Generalized Bags, Bag Relations, and Applications to Data Analysis and Decision Making

Sadaaki Miyamoto

Department of Risk Engineering
Faculty of Systems and Information Engineering
University of Tsukuba, Ibaraki 305-8573, Japan
miyamoto@risk.tsukuba.ac.jp

Abstract. Bags alias multisets have long been studied in computer science, but recently more attention is paid on bags. In this paper we consider generalized bags which include real-valued bags, fuzzy bags, and a region-valued bags. Basic definitions as well as their properties are established; advanced operations such as s-norms, t-norms, and their duality are also studied. Moreover bag relations are discussed which has max-plus and max-min algebras as special cases. The reason why generalized bags are useful in applications is described. As two applications, bag-based data analysis and decision making based on convex function optimization related to bags are discussed.

Keywords: Generalized bag, s-norm, bag relation, data analysis, decision making, convex function.

1 Introduction

Bags which are also called multisets have long been studied by computer scientists as a basic data structure [6,10]. More recently, Calude and others [2] showed various aspects of multiset processing including a new paradigm of computation.

Since Yager [32] have proposed fuzzy bag, its theory and applications have been studied by several researchers [7,8,26,27,28,29,33] in the field of soft computing.

The author has redefined and re-established basic operations for fuzzy bags [14,15,16,17,20], and considered further generalizations [19]. Moreover we have applied fuzzy bags to data clustering [18] and text data analysis [21].

In this paper we overview bags and their generalizations with basic operations and their fundamental properties. Advanced operations such as s-norms and bag relations with new compositions are considered. We also discuss applications of generalized bags to data analysis, where classical methods as well as more recent techniques using kernel functions [30] are considered. Moreover a decision making aspect based on optimization of convex functions derived from set operations is considered, which are inspired from toll sets [4,13].

V. Torra, Y. Narukawa, and M. Inuiguchi (Eds.): MDAI 2009, LNAI 5861, pp. 37–54, 2009.

Although we show many propositions, we omit the proofs for the most part, as they are straightforward.

2 Bags and Generalized Bags

We begin with a review of classical bags and their generalizations.

2.1 Crisp Bags

Assume that the universal set $X = \{x_1, \ldots, x_n\}$ is finite for simplicity. A (crisp) bag M of $X = \{x_1, \ldots, x_n\}$ is characterized by a function $C_M(\cdot)$ (called count of M) whereby a natural number including zero corresponds to each $x \in X$ ($C_M : X \to \boldsymbol{N}$), where $\boldsymbol{N} = \{0, 1, 2, \ldots\}$. $C_M(\cdot)$ is called a *count function.*

We may express a crisp bag as

$$M = \{k_1/x_1, \ldots, k_n/x_n\}$$

or

$$M = \{\overbrace{x_1, \ldots, x_1}^{k_1}, \ldots, \overbrace{x_n, \ldots, x_n}^{k_n}\}.$$

In this way, an element of X may appear more than once in a bag.

Example 1. Consider an example in which $X = \{a, b, c, d\}$ and

$$C_M(a) = 2, \quad C_M(b) = 3, \quad C_M(c) = 1, \quad C_M(d) = 0.$$

In other words, $M = \{a, a, b, b, b, c\}$. This means that a, b, c, and d are included 2, 1, 3, and 0 times, respectively, in M. We can write $M = \{2/a, 3/b, 1/c\}$, ignoring an element of zero occurrence. Other expressions such as $M = \{3/b, 2/a, 1/c\}$ and $M = \{c, a, b, b, a, b\}$ are also used.

Basic relations and operations for crisp bags:

1. (inclusion): $M \subseteq N \Leftrightarrow C_M(x) \le C_N(x), \quad \forall x \in X.$
2. (equality): $M = N \Leftrightarrow C_M(x) = C_N(x), \quad \forall x \in X.$
3. (union): $C_{M \cup N}(x) = \max\{C_M(x), C_N(x)\}.$
4. (intersection): $C_{M \cap N}(x) = \min\{C_M(x), C_N(x)\}.$
5. (addition or sum): $C_{M+N}(x) = C_M(x) + C_N(x).$
6. (scalar multiplication): $C_{\alpha M} = \alpha C_M(x)$, where α is a nonnegative integer.
7. (Cartesian product): Let P is a bag of Y. $C_{M \times P}(x, y) = C_M(x) C_P(y).$

We use $\vee$ and $\wedge$ for max and min, respectively. Note that the relations and operations are similar to those for fuzzy sets. However, bags have the addition operation that fuzzy sets do not have, and the Cartesian product for bags is different from that for fuzzy sets.

2.2 *R*-Bags, *F*-Bags, and *G*-Bags

We discuss three generalizations. The first generalization to real-valued bags is simple, but useful in applications, and second is fuzzy bags, while the third is a minimum extension including the former two. We call them R-bags, F-bags, and G-bags for simplicity.

***R*-Bags.** The first generalization is straightforward. We assume a count function can take an arbitrary positive real value. Moreover the value of infinity should be included into the range of a count function, as we show its usefulness later. Thus, $C_M\colon X \to [0, +\infty]$ (note $[0, +\infty] = [0, \infty) \cup \{+\infty\}$). Since count function takes real values, we say real-valued bags, or shortly R-bags. The above definitions of basic relations and operations 1–7 are unchanged.

***F*-Bags.** Fuzzy bags are abbreviated as F-bags here. They were first studied by Yager [32], and basic relations and operations have been reconsidered by the authors [14,15].

In a fuzzy bag an element of X may occur more than once with possibly the same or different membership values.

Example 2. Consider a fuzzy bag

$$A = \{(a, 0.2), (a, 0.3), (b, 1), (b, 0.5), (b, 0.5)\}$$

of $X = \{a, b, c, d\}$, which means that a with the membership 0.2, a with 0.3, b with the membership 0.5, and two b's with 0.5 are contained in A.

We may write

$$A = \{\{0.2, 0.3\}/a, \{1, 0.5, 0.5\}/b\}$$

in which the bag of membership $\{0.2, 0.3\}$ corresponds to a and $\{1, 0.5, 0.5\}$ corresponds to b. Thus, $C_A(x)$ is a bag of the unit interval [32].

For an $x \in X$, the membership sequence is defined to be the decreasingly ordered sequence of elements in $C_A(x)$. It is denoted by

$$\mu_A^1(x), \mu_A^2(x), \ldots, \mu_A^p(x),$$

$(\mu_A^1(x) \geq \mu_A^2(x) \geq \cdots \geq \mu_A^p(x))$.

When we handle a finite number of fuzzy bags in a finite universal set, the length p of the membership sequences is set to be a constant for all members and for all the concerned fuzzy bags, by appending appropriate numbers of 0 at the end of the membership sequences.

Example 3. For the above example, we can set $p = 3$, $\mu_A^1(a) = 0.3$, $\mu_A^2(a) = 0.2$, $\mu_A^3(a) = 0$, $\mu_A^1(b) = 1$, $\mu_A^2(b) = \mu_A^3(b) = 0.5$, $\mu_A^1(c) = \mu_A^2(c) = \mu_A^3(c) = \mu_A^1(d) = \mu_A^2(d) = \mu_A^3(d) = 0$. By the representation of the membership sequence,

$$A = \{(0.3, 0.2)/a, (1, 0.5, 0.5)/b\},$$

or appending 0,

$$A = \{(0.3, 0.2, 0)/a, (1, 0.5, 0.5)/b, (0, 0, 0)/c, (0, 0, 0)/d\}.$$

The followings are basic relations and operations for fuzzy bags [14].

1. **inclusion:**
   $A \subseteq B \Leftrightarrow \mu^j_A(x) \le \mu^j_B(x),\ j = 1, \dots, p,\ \ \forall x \in X.$
2. **equality:**
   $A = B \Leftrightarrow \mu^j_A(x) = \mu^j_B(x),\ j = 1, \dots, p,\ \ \forall x \in X.$
3. **sum:**
   $A + B$ is defined by the sum operation in $X \times [0, 1]$ for crisp bags [32].
4. **union:**
   $\mu^j_{A \cup B}(x) = \mu^j_A(x) \vee \mu^j_B(x),\ j = 1, \dots, p,\ \ \forall x \in X.$
5. **intersection:**
   $\mu^j_{A \cap B}(x) = \mu^j_A(x) \wedge \mu^j_B(x),\ j = 1, \dots, p,\ \ \forall x \in X.$

***G*-Bags.** A further generalization of fuzzy bags is useful from theoretical viewpoint. It has been studied by the author [19] and is called G-bags here (this name is an abbreviation of generalized bags).

We introduce a G-bag using a closed region on a first quadrant $[0, +\infty]^2$ of a plane. The horizontal and vertical axes are called y-axis and z-axis, respectively. We define

$$C_A(x) = \nu_A(x) \tag{1}$$

where $\nu_A(x)$ is a closed region of $[0, +\infty]^2$ that satisfies the following conditions.

(I) For each $y \in [0, +\infty]$, the intersection between $\nu_A(x)$ and $\{y\} \times [0, +\infty]$ (the vertical line starting from y) is either empty or a segment starting from 0 and ending up to a point. We call this point $Z\nu_A(y; x)$. Thus,

$$\nu_A(x) \cap (\{y\} \times [0, +\infty]) = \{y\} \times [0, Z\nu_A(y; x)].$$

$Z\nu_A(y; x)$ as a function of y is monotonically non-increasing and

$$\lim_{y \to \infty} Z\nu_A(y; x) = 0.$$

(II) For each $z \in [0, +\infty]$, the intersection between $\nu_A(x)$ and $[0, +\infty] \times \{z\}$ (the horizontal line starting from z) is either empty or a segment starting from 0 and ending up to a point. We call this point $Y\nu_A(z; x)$. Thus,

$$\nu_A(x) \cap ([0, +\infty] \times \{z\}) = [0, Y\nu_A(z; x)] \times \{z\}.$$

$Y\nu_A(z; x)$ as a function of z is monotonically non-increasing and

$$\lim_{z \to \infty} Y\nu_A(z; x) = 0.$$

We illustrate an example of $\nu_A(x)$ in Figure 1. Note that when we are given either one of $Z\nu_A(y; x)$ or $Y\nu_A(z; x)$, $\nu_A(x)$ can uniquely be determined.

The basic relations and operations for two G-bags are defined as follows.

(I) (inclusion)

$$A \subseteq B \Leftrightarrow \nu_A(x) \subseteq \nu_B(x),\ \ \forall x \in X.$$

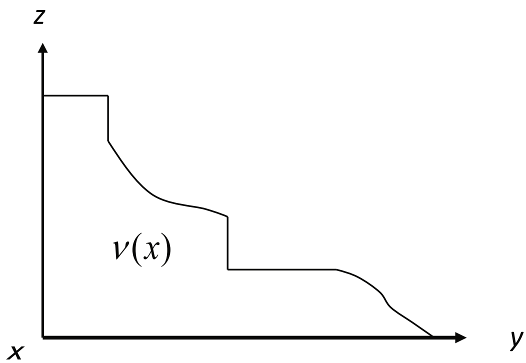

Fig. 1. Region $\nu(x)$ as a count function for G-bag

(II) (equality)

$$A = B \Leftrightarrow \nu_A(x) = \nu_B(x),\ \forall x \in X.$$

(III) (sum) Define

$$Y\nu_{A+B}(z;x) = Y\nu_A(z;x) + Y\nu_B(z;x)$$

and derive $\nu_{A+B}(x)$ from $Y\nu_{A+B}(z;x)$.

