MULTI-SOURCE MULTI-ATTRIBUTE DATA FUSION

Tatjana Petković and Risto Lahdelma

Department of Information Technology and Turku Centre for Computer Science,
University of Turku, Lemminkisenkatu 14 A, FIN-20520 Turku, FINLAND
tatpet@utu.fi, risto.lahdelma@cs.utu.fi

ABSTRACT

A framework for a multi-attribute multi-source data fusion is described. Uncertainty of data is modeled by fuzzy sets. The problem is considered on a general level. Credibility of sources, compatibility of data, as well as their reasonableness, are taken into consideration. The proposed method is illustrated by an example.

1. INTRODUCTION

The problem studied here can be described as follows: Assume that from sources $s_1, s_2, \ldots, s_m$ information on attributes $\mathrm{Att}_1, \mathrm{Att}_2, \ldots, \mathrm{Att}_n$ of an object from a set X is obtained. Not all sources give information on all attributes, and the information may be imprecise. In the extreme case, every source may give information only on one of the attributes. Besides that, we may be supplied with previous knowledge on the object, and this will be thought of as reasonableness. Proximity relations are used to model similarity between elements in attribute spaces. The problem is to find an element from a finite set of alternatives that fits the best with the information we have. As a tool to represent uncertainty of the information fuzzy sets are chosen. The work is based on the framework for multi-source data fusion described by Yager in [3]. However, Yager's model does not deal with multiple attributes and assumes that the information obtained from the sources is precise, i.e., these are single-element crisp sets.

Let us recall here some general terminology and fix notation. Whenever not relevant, arguments of functions will be denoted by $x, x_1, x_2, \ldots, y, y_1, y_2, \ldots$, or by A, $B, \ldots$ when it is important to emphasize that they are (fuzzy) sets. A function $f(x_1, \ldots, x_n)$, $n \in \mathbb{N}$, is *strictly monotone* if $x_j < y_j$ for some $j \in \{1, \ldots, n\}$ and $x_i = y_i$ for all $i \neq j$ implies that $f(x_1, \ldots, x_n) < f(y_1, \ldots, y_n)$, and it is *monotone* if $x_j \leq y_j$ for every j implies that $f(x_1, \ldots, x_n) \leq f(y_1, \ldots, y_n)$. The function f is *symmetric* in the i-th and j-th arguments if its value remains unchanged when these two arguments are swapped, and it is symmetric if this holds for any two arguments.

By $\mathbb{N}$ we mean the set of natural numbers, and I denotes the unit interval $[0, 1]$. It is rather easy to notice that this interval can be replaced by any complete lattice, but for the sake of simplicity we use the unit interval. A *fuzzy subset* A of a set X is associated with a *membership function* $\mu_A : X \to I$. Without danger of confusion, we usually write $A(x)$ instead of $\mu_A(x)$. The set of all fuzzy subsets of a set X is denoted by I^X since they are functions from X to I. The *intersection* of two fuzzy subsets A and B is the fuzzy set $A \cap B$ with the membership function $(A \cap B)(x) = A(x) \wedge B(x)$, where $\wedge$ stands for the $\min$ operator. The *height* of a fuzzy subset A of X is defined by $\mathrm{height}(A) = \max_{x \in X} A(x)$. Usually fuzzy sets in the paper are assumed to be *normalized*, i.e., they have height 1.

A *fuzzy binary relation* on X, i.e., a fuzzy subset of $X \times X$, is a *proximity relation* if it is reflexive and symmetric, but not necessarily transitive. It defines to what extent two values from X can be considered close or similar to each other. For example, we may assume two towns close if their distance is less than 100 km. This relation is reflexive and symmetric, but obviously not transitive. Recall here that a fuzzy relation $\rho \in I^{X \times X}$ is

– *reflexive* if $\rho(x, x) = 1$,
– *symmetric* if $\rho(x, y) = \rho(y, x)$,
– *transitive* if $\rho(x, z) \geq \rho(x, y) \wedge \rho(y, z)$,

for any $x, y, z \in X$.

Usually n will denote the number of attributes and m the number of sources.

