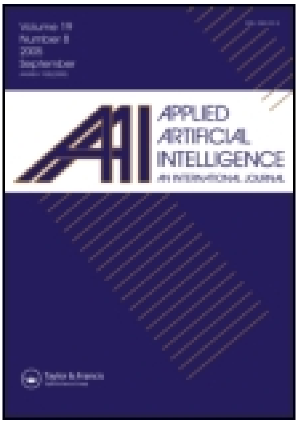


Applied Artificial Intelligence: An International Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uaai20

Automatically constructing multi-relationship fuzzy concept networks for document retrieval

Yih-JEN Horng [a], Shyi-Ming Chen [b] & Chia-Hoang Lee [a]

[a] Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan, R. O. C.

[b] Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, R. O. C.

Published online: 30 Nov 2010.

Applied Artificial Intelligence, 17:303–328, 2003

0883-9514/03 $12.00 +.00
DOI: 10.1080/08839810390198664

AUTOMATICALLY CONSTRUCTING MULTI-RELATIONSHIP FUZZY CONCEPT NETWORKS FOR DOCUMENT RETRIEVAL

YIH-JEN HORNG
Department of Computer and Information Science,
National Chiao Tung University, Hsinchu, Taiwan, R. O. C.

SHYI-MING CHEN
Department of Computer Science and Information Engineering,
National Taiwan University of Science and Technology,
Taipei, Taiwan, R. O. C.

CHIA-HOANG LEE
Department of Computer and Information Science,
National Chiao Tung University, Hsinchu, Taiwan, R. O. C.

Although the knowledge bases incorporated in existing information retrieval systems can enhance retrieval effectiveness, many of them are built by domain experts. It is obvious that the construction of such knowledge bases requires a large amount of human effort. In this paper, an intelligent fuzzy information retrieval system with an automatically constructed knowledge base is presented; the knowledge base is represented by a multi-relationship fuzzy concept network. The multi-relationship fuzzy concept network can describe four kinds of context-independent and context-dependent fuzzy relationships, i.e., ''fuzzy positive association'' relationship, ''fuzzy negative association'' relationship, ''fuzzy generalization'' relationship, and ''fuzzy specialization'' relationship between concepts. The users of the fuzzy information retrieval system can submit a fuzzy contextual query which specifies the search context in the query formula. The fuzzy information retrieval system retrieves documents whose contents are relevant to the user's query by some kinds of fuzzy relationships for the specified search context of the user's query. The proposed fuzzy information retrieval method is more intelligent and more flexible than the existing methods due to the fact that it can construct multi-relationship fuzzy concept networks automatically and it can provide contextual search capability to allow the users to specify fuzzy contextual queries in a more intelligent and flexible manner.

This work was supported in part by the National Science Council, Republic of China, under Grant NSC 90-2213-E-011-054.

Address correspondence to Professor Shyi-Ming Chen, Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, R.O.C.

INTRODUCTION

Most of the existing information retrieval systems are based on the Boolean logic model and assume that documents and user's queries can be represented precisely by index terms (Salton and Mcgill 1983). Moreover, documents are retrieved only when they contain the index terms specified in the user's queries. However, this approach will neglect other relevant documents that do not contain the index terms specified in the user's queries. This problem could be overcome by incorporating a knowledge base which depicts the relationships between index terms into the existing information retrieval systems. Many intelligent information retrieval systems have been proposed to retrieve documents intelligently by incorporating knowledge bases into the systems (Bezdek et al. 1986; Bhatia and Deogun 1998; Chang and Chen 1998; Chen and Lu 1997; Chen and Horng 1999; Chen et al. 1997; Chen and Wang 1995; Horng et al. 2000; Horng et al. 2001; Kracker 1992; Liang and Chang 1999; Lin et al. 1999; Lucarella and Morara 1991). Lucarella and Morara (1991) presented a knowledge-based fuzzy information retrieval system based on the fuzzy set theory (Zadeh 1965), where the knowledge base is represented by a concept network. The concept networks can depict the relationships between concepts which are defined as meaningful entities (e.g., index terms [Kim and Cho 2001] or classes of documents [Liang and Chang 1999]) in a specific domain. In Chen and Wang (1995) in order to increase the speed of inferences through concept networks, we have presented a method for document retrieval using knowledge-based fuzzy information retrieval techniques, where concept matrices are used to model fuzzy concept networks. By calculating the transitive closure of a concept matrix, the implicit degrees of relationships between concepts can be obtained. The fuzzy information retrieval systems can deal with the user's query more efficiently. However, the information retrieval methods presented in Chen and Wang (1995) and Lucarella and Morara (1991) all assumed that the link strengths between any two concept nodes or between a concept node and a document node in a concept network are specified by experts. This assumption may be impractical when the application domain contains a large amount of concepts and documents. In this case, the construction of the corresponding knowledge base will require a large amount of human effort.

To overcome the drawback of the information retrieval methods presented in Chen and Wang (1995) and Lucarella and Morara (1991), some methods have been proposed to construct concept networks automatically (Chang and Chen 1998; Liang and Chang 1999). In Chang and Chen (1998), a method for supporting conceptual and neighborhood queries on the World Wide Web (WWW) is presented, where the links between homepages and the HyperText Markup Language (HTML) structures of homepages are used to derive relationships between concepts. In Liang and Chang (1999), a three-level concept network architecture is presented, where the nodes of the bottom level stand for documents and the nodes of the other two levels stand for concepts. The concepts of the top level are formed by clustering similar concepts of the middle level. The relationship between two concepts is obtained by comparing the keywords involved in these two concepts. The more that common keywords appear in these two concepts, the higher the degree of relationship between these two concepts.

However, the information retrieval methods presented in (Chang and Chen 1998; Chen and Wang 1995; Liang and Chang 1999; Lucarella and Morara 1991) all assume that each pair of concepts in a concept network only can be related to each other by one kind of relationship. That is, these methods assume that the relationship between any two concepts is invariable in different cases. However, two concepts may have different relationships from various perspectives. For example, *"Internet"* and *"Intranet"* are two concepts concerning computer network architectures. They are antonyms since they represent different scales of computer networks. On the other hand, they are regarded as synonymous concepts due to the fact that they both introduce techniques to connect computers together. Furthermore. *"Intranet"* is involved in *"Internet"* since the former can be deemed as a portion of the latter.

In this paper, we present a method to automatically construct multi-relationship fuzzy concept networks based on training documents for fuzzy information retrieval. In a multi-relationship fuzzy concept network, there are four kinds of fuzzy relationships (Kracker 1992) (i.e., "fuzzy positive association" relationship, "fuzzy negative association" relationship, "fuzzy generalization" relationship and "fuzzy specialization" relationship) to describe possible semantic relationships between concepts. These fuzzy relationships can exist simultaneously between any two concepts and each has its own degree of strength. Users of the fuzzy information retrieval systems based on multi-relationship fuzzy concept networks can submit a query in which a search context is involved to provide the user's perspective on the fuzzy relationships between concepts. Documents are retrieved if they contain concepts that have fuzzy relationships with the concepts specified in the user's query. The proposed information retrieval method is more intelligent and more flexible than the existing methods due to the fact that it can automatically construct multi-relationship fuzzy concept networks and it can provide contextual search capability to allow the users to specify fuzzy contextual queries in a more intelligent and more flexible manner.