(IV) (union) Define

$$\nu_{A\cup B}(x) = \nu_A(x) \cup \nu_B(x).$$

(V) (intersection) Define

$$\nu_{A\cap B}(x) = \nu_A(x) \cap \nu_B(x).$$

(VI) (α-cut and ℓ-cut) Let a G-bag be A and $\alpha \in [0,1]$ and $\ell \in [0,+\infty)$ are given. An α-cut $[A]_\alpha$ is an R-bag with count function

$$C_{[A]_\alpha}(x) = Y\nu_A(\alpha;x).$$

On the other hand, a ℓ-cut $\langle A\rangle^\ell$ is a fuzzy set with membership

$$\mu_{\langle A\rangle^\ell}(x) = Z\nu_A(\ell;x)$$

A crisp bag, an R-bag, and a fuzzy bag can be regarded as a special case of G-bags by taking regions under the bars defined from count functions. If we have an R-bag with $C_M(x) = a$, then we define

$$\nu_A(x) = \{(y,z) : 0 \le y \le a,\ 0 \le z \le 1\} \tag{2}$$

and this R-bag is transformed into an equivalent G-bag. If we have a fuzzy bag, then we define

$$\nu_A(x) = \bigcup_{i=1}^{\infty} \{\ (y,z)\ :\ i-1 \le y \le i,\ 0 \le z \le \mu_A^i(x)\ \} \tag{3}$$

We have the next proposition.

Proposition 1. *Let A and B be arbitrary G-bags of X.*

$$[A+B]_\alpha = [A]_\alpha + [B]_\alpha, \tag{4}$$
$$[A \cup B]_\alpha = [A]_\alpha \cup [B]_\alpha, \tag{5}$$
$$[A \cap B]_\alpha = [A]_\alpha \cap [B]_\alpha, \tag{6}$$
$$\langle A \cup B\rangle^\ell = \langle A\rangle^\ell \cup \langle B\rangle^\ell, \tag{7}$$
$$\langle A \cap B\rangle^\ell = \langle A\rangle^\ell \cap \langle B\rangle^\ell. \tag{8}$$

The proof is straightforward and omitted.

Note 1. G-bags have a close relation to fuzzy interval-valued bags of which future studies are expected, but we omit the detail (see [19]).

2.3 Complement, *s*-Norm and *t*-Norm

This section is mainly concerned with R-bags. We state propositions without proofs. They are found in [22].

Complementation of *R*-Bags. A function $\mathcal{N}\colon [0,+\infty] \to [0,+\infty]$ with the next properties is used to define a complementation operation:

(i) $\mathcal{N}(0) = +\infty, \quad \mathcal{N}(+\infty) = 0.$
(ii) $\mathcal{N}(x)$ is strictly monotonically decreasing on $(0,+\infty)$.

A typical example is $\mathcal{N}(x) = const/x$ with $const > 0$.

An operation for the complement is then defined:

9.(complement):
$C_{\bar{M}}(x) = \mathcal{N}(C_M(x))$.

This operation justifies the generalization into R-bags, i.e., even when we start from crisp bags, the result of complementation is generally real-valued.

We immediately have the next two propositions; the proof is easy and omitted.

Proposition 2. *For arbitrary R-bags M, N, the next properties are valid:*

$$\overline{(\overline{M})} = M \tag{9}$$
$$\overline{M \cup N} = \bar{M} \cap \bar{N}, \qquad \overline{M \cap N} = \bar{M} \cup \bar{N}. \tag{10}$$

Proposition 3. *Let an empty bag $\emptyset$ and the maximum bag* **Infinity** *in R-bags be*

$$C_\emptyset(x) = 0, \quad \forall x \in X, \tag{11}$$
$$C_{\mathbf{Infinity}}(x) = +\infty, \quad \forall x \in X. \tag{12}$$

Then we have

$$\bar{\emptyset} = \mathbf{Infinity}, \quad \overline{\mathbf{Infinity}} = \emptyset. \tag{13}$$

***s*-Norms and *t*-Norms.** There have been studies on t-norms for crisp bags [11,3], but generalization into R-bags admits a broader class of s-norms and t-norms. For this purpose we introduce two functions $t(a,b)$ and $s(a,b)$ like those in fuzzy sets, but the definitions are different.

Definition 1. *Two functions* $t\colon [0,+\infty] \times [0,+\infty] \to [0,+\infty]$ *and* $s\colon [0,+\infty] \times [0,+\infty] \to [0,+\infty]$ *having the next properties (I)–(IV) are called a t-norm and an s-norm for R-bags, respectively. An s-norm is also called a t-conorm for bags.*

(I)[monotonicity] For $a \le c$, $b \le d$,

$$t(a,b) \le t(c,d),$$
$$s(a,b) \le s(c,d).$$

(II)[symmetry]

$$t(a,b) = t(c,d), \quad s(a,b) = s(b,a).$$

(III)[associativity]

$$t(t(a,b),c) = t(a,t(b,c)),$$
$$s(s(a,b),c) = s(a,s(b,c)).$$

(IV)[boundary condition]

$$t(0,0) = 0, \quad t(a,+\infty) = t(+\infty,a) = a,$$
$$s(+\infty,+\infty) = +\infty, \quad s(a,0) = s(0,a) = a.$$

A purpose to introduce such norms for bags is to generalize the intersection and union operations. First we note that $s(a,b) = a+b$, $s(a,b) = \max\{a,b\}$, and $t(a,b) = \min\{a,b\}$ satisfy the above conditions (I)–(IV). Thus the s-norms and t-norm represent the operations of addition, union, and intersection.

We moreover introduce a generating function $g(x)$ for s-norm.

Definition 2. *A function* $g\colon [0,+\infty] \to [0,+\infty]$ *is called a generating function for s-norm if it satisfies the next (i)–(iii):*

(i) it is strictly monotonically increasing,
(ii) $g(0) = 0, \quad g(+\infty) = +\infty$,
(iii) $g(x+y) \ge g(x) + g(y), \quad \forall x, y \in [0,+\infty]$.

We have the next proposition.

Proposition 4. *Let*

$$s(a,b) = g^{-1}(g(a) + g(b)). \tag{14}$$

Then $s(a,b)$ *is an s-norm.*

An example of the generation function is

$$g(x) = x^p \quad (p \ge 1). \tag{15}$$

Moreover, note the following.

Proposition 5. *Let $s(a,b)$ is an s-norm and $\mathcal{N}$ is a complementation operator. Then*

$$t(a,b) = \mathcal{N}(s(\mathcal{N}(a), \mathcal{N}(b))) \tag{16}$$

is a t-norm. Suppose $t(a,b)$ is a t-norm, then

$$s(a,b) = \mathcal{N}(t(\mathcal{N}(a), \mathcal{N}(b))) \tag{17}$$

is an s-norm.

If a pair of t-norm and s-norm has the above property stated in Proposition 5, we say (s,t) has the *duality* of norm and conorm. The duality has the next property.

Proposition 6. *Suppose $s_0(a,b)$ is an s-norm and $t_0(a,b)$ is derived from $s_0(a,b)$ by the operation (16). Let*

$$s(a,b) = \mathcal{N}(t_0(\mathcal{N}(a), \mathcal{N}(b)))$$

Then $s(a,b) = s_0(a,b)$. Suppose also that $t_0(a,b)$ is a t-norm and $s_0(a,b)$ is derived from $t_0(a,b)$ by the operation (16). Let

$$t(a,b) = \mathcal{N}(s_0(\mathcal{N}(a), \mathcal{N}(b)))$$

Then $t(a,b) = t_0(a,b)$.

We apply s-norm and t-norm to define bag operations $M\mathcal{S}N$ and $M\mathcal{T}N$, respectively.

$$C_{M\mathcal{S}N}(x) = s(C_M(x), C_N(x)). \tag{18}$$

$$C_{M\mathcal{T}N}(x) = t(C_M(x), C_N(x)). \tag{19}$$

Let us consider examples of s-norms and t-norms.

Example 4. The standard operators

$$s(a,b) = \max\{a,b\} \tag{20}$$

$$t(a,b) = \min\{a,b\} \tag{21}$$

are an s-norm and a t-norm, respectively. This pair has the duality stated in Propositions 5 and 6. Note, however, that s-norm (20) does not have a generating function that satisfies (14), while the next example uses the generating function.

Example 5. Let $g(x)$ be given by (15). Then we have

$$s(a,b) = (a^p + b^p)^{\frac{1}{p}}, \tag{22}$$

$$t(a,b) = (a^{-p} + b^{-p})^{-\frac{1}{p}}. \tag{23}$$

are an s-norm and a t-norm, respectively. This pair has the duality stated in Proposition 5 when $\mathcal{N} = const/x$ is used.

The second example has interesting properties. First, $s(a,b) = a + b$ is a particular case of (22) for $p = 1$. Moreover $s(a,b) = \max\{a,b\}$ and $t(a,b) = \min\{a,b\}$ are obtained from (22) and (23) when $p \to +\infty$.

Generalization to G-Bags. Apparently the complementation $\mathcal{N}$ cannot be generalized to G-bags. However, it is possible to define s-norms and t-norms as follows.

Definition 3. *Given two G-bags A, B of X, and an s-norm and t-norm, we define $Z\nu_{A\mathcal{S}B}(z;x)$ and $Z\nu_{A\mathcal{T}B}(z;x)$ are defined by*

$$Y\nu_{A\mathcal{S}B}(z;x) = s(Y\nu_A(z;x), Y\nu_B(z;x)), \tag{24}$$

$$Y\nu_{A\mathcal{T}B}(z;x) = t(Y\nu_A(z;x), Y\nu_B(z;x)), \tag{25}$$

Using $Y\nu_{A\mathcal{S}B}(z;x)$ and $Y\nu_{A\mathcal{T}B}(z;x)$, we generate $\nu_{A\mathcal{S}B}(x)$ and $\nu_{A\mathcal{T}B}(x)$.

The next proposition justifies the above definition.

Proposition 7.

$$[A\,\mathcal{S}\,B]_\alpha = [A]_\alpha\,\mathcal{S}\,[B]_\alpha, \tag{26}$$

$$[A\,\mathcal{T}\,B]_\alpha = [A]_\alpha\,\mathcal{T}\,[B]_\alpha \tag{27}$$

3 Bag Relations for Generalized Bags

A bag relation is a concept that corresponds to fuzzy relation. We define algebras for bag relations for R-bags, and then generalize them to G-bags. Proofs of the propositions in this section are shown in [22].

3.1 Max-s and Max-t Algebras

Let us introduce a new notation of $\boxplus$ and $\boxdot$ for

$$a \boxplus b = \max\{a, b\}, \quad a \boxdot b = s(a, b) \tag{28}$$

where $s(a, b)$ is an s-norm for R-bags. We call this *max-s algebra*.

