

Research Article

OMFM: A Framework of Object Merging Based on Fuzzy Multisets

Lin Yue,^{1,2,3} Wanli Zuo,^{1,2} Lizhou Feng,^{1,2} and Lin Guo^{1,2}

¹ College of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China

² Key Laboratory of Symbol Computation and Knowledge Engineering of The Ministry of Education, Changchun, Jilin 130012, China

³ School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia

Correspondence should be addressed to Wanli Zuo; wanli@jlu.edu.cn

Received 2 May 2014; Accepted 16 October 2014; Published 10 November 2014

Academic Editor: Pandian Vasant

Copyright © 2014 Lin Yue et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Information fusion is a process of merging information from multiple sources into a new set of information. Existing work on information fusion is applicable in various scenarios such as multiagent system, group decision making, and multidocument summarization. This paper intends to develop an effective framework to solve object merging problem based on fuzzy multisets. The objects defined in this paper are data segments in document fusion task, referring to the concepts with semantic-related terms of different semantic relations embedded. The fundamental operation is the merge function mapping data segments in multiple fuzzy multisets onto one object, which is a solution. Under this framework, we define quality measures of purity and entropy to quantify the quality of the solutions, balancing accurateness, and completeness of the results. Merge function that yields this kind of solutions is VI-optimal merge function and a series of theoretical properties concerning it are studied. Finally, we investigate the proposed framework in a special application scenario (i.e., document fusion) which is related to the task of multidocument summarization and show how the framework works with illustrative example.

1. Introduction

As an important research area, information fusion is a process of merging information from multiple sources into a new set of information. There are many applications in this research area such as heterogeneous database, multiagent system, group decision making, and multidocument summarization. Under different application scenarios, different principles and procedures are utilized to solve the problems. Many classical mathematical theories of aggregation operators [1–5] have been developed for multiagent system and group decision making system, and the information that aggregation operators try to fuse typically expresses facts of opinion or score of an agent. Besides these researches, a fair amount of work focused on the situation where the source is regarded as a propositional belief [6–8]. The existence of nonfactual knowledge like integrity constraints and inference rules makes the difference between these two theories. As a result, a lot of work has been done in the heterogeneous database area on first-order theory. Another type of fusion is that each source

presents knowledge by means of a possibility distribution [9], in this case, the imperfection of incorrectness, uncertainty, and incompleteness in the data should be coped with. The main challenge is how to deal with conflicting information provided by different sources.

To address the issues in the third type of fusion, a framework of object merging has been investigated by using multiset theory currently, which could be utilized to solve the problem of multidocument summarization (MDS) [10]. Also, object merging is a hot spot of research in many domains with good prospect for application. The framework of multiset merging for MDS has defined the merge function which maps the objects in multisets onto a single object and has got some foregoing results which cannot be considered as a final summarization yet [11, 12], owing to the fact that these foregoing results are just some keywords without any relation among them, not mention to context of co-text, context of culture, context of situation, and so forth. The essential reason for this result is that the framework defined the quality measures with the multiplicity of element as the measure of

important element. In other words, the multiplicity is equal to term frequency which is just shallow text feature. When performing source selection in MDS, the traditional method transformed one document into the representation of a vector of words or a multiset of words, which are just simple settings. Other progressive approaches should be proposed, which are semantically richer than using words as source representation. In short, the problem of processing coreferent objects has not been deeply investigated at present. On one hand, merging of nonquantitative objects, especially the objects with semantic information, has not been proposed. On the other hand, object merging functions and the rationality of merging still need to be further investigated.

Within the scope of our paper, we also focus on the problem of object merging in information fusion, and our work should be treated as an extension of the framework mentioned hereinbefore. There are many differences between these two works. The basic difference concerning the definition of coreferent objects: coreferent objects in paper [11] are the objects describing the same entity in the real world, while in our paper the object we're discussing is a piece of data or information, which could be used to denote the same concept with semantic-related terms of different semantic relations embedded. Then, fuzzy multiset theory is investigated in our paper, in which membership degree function and length function are used to describe both uncertainty and repeatability of the natural language. When performing fusion in practical situations, the object merging process has considered deep text features of semantic relations such as hypernym, synonym, and antonym. Moreover, two quality measures (purity [13] and entropy [14, 15]) widely used in the text mining literature are adopted to quantify the result of a merge function. Thus, the behavior of the merge functions we defined in this paper can be characterized by the behavior of the quality measures. With this strategy, we can get an optimal merge result. The possible application of this work is document fusion [16], where a collection of textual documents is used to produce the shortest description containing all information found within the document set, but without repetition. Existing solutions for this problem normally focused on statistical methods or heuristics methods used in multidocument summarization [17, 18]. In this paper, object merging based on fuzzy multisets (OMFM) is definitely a meaningful attempt, where a source set of multiple documents is denoted as a multiset and each document is denoted as a fuzzy multiset of multiple concepts.

This paper is organized as follows. In Section 2, we review mathematical preliminaries. Furthermore, the general framework of objects and object merging are proposed in Section 3, and definition of the quality measures and construction of merge functions are introduced in Section 4. Next, demonstration about how our framework works on practical problem (i.e., document fusion) with illustrative example is presented in Section 5. Finally, in Section 6, we give the conclusion and future work to the proposed framework OMFM.

2. Preliminaries

In mathematics, fuzzy set introduced by Zadeh in 1965 is set whose elements have degrees of membership which is an extension of the classical notion of set [19, 20]. Fuzzy set theory is very useful to deal with problems that are not easily handled by classical computing techniques. On the other hand, the use of membership degrees instead of real numbers to represent memberships also provides a mean to measure the possible uncertainty in languages computational theory. The notion of multiset is a generalization of the classical notion of set in which members are allowed to appear more than once. As a data structure, multiset stands in between strings where a linear ordering of symbols is presented and sets where no ordering is considered. Combined with the notion of fuzzy set, multiset is generalized to fuzzy multiset [21], which could describe both uncertainty and repeatability of the natural language. Consider one language modeling problem: given some sentences, identify the concepts and words which are similar or identical, and merge these objects to get a condensed description. This task is a challenging natural language problem with large amounts of diverse and compositional data. To solve this problem, we extend fuzzy multiset to produce a language model which maps data segments in multiple fuzzy multisets onto one object, where different semantic relations for one concept are treated as repeated elements with different membership degree in fuzzy multisets. In this section, mathematical theories of fuzzy set, multiset, and fuzzy multiset will be briefly reviewed.

Definition 1 (membership function). The membership function $\mu_A : U \rightarrow [0, 1]$ indicates the degree of u belonging to A . $\mu_A(u) = 1$ indicates that element u completely belongs to set A ; that is, $u \in A$ is the concept of traditional set.

Definition 2 (fuzzy set). The membership function over $U = \{u_1, u_2, \dots, u_n\}$ defines a fuzzy set, which is represented as A . Fuzzy set A with elements u_1, u_2, \dots, u_n can be denoted as $A = \{u_i, (\mu_A(u_i)) \mid u_i \in U\}$.

According to the definition of fuzzy set, to what extent an object belongs to a set is not fixed any more, and the membership of each object falls in the range of interval $[0, 1]$.

2.1. Multiset

Definition 3 (multiset). A multiset M over the universe $U = \{u_1, u_2, \dots, u_n\}$ is defined as $M = \{k_1/u_1, k_2/u_2, \dots, k_n/u_n\}$, where $k_i = \text{Count}_M(u_i)$, $u_i \in U$. For $u_i \in U$, $\text{Count}_M(u_i)$ denotes the multiplicity of u_i in M .

The cardinality of a multiset M is given by $|M| = \sum_{u_i \in U} \text{Count}_M(u_i)$ [22].

A multiset M also could be denoted as $M = \left\{ \overbrace{u_1, u_1, \dots, u_1}^{k_1}, \overbrace{u_2, u_2, \dots, u_2}^{k_2}, \dots, \overbrace{u_n, u_n, \dots, u_n}^{k_n} \right\}$. The set of all multisets drawn from a universe U is denoted as $\mathfrak{M}(U)$.

There are some basic operators and relations of multiset below:

Inclusion:

$$M_{(P)} \subseteq M_{(Q)} \iff \text{Count}_{M_{(P)}}(u) \leq \text{Count}_{M_{(Q)}}(u), \quad (1)$$

$$\forall u \in U.$$

Equality:

$$M_{(P)} = M_{(Q)} \iff \text{Count}_{M_{(P)}}(u) = \text{Count}_{M_{(Q)}}(u), \quad (2)$$

$$\forall u \in U.$$

Intersection:

$$\text{Count}_{(M_{(P)} \cap M_{(Q)})}(u) = \min \{ \text{Count}_{M_{(P)}}(u), \text{Count}_{M_{(Q)}}(u) \}, \quad (3)$$

$$\forall u \in U.$$

Union:

$$\text{Count}_{(M_{(P)} \cup M_{(Q)})}(u) = \max \{ \text{Count}_{M_{(P)}}(u), \text{Count}_{M_{(Q)}}(u) \}, \quad (4)$$

$$\forall u \in U.$$

Addition:

$$\text{Count}_{(M_{(P)} \oplus M_{(Q)})}(u) = \text{Count}_{M_{(P)}}(u) + \text{Count}_{M_{(Q)}}(u), \quad (5)$$

$$\forall u \in U.$$

Definition 4 (k -cut set of multiset). The k -cut set of a multiset M is denoted as M_k and given by $M_k = \{u \mid u \in U \wedge \text{Count}_M(u) \geq k\}$.

Note that the difference between the notation $M_{(k)}$ and M_k is that the former one means assigning an index k to the multiset M and the latter one means the k -cut set of the multiset M [23].

