXISTING AVERAGING OPERATORS

In this section, we indicated some drawbacks of the AND/OR query operators based on GFNGMA operation [24] and the AND/OR query operators based on GFNQMA operation [29] by some examples.

Geometric-Mean Averaging Operators

In formulas (3) and (4), we can understand that the value $\hat{w}_{ij}$ of GFNGMA operators in the AND and OR operations are the same as $\left[\prod_{j=1}^{m} \hat{w}_{ij}\right]^{\frac{1}{m}}$. However, the system can't distinguish the degrees of satisfaction of the documents relate to the queries for the AND and OR operations. It shown as follows:

Example 1: Assume that the parameter α = 1, and there is a document d_1, and there are two queries q_1 and q_2, shown as follows:

d_1 = {(Bioinformatics, $\tilde{e}_{11}$), (Retrieval, $\tilde{e}_{12}$), (Algorithm, $\tilde{e}_{13}$)},
q_1 = Bioinformatics AND Retrieval AND Algorithms,
q_2 = Bioinformatics OR Retrieval OR Algorithms,

where $\tilde{e}_{11}$ = (a_{11}, b_{11}, c_{11}, d_{11}; $\hat{w}_{11}$) = (0.1, 0.2, 0.3, 0.4; 0.9), $\tilde{e}_{12}$ = (a_{12}, b_{12}, c_{12}, d_{12}; $\hat{w}_{12}$) = (0.1, 0.2, 0.3, 0.4; 0.85) and $\tilde{e}_{13}$ = (a_{13}, b_{13}, c_{13}, d_{13}; $\hat{w}_{13}$) = (0.1, 0.2, 0.3, 0.4; 0.93) are generalized fuzzy numbers means the degrees of strength of the index terms t_1 = "Bioinformatics", t_2 = "Retrieval", and t_3 = "Algorithms" in document d_1. If we use formulas (3) and (4) to deal with Example 3.1, we can calculate the degrees of

ISBN 978-967-0850-51-1

satisfaction of the document d_1 relate to the queries q_1 and q_2 for the AND and OR operations, shown as follows:

$\tilde{F}$ (d_1, q_1) = $\tilde{F}$ (d_1, Bioinformatics AND Retrieval AND Algorithms)

$$= (\wedge_{j=1}^{3} a_{1j}, \wedge_{j=1}^{3} b_{1j}, \wedge_{j=1}^{3} c_{1j}, \wedge_{j=1}^{3} d_{1j}; [\prod_{j=1}^{3} \hat{w}_{1j}]^{\frac{1}{3}})$$

$$= (0.1, 0.2, 0.3, 0.4; 0.8927).$$

$\tilde{F}$ (d_1, q_2) = $\tilde{F}$ (Bioinformatics OR Retrieval OR Algorithms)

$$= (\vee_{j=1}^{3} a_{1j}, \vee_{j=1}^{3} b_{1j}, \vee_{j=1}^{3} c_{1j}, \vee_{j=1}^{3} d_{1j}; [\prod_{j=1}^{3} \hat{w}_{1j}]^{\frac{1}{3}})$$

$$= (0.1, 0.2, 0.3, 0.4; 0.8927).$$

Hence, the degree of satisfaction of the document d_1 relate to the query q_1 for the AND operation and the document d_1 relate to the query q_2 for the OR operation are the same as (0.1, 0.2, 0.3, 0.4; 0.8927). Thus, the system can't distinguish the degrees of satisfaction of the documents relate to the queries for the AND and OR operations.

Generalized Fuzzy-Number Quadratic-Mean Averaging Operators

Example 2: Assume that the parameter α = 1, and there are two documents d_4 and d_5, and there is a query q_4, shown as follows:

d_2 = {(Bioinformatics, $\tilde{e}_{21}$), (Retrieval, $\tilde{e}_{22}$)},
d_3 = {(Bioinformatics, $\tilde{e}_{31}$), (Retrieval, $\tilde{e}_{32}$)},
q_3 = Bioinformatics AND Retrieval AND Algorithms,

where $\tilde{e}_{21}$ = (a_{21}, b_{21}, c_{21}, d_{21}; $\hat{w}_{21}$) = (0.7, 0.7, 0.7, 0.8; 0.5), $\tilde{e}_{22}$ = (a_{22}, b_{22}, c_{22}, d_{22}; $\hat{w}_{22}$) = (0.2, 0.2, 0.2, 0.3; 0.5) and $\tilde{e}_{31}$ = (a_{31}, b_{31}, c_{31}, d_{31}; $\hat{w}_{31}$) = (0.5, 0.6, 0.6, 0.6; 0.5), $\tilde{e}_{32}$ = (a_{32}, b_{32}, c_{32}, d_{32}; $\hat{w}_{32}$) = (0.3, 0.3, 0.4, 0.4; 0.5) are generalized fuzzy numbers means the degrees of strength of the index terms t_1 = "Bioinformatics", t_2 = "Retrieval" in documents d_2 and d_3. If we use formulas (5) and (6) to deal with this example, we can get the degrees of satisfaction of the documents d_2 and d_3 relate to the query q_3 for the AND operation as follows:

$\tilde{F}$ (d_2, q_3) = $\tilde{F}$ (d_2, Bioinformatics AND Retrieval AND Algorithms)

$$= (\wedge_{j=1}^{3} a_{2j}, \wedge_{j=1}^{3} b_{2j}, \wedge_{j=1}^{3} c_{2j}, \wedge_{j=1}^{3} d_{2j}; [\prod_{j=1}^{3} \hat{w}_{2j}]^{\frac{1}{3}})$$

$$= (0.429968, 0.429968, 0.486725, 0.528606; 0.5).$$

$\tilde{F}$ (d_3, q_3) = $\tilde{F}$ (d_3, Bioinformatics AND Retrieval AND Algorithms)

$$= (\wedge_{j=1}^{3} a_{3j}, \wedge_{j=1}^{3} b_{3j}, \wedge_{j=1}^{3} c_{3j}, \wedge_{j=1}^{3} d_{3j}; [\prod_{j=1}^{3} \hat{w}_{3j}]^{\frac{1}{3}})$$

$$= (0.396878, 0.442759, 0.49667, 0.49667; 0.5).$$

ISBN 978-967-0850-51-1

Furthermore, we can calculate the ranking values of generalized fuzzy numbers by formula (5). The ranking value of R($\tilde{D}_{(d_2, q_3)}$) = 0.927164 and the ranking value of R($\tilde{D}_{(d_3, q_3)}$) = 0.927959. Hence, the degree of satisfaction in document d_2 is higher than the degree of satisfaction in document d_3, and the system will retrieve document d_3. However, it will not coincide with the intuition of the human being.

NEW FUZZY-NUMBER AND/OR QUERY OPERATORS BASED ON THE EXTENDED QUADRATIC-MEAN OPERATOR

In [7], Miyamoto pointed out that linguistic queries processing is an important research topic in fuzzy information retrieval. It is more flexible and more useful when using the generalized fuzzy numbers or interval fuzzy numbers to presented linguistic queries for dealing with the linguistic queries processing problem. In this section, we will propose the new generalized fuzzy-numbers AND/OR query operators to deal with queries represented by generalized fuzzy numbers problems. In the following, we use the $\breve{\wedge}$ and $\breve{\vee}$ to presented formulas (7) and (8), repetitively. Assume there are n documents $d_1, d_2, \ldots, d_n$ and there are m index terms $t_1, t_2, \ldots, t_m$. The value of $\tilde{e}_{ij}$ is a generalized fuzzy numbers means the degree of strength of term t_j in document d_1, where $\tilde{e}_{ij} = (a_{ij}, b_{ij}, c_{ij}, d_{ij}; \hat{w}_{ij})$, $0 \leqq a_{ij} \leqq b_{ij} \leqq c_{ij} \leqq d_{ij} \leqq 1$, $0 \leqq \hat{w}_{ij} \leqq 1$, $0 \leqq i \leqq 1$, $0 \leqq j \leqq 1$, and $\hat{w}_{ij}$ means the degree of confidence of the value $\tilde{e}_{ij}$. The new generalized fuzzy-numbers AND/OR query operators based on extended quadratic mean operator are shown as follows:

$\tilde{F}$ (d_i, q_{AND}) = $\tilde{F}$ (d_i, t_1 AND t_2 AND … AND t_m)

$$= (\breve{\vee}_{j=1}^{m} a_{ij}, \breve{\vee}_{j=1}^{m} b_{ij}, \breve{\vee}_{j=1}^{m} c_{ij}, \breve{\vee}_{j=1}^{m} d_{ij}; 1 - [\prod_{j=1}^{m} \hat{w}_{ij}]^{\frac{1}{m}})$$

$$= [Min(a_{ij})(1-\alpha) - \alpha \times \left(4 - \sqrt{\frac{\sum_{j=1}^{m} (4 - a_{ij})^2}{m}} \right)],$$

$$[Min(a_{ij})(1-b) - \alpha \times \left(4 - \sqrt{\frac{\sum_{j=1}^{m} (4 - b_{ij})^2}{m}} \right)],$$

$$[Min(a_{ij})(1-c) - \alpha \times \left(4 - \sqrt{\frac{\sum_{j=1}^{m} (4 - c_{ij})^2}{m}} \right)],$$

$$[Min(d_{ij})(1-d) - \alpha \times \left(4 - \sqrt{\frac{\sum_{j=1}^{m} (4 - d_{ij})^2}{m}} \right)];[\prod_{j=1}^{m} (\hat{w}_{ij})]^{\frac{1}{m}}),$$

(9)

ISBN 978-967-0850-51-1

$$\tilde{F}(d_i, q_{OR}) = \tilde{F}(d_i, t_1 \text{ OR } t_2 \text{ OR } \ldots \text{ OR } t_m)$$

$$= (\breve{\wedge}_{j=1}^{m} a_{ij}, \breve{\wedge}_{j=1}^{m} b_{ij}, \breve{\wedge}_{j=1}^{m} c_{ij}, \breve{\wedge}_{j=1}^{m} d_{ij}; 1 - [\prod_{j=1}^{m} (1 - \hat{w}_{ij})]^{\frac{1}{m}})$$

$$= ([Max(a_{ij})(1-\alpha) + \alpha \times \left(\sqrt{\frac{\sum_{j=1}^{m} (a_{ij}+3)^2}{m}} - 3 \right)],$$

$$[Max(b_{ij})(1-\alpha) + \alpha \times \left(\sqrt{\frac{\sum_{j=1}^{m} (b_{ij}+3)^2}{m}} - 3 \right)],$$

$$[Max(c_{ij})(1-\alpha) + \alpha \times \left(\sqrt{\frac{\sum_{j=1}^{m} (c_{ij}+3)^2}{m}} - 3 \right)],$$

$$[Min(d_{ij})(1-\alpha) + \alpha \times \left(\sqrt{\frac{\sum_{j=1}^{m} (d_{ij}+3)^2}{m}} - 3 \right)]; 1 - [\prod_{j=1}^{m} (1 - \hat{w}_{ij})]^{\frac{1}{m}}), \qquad (10)$$

where $\alpha \in \{0, 1\}$, $1 \leqq i \leqq n$, $1 \leqq j \leqq m$, $\tilde{F}(d_i, q_{AND}) \in [0, 1]$ and $\tilde{F}(d_i, q_{OR}) \in [0, 1]$. The degree of satisfaction $\tilde{F}(d_i, q_{AND})$ of document d_i relate to the query q_{AND} and the degree of satisfaction $\tilde{F}(d_i, q_{OR})$ of document d_i relate to the query q_{OR} are generalized fuzzy numbers.

Then we use formula (9) and (10) to deal with Example 1, we can calculated $\tilde{F}(d_1, q_1) = \tilde{F}$ (d_1, Bioinformatics AND Retrieval AND Algorithms)

$$= (\tilde{\wedge}_{j=1}^{3} a_{1j}, \tilde{\wedge}_{j=1}^{3} b_{1j}, \tilde{\wedge}_{j=1}^{3} c_{1j}, \tilde{\wedge}_{j=1}^{3} d_{1j}; [\prod_{j=1}^{3} \hat{w}_{1j}]^{\frac{1}{3}})$$

$$= (0.1, 0.2, 0.3, 0.4; 0.89271903).$$

$\tilde{F}(d_1, q_2) = \tilde{F}$ (Bioinformatics OR Retrieval OR Algorithms)

$$= (\tilde{\vee}_{j=1}^{3} a_{1j}, \tilde{\vee}_{j=1}^{3} b_{1j}, \tilde{\vee}_{j=1}^{3} c_{1j}, \tilde{\vee}_{j=1}^{3} d_{1j}; 1 - [\prod_{j=1}^{3} (1 - \hat{w}_{1j})]^{\frac{1}{3}})$$

$$= (0.1, 0.2, 0.3, 0.4; 0.8983603).$$

Hence, the degree of satisfaction of the document d_1 relate to the query q_1 for the AND operation and the document d_1 relate to the query q_2 for the OR operation are different. Thus, the system can distinguish the degrees of satisfaction of the documents relate to the queries for the AND/OR operations.