It is easy to see that the following properties hold.

$$a \boxplus b = b \boxplus a, \tag{29}$$

$$a \boxplus (b \boxplus c) = (a \boxplus b) \boxplus c, \tag{30}$$

$$a \boxplus 0 = a, \tag{31}$$

$$a \boxdot b = b \boxdot a, \tag{32}$$

$$a \boxdot (b \boxdot c) = (a \boxdot b) \boxdot c, \tag{33}$$

$$a \boxdot 0 = a. \tag{34}$$

We moreover define $\boxplus$ and $\boxdot$ for

$$a \boxplus b = \max\{a, b\}, \quad a \boxdot b = t(a, b) \tag{35}$$

where $t(a, b)$ is a t-norm for R-bags. The latrter is called *max-t algebra*. We see that (29)–(33) hold, while (34) should be replaced by

$$a \boxdot +\infty = a. \tag{36}$$

We have the following.

Proposition 8. *Let a, b, c be real numbers. Then*

$$a \boxdot (b \boxplus c) = (a \boxdot b) \boxplus (a \boxdot c). \tag{37}$$

where $\boxdot$ is either an s-norm or a t-norm.

3.2 Bag Relations

Assume that all bags in this section are R-bags, unless otherwise stated.

Definition 4. *A bag relation R on $X \times Y$ is a bag R of $X \times Y$. The count function is denoted by $R(x, y)$ instead of $C_R(x, y)$ for simplicity.*

We define composition operation using max-s or max-t algebra.

Definition 5. *Let X, Y, Z be three universal sets. Assume R is a bag relation of $X \times Y$ and S is a bag relation of $Y \times Z$. Then a max-s composition $R \circ S$ is defined as follows.*

$$(R \circ S)(x, z) = \boxplus_{y \in Y}\{R(x, y) \boxdot S(y, z)\} \tag{38}$$

where $\boxdot$ is defined by an s-norm. Note that

$$\boxplus_{y \in \{a_1,\ldots,a_L\}} = a_1 \boxplus a_2 \boxplus \cdots \boxplus a_L.$$

A max-t composition is defined by the same equation (49) except that $\boxdot$ uses a t-norm.

Note also that the addition is straightforward

$$(R_1 \boxplus R_2)(x, y) = R_1(x, y) \boxplus R_2(x, y), \tag{39}$$

for bag relations on $X \times Y$.

We have the following.

Proposition 9. *The composition satisfies the associative property*

$$(R \circ S) \circ T = R \circ (S \circ T). \tag{40}$$

and the distributive property

$$(R_1 \boxplus R_2) \circ S = (R_1 \circ S) \boxplus (R_2 \circ S), \tag{41}$$

$$R \circ (S_1 \boxplus S_2) = (R \circ S_1) \boxplus (R \circ S_2). \tag{42}$$

In short, the composition is calculated like ordinary matrix calculations when the universes are finite.

We introduce the unit relations for the max-s and max-t compositions. For this purpose we define O_{XY} and Ω_{XY} on $X \times Y$.

$$O_{XY}(x, y) = 0, \quad \forall (x, y) \in X \times Y, \tag{43}$$

$$\Omega_{XY}(x, y) = +\infty, \quad \forall (x, y) \in X \times Y. \tag{44}$$

Frequently we omit the subscripts like O and Ω when we have no ambiguity. We then have

Proposition 10. *Assume that the max-s algebra is used. For arbitrary bag relation R on $X \times Y$,*

$$R \boxplus O = O \boxplus R = R, \tag{45}$$

$$R \circ O = O \circ R = R. \tag{46}$$

In contrast, assume that the max-t algebra is used. For arbitrary bag relation R on $X \times Y$,

$$R \boxplus O = O \boxplus R = R, \tag{47}$$

$$R \circ \Omega = \Omega \circ R = R. \tag{48}$$

Note 2. Max-s algebra is a generalization of max-plus algebra [5] and max-t algebra generalizes max-min algebra [12].

3.3 Relations of G-Bags

It is possible to generalize bag relations to G-bags. The idea is the same as that for s-norms of G-bags.

Definition 6. *Let X, Y, Z be three universal sets. Assume R is a G-bag relation of $X \times Y$ and S is a G-bag relation of $Y \times Z$. Then a max-s composition $R \circ S$ is defined as follows.*

$$Y\nu_{(R \circ S)}(w; x, z) = \boxplus_{y \in Y} \{Y\nu_R(w; x, y) \boxdot Y\nu_S(w; y, z)\} \tag{49}$$

where $\boxdot$ is defined by an s-norm.

A max-t composition is defined by the same equation (49) except that $\boxdot$ uses a t-norm. Using $Y\nu_{(R \circ S)}(w; x, z)$, we generate $\nu_{(R \circ S)}(x, z)$.

This definition is justified by the next proposition.

Proposition 11. *Let R is a G-bag relation of $X \times Y$ and S is a G-bag relation of $Y \times Z$. Assume $R \circ S$ is either max-s or max-t composition. We then have*

$$[R \circ S]_\alpha = [R]_\alpha \circ [S]_\alpha. \tag{50}$$

4 Data Analysis Based on Bag Models

We briefly overview bag-based models for data analysis. A typical bag model is used in document analysis, where frequency of a term in a document is regarded as a bag.

A less-known but useful model is fuzzy bags (F-bags), that is, weighted terms with many occurrences. Since discussion of F-bags include that of classical bags, we focus on F-bags.

4.1 Distance between *F*-Bags

An important point in data analysis is the measurement of a distance between two F-bags. We consider two distances. For this purpose we introduce additional symbols.

(i) Cardinal number of F-bag:

$$|A| = \sum_{x \in X} \sum_{j} \mu_A^j(x). \tag{51}$$

(ii) Product:

$$\mu_{A \cdot B}^j(x) = \mu_A^j(x) \mu_B^j(x), \quad j = 1, 2, \ldots \tag{52}$$

We now define two distances:

$$d_1(A, B) = |A \cup B| - |A \cap B|, \tag{53}$$

$$d_2(A, B) = |A \cdot A| + |B \cdot B| - 2|A \cdot B|. \tag{54}$$

The next proposition is useful.

Proposition 12.

$$d_1(A, B) = \sum_{x \in X} \sum_{j} |\mu_A^j(x) - \mu_B^j(x)|, \tag{55}$$

$$d_2(A, B) = \sum_{x \in X} \sum_{j} |\mu_A^j(x) - \mu_B^j(x)|^2. \tag{56}$$

Proof. We have

$$\begin{aligned} d_1(A, B) &= |A \cup B| - |A \cap B| \\ &= \sum_{x \in X} \sum_{j} \max\{\mu_A^j(x), \mu_B^j(x)\} - \sum_{x \in X} \sum_{j} \min\{\mu_A^j(x), \mu_B^j(x)\} \\ &= \sum_{x \in X} \sum_{j} |\mu_A^j(x) - \mu_B^j(x)|. \end{aligned}$$

$$\begin{aligned} d_2(A, B) &= |A \cdot A| + |B \cdot B| - 2|A \cdot B| \\ &= \sum_{x \in X} \sum_{j} \{(\mu_A^j(x))^2 + (\mu_B^j(x))^2\} - 2 \sum_{x \in X} \sum_{j} \mu_A^j(x) \mu_B^j(x) \\ &= \sum_{x \in X} \sum_{j} |\mu_A^j(x) - \mu_B^j(x)|^2. \end{aligned}$$

□

Miyamoto [18] applied these measures to fuzzy c-means clustering of documents and terms when terms have weights. The weighted terms with many occurrences were interpreted as fuzzy bags. It should be noted that cluster centers for both $d_1(A, B)$ and $d_2(A, B)$ are well-defined fuzzy bags and their calculations are not difficult [18].

Kernel Functions. Recently, kernel functions have been remarked by many researchers (e.g. [31,30]). It is possible to apply kernel functions to the set of F-bags. An effect of kernel functions is that nonlinear classification boundaries are easily obtained. The best-known kernel is the Gaussian kernel:

$$K(x, y) = \exp(-\lambda \|x - y\|^2)$$

where $\|x - y\|$ is the Euclidean distance between two points of the Euclidean space. When we use a kernel function, we assume an implicit mapping $\phi(\cdot)$ from a data space into an implicit high-dimensional space. Note that the high-dimensional space and mapping ϕ need not be known, but their inner product $\langle \phi(x), \phi(y) \rangle$ is given by an explicit kernel function:

$$K(x, y) = \langle \phi(x), \phi(y) \rangle.$$

For F-bags, it is not difficult to see that

$$K(A, B) = \exp(-\lambda d_2(A, B)) \tag{57}$$

is a positive-definite kernel, and hence we can use this kernel to data analysis of F-bags. On the other hand, $K(A, B) = \exp(-\lambda d_1(A, B))$ does not necessarily define a positive-definite kernel. Mizutani *et al.* [24] applied the Gaussian kernel to a set of documents and performed kernel fuzzy c-means clustering [23]. The results showed the kernel function better separates clusters than the ordinary fuzzy c-means clustering.

There is another point that kernel functions are useful. The original set of fuzzy bags is not a vector space, but after the mapping, F-bags are represented as points in a high-dimensional space. It is true that the high-dimensional space itself is invisible, but the method of kernel principal components [30] projects the points onto a low-dimensional subspace. Using such a method, we can visualize F-bags as points on a plane when two principal axes are used.

5 Application to Decision Making Using *R*-Bags

A classical work by Bellman and Zadeh [1] showed how fuzzy set framework is used in decision making, where an objective and a constraint are represented by fuzzy sets and a point that maximizes the membership of their intersection should be an optimal solution.

When we contrast bags and fuzzy sets, we should study decision making using bags instead of fuzzy sets, and consider if we have an essential difference between the two approaches.

Example 6. Let us review a simple example in the framework of fuzzy decision making. For simplicity we handle an objective and a constraint, but generalization to many objectives and constraints are straightforward.

An objective is represented by a fuzzy set G of X, while a *soft* constraint C is also a fuzzy set of the same universe. A larger $G(x)$ (we write $G(x)$ instead of $\mu_G(x)$) means that the objective is better satisfied. In the same way, a larger $C(x)$ means that the constraint is more satisfied. Hence we should consider maximization:

$$\max_{x \in X} (G \cap C)(x) \tag{58}$$

because both the objective and the constraint should be satisfied. That is, decision should be fuzzy set $D = G \cap C$.

If we use bags in the above formulation in just the same way, the result is the same as that by fuzzy sets. Only difference is that we do not have the ceiling of unity when handling bags.

We have, however, another formulation which is complementary to fuzzy decision making. It is more classical and yet employs a feature of fuzzy decision making. We describe an example using R-bags.

Example 7. Let G and C be R-bags of X, but they have different meanings:

- $G(x) = n$ means that n people are unsatisfied concerning the objective.
- $C(x) = m$ means that m people say the constraint is unsatisfied.
- We should minimize the number of unsatisfied people.

Since we have two bags G and C, minimization of

$$D(x) = G(x) + C(x) \tag{59}$$

is reasonable. Note that we write $D(x)$ instead of $C_D(x)$ for simplicity.

The above equation (59) means that total number of unsatisfied people is estimated to be $D(x) = G(x) + C(x)$ when decision variable is x. This means that there is no overlap between people unsatisfied to G and those unsatisfied to C.

In contrast, if we consider maximum overlap between those people unsatisfied to G or C, the decision is represented by

$$D(x) = \max\{G(x), C(x)\} = (G \cup C)(x) \tag{60}$$

These equations show a complementary formulation to that of fuzzy decision (58). If we should handle multiple goals $G_1, \ldots, G_m$ and constraints $C_1, \ldots, C_n$, we consider either

$$D(x) = \sum_{i=1}^{m} G_i(x) + \sum_{j=1}^{n} C_j(x), \tag{61}$$

or

$$\begin{aligned} D(x) &= \max_{1 \le i \le m, 1 \le j \le n} \{G_i(x), C_j(x)\} \\ &= (G_1 \cup \cdots \cup G_m \cup C_1 \cup \cdots \cup C_n)(x). \end{aligned} \tag{62}$$

Moreover we can use an s-norm of Minkowski type as a generalization:

$$D(x) = (G_1(x)^p \cup \cdots \cup G_m(x)^p \cup C_1(x)^p \cup \cdots \cup C_n(x)^p)^{\frac{1}{p}}, \quad (p \ge 1). \tag{63}$$

Note that (63) includes (61) when $p = 1$ and also approaches (62) as $p \to \infty$.

Note that in any case of (61), (62), and (63), we consider the minimization:

$$\min_{x \in X} D(x). \tag{64}$$

5.1 Convexity of Bags

In order to handle convex functions, we assume $X = \boldsymbol{R}^h$, the h-dimensional Euclidean space. Note that a convex function F defined on $X = \boldsymbol{R}^h$ means that

$$F(\lambda x + (1-\lambda)y) \le (\lambda F(x) + (1-\lambda)F(y)$$

for all $x, y \in X$ and all $\lambda \in [0,1]$. A necessary and sufficient condition for the convexity of $F(x)$ is that its epigraph

$$\mathrm{epi}(F) = \{(x,\beta) \in \boldsymbol{R}^{h+1} : F(x) \le \beta\}$$

is a convex set.

A drawback in fuzzy decision making is that we cannot use the theory of convex functions, i.e., even when we handle convex fuzzy sets, they are quasi-convex but never convex, since the membership of a fuzzy set is limited to $[0,1]$.

In contrast, we can assume convex R-bags G and C, since the membership value is in $[0,+\infty]$. It is easy to see the next properties are valid.

Proposition 13. *Assume R-bags G and C of $X = \boldsymbol{R}^h$ are convex. Then,*

$$D(x) = G(x) + C(x) \tag{65}$$

and

$$D'(x) = (G \cup C)(x) \tag{66}$$

are convex functions.

Proof. The convexity of $D(x)$ from (65) follows from the well-known fact that addition of two convex functions are also convex. The convexity of $D'(x)$ from (66) is based on the property that the intersection of two convex epigraphs is convex. □

We moreover have the next proposition.

Proposition 14. *Assume R-bags G and C of $X = \boldsymbol{R}^h$ are convex. Then,*

$$D''(x) = (G(x)^p + C(x)^p)^{\frac{1}{p}}, \quad (p \ge 1) \tag{67}$$

is convex.

Proof. We first note the Minkowski inequality [9]:

$$((a_1+b_1)^p + (a_2+b_2)^p)^{\frac{1}{p}} \le (a_1^p + a_2^p)^{\frac{1}{p}} + (b_1^p + b_2^p)^{\frac{1}{p}}$$

for $a_1, a_2, b_1, b_2 \ge 0$. noting that $G(x)$ and $C(x)$ are convex, we have

$$\begin{aligned} D''(\lambda x + (1-\lambda)y) &= [G(\lambda x + (1-\lambda)y)^p + C(\lambda x + (1-\lambda)y)^p]^{\frac{1}{p}} \\ &\le [(\lambda G(x) + (1-\lambda)G(y))^p + (\lambda C(x) + (1-\lambda)C(y))^p]^{\frac{1}{p}}. \end{aligned}$$

Using the Minkowski inequality, we have

$$\begin{aligned} D''(\lambda x + (1-\lambda)y) &\le [(\lambda G(x))^p + (\lambda C(x))^p]^{\frac{1}{p}} + [((1-\lambda)G(y))^p + ((1-\lambda)C(y))^p]^{\frac{1}{p}} \\ &= \lambda(G(x)^p + C(x)^p)^{\frac{1}{p}} + (1-\lambda)(G(y)^p + C(y)^p)^{\frac{1}{p}} \\ &= \lambda D''(x) + (1-\lambda)D''(y). \end{aligned}$$

Thus the convexity of $D''(x)$ is proved. □

It is straightforward to generalize the above propositions to decisions with multiple objectives and constraints using (61), (62), and (63). We omit the detail.

Thus if we use R-bags, we can handle convex decision functions.

6 Conclusion

We have overviewed generalizations of classical bags. Three types of generalizations have been studied. For R-bags, complementation, s-norm and t-norm, and bag relations have directly been defined, while they are more complicated for G-bags. Fuzzy bags (F-bags) can be handled as a special case of G-bags. Using s-norms and t-norms, we have defined max-s and max-t compositions for bag relations.

We have shown applications of F-bags to data analysis with discussion of kernel functions. It has been known that kernel-based methods of data analysis work well in many applications, and hence more studies are necessary concerning this topic.

Moreover it was shown that decision functions can be convex using s-norms in contrast to fuzzy decision making, where convexity property does not hold.

We have omitted many other applications, for example, application to rough sets [25] is also possible and we can develop rough bags and their generalizations (see, e.g., [19]).