2. DESCRIPTION OF THE FRAMEWORK

Let X be the universe of a variable a that represents a fused value. Each of the elements of the universe has attributes that are fuzzy subsets of X_j, $j \in \{1, \ldots, n\}$, $n \in \mathbb{N}$. The function $\pi_j : X \to I^{X_j}$ assigns to an element $a \in X$ the (fuzzy) value $\pi_j(a)$ of its j-th attribute. From sources $s_1, s_2, \ldots, s_m$, $m \in \mathbb{N}$, information on some of the attributes of a is obtained. The information on Att_j given by source s_i is a fuzzy subset of X_j denoted by A_{ij}, $1 \leq i \leq m, 1 \leq j \leq n$. Information received from sources is represented as a matrix with elements that are ordered pairs $[(c_{ij}, A_{ij})]_{m \times n}$, where the first component is from $I \setminus \{0\}$ being the credibility of the source s_i for the information on Att_j, and the second component is a fuzzy subset of X_j representing the information. As noted earlier, not all sources give information on all attributes, hence the matrix may have

The work has been supported by the Academy of Finland, decision numbers 208824 and 213905

	Att_1	Att_2	$\cdots$	Att_j	$\cdots$	Att_n
s_1	(c_{11}, A_{11})	(c_{12}, A_{12})		(c_{1j}, A_{1j})		(c_{1n}, A_{1n})
s_2	(c_{21}, A_{21})	(c_{22}, A_{22})		(c_{2j}, A_{2j})		(c_{2n}, A_{2n})
$\vdots$						
s_i	(c_{i1}, A_{i1})	(c_{i2}, A_{i2})		(c_{ij}, A_{ij})		(c_{in}, A_{in})
$\vdots$						
s_m	(c_{m1}, A_{m1})	(c_{m2}, A_{m2})		(c_{mj}, A_{mj})		(c_{mn}, A_{mn})

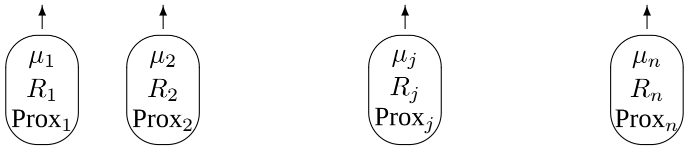

Figure 1

empty places. However, we assume, without loosing generality, that every source gives information on at least one of the attributes, and vice versa, for every attribute there is at least one source giving information. Besides that, proximity relations Prox_j on X_j, $1 \le j \le n$, are given. Any previous knowledge on some attributes of our object is considered as reasonableness, and it is represented as fuzzy subsets R_j of sets X_j, $1 \le j \le n$. In case of missing this kind of information, we may assume the entire X_j to be reasonable, that is $R_j = X_j$.

When dealing with fuzzy sets, it is often desirable to 'measure' the amount of information they contain. For example, when asking about somebody's age, we would rather hear an answer 'less than 22' than 'young'. For this reason, we are encouraged to involve a *measure of specificity* in the computation. This means that to every fuzzy subset A of X_j a value $\mu_j(A) \in I$ is assigned in such a way that the bigger the set $\{x \mid A(x) \neq 0\}$ the smaller the value $\mu_j(A)$ is. In extreme cases, $\mu_j(X_j) = 0$ and $\mu_j(\{x\}) = 1$. In order words, $\mu'(A) = 1 - \mu(A)$ is a *monotonic fuzzy measure* (see [4]).

The framework is illustrated in Figure 1. The computation can be roughly described as follows. The crucial part of the computation happens on the level of a cell in this scheme. Namely, when checking an element $a \in X$, support of its j-th attribute $\mathrm{Supp}_{ij}(\pi_j(a))$ by the information received from the i-th source is computed. However, this value depends on all values appearing in its row and influencing its column. The next step is aggregating the obtained values from the same column, i.e., calculating support $\mathrm{Supp}_j(\pi_j(a))$ of the j-th attribute of a by the information on this attribute received from all sources. Finally, the overall support $\mathrm{Supp}(a)$ for $a \in X$ is an aggregate of supports of its attributes in each column.