2.2. Fuzzy Multiset. Combined with the concept of fuzzy set, the traditional concept of multiset could be denoted as

$$\widetilde{M} = \left\{ \begin{array}{l} \overbrace{\frac{\mu_{\widetilde{M}}^1(u_1)}{u_1}, \frac{\mu_{\widetilde{M}}^2(u_1)}{u_1}, \dots, \frac{\mu_{\widetilde{M}}^{k_1}(u_1)}{u_1}}^{k_1}, \\ \overbrace{\frac{\mu_{\widetilde{M}}^1(u_2)}{u_2}, \frac{\mu_{\widetilde{M}}^2(u_2)}{u_2}, \dots, \frac{\mu_{\widetilde{M}}^{k_2}(u_2)}{u_2}}^{k_2}, \dots, \\ \overbrace{\frac{\mu_{\widetilde{M}}^1(u_n)}{u_n}, \frac{\mu_{\widetilde{M}}^2(u_n)}{u_n}, \dots, \frac{\mu_{\widetilde{M}}^{k_n}(u_n)}{u_n}}^{k_n} \end{array} \right\}, \quad (6)$$

where $\mu_{\widetilde{M}}^j(u_i) \in [0, 1]$, $j = 1, 2, \dots$, represents the membership degree of element u_i recurring in multiset for the j th

time. According to the traditional concept of set, the membership degree of all elements in a multiset is 1. For $\mu_{\widetilde{M}}^j(u_i) \in [0, 1]$, the traditional sense of multiset is generalized to fuzzy multiset.

Definition 5 (fuzzy multiset). A fuzzy multiset M over the universe $U = \{u_1, u_2, \dots, u_n\}$ is defined as

$$\widetilde{M} = \left\{ \begin{array}{l} \frac{(\mu_{\widetilde{M}}^1(u_1), \mu_{\widetilde{M}}^2(u_1), \dots, \mu_{\widetilde{M}}^{k_1}(u_1))}{u_1}, \\ \frac{(\mu_{\widetilde{M}}^1(u_2), \mu_{\widetilde{M}}^2(u_2), \dots, \mu_{\widetilde{M}}^{k_2}(u_2))}{u_2}, \dots, \\ \frac{(\mu_{\widetilde{M}}^1(u_n), \mu_{\widetilde{M}}^2(u_n), \dots, \mu_{\widetilde{M}}^{k_n}(u_n))}{u_n} \end{array} \right\}, \quad (7)$$

where $\{\mu_{\widetilde{M}}^1(u_i), \mu_{\widetilde{M}}^2(u_i), \dots, \mu_{\widetilde{M}}^{k_i}(u_i)\}$ is the degree of membership of element u_i ($\mu_{\widetilde{M}}^j(u_i) \in [0, 1]$).

Definition 6 (length function). The concept of multiplicity $\text{Count}_M(u_i)$ in multiset generalized to fuzzy multiset is length function, which is denoted as $L(u_i, \widetilde{M}) = \max\{j : \mu_{\widetilde{M}}^j(u_i) \neq 0\}$ [24].

The number of occurrence or cardinality of a fuzzy multiset \widetilde{M} is given by $|\widetilde{M}| = \sum_{u_i \in U} \sum_{j=1}^{L(u_i, \widetilde{M})} \mu_{\widetilde{M}}^j(u_i)$ [22].

There are some basic operators and relations of fuzzy multiset below:

Inclusion:

$$\widetilde{M}_{(P)} \subseteq \widetilde{M}_{(Q)} \iff \mu_{\widetilde{M}_{(P)}}^j(u_i) \leq \mu_{\widetilde{M}_{(Q)}}^j(u_i), \quad (8)$$

$$j = 1, 2, \dots, L(u_i), \quad \forall u_i \in U.$$

Equality:

$$\widetilde{M}_{(P)} = \widetilde{M}_{(Q)} \iff \mu_{\widetilde{M}_{(P)}}^j(u_i) = \mu_{\widetilde{M}_{(Q)}}^j(u_i), \quad (9)$$

$$j = 1, 2, \dots, L(u_i), \quad \forall u_i \in U.$$

Addition: For $\widetilde{M}_{(P)} = \{(u_i, \mu_i), \dots, (u_k, \mu_k)\}$, $\widetilde{M}_{(Q)} = \{(u_s, \mu_s), \dots, (u_r, \mu_r)\}$, $\widetilde{M}_{(P)} \oplus \widetilde{M}_{(Q)} = \{(u_i, \mu_i), \dots, (u_k, \mu_k), (u_s, \mu_s), \dots, (u_r, \mu_r)\}$.

Intersection:

$$\mu_{\widetilde{M}_{(P)} \cap \widetilde{M}_{(Q)}}^j(u_i) = \min \{ \mu_{\widetilde{M}_{(P)}}^j(u_i), \mu_{\widetilde{M}_{(Q)}}^j(u_i) \}, \quad (10)$$

$$j = 1, 2, \dots, L(u_i), \quad \forall u_i \in U.$$

Union:

$$\mu_{\widetilde{M}_{(P)} \cup \widetilde{M}_{(Q)}}^j(u_i) = \max \{ \mu_{\widetilde{M}_{(P)}}^j(u_i), \mu_{\widetilde{M}_{(Q)}}^j(u_i) \}, \quad (11)$$

$$j = 1, 2, \dots, L(u_i), \quad \forall u_i \in U.$$

Note that when performing any operator for two fuzzy multisets, the length of the membership degree sequences $\mu_{\widetilde{M}_{(p)}}^1(u_i), \mu_{\widetilde{M}_{(p)}}^2(u_i), \dots, \mu_{\widetilde{M}_{(p)}}^j(u_i)$ and $\mu_{\widetilde{M}_{(Q)}}^1(u_i), \mu_{\widetilde{M}_{(Q)}}^2(u_i), \dots, \mu_{\widetilde{M}_{(Q)}}^j(u_i)$, will be set up to be similar. For this reason, adequate number of zeros is affixed. For $u_i \in U$, the membership degree is arranged by the decreasing order of the elements in the sequence, where $\mu_{\widetilde{M}_{(p)}}^1(u_i) \geq \mu_{\widetilde{M}_{(p)}}^2(u_i) \geq \dots \geq \mu_{\widetilde{M}_{(p)}}^j(u_i)$ and $\mu_{\widetilde{M}_{(Q)}}^1(u_i) \geq \mu_{\widetilde{M}_{(Q)}}^2(u_i) \geq \dots \geq \mu_{\widetilde{M}_{(Q)}}^j(u_i)$.

Example 7. A fuzzy multiset $\widetilde{M} = \{(0.4, 0.2)/a, (1, 0.7, 0.7)/b, (0.7, 0.2)/c, (0.4)/d\}$ over U ; the membership degree of element a is $\{0.2, 0.4\}$; the membership degree of element b is $\{1, 0.7, 0.7\}$; the membership degree of element c is $\{0.7, 0.2\}$; the membership degree of element d is $\{0.4\}$, where $L(a, \widetilde{M}) = 2$, $L(b, \widetilde{M}) = 3$, $L(c, \widetilde{M}) = 2$, and $L(d, \widetilde{M}) = 1$.

The set of all fuzzy multisets drawn from a universe U is denoted as $\mathfrak{M}(U)$.

Definition 8 (λ -cut set of fuzzy multiset). The λ -cut set of a fuzzy multiset \widetilde{M} is denoted as \widetilde{M}_λ and given by

$$\widetilde{M}_\lambda = \{u \mid u \in U \wedge \mu_{\widetilde{M}}(u) \geq \lambda\}. \quad (12)$$

Note that the difference between the notation $\widetilde{M}_{(\lambda)}$ and notation \widetilde{M}_λ is that \widetilde{M}_λ is preserved for the λ -cut set of the fuzzy multiset \widetilde{M} , while $\widetilde{M}_{(\lambda)}$ means assigning an index λ to the fuzzy multiset \widetilde{M} .

3. Objects and Object Merging

3.1. The General Framework. We have reviewed the most relevant definitions in the previous section. As we've mentioned earlier, the framework in our paper extends the work in paper [11], so now we will introduce some work basis below. The bases involve the redefinitions of coreferent objects and merge function in OMFM, and a brief review of properties of preservation and majority rule in [11].

The bases involve the redefinitions of coreferent objects and merge function in OMFM.

Reference function $\rho : U \rightarrow \varepsilon$ is formalized to describe a concept in the real world, where ε symbolizes the real world. By definition, two concepts are called coreferent if they describe the same real world concept.

Definition 9 (coreferent objects). Let U be a universe set of concepts. Two concepts u_1 and u_2 are coreferent if and only if $\rho(u_1) = \rho(u_2)$.

By the definition above, two objects that describe the same real world concept with semantic-related terms of different semantic relations embedded are formalized axiomatically. Here, we consider the context as the baseline: when describing a theme in a document, some semantic-related terms relating to this concept will be used to extend the theme.

Definition 10 (merge function). The merge function in OMFM is represented by function $\omega : \mathfrak{M}(U) \rightarrow U$.

Mapping the fuzzy multisets of objects onto a single object is the job of merge function in our work, and these functions are often idempotent; that is, $\omega(\widetilde{u}, \widetilde{u}, \dots, \widetilde{u}) = \widetilde{u}$, $\forall \widetilde{u} \in U$. This conclusion is also suitable in this paper and corresponding proof will be given in the following section.

A brief review of two important properties.

Property 1 (preservation). A merge function is preservative when merge function only selects one of the elements from the source set, the property of preservation in OMFM is denoted as

$$\forall \widetilde{M} \in \mathfrak{M}(U) : \omega(\widetilde{M}) \in \widetilde{M}. \quad (13)$$

Property 2 (majority rule). If the multiplicity value of an element is larger than the half of cardinality value of the source set then this element must be selected by the merge function, which is denoted as

$$\forall M \in \mathfrak{M}(U) : \left(\text{Count}_M(u) > \frac{|M|}{2}, \exists u \in U \right) \implies (\omega(M) = u). \quad (14)$$

The majority rule above is an important property for merge function in multiset that was further studied in [25] and a weaker version has been proved in [26]. By now, the majority rule is not extended deeply in fuzzy multiset as it does not apply in general, but the preservation rule will be elaborated in our paper.

3.2. Merging of Fuzzy Multisets. Within the scope of OMFM, we focus on the case of object merging of compound a multiset and multiple fuzzy multisets with the function of the type below:

$$\omega : \mathfrak{M}(\mathfrak{M}(U)) \longrightarrow \mathfrak{M}(U), \quad (15)$$

where the elements of $\mathfrak{M}(U)$ are denoted as $\widetilde{M}_{(p)}$, and the elements of $\mathfrak{M}(\mathfrak{M}(U))$ are $M_{(Q)}$. Here, the multiset $M_{(Q)}$ could be denoted as $M_{(Q)} = \{\widetilde{M}_{(p_1)}, \widetilde{M}_{(p_2)}, \dots, \widetilde{M}_{(p_n)}\}$, where $\text{Count}_{M_{(Q)}}(\widetilde{M}_{(p_i)})$ denotes the multiplicity of $\widetilde{M}_{(p_i)}$ in $M_{(Q)}$.

The fundamental operator is mapping the data segments of fuzzy multisets onto one object, which is called a solution. In following sections the symbol φ is used to represent a random solution of a given merge function; that is, $\omega(M) = \varphi$.

Example 11. Given a fuzzy multiset over $U = \{(0.1, 0.6)/a, (0.6, 1)/b, (0.5, 0.6, 0.8)/c, (0.5, 0.9)/d\}$ and the multiset $M_{(Q)} = \{\widetilde{M}_{(p_1)}, \widetilde{M}_{(p_1)}, \widetilde{M}_{(p_2)}, \widetilde{M}_{(p_3)}\}$, such that

$$\begin{aligned} \widetilde{M}_{(p_1)} &= \{(a, 0.1), (a, 0.6), (b, 1), (c, 0.8)\}, \\ \widetilde{M}_{(p_2)} &= \{(b, 0.1), (b, 0.6), (b, 1), (c, 0.5)\}, \\ \widetilde{M}_{(p_3)} &= \{(a, 0.1), (c, 0.6), (d, 0.9), (d, 0.5)\}. \end{aligned} \quad (16)$$

Here, consider two merge functions we have mentioned in Section 2, $\bar{\omega}_1 = \cup, \bar{\omega}_2 = \cap$. Then, we can obtain that

$$\begin{aligned}\bar{\omega}_1(M_{(Q)}) &= \bigcup_{\bar{M}_{(p_i)} \in \bar{M}_{(Q)}} (\bar{M}_{(p_i)}) = \varphi_{(1)} \\ &= \left\{ \frac{(0.6, 0.1)}{a}, \frac{(1, 0.6, 0.1)}{b}, \frac{(0.8)}{c}, \frac{(0.9, 0.5)}{d} \right\}, \\ \bar{\omega}_2(M_{(Q)}) &= \bigcap_{\bar{M}_{(p_i)} \in \bar{M}_{(Q)}} (\bar{M}_{(p_i)}) = \varphi_{(2)} = \left\{ \frac{0.5}{c} \right\}.\end{aligned}\quad (17)$$

Here, the fuzzy multiset $\bar{M}_{(p_i)}$ is referred to as source or source fuzzy multiset; the multiplicity of different source fuzzy multiset is not considered by merge function.

The case is $(\bar{\mathfrak{M}}(U), \subseteq)$ is not an upper bounded lattice. The normalization criterion that is needed when performing merge functions is usually omitted by fuzzy multiset theory. Therefore, we show another property below.

Property 3 (boundedness). A bounded merge function $\bar{\omega}$ over $\bar{\mathfrak{M}}(U)$ should satisfy the following constraint:

$$\begin{aligned}\forall M_{(Q)} \in \bar{\mathfrak{M}}(\bar{\mathfrak{M}}(U)) : \forall u \in U : \forall \bar{M}_{(p)} \in M_{(Q)} : \\ \mu_{\bar{\omega}(M_{(Q)})}(u) \leq \mu_{\bar{M}_{(p)}}(u).\end{aligned}\quad (18)$$

It indicates that the merge function selects one of the elements from all the source sets. A corresponding inference is that

$$\begin{aligned}\forall M_{(Q)} \in \bar{\mathfrak{M}}(\bar{\mathfrak{M}}(U)) : \forall u \in U : \\ (\forall \bar{M}_{(p)} \in M_{(Q)} : (\mu(u) \notin \bar{M}_{(p)})) \implies \mu(u) \notin \bar{\omega}(M_{(Q)}).\end{aligned}\quad (19)$$

This inference explains that any element not belonging to any source set should not exist in the outcomes of a bounded merge function. We could easily get this natural property just from the observation, because element u with membership degree $\mu(u)$ should be mixed into a solution arbitrarily. Also, it is a weaker notion of preservation. Besides, we also formulate the enforcing preservation:

$$\exists \bar{M}_{(p)} \in M_{(Q)} : \forall u \in \bar{\omega}(M_{(Q)}) : \mu_{\bar{\omega}(M_{(Q)})}(u) = \mu_{\bar{M}_{(p)}}(u).\quad (20)$$

Then, Property 3 is equivalent to indicating

$$\exists \bar{M}_{(p)} \in M_{(Q)} : \forall x \in \bar{\omega}(M_{(Q)}) : \mu_{\bar{\omega}(M_{(Q)})}(x) \neq 0.\quad (21)$$

Paper [11] has pointed out that keeping the weaker version of Property 2 in the situation of multiset is advantageous. They take multidocument summarization (*MDS*) as an example to explain that keeping a strict preservation would lead to a bad result in practical situations, that is, one of the documents itself would be the summary of the entire document

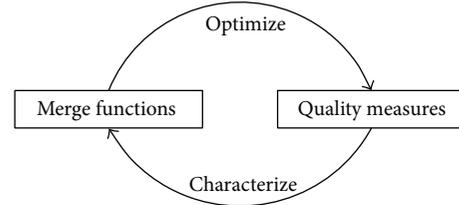


FIGURE 1: The relationship between the merger functions and quality measures.

set. While the task of document fusion (*DF*) is to generate a text containing all the information in entire document set. So, a weaker version of Property 2 is also advantageous in our framework. The bounded merge function of fuzzy multiset will be further elaborated in subsequent sections.

Theorem 12. *The functions \cup , \cap , and \oplus are bounded.*

Proof. We can get the proof from the case that for any $M_{(Q)} \in \bar{\mathfrak{M}}(\bar{\mathfrak{M}}(u))$

$$\begin{aligned}\min_{\bar{M}_{(p)} \in M_{(Q)}} \mu_{\bar{M}_{(p)}}(u) &\leq \max_{\bar{M}_{(p)} \in M_{(Q)}} \mu_{\bar{M}_{(p)}}(u) \leq \sum_{\bar{M}_{(p)} \in M_{(Q)}} \mu_{\bar{M}_{(p)}}(u) \\ &\leq \sum_{\bar{M}_{(p)} \in M_{(Q)}} \left(\text{Count}_{M_{(Q)}}(\bar{M}_{(p)}) \cdot \mu_{\bar{M}_{(p)}}(u) \right), \\ &\forall u \in U, \forall \bar{M}_{(p)} \in M_{(Q)}.\end{aligned}\quad (22)$$

□

4. Optimal Merging of Fuzzy Multisets

4.1. Quality Measures. The purpose of defining quality measures is to construct the merge functions that could get good performance for object merging in multiple fuzzy multisets. On one hand, the behavior of the merge functions we defined could be characterized by the value of the quality measures. On the other hand, adjusting a merge function could also optimize a balance between accurateness and completeness of a given solution to get a higher value of quality measures. The relationship between the merge functions and quality measures can be shown in Figure 1.

Within the scope of our paper, we adopted two quality measures widely used in the text mining literature: the first one is purity [13], and the second one is entropy [14, 15]. Information entropy is a concept used to measure the amount of information in the information theory, which is often taken as a measure of “disorder;” that is, the higher the value of entropy, the higher the extent of disorder; information purity is a measure of correlation between a system and its environment, where a higher value of purity means that a system is more relevant to its environment. Both of the two measures fall into range interval $[0, 1]$. Basically, the maximum purity and minimum entropy of results are the goals we try to achieve. Nevertheless, when we try to analyze the effect of a merge function, we should be able to analyze the effect

at fundamental level of the elements. So, some local quality measures will be introduced first.

Definition 13 (local precision). Given a multiset $M_{(Q)} = \{\widetilde{M}_{(p_1)}, \widetilde{M}_{(p_2)}, \dots, \widetilde{M}_{(p_n)}\}$, the local precision of the element u could be defined as

$$P^*: U \times \mu \longrightarrow [0, 1]: (u, \lambda) \longmapsto P^*(u, \lambda | M_{(Q)}), \quad (23)$$

such that

$$\begin{aligned} P^*(u, \lambda | M_{(Q)}) &= \frac{1}{|M_{(Q)}|} \cdot \sum_{(\widetilde{M}_{(p)} \in M_{(Q)}) \wedge (\mu_{\widetilde{M}_{(p)}}(u) = \lambda)} \text{Count}_{M_{(Q)}}(\widetilde{M}_{(p)}). \end{aligned} \quad (24)$$

The local precision judges the accurateness of adding the element u with the membership degree λ into the solution. Here, $M_{(Q)}$ is a multiset of sources. P^* judges the proportion of fuzzy multisets where the membership degree of element u is λ .

Example 14. Given the multiset $M_{(Q)}$ in Example 11. The local precision of adding an element c into a solution with membership degree 0.5 is calculated. When $(\widetilde{M}_{(p)} \in M_{(Q)}) \wedge (\mu_{\widetilde{M}_{(p)}}(u) = 0.5)$, $\text{Count}_{M_{(Q)}}(\widetilde{M}_{(p_1)}) = 0$, $\text{Count}_{M_{(Q)}}(\widetilde{M}_{(p_2)}) = 1$, $\text{Count}_{M_{(Q)}}(\widetilde{M}_{(p_3)}) = 0$, and $|M_{(Q)}| = 4$:

$$P^*(c, 0.5 | M_{(Q)}) = \frac{1}{4} \cdot 1 = 0.25. \quad (25)$$

Property 4 (monotony of P^*). Local precision P^* is a decreasing function in accordance with the membership degree threshold λ :

$$\begin{aligned} \forall u \in U: \forall (\lambda_1, \lambda_2) \in [0, 1]^2: \\ (\lambda_1 \geq \lambda_2) \implies P^*(u, \lambda_1 | M) \leq P^*(u, \lambda_2 | M). \end{aligned} \quad (26)$$

The monotony of P^* is a natural property. The lower membership degree means more sources will be added into the solution, owing to the fact that higher membership degree indicates relative simple relations related one concept (say the synonym of one word), and lower membership degree indicates more unspecific and more layered descriptions concerning one concept. As a result, we will get more complete information with higher precision.

Definition 15 (purity). Purity is computed using the maximal local precision value for each element in the solution φ as follows:

$$\text{Purity}: \widetilde{\mathcal{M}}(U) \longrightarrow [0, 1]: \varphi \longmapsto \text{Purity}(\varphi | M_{(Q)}), \quad (27)$$

such that

$$\begin{aligned} \text{Purity}(\varphi, M_{(Q)}) &= \sum_{(\widetilde{M}_{(p_j)} \in M_{(Q)}) \wedge (\mu_{\widetilde{M}_{(p_j)}}(u_k) = \lambda)} \\ &\times \left(|M_{(Q)}| \left(\sum_{\widetilde{M}_{(p_j)} \in M_{(Q)}} \sum_{u_i \in \widetilde{M}_{(p_j)}} L(u_i; \widetilde{M}_{(p_j)}) \right. \right. \\ &\quad \left. \left. \cdot \text{Count}_{M_{(Q)}}(\widetilde{M}_{(p_j)}) \right)^{-1} \right) \\ &\cdot P^*(u_k, \lambda | M_{(Q)}). \end{aligned} \quad (28)$$

Example 16. Given the multiset $M_{(Q)}$ of Example 11 and solution $\varphi = \{0.6/a, 1/b, 0.5/c\}$, we could obtain the local precision of all elements in this solution

$$\begin{aligned} P^*(a, 0.6 | M_{(Q)}) &= \frac{1}{4} \cdot 2 = 0.5, \\ P^*(b, 1 | M_{(Q)}) &= \frac{1}{4} \cdot 3 = 0.75, \\ P^*(c, 0.5 | M_{(Q)}) &= \frac{1}{4} \cdot 1 = 0.25. \end{aligned} \quad (29)$$

Note that we can also set λ of each element in solution a different value.

Then, we get the purity

$$\text{Purity}(\varphi | M_{(Q)}) = 0.375. \quad (30)$$

Definition 17 (local entropy). The local entropy of each fuzzy multiset in $M_{(Q)}$ is calculated as

$$E^*: U \times \mu \longrightarrow [0, 1]: (u, \lambda) \longmapsto E^*(u, \lambda | M_{(Q)}), \quad (31)$$

such that

$$E^*(u, \lambda | M_{(Q)}) = -P^*(u, \lambda | M_{(Q)}) \cdot \log P^*(u, \lambda | M_{(Q)}). \quad (32)$$

Property 5 (monotony of E^*). Local precision E^* is an increasing function in accordance with the degree of membership threshold λ when $0.5 \leq P^* \leq 1$, a decreasing function in accordance with the degree of membership threshold λ when $0 \leq P^* \leq 0.5$:

$$\begin{aligned} \forall u \in U: \forall (\lambda_1, \lambda_2) \in [0, 1]^2: 0.5 \leq \forall (P_1^*, P_2^*) \leq 1: \\ (\lambda_1 \geq \lambda_2) \implies E^*(u, \lambda_1 | M) \geq E^*(u, \lambda_2 | M), \\ \forall u \in U: \forall (\lambda_1, \lambda_2) \in [0, 1]^2: 0 \leq \forall (P_1^*, P_2^*) \leq 0.5: \\ (\lambda_1 \geq \lambda_2) \implies E^*(u, \lambda_1 | M) \leq E^*(u, \lambda_2 | M). \end{aligned} \quad (33)$$

Property 5 implies that the variation trend of local entropy is impacted by both fuzziness and proportion of an

element x in a solution; that is, neither excessively detailed or excessively brief information, nor more sources or less sources contained in the solution is appropriate to enrich the information of a fusion system. The proofs of these natural properties are omitted here. Back to our approach, the important connection exists between local precision and local entropy is also reflected by this property.

Definition 18 (total entropy). The total entropy of $M_{(Q)}$ is calculated as

$$\text{Entropy: } \overline{\mathfrak{M}}(U) \longrightarrow [0, 1] : \varphi \longmapsto \text{Entropy}(\varphi | M_{(Q)}), \quad (34)$$

such that

$$\begin{aligned} & \text{Entropy}(\varphi | M_{(Q)}) \\ &= \sum_{(\overline{M}_{(p_j)} \in M_{(Q)}) \wedge (\mu_{\overline{M}_{(p_j)}}(u_k) = \lambda)} \\ & \times \left(|M_{(Q)}| \left(\sum_{\overline{M}_{(p_j)} \in M_{(Q)}} \sum_{u_i \in \overline{M}_{(p_j)}} L(u_i; \overline{M}_{(p_j)}) \right. \right. \\ & \quad \left. \left. \cdot \text{Count}_{M_{(Q)}}(\overline{M}_{(p_j)}) \right)^{-1} \right) \\ & \cdot E^*(u_k, \lambda | M_{(Q)}). \end{aligned} \quad (35)$$

Example 19. Given the multiset $M_{(Q)}$ of Example 11 and $\varphi = \{0.6/a, 1/b, 0.5/c\}$, we obtain the local precision of all elements in this solution

$$\begin{aligned} E^*(a, 0.6 | M_{(Q)}) &= 0.3466, \\ E^*(b, 1 | M_{(Q)}) &= 0.2158, \\ E^*(c, 0.5 | M_{(Q)}) &= 0.3466. \end{aligned} \quad (36)$$

Then, we get the total entropy

$$\text{Entropy}(\varphi | M_{(Q)}) = 0.2273. \quad (37)$$

The purity and entropy can, respectively, express the quality measures, but the variation scales between them may be unequal. As mentioned above, the maximum purity and minimum entropy of results are the goals we try to achieve. Therefore, we try to investigate an index with the similar variation scales.

Definition 20 (validation index value). Given a multiset of sources $M_{(Q)} = \{\overline{M}_{(p_1)}, \overline{M}_{(p_2)}, \dots, \overline{M}_{(p_n)}\}$, the VI-value of the validation index is calculated as

$$\text{VI: } \overline{\mathfrak{M}}(U) \longrightarrow [0, 1] : \varphi \longmapsto \text{VI}(\varphi | M_{(Q)}), \quad (38)$$

such that

$$\text{VI}(\varphi | M_{(Q)}) = \text{Purity}(\varphi | M_{(Q)}) - \frac{\text{Entropy}(\varphi | M_{(Q)})}{C}. \quad (39)$$

Next, the rationality of this index will be shown. Generally, a brilliant result is generated by the higher value of the purity and the lower value of the entropy. That is to say, if the discrepancy between these two values is large, the value of the validation index is large and a good result can be determined by this validation index. That is to say, a balance between purity and entropy is expressed by validation index. In the case where the variation scales of these two values are similar, we propose a constant value C which could change the similar variation scales of purity value and entropy value. In practice, we determine the most significant C singular values by selecting the best VI, and it is kind of an empirical value which could be achieved during the simulation and modified through iterated procedure. But how to determine the value of this constant is not the problem we really care about now, we have not discussed this problem deeply in this paper. In our future work, we will explore this problem deeply with experimental analysis.

Note that for any solution φ , $\text{VI}(\varphi | M) \neq 0$ if and only if the local precisions of all elements in this solution differ from zero.

4.2. Optimization of Quality. The effect of a merge function can be judged by quality measures introduced in previous phase. And then we try to investigate the solutions optimizing the values of the quality measures. This type of optimization problem also appears in other research fields, paper [27] utilized the transitive closure as the effective mechanism transforming a matrix into fuzzy equivalence relation, by this way, finding the approximate partitions of data sequences. It is a classic example in the field of fuzzy set theory. Another example involved searching approximate minimum-distance by transforming a fuzzy reciprocal relation with a transitive reciprocal relation [28]. That is to say, the optimization mechanism could not be one of a kind. At the next step, we will concentrate on maximum quality generated from VI-value (the maximization of the purity and the minimization of the entropy). The difficulty of this step is to find the solution φ which gets the best VI-value. Therefore, the main task here is to define and investigate a suitable merge function.

Definition 21 (VI-value merge function). A VI-value merge function ω over $\overline{\mathfrak{M}}(U)$ should satisfy the following constraint:

$$\forall M \in \overline{\mathfrak{M}}(\overline{\mathfrak{M}}(U)) : \omega(M) = \arg \max_{\varphi \in \overline{\mathfrak{M}}(U)} \text{VI}(\varphi | M) \quad (40)$$

constrained by

$$\left(\max_{\varphi \in \overline{\mathfrak{M}}(U)} \text{VI}(\varphi | M) = 0 \right) \implies \omega(M) = \phi. \quad (41)$$

At this step, some properties of VI-optimal merge function will be studied further. A notable point is that there may appear several solutions sharing one maximum VI-value. With the definition of the merge function, how to select the unique solution is an important task here. Therefore, a selection criterion that selects one solution from the optimal solutions set is needed when performing these merge functions. With the special application area of OMFM, we will

show the details in illustrative examples. Another problem is a solution φ that has $\text{VI}(\varphi \mid M) \neq 0$ does not always exist. Hence, the notion of invalid solution is given below.

Definition 22 (invalid solution). Assume a VI-optimal merge function $\bar{\omega}$ and a fuzzy multiset of sources $M \in \mathfrak{M}(\mathfrak{M}(U))$.

A multiset $\varphi \in \mathfrak{M}(U)$ is defined as an invalid solution of $\bar{\omega}(M)$ if

$$\bar{\omega}(M) = \varphi \wedge \text{VI}(\varphi \mid M) = 0. \quad (42)$$

A solution φ of a VI-optimal merge function $\bar{\omega}$ that is not invalid is called a valid solution. Notice the differences between invalid solution and valid solution. Then, we will introduce another significant theorem.

Theorem 23. Any solution that is a real subset of the source intersection or a real superset of the source union has that

$$\begin{aligned} \forall M_{(Q)} \in \mathfrak{M}(\mathfrak{M}(U)) : \forall \varphi \in \mathfrak{M}(U) : \\ \left(\varphi \supset \bigcup_{\bar{M}_{(p)} \in M_{(Q)}} \bar{M}_{(p)} \right) \vee \left(\varphi \subset \bigcap_{\bar{M}_{(p)} \in M_{(Q)}} \bar{M}_{(p)} \right) \\ \implies (\text{VI}(\varphi \mid M_{(Q)}) = 0). \end{aligned} \quad (43)$$

Proof. Assume a fuzzy multiset of source $M_{(Q)} = \{\bar{M}_{(p_1)}, \bar{M}_{(p_2)}, \dots, \bar{M}_{(p_n)}\}$.

(1) A solution φ that satisfies

$$\varphi \supset \bigcup_{\bar{M}_{(p)} \in M_{(Q)}} \bar{M}_{(p)}. \quad (44)$$

Also it satisfies

$$\text{Purity}(\varphi \mid M_{(Q)}) = 0. \quad (45)$$

Owing to the case of all the elements of the solution φ would generate a local precision equivalent to 0, then

$$\text{VI}(\varphi \mid M_{(Q)}) = 0. \quad (46)$$

(2) A solution φ that satisfies

$$\varphi \subset \bigcap_{\bar{M}_{(p)} \in M_{(Q)}} \bar{M}_{(p)} \quad (47)$$

also satisfies

$$\text{Purity}(\varphi \mid M_{(Q)}) = 0. \quad (48)$$

Owing to the case of all the elements of the solution φ would generate a local precision equivalent to 0, then

$$\text{VI}(\varphi \mid M_{(Q)}) = 0. \quad (49)$$

□

The conclusion here is that a valid solution of VI-optimal merge function should include the intersection of the sources and should be included by the union of the source. In view of this point, we define the intersection of the sources as the lower bound and the union of the sources as the upper bound. The formalized definition is shown as

$$\begin{aligned} \bar{\varphi} &= \bigcup_{\bar{M}_{(p)} \in M_{(Q)}} \bar{M}_{(p)}, \\ \underline{\varphi} &= \bigcap_{\bar{M}_{(p)} \in M_{(Q)}} \bar{M}_{(p)}, \end{aligned} \quad (50)$$

where the lower bound is denoted as $\underline{\varphi}$ and the upper bound is denoted as $\bar{\varphi}$. Hence, we shall only consider solution φ that satisfies $\underline{\varphi} \subseteq \varphi \subseteq \bar{\varphi}$ in the following section.

Theorem 24. An VI-optimal merge function is idempotent.

Proof. Assume the fuzzy multiset $M_{(Q)} = \{\bar{M}_{(p_1)}, \bar{M}_{(p_2)}, \dots, \bar{M}_{(p_n)}\}$ satisfies $\underline{\varphi} = \bar{\varphi}$.

Thus, for $\varphi = \underline{\varphi} = \bar{\varphi}$, we have that

$$P^*(u, \lambda \mid M_{(Q)}) = \text{Purity}(\varphi \mid M_{(Q)}) = \text{VI}(\varphi \mid M_{(Q)}) = 1. \quad (51)$$

The corresponding proof is also shown when applying the previous theorem. □

Theorem 25. A VI-optimal merge function $\bar{\omega}$ is bounded.

Proof. With Theorems 12 and 23, we could get this conclusion. □

An important point is that VI-optimal merge functions do not satisfy the property of preservation. Nevertheless, due to the theorem we just proved above, they are bounded undoubtedly and boundedness offering a weaker version of preservation is shown in previous section. Besides the theorem of boundedness, several interesting theorems relevant to VI-optimal merge function need to be mentioned here. One of them is the theorem of VI-optimality invariance when scaling of multiplicity of the sources below.

Theorem 26. Assume a fuzzy multiset $M_{(Q)} = \{\bar{M}_{(p_1)}, \bar{M}_{(p_2)}, \dots, \bar{M}_{(p_n)}\}$ and a merge function $\bar{\omega}$. A conclusion could be got that

$$\bar{\omega}(M_{(Q)}) = \bar{\omega}(k \cdot M_{(Q)}), \quad \forall k \in \mathbb{N}_{>0}. \quad (52)$$

Proof. Several facts could be got that

$$\begin{aligned} P^*(u, \lambda \mid k \cdot M_{(Q)}) \\ = \frac{1}{|k \cdot M_{(Q)}|} \cdot \sum_{(\bar{M}_{(p)} \in k \cdot M_{(Q)}) \wedge (\mu_p(u) = \lambda)} k \cdot \text{Count}_{M_{(Q)}}(\bar{M}_{(p)}), \end{aligned}$$

$$\forall u \in \bar{\varphi},$$

$$\begin{aligned}
& P^*(u, \lambda | k \cdot M_{(Q)}) \\
&= \frac{1}{k \cdot |M_{(Q)}|} \cdot \sum_{(\bar{M}_{(p)} \in M_{(Q)}) \wedge (\mu_p(u)=\lambda)} k \cdot \text{Count}_{M_{(Q)}}(\bar{M}_{(p)}), \\
& \quad \forall u \in \bar{\varphi}, \\
& P^*(u, \lambda | k \cdot M_{(Q)}) \\
&= \frac{1}{|M_{(Q)}|} \cdot \sum_{(\bar{M}_{(p)} \in M_{(Q)}) \wedge (\mu_p(u)=\lambda)} \text{Count}_{M_{(Q)}}(\bar{M}_{(p)}), \\
& \quad \forall u \in \bar{\varphi}, \\
& P^*(u, \lambda | k \cdot M_{(Q)}) = P^*(u, \lambda | M_{(Q)}), \quad \forall u \in \bar{\varphi}.
\end{aligned} \tag{53}$$

And on the other hand,

$$\begin{aligned}
& E^*(u, \lambda | k \cdot M_{(Q)}) \\
&= -P^*(u, \lambda | k \cdot M_{(Q)}) \cdot \log P^*(u, \lambda | k \cdot M_{(Q)}), \\
& \quad \forall u \in \bar{\varphi}, \\
& E^*(u, \lambda | k \cdot M_{(Q)}) \\
&= -P^*(u, \lambda | M_{(Q)}) \cdot \log P^*(u, \lambda | M_{(Q)}), \quad \forall u \in \bar{\varphi}, \\
& E^*(u, \lambda | k \cdot M_{(Q)}) = E^*(u, \lambda | M_{(Q)}), \quad \forall u \in \bar{\varphi}.
\end{aligned} \tag{54}$$

□

Theorem 27. Assume a fuzzy multiset $M_{(Q)} = \{\bar{M}_{(p_1)}, \bar{M}_{(p_2)}, \dots, \bar{M}_{(p_n)}\}$ is a VI-optimal merge function $\bar{\omega}$ and a scaling parameter $k \in \mathbb{N}_{>1}$. If the solution φ is VI-optimal with regard to the sources $M_{(Q)}$, then φ is VI-optimal in regard to the sources $M_{(Q)} = \{k \cdot \bar{M}_{(p_1)}, \dots, k \cdot \bar{M}_{(p_n)}\}$.

Proof. We could get the corollary in last theorem. □

5. An Application: Document Fusion with Illustrative Example

5.1. Document Fusion. One possible application for this fuzzy multiset framework is document fusion. It involves the merging of elements with the different relations embedded. When it comes to document fusion, we have to introduce multidocument summarization briefly. Document fusion and multidocument summarization are two relevant areas. The important difference between these two areas is that, for multidocument summarization, the main task is to generate the shortest description containing the most relevant information, while for document fusion, the focus is to generate the shortest description containing all information contained in the whole document set excluding the redundancy [15, 16]. It is like that multidocument summarization is the intersection of the documents and document fusion is the union of the documents. Unlike multidocument summarization system, there is no

organization like DUC (Document Understand Conference) [29] providing “ideal” datasets for document fusion research yet, with which multiple documents under same subject and ideal summarization results for testing can be achieved. In addition, intrinsic and extrinsic evaluations in multidocument summarization system could not be suitable in fusion task: intrinsic evaluation where evaluation is done by human on accessing the quality of the fused documents itself makes the evaluation process subjective [30], and on the other hand, the difficulty in intrinsic evaluation of document fusion systems is that there is no existing collection of human written fusion results of multiple documents, serving as a gold standard for such evaluations by now; and extrinsic evaluation where the result of the document fused is evaluated by the completion of a specific task makes the evaluation process more complicated. Thus, there are no standard methods used to estimate the work in fusion task like in some document summarization tasks [31–33]. Given the problems we mentioned above, the evaluation that we performed is limited to date. To demonstrate our work, an example of an article cluster concerning the spoilage problem complaints of the dairy products on a particular brand has been selected from “315 consumption complaint” website to show the general fusion process and results by utilizing our framework. Although we use Chinese text for illustration, it is worth mentioning that there is not any fundamental difference between Chinese and English or other language under this framework.

The work of our paper is to propose a framework for document fusion, so we are not only aiming to get keywords, but for comprehensive information. Here, we just try to consider the situation of fuzzy multiset. With such extensions, the membership degree could be used to show the importance and fuzziness of an element, which makes the document representation more granular and semantically richer than multiset merging model in paper [11]. Assigning different weights to the same element also makes sense, when considering the situation that semantic-related terms with different semantic relations are used to identify the concept, which is semantically richer than just using words. Under our framework, semantic methods and statistical methods could be combined and used in many domains.

5.2. Illustrative Example. The main processing that needs to be performed is to get the Extra Strong, Strong, and Medium Strong relations of every concept in each article by using HowNet [34]. As a common-sense knowledge base, HowNet unveils interconceptual and interattribute relations of concepts. In HowNet, every concept of a word or phrase and its description form one entry with relations such as hypernym, hyponym, synonym, antonym, meronym, and Holonym (descriptions for these relations could be seen in Table 1), existing in HowNet and presented in DEF (concept definition) as shown in Box 1.