CONCEPT NETWORKS AND MULTI-RELATIONSHIP FUZZY CONCEPT NETWORKS

In the following, we briefly review the definitions of concept networks (Lucarella and Morara 1991) and multi-relationship fuzzy concept networks

(Horng et al. 2000). In Lucarella and Morara (1991), a knowledge-based fuzzy information retrieval method was proposed, where the knowledge base is represented by a concept network which consists of nodes and directed links. In a concept network, each node represents a concept or a document. A link connecting two concept nodes means that these two concepts are semantically related. A link directed from a concept node to a document node means that the content of this document contains this concept. Each directed link is associated with a degree μ, where $\mu \in [0, 1]$, indicating the degree of strength of the relationship between two concepts or the degree of strength that a document contains a concept. For example, Figure 1 shows a concept network where $c_1, c_2, \ldots,$ and c_5 are concepts; d_1 and d_2 are documents. For Figure 1, we can see that document d_1 contains concepts c_1 and concept c_2 with the degrees of strength 0.9 and 0.3, respectively; concept c_3 is related to concept c_1 and concept c_4 with the degrees of strength 0.8 and 0.5, respectively.

In a multi-relationship fuzzy concept network, there are four kinds of possible fuzzy relationships (Kracker 1992) between concepts and each has its own degree of strength. The semantics of the four kinds of possible fuzzy relationships between concepts are shown as follows:

1. Fuzzy positive association. It relates concepts which are similar to each other.
2. Fuzzy negative association. It relates concepts which have complementary, incompatible, or antonymous meanings in some contexts.

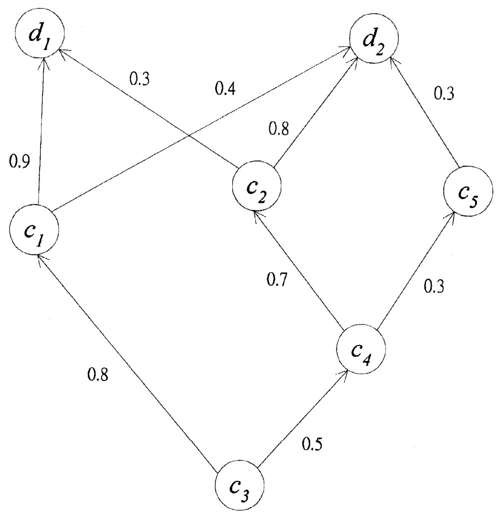


FIGURE 1. A concept network.

3. Fuzzy generalization. It relates a concept to another concept if the former is a part of the latter or the former is a kind of the latter.
4. Fuzzy specialization. It is the inverse of the fuzzy generalization relationship.

The fuzzy relationships between concepts are defined formally as follows (Kracker 1992):

Definition 1. Let C be a set of concepts in a multi-relationship fuzzy concept network. Then,

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

In the following, we briefly review the definitions of multi-relationship fuzzy concept networks we presented in (Horng et al. 2000).

Definition 2. 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 and is associated with a four-tuple $(\langle\mu_P,P\rangle,\langle\mu_N,N\rangle,\langle\mu_G,G\rangle,\langle\mu_S,S\rangle)$, where μ_P indicates that concept c_i and concept c_j are related by the "fuzzy positive association" relationship with a degree μ_P; μ_N indicates that concept c_i and concept c_j are related by the "fuzzy negative association" relationship with a degree μ_N; μ_G indicates that concept c_i is related to a more general concept c_j with a degree μ_G; μ_S indicates that concept c_i is related to a more specific concept c_j with a degree μ_S, where $\mu_P \in [0,1], \mu_N \in [0,1], \mu_G \in [0,1]$ 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 the fuzzy relationship r, where $\mu_r \in [0,1]$ and $r \in \{P, N, G, S\}$.
2. $c_i \xrightarrow{\mu} d_j$ means that document d_j contains concept c_i with a degree of strength μ, where $\mu \in [0,1]$.

For example, Figure 2 shows a multi-relationship fuzzy concept network, where $c_1, c_2, \ldots, c_5$ are concepts, and d_1, d_2, and d_3 are documents. From

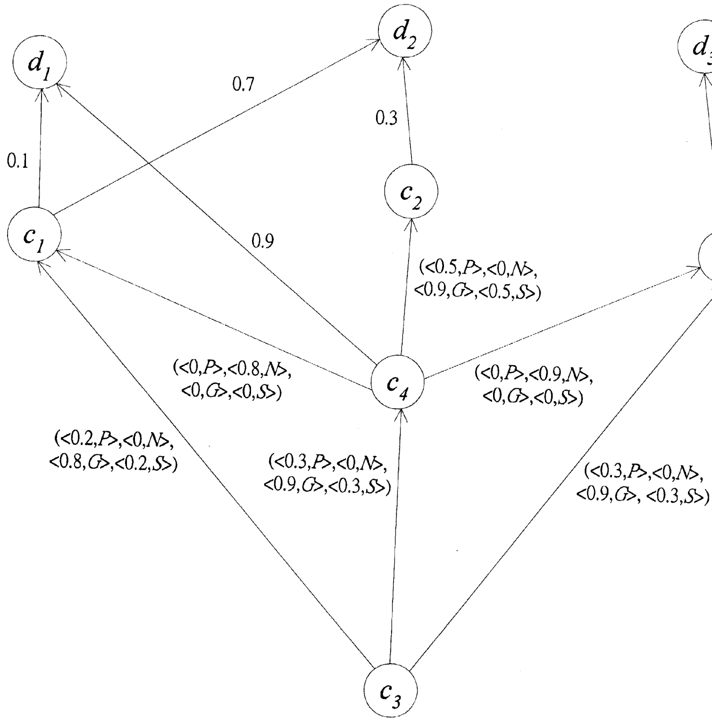


FIGURE 2. A multi-relationship fuzzy concept network.

Figure 2, we can see that document d_2 contains concept c_1 and concept c_2 with degrees of strength 0.7 and 0.3, respectively; concept c_3 is related to concept c_1 by the "fuzzy positive association" relationship with the degree of strength 0.2, by the "fuzzy generalization" relationship with the degree of strength 0.8, and by the "fuzzy specialization" relationship with the degree of strength 0.2.

AUTOMATICALLY CONSTRUCTING MULTI-RELATIONSHIP FUZZY CONCEPT NETWORKS

In the following, we present a method to automatically construct multi-relationship fuzzy concept networks. Liang and Chang (1999), defined the concepts in concept networks as classes of documents. In Horng et al. (2000), we also defined the concepts in multi-relationship fuzzy concept networks as classes of documents. The multi-relationship fuzzy concept networks can be used as a knowledge base for Web-page retrieval on the World Wide Web (WWW). In order to further analyze the contents of Web documents, each document is grabbed by a Web spider from the Web and stored in a document database. Every document in the document database is related to one or more concepts in the concept set C. The contents of the documents in the document database are then broken into words. If document d_i is related to concept c_j, then the words contained in document d_i are also contained in concept c_j. Therefore, each concept contains a set of words derived from a set of documents containing the words. By comparing the set of words contained in each concept, we can obtain the fuzzy relationships between concepts. However, since there is usually a large amount of words contained in the document set, we employ a "word extractor" to discard insignificant words to reduce the word space. First, the word extractor deletes the "HyperText Markup Language" (HTML) tags since they are responsible for the format but not the contents of the Web documents. Second, the words appear in a *stoplist* (i.e., the words that appear with high frequency in all documents) (Salton 1971) are also eliminated by the word extractor. Finally, the word extractor stems each remaining word to its root form (Frakes 1992). The collection of these root-formatted words forms a word set W for the document set. The flow chart for automatically constructing a multi-relationship fuzzy concept network is shown in Figure 3.

The formula for calculating the weight of a word in a document is based on the normalized $TF \times IDF$ (i.e., Term Frequency multiplied by Inverse Document Frequency) weighting method (Salton and Buckley 1988; Salton

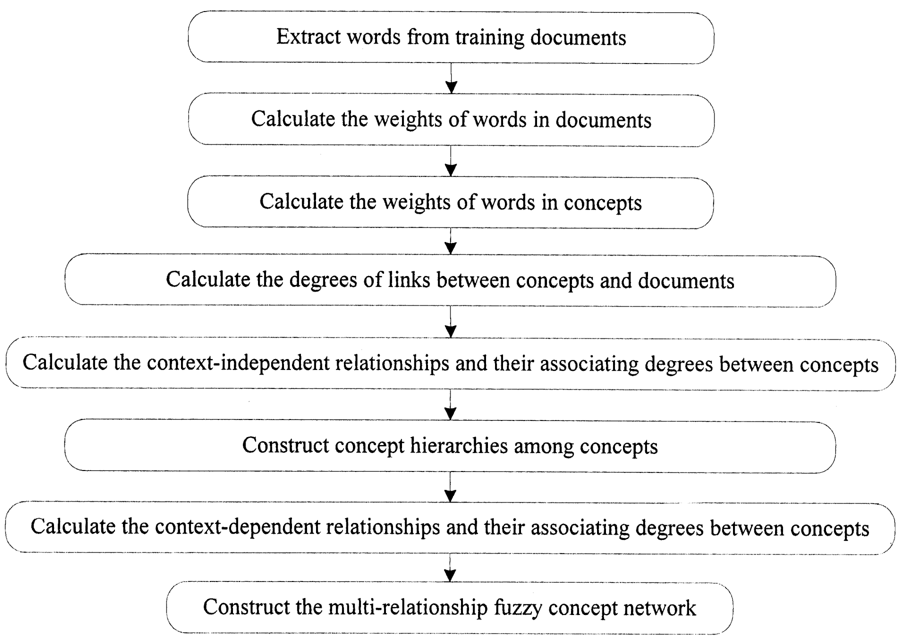


FIGURE 3. The flow chart for automatically constructing a multi-relationship fuzzy concept network.

and Mcgill 1983). That is, if word t appears more frequently than other words contained in document d_i, then the weight of word t should be larger than the weights of other words. Thus, the weight of word t in document d_i is proportional to its occurrence frequency in document d_i. However, if word t also appears frequently in other documents in the corpus, then the importance of word t to the document d_i is reduced. That is, the weight of word t in document d_i is in inverse proportional to its occurrence frequency in all documents in the corpus. The weight "$w\_word\_document(t, d_i)$" of word t in document d_i is calculated as follows:

$$w\_word\_document(t, d_i) = \frac{\left(0.5 + 0.5 \frac{tf_{it}}{\underset{k=1,2,\ldots,L}{\mathrm{Max}} tf_{ik}}\right) \log \frac{N}{df_t}}{\underset{j=1,2,\ldots,L}{\mathrm{Max}} \left\{ \left(0.5 + 0.5 \frac{tf_{it}}{\underset{k=1,2,\ldots,L}{\mathrm{Max}} tf_{ik}}\right) \log \frac{N}{df_j} \right\}}, \quad (1)$$

where tf_{it} is the frequency of word t appearing in document d_i, df_t is the number of documents containing word t, L is the number of words contained in document d_i, and N is the number of documents in the corpus. The larger the value of $w\_word\_document(t, d_i)$, the more important the word t to document d_i. From formula (1), we can see that the value of $w\_word\_document(t, d_i)$ is normalized and its value is between zero and one.

After the weights of words in documents are obtained, the weight "$w\_word\_concept(t, c)$" of word t in concept c can be calculated as follows:

$$w\_word\_concept(t, c) = \frac{\sum_{i=1}^{m} w\_word\_concept(t, d_i)}{m}, \quad (2)$$

where m is the number of documents which contain word t and belong to concept c. From formula (2), we can see that the larger the weight of word t is contained in documents which contains concept c, the larger the weight of word t is in concept c. If $w\_word\_concept(t, c) > 0$, then we say that word t is contained in concept c.

If most of the words contained in document d_j having high weights in concept c, then the weight of document d_i to concept c should be high. The weight "$w\_document\_concept(d_i, c)$" of document d_i to concept c is calculated as follows:

$$w\_document\_concept(d_i, c) = \frac{\sum_{j=1}^{k} w\_word\_concept(t_j, c)}{k}, \quad (3)$$

where k is the number of words contained in document d_i.

Since each concept contains particular words in the word set W, we can use a mapping function M to represent each concept by showing its corresponding fuzzy subset in the word set W. The mapping function M is shown as follows:

$$M(c_i) = w_{i1}/t_1 + w_{i2}/t_2 + \cdots + w_{ih}/t_h, \tag{4}$$

where $M\colon C \to [0,1]^W$, w_{ij} is the weight of word t_j in concept c_i, and h is the number of words in the word set W. If $w_{ij} > 0$, then word t_j is contained in concept c_i. Let $|M(c_i)| = \sum_{j=1,2,\ldots,h} w_{ij}$, then $|M(c_i)|$ is called the *cardinality* of the corresponding fuzzy subset $M(c_i)$ of concept c_i in the word set W.

In the following, we present a method for deciding fuzzy relationships between concepts. Let concept c_i and concept c_j be any two arbitrary concepts in the concept set C. Let's consider the following conditions:

Case 1. If concepts c_i and concept c_j contain different words, then they are not related.

Case 2. If concept c_i and concept c_j contain almost the same words, but the weights of the words in concept c_i are larger than those in concept c_j, then concept c_i is said to dominate concept c_j and should be more general than concept c_j.

Case 3. If concept c_i and concept c_j contain almost the same words, but the weights of the words in concept c_i are smaller than those in concept c_j, then concept c_i is said to be dominated by concept c_j and should be more specific than concept c_j.

Case 4. If most words contained in concept c_j are also contained in concept c_i, but many words contained in concept c_i are not contained in concept c_j, then concept c_i concerns more aspects than concept c_j and should be more general than concept c_j.

Case 5. If most words contained in concept c_i are also contained in concept c_j, but many words contained in concept c_j are not contained in concept c_i, then concept c_i concerns fewer aspects than concept c_j and should be more specific than concept c_j.

Case 6. If concept c_i and concept c_j contain almost the same words, and the weights of the words are similar in both concepts, then these two concepts should be similar to each other and have a fuzzy positive association relationship.

Therefore, we can decide the fuzzy relationships and the associated degrees between concepts by comparing their corresponding fuzzy subsets in the word set W. The degree that concept c_j is more general than concept c_i is denoted as $G(c_i, c_j)$ and is equal to the degree of subsethood of $M(c_i)$ in $M(c_j)$ (i.e., the degree that $M(c_i)$ is contained in $M(c_j)$). A method to calculate

$G(c_i, c_j)$ is by using a commonly used fuzzy subsethood measure (Kosko 1992; Young 1996) shown as follows:

$$G(c_i, c_j) = \begin{cases} \frac{|M(c_i) \cap M(c_j)|}{|M(c_i)|} = \frac{\sum_{k=1}^{h} \min(w_{ki}, w_{kj})}{\sum_{k=1}^{h} w_{ki}}, & \text{if } M(c_i) \neq \phi \\ 1, & \text{if } M(c_j) = \phi \end{cases} \tag{5}$$

where w_{ki} is the weight of word t_k in concept c_i, w_{kj} is the weight of word t_k in concept c_j, and h is the number of words in the word set W. However, formula (5) cannot effectively reveal the generality of concept c_j over concept c_i if concept c_i and concept c_j meet the condition "Case 4" described previously. Let us consider the following example.

Example 1. Assume that there are seven words $t_1, t_2, \ldots,$ and t_7 in the word set W, and assume that the corresponding fuzzy subsets $M(c_i)$ and $M(c_j)$ of concept c_i and concept c_j in the word set W are shown as follows:

$$M(c_i) = 0.3/t_2 + 0.3/t_3 + 0.4/t_4 + 0.4/t_5 + 0.3/t_6,$$
$$M(c_j) = 0.8/t_3 + 0.9/t_4,$$

where the membership functions of $M(c_i)$ and $M(c_j)$ are shown in Figure 4.

From Figure 4, we can see that concept c_i contains all the words contained in concept c_j (i.e., the words t_3 and t_4), but the words t_2, t_5, and t_6 contained in concept c_i are not contained in concept c_j. Therefore, concept c_i should be more general than concept c_j. Based on formula (5), $G(c_i, c_j)$ and $G(c_j, c_i)$ can be calculated as follows:

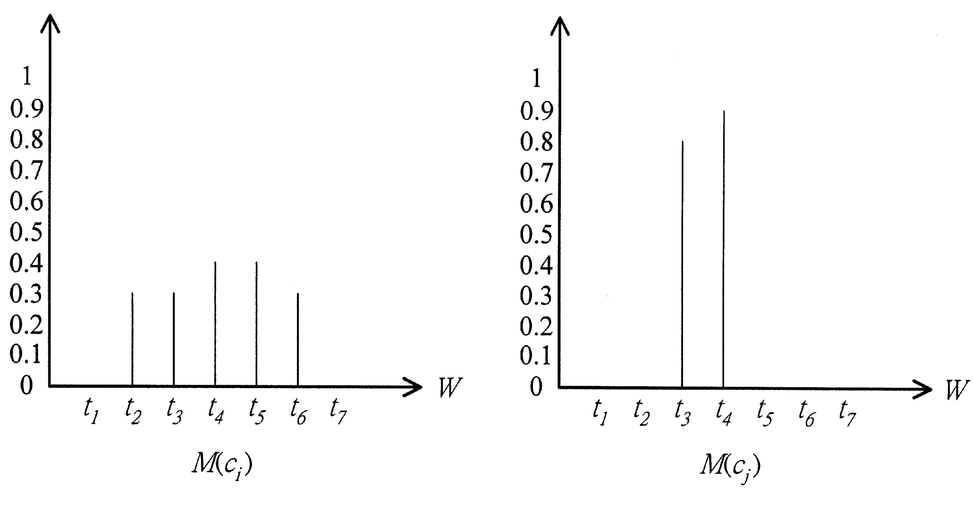


FIGURE 4. The corresponding fuzzy subsets of concept c_i and concept c_j in the word set W.

$$G(c_i, c_j) = \frac{0.3 + 0.4}{0.3 + 0.3 + 0.4 + 0.4 + 0.3} = 0.41,$$

$$G(c_j, c_i) = \frac{0.3 + 0.4}{0.8 + 0.9} = 0.41.$$

That is, $G(c_i, c_j)$ and $G(c_j, c_i)$ have the same value. Therefore, if based on formula (5), we cannot know which concept is more general than the other one.

In this paper, in order to overcome the drawback of formula (5), we calculate $G(c_i, c_j)$ by using the following formula:

$$G(c_i, c_j) = \begin{cases} \left(\frac{|M(c_i) \cap M(c_j)|}{|M(c_i)|}\right)^{\frac{WC(c_j)}{\max(WC(c_i), WC(c_j))}} \\ \quad = \left(\frac{\sum_{k=1}^{h} \min(w_{ki}, w_{kj})}{\sum_{k=1}^{h} w_{ki}}\right)^{\frac{WC(c_j)}{\max(WC(c_i), WC(c_j))}}, & \text{if } M(c_j) \neq \phi \\ 1, & \text{if } M(c_j) = \phi \end{cases} \tag{6}$$

where w_{ki} is the weight of word t_k in concept c_i, w_{kj} is the weight of word t_k in concept c_j, $WC(c_i)$ is the number of words contained in concept c_i, $WC(c_j)$ is the number of words contained in concept c_j, and h is the number of words in the word set W. The advantage of formula (6) is that it takes the number of words contained in each concept into account. Therefore, if concept c_j contains more words than concept c_i does, then the value of $G(c_i, c_j)$ will be increased.

Example 2. Let us consider the concepts c_i and c_j shown in Example 1. According to formula (6), $G(c_i, c_j)$ and $G(c_j, c_i)$ can be calculated as follows:

$$G(c_i, c_j) = \left(\frac{0.3 + 0.4}{0.3 + 0.3 + 0.4 + 0.4 + 0.3}\right)^{\frac{5}{5}} = 0.41,$$

$$G(c_j, c_i) = \left(\frac{0.3 + 0.4}{0.8 + 0.9}\right)^{\frac{2}{5}} = 0.7.$$

Since $G(c_j, c_i)$ is larger than $G(c_i, c_j)$, we can easily see that concept c_i is more general than concept c_j.

The degree of concept c_j contained in concept c_i (i.e., concept c_j is more specific than concept c_i) is denoted as $S(c_i, c_j)$. Since the fuzzy specialization relationship is the inverse of the fuzzy generalization relationship, we let

$$S(c_i, c_j) = G(c_j, c_i). \tag{7}$$

Moreover, the degree of fuzzy positive association relationship between concept c_i and concept c_j, denoted as $P(c_i, c_j)$, is calculated as follows:

$$P(c_i, c_j) = \min(G(c_i, c_j), S(c_i, c_j)). \tag{8}$$

From formula (8), we can see that if both the values of $G(c_i, c_j)$ and $S(c_i, c_j)$ are large, then the value of $P(c_i, c_j)$ will also be large. This means that if concept c_i and concept c_j contain almost the same words with similar weights, then the degree of fuzzy positive association relationship between concept c_i and concept c_j is large.

From the previous discussions, we can obtain the degrees of fuzzy generalization relationships, fuzzy specialization relationships, and fuzzy positive association relationships between concepts in a multi-relationship fuzzy concept network based on formula (6), formula (7), and formula (8), respectively. However, the degrees of fuzzy negative association relationships between concepts cannot be directly derived by comparing their corresponding fuzzy subsets in the word set W. This is because the fuzzy negative association relationships is a kind of context-dependent fuzzy relationship, i.e., the fuzzy negative association relationship between concepts must be discussed in a specific context. On the other hand, the other three fuzzy relationships (i.e., fuzzy generalization relationships, fuzzy specialization relationships, and fuzzy positive association relationships) are context-independent fuzzy relationships.

After the context-independent fuzzy relationships between concepts have been obtained, one approach to discuss the context-dependent fuzzy relationships between concepts is to construct concept hierarchies among the concepts, where general concepts are placed at higher levels in the concept hierarchies. Any node in a concept hierarchy can be selected as a context concept. Assume that concept c_i and concept c_j are both descendants of the context concept. If concept c_i and concept c_j are in different branches of the context concept and the degree of the fuzzy positive association relationship $P(c_i, c_j)$ between concept c_i and concept c_j is low (i.e., concept c_i and concept c_j are not very similar to each other), then concept c_i and concept c_j should represent different parts of the context concept. Therefore, concept c_i and concept c_j have a "fuzzy negative association" relationship.

In the following, we present an algorithm to construct concept hierarchies based on the "fuzzy positive association" relationship, the "fuzzy generalization" relationship, and the "fuzzy specialization" relationship between concepts. Let α be a threshold value, where $\alpha \in [0, 1]$. Assume that

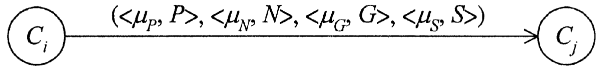
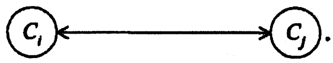


FIGURE 5. Fuzzy relationships between concepts c_i and c_j.

the fuzzy relationships between concepts c_i and c_j are shown in Figure 5, where $\mu_P \in [0,1], \mu_N \in [0,1], \mu_G \in [0,1]$, and $\mu_S \in [0,1]$. Then

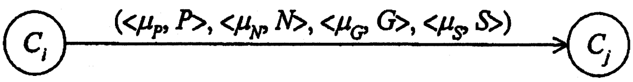
Case 1. If $\mu_G \geq \alpha$ and $\mu_S \geq \alpha$, where $\alpha \in [0,1]$, then concepts c_i and c_j are synonymous concepts, and they should be put in the same concept class, denoted as



Case 2. If $\mu_G \geq \alpha$ and $\mu_S < \alpha$, where $\alpha \in [0,1]$, then concept c_j is more general than concept c_i, denoted as

Case 3. If $\mu_G < \alpha$ and $\mu_S \geq \alpha$, where $\alpha \in [0,1]$, then concept c_i is more general than concept c_j, denoted as

Case 4. If $\mu_G < \alpha$ and $\mu_S < \alpha$, where $\alpha \in [0,1]$, then concepts c_i and c_j *are regarded as not having a generalization relationship under the threshold value* α, *where* $\alpha \in [0,1]$, *and they should not be put in the same concept class.*



Let $G(c_i, c_j)$ be the degree of fuzzy generalization relationship from concept c_i to concept c_j, and let $S(c_i, c_j)$ be the degree of fuzzy specialization relationship from concept c_i to concept c_j, where $G(c_i, c_j) \in [0,1]$ and $S(c_i, c_j) \in [0,1]$. The concept hierarchy construction algorithm is now presented as follows:

Concept Hierarchy Construction Algorithm

If both $G(c_i, c_j)$ and $S(c_i, c_j)$ are larger than the threshold value α, where $\alpha \in [0,1]$,
  then concept c_i and concept c_j are synonymous concepts and they should be put in the same concept class;
if $G(c_i, c_j)$ is larger than the threshold value α and $S(c_i, c_j)$ is smaller than the threshold value α, where $\alpha \in [0,1]$,
  then concept c_j is more general than concept c_i and should be a parent of concept c_i in the concept hierarchy;

if $G(c_i, c_j)$ is smaller than the threshold value α and $S(c_i, c_j)$ is larger than the threshold value α, where $\alpha \in [0, 1]$,
then concept c_i is more general than concept c_j and should be a parent of concept c_j in the concept hierarchy;
if both $G(c_i, c_j)$ and $S(c_i, c_j)$ are smaller than the threshold value α, where $\alpha \in [0, 1]$,
then do nothing.

It should be noted that if both $G(c_i, c_j)$ and $S(c_i, c_j)$ are larger than the threshold value α, where $\alpha \in [0, 1]$, then $P(c_i, c_j)$ is also larger than the threshold value α due to the fact that $P(c_i, c_j)$ is defined as the minimum of $G(c_i, c_j)$ and $S(c_i, c_j)$. Therefore, concept c_i and concept c_j are very similar to each other and should be marked as in the same concept class. A concept class is defined as a collection of synonymous concepts. The proposed concept hierarchy construction algorithm may result in more than one concept hierarchy, where each concept hierarchy has a different root concept. However, except for root concepts, each concept may belong to more than one concept hierarchy at the same time. Moreover, although the concept hierarchies are constructed automatically, they can be further manually revised by human experts to make them more appropriate.

Definition 3. Assume that there is a path from concept c_i to concept c_j in a concept hierarchy shown as follows:

$$c_i \rightarrow c_1 \rightarrow c_2 \rightarrow \cdots \rightarrow c_k \rightarrow c_j,$$

where the symbol "$\rightarrow$" represents a directed link from a child to its parent. Then, concept c_j is an *ancestor* of concept c_i. The distance "*distance_ij*" between concept c_i and concept c_j is defined as the number of links in this path. If there are multiple paths from concept c_i to concept c_j in a concept hierarchy, then the distance *distance_ij* between concept c_i and concept c_j is defined as the number of links in the longest path from concept c_i to concept c_j.

Definition 4. Assume that concept c_h is both an ancestor of concept c_i and concept c_j in a concept hierarchy, then concept c_h is called a *common ancestor* of concept c_i and concept c_j.

Assume that concept c_h is selected as the context concept. If concept c_h is a common ancestor of concept c_i and concept c_j in a concept hierarchy, and concept c_i and concept c_j are in different branches of concept c_h and they are not in the same concept class, then there is a "fuzzy negative association" relationship between concept c_i and concept c_j. The degree of the fuzzy negative association relationship $N_{c_h}(c_i, c_j)$ between concept c_i and concept c_j when the context concept is concept c_h is calculated as follows:

$$N_{c_h}(c_i, c_j) = [\min(G(c_i, c_h), G(c_j, c_h))]^{distance\_ih+distanc\_jh-1}. \quad (9)$$

Because the context concept c_h may be different in different search contexts, there may be several possible values of $N_{c_h}(c_i, c_j)$. The overall degree of the fuzzy negative association relationship between concept c_i and concept c_j, denoted as $N(c_i, c_j)$, is calculated as follows:

$$N(c_i, c_j) = \max_{c_h \in C} N_{c_h}(c_i, c_j). \quad (10)$$

In the process of automatically constructing a multi-relationship fuzzy concept network, we can see that the degrees of fuzzy positive association relationships between concepts, the degrees of fuzzy negative association relationships between concepts, and the degrees of fuzzy specialization relationships between concepts are all derived based on the degrees of fuzzy generalization relationships between concepts. Therefore, the formula for calculating the degrees of fuzzy generalization relationships between concepts plays an important role in the process of automatically constructing a multi-relationship fuzzy concept network. In the above, we have presented two different formulas (i.e., formula (5) and formula (6)) for calculating the degrees of fuzzy generalization relationships between concepts. From Example 1 and Example 2, we can see that the proposed method (i.e., formula (6)) is better than formula (5) to calculate the degrees of generalization between concepts.

FUZZY INFERENCE BASED ON MULTI-RELATIONSHIP FUZZY CONCEPT NETWORKS

Although a multi-relationship fuzzy concept network explicitly describes the fuzzy relationships and their associated degrees of strength between concepts, some implicit links do not reveal in the multi-relationship fuzzy concept network. For example, let us consider the multi-relationship fuzzy concept network shown in Figure 6. It contains three concepts c_1, c_2, and c_3 and three explicit directed links l_1, l_2, and l_3 between these concepts. Figure 6 shows that besides the explicit directed link l_3 from concept c_1 to concept c_3, which is constructed from the proposed multi-relationship fuzzy concept network construction method, there also exists an implicit directed link l_4 between these two concepts, which is inferred through the directed link l_1 from concept c_1 to concept c_2 and through the directed link l_2 from concept c_2 to concept c_3. In this case, the actual degrees of strength of relationships between concept c_1 and concept c_3 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 for inferring the fuzzy relationships and the

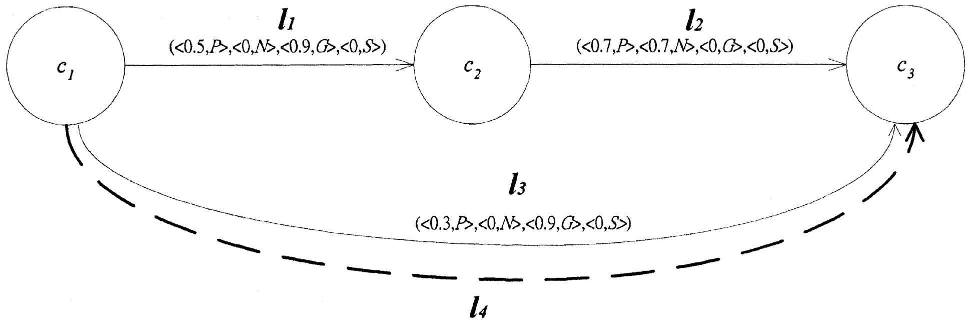


FIGURE 6. Inferring the fuzzy relationships and the associated degrees of the implicit directed link l_4 from concept c_1 to concept c_3.

associated degrees of the 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^r_{ij} where $\mu^r_{ij} \in [0, 1]$, and if the degree of fuzzy relationship r between concept c_j and concept c_k is μ^r_{jk}, where $\mu^r_{jk} \in [0,1]$, then the degree μ^r_{ik} of fuzzy relationship r between concept c_i and concept c_k can be inferred as follows:

$$\mu^r_{ik} = \mu^r_{ij} \wedge \mu^r_{jk}, \tag{11}$$

where $\wedge$ is the minimum operator and $\mu^r_{ik} \in [0,1]$. Furthermore, if the degree of fuzzy relationship r between concept c_1 and concept c_2 is μ^r_{12}, the degree of fuzzy relationship r between concept c_2 and concept c_3 is $\mu^r_{23}, \ldots$, and the degree of fuzzy relationship r between concept c_{n-1} and concept c_n is $\mu^r_{(n-1)n}$, where $\mu^r_{12} \in [0,1]$, $\mu^r_{23} \in [0,1], \ldots$, and $\mu^r_{(n-1)n} \in [0,1]$, then the degree of fuzzy relationship r between concept c_1 and concept c_n is μ^r_{1n}, where $\mu^r_{1n} \in [0,1]$ and

$$\mu^r_{1n} = \mu^r_{12} \wedge \mu^r_{23} \wedge \cdots \wedge \mu^r_{(n-1)n}. \tag{12}$$

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

$$\mu^r_{1n} = \max(\mu^{r(1)}_{1n}, \mu^{r(2)}_{1n}, \ldots, \mu^{r(h)}_{1n}) \tag{13}$$

where $\mu^{r(i)}_{1n}$ 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 $1 \leq i \leq h$. Therefore, based on formula (11) and Figure 6, we can see that the degrees of strength of fuzzy relationships associated with the implicit link l_4 are $\mu_P = 0.5$

(i.e., the degree of strength of the "fuzzy positive association" relationship is equal to 0.5) and $\mu_N = \mu_G = \mu_S = 0$ (i.e., the degrees of strength of the "fuzzy negative association" relationship, the "fuzzy generalization" relationship, and the "fuzzy specialization" relationship are all equal to 0). Furthermore, based on formula (13), the aggregated degrees of strength of fuzzy relationships between the concept c_1 and the concept c_3 are $\mu_P = 0.5$, $\mu_G = 0.9, \mu_N = 0$, and $\mu_S = 0$.

To obtain the implicit relationships between concepts, we adopt the method presented in Chen and Wang (1995) by modeling a multi-relationship fuzzy concept network using four concept matrices due to the fact that there are four kinds of fuzzy relationships in a multi-relationship fuzzy concept network, where each concept matrix is used to model one kind of fuzzy relationship. After calculating the transitive closures of these concept matrices, the implicit relationships and their associating degrees of strength can be obtained.

Definition 5. A concept matrix U_r is a fuzzy matrix (Kandel 1986),

$$U_r = \begin{array}{c} \\ c_1 \\ c_2 \\ \vdots \\ c_n \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_n \end{array} \\ \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nn} \end{bmatrix} \end{array},$$

where the element u_{ij} represents the degree of strength of fuzzy relationship r between concept c_i and concept c_j, where $r \in \{P, N, G, S\}, u_{ij} \in [0, 1]$, $1 \leq i \leq n$, and $1 \leq j \leq n$. The elements of U_P are obtained by formula (8); the elements of U_N are obtained by formula (10); the elements of U_G are obtained by formula (6); the elements of U_S are obtained by formula (7). If fuzzy relationship r is reflexive (i.e., r is the fuzzy positive association relationship), then $u_{ii} = 1$. Otherwise, $u_{ii} = 0$. If fuzzy relationship r is symmetric (i.e., r is the fuzzy positive association relationship or the fuzzy negative association relationship), then $u_{ij} = u_{ji}$. If $u_{ij} = u_{ji} = 0$, then concept c_i and concept c_j are not related by fuzzy relationship r, where $r \in \{P, N, G, S\}$.

Definition 6. Assume that U_r is a concept matrix, where $r \in \{P, N, G, S\}$. If fuzzy relationship r is nontransitive (i.e., r is the fuzzy negative association relationship), then we let the transitive closure U_r^* of U_r be itself, i.e., $U_r^* = U_r$. If fuzzy relationship r is transitive (i.e., r is the fuzzy positive association relationship, the fuzzy generalization relationship, or the fuzzy specialization relationship), then the transitive closure U_r^* (Chen and Wang 1995) of the concept U_r is defined as follows. Let

$$U_r^2 = U_r \otimes U_r$$
$$= \begin{bmatrix} \bigvee\limits_{i=1,2,\ldots,n} (u_{1i} \wedge u_{i1}) & \bigvee\limits_{i=1,2,\ldots,n} (u_{1i} \wedge u_{i2}) & \cdots & \bigvee\limits_{i=1,2,\ldots,n} (u_{1i} \wedge u_{in}) \\ \bigvee\limits_{i=1,2,\ldots,n} (u_{2i} \wedge u_{i1}) & \bigvee\limits_{i=1,2,\ldots,n} (u_{2i} \wedge u_{i2}) & \cdots & \bigvee\limits_{i=1,2,\ldots,n} (u_{2i} \wedge u_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ \bigvee\limits_{i=1,2,\ldots,n} (u_{ni} \wedge u_{i1}) & \bigvee\limits_{i=1,2,\ldots,n} (u_{ni} \wedge u_{i2}) & \cdots & \bigvee\limits_{i=1,2,\ldots,n} (u_{ni} \wedge u_{in}) \end{bmatrix}, \tag{14}$$

where "$\vee$" is the maximum operator and "$\wedge$" is the minimum operator. Then, there exists a positive integer k, where $k \leq n-1$, such that $U_r^k = U_r^{k+1} = U_r^{k+2} = \cdots$. Let $U_r^* = U_r^k$, then U_r^* is called the transitive closure of the concept matrix U_r.

After calculating the transitive closures of the concept matrices, the actual degrees of strength of fuzzy relationships between concepts can be obtained. In the following, we use a document descriptor matrix to represent the degrees of concepts in documents. The definition of the document descriptor matrix is shown as follows.

Definition 7. Let D be a set of documents, $D = \{d_1, d_2, \ldots, d_m\}$, and let C be a set of concepts, $C = \{c_1, c_2, \ldots, c_n\}$. The document descriptor matrix T is shown as follows:

$$T = \begin{array}{c} \\ d_1 \\ d_2 \\ \vdots \\ d_m \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_n \end{array} \\ \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \end{bmatrix} \end{array},$$

where m is the number of documents, n is the number of concepts, t_{ij} indicates the weight of concept c_j in document d_i which is obtained by formula (3), $t_{ij} \in [0, 1]$, $1 \leq i \leq m$, and $1 \leq j \leq n$.