Elements from X will be compared on the basis of support their attributes get from the available information, i.e., for every $a \in X$ the value $\mathrm{Supp}(a) \in I$ is defined by

$$\mathrm{Supp}(a) = F(\mathrm{Supp}_1(\pi_1(a)), \ldots, \mathrm{Supp}_n(\pi_n(a))), \quad (1)$$

where $\mathrm{Supp}_j(\pi_j(a))$ measures how much the j-th attribute of a is supported by the information we have received from the sources. The function $F : I^n \to I$ is strictly monotone, symmetric, and satisfies $F(0, 0, \ldots, 0) = 0$ and $F(1, 1, \ldots, 1) = 1$. In order words, $\mathrm{Supp}(a)$ represents to what extent attributes of a match the information received from the sources and the previous knowledge.

Using reasonableness R_j, we can compute *reasonability* of a fuzzy subset A of X_j, $1 \le j \le n$, in notation $r_j(A)$, as the height of the intersection $A \cap R_j$, i.e.,

$$r_j(A) = \max_{x \in X_j}(A(x) \wedge R_j(x)). \quad (2)$$

Another constraint for usefulness of a fuzzy set A is its measure $\mu_j(A)$. Now the measure of *usefulness* of A, in notation $u_j(A)$, is defined by

$$u_j(A) = \mu_j(A) \wedge r_j(A). \quad (3)$$

Roughly speaking, $u_j(A)$ represents the amount of useful information contained in A. It is clear that $u_j(A) = \mu_j(A)$ in case we do not have any information on reasonability. In the other extreme case, when $A = X_j$, we get $u_j(A) = 0$. Let us note that $u_j(A) = 0$ if either A is too big or it does not contain reasonable information, i.e., $r_j(A) = 0$. The later will have a serious consequence on Supp.

Let us now consider the function $\mathrm{Supp}_j : I^{X_j} \to I$ defined by

$$\mathrm{Supp}_j(A) = f(\mathrm{Supp}_{1j}(A), \ldots, \mathrm{Supp}_{mj}(A)) \quad (4)$$

The function f is defined for all finite sequences of elements from I since the number of sources for different attributes may be different, i.e., $f : \cup_{m \in \mathbb{N}} I^m \to I$. This function is strictly monotone and symmetric since the order of the sources does not matter. It is also idempotent in the sense that $f(x, x, \ldots, x) = x$ since if all sources agree on support for an attribute, then the overall support should have the same value. Moreover, the condition

$f(x_1, \ldots, x_m, f(x_1, \ldots, x_m)) = f(x_1, \ldots, x_m)$ holds, meaning that if we happen to get a new piece of information from a new source on the same attribute and the support for the new information equals the support of the previously computed fused value, then the result should be the same. Therefore, this function is mean-like.

The function $\text{Supp}_{ij} : I^{X_j} \to I$ represents the support a fuzzy subset A of X_j gets as a value for Att_j according to the available information, in particular the one obtained from source s_i. The value $\text{Supp}_{ij}(A)$ of a fuzzy subset A of X_j depends on:

• *Compatibility* between fuzzy sets A and A_{ij} is denoted by $\text{Comp}_j(A, A_{ij})$. Clearly, it measures how much the information on the j-th attribute represented by set A is compatible with the information we have received from source s_i. Compatibility is a function of a given proximity relation Prox_j on X_j. The proximity relation can be extended to fuzzy subsets by

$$\text{Prox}_j(A, B) = \max_{x,y \in X_j} (A(x) \wedge \text{Prox}_j(x, y) \wedge B(y)). \quad (5)$$

If both sets A and B are crisp, this value becomes $\max_{x \in A, y \in B} \text{Prox}_j(x, y)$, and for two crisp values the extended proximity equals their proximity. For that reason we are allowed to use the same notation for both essentially different relations.