Overall, bags have great potential to produce new useful tools in soft computing. There are many unsolved problems both in theory and applications. Many future researches are needed.

Acknowledgment

This study has partly been supported by the Grant-in-Aid for Scientific Research, Japan Society for the Promotion of Science, No.19650052 and No.19300074.

References

1. Bellman, R.E., Zadeh, L.A.: Decision-making in a fuzzy environment. Management Sci. 17, B-141–B164 (1970)
2. Calude, C.S., Păun, G., Rozenberg, G., Salomaa, A. (eds.): Multiset Processing. LNCS, vol. 2235. Springer, Heidelberg (2001)

3. Casasnovas, J., Mayor, G.: Discrete t-norms and operations on extended multisets. Fuzzy Ses and Systems 159, 1165–1177 (2008)
4. Dubois, D., Prade, H.: Toll sets. In: IFSA 1991 Brussels, Artificial Intelligence, pp. 21–24 (1991)
5. Heidergott, B., Olsder, G.J., van der Woude, J.: Max Plus at Work. Princeton University Press, Princeton (2006)
6. Knuth, D.E.: The Art of Computer Programming, Seminumerical Algorithms, vol. 2. Addison-Wesley, Reading (1969)
7. Li, B., Peizhang, W., Xihui, L.: Fuzzy bags with set-valued statistics. Comput. Math. Applic. 15, 811–818 (1988)
8. Li, B.: Fuzzy bags and applications. Fuzzy Sets and Systems 34, 61–71 (1990)
9. Lusternik, L.A., Sobolev, V.J.: Elements of Functional Analysis. Hindustan Publishing (1961)
10. Manna, Z., Waldinger, R.: The Logical Basis for Computer Programming, Deductive Reasoning, vol. 1. Addison-Wesley, Reading (1985)
11. Mayor, G., Torrens, J.: Triangular norms on discrete settings. In: Klement, E.P., Mesiar, R. (eds.) Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms, pp. 189–230. Elsevier, Amsterdam (2005)
12. Miyamoto, S.: Fuzzy Sets in Information Retrieval and Cluster Analysis. Kluwer, Dordrecht (1990)
13. Miyamoto, S.: Complements, t-norms, s-norms, and applications of toll sets. Journal of Japan Society for Fuzzy Theory and Systems 6(1), 130–140 (1994) (in Japanese)
14. Miyamoto, S.: Basic operations of fuzzy multisets. Journal of Japan Society for Fuzzy Theory and Systems 8(4), 639–645 (1996) (in Japanese)
15. Miyamoto, S.: Fuzzy multisets with infinite collections of memberships. In: Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA 1997), Prague, Czech, June 25-30, vol. 1, pp. 61–66 (1997)
16. Miyamoto, S., Kim, K.S.: An image of fuzzy multisets by one variable function and its application. J. Japan Society for Fuzzy Theory and Systems 10(1), 157–167 (1998) (in Japanese)
17. Miyamoto, S., Kim, K.S.: Multiset-valued images of fuzzy sets. In: Proceedings of the Third Asian Fuzzy Systems Symposium, Masan, Korea, June 18-21, pp. 543–548 (1998)
18. Miyamoto, S.: Information clustering based on fuzzy multisets. Information Processing and Management 39(2), 195–213 (2003)
19. Miyamoto, S.: Generalizations of multisets and rough approximations. International Journal of Intelligent Systems 19(7), 639–652 (2004)
20. Miyamoto, S.: Remarks on basics of fuzzy sets and fuzzy multisets. Fuzzy Sets and Systems 156(3), 427–431 (2005)
21. Miyamoto, S., Kawasaki, Y.: Text Analysis Based on Bag Neighborhood Model and Application to Incident Reports. In: Da Ruan, et al. (eds.) Computational Intelligence in Decision and Control, Proc. of the 8th Intern. FLINS Conf., Madrid, Spain, September 21-24, pp. 1015–1020 (2008)
22. Miyamoto, S.: Operations for Real-Valued Bags and Bag Relations. In: Proc. of IFSA (to appear, 2009)
23. Miyamoto, S., Ichihashi, H., Honda, K.: Algorithms for Fuzzy Clustering. Springer, Heidelberg (2008)
24. Mizutani, K., Inokuchi, R., Miyamoto, S.: Algorithms of nonlinear document clustering based on fuzzy multiset model. International Journal of Intelligent Systems 23(2), 176–198 (2008)

25. Pawlak, Z.: Rough Sets. Kluwer, Dordrecht (1991)
26. Ramer, A., Wang, C.-C.: Fuzzy multisets. In: Proc. of 1996 Asian Fuzzy Systems Symposium, Kenting, Taiwan, December 11-14, pp. 429–434 (1996)
27. Rebai, A.: Canonical fuzzy bags and bag fuzzy measures as a basis for MADM with mixed non cardinal data. European J. of Operational Res. 78, 34–48 (1994)
28. Rebai, A., Martel, J.-M.: A fuzzy bag approach to choosing the "best" multiattributed potential actions in a multiple judgment and non cardinal data context. Fuzzy Sets and Systems 87, 159–166 (1997)
29. Rocacher, D.: On fuzzy bags and their application to flexible querying. Fuzzy Sets and Systems 140(1), 93–110 (2003)
30. Schölkopf, B., Smola, A.J.: Learning with Kernels. The MIT Press, Cambridge (2002)
31. Vapnik, V.N.: Statistical Learning Theory. Wiley, New York (1998)
32. Yager, R.R.: On the theory of bags. Int. J. General Systems 13, 23–37 (1986)
33. Yager, R.R.: Cardinality of fuzzy sets via bags. Mathl. Modelling 9(6), 441–446 (1987)