FUZZY QUERY PROCESSING FOR DOCUMENT RETRIEVAL BASED ON MULTI-RELATIONSHIP FUZZY CONCEPT NETWORKS

In the following, we present fuzzy query processing techniques for document retrieval based on multi-relationship fuzzy concept networks. The users of fuzzy information retrieval systems based on the multi-relationship fuzzy concept networks can submit contextual queries. That is, the users can specify a concept as the search context in the query formula. According to the search context, documents whose contents are related to the concepts listed in the user's query by the context-dependent fuzzy relationship (i.e., "fuzzy negative association" relationship) will be retrieved by the fuzzy information retrieval system. On the other hand, if the users wish the fuzzy information retrieval system to retrieve documents whose contents are related to the concepts specified in the query by only context-independent fuzzy relationships, the user does not need to specify the search context in the query formula. In the following, we propose a method for processing users' queries. The user's query Q has the following format:

$$Q = \{c_T, (c_1, r_1, x_1), (c_2, r_2, x_2), \ldots, (c_n, r_n, x_n)\},$$

where $c_T \in C$ is a concept representing the search context, c_i is a concept in a multi-relationship fuzzy concept network, $r_i \in \{P, N, G, S\}$ indicates the desired fuzzy relationship of concept c_i to the retrieved documents, $x_i \in [0, 1]$ indicates the desired degree of relationship between concept c_i and the retrieved documents, $1 \leq i \leq n$, and n is the number of concepts in the multi-relationship fuzzy concept network. In a user's query Q, the context concept can be ignored due to the fact that the user wants the fuzzy information retrieval system to retrieve documents whose contents are related to the concepts specified in the query by only context-independent fuzzy relationships, where the context concept "$c_T = \varepsilon$" indicating the search context has been ignored. If $x_i = 0$, then it indicates that documents desired by the user do not have concept c_i. If the user considers that some concepts may be ignored (i.e., to include those concepts or not would have no substantial effect on the retrieval result), then the user does not have to assign fuzzy relationships and degrees of documents with respect to such concepts in the query, where the symbol "$-$" is used for labeling an ignored concept.

The user's query Q can be represented by a query descriptor vector $\bar{q}$ as follows:

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

where $x_i \in [0, 1]$ indicates the desired degree of relationship between concept c_i and the retrieved documents, $1 \leq i \leq n$, and n is the number of concepts in a multi-relationship fuzzy concept network. The query descriptor vector $\bar{q}$ is then expanded to the expanded query descriptor vector $\overline{q^*}$ by the proposed query-vector expanding algorithm which adds more related concepts into the query formula and then the information retrieval system can retrieve more relevant documents. In the following, we present the proposed query-vector expanding algorithm. Assume that four concept matrices U_P, U_N, U_G, and

U_S are used to model a multi-relationship fuzzy concept network, and the transitive closures U_P^*, U_N^*, U_G^*, and U_S^* of the concept matrices U_P, U_N, U_G, and U_S, respectively, are given. The query-vector expanding algorithm is now presented as follows.

Query-Vector Expanding Algorithm

If $c_T =$ 'ε' **then** /* expanded by context-independent fuzzy relationships */
  begin
    for $i = 1$ **to** n **do** /* n is the number of concepts in a multi-relationship fuzzy concept network */
    begin
      if $r_i = P$ **then**
      for $j = 1$ **to** n **do**
        if $U_P^*(c_i, c_j) > 0$ **then**

$$\overline{q_j^*} = \begin{cases} \max(x_j, U_P^*(c_i, c_j)), & \text{if } x_j \neq \text{“}-\text{”} \\ U_P^*(c_i, c_j), & \text{if } x_j = \text{“}-\text{”} \end{cases}$$

      if $r_i = G$ **then**
        for $j = 1$ **to** n **do**
          if $U_G^*(c_i, c_j) > 0$ **then**

$$\overline{q_j^*} = \begin{cases} \max(x_j, U_G^*(c_i, c_j)), & \text{if } x_j \neq \text{“}-\text{”} \\ U_G^*(c_i, c_j), & \text{if } x_j = \text{“}-\text{”} \end{cases}$$

      if $r_i = S$ **then**
        for $j = 1$ **to** n **do**
          if $U_S^*(c_i, c_j) > 0$ **then**

$$\overline{q_j^*} = \begin{cases} \max(x_j, U_S^*(c_i, c_j)), & \text{if } x_j \neq \text{“}-\text{”} \\ U_S^*(c_i, c_j), & \text{if } x_j = \text{“}-\text{”} \end{cases}$$

    end
  end
else /* expanded by context-dependent fuzzy relationships (i.e., fuzzy negative association relationship) */
  begin
    if $r_i = N$ **then**
      for $j = 1$ **to** n **do**
        if concept c_i and concept c_j are in different branches of concept c_t and $U_N^*(c_i, c_j) > 0$
        then

$$\overline{q_j^*} = \begin{cases} \max(x_j, U_N^*(c_i, c_j)), & \text{if } x_j \neq \text{“}-\text{”} \\ U_N^*(c_i, c_j), & \text{if } x_j = \text{“}-\text{”} \end{cases}$$

end.

Let x and y be two values, where $x \in [0, 1]$ and $y \in [0, 1]$. The degree of similarity between x and y can be evaluated by the function H (Chen and Wang 1995):

$$H(x, y) = 1 - |x - y|, \tag{15}$$

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

Assume that the document descriptor vector $\bar{D}_i$ (i.e., the ith row of the document descriptor matrix T) and the expanded query descriptor vector $\overline{q^*}$ are represented as follows:

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

$$\overline{q^*} = \langle x_1, x_2, \ldots, x_n \rangle,$$

where $s_{ij} \in [0, 1]$, $x_i \in [0, 1]$, $1 \leq j \leq n$, $1 \leq i \leq m$, n is the number of concepts, and m is the number of documents. Then, based on Chen and Wang (1995), the degree of satisfaction $DS(d_i)$ that document d_i satisfies the user's query Q can be evaluated as follows:

$$DS(d_i) = \frac{\sum_{\overline{q^*}(j) \neq ''-'' \text{ and } j=1,\ldots,n} H(s_{ij}, x_j)}{k}, \tag{16}$$

where $DS(d_i) \in [0, 1]$, $1 \leq i \leq n$, and k is the number of concepts not ignored by the user's query. The larger the value of $DS(d_i)$, the more the degree of satisfaction that the document d_i satisfies the user's query. A document d_i is retrieved if $DS(d_i)$ is larger than a retrieval threshold value β given by the user, where $\beta \in [0, 1]$ and $1 \leq i \leq n$. The retrieved documents are then ranked in a decreasing sequence according to their degrees of satisfaction with respect to the user's query.

Example 3. Assume that there is a multi-relationship fuzzy concept network as shown in Figure 2, where c_1, $c_2, \ldots,$ and c_5 are concepts, and d_1, d_2 and d_3 are documents. The corresponding concept matrices U_P, U_N, U_G, and U_S of the multi-relationship fuzzy concept network shown in Figure 2 are shown as follows:

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

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

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

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

Based on Definition 6, the transitive closures of these four concept matrices can be obtained shown as follows, where U_P^* is the transitive closure of U_P, U_N^* is the transitive closure of U_N, U_G^* is the transitive closure of U_G, and U_S^* is the transitive closure of U_S (note that $U_N^* = U_N$ due to the fact that the fuzzy negative association is not a transitive relationship):

$$U_P^* = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.3 & 0.5 & 0.3 \\ 0.2 & 0.3 & 1 & 0.3 & 0.3 \\ 0.2 & 0.5 & 0.3 & 1 & 0.3 \\ 0.2 & 0.3 & 0.3 & 0.3 & 1 \end{bmatrix},$$

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

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

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

Furthermore, we can use a document descriptor matrix T to model the degrees of documents containing the concepts in the multi-relationship fuzzy concept network shown in Figure 2 as follows:

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

Case 1. Assume that the user's query Q is as follows:

$$Q = \{c_3, (c_1, \text{-}, \text{-}), (c_2, \text{-}, \text{-}), (c_3, \text{-}, \text{-}), (c_4, N, 0.8), (c_5, \text{-}, \text{-})\}.$$

It means the user wants to retrieve documents whose contents are related to concept c_4 by the fuzzy negative association relationship with the degrees 0.8 when the search context is concept c_3. Then, the corresponding query descriptor vector $\bar{q}$ is as follows:

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

Then, based on the proposed query-vector expansion algorithm, we can derive the expanded query descriptor vector $\overline{q^*}$ as follows:

$$\overline{q^*} = [0.8, \text{-}, \text{-}, 0.8, 0.8].$$

Based on formula (16), we can calculate the degree of satisfaction that document d_i, where $1 \leq i \leq 3$, satisfies the user's query shown as follows:

$$\begin{aligned} DS(d_1) &= \frac{(1-|0.1-0.8|)+(1-|0.9-0.8|)+(1-|0-0.8|)}{3} \\ &= 0.47, \\ DS(d_2) &= \frac{(1-|0.7-0.8|)+(1-|0-0.8|)+(1-|0-0.8|)}{3} \\ &= 0.43, \\ DS(d_3) &= \frac{(1-|0-0.8|)+(1-|0-0.8|)+(1-|1-0.8|)}{3} \\ &= 0.4, \end{aligned}$$

where $DS(d_i)$ is the degree of satisfaction of document d_i with respect to the user's query, and $1 \leq i \leq 3$. Assume that the retrieval threshold value β given by the user is 0.4, then all the documents will be retrieved in this example due to the fact that their degrees of satisfaction with respect to the user's query are larger than 0.4. The order from the retrieved document with the largest degree of satisfaction to that with the smallest degree of satisfaction is $d_1 > d_2 > d_3$. In this case, document d_1 is the best choice for the user's query.

Case 2. Assume that the user's query Q is as follows:

$$Q = \{\varepsilon, (c_1, \text{-}, \text{-}), (c_2, \text{-}, \text{-}), (c_3, \text{-}, \text{-}), (c_4, G, 0.8), (c_5, \text{-}, \text{-})\}.$$

It means the user wants to retrieve documents whose contents are related to concept c_4 by the fuzzy generalization relationship with the degree 0.8. The corresponding query descriptor vector $\bar{q}$ is as follows:

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

Then, based on the proposed query-vector expansion algorithm, we can derive the expanded query descriptor vector $\overline{q^*}$ as follows:

$$\overline{q^*} = [\text{-}, 0.8, \text{-}, 0.8, \text{-}].$$

Based on formula (16), we can calculate the degree of satisfaction of document d_i, where $1 \leq i \leq 3$, satisfies the user's query shown as follows:

$$\begin{aligned} DS(d_1) &= \frac{(1-|0-0.8|)+(1-|0.9-0.8|)}{2} \\ &= 0.55, \\ DS(d_2) &= \frac{(1-|0.3-0.8|)+(1-|0-0.8|)}{2} \\ &= 0.35, \\ DS(d_3) &= \frac{(1-|0-0.8|)+(1-|0-0.8|)}{2} \\ &= 0.2, \end{aligned}$$

where $DS(d_i)$ is the degree of satisfaction of document d_i with respect to the user's query, and $1 \leq i \leq 3$. Assume that the retrieval threshold value β given by the user is 0.3, then document d_3 will not be retrieved in this example due to the fact that its degree of satisfaction with respect to the user's query is smaller than 0.3. The order from the retrieved document with the largest degree of satisfaction to that with the smallest degree of satisfaction is $d_1 > d_2$. In this case, document d_1 is the best choice for the user's query.

CONCLUSIONS

In this paper, we have presented a method to automatically construct multi-relationship fuzzy concept networks for fuzzy information retrieval systems for document retrieval, where multi-relationship fuzzy concept networks are used as knowledge bases of fuzzy information retrieval systems. Two methods for calculating the degrees of fuzzy generalization relationships between concepts are discussed. Four concept matrices are then used to model the multi-relationship fuzzy concept networks. The implicit degrees of relationships between concepts can be obtained by calculating the transitive closures of the concept matrices. The users of the fuzzy information retrieval systems based on the multi-relationship fuzzy concept networks can submit a contextual query in which the search context is involved. We also presented a query-vector expanding algorithm to expand the query descriptor vectors. Documents are retrieved if they contain concepts that have the specified fuzzy relationships with the concepts contained in the user's query in the user-specified search context. The proposed fuzzy information retrieval method is more intelligent and more flexible than the existing methods due to the fact that it can automatically construct multi-relationship fuzzy concept networks and it can provide contextual search capability to allow users to specify fuzzy contextual queries in a more intelligent and flexible manner.

REFERENCES

Bezdek, J. C., G. Biswas, and L. Y. Huang. 1986. Transitive closures of fuzzy thesauri for information-retrieval systems. *International Journal of Man-Machine Studies* 25(3):343–356.

Bhatia, S. K., and J. S. Deogun. 1998. Conceptual clustering in information retrieval. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics* 28(3):427-435.

Chang, C. S., and A. L. P. Chen. 1998. Supporting conceptual and neighborhood queries on the World Wide Web. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics* 28(2):300–308.

Chen, C. L. P., and Y. Lu. 1997. FUZZ: A fuzzy-based concept formation system that integrates human categorization and numerical clustering. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics* 27(1):79–94.

Chen, S. M., and Y. J. Horng. 1999. Fuzzy query processing for document retrieval based on extended fuzzy concept networks. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics* 29(1):126–135.

Chen, S. M., W. H. Hsiao, and Y. J. Horng. 1997. A knowledge-based method for fuzzy query processing for document retrieval. *Cybernetics and Systems: An International Journal* 28(1):99–119.

Chen, S. M., and J. Y. Wang. 1995. Document retrieval using knowledge-based fuzzy information retrieval techniques. *IEEE Transactions on Systems, Man, and Cybernetics* 25(6):793–803.

Frakes, W. B. 1992. Stemming algorithms. In *Information Retrieval: Data Structure & Algorithms*, eds. W. B. Frakes, and R. Baeza-Yates. New Jersey: Prentice Hall.

Horng, Y. J., S. M. Chen, and C. H. Lee. 2000. A fuzzy information retrieval method based on multi-relationship fuzzy concept networks. In *Proceedings of the 2000 International Computer Symposium: Workshop on Artificial Intelligence*, pages 79–86, Chiayi, Taiwan, Republic of China.

Horng, Y. J., S. M. Chen, and C. H. Lee. 2001. Automatically constructing multi-relationship fuzzy concept networks in fuzzy information retrieval systems. In *Proceedings of the 10th IEEE International Conference on Fuzzy Systems*, Melbourne, Australia.

Kandel, A. 1986. *Fuzzy Mathematical Techniques with Applications*. CA: Addison-Wesley.

Kim, K. J., and S. B. Cho. 2001. A personalized web search engine using fuzzy concept network with link structure. In *Proceedings of the Joint 9th IFSA Congress and 20th NAFIPS International Conference*, pages 81–86 Vancouver, Canada.

Kracker, M. 1992. A fuzzy concept network model and its applications. In *Proceedings of the First IEEE International Conference on Fuzzy Systems*, pages 761–768, San Diego, U. S. A.

Kosko, B. 1992. *Neural Networks and Fuzzy Systems*. New Jersey: Prentice-Hall.

Liang, T., and C. C. Chang. 1999. Chinese textual retrieval based on fuzzy concept networks. In *Proceedings of National Computer Symposium*, Tamsui, Taiwan, Republic of China, Vol. 1, 61–67.

Lin, C. C., S. Y. Tseng, and P. M. Chen. 1999. A fuzzy document retrieval system based on concept networks and cluster analysis. *Soochow Journal of Economics and Business* (25):39–60.

Lucarella, D., and R. Morara. 1991. FIRST: Fuzzy information retrieval system. *Journal of Information Science* 17(1):81–91.

Salton, G. 1971. *The SMART Retrieval System: Experiments in Automatic Document Processing*. New Jersey: Prentice Hall.

Salton, G., and C. Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing and Management* 24(5):513–523.

Salton, G., and M. J. Mcgill. 1983. *Introduction to Modern Information Retrieval*. New York: McGraw-Hill.

Young, V. R. 1996. Fuzzy subsethood. *Fuzzy Sets and Systems* 77(3):371–384.

Zadeh, L. A. 1965. Fuzzy sets. *Information and Control* 8:338–353.