• *Credibility* $c_{ij} \in I$ of source s_i with respect to attribute Att_j is one of the inputs. It may originally be computed as a function $c(c^1_{ij}, c^2_{ij})$, where c^1_{ij} is the credibility that decision maker assigns to the source s_i and for the attribute Att_j, c^2_{ij} is the self-confidence of the source s_i on the information A_{ij}. However, it is possible that during processing, the value c_{ij} needs to be modified. First of all, usefulness of the information A_{ij} can modify credibility. Further, to every source we assign *reliability*, in notation $\text{rel}(i)$. This value depends on reasonableness and credibility of all information we have got from the source. In case the source always gives reasonable answers, i.e. $r_j(A_{ij}) \neq 0$, then $\text{rel}(i)$ does not influence the credibilities. In case we get unreasonable information from the source, still this should not make a big influence on the reliability of the source if the credibility for the information was not big, but in case it was bigger than a certain threshold, that should certainly decrease our overall confidence in the source, and all credibilities assigned to it. Moreover, we may even decide to ignore all the information we have got from the source, i.e., to remove it from our matrix. Therefore,

$$\text{rel}(i) = p(\text{dif}(c_{i1}, r_1(A_{i1})), \ldots, \text{dif}(c_{in}, r_n(A_{in}))) \quad (6)$$

where p is monotone, symmetric, $\text{rel}(i) \geq c_{ij}$ for each j in case all $r_j(A_{ij}) \neq 0$ and $p(x_1, \ldots, x_n) = 0 \Leftrightarrow x_j = 0$ for some j. The function $\text{dif} : I^2 \to I$ measures the importance of the difference between the values of credibility and reasonableness, and it satisfies the condition $\text{dif}(c, r) = 0 \Leftrightarrow r = 0$ and $c \geq \alpha$ for a given threshold $\alpha \in I$. Finally, the new credibility c'_{ij} is computed as

$$c'_{ij} = h(c_{ij}, u_j(A_{ij}), \text{rel}(i)) \quad (7)$$

It is monotone, min-like, symmetric in the first two arguments, and the last argument influences only in case it is 0. Thus, the original credibility is modified by the middle argument, which is concerned with the credibility of the information, and the last, which takes care of the source in general. Information with credibility 0 should not influence the final result, and hence if $c'_{ij} = 0$ then the information on the place ij is discarded from further consideration. In this way an unreliable source becomes removed from the scheme.

After this, $\text{Supp}_{ij}(A)$ can be computed using

$$\text{Supp}_{ij}(A) = u_j(A) \wedge g(c'_{ij}, \text{Comp}_j(A, A_{ij})), \quad (8)$$

where $g : I^2 \to I$ is monotone, and it satisfies conditions $g(1, 1) = 1$ and $g(x, y) = 0$ if and only if $x = 0$ or $y = 0$, that is, this is a min-like function. It can be viewed as local support of the source s_i to the value A for Att_j.

Let us now consider again function F from (1). If $r_j(\pi_j(a)) = 0$ for some j, i.e., if the j-th attribute of a does not match the previous knowledge on this attribute, then F should be defined so that $\text{Supp}(a) = 0$, i.e., for

$$r(a) = \wedge_{j=1}^n r_j(\pi_j(a)) \quad (9)$$

we have

$$r(a) = 0 \Rightarrow \text{Supp}(a) = 0.$$

On the other hand, even if $r(a) \neq 0$, still it is possible that $\pi_j(a)$ does not match the information on the j-th attribute we have got from all reliable sources. This means that for

$$P(a) = \wedge_{j=1}^n \vee_{i=1}^m \text{Comp}_j(\pi_j(a), A_{ij})$$

the following holds

$$P(a) = 0 \Rightarrow \text{Supp}(a) = 0.$$

This can be handled by correcting reasonabilities $r_j(\pi_j(a))$ in this way

$$r'_j(\pi_j(a)) = r_j(\pi_j(a)) \wedge \text{sgn}(r(a) \wedge P(a)),$$

i.e., new reasonabilities are either the same as before or all are 0 if one of the attributes of a is not reasonable or not compatible with the information obtained from sources. Continuing computation with these reasonabilities in case they are 0 implies that all $\text{Supp}_{ij}(\pi_j(a))$ are 0, and so $\text{Supp}(a) = 0$.