Then we use the formula(9) to deal with the Example 2, we can calculated $\tilde{F}(d_2, q_3) = \tilde{F}$ (d_2, Bioinformatics AND Retrieval AND Algorithms)

$$= (\tilde{\wedge}_{j=1}^{3} a_{2j}, \tilde{\wedge}_{j=1}^{3} b_{2j}, \tilde{\wedge}_{j=1}^{3} c_{2j}, \tilde{\wedge}_{j=1}^{3} d_{2j}; [\prod_{j=1}^{3} \hat{w}_{2j}]^{\frac{1}{3}})$$

$$= (0.441208, 0.441208, 0.49429, 0.540954; 0.5).$$

ISBN 978-967-0850-51-1

$\tilde{F}$ (d_3, q_3) = $\tilde{F}$ (d_3, Bioinformatics AND Retrieval AND Algorithms)

$$= (\bigwedge_{j=1}^{3} a_{3j}, \bigwedge_{j=1}^{3} b_{3j}, \bigwedge_{j=1}^{3} c_{3j}, \bigwedge_{j=1}^{3} d_{3j}; [\prod_{j=1}^{3} \hat{w}_{3j}]^{\frac{1}{3}})$$

$$= (0.398611, 0.446832, 0.498572, 0.498572; 0.5).$$

Furthermore, we can calculate the ranking values of generalized fuzzy numbers by formula (9). The ranking value of R($\tilde{D}_{(d_2, q_3)}$) = 0.9247164 and the ranking value of R($\tilde{D}_{(d_3, q_3)}$) = 0.933349. Hence, the degree of satisfaction in document d_2 is higher than the degree of satisfaction in document d_3, and the system will retrieve document d_2. However, it will coincide with the intuition of the human being.

CONCLUSIONS

In this paper, we point out that some drawbacks in existing averaging operators to deal with queries represented by generalized fuzzy-numbers. Therefore, we propose novel generalized fuzzy-numbers AND/OR query operators to overcome the existing methods. The proposed novel generalized fuzzy-numbers AND/OR query operators are more flexible and more useful than the existing averaging operators to deal with novel generalized fuzzy numbers.

ACKNOWLEDGMENTS

The quality of this experiment was greatly enhanced by the gracious assistance of Professor Shi-Jay Chen. Thanks to Professor Shi-Jay Chen at National United University, Taiwan, whose advice on experimental design and help with subsequent analysis was invaluable.

REFERENCES

1. Kenney J. F. and Keeping E. S. (1962). Mathematics of Statistics, Princeton, *NJ: Van Nostrand*, Part I, 3rd ed, 17.
2. Zadeh L. A. (1965). Fuzzy sets, *Information and Control*, 8,338-353.
3. Waller W. G. and Kraft D. H. (1979). A mathematical model of a weighted Boolean retrieval system, *Information Processing and Management*, 15 (5), 235-245.
4. Salton G., Fox E. A., and Wu H. (1983). Extended Boolean information retrieval, Communications of the ACM, 26 (12), 1022-1036.
5. Chen S. H. (1985). Ranking fuzzy numbers with maximizing set and minimizing set, *Fuzzy Sets and Systems*, 17 (1), 113-129.
6. Gorzałczany M. B. (1987). A method of inference in approximate reasoning based on interval-valued fuzzy sets, *Fuzzy Sets and Systems*, 21 (1), 1-17.
7. Smith M. E. (1990). Aspects of the P-Norm Model of Information Retrieval: Syntactic Query Generation, Efficiency and Theoretical Properties, Ph. D Thesis. TR 90-1128.
8. Miyamoto S. (1990). Fuzzy Sets in Information Retrieval and Cluster Analysis, *Springer*, 4.

ISBN 978-967-0850-51-1

9. Gupta M. M. and Qi J. (1991). Theory of T-norms and fuzzy inference methods, *Fuzzy Sets and Systems*, 40 (3), 431-450.
10. Kim M. H., Lee J. H. and Lee Y. J. (1993). Analysis of fuzzy operators for high quality information retrieval, *Information Processing Letters*, 46 (5), 251-256.
11. Lee J. H. (1994). Properties of extended Boolean models in information retrieval, *Proceedings of the Seventeenth Annual ACM Conference on Research and Development in Information Retrieval*, 182-190.
12. Patyra M. J. (1995). A degenerated fuzzy-number processing system based on artificial neural networks, *Information Sciences*, 86 (4), 211-226.
13. Liu S. Y. and Chen J. G. (1995). Development of a machine troubleshooting expert system via fuzzy multiattribute decision-making approach, *Expert Systems with Applications*, 8 (1), 187-201.
14. Chiang D. A., Chow L. R., and Hsien N. C. (1997). Fuzzy information in extended fuzzy relational databases, *Fuzzy Sets and Systems*, 92 (1), 1-20.
15. Devedzic G. B. and Pap E. (1999). Multi-criteria-multistages linguistic evaluation and ranking of machine tools, *Fuzzy Sets and Systems*, 102 (4), 451-461.
16. Chiadamrong N. (1999). An integrated fuzzy multi-criteria decision making method for manufacturing strategies selection, *Computers and Industrial Engineering*, 37 (1-2) ,433-436.
17. Rocacher D. (2001). On the use of fuzzy numbers in flexible querying, *Proceeding of the IFSA World Congress and 20th NAFIPS International Conference*, vol. 4, pp. 2440–2445.
18. Frigui H. (2001). Interactive image retrieval using fuzzy sets, *Pattern Recognition Letters*, 22 (9), 1021-1031.
19. Chen S. M., Horng Y. J. and Lee C. H. (2001). Document retrieval using fuzzy-valued concept networks, *IEEE Transactions on Systems*, Man and Cybernetics-Part B: Cybernetics 31 (1), 111 –118.
20. Hong D. H. and Lee S. (2002). Some algebraic properties and a distance measure for interval-valued fuzzy numbers, *Information Sciences* 148 (1), 1-10.
21. Yao J. S. and Lin F. T. (2002). Constructing a fuzzy flow-shop sequencing model based on statistical data, *International Journal of Approximate Reasoning* 29 (3), 215-234.
22. Chen S. J. and Chen S. M. (2002). A new method for fuzzy information retrieval based on geometric-mean averaging operators, *Proceedings of the 2002 International Computer Symposium: Workshop on Artificial Intelligence*.
23. Chen S. J. and Chen S. M. (2003). A new fuzzy query processing method for handling fuzzy-number information retrieval problems, *Proceedings of the 14th International Conference on Information Management*.
24. Chen S. J. and Chen S. M. (2003). Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers, IEEE Transactions on Fuzzy Systems 11 (1) 45-56.
25. Hong W. S., Chen S. M., and Chen S. J. (2005). Fuzzy information retrieval based on weighted power-mean average operators, *Proceedings of the Tenth International Symposium on Artificial Life and Robotics*.
26. Chen S. J. and Chen S. M. (2007). Fuzzy query processing for document retrieval based on GFNGMA operators, *Intelligent Automation & Soft Computing*, vol. 13, no. 2, 171-196.

ISBN 978-967-0850-51-1

27. Chen S. J. and Chen S. M. (2007). Fuzzy risk analysis based on the ranking of generalized trapezoidal fuzzy numbers, Applied Intelligence: An International Journal, 26 (1), 1-11.
28. Chen S. J. and Chu H. C. (2010). A new method for fuzzy information retrieval based on quadratic-mean averaging operators, *Proceedings of the e-CASE & e-Tech International Conference*, 2487-2513.
29. Chen S. J. and Chu H. C. (2010). A new method for fuzzy query processing based on GFNQMA operators, *Proceedings of the 2010 International Conference on Electronics and Information Engineering (ICEIE 2010)*.
30. Chen S. J. and Shih Y. Y. (2015). Novel Fuzzy Query Processing Method for Information Retrieval, *Proceedings of the e-CASE & e-Tech International Conference*.

ISBN 978-967-0850-51-1