Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT
Fuzzy Sets and Systems 140 (2003) 183–205
FUZZY sets and systems
www.elsevier.com/locate/fss

Fuzzy information retrieval based on multi-relationship fuzzy concept networks

Shyi-Ming Chen[a,*], Yih-Jen Horng[b], Chia-Hoang Lee[b]

[a] *Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, ROC*
[b] *Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan, ROC*

Received 16 June 2002; received in revised form 3 September 2002; accepted 24 September 2002

Abstract

In this paper, we present a new method for fuzzy information retrieval based on multi-relationship fuzzy concept networks. There are four kinds of fuzzy relationships in a multi-relationship fuzzy concept network, i.e., "fuzzy positive association" relationship, "fuzzy negative association" relationship, "fuzzy generalization" relationship and "fuzzy specialization" relationship. By performing fuzzy inferences based on the multi-relationship fuzzy concept network, the fuzzy information retrieval system can retrieve documents containing concepts that are not directly specified by the user but are somehow related to the user's query. In order to perform fuzzy inferences more efficiently, we use concept matrices to represent the degrees of fuzzy relationships between concepts in a multi-relationship fuzzy concept network. By calculating the transitive closures of concept matrices, the implicit degrees of fuzzy relationships between concepts are obtained. Multiple degrees of satisfaction that a document satisfies the user's query with respect to the fuzzy relationships between concepts are calculated. These satisfaction degrees are aggregated according to the user's specification to find the most relevant documents with respect to the user's query. The proposed fuzzy information retrieval method is more flexible and more intelligent than the one we presented in (IEEE Trans. Systems Man Cybernet.—Part B: Cybernet. 29(1) (1999) 126).
© 2002 Elsevier B.V. All rights reserved.

Keywords: Concept matrices; Document descriptor matrices; Fuzzy information retrieval; Multi-relationship fuzzy concept networks; OWA operators

* Corresponding author.

0165-0114/03/$ - see front matter © 2002 Elsevier B.V. All rights reserved.
PII: S0165-0114(02)00464-5

1. Introduction

In [19], Lucarella pointed out that imprecise and uncertain information comes from three major aspects in an information retrieval system environment including the representations of users' queries, the representations of documents, and the relevance relationships between users' queries and documents. In order to represent and process the imprecise and uncertain information in information retrieval systems, the fuzzy set theory [28,29] has been applied to information retrieval systems [1,2,4–7,11,13–15,17–21]. In [2], Bordogna et al. proposed a fuzzy document representation method which can support user adaptation in information retrieval systems. In [1], Bezdek et al. presented a method to calculate the transitive closures of fuzzy thesauri for information retrieval systems. In [7], we used concept matrices and document descriptor matrices to model concept networks for fuzzy information retrieval. In [6], we proposed an information retrieval method based on interval valued fuzzy concept networks, where the degrees of association between concepts are represented by interval values. In [4], we proposed an information retrieval method based on extended fuzzy concept networks, where the relationships between concepts can be one of the four fuzzy relationships, i.e., fuzzy positive association relationship, fuzzy negative association relationship, fuzzy generalization relationship and fuzzy specialization relationship [16]. In [5], we proposed an information retrieval method based on fuzzy valued concept networks, where the degrees of association between concepts are represented by fuzzy numbers. In [11], Jing et al. proposed an approach to construct collection-dependent association thesauri automatically using large full-text document collections. In [13], Kim et al. proposed a personalized web search engine using fuzzy concept networks with the link structure. In [14], Kim et al. proposed a query term expansion and reweighting method which consider the term co-occurrence within the feedbacked documents. In [15], Kóczy et al. proposed a method using fuzzy tolerance and similarity relations in fuzzy information retrieval systems. In [17], Kraft et al. explored several ways of using fuzzy clustering techniques in information retrieval systems, where the most important one is to capture the relationships among index terms. They used the fuzzy logic rules to represent the association relationships between index terms and to form the basis of the association mechanism. In [18], Larsen et al. presented a method to use fuzzy relational thesauri for solving classificatory problem in information retrieval systems and expert systems. In [19], Luearella discussed the design and implementation issues for a fuzzy information retrieval system using a knowledge-based approach. In [20], Lucarella et al. presented a concept network structure which acts as a knowledge base for fuzzy information retrieval. In [21], Miyamoto proposed two approaches for information retrieval through fuzzy associations.

However, the methods presented in [6,7,20] are restricted since the actual relationships between concepts cannot be explicitly specified in their concept network models. In [4], in order to overcome the drawbacks of [6,7,20], we have extended the fuzzy concept networks in [6,7] based on [16] to let the concepts be explicitly linked by one of the four fuzzy relationships, i.e., fuzzy positive association relationship, fuzzy negative association relationship, fuzzy generalization relationship and fuzzy specialization relationship. Furthermore, in [4], we also proposed some methods to deal with users' queries based on the extended fuzzy concept networks for fuzzy information retrieval. However, there are still some drawbacks in [6], i.e., in a extended fuzzy concept network, each pair of concepts can be related to each other by only one kind of relationship. However, since the relationship between concepts may vary from different perspectives, the precise relationship between concepts may be difficult to define. For example, the "internet" and the "intranet" are two concepts regarding computer networks and they are antonym concepts if the network size is concerned. However, they also can be regarded as similar concepts due to the fact that they essentially deal with connecting computers together. In [22], Miyamoto pointed out that there are no theoretical reasons to avoid multiple fuzzy relationships defined at the same time between each concept pair. Therefore, if we can let the concept pairs in fuzzy concept networks have multiple relationships simultaneously, then the fuzzy information retrieval systems can deal with the users' queries in a more flexible and more intelligent manner.

In this paper, we extend the work we presented in [4] to allow multiple fuzzy relationships between each pair of concepts in fuzzy concept networks, where each relationship has its own strength. Multiple degrees of satisfaction that a document satisfies the user's query are individually calculated when considering each kind of fuzzy relationship between concepts. Then, these multi-relationship satisfaction degrees are aggregated according to the user's specification to find the most relevant documents with respect to the user's query. Since the concept pairs in a multi-relationship fuzzy concept network can have multiple relationships simultaneously, the proposed fuzzy information retrieval method is more flexible and more intelligent than the existing methods.

The rest of this paper is organized as follows. In Section 2, we briefly review the definitions of concept networks and extended fuzzy concept networks from [4,20], respectively. In Section 3, we present the definitions of multi-relationship fuzzy concept networks. In Section 4, we present a fuzzy information retrieval method based on multi-relationship fuzzy concept networks. In Section 5, we use an example to illustrate the proposed fuzzy information retrieval method. The conclusions are discussed in Section 6.

2. Concept networks and extended fuzzy concept networks

In [20], Lucarella et al. proposed a fuzzy information retrieval method based on concept networks. A concept network includes nodes and directed links, where each node represents a concept or a document. A link associated with a real value μ between zero and one connecting two distinct concept nodes means that these two concepts are semantically related with strength μ, where $\mu \in [0, 1]$. A link with a real value μ between zero and one connecting a concept node and a document node means that the content of this document contains the linked concept with strength μ, where $\mu \in [0, 1]$. For more details about concept networks, please refer to [20].

In [4], we have presented the definitions of extended fuzzy concept networks. The extended fuzzy concept networks are more general than the concept networks presented in [20]. In an extended fuzzy concept network, there are four kinds of fuzzy relationships between concepts, i.e., fuzzy positive association, fuzzy negative association, fuzzy generalization and fuzzy specialization [16]. The fuzzy relationships between concepts are reviewed from [4] as follows:

(1) *Fuzzy positive association*: It relates concepts which have a fuzzy similar meaning (e.g., person ↔ individual) in some contexts.
(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., big ↔ small) in some contexts.

(3) *Fuzzy generalization*: A concept is regarded as a fuzzy generalization of another concept if it consists of that concept (e.g., machine → screw) or it includes that concept (e.g., vehicle → car) in a partitive sense.
(4) *Fuzzy specialization*: It is the inverse of the fuzzy generalization relationship. That is, a concept is regarded as a fuzzy specialization of another concept if it is a part of that concept (e.g., screw → machine) or it is a kind of that concept (e.g., car → vehicle).

The properties of these fuzzy relationships are summarized as follows [4]:

Definition 2.1. Let C be a set of concepts. Then,

(1) "Fuzzy positive association" P is a fuzzy relation, $P: C\times C \to [0,1]$, which is reflexive, symmetric, and max-*-transitive.
(2) "Fuzzy negative association" N is a fuzzy relation, $N: C\times C \to [0,1]$, which is anti-reflexive, symmetric, and max-*-nontransitive.
(3) "Fuzzy generalization" G is a fuzzy relation, $G: C\times C \to [0,1]$, which is anti-reflexive, anti-symmetric, and max-*-transitive.
(4) "Fuzzy specialization" S is a fuzzy relation, $S: C\times C \to [0,1]$, which is anti-reflexive, anti-symmetric, and max-*-transitive.

An extended fuzzy concept network consists of nodes and directed links, where each node denotes a concept or a document. Each directed link connects two concepts or connects from a concept c_i to a document d_j and has one of the following formats:

(1) $c_i \xrightarrow{(\mu,P)} c_j$ means that there is a fuzzy positive association relationship between concept c_i and concept c_j, and the associated degree is μ, where $\mu \in [0,1]$.
(2) $c_i \xrightarrow{(\mu,N)} c_j$ means that there is a fuzzy negative association relationship between concept c_i and concept c_j, and the associated degree is μ, where $\mu \in [0,1]$.
(3) $c_i \xrightarrow{(\mu,G)} c_j$ means that concept c_i is more general than concept c_j, and the associated degree is μ, where $\mu \in [0,1]$.
(4) $c_i \xrightarrow{(\mu,S)} c_j$ means that concept c_i is more special than concept c_j, and the associated degree is μ, where $\mu \in [0,1]$.
(5) $c_i \xrightarrow{(\mu,P)} d_j$ means that there is a fuzzy positive association relationship between concept c_i and document d_j, and the associated degree is μ, where $\mu \in [0,1]$ (i.e., document d_j possesses concept c_i with the degree $\mu \times 100\%$).
(6) $c_i \xrightarrow{(\mu,N)} d_j$ means that there is a negative association relationship between concept c_i and document d_j, and the relevance degree is μ, where $\mu \in [0,1]$ (i.e., document d_j possesses a concept which is $\mu \times 100\%$ complementary to the concept c_i).
(7) $c_i \xrightarrow{(\mu,G)} d_j$ means that there is a generalization relationship between concept c_i and document d_j, and the relevance degree is μ, where $\mu \in [0,1]$ (i.e., concept c_i is more general than the concept possessed by document d_j with the degree of $\mu \times 100\%$).

(8) $c_i \xrightarrow{(\mu, G)} d_j$ means that there is a specialization relationship between concept c_i and document d_j, and the relevance degree is μ, where $\mu \in [0,1]$ (i.e., concept c_i is more special than the concept possessed by document d_j with the degree of $\mu \times 100\%$).

For more details about extended fuzzy concept networks, please refer to [4].

3. Multi-relationship fuzzy concept networks

In this section, we present the definitions of multi-relationship fuzzy concept networks [10]. The concepts of multi-relationship fuzzy concept networks are similar to the concepts of semantic networks [23] due to the fact that they both consist of nodes and directed links. Each link directed from one node to another is associated with a label to indicate the relationship between these two nodes. However, there are two main differences between the concepts of multi-relationship fuzzy concept networks and the concepts of semantic networks. Firstly, the relationships between nodes in a semantic network can be arbitrarily defined. But in a multi-relationship fuzzy concept network, the types of relationships are restricted. Therefore, the complexity of inferences based on the multi-relationship fuzzy concept network can be reduced. Secondly, in a semantic network, nodes are linked by only one relationship and the relationship is crisp. In a multi-relationship fuzzy concept network, the concepts can be related to other concepts by more than one relationship at the same time, each with its own degree. Therefore, a multi-relationship fuzzy concept network has the capability to represent various relationships between the same pair of concepts in different point of views. Our aim of developing the multi-relationship fuzzy concept networks is to provide a more powerful knowledge representation method which is more appropriate than the semantic networks in the information retrieval environment.

We use four kinds of fuzzy relationships [16] (i.e., "fuzzy positive association" relationship, "fuzzy negative association" relationship, "fuzzy generalization" relationship and "fuzzy specialization" relationship) to describe the possible relationships between concepts in a multi-relationship fuzzy concept network.

Definition 3.1. A multi-relationship fuzzy concept network is denoted as *MRFCN* (E, L), where E is a set of nodes and each node stands for a concept or a document; L is a set of directed edges between nodes. If $\ell \in L$, then the directed edge ℓ has the following two formats:

(1) $c_i \xrightarrow{(\langle \mu_P, P\rangle, \langle \mu_N, N\rangle, \langle \mu_G, G\rangle, \langle \mu_S, S\rangle)} c_j$ means that the directed edge ℓ connects from concept c_i to concept c_j associated with a quadruple $(\langle \mu_P, P\rangle, \langle \mu_N, N\rangle, \langle \mu_G, G\rangle, \langle \mu_S, S\rangle)$, where μ_P indicates the degree of "fuzzy positive association" relationship P between concept c_i and concept c_j (i.e., concept c_i is similar to concept c_j with degree μ_P), μ_N indicates the degree of "fuzzy negative association" relationship N between concept c_i and concept c_j (i.e., concept c_i and concept c_j are complementary, incompatible or antonyms with degree μ_N), μ_G indicates the degree of "fuzzy generalization" relationship G between concept c_i and concept c_j (i.e., concept c_i is more general than concept c_j with degree μ_G), μ_S indicates the degree of "fuzzy specialization" relationship S between concept c_i and concept c_j (i.e., concept c_i is more special than concept c_j with degree μ_S), $\mu_P \in [0,1]$, $\mu_N \in [0,1]$, $\mu_G \in [0,1]$, and $\mu_S \in [0,1]$. The larger the value of μ_r, the more the concept c_i is related to concept c_j by fuzzy

relationship r, where $r \in \{P, N, G, S\}$. If $\mu_r = 0$, then concept c_i is not related to concept c_j by fuzzy relationship r, where $r \in \{P, N, G, S\}$.

(2) $c_i \xrightarrow{\mu} d_j$ means that the directed edge ℓ connects from concept c_i to document d_j associated with a degree μ, indicating the strength of document d_j containing concept c_i, where $\mu \in [0, 1]$. The larger the value of μ, the more the document d_j contains concept c_i.

The purpose of allowing the relationships between the same pair of concepts being multiple defined is to extend the generality of the knowledge base architecture. In different application domains, the kind of relationship between any two concepts may be defined differently from various points of views. Thus, if only one kind of relationship among several possible relationships between concepts can be kept in the knowledge base, then the resulting knowledge base is restricted and only appropriate for some specific application domains.

Generally speaking, in a multi-relationship fuzzy concept network, simple concepts are usually related by only one kind of relationship, but comprehensive concepts may be related by several kinds of relationships. Fox example, "*male*" and "*female*" are two simple concepts which represent the gender of creatures and can be related to each other only by the "fuzzy negative association" relationship. On the other hand, "*man*" and "*woman*" are two more comprehensive concepts representing not only the gender of creatures but also the sort of creatures (i.e., human beings). Therefore, "*man*" and "*woman*" can be related to each other not only by the "fuzzy negative association" relationship from the aspect of gender they represent, but also by the "fuzzy positive association" relationship due to the fact that they represent the same sort of creatures.

In the following, we use an example to illustrate a complete multi-relationship fuzzy concept network.

Example 3.1. Assume that there is a multi-relationship fuzzy concept network as shown in Fig. 1, where $c_1, c_2, \ldots,$ and c_6 are concept nodes, and d_1, d_2 and d_3 are document nodes. Each concept node stands for a concept as shown in Table 1, and each document node stands for a document as shown in Table 2.

From Fig. 1, we can see that concept c_1 (i.e., *Security and Encryption*) is both contained in document d_1 (with degree $= 0.1$) and document d_2 (with degree $= 0.7$). The associated degrees indicate that document d_2 contains more about the topic of security and encryption than document d_1. Concept c_2 (i.e., *Internet*) is related to concept c_5 (i.e., *Intranet*) by the "fuzzy positive association" relationship (with degree $= 0.7$) and also by the "fuzzy negative association" relationship (with degree $= 0.7$) due to the fact that *Internet* and *Intranet* are two similar concepts both about connecting computers, but they are also antonyms to each other from the aspect of their network sizes.

In a multi-relationship fuzzy concept network, the relationships and their associated degrees between concepts are specified by domain experts. However, sometimes the domain experts may forget to set some relationships and their associated degrees between concepts. Moreover, even if the relationships are given by the domain experts, some implicit links do not show in the multi-relationship fuzzy concept network. For example, consider the multi-relationship fuzzy concept network shown in Fig. 2. It contains three concepts "*Computer Science*", "*Network*" and "*Intranet*", and three explicit directed links l_1, l_2 and l_3 between these concepts.

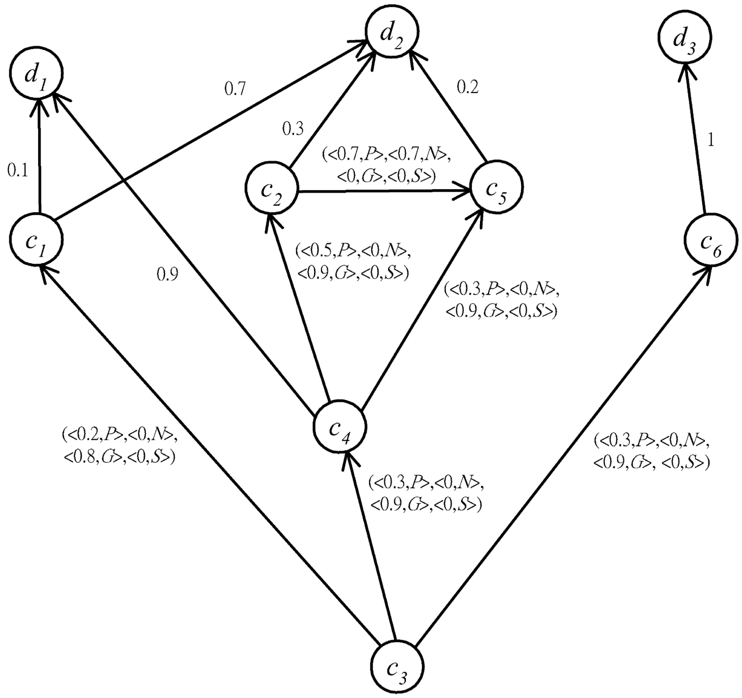

Fig. 1. A multi-relationship fuzzy concept network of Example 3.1.

Table 1
Concept nodes and their corresponding concepts

Concept nodes	Concepts
c_1	*Security and Encryption*
c_2	*Internet*
c_3	*Computer Science*
c_4	*Networks*
c_5	*Intranet*
c_6	*Artificial Intelligence*

Table 2
Document nodes and their corresponding documents

Document nodes	Document titles
d_1	*Computer Networks*
d_2	*Internet and Intranet Security*
d_3	*Artificial Intelligence*

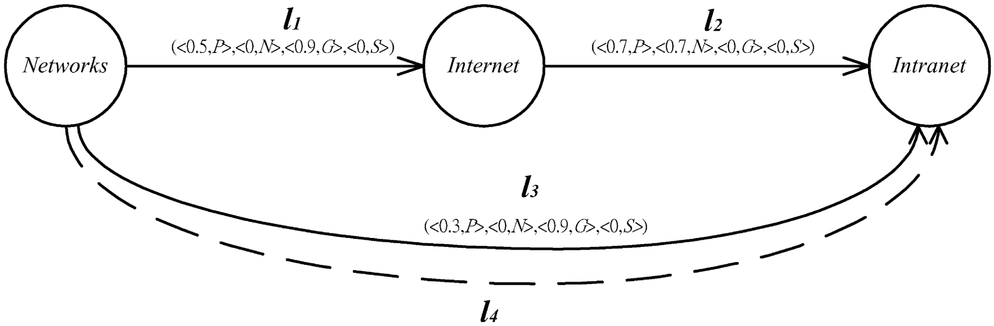

Fig. 2. Inferring the implicit directed link l_4 from the concept "*Networks*" to the concept "*Intranet*".

Fig. 2 shows that besides the explicit directed link l_3 specified by the domain experts from the concept "*Networks*" to the concept "*Intranet*", there also exists an implicit directed link l_4 between these two concepts which is inferred through the directed link l_1 from the concept "*Networks*" to the concept "*Internet*" and through the directed link l_2 from the concept "*Internet*" to the concept "*Intranet*". In this case, the actual degrees of relationships between the concept "*Networks*" and the concept "*Intranet*" should be calculated by aggregating the degrees of relationships associated with the directed link l_3 and the ones associated with the directed link l_4. The methods of inferring implicit directed links and aggregating the degrees of relationships between concepts are described as follows. In a multi-relationship fuzzy concept network, if fuzzy relationship r is transitive, i.e., $r \in \{P, G, S\}$, and the degree of fuzzy relationship r between concept c_i and concept c_j is μ_{ij}^r, where $\mu_{ij}^r \in [0,1]$, and if the degree of fuzzy relationship r between concept c_j and concept c_k is μ_{jk}^r, where $\mu_{jk}^r \in [0,1]$, then the degree μ_{ik}^r of fuzzy relationship r between concept c_i and concept c_k can be inferred as follows:

$$\mu_{ik}^r = \quad \mu_{ij}^r \bullet \mu_{jk}^r, \tag{1}$$

where $\bullet$ is the arithmetic product and $\mu_{ik}^r \in [0,1]$. Furthermore, if the degree of fuzzy relationship r between concept c_1 and concept c_2 is μ_{12}^r, the degree of fuzzy relationship r between concept c_2 and concept c_3 is $\mu_{23}^r, \ldots,$ and the degree of fuzzy relationship r between concept c_{n-1} and concept c_n is $\mu_{(n-1)n}^r$, where $\mu_{12}^r \in [0,1]$, $\mu_{23}^r \in [0,1], \ldots,$ and $\mu_{(n-1)n}^r \in [0,1]$, then the degree of fuzzy relationship r between concept c_1 and concept c_n is μ_{1n}^r, where $\mu_{1n}^r \in [0,1]$ and

$$\mu_{1n}^r = \mu_{12}^r \bullet \mu_{23}^r \bullet \cdots \bullet \mu_{(n-1)n}^r. \tag{2}$$

If there are h routes between concept c_1 and concept c_n, then the actual degree of fuzzy relationship r between concept c_1 and concept c_n can be calculated as follows:

$$\mu_{1n}^r = \max(\mu_{1n}^{r(1)}, \mu_{1n}^{r(2)}, \ldots, \mu_{1n}^{r(h)}), \tag{3}$$

where $\mu_{1n}^{r(i)}$ denotes the evaluated degree of fuzzy relationship r of the ith route that started from concept c_1 and ended at concept c_n, and $1 \leqslant i \leqslant h$.

Therefore, based on formula (1) and Fig. 2, we can see that the degrees of fuzzy relationships associated with the implicit link l_4 are $\mu_P = 0.35$ (i.e., the degree of the "fuzzy positive association" relationship is equal to 0.35) and $\mu_N = \mu_G = \mu_S = 0$ (i.e., the degrees of the "fuzzy negative association" relationship, the "fuzzy generalization" relationship and the "fuzzy specialization" relationship are all equal to 0). Furthermore, based on formula (3), the actual degrees of fuzzy relationships between the concept "*Networks*" and the concept "*Intranet*" are $\mu_P = 0.35, \mu_G = 0.9, \mu_N = 0$ and $\mu_S = 0$.

In a multi-relationship fuzzy concept network, the document descriptor of a document consists of concepts having a directed link to this document. Let C be a set of concepts in a multi-relationship fuzzy concept network, $C = \{c_1, c_2, \ldots, c_n\}$, where n is the number of concepts, and let D be a set of documents in a multi-relationship fuzzy concept network, $D = \{d_1, d_2, \ldots, d_m\}$, where m is the number of documents. The document descriptor of a document d_i, $d_i \in D$, is defined as a fuzzy subset in C:

$$d_i = \{(c_j, f_{d_i}(c_j)) \mid c_j \in C\},$$

where $f_{d_i}(c_i)$, $f_{d_i}: C \to [0, 1]$, denotes the degree in which document d_i contains concept c_j. Based on the fuzzy inference through the links of the multi-relationship fuzzy concept network, the document descriptor can be expanded to derive the membership degrees of concepts that originally have zero membership degrees in the document descriptor but are related to the concepts having non-zero membership degrees in the document descriptor. However, since concepts may be related to each other by four possible kinds of fuzzy relationships, the document descriptor should be expanded by each kind of fuzzy relationship. That is, there will be four expanded document descriptors for each document, and each expanded document descriptor consists of concepts related to this document by one of the four kinds of fuzzy relationships. The method of inferring the relationships and their associated degrees between concepts and documents is described as follows. Let C_{d_i} be a set of concepts of document d_i which have nonzero membership degree as shown in Fig. 3, where $C_{d_i} = \{c_j, c_x, \ldots, c_y\}$. From Fig. 3, we can see that the degrees of document d_i containing concepts $c_j, c_x, \ldots,$ and c_y are $\mu_{ij}, \mu_{ix}, \ldots,$ and μ_{iy}, respectively, where $\mu_{ij} \in [0, 1]$, $\mu_{ix} \in [0, 1], \ldots,$ and $\mu_{iy} \in [0, 1]$. Furthermore, from Fig. 3, we can see that the degree of fuzzy relationship r between concept c_j and concept c_k is μ^r_{jk}, the degree of fuzzy relationship r between concept c_x and concept c_k is $\mu^r_{xk}, \ldots,$ and the degree of fuzzy relationship r between concept c_y and concept c_k is μ^r_{yk}, where $c_j \in C_{d_i}$, $c_x \in C_{d_i}, \ldots,$ $c_y \in C_{d_i}$, $r \in \{P, N, G, S\}$, $\mu^r_{jk} \in [0, 1]$, $\mu^r_{xk} \in [0, 1], \ldots,$ and $\mu^r_{yk} \in [0, 1]$. Then, the degree μ^r_{ik} of document d_i relating to concept c_k by fuzzy relationship r can be evaluated as follows:

$$\mu^r_{ik} = \max(\mu_{ij} \bullet \mu^r_{jk}, \mu_{ix} \bullet \mu^r_{xk}, \ldots, \mu_{iy} \bullet \mu^r_{yk}), \tag{4}$$

where $\mu^r_{ik} \in [0, 1]$, $\mu_{ij} \in [0, 1]$, $\mu^r_{jk} \in [0, 1]$, $\mu_{ix} \in [0, 1]$, $\mu^r_{xk} \in [0, 1], \ldots, \mu_{iy} \in [0, 1]$, $\mu^r_{yk} \in [0, 1]$, and $r \in \{P, N, G, S\}$. The expanded document descriptors of the documents will form the basis for document retrieval.

4. Fuzzy query processing for document retrieval based on multi-relationship fuzzy concept networks

In this section, we present a new method for fuzzy query processing for document retrieval based on multi-relationship fuzzy concept networks. In order to increase the inference efficiency, we adopt the method we presented in [7] to model the multi-relationship fuzzy concept network by several

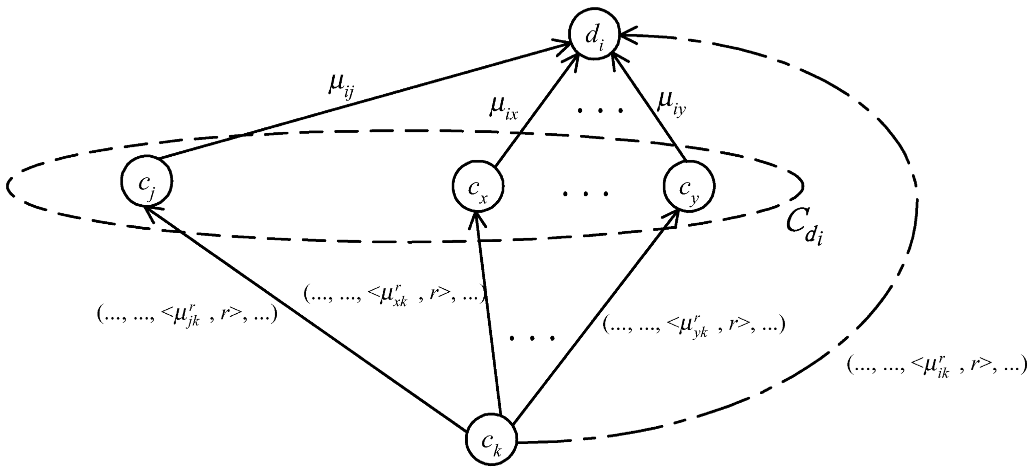

Fig. 3. A multi-relationship fuzzy concept network.

concept matrices. We use four concept matrices to represent the degrees of fuzzy relationships between concepts in a multi-relationship fuzzy concept network; each concept matrix describes the degrees of one kind of fuzzy relationship (i.e., “fuzzy positive association” relationship, “fuzzy negative association” relationship, “fuzzy generalization” relationship and “fuzzy specialization” relationship) between concepts. By computing the transitive closures of these concept matrices, the implicit degrees of fuzzy relationships between concepts can be obtained.

Definition 4.1. Let C be the set of concepts in a multi-relationship fuzzy concept network, where $C=\{c_1,c_2,\dots,c_n\}$ and n is the number of concepts. A concept matrix V_r is a fuzzy matrix [12]:

$$V_r = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix},$$

where $r\in\{P,N,G,S\}, n$ is the number of concepts, $v_{ij}\in[0,1]$, $1\leqslant i\leqslant n$, and $1\leqslant j\leqslant n$. The element $V_r(c_i,c_j)$ denotes the degree of fuzzy relationship r from concept c_i to concept c_j in the multi-relationship fuzzy concept network, where $r\in\{P,N,G,S\}$. If the fuzzy relationship r is reflexive, then $V_r(c_i,c_i)=1$. Otherwise, $V_r(c_i,c_i)=0$, where $1\leqslant i\leqslant n$. If the fuzzy relationship r is symmetric, then $V_r(c_i,c_j)=V_r(c_j,\ c_i)$. If $V_r\ (c_i,c_j)=0$, then the degree between concept c_i and concept c_j is not defined explicitly by the experts in the multi-relationship fuzzy concept network.

Note that the fuzzy generalization relationship is the inverse of the fuzzy specialization relationship. If there is a fuzzy generalization relationship from concept c_i to concept c_j, then there will exist a fuzzy specialization relationship from concept c_i to concept c_i, and vice versa. Thus, if there is a fuzzy generalization from concept c_i to concept c_j with degree μ, then we let $V_S(c_j,c_i)=V_G(c_i,c_j)=\mu$;

if there is a fuzzy specialization from concept c_i to concept c_j with degree μ, then we let $V_G(c_j, c_i) = V_S(c_i, c_j) = \mu$, where $\mu \in [0, 1]$.

Definition 4.2. Assume that V_r is a concept matrix, where $r \in \{P, N, G, S\}$. If the fuzzy relationship r is nontransitive, then we let the transitive closure V_r^* of V_r be itself, i.e., $V_r^* = V_r$. If the fuzzy relationship r is transitive, then the transitive closure V_r^* is defined as follows. Let

$$V_r^2 = V_r \otimes V_r = \begin{bmatrix} \bigvee\limits_{i=1,\dots,n} (v_{1i} \bullet v_{i1}) & \bigvee\limits_{i=1,\dots,n} (v_{1i} \bullet v_{i2}) & \cdots & \bigvee\limits_{i=1,\dots,n} (v_{1i} \bullet v_{in}) \\ \bigvee\limits_{i=1,\dots,n} (v_{2i} \bullet v_{i1}) & \bigvee\limits_{i=1,\dots,n} (v_{2i} \bullet v_{i2}) & \cdots & \bigvee\limits_{i=1,\dots,n} (v_{2i} \bullet v_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ \bigvee\limits_{i=1,\dots,n} (v_{ni} \bullet v_{i1}) & \bigvee\limits_{i=1,\dots,n} (v_{ni} \bullet v_{i2}) & \cdots & \bigvee\limits_{i=1,\dots,n} (v_{ni} \bullet v_{in}) \end{bmatrix}, \tag{5}$$

where "$\bigvee$" is the maximum operator and "$\bullet$" is the arithmetic product. Moreover, let

$$V_r^k = V_r^{k-1} \cup V_r, \tag{6}$$

where "$\cup$" is the union operator (i.e., $V_r^k(c_i, c_j) = \max(V_r(c_i, c_j), V_r^{k-1}(c_i, c_j))$ and the powers of V_r on the right-hand side are computed by formula (5). Then, there exists an integer p, where $p \geqslant 2$, which satisfies $V_r{}^p = V_r{}^{p+1}$. Let $V_r^* = V_r{}^p$, then V_r^* is called the transitive closure of V_r.

We also use a document descriptor matrix to represent the degree of documents containing concepts in a multi-relationship fuzzy concept network. By multiplying the document descriptor matrix by the transitive closures of the concept matrices, we can obtain the expanded document descriptor matrices which represent the degrees of fuzzy relationships between concepts and documents.

Definition 4.3. Let D be the set of documents, $D = \{d_1, d_2, \dots, d_m\}$, and let C be the set of concepts, $C = \{c_1, c_2, \dots, c_n\}$, in a multi-relationship fuzzy concept network. The document descriptor matrix U is shown as follows:

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mn} \end{bmatrix},$$

where m denotes the number of documents, n denotes the number of concepts, u_{ij} stands for the degree of document d_i containing concept c_j, $u_{ij} \in [0, 1]$, $1 \leqslant i \leqslant m$, and $1 \leqslant j \leqslant n$. If $U(d_i, c_j) = 0$, then the degree of document d_i containing concept c_j is not defined explicitly by the experts in the multi-relationship fuzzy concept network.

Definition 4.4. Assume that U is a document descriptor matrix and V_r^* is the transitive closure of concept matrix V_r, where $r \in \{P, N, G, S\}$. Let

$$U_r^* = U \otimes V_r^* = \begin{bmatrix} \bigvee_{i=1,\ldots,n} (u_{1i} \bullet v_{i1}) & \bigvee_{i=1,\ldots,n} (u_{1i} \bullet v_{i2}) & \cdots & \bigvee_{i=1,\ldots,n} (u_{1i} \bullet v_{in}) \\ \bigvee_{i=1,\ldots,n} (u_{2i} \bullet v_{i1}) & \bigvee_{i=1,\ldots,n} (u_{2i} \bullet v_{i2}) & \cdots & \bigvee_{i=1,\ldots,n} (u_{2i} \bullet v_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ \bigvee_{i=1,\ldots,n} (u_{mi} \bullet v_{i1}) & \bigvee_{i=1,\ldots,n} (u_{mi} \bullet v_{i2}) & \cdots & \bigvee_{i=1,\ldots,n} (u_{mi} \bullet v_{in}) \end{bmatrix}, \quad (7)$$

where u_{ij} is an element of U, $1 \leqslant i \leqslant m$, and $1 \leqslant j \leqslant n$; v_{ij} is an element of V_r^*, $1 \leqslant i \leqslant n$, and $1 \leqslant j \leqslant n$; "$\bigvee$" is the maximum operator and "$\bullet$" is the arithmetic product. In this case, U_r^* is referred to as an expanded document descriptor matrix of document descriptor matrix U by fuzzy relationship r, where $r \in \{P, N, G, S\}$. Since there are four kinds of fuzzy relationships defined in a multi-relationship fuzzy concept network, we can obtain four expanded document descriptor matrices. These four expanded document descriptor matrices are then used as a basis for similarity measures between the users' queries and the documents.

A user's query is represented by a query descriptor Q expressed as a fuzzy subset of the collection of concepts by the following expression:

$$Q = \{(c_1, x_1), (c_2, x_2), \ldots, (c_n, x_n)\},$$

where $x_i \in [0, 1]$ denotes the relevance degree of the query descriptor Q with respect to concept c_i, i.e., the degree that the user thinks concept c_i should be contained in the retrieved documents. If $x_i = 0$, then it indicates that documents desired by the user do not possess concept c_i. If $x_i =$ "-", then it indicates that the degree of the desired documents with respect to concept c_i can be neglected. Furthermore, the user's query Q can be represented by a query descriptor vector $\bar{q}$ [7] shown as follows:

$$\bar{q} = \langle x_1, x_2, \ldots, x_n \rangle,$$

where $x_i \in [0, 1]$ indicates the desired degree of the document with respect to concept c_i and $1 \leqslant i \leqslant n$.

Before introducing the function for calculating the degree of similarity between document descriptors and users' queries, we first introduce a function for calculating the degree of similarity between two real values between zero and one. Let x and y be two values, where $x \in [0, 1]$ and $y \in [0, 1]$. Then, the degree of similarity between x and y can be evaluated by the function T [7]:

$$T(x, y) = 1 - |x - y|, \quad (8)$$

where $T(x, y) \in [0, 1]$. The larger the value of $T(x, y)$, the more the similarity between x and y.

Assume that the document descriptor vector $\overline{dr_i}$ (i.e., the ith row of the expanded document descriptor matrix U_r^*, where $r \in \{P, N, G, S\}$), and the query descriptor vector $\bar{q}$ are represented as follows:

$$\overline{dr_i} = \langle s_{i1}, s_{i2}, \ldots, s_{in} \rangle,$$

$$\bar{q} = \langle x_1, x_2, \ldots, x_n \rangle,$$

where $s_{ij} \in [0,1]$, $x_i \in [0,1]$, $1 \leqslant j \leqslant n$, $1 \leqslant i \leqslant m$, n is the number of concepts, m is the number of documents, and $r \in \{P, N, G, S\}$. Let $\overline{q(j)}$ be the jth element of the query descriptor vector $\bar{q}$. If $\overline{q(j)} =$ "-", then it indicates that concept c_j is neglected by the user's query. The degree of satisfaction $DS_r(d_i)$ that document d_i satisfies the user's query Q by fuzzy relationship r can be evaluated as follows [7]:

$$DS_r(d_i) = \frac{\sum_{\overline{q(j)} \neq \text{"-"} \text{ and } j=1,2,\ldots,n} T(s_{ij}, x_j)}{k}, \tag{9}$$

where $DS_r(d_i) \in [0,1]$, $1 \leqslant i \leqslant m$, and k is the number of concepts not neglected by the user's query. The larger the value of $DS_r(d_i)$, the more the degree of satisfaction that document d_i satisfies the user's query by fuzzy relationship r, $r \in \{P, N, G, S\}$.

The degrees of satisfaction that the document satisfies the user's query by different fuzzy relationships are then aggregated to obtain the overall satisfaction that the document satisfies the user's query. The user can choose the aggregation operation according to his/her needs in one of the following ways: (1) by setting an importance weight to each degree of satisfaction $DS_r(d_i)$, (2) by setting an importance order of all fuzzy relationships, or (3) by using some predefined simple linguistic quantifiers to specify how many fuzzy relationships should be taken into account. In this paper, we utilize the OWA operators [27] to formalize the linguistic quantifiers. In the following, we first introduce the OWA aggregation operators, and then introduce the three aggregation approaches.

In [27], Yager proposes a family of mean-like operators which are used to deal with multi-criteria decision-making problems. The arguments of these operators are weighted according to their order made by sorting the arguments and then averaged according to their weights, so these operators are named ordered weighted averaging (OWA) operators. By giving different weighting vectors, the OWA operators lie between the choices of the minimum and the maximum of the arguments. We briefly review the OWA operators [27] as follows.

Definition 4.5. An OWA operator that takes n input arguments is a mapping

$$F : R^n \rightarrow R,$$

which has a weighting vector W of dimension n associated with it. The weighting vector W has the following properties:

$$w_j \in [0,1],$$

$$\sum_{j=1}^{n} w_j = 1.$$

Moreover,

$$F(a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} w_j b_j, \tag{10}$$

where b_j is the jth largest value of the input arguments $a_1, a_2, \ldots,$ and a_n, and $1 \leqslant j \leqslant n$.

If B is a vector consisting of the ordered arguments a_i, which is called the ordered argument vector, and W^{T} is the transpose of the weighting vector, then the OWA aggregation can also be expressed as

$$F_w(a_1, a_2, \ldots, a_n) = W^{\mathrm{T}} B. \tag{11}$$

Example 4.1. Assume that we have four argument variables shown as follows:

$$a_1 = 0.5,$$

$$a_2 = 0.3,$$

$$a_3 = 0.7,$$

$$a_4 = 0.8$$

and we want to aggregate them using the weighting vector W, where

$$W = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.2 \\ 0.1 \end{bmatrix}.$$

Then, we can get the ordered argument vector B as follows:

$$B = \begin{bmatrix} 0.8 \\ 0.7 \\ 0.5 \\ 0.3 \end{bmatrix}$$

and the aggregation result is as follows:

$$F_w(0.5, 0.3, 0.7, 0.8) = (0.4)(0.8) + (0.3)(0.7) + (0.2)(0.5) + (0.1)(0.3) = 0.66.$$

The OWA operators are also applied in the information retrieval field. In [8], Damiani et al. proposed a fuzzy retrieval model to retrieve reusable components containing the features needed by users. Since each feature contributes different weights to components under different contexts (categories), they used the OWA operator associated with *context weights* to obtain the aggregated weights of the reusable components with respect to the features needed by the user. In [3], Bordogna et al. proposed a document retrieval model where the documents are divided into subparts. They used the OWA operator to aggregate the significances of the term with respect to different subparts of the document by some linguistic quantifiers which are formalized as corresponding weighting vectors.

Let D be a set of documents and let C be a set of concepts defined in a multi-relationship fuzzy concept network, $D=\{d_1,d_2,\ldots,d_m\}$ and $C=\{c_1,c_2,\ldots,c_n\}$. The degree of satisfaction that document d_i satisfies the user's query by fuzzy relationships r is denoted as $DS_r(d_i)$, where $1\leqslant i\leqslant m$ and $r\in\{P,N,G,S\}$. The three aggregation methods are described as follows:

Case 1: The user assigns an importance weight w_r to each degree of satisfaction $DS_r(d_i)$ that document d_i satisfies the user's query by fuzzy relationships r, where $0\leqslant w_r\leqslant 1$ and $r\in\{P,N,G,S\}$. The larger the value of w_r, the more important the fuzzy relationship r. Moreover, the summation of these importance weights must be equal to one (i.e., $w_P+w_N+w_G+w_S=1$). Then, the aggregated degree of satisfaction $DS(d_i)$ that document d_i satisfies the user's query can be calculated as follows:

$$DS(d_i)=w_P\times DS_P(d_i)+w_N\times DS_N(d_i)+w_G\times DS_G(d_i)+w_S\times DS_S(d_i). \tag{12}$$

Case 2: The user ranks the four kinds of fuzzy relationships in a decreasing importance order represented by the following expression:

$$r_1 > r_2 > r_3 > r_4,$$

where $r_j\in\{P,N,G,S\}$, $r_1\neq r_2\neq r_3\neq r_4$, $1\leqslant j\leqslant 4$, and the fuzzy relationship on the left-hand side of "$>$" is more important than the one on the right-hand side of "$>$". Therefore, r_j is the jth important fuzzy relationship among all fuzzy relationships. Then, the fuzzy information retrieval system will assign an importance weight w_{r_j} to each degree of satisfaction $DS_{r_j}(d_i)$ that document d_i satisfies the user's query by fuzzy relationships r_j as follows:

$$w_{r_j}=\frac{5-j}{\sum\limits_{j=1,2,\ldots,4} j}, \tag{13}$$

where $0\leqslant w_{r_j}\leqslant 1$, and $r_j\in\{P,N,G,S\}$. Based on formula (13), we can see that $w_{r_1}=4/10$, $w_{r_2}=3/10$, $w_{r_3}=2/10$ and $w_{r_4}=1/10$. Thus, the degree of satisfaction that document d_i satisfies the user's query by the jth important fuzzy relationship will obtain the jth large importance weight; moreover, the summation of all importance weights will be equal to one (i.e., $w_{r_1}+w_{r_2}+w_{r_3}+w_{r_4}=1$). Then, the aggregated degree of satisfaction $DS(d_i)$ that document d_i satisfies the user's query can be calculated as follows:

$$DS(d_i)=\tfrac{4}{10}DS_{r_1}(d_i)+\tfrac{3}{10}DS_{r_2}(d_i)+\tfrac{2}{10}DS_{r_3}(d_i)+\tfrac{1}{10}DS_{r_4}(d_i), \tag{14}$$

where $r_j\in\{P,N,G,S\}$, $r_1\neq r_2\neq r_3\neq r_4$, and $1\leqslant j\leqslant 4$.

Case 3: The users give a predefined linguistic quantifier which can be formalized as a corresponding weighting vector of the OWA operators. Based on [3], we defined three linguistic quantifiers, i.e., "top one fuzzy relationship", "top t fuzzy relationships" and "top p percent fuzzy relationships" in the fuzzy information retrieval system, where $1\leqslant t\leqslant 4$ and $1\leqslant p\leqslant 100$. If the linguistic quantifier given by the user is "top one fuzzy relationship", then it means that only the fuzzy relationship with the largest relevance degree between the documents and the user's query is considered. Then, the corresponding weighting vector W is

$$W = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

If the linguistic quantifier given by the user is "top t fuzzy relationships", where $1 \leqslant t \leqslant 4$, then it means that only the relationships within the top tth relevance degree between the documents and the user's query are considered. Then, the corresponding weighting vector W is

$$W = \begin{array}{c} w_1 \\ \vdots \\ w_t \\ w_{t+1} \\ \vdots \\ w_4 \end{array} \begin{bmatrix} 1/t \\ \vdots \\ 1/t \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

where (1) $w_j = 1/t$ when $1 \leqslant j \leqslant t$, (2) $w_j = 0$ when $t < j \leqslant 4$, where $1 \leqslant j \leqslant 4$ and $1 \leqslant t \leqslant 4$.

If the linguistic quantifier given by the user is "top p percent fuzzy relationships", then it means that only the relationships within the top p percent of the relevance degrees between the documents and users' queries are considered, where $1 \leqslant p \leqslant 100$. Then, the corresponding weighting vector W is

$$W = \begin{array}{c} w_1 \\ \vdots \\ w_l \\ w_{l+1} \\ \vdots \\ w_4 \end{array} \begin{bmatrix} 1/l \\ \vdots \\ 1/l \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

where $l = \lceil p \times 4/100 \rceil = \lceil p/25 \rceil$, $\lceil x \rceil$ denotes the operator that takes the least integer greater than or equal to the argument x, and (1) $w_j = 1/l$ when $1 \leqslant j \leqslant l$, (2) $w_j = 0$ when $l < j \leqslant 4$, where $1 \leqslant j \leqslant 4$ and $1 \leqslant l \leqslant 4$.

After performing the ordering process according to the input arguments of the OWA operators, we can get the ordered argument vector B shown as follows:

$$B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix},$$

where b_j is the jth largest value of the $DS_r(d_i)$. The aggregated degree of satisfaction $DS(d_i)$ that document d_i satisfies the user's query can be calculated as follows:

$$DS(d_i) = F_w(DS_r(d_i)) = \sum_{j=1}^{4} w_j b_j, \tag{15}$$

where w_j is the jth element of W, b_j is the jth element of B, $1 \leqslant w_j \leqslant 4$, and $1 \leqslant b_j \leqslant 4$.

Then, the documents in the document set D are arranged in a decreasing order according to their aggregated degrees of satisfaction to the user's query. The user may assign an information retrieval threshold value α, where $\alpha \in [0, 1]$, such that the documents with the aggregated degrees of satisfaction less than α will not be retrieved by the fuzzy information retrieval system.

Users familiar with the meanings of the four kinds of fuzzy relationships (i.e., "fuzzy positive association" relationship, "fuzzy negative association" relationship, "fuzzy generalization" relationship and "fuzzy specialization" relationship) can choose the first two ways of the aggregation methods and can express their preferences on fuzzy relationships by giving different weights to the $DS_r(d_i)$'s or by specifying an importance order to these four kinds of fuzzy relationships. Therefore, if a document has larger degrees of satisfaction with respect to the user's query by some user's preferable fuzzy relationships, then it tends to get a larger aggregated degree of satisfaction with respect to the user's query. Thus, the semantic of the aggregated degree of satisfaction of the first two ways of the aggregation methods can be controlled by users. On the other hand, if the users are not familiar with the meanings of the four kinds of fuzzy relationships or have no preferences on the fuzzy relationships, then they can use the last way of the aggregation methods. Therefore, if a document has larger degrees of satisfaction with respect to the user's query by some fuzzy relationships (no matter what kind of fuzzy relationships), then it tends to get a larger aggregated degree of satisfaction with respect to the user's query.

5. An example

In this section, we use an example to illustrate the fuzzy information retrieval process of the proposed method.

Example 5.1. Consider the multi-relationship fuzzy concept network shown in Fig. 1, where there are six concept nodes $c_1, c_2, \ldots,$ and c_6 and three documents nodes d_1, d_2 and d_3 in the multi-relationship fuzzy concept network, where the corresponding concepts of the concept nodes are shown in Table 1 and the corresponding document titles of the document nodes are shown in Table 2. The corresponding concept matrices V_P, V_N, V_G and V_S are shown as follows:

$$V_P = \begin{bmatrix} 1 & 0 & 0.2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.5 & 0.7 & 0 \\ 0.2 & 0 & 1 & 0.3 & 0 & 0.3 \\ 0 & 0.5 & 0.3 & 1 & 0.3 & 0 \\ 0 & 0.7 & 0 & 0.3 & 1 & 0 \\ 0 & 0 & 0.3 & 0 & 0 & 1 \end{bmatrix},$$

$$V_N = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$V_G = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.8 & 0 & 0 & 0.9 & 0 & 0.9 \\ 0 & 0.9 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$V_S = \begin{bmatrix} 0 & 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0 & 0 \end{bmatrix}.$$

Based on Definition 4.2, the transitive closures of these four concept matrices can be obtained shown as follows, where V_P^* is the transitive closure of V_P, V_N^* is the transitive closure of V_N, V_G^* is the transitive closure of V_G, and V_S^* is the transitive closure of V_S (note that $V_N = V_N^*$ due to the fact that fuzzy negative association is not a transitive relationship):

$$V_P^* = \begin{bmatrix} 1 & 0.03 & 0.2 & 0.06 & 0.018 & 0.06 \\ 0.03 & 1 & 0.15 & 0.5 & 0.7 & 0.045 \\ 0.2 & 0.15 & 1 & 0.3 & 0.09 & 0.3 \\ 0.06 & 0.5 & 0.3 & 1 & 0.35 & 0.09 \\ 0.18 & 0.7 & 0.09 & 0.35 & 1 & 0.027 \\ 0.06 & 0.045 & 0.3 & 0.09 & 0.027 & 1 \end{bmatrix},$$

$$V_N^* = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$V_G^* = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.8 & 0.81 & 0 & 0.9 & 0.81 & 0.9 \\ 0 & 0.9 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$V_S^* = \begin{bmatrix} 0 & 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0.81 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & 0.81 & 0.9 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0 & 0 \end{bmatrix}.$$

Furthermore, we can use a document descriptor matrix U to model the degrees of documents containing concepts in the multi-relationship fuzzy concept network shown as follows:

$$U = \begin{bmatrix} 0.1 & 0 & 0 & 0.9 & 0 & 0 \\ 0.7 & 0.3 & 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then, based on Definition 4.4, the four expanded document descriptor matrices U_P^*, U_N^*, U_G^* and U_S^* can be obtained as follows:

$$U_P^* = \begin{bmatrix} 0.1 & 0.45 & 0.27 & 0.9 & 0.27 & 0.081 \\ 0.7 & 0.3 & 0.14 & 0.15 & 0.2 & 0.042 \\ 0.06 & 0.045 & 0.3 & 0.09 & 0.027 & 1 \end{bmatrix},$$

$$U_N^* = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.14 & 0 & 0 & 0.21 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$U_G^* = \begin{bmatrix} 0 & 0.81 & 0 & 0 & 0.81 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$U_S^* = \begin{bmatrix} 0 & 0 & 0.81 & 0 & 0 & 0 \\ 0 & 0 & 0.56 & 0.27 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0 & 0 \end{bmatrix}.$$

Assume that the user's query descriptor vector $\bar{q}$ is

$$\bar{q} = [0.5, 0.8, -, -, -, -],$$

which means the user wants to retrieve documents whose contents are related to "*Security and Encryption*" and "*Internet*" with the degrees 0.5 and 0.8, respectively. Then, based on formula (9), we can calculate the degree of satisfaction that document d_j satisfies the user's query by fuzzy relationship r, where $r \in \{P, N, G, S\}$. The degree of satisfaction that document d_1 satisfies the user's query by fuzzy relationship r, $r \in \{P, N, G, S\}$, can be calculated as follows:

$$DS_P(d_1) = \frac{(1 - |0.1 - 0.5|) + (1 - |0.45 - 0.8|)}{2} = \frac{1.25}{2} = 0.625,$$

$$DS_N(d_1) = \frac{(1-|0-0.5|)+(1-|0-0.8|)}{2} = \frac{0.7}{2} = 0.35,$$

$$DS_G(d_1) = \frac{(1-|0-0.5|)+(1-|0.81-0.8|)}{2} = \frac{1.49}{2} = 0.745,$$

$$DS_S(d_1) = \frac{(1-|0-0.5|)+(1-|0-0.8|)}{2} = \frac{0.7}{2} = 0.35.$$

Similarly, the degrees of satisfaction that documents d_2 and d_3 satisfy the user's query by each fuzzy relationship r, $r \in \{P, N, G, S\}$, are calculated as follows:

$$DS_P(d_2) = 0.65, \quad DS_N(d_2) = 0.65, \quad DS_G(d_2) = 0.42 \quad \text{and} \quad DS_S(d_2) = 0.35,$$

$$DS_P(d_3) = 0.4025, \quad DS_N(d_2) = 0.35, \quad DS_G(d_2) = 0.35 \quad \text{and} \quad DS_S(d_2) = 0.35.$$

The degrees of satisfaction that a document satisfies the user's query by different fuzzy relationships are then aggregated to obtain the aggregated satisfaction that the document satisfies the user's query. The user could choose the aggregation operations by the following three ways:

Case 1: The user assigns the importance weights directly to each degree of satisfaction that the document satisfies the user's query by different fuzzy relationships. Assume that the importance weights specified by the user are as follows:

$$w_P = 0.8,$$

$$w_N = 0.2,$$

$$w_G = 0,$$

$$w_S = 0,$$

where w_P, w_N, w_G and w_S are the user-specified weights assigned to the degrees of satisfaction that the document satisfies the user's query by "fuzzy positive association" relationship, "fuzzy negative association" relationship, "fuzzy generalization" relationship and "fuzzy specialization" relationship, respectively. Based on formula (12), the degree of satisfaction $DS(d_i)$ that each document d_i, where $1 \leqslant i \leqslant 3$, satisfies the user's query can be evaluated as follows:

$$DS(d_1) = 0.8 \times 0.625 + 0.2 \times 0.35 + 0 \times 0.745 + 0 \times 0.35 = 0.57,$$

$$DS(d_2) = 0.8 \times 0.65 + 0.2 \times 0.42 + 0 \times 0.35 + 0 \times 0.35 = 0.604,$$

$$DS(d_3) = 0.8 \times 0.4025 + 0.2 \times 0.35 + 0 \times 0.35 + 0 \times 0.35 = 0.392,$$

where $DS(d_i)$ is the degree of satisfaction of document d_i with respect to the user's query, and $1 \leqslant i \leqslant 3$. The order from the document with the largest degree of satisfaction to that with the smallest degree of satisfaction is $d_2 > d_1 > d_3$.

Case 2: The user ranks the four kinds of fuzzy relationships in a decreasing importance order. Assume that the order specified by the user is

$$G > P > N > S.$$

Based on formula (14), the degree of satisfaction $DS(d_i)$ of each document d_i can be calculated as follows:

$$DS(d_1) = 0.3 \times 0.625 + 0.2 \times 0.35 + 0.4 \times 0.745 + 0.1 \times 0.35 = 0.5905,$$

$$DS(d_2) = 0.3 \times 0.65 + 0.2 \times 0.42 + 0.4 \times 0.35 + 0.1 \times 0.35 = 0.454,$$

$$DS(d_3) = 0.3 \times 0.4025 + 0.2 \times 0.35 + 0.4 \times 0.35 + 0.1 \times 0.35 = 0.36575,$$

where $DS(d_i)$ is the degree of satisfaction of document d_i with respect to the user's query, and $1 \leqslant i \leqslant 3$. The order from the document with the largest degree of satisfaction to that with the smallest degree of satisfaction is $d_1 > d_2 > d_3$.

Case 3: The user gives a predefined linguistic quantifier which can be formalized as a corresponding weight vector of the OWA operators.

(1) Assume that the user uses the linguistic quantifier "top one relationship" which can be represented by the transpose w^{T} of the weighting vector, where $w^{\mathrm{T}} = [1\ 0\ 0\ 0]$. Based on formula (15), after applying the OWA aggregation operations, the degree of satisfaction $DS(d_i)$ of each document d_i, where $1 \leqslant i \leqslant 3$, with respect to the user's query can be evaluated as follows:

$$DS(d_1) = 0 \times 0.625 + 0 \times 0.35 + 1 \times 0.745 + 0 \times 0.35 = 0.745,$$

$$DS(d_2) = 1 \times 0.65 + 0 \times 0.42 + 0 \times 0.35 + 0 \times 0.35 = 0.65,$$

$$DS(d_3) = 1 \times 0.4025 + 0 \times 0.35 + 0 \times 0.35 + 0 \times 0.35 = 0.4025,$$

where $DS(d_i)$ is the degree of satisfaction of document d_i with respect to the user's query, and $1 \leqslant i \leqslant 3$. The order from the document with the largest degree of satisfaction to that with the smallest degree of satisfaction is $d_1 > d_2 > d_3$.

(2) Assume that the user uses the linguistic quantifier "top 2 fuzzy relationships" which can be represented by the transpose w^{T} of the weighting vector, where $w^{\mathrm{T}} = [0.5\ 0.5\ 0\ 0]$. Based on formula (15), after applying the OWA aggregation operations, the degree of satisfaction $DS(d_i)$ of each document d_i, where $1 \leqslant i \leqslant 3$, with respect to the user's query can be evaluated as follows:

$$DS(d_1) = 0.5 \times 0.625 + 0 \times 0.35 + 0.5 \times 0.745 + 0 \times 0.35 = 0.685,$$

$$DS(d_2) = 0.5 \times 0.65 + 0.5 \times 0.42 + 0 \times 0.35 + 0 \times 0.35 = 0.535,$$

$$DS(d_3) = 0.5 \times 0.4025 + 0.5 \times 0.35 + 0 \times 0.35 + 0 \times 0.35 = 0.37625,$$

where $DS(d_i)$ is the degree of satisfaction of document d_i with respect to the user's query, and $1 \leqslant i \leqslant 3$. The order from the document with the largest degree of satisfaction to that with the smallest degree of satisfaction is $d_1 > d_2 > d_3$.

(3) Assume that the user uses the linguistic quantifier "top 75 percent fuzzy relationships" which can be represented by the transpose w^{T} of the weighting vector, where $w^{\mathrm{T}} = [0.33\ 0.33\ 0.33\ 0]$. Based on formula (15), after applying the OWA aggregation operations, the degree of satisfaction $DS(d_i)$ of each document d_i, where $1 \leqslant i \leqslant 3$, with respect to the user's query can be evaluated

as follows:

$$DS(d_1) = 0.33 \times 0.625 + 0.33 \times 0.35 + 0.33 \times 0.745 + 0 \times 0.35 = 0.5676,$$

$$DS(d_2) = 0.33 \times 0.65 + 0.33 \times 0.42 + 0.33 \times 0.35 + 0 \times 0.35 = 0.4686,$$

$$DS(d_3) = 0.33 \times 0.4025 + 0.33 \times 0.35 + 0.33 \times 0.35 + 0 \times 0.35 = 0.3638,$$

where $DS(d_i)$ is the degree of satisfaction of document d_i with respect to the user's query, and $1 \leqslant i \leqslant 3$. The order from the document with the largest degree of satisfaction to that with the smallest degree of satisfaction is $d_1 > d_2 > d_3$.

6. Conclusions

In this paper, we have extended the work we presented in [4] to present a new method for fuzzy information retrieval based on multi-relationship fuzzy concept networks, where the concepts in a multi-relationship fuzzy concept network can be related to other concepts by more than one relationship at the same time, and each relationship has its associated degree between zero and one. There are four fuzzy kinds of relationships in a multi-relationship fuzzy concept network, i.e., "fuzzy positive association" relationship, "fuzzy negative association" relationship, "fuzzy generalization" relationship and "fuzzy specialization" relationship. We also have presented an information retrieval method to deal with the users' fuzzy queries based on the multi-relationship fuzzy concept networks. Because the concepts in a multi-relationship fuzzy concept network can be related to each other by more than one relationship at the same time, the proposed fuzzy information retrieval method based on the multi-relationship fuzzy concept network model is more flexible and more intelligent than the one we presented in [4]. In this paper, we assume that the relationships and their associated degrees between concepts in a multi-relationship fuzzy concept network are specified by domain experts. In the future, we will provide an automatic multi-relationship fuzzy concept network construction mechanism for fuzzy information retrieval. Moreover, to develop a more efficient method for calculating the degree of satisfaction that a document satisfies the user's fuzzy query is worth of future research.

Acknowledgements

The authors would like to thank the referees for providing very helpful comments and suggestions. Their insight and comments led to a better presentation of the ideas expressed in this paper. This work was supported in part by the National Science Council, Republic of China, under Grant NSC-91-2213-E-011-052.

References

[1] J.C. Bezdek, G. Biswas, L.Y. Huang, Transitive closures of fuzzy thesauri for information-retrieval systems, Internat. J. Man–Mach. Stud. 25 (1986) 343–356.

[2] G. Bordogna, P. Carrara, G. Pasi, A fuzzy document representation supporting user adaptation in information retrieval, Proc. 2nd IEEE Internat. Conf. on Fuzzy Systems, Vol. 2, San Francisco, USA, 1993, pp. 974–979.
[3] G. Bordogna, G. Pasi, Controlling retrieval through a user-adaptive representation of documents, Internat. J. Approx. Reason. 12 (3) (1995) 317–339.
[4] S.M. Chen, Y.J. Horng, Fuzzy query processing for document retrieval based on extended fuzzy concept networks, IEEE Trans. Systems Man Cybernet.—Part B: Cybernet. 29 (1) (1999) 126–135.
[5] S.M. Chen, Y.J. Horng, C.H. Lee, Document retrieval using fuzzy valued concept networks, IEEE Trans. Systems Man and Cybernet.—Part B: Cybernet. 31 (1) (2001) 111–118.
[6] S.M. Chen, W.H. Hsiao, Y.J. Horng, A knowledge-based method for fuzzy query processing for document retrieval, Cybernet. Systems Internat. J. 28 (1) (1997) 99–119.
[7] S.M. Chen, J.Y. Wang, Document retrieval using knowledge-based fuzzy information retrieval techniques, IEEE Trans. Systems Man Cybernet. 25 (5) (1995) 793–803.
[8] E. Damiani, M.G. Fugini, Automatic thesaurus construction supporting fuzzy retrieval of reusable components, Proc. 1995 ACM Symp. on Applied Computing, Nashville, USA, 1995, pp. 542–547.
[9] Y.J. Horng, S.M. Chen, C. H. Lee, A fuzzy information retrieval method based on multi-relationship fuzzy concept networks, Proc. 2000 Internat. Computer Symp.: Workshop on Artificial Intelligence, Chiayi, Taiwan, Republic of China, 2000, pp. 79–86.
[10] Y.J. Horng, S.M. Chen, C.H. Lee, Automatically constructing multi-relationship fuzzy concept networks in fuzzy information retrieval systems, Proc. 10th IEEE Internat. Conf. on Fuzzy Systems, Vol. 2, Melbourne, Australia, 2001.
[11] Y. Jing, W.B. Croft, An association thesaurus for information retrieval, Proc. 1994 RIAO Internat. Conf., New York, USA, 1994, pp. 146–160.
[12] A. Kandel, Fuzzy Mathematical Techniques with Applications, Addison-Wesley, CA, 1986.
[13] K.J. Kim, S.B. Cho, A personalized web search engine using fuzzy concept network with link structure, Proc. Joint 9th IFSA World Cong. and 20th NAFIPS Internat. Conf., Vol. 1, Vancouver, Canada, 2001, pp. 81–86.
[14] B.M. Kim, J.Y. Kim, J. Kim, Query term expansion and reweighting using term co-occurrence similarity and fuzzy inference, Proc. Joint 9th IFSA World Cong. and 20th NAFIPS Internat. Conf., Vol. 2, Vancouver, Canada, 2001, pp. 715–720.
[15] L.T. Kóczy, T.D. Gedeon, J.A. Kóczy, Fuzzy tolerance relations and relational maps applied to information retrieval, Fuzzy Sets and Systems 126 (1) (2002) 49–61.
[16] M. Kracker, A fuzzy concept network model and its applications, Proc. First IEEE Internat. Conf. on Fuzzy Systems, San Diego, USA, 1992, pp. 761–768.
[17] D.H. Kraft, J. Chen, A. Mikulcic, Combining fuzzy clustering and fuzzy inferencing in information retrieval, Proc. 9th IEEE Internat. Conf. on Fuzzy Systems, Vol. 1, San Antonio, TX, USA, 2000, pp. 375–380.
[18] H.L. Larsen, R.R. Yager, The use of fuzzy relational thesauri for classificatory problem solving in information retrieval and expert systems, IEEE Trans. Systems Man Cybernet. 23 (1) (1993) 31–41.
[19] D. Lucarella, Uncertainty in information retrieval: an approach based on fuzzy sets, Proc. 9th Annu. Internat. Phoenix Conf. on Computer and Communications, Arizona, USA, 1990, pp. 809–814.
[20] D. Lucarella, R. Morara, FIRST: fuzzy information retrieval system, J. Inform. Sci. 17 (1) (1991) 81–91.
[21] S. Miyamoto, Two approaches for information retrieval through fuzzy associations, IEEE Trans. Systems Man Cybernet. 19 (1) (1989) 123–130.
[22] S. Miyamoto, Fuzzy Sets in Information Retrieval and Cluster Analysis, Kluwer Academic Publishers, Netherlands, 1990.
[23] R. Quillian, Semantic memory, in: M. Minsky (Ed.), Semantic Information Processing, MIT Press, Cambridge, MA, 1968.
[24] T. Radecki, Generalized Boolean methods of information retrieval, Internat. J. Man–Machine Stud. 18 (5) (1983) 409–439.
[25] G. Salton, M.J. Mcgill, Introduction to Modern Information Retrieval, McGraw-Hill, New York, 1983.
[26] V. Tahani, A fuzzy model of document retrieval system, Inform. Process. Management 12 (3) (1976) 177–187.
[27] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decisionmaking, IEEE Trans. Systems Man Cybernet. 18 (1) (1988) 183–190.
[28] L.A. Zadeh, Fuzzy sets, Inform. and Control 8 (1965) 338–353.
[29] H.J. Zimmermann, Fuzzy Set Theory and its Applications, Kluwer Academic Publishers, Boston, 1990.