3. OUTPUTS

Let us assume that outputs belong to a final set of alternatives X. In case we are supplied with precise information on all attributes of all alternatives, we are able to order linearly elements from X with

$$a \leq b \Leftrightarrow \text{Supp}(a) \leq \text{Supp}(b).$$

The set of all elements with maximal $\text{Supp}(a)$ will form a complete solution to the problem in this case.

However, in the spirit of dealing with imprecise information, we cannot expect to have exact values for all attributes of all alternatives, some part of this information maybe completely missing. In this case we may compute the amount of available information about an alternative, that is

$$\mu(a) = \mu_1(\pi_1(a)) + \ldots + \mu_n(\pi_n(a)).$$

Now a partial order, i.e., a reflexive, antisymmetric and transitive relation, can be defined on X by

$$a \leq b \quad \Leftrightarrow \quad r(a) \wedge P(a) = 0 \text{ or } (\mu(a) = \mu(b) \text{ and } \text{Supp}(a) \leq \text{Supp}(b)). \tag{10}$$

This means that an element a with one of the attributes not reasonable or not compatible with the information obtained from sources, is smaller than any other, or two elements can be compared if there is equal amount of information on their attributes. This seems rather strict and is rarely the case. Hence, we may extend this relation by

$$a \preceq b \Leftrightarrow r(a) \wedge P(a) = 0 \text{ or } (|\mu(a) - \mu(b)| \leq \alpha \text{ and } \text{Supp}(a) \leq \text{Supp}(b))$$

for some $\alpha \in I$. On the basis of this relation elements of X can be compared, although it is not transitive.

Another way to compare elements is by computing their *relative acceptability*, in notation acc(a), by

$$\text{acc}(a) = \text{Supp}(a)/\mu(a). \tag{11}$$

This results in a total order defined by

$$a \lesssim b \Leftrightarrow \text{acc}(a) \leq \text{acc}(b).$$

This order extends the partial order $\leq$ defined in (10).

4. POSSIBLE CHOICES FOR FUNCTIONS

In the following table we give an overview of the functions appearing in the model, with short explanation and reference to equations where they are applied.

function	description	ref.
$F(x_1, \ldots, x_n)$	overall support	(1)
$f(x_1, \ldots, x_n)$	support for attribute Att$_j$	(4)
$p(x_1, \ldots, x_n)$	reliability of source s_i	(6)
dif(x, y)	difference between credibility and reasonability	(6)
$h(x, y, z)$	modified credibility	(7)
$g(x, y)$	support at the place ij	(8)
Comp$_j(A, B)$	compatibility of two sets	(5,8)
$r_j(A)$	j-reasonability of a fuzzy set	(2)
$\mu_j(A)$	j-measure of A	
Prox$_j(A, B)$	proximity between fuzzy sets	(5)
rel(i)	reliability of source s_i	(6)
$r(a)$	reasonableness of alternative	(9)
$r'_j(\pi_j(a))$	modified reasonableness	
$\pi_j(a)$	fuzzy value of Att$_j$ of a	

We have already discussed conditions that functions used in (1)–(8) must satisfy. The function F in (1) can be either product or average. Anyway, product would give Supp$(a) = 0$ even if a is a reasonable solution, but we do not have information on one of its attributes. This would not be in the spirit of dealing with imprecise data, hence our choice here is the average function.

Here we list functions that satisfy previously given conditions, and hence are possible choices for the functions.

$$\begin{aligned}
&F(x_1, \ldots, x_n) = 1/n\Sigma_{i=1}^n x_n \\
&f(x_1, \ldots, x_n) = 1/n\Sigma_{i=1}^n x_n \\
&p(x_1, \ldots, x_n) = \text{sgn}(x_1) \wedge \ldots \wedge \text{sgn}(x_n) \\
&\text{dif}(x, y) = \begin{cases} x, & y > 0 \\ 0, & y = 0 \text{ and } x \geq 0.5 \end{cases} \\
&h(x, y, z) = x \wedