Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 12, December 2024

A New Fuzzy Query Processing Method for Document Retrieval Based on Extended Fuzzy Concept Networks

Shi-Jay Chen
Department of Information Management,
National United University,
Miaoli, Taiwan

***Abstract*—This study proposes a new mechanism based on extended fuzzy concept networks for fuzzy query processing of document retrieval and we use a relevance matrix and a relation matrix to model extended fuzzy concept networks. This mechanism combines the document descriptor relevance matrix defined by the expert with the user's query descriptor based on different weights for obtaining a matrix called a satisfaction matrix. This mechanism uses the AND operator of the quadratic-mean averaging operators to calculate the AND operation of all components in each row of the satisfaction matrix. Finally, ranking the degrees of satisfaction of each satisfaction matrix obtains documents more suitable for the user's needs.**

***Keywords:* Document retrieval, Fuzzy query processing, Quadratic-Mean Averaging (QMA) operators, Extended fuzzy concept networks, Relevance matrix, Relation matrix.**

1. INTRODUCTION

Recently, several researchers (Chen and Wang 1995; Her and Ke 1983; Horng and Chen 1999; Kamel et al. 1990; Lucarella and R.Morara 1991; Miyamoto 1990; Moradi et al. 2008; Murai et al. 1989; Radechi 1977; Tadechi 1979; Tahani 1976; Zemankova 1989) dealt with document retrieval processing problems based on fuzzy set theory presented by Zadeh (1965). Lucarella et al. (1991) presented an information retrieval method based on fuzzy concept networks. However, there is only one kind of fuzzy relationship between concepts in concept networks (Lucarella et al. 1991) (i.e., a fuzzy positive association relation). Kracker (1992) presented an extended fuzzy concept network model and its applications that have four kinds of fuzzy relationships between concepts in the concept networks for database queries (i.e., fuzzy positive association, fuzzy negative association, fuzzy generalization, and fuzzy specialization). Furthermore, Horng and Chen (1999) and Moradi et al. (2008) presented information retrieval systems that deal with document retrieval based on extended fuzzy concept networks. However, these methods based on fuzzy concept networks do not satisfy efficiency or effectiveness. For example, the general user cannot define the degree of relevance and fuzzy relationship between concepts and documents as precisely as can an expert.

This paper proposes a new mechanism for dealing with document retrieval based on extended fuzzy concept networks. The rest of the paper is organized as follows. Section 2 briefly reviews the principles of concept networks presented by Lucarella et al. (1991) and extended fuzzy concept networks presented by Kracker (1992). Section 3 proposes a new mechanism for fuzzy query processing for document retrieval based on extended fuzzy concept networks and we use relevance matrix and relation matrix to model extended fuzzy concept networks. Section 4 discusses the conclusions.

2. PRELIMINARY

2.1. Concept networks

Lucarella et al. (1991) presented a fuzzy information retrieval method based on concept networks. A concept network consists of nodes and directed links where each node presents a concept or a document; each directed link connects two concepts or directs from one concept C_i to one document d_j and is labeled with a real value between zero and one. If $C_i \overset{\mu}{\rightarrow} C_j$, it indicates that the degree of relevance from concept C_i to concept C_j is μ where $\mu \in [0,1]$. If $C_i \overset{\mu}{\rightarrow} d_j$, it indicates that the degree of relevance of concept C_i with respect to document d_j is μ where $\mu \in [0,1]$. For example, Fig. 1 presents a concept network where $C_1, C_2, \dots, C_7$ are concepts; d_1, d_2, d_3, d_4 are documents. Fig. 1 shows documents d_1, d_2, d_3, d_4 as a fuzzy subset of concepts,

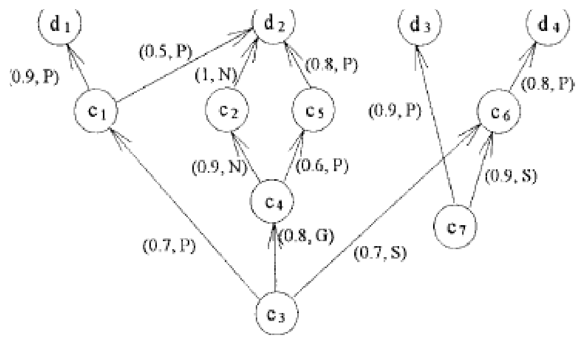

Fig. 1. A concept network

where $d_1 = \{(C_1,0.9)\}$, $d_2 = \{(C_1,0.6),(C_2,1),(C_5,0.8)\}$, $d_3 = \{(C_7,0.9)\}$, $d_4 = \{(C_6,0.8)\}$, and 0.6 presents the relevance value of the document d_2 with respect to concept C_1.

2.2. Extended Fuzzy Concept Networks

There is only one kind of fuzzy relationship between concepts in the concept networks presented by Lucarella et al. (1991) (i.e.,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 12, December 2024

a fuzzy positive association relation). Kracker (1992) presented an extended fuzzy concept network model and its applications for database queries that have four kinds of fuzzy relationships between concepts in a concept network (i.e., fuzzy positive association, fuzzy negative association, fuzzy generalization, and fuzzy specialization). Horng and Chen (1999) and Moradi et al. (2008) presented information retrieval systems for dealing with document retrieval based on extended fuzzy concept networks. The fuzzy relationships between concepts are described by Kracker (1992) as follows:

(1) Fuzzy Positive Association, it relates concepts which in some contexts have a fuzzy similar meaning (e.g., person ↔ individual).

(2) Fuzzy Negative Association, it relates concepts which are fuzzy complementary (e.g., male ↔ female), fuzzy incompatible (e.g., unemployed ↔ freelance) or fuzzy antonyms (e.g., large ↔ small).

(3) Fuzzy Generalization, one concept that is regarded as a fuzzy generalization of another concept if it consists of that concept (e.g., vehicle ↔ car).

(4) Fuzzy Specialization, the inverse of the fuzzy generalization relationship. That is, one concept that is regarded as a fuzzy specialization of another concept if it parts of that concept (e.g., car ↔ vehicle)

The fuzzy relationships between concepts introduced can be formally described by Kracker (1992) as follows:

Definition 2.1: Let C be the universal set of all concepts, then

(1) Fuzzy Positive Association (P) is a fuzzy relation μ_p, $\mu_p: C \times C \rightarrow [0,1]$, which is reflexive, sysmmetric, and max-*-transitive.

(2) Fuzzy Negative Association (N) is a fuzzy relation μ_N, $\mu_N: C \times C \rightarrow [0,1]$, which is anti-reflexive, sysmmetric, and max-*-nontransitive.

(3) Fuzzy Generalization (G) is a fuzzy relation μ_G, $\mu_G: C \times C \rightarrow [0,1]$, which is anti-reflexive, anti-sysmmetric, and max-*-transitive.

(4) Fuzzy Specialization (S) is a fuzzy relation μ_S, $\mu_S: C \times C \rightarrow [0,1]$, which is anti-reflexive, anti-sysmmetric, and max-*-transitive

Definition 2.2: An extended fuzzy concept network consists of nodes and directed links. Each node presents a concept of a document. Each directed link connects two concepts or directs from a concept to a document presented by Kracker (1992) and Horng and Chen (1999). If

(1) $c_i \overset{(\mu,P)}{\rightarrow} c_j$, then there is a positive association relationship between concept c_i and concept c_j, and the relevance degree is μ where $\mu \in [0,1]$.

(2) $c_i \overset{(\mu,N)}{\rightarrow} c_j$, then there is a negative association relationship between concept c_i and concept c_j, and the relevance degree is μ where $\mu \in [0,1]$.

(3) $c_i \overset{(\mu,G)}{\rightarrow} c_j$, then concept c_i is more general than concept c_j, and the degree of generalization is μ where $\mu \in [0,1]$.

(4) $c_i \overset{(\mu,S)}{\rightarrow} c_j$, then concept c_i is more special than concept c_j, and the degree of specialization is μ where $\mu \in [0,1]$.

(5) $c_i \overset{(\mu,Z)}{\rightarrow} c_j$, then concept c_i and concept c_j is not defined by the expert explicitly.

(6) $c_i \overset{(\mu,P)}{\rightarrow} d_j$, then there is a positive association relationship between concept c_i and document d_j, and the relevance degree is μ where $\mu \in [0,1]$.

(7) $c_i \overset{(\mu,N)}{\rightarrow} d_j$, then there is a negative association relationship between concept c_i and document d_j, and the relevance degree is μ where $\mu \in [0,1]$.

(8) $c_i \overset{(\mu,Z)}{\rightarrow} d_j$, then concept c_i and document d_j is not defined by the expert explicitly.

For example, Fig. 2 presents an extended concept network where $c_1, c_2, \ldots, c_7$ are concepts; d_1, d_2, d_3, d_4 are documents. Fig. 2 expresses documents d_1, d_2, d_3, d_4 as a fuzzy subset of concepts as follows:

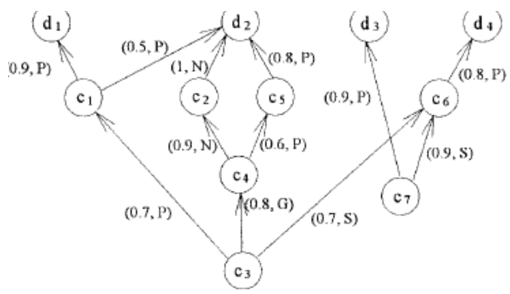

Fig. 2. An extended fuzzy concept network.

where $d_1 = \{(c_1,0.9,P)\}$, $d_2 = \{(c_1,0.5,P),(c_2,1,N),(c_5,0.8,P)\}$, $d_3 = \{(c_7,0.9,P)\}$, $d_4 = \{(c_6,0.8,P)\}$, 0.5 presents the relevance value of the document d_2 with respect to concept c_1, and P presents fuzzy positive association of the fuzzy relationship of the document d_2 with respect to concept c_1.

3. A PROPOSED MECHANISM FOR A DOCUMENT RETRIEVAL METHOD BASED ON EXTENDED FUZZY CONCEPT NETWORKS

This section proposes a new mechanism of fuzzy query processing for document retrieval based on extended fuzzy concept networks. Fig. 3 shows the new mechanism for dealing with document retrieval based on extended fuzzy concept networks. The first step models the matrices (i.e., relevance matrix and relation matrix) between concepts and concepts, and the second step models the matrices (i.e., relevance matrix and relation matrix) between concepts and documents. The third step presents the user's query descriptor vectors. The forth step combines the document descriptor relevance matrix defined by the expert with the user's query descriptor using different weights to obtain a satisfaction matrix. The last step ranks the degrees of satisfaction to which each document satisfies the user's query descriptor.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 12, December 2024

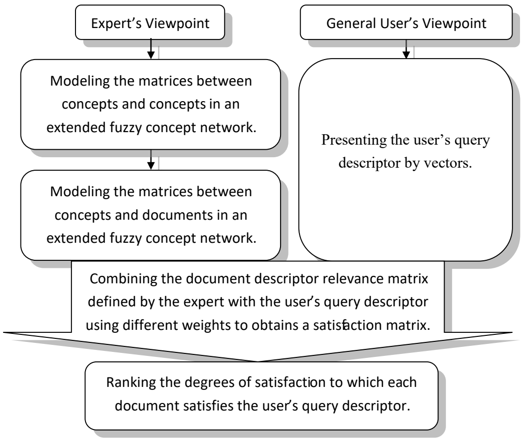

Fig. 3. A new mechanism of fuzzy query processing for document retrieval based on extended fuzzy concept networks

3.1. Modelling the matrices between concepts and concepts in an extended fuzzy concept network

Definition 3.1: A relevance matrix V is a fuzzy matrix presented by Kandel (1986) in which the element $v_{ij} \in [0,1]$ presents the relevance degree between concept c_i and concept c_j as follows:

$$V = \begin{array}{c} \\ c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_1 & \cdots & c_n \end{array} \\ \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix} \end{array},$$

where n is the number of concepts, $v_{ij} \in [0,1]$, $1 \le i \le n$ and $1 \le j \le n$. If $v_{ij} = 0$, the relevance degree between concept c_i and concept c_j is not defined by the expert explicitly. A positive integer p exists where $p \le n-1$, such that $V^P = V^{P+1} = V^{P+2} = \ldots$. Let $V^* = V^P$, then V^* is called the transitive closure of relevance matrix as follows:

$$V^* = V \otimes V = \begin{bmatrix} \bigvee_{i=1,\ldots,n} (v_{1i} \wedge v_{i1}) & \bigvee_{i=1,\ldots,n} (v_{1i} \wedge v_{i2}) & \cdots & \bigvee_{i=1,\ldots,n} (v_{1i} \wedge v_{in}) \\ \bigvee_{i=1,\ldots,n} (v_{2i} \wedge v_{i1}) & \bigvee_{i=1,\ldots,n} (v_{2i} \wedge v_{i2}) & \cdots & \bigvee_{i=1,\ldots,n} (v_{2i} \wedge v_{in}) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \bigvee_{i=1,\ldots,n} (v_{ni} \wedge v_{i1}) & \bigvee_{i=1,\ldots,n} (v_{ni} \wedge v_{i2}) & \cdots & \bigvee_{i=1,\ldots,n} (v_{ni} \wedge v_{in}) \end{bmatrix}, \quad (1)$$

where $\vee$ is the maximum operator and $\wedge$ is the minimum operator.

Definition 3.2: A relation matrix R is a fuzzy matrix in which the element $r_{ij} \in \{P, N, G, S, Z\}$ presents the fuzzy relationship between concept c_i and concept c_j, and P, N, G, S indicated that fuzzy positive association, fuzzy negative association, fuzzy generalization, and fuzzy specialization, respectively as follows:

$$R = \begin{array}{c} \\ c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_1 & \cdots & c_n \end{array} \\ \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \end{array},$$

where n is the number of concepts, $r_{ij} \in \{P, N, G, S, Z\}$, $1 \le i \le n$ and $1 \le j \le n$. If $r_{ij} = Z$, the fuzzy relationship between concept c_i and concept c_j is not defined by the expert explicitly. A positive integer p exists where $p \le n-1$, such that $R^P = R^{P+1} = R^{P+2} = \ldots$. Let $R^* = R^P$, then R^* is called the transitive closure of relevance matrix as follows:

$$R^* = R \otimes R = \begin{bmatrix} \check{\vee}_{i=1,\ldots,n} (r_{1i} \breve{\wedge} r_{i1}) & \check{\vee}_{i=1,\ldots,n} (r_{1i} \breve{\wedge} r_{i2}) & \cdots & \check{\vee}_{i=1,\ldots,n} (r_{1i} \breve{\wedge} r_{in}) \\ \check{\vee}_{i=1,\ldots,n} (r_{2i} \breve{\wedge} r_{i1}) & \check{\vee}_{i=1,\ldots,n} (r_{2i} \breve{\wedge} r_{i2}) & \cdots & \check{\vee}_{i=1,\ldots,n} (r_{2i} \breve{\wedge} r_{in}) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \check{\vee}_{i=1,\ldots,n} (r_{ni} \breve{\wedge} r_{i1}) & \check{\vee}_{i=1,\ldots,n} (r_{ni} \breve{\wedge} r_{i2}) & \cdots & \check{\vee}_{i=1,\ldots,n} (r_{ni} \breve{\wedge} r_{in}) \end{bmatrix}, \quad (2)$$

where $\check{\vee}$ is the operation of choosing the highest priority fuzzy relationship and $\breve{\wedge}$ is the operation of choosing the combination of two relationships according to Table I presented by Kracker (1992) and Horng and Chen (1999). Moreover, in Table I, we let the five different fuzzy relationships have different priorities (i.e., the negative associations (N) has the highest priority, the positive associations (P) has the second highest priority, the relationships (Z) not defined by the expert explicitly is lower, and the priority of the generalization (G) and the specialization (S) are the lowest priority). In Table I, the combination of the high priority relationship and the low priority relationship results in a relationship of high priority except that the combination of the generalization (G) and the specialization (S) is a positive association (P), and the combination of the negative associations (N) with itself is a positive association (P).

TABLE I. THE COMBINATION OF FUZZY RELATIONSHIPS IN A RELATION MATRIX

	P	**N**	**G**	**S**	**Z**
P	P	N	P	P	P
N	N	P	N	N	N
G	P	N	G	P	Z
S	P	N	P	S	Z
Z	P	N	Z	Z	Z

3.2. Modelling the matrices between concepts and documents in an extended fuzzy concept network

Definition 3.3: Let P be a set of documents where $P = \{d_1, d_2, \ldots, d_m\}$, and let C be a set of concepts where $C = \{c_1, c_2, \ldots, c_n\}$. The document descriptor relevance matrix D as follows:

$$D = \begin{array}{c} \\ d_1 \\ d_2 \\ \vdots \\ d_m \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_1 & \cdots & c_n \end{array} \\ \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mn} \end{bmatrix} \end{array},$$

where m is the number of documents, n is the number of concepts, u_{ij}presents the relevance degree between document d_i and conceptc_j, $u_{ij} \in [0,1]$, $1 \le i \le m$ and $1 \le j \le n$.

Definition 3.4: Let P be a set of documents where $P = \{d_1, d_2, ..., d_m\}$, and let C be a set of concepts where $C = \{c_1, c_2, ..., c_n\}$. The document descriptor relation matrix M as follows:

$$M = \begin{array}{c} \\ d_1 \\ d_2 \\ \vdots \\ d_m \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_1 & \cdots & c_n \end{array} \\ \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} \end{bmatrix} \end{array},$$

where m is the number of documents, n is the number of concepts, s_{ij}presents the fuzzy relationship between concept c_j and documentd_i, $s_{ij} \in \{P, N, Z\}$, $1 \le i \le m$ and$1 \le j \le n$.

Horng and Chen (1999) indicated that the document descriptor relevance matrix D and the document descriptor relation matrix M are given subjectively by expert. However, the expert may somehow forget to set some relevance degree and fuzzy relationship between concepts and documents. So, we can obtain the implicit relevance degree between concepts and documents by calculating the document descriptor relevance matrix $D^* = D \otimes V^*$ and the implicit fuzzy relationship between concepts and documents by calculating the document descriptor relation matrix$M^* = M \otimes R^*$.

3.3. Presenting the user's query descriptor by vectors

The user's query descriptor Q can be expressed as follows:

$$Q = \{(c_1, (x_1, y_1), (c_2, (x_2, y_2), ..., (c_i, (x_i, y_i), ..., (c_n, (x_n, y_n)\},$$

where x_i presents the desired relevance degree of the concept c_i with respect to a document d, where$x_i \in [0, 1]$, and y_i presents the desired fuzzy relationships of the concept c_i with respect to a document d where$y_i \in$ {P,N,-},$1 \le i \le n$.

The user's query descriptor Q can also be expressed as a query descriptor relevance vector $\overline{qv}$ and a query descriptor relation vector $\overline{qr}$ as follows:

$$\overline{qv} = (x_1, x_2, ..., x_i, ..., x_n),$$

$$\overline{qr} = (y_1, y_2, ..., y_i, ..., y_n).$$

In a query descriptor relevance vector $\overline{qv}$, if $x_i = 0$, it indicates that desired document d by the general user does not possess concept c_i If $y_i =$ "-" , it indicates that the relevance degree of the concept c_i with respect to the desired document d can be neglected.

3.4. Combining the document descriptor relevance matrix defined by the expert with the user's query descriptor using different weights to obtain a satisfaction matrix

In the following, we use a formula based on weighted power mean to calculate the degrees of weighted T between expert and general user, and use the document descriptor relevance matrix D^* defined by the expert and the user's query descriptor Q in extended fuzzy concept networks as follows:

$$T = [\frac{1}{m^2} \sum_{k=1}^{m} (2m - 2k + 1)\mu_a]^{\frac{1}{m}}, \quad (3)$$

where m presents the number of experts and general users, $k$presents the priority of expert and general user, and μ_a presents the relevance value, where μ_{expert} presents the relevance value of the document descriptor relevance matrix D defined by the expert, and μ_{user} presents the relevance value of the user's query descriptor$Q$,$a \in$ {expert,user}. For example, assume that there are one expert and one general user, and the expert is first in priority and the general user is second in priority in the document retrieval system. We can understand that the degree of weighted T between expert and general user based on formula (3) as follows:

$$T = [\frac{1}{2^2}(2 \times 2 - 2 \times 1 + 1) \times \mu_{\text{expert}} + \frac{1}{2^2}(2 \times 2 - 2 \times 2 + 1) \times \mu_{\text{user}}]^{\frac{1}{2}} = [\frac{3}{4}\mu_{\text{expert}} + \frac{1}{4}\mu_{\text{user}}]^{\frac{1}{2}}.$$

The expert get the degree of weighted is 0.75 and the general user get the degree of weighted is 0.25.

Let (x, s) and (y, t) be two pairs of values, where$x \in [0, 1]$, $y \in [0, 1]$, $s \in \{P, N, Z\}$, and$t \in$. Assume that the document descriptor relevance vector $\overline{dr_i}$ (i.e., the ith row of the document descriptor relevance matrixD^*), the document descriptor relation vector $\overline{dr_i}$ (i.e., the ith row of the document descriptor relation matrix M^*), the query descriptor relevance vector $\overline{qv}$ and the query descriptor relation vector $\overline{qr}$ are presented as follows:

$$\overline{dv_i} = (x_{i1}, x_{i2}, ..., x_{in}),$$

$$\overline{dr_i} = (s_{i1}, s_{i2}, ..., s_{in}),$$

$$\overline{qv} = (y_1, y_2, ... y_n),$$

$$\overline{qr} = (t_1, t_2, ..., t_n),$$

Where $x_{ij} \in [0, 1]$, $y_i \in [0, 1]$, $s_{ij} \in \{P, N, Z\}$, and$t_i \in$, $1 \le i \le m$ and $1 \le j \le n$, m is the number of documents, n is the number of concepts. $\overline{qv}$ and $\overline{qr}$ are defined by the general user. The degree of weighted $T((x, s), (y, t))$ between (x, s) and (y, t) as follows:

$$T((x, s), (y, t)) = \begin{cases} 0 & \text{if } s \neq t, \\ [\frac{1}{m^2} \sum_{k=1}^{m} (2m - 2k + 1)\mu_a]^{\frac{1}{m}} & \text{if } s = t, \end{cases} \quad (4)$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 12, December 2024

where $T((x,s),(y,t)) \in [0,1]$, m presents the number of experts and general users, k presents the priority of expert and general user, and μ_a presents the relevance value, where μ_{expert} presents the relevance value of the document descriptor relevance matrix D^* defined by the expert, and μ_{user} presents the relevance value of the user's query descriptor Q, $a \in$ {expert,user}. If $y =$ "-" or $t =$ "-" , it indicates that concept is neglected by the user's query descriptor. Based on formula (4), we get a matrix called "satisfaction matrix" *SM* that combines the document descriptor relevance matrix D^* defined by the expert with the user's query descriptor Q in extended fuzzy concept networks as follows:

$$SM = \begin{array}{c} \\ d_1 \\ d_2 \\ \vdots \\ d_m \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_1 & \cdots & c_n \end{array} \\ \begin{bmatrix} \mu_{11}^* & \mu_{12}^* & \cdots & \mu_{1n}^* \\ \mu_{21}^* & \mu_{22}^* & \cdots & \mu_{2n}^* \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{m1}^* & \mu_{m2}^* & \cdots & \mu_{mn}^* \end{bmatrix} \end{array},$$

where μ_{ij}^* presents the degree of satisfaction between concept c_j and document d_i from the document descriptor relevance matrix D defined by the expert and the user's query descriptor Q using formula (4) (e.g., $T((x,s),(y,t))=\mu_{ij}^*$), $\mu_{ij}^* \in [0,1]$, $1 \le i \le m$ and $1 \le j \le n$, m is the number of documents, $n$is the number of concepts.

3.5. Ranking the degrees of satisfaction to which each document satisfies the user's query descriptor

In the following, we use quadratic-mean averaging (QMA) operators presented by Chen and Chu (2010) to deal with AND operation in document retrieval based on extended fuzzy concept networks. Furthermore, we calculate the degree of satisfaction to which each document satisfies the user's query descriptor Q for ranking the desired documents for general user needs. According to satisfaction matrixSM, we can use formula (5) to calculate the degree of satisfaction to which document d_i satisfies the user's query descriptor Q as follows:

$$RS_{AND(d_i,qv)} = 2 - \sqrt{\frac{\sum_{j=1}^{n}(2-\mu_{ij}^*)^2}{n-k}}, \qquad (5)$$

where $RS_{AND(d_i,qv)} \in [0,1]$, $1 \le i \le m$ and $1 \le j \le n$, $RS_{AND(d_i,qv)}$ presents the degree of satisfaction to which document d_i satisfies the user's query descriptor Q of AND operation, and μ_{ij}^* presents the degree of satisfaction between concept c_j and document d_i from the documentt descriptor relevance matrix D^* defined by the expert and the user's query descriptor Q, where $\mu_{ij}^* \in [0,1]$, $1 \le i \le m$ and $1 \le j \le n$, k is the number of concepts neglected by the user's query descriptor, m is the number of documents, and n is the number of concepts. The larger the value of $\text{RS}_{\text{AND}(d_i,\text{qv})}$, the greater the degree of satisfaction to which document d_i satisfies the user's query descriptor Q.

Example 3.1: Assume that there is an extended fuzzy concept network as shown in Fig. 4. We can model the extended fuzzy concept network with relevance matrix V and relation matrix R as follows:

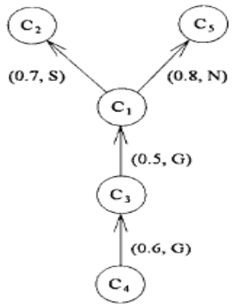

Fig. 4. An extended fuzzy concept network of Example 3.1..

$$V = \begin{bmatrix} 1 & 0.7 & 0.5 & 0 & 0.8 \\ 0.7 & 1 & 0 & 0 & 0 \\ 0.5 & 0 & 1 & 0.6 & 0 \\ 0 & 0 & 0.6 & 1 & 0 \\ 0.8 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R = \begin{bmatrix} P & S & S & Z & N \\ G & P & Z & Z & Z \\ G & Z & P & S & Z \\ Z & Z & G & P & Z \\ N & Z & Z & Z & P \end{bmatrix}.$$

Based on the previous discussion, we can calculate the transitive closure of relevance matrix V^* and the transitive closure of relation matrix R^* based on formulas (1) and (2) as follows:

$$V^* = \begin{bmatrix} 1 & 0.7 & 0.5 & 0.5 & 0.8 \\ 0.7 & 1 & 0.5 & 0.5 & 0.7 \\ 0.5 & 0.5 & 1 & 0.6 & 0.5 \\ 0.5 & 0.5 & 0.6 & 1 & 0.5 \\ 0.8 & 0.7 & 0.5 & 0.5 & 1 \end{bmatrix},$$

$$R^* = \begin{bmatrix} P & N & N & N & P \\ N & P & P & P & N \\ N & P & P & P & N \\ N & P & P & P & N \\ P & N & N & N & P \end{bmatrix}.$$

Assume that there are five documents in a fuzzy information retrieval system, and the document descriptor relevance matrix D and the document descriptor relation matrix M as follows:

$$D = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0.5 & 1 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0.6 & 0 \\ 0.8 & 1 & 1 & 1 & 0 \\ 0.4 & 0.9 & 0 & 0 & 1 \end{bmatrix},$$

$$M = \begin{bmatrix} P & P & P & Z & Z \\ P & P & Z & P & Z \\ Z & Z & Z & P & Z \\ P & P & P & P & Z \\ P & P & Z & Z & N \end{bmatrix}.$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 12, December 2024

Then, based on the previous discussion, we can calculate the document descriptor relevance matrix D^* by $D^* = D \otimes V^*$ and the document descriptor relation matrix M^* by $M^* = M \otimes R^*$ as follows:

$$D^* = D \otimes V^* = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0.5 & 1 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0.6 & 0 \\ 0.8 & 1 & 1 & 1 & 0 \\ 0.4 & 0.9 & 0 & 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0.7 & 0.5 & 0.5 & 0.8 \\ 0.7 & 1 & 0.5 & 0.5 & 0.7 \\ 0.5 & 0.5 & 1 & 0.6 & 0.5 \\ 0.5 & 0.5 & 0.6 & 1 & 0.5 \\ 0.8 & 0.7 & 0.5 & 0.5 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0.6 & 0.8 \\ 0.7 & 1 & 0.6 & 0.7 & 0.7 \\ 0.5 & 0.5 & 0.6 & 0.6 & 0.5 \\ 0.8 & 1 & 1 & 1 & 0.8 \\ 0.8 & 0.9 & 0.5 & 0.5 & 1 \end{bmatrix},$$

$$M^* = M \otimes R^* = \begin{bmatrix} P & P & P & Z & Z \\ P & P & Z & P & Z \\ Z & Z & Z & P & Z \\ P & P & P & P & Z \\ P & P & Z & Z & N \end{bmatrix} \otimes \begin{bmatrix} P & N & N & N & P \\ N & P & P & P & N \\ N & P & P & P & N \\ N & P & P & P & N \\ P & N & N & N & P \end{bmatrix} = \begin{bmatrix} N & P & P & P & N \\ N & P & P & P & N \\ N & P & P & P & N \\ N & P & P & P & N \\ P & N & N & N & P \end{bmatrix}.$$

If the user's query descriptor Q_1 presents by the query descriptor relevance vector $\overline{qv_1}$ and the query descriptor relation vector $\overline{qr_1}$ as follows:

$$\overline{qv_1} = \{0.6, 1.0, 0.8, -, 0.7\},$$

$$\overline{qr_1} = \{N, P, P, -, N\},$$

Based on formula (4), the satisfaction matrix SM_1 that combines the document descriptor relevance matrix D^* defined by the expert with the user's query descriptor Q_1 as follows:

$$SM_1 = \begin{bmatrix} 0.94868 & 1 & 0.97468 & - & 0.88034 \\ 0.82158 & 1 & 0.80623 & - & 0.83666 \\ 0.72457 & 0.79057 & 0.80623 & - & 0.74162 \\ 0.86603 & 1 & 0.97468 & - & 0.88034 \\ 0 & 0 & 0 & - & 0 \end{bmatrix},$$

Furthermore, based on formula (5), the degree of satisfaction to which document d_i with respect to the user's query descriptor Q_1 can be calculated as follows:

$$RS_{AND(d_1,qr_1)} = 2 - \sqrt{\frac{(2-0.94868)^2 + (2-1)^2 + (2-0.97468)^2 + (2-0.88034)^2}{5-1}} = 0.949977,$$

$$RS_{AND(d_2,qr_1)} = 2 - \sqrt{\frac{(2-0.82158)^2 + (2-1)^2 + (2-0.80623)^2 + (2-0.83666)^2}{5-1}} = 0.863435,$$

$$RS_{AND(d_3,qr_1)} = 2 - \sqrt{\frac{(2-0.72457)^2 + (2-0.79057)^2 + (2-0.80623)^2 + (2-0.74162)^2}{5-1}} = 0.765289,$$

$$RS_{AND(d_4,qr_1)} = 2 - \sqrt{\frac{(2-0.86603)^2 + (2-1)^2 + (2-0.97468)^2 + (2-0.88034)^2}{5-1}} = 0.928692,$$

$$RS_{AND(d_5,qr_1)} = 2 - \sqrt{\frac{(2-0)^2 + (2-0)^2 + (2-0)^2 + (2-0)^2}{5-1}} = 0.$$

Hence, we can understand that the documents that satisfy the user's query descriptor are d_1, d_4, d_2, d_3, d_5. In this case, document d_1 is the best choice for the user's query descriptor Q_1, because it has the largest retrieval status value.

4. CONCLUSIONS

This paper presents the concepts of extended fuzzy concept networks in which four kinds of fuzzy relationships exist between concepts in the concept networks (i.e., fuzzy positiveassociation, fuzzy negative association, fuzzy generalization, and fuzzy specialization). We also propose a new mechanism for dealing with document retrieval based on extended fuzzy concept networks. Hence, the proposed method is a more useful fuzzy information retrieval method for dealing with document retrieval because it provides different weights for experts and general users, and coincides with human intuition.

ACKNOWLEDGMENT

This research was funded by the 2024 Annual Key Development Program of National United University (Project No. LC113005). The authors gratefully acknowledge this financial support.

REFERENCES

[1] Chen, S. J. and Chu, H. C. "A new method for fuzzy information retrieval based on quadratic-mean averaging operators," in Processing of the e-CASE & e-Tech International Conference, (2010): 2487-2513.
[2] Chen, S. M. and Horng, Y. J. "Finding inheritance hierarchies in interval-valued fuzzy concept-networks," Fuzzy Sets Syst., vol. 84, no. 1 (1996): 75–83.
[3] Chen, S. M. and Wang, J. Y. "Document retrieval using knowledge based fuzzy information retrieval techniques," IEEE Trans. Syst., Man, Cybern., vol. 25 (1995): 793–803.
[4] Her, G. T. and Ke, J. S. "A fuzzy information retrieval system model," in Proc. 1983 National Computer Symp., Taiwan, R.O.C., 147–151, 1983.
[5] Horng, Y. J. and Chen, S. M. "Document retrieval based on extended fuzzy concept networks," in Proc. 4th Nat. Conf. Defense Management, Taipei, Taiwan, R.O.C., vol. 2 (1996): 1039–1050.
[6] Horng, Y. J. and Chen, S. M. "Fuzzy query processing for document retrieval based on extended fuzzy concept networks," IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, vol. 29, no. 1 (1999), 96-104.
[7] Horng, Y. J., Chen, S. M. and Lee, C. H. "A fuzzy information retrieval method using fuzzy-valued concept networks," Proceedings of the IEEE 10th International Conference on Tools with Artificial Intelligence, Taipei, Taiwan, Republic of China, 104-111, November 1998.
[8] Horng, Y. J., Chen, S. M. and Lee, C. H. "Automatically constructing multi-relationship fuzzy concept networks for document retrieval," Applied Artificial Intelligence: An International Journal, vol. 17, no. 4 (2003): 303-328.
[9] Kandel, S. "Fuzzy Mathematical Techniques with Applications." Reading, MA: Addison-Wesley, 1986.
[10] Kamel, M., Hadfield, B., and Ismail, M. "Fuzzy query processing using clustering techniques," Inf. Process. Manage., vol. 26, no. 2 (1990), 279–293.
[11] Kim, M. H., Lee, J. H. and Lee, Y. J. "Analysis of fuzzy operators for high quality information retrieval," Information Processing Letters 46 (6), 251-256, 1993.
[12] Kracker, M. "A fuzzy concept network model and its applications," in Proc. 1st IEEE Int. Conf. Fuzzy Systems, 761–768, 1992.
[13] Kraft, D. H. and Buell, D. A. "Fuzzy sets and generalized Boolean retrieval systems," Int. J. Man-Mach. Stud., vol. 19, no. 1 (1983): 45–56.
[14] Lucarella, D. and Morara, R. "FIRST: Fuzzy information retrieval system," J. Inf. Sci., vol. 17 (1991): 81–91.
[15] Miyamoto, S. "Information retrieval based on fuzzy associations," Fuzzy Sets Syst., vol. 38 (1990): 191–205.
[16] Moradi, P., Ebrahim, M. and Ebadzadeh, M. M. "Personalizing results of information retrieval systems using extended fuzzy concept networks," Iran., 2008.
[17] Murai, T., Miyakoshi, M. and Shimbo, M. "A fuzzy document retrieval method based on two-valued indexing," Fuzzy Sets Syst., vol. 30 (1989), 103–120.
[18] Radechi, T. "Mathematical model of time effective information retrieval system based on the theory of fuzzy set," Inf. Process. Manage., vol. 13 (1977): 109–116.
[19] Radechi, T. "Fuzzy set theoretical approach to document retrieval," Inf. Process. Manage., vol. 15 (1979), 247–259.
[20] Salton, G. and Mcgill, M. J. "Introduction to Modern Information Retrieval." New York: McGraw-Hill, 1983.
[21] Tahani, V. "A fuzzy model of document retrieval system," Inf. Process. Manage., vol. 12 (1976): 177–187.
[22] Wang, J. Y. and Chen, S. M. "A knowledge-based method for fuzzy information retrieval," in Proc. 1st Asian Fuzzy Systems Symp., Singapore, 1993.
[23] Zadeh, L. A. "Fuzzy sets," Inf. Contr., vol. 8 (1965): 338–353.
[24] Zemankova, M. "FIIS: A fuzzy intelligent information system," Data Eng., vol. 12, no. 2 (1989): 11–20, 1989.