When performing English text, a large lexical database of English, WordNet, could be used to identify these relations instead of HowNet. Here, textual intention structure is determined by three relations of every concept. As an indicator in

TABLE 1: Descriptions for different semantic relations.

Relations	Descriptions
Hypernym	Also known as a superordinate, which is a word referring to broad categories or general concepts. For example, “musical instrument” is a hypernym of “guitar” because a guitar is a musical instrument.
Hyponym	A word or phrase whose semantic field is included within that of another word, a hyponym shares a type-of relationship with its hypernym, for example, “pigeon,” “crow,” “eagle,” and “seagull” are all hyponyms of “bird.”
Synonym	A word with the same or similar meaning of another word.
Antonym	A word which has the opposite meaning as another word.
Meronym	Meronym denotes a constituent part of or a member of something. For example, “finger” is a meronym of “hand” because a finger is part of a hand. Similarly, “wheels” is a meronym of “automobile.”
Holonym	Holonym defines the relationship between a term denoting the whole and a term denoting a part of or a member of the whole. For example, “tree” is a holonym of “bark,” of “trunk” and of “limb.”

taste:	DEF = attribute, taste, & edible
shape:	DEF = attribute, form, & physical
bearing:	DEF = attribute, behavior, & human

Box 1: Examples for concept definition.

linguistic segments, three relation segments of every concept tend to indicate the theme segments. That is to say, once three relations of every concept have been confirmed, the corresponding linguistic segments will have determinate tendency. Each concept is defined as an element in fuzzy multisets and the three relation segments that are included in each concept determine the different membership degree of each element as shown in Table 2.

In the following example, three semantic segments concerning different concepts x , y , and z will be used in merge process. The corresponding fuzzy multisets obtained from the source set are shown below:

$$\widetilde{M}_{(p_1)} = \{(x, 1.0), (x, 0.5), (x, 0.2), (y, 1.0), (y, 0.5), (y, 0.2), (z, 0.5), (z, 0.2)\},$$

$$\widetilde{M}_{(p_2)} = \{(x, 1.0), (x, 0.5), (x, 0.2), (y, 1.0), (y, 0.5), (y, 0.2), (z, 0.2)\},$$

$$\widetilde{M}_{(p_3)} = \{(y, 1.0), (y, 0.5), (y, 0.2), (z, 0.5), (z, 0.2)\},$$

TABLE 2: Three relations for constructing concepts.

	Composition of three relations	Membership degree
Extra Strong	Identity	1
Strong	① Synonym ② Hypernym, Hyponym, Antonym	0.5
Medium Strong	Meronym, Holonym	0.2

$$\widetilde{M}_{(p_4)} = \{(x, 0.5), (x, 0.2), (y, 1.0), (y, 0.5), (y, 0.2), (z, 0.5), (z, 0.2)\},$$

$$\widetilde{M}_{(p_5)} = \{(x, 0.5), (x, 0.2), (y, 0.5), (z, 0.5), (z, 0.2)\},$$

$$\widetilde{M}_{(p_6)} = \{(x, 1.0), (x, 0.2), (y, 1.0), (y, 0.5), (y, 0.2), (z, 0.2)\},$$

$$\widetilde{M}_{(p_7)} = \{(y, 1.0), (y, 0.2), (z, 1.0), (z, 0.5), (z, 0.2)\},$$

$$\widetilde{M}_{(p_8)} = \{(y, 1.0), (y, 0.5), (y, 0.2), (z, 1.0), (z, 0.2)\},$$

$$\widetilde{M}_{(p_9)} = \{(x, 0.5), (x, 0.2), (y, 1.0), (y, 0.5), (y, 0.2)\}, \quad (55)$$

where $(x, 1.0)$ denotes there are Extra Strong semantic relations identified in concept x .

As we've mentioned above, document fusion is to produce the shortest description containing all information found within the document set, but without repetition. The solution that we need is the solution concluding all the key concepts (x , y , and z in this example) and these concepts are constructed by three relations (Extra Strong, Strong, and Medium Strong relations). Let us consider the solution $\varphi = \{(x, 1), (y, 0.5), (z, 0.2)\}$. We get the local precision of all elements in this solution

$$P^*(x, \varphi(x) | M) = P^*(x, 1 | M) = 0.333,$$

$$P^*(y, \varphi(y) | M) = P^*(y, 0.5 | M) = 0.889, \quad (56)$$

$$P^*(z, \varphi(z) | M) = P^*(z, 0.2 | M) = 0.889,$$

such that

$$\text{Purity}(\varphi | M) = 0.352. \quad (57)$$

And we calculate the local entropy

$$E^*(x, \varphi(x) | M) = E^*(x, 1 | M) = 0.366,$$

$$E^*(y, \varphi(y) | M) = E^*(y, 0.5 | M) = 0.105, \quad (58)$$

$$E^*(z, \varphi(z) | M) = E^*(z, 0.2 | M) = 0.105,$$

such that

$$\text{Entropy}(\varphi | M) = 0.096. \quad (59)$$

TABLE 3: VI-values of solutions (concept with single relation embedded).

φ	$P^*(x, \varphi(x) M)$	$P^*(y, \varphi(y) M)$	$P^*(z, \varphi(z) M)$	Purity(φM)	$(C = 1.5)$ VI(φM)
	$E^*(x, \varphi(x) M)$	$E^*(y, \varphi(y) M)$	$E^*(z, \varphi(z) M)$	Entropy(φM)	
$\varphi_1 = \{0.2/x, 0.5/y, 0.5/z\}$	0.667	0.889	0.556	0.352	0.274
	0.270	0.105	0.326	0.117	
$\varphi_2 = \{0.2/x, 0.5/y, 0.2/z\}$	0.667	0.889	0.889	0.408	0.355
	0.270	0.105	0.105	0.080	
$\varphi_3 = \{0.2/x, 0.5/y, 1.0/z\}$	0.667	0.889	0.222	0.296	0.217
	0.270	0.105	0.334	0.118	
$\varphi_4 = \{0.5/x, 0.5/y, 0.5/z\}$	0.556	0.889	0.556	0.334	0.250
	0.326	0.105	0.326	0.126	
$\varphi_5 = \{0.5/x, 0.5/y, 0.2/z\}$	0.556	0.889	0.889	0.389	0.330
	0.326	0.105	0.105	0.089	
$\varphi_6 = \{0.5/x, 0.5/y, 1.0/z\}$	0.556	0.889	0.222	0.278	0.193
	0.326	0.105	0.334	0.128	
$\varphi_7 = \{1.0/x, 0.5/y, 0.5/z\}$	0.333	0.889	0.556	0.296	0.207
	0.366	0.105	0.326	0.133	
$\varphi_8 = \{1.0/x, 0.5/y, 0.2/z\}$	0.333	0.889	0.889	0.352	0.288
	0.366	0.105	0.105	0.096	
$\varphi_9 = \{1.0/x, 0.5/y, 1.0/z\}$	0.333	0.889	0.222	0.241	0.152
	0.366	0.105	0.334	0.134	

Then, we have got the validation index with setting the constant value $C = 1.5$:

$$VI(\varphi | M) = 0.352 - \frac{0.096}{C} = 0.352 - \frac{0.096}{1.5} = 0.288. \quad (60)$$

On the same principle, we present the local precision, purity, local entropy, entropy, and VI-value of the solutions with only one semantic relation embedded in each concept. If every concept is treated of equal importance with single relation embedded, in this case, we have got a maximal VI-value 0.355 (see Table 3).

As mentioned in former section, a valid solution φ of a VI-optimal merge function should follow the constraint below:

$$\begin{aligned} \overline{\varphi} &= \cup \widetilde{M} = \left\{ \frac{(1.0, 0.5, 0.2)}{x}, \frac{(1.0, 0.5, 0.2)}{y}, \frac{(1.0, 0.5, 0.2)}{z} \right\}, \\ \underline{\varphi} &= \cap \widetilde{M} = \left\{ \frac{0.5}{y} \right\}, \\ \underline{\varphi} &\subseteq \varphi \subseteq \overline{\varphi}. \end{aligned} \quad (61)$$

We should only consider solution φ that satisfies the constraint mentioned above in practical application. So, more complicated solutions could also be considered. For example, when concept y is an important concept which needs to be described explicitly in the fusion result, more details as Medium Strong relation in Table 2 should be contained to construct the description. If the concept is not treated of equal importance with single relation embedded, in this case, we've got two maximal VI-value of 0.491 (see Table 4). Generally,

for any solution φ , we can calculate the VI-values and choose the solution based on the observation of maximal VI-value. Within the scope of this paper, a tie-breaking criterion does not always exist, the accessorial choice criterion that helps selecting a solution from the set of optimal solutions is necessary. From the observations of φ_{10} and φ_{19} , we need to decide the merging function by actual requirement or finding a new solution with more semantic relations embedded.

More examples with considering concept z and concept x as important concepts which needs to be describe explicitly in the fusion result, will be shown in Tables 5 and 6. When concept z and concept x are with similar scale of multiple relations embedded, we could get equivalent maximum VI-value 0.410 from φ_{28} and φ_{38} , which means these two strategies get the same effect. If only one concept is with multiple relations embedded, we should consider φ_{10} to φ_{45} and get the maximum VI-value 0.491 from φ_{10} or φ_{19} . Both the situations needs to be further selected by considering the specific application: For φ_{28} and φ_{38} , two solutions selected strong and medium strong relations for concept z and x , so the importance of concept should be further considered, that is to say, if concept x is more importance for fusion, φ_{38} should be selected. For φ_{10} or φ_{19} , when the fusion results need to be described with more details, corresponding solution φ_{10} should be selected; otherwise, φ_{19} would also be a candidate selection for fusion.

Figure 2 shows the distribution of the VI-values with different scales of semantic relations embedded using box-and-whisker plot and evaluates the effectiveness of the merge function. The observations on partial of VI-values are presented here. The top bar stands for the maximum observation value; the bottom bar represents the lowest observation value. The bottom of the box is the lower quartile with 25% of

TABLE 4: VI-values of solutions (concept y with multiple relations embedded).

φ	$P^*(x, \varphi(x) M)$	$P^*(y, \varphi(y) M)$	$P^*(z, \varphi(z) M)$	Purity(φM)	$(C = 1.5)$ VI(φM)	
	$E^*(x, \varphi(x) M)$	$E^*(y, \varphi(y) M)$	$E^*(z, \varphi(z) M)$	Entropy(φM)		
$\varphi_{10} = \{0.2/x, (0.2, 0.5)/y, 0.2/z\}$	0.667 0.270	0.889 0.105	0.889 0.105	0.889 0.105	0.556 0.098	0.491
$\varphi_{11} = \{0.2/x, (0.2, 0.5)/y, 0.5/z\}$	0.667 0.270	0.889 0.105	0.889 0.105	0.556 0.326	0.500 0.134	0.411
$\varphi_{12} = \{0.2/x, (0.2, 0.5)/y, 1.0/z\}$	0.667 0.270	0.889 0.105	0.889 0.105	0.222 0.334	0.445 0.136	0.354
$\varphi_{13} = \{0.5/x, (0.2, 0.5)/y, 0.2/z\}$	0.556 0.326	0.889 0.105	0.889 0.105	0.889 0.105	0.537 0.107	0.466
$\varphi_{14} = \{0.5/x, (0.2, 0.5)/y, 0.5/z\}$	0.556 0.326	0.889 0.105	0.889 0.105	0.556 0.326	0.482 0.144	0.386
$\varphi_{15} = \{0.5/x, (0.2, 0.5)/y, 1.0/z\}$	0.556 0.326	0.889 0.105	0.889 0.105	0.222 0.334	0.426 0.145	0.329
$\varphi_{16} = \{1.0/x, (0.2, 0.5)/y, 0.2/z\}$	0.333 0.366	0.889 0.105	0.889 0.105	0.889 0.105	0.500 0.114	0.424
$\varphi_{17} = \{1.0/x, (0.2, 0.5)/y, 0.5/z\}$	0.333 0.366	0.889 0.105	0.889 0.105	0.556 0.326	0.445 0.150	0.345
$\varphi_{18} = \{1.0/x, (0.2, 0.5)/y, 1.0/z\}$	0.333 0.366	0.889 0.105	0.889 0.105	0.222 0.334	0.389 0.152	0.288
$\varphi_{19} = \{0.2/x, (1.0, 0.5)/y, 0.2/z\}$	0.667 0.270	0.889 0.105	0.889 0.105	0.889 0.105	0.556 0.098	0.491
$\varphi_{20} = \{0.2/x, (1.0, 0.5)/y, 0.5/z\}$	0.667 0.270	0.889 0.105	0.889 0.105	0.556 0.326	0.500 0.134	0.411
$\varphi_{21} = \{0.2/x, (1.0, 0.5)/y, 1.0/z\}$	0.667 0.270	0.889 0.105	0.889 0.105	0.222 0.334	0.445 0.136	0.354
$\varphi_{22} = \{0.5/x, (1.0, 0.5)/y, 0.2/z\}$	0.556 0.326	0.889 0.105	0.889 0.105	0.889 0.105	0.537 0.107	0.466
$\varphi_{23} = \{0.5/x, (1.0, 0.5)/y, 0.5/z\}$	0.556 0.326	0.889 0.105	0.889 0.105	0.556 0.326	0.482 0.144	0.386
$\varphi_{24} = \{0.5/x, (1.0, 0.5)/y, 1.0/z\}$	0.556 0.326	0.889 0.105	0.889 0.105	0.222 0.334	0.426 0.145	0.329
$\varphi_{25} = \{1.0/x, (1.0, 0.5)/y, 0.2/z\}$	0.333 0.366	0.889 0.105	0.889 0.105	0.889 0.105	0.500 0.114	0.424
$\varphi_{26} = \{1.0/x, (1.0, 0.5)/y, 0.5/z\}$	0.333 0.366	0.889 0.105	0.889 0.105	0.556 0.326	0.445 0.150	0.345
$\varphi_{27} = \{1.0/x, (1.0, 0.5)/y, 1.0/z\}$	0.333 0.366	0.889 0.105	0.889 0.105	0.222 0.334	0.389 0.152	0.288

data less than this value, the top of the box is the upper quartile with 25% of data greater than this value, and the red line in the box is the median observation value with 50% of data greater than this value. The x -axis presents the functions we examined (single relation embedded in single concept as shown in Table 3 multiple relations embedded in single concept as shown in Tables 4, 5, and 6, and the upper solution and lower solution). The y -axis presents the VI-values. By observing the range of VI-values, we see that, for all VI-values, there is a wide range of values that are achieved by different merge functions. Thus, simple strategies are not likely to work well. Observing the performance of the multiple relations embedded strategy, we see that

multiple relations embedded surpassed over only one relation embedded function, as it produces very good results on VI-values we examined. Thus, it is enough to optimize merge function using different scale of multiple relations embedded in concept, to obtain high quality results.

We have not performed all the merge functions with corresponding VI-values in our illustrative example yet. The data and explanation we give above is just to show the proposed framework vividly. Now, we've got the conclusion that we could get a corresponding VI-value for any solution. On the other hand, we could use the VI-value to select the best solution. As our framework is flexible enough to generate shortest description containing all information found within

TABLE 5: VI-values of solutions (concept z with multiple relations embedded).

φ	$P^*(x, \varphi(x) M)$		$P^*(y, \varphi(y) M)$		$P^*(z, \varphi(z) M)$		Purity(φM)	$(C = 1.5)$ VI(φM)
	$E^*(x, \varphi(x) M)$		$E^*(y, \varphi(y) M)$		$E^*(z, \varphi(z) M)$	Entropy(φM)		
$\varphi_{28} = \{0.2/x, 0.5/y, (0.2, 0.5)/z\}$	0.667	0.556	0.889	0.105	0.889	0.556	0.500	0.410
	0.270	0.326	0.105	0.105	0.105	0.326	0.134	
$\varphi_{29} = \{0.2/x, 0.5/y, (0.2, 1.0)/z\}$	0.667	0.556	0.889	0.105	0.889	0.222	0.445	0.355
	0.270	0.326	0.105	0.105	0.105	0.334	0.136	
$\varphi_{30} = \{0.2/x, 0.5/y, (0.5, 1.0)/z\}$	0.667	0.556	0.889	0.105	0.556	0.222	0.389	0.274
	0.270	0.326	0.105	0.105	0.326	0.334	0.173	
$\varphi_{31} = \{0.5/x, 0.5/y, (0.2, 0.5)/z\}$	0.556	0.556	0.889	0.105	0.889	0.556	0.482	0.386
	0.326	0.326	0.105	0.105	0.105	0.326	0.144	
$\varphi_{32} = \{0.5/x, 0.5/y, (0.2, 1.0)/z\}$	0.556	0.556	0.889	0.105	0.889	0.222	0.426	0.329
	0.326	0.326	0.105	0.105	0.105	0.334	0.145	
$\varphi_{33} = \{0.5/x, 0.5/y, (0.5, 1.0)/z\}$	0.556	0.556	0.889	0.105	0.556	0.222	0.371	0.250
	0.326	0.326	0.105	0.105	0.326	0.334	0.182	
$\varphi_{34} = \{1.0/x, 0.5/y, (0.2, 0.5)/z\}$	0.333	0.556	0.889	0.105	0.889	0.556	0.445	0.345
	0.366	0.326	0.105	0.105	0.105	0.326	0.150	
$\varphi_{35} = \{1.0/x, 0.5/y, (0.2, 1.0)/z\}$	0.333	0.556	0.889	0.105	0.889	0.222	0.389	0.288
	0.366	0.326	0.105	0.105	0.105	0.334	0.152	
$\varphi_{36} = \{1.0/x, 0.5/y, (0.5, 1.0)/z\}$	0.333	0.556	0.889	0.105	0.556	0.222	0.333	0.207
	0.366	0.326	0.105	0.105	0.326	0.334	0.189	

TABLE 6: VI-values of solutions (concept x with multiple relations embedded).

φ	$P^*(x, \varphi(x) M)$		$P^*(y, \varphi(y) M)$		$P^*(z, \varphi(z) M)$		Purity(φM)	$(C = 1.5)$ VI(φM)
	$E^*(x, \varphi(x) M)$		$E^*(y, \varphi(y) M)$		$E^*(z, \varphi(z) M)$	Entropy(φM)		
$\varphi_{37} = \{(0.2, 0.5)/x, 0.5/y, 0.5/z\}$	0.667	0.556	0.889	0.105	0.556	0.445	0.445	0.331
	0.270	0.326	0.105	0.105	0.326	0.171	0.171	
$\varphi_{38} = \{(0.2, 0.5)/x, 0.5/y, 0.2/z\}$	0.667	0.556	0.889	0.105	0.889	0.500	0.500	0.410
	0.270	0.326	0.105	0.105	0.105	0.134	0.134	
$\varphi_{39} = \{(0.2, 0.5)/x, 0.5/y, 1.0/z\}$	0.667	0.556	0.889	0.105	0.222	0.389	0.389	0.274
	0.270	0.326	0.105	0.105	0.334	0.173	0.173	
$\varphi_{40} = \{(0.2, 1.0)/x, 0.5/y, 0.5/z\}$	0.667	0.333	0.889	0.105	0.556	0.408	0.408	0.289
	0.270	0.366	0.105	0.105	0.326	0.178	0.178	
$\varphi_{41} = \{(0.2, 1.0)/x, 0.5/y, 0.2/z\}$	0.667	0.333	0.889	0.105	0.889	0.463	0.463	0.372
	0.270	0.366	0.105	0.105	0.105	0.136	0.136	
$\varphi_{42} = \{(0.2, 1.0)/x, 0.5/y, 1.0/z\}$	0.667	0.333	0.889	0.105	0.222	0.352	0.352	0.233
	0.270	0.366	0.105	0.105	0.334	0.179	0.179	
$\varphi_{43} = \{(1.0, 0.5)/x, 0.5/y, 0.5/z\}$	0.333	0.556	0.889	0.105	0.556	0.389	0.389	0.264
	0.366	0.326	0.105	0.105	0.326	0.187	0.187	
$\varphi_{44} = \{(1.0, 0.5)/x, 0.5/y, 0.2/z\}$	0.333	0.556	0.889	0.105	0.889	0.445	0.445	0.345
	0.366	0.326	0.105	0.105	0.105	0.150	0.150	
$\varphi_{45} = \{(1.0, 0.5)/x, 0.5/y, 1.0/z\}$	0.333	0.556	0.889	0.105	0.222	0.333	0.333	0.207
	0.366	0.326	0.105	0.105	0.334	0.189	0.189	

the document sets with different levels of details depending on practical requirement. With this framework, neither the relations between the keywords nor the context of co-text will be lost. To generate a moderately fluent semantic fusion result from a collection of documents, sentence planning and regeneration are then used to combine the segments together to form a coherent whole. In our paper, the framework solved basic issues on developing a document fusion system. (1) The documents are fused on more levels of granularity, as we assign different weights to different semantic relations

embedding in the same element of concept. (2) Meanwhile, taking into account semantic relations in the fusion progress ensures the readability of the fused document.

6. Conclusion and Future Work

We have presented a framework OMFM to map the fuzzy multisets of objects into one object. Our framework for merging multiple fuzzy multisets of documents is an interesting

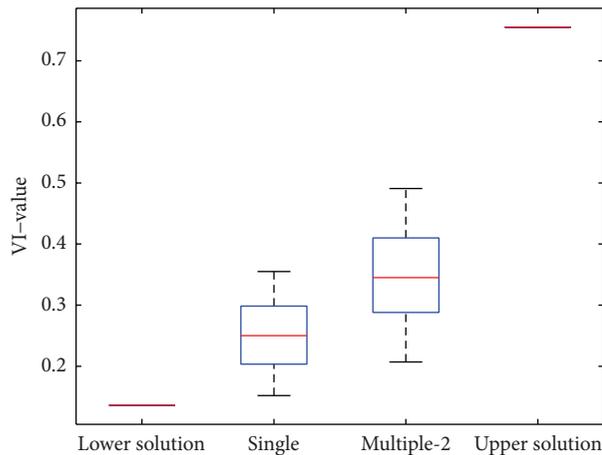


FIGURE 2: The distribution of VI-value with different scales of semantic relations embedded.

work, where a document set is modeled as a multiset of documents and each document is modeled as a fuzzy multiset of concepts. Also, OMF is an extension of the work in paper [11], which could describe both uncertainty and repeatability of the natural language by using the membership degree as the semantic fuzziness of the objects. The quality measures widely used in the text mining literature are defined to quantify the result of a merge function: purity (a measure of correctness) and entropy (a measure of completeness), where the maximum purity is got by upper solution, and the minimum entropy is got by lower solution. Then, we have constructed VI-optimal merge function to get the best solution, where both the higher purity and the lower entropy could be achieved simultaneously. Moreover, we have proved the properties related to constraints of merging problem. Finally, how to settle the problem in document fusion application using OMF has shown the practicality and effectiveness of our work.

With comparatively higher theoretical value and prospect of application, object merging problem will become a hot spot of research in many domains. Future work will further focus on experimental research and applying this framework to solving more relevant problems.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research has been supported by the National Natural Science Foundation of China (Grant no. 61472049), the National Natural Science Foundation for Young Scholars of China (Grant no. 61300148), and the Key Scientific and Technological Project of Jilin Province (Grant no. 20130206051GX). Besides, Dr. Lin Yue has been awarded a scholarship under the State Scholarship Fund to pursue her study at the University of Queensland as a joint Ph.D. Student, and this

work also has been awarded by China Scholarship Council (CSC). Finally, The authors really appreciate the anonymous reviewers for their constructive comments which have made substantial improvements to this paper.

References

- [1] D. Dubois and H. M. Prade, *Fundamentals of Fuzzy Sets*, Kluwer Academic, 2000.
- [2] R. R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 18, no. 1, pp. 183–190, 1988.
- [3] M. Mursaleen and A. K. Noman, "Applications of Hausdorff measure of noncompactness in the spaces of generalized means," *Mathematical Inequalities & Applications*, vol. 16, no. 1, pp. 207–220, 2013.
- [4] G. Beliakov and S. James, "On extending generalized Bonferroni means to Atanassov orthopairs in decision making contexts," *Fuzzy Sets and Systems*, vol. 211, pp. 84–98, 2013.
- [5] R. R. Yager and A. Rybalov, "Uninorm aggregation operators," *Fuzzy Sets and Systems*, vol. 80, no. 1, pp. 111–120, 1996.
- [6] J. Lin and A. O. Mendelzon, "Merging databases under constraints," *International Journal of Cooperative Information Systems*, vol. 7, no. 1, pp. 55–76, 1998.
- [7] S. Konieczny and R. P. Pérez, "Merging information under constraints: a logical framework," *Journal of Logic and Computation*, vol. 12, no. 5, pp. 773–808, 2002.
- [8] A. M. Pitts, "Nominal logic, at first order theory of names and binding," *Information and Computation*, vol. 186, no. 2, pp. 165–193, 2003.
- [9] S. Destercke, D. Dubois, and E. Chojnacki, "Possibilistic information fusion using maximal coherent subsets," *IEEE Transactions on Fuzzy Systems*, vol. 17, pp. 79–92, 2009.
- [10] K. McKeown, R. J. Passonneau, D. K. Elson, A. Nenkova, and J. Hirschberg, "Do summaries help?" in *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)*, pp. 210–217, ACM, August 2005.
- [11] A. Bronselaer, D. van Britsom, and G. de Tré, "A framework for multiset merging," *Fuzzy Sets and Systems*, vol. 191, pp. 1–20, 2012.
- [12] A. Bronselaer and G. D. Tré, "Aspects of object merging," in *Proceedings of the Annual North American Fuzzy Information Processing Society Conference (NAFIPS '10)*, IEEE, July 2010.
- [13] P. Pantel and D. Lin, "Document clustering with committees," in *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, 2002.
- [14] M. Steinbach, G. Karypis, and V. Kumar, "A comparison of document clustering techniques," in *Proceedings of the KDD Workshop on Text Mining*, Boston, Mass, USA, 2000.
- [15] V. Dodonov, "Purity- and entropy-bounded uncertainty relations for mixed quantum states," *Journal of Optics B: Quantum and Semiclassical Optics*, vol. 4, no. 3, pp. S98–S109, 2002.
- [16] C. Monz, "Document fusion for comprehensive event description," in *Proceedings of the Workshop on Human Language Technology and Knowledge Management*, Association for Computational Linguistics, 2001.
- [17] J. Azzopardi and C. Staff, "Fusion of news reports using surface-based methods," in *Proceedings of the 26th International Conference on Advanced Information Networking and Applications*

- Workshops (WAINA '12)*, pp. 809–814, Fukuoka, Japan, March 2012.
- [18] D. R. Radev, “A common theory of information fusion from multiple text sources step one: cross-document structure,” in *Proceedings of the 1st SIGdial Workshop on Discourse and Dialogue*, vol. 10, Association for Computational Linguistics, 2000.
- [19] L. A. Zadeh, “Fuzzy sets,” *Information and Computation*, vol. 8, pp. 338–353, 1965.
- [20] H.-J. Zimmermann, *Fuzzy Set Theory—and Its Applications*, Kluwer Academic, 4th edition, 2001.
- [21] S. Miyamoto, “Multisets and fuzzy multisets,” in *Soft Computing and Human-Centered Machines*, pp. 9–33, Springer, 2000.
- [22] J. Casasnovas and F. Rosselló, “Scalar and fuzzy cardinalities of crisp and fuzzy multisets,” *International Journal of Intelligent Systems*, vol. 24, no. 6, pp. 587–623, 2009.
- [23] T. Y. Nishida, “Multiset and K-subset transforming systems,” in *Proceedings of the Workshop on Multiset Processing*, pp. 203–217, 2000.
- [24] S. Miyamoto, “Fuzzy multisets and their generalizations,” in *Multiset Processing*, vol. 2235 of *Lecture Notes in Computer Science*, pp. 225–235, Springer, Berlin, Germany, 2001.
- [25] J. Lin and A. O. Mendelzon, “Knowledge base merging by majority,” in *Dynamic Worlds*, pp. 195–218, Springer, Berlin, Germany, 1999.
- [26] A. Bronselaer, G. De Tré, and D. Van Britsom, “Multiset merging: the majority rule,” in *Eurofuse 2011*, pp. 279–292, Springer, Berlin, Germany, 2012.
- [27] Y.-J. Wang, “A clustering method based on fuzzy equivalence relation for customer relationship management,” *Expert Systems with Applications*, vol. 37, no. 9, pp. 6421–6428, 2010.
- [28] S. Freson, H. de Meyer, and B. de Baets, “An algorithm for generating consistent and transitive approximations of reciprocal preference relations,” in *Computational Intelligence for Knowledge-Based Systems Design*, vol. 6178 of *Lecture Notes in Computer Science*, pp. 564–573, 2010.
- [29] <http://duc.nist.gov>.
- [30] R. Barzilay, K. R. McKeown, and M. Elhadad, “Information fusion in the context of multi-document summarization,” in *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, Association for Computational Linguistics, 1999.
- [31] Y. Chen, X. Lou, and J. Pan, “Research on multi-document summarization using lexical cohesion,” in *Proceedings of the International Conference on Web Information Systems and Mining (WISM '09)*, pp. 118–122, Shanghai, China, November 2009.
- [32] C. Dang and X. Luo, “WordNet-based document summarization,” in *Proceedings of the 7th WSEAS International Conference on Applied Computer & Applied Computational Science (ACACOS '08)*, 2008.
- [33] K. R. McKeown, J. L. Klavans, V. Hatzivassiloglou, R. Barzilay, and E. Eskin, “Towards multidocument summarization by reformulation: progress and prospects,” in *Proceedings of the 16th National Conference on Artificial Intelligence (AAAI '99)*, pp. 453–460, July 1999.
- [34] <http://www.keenage.com/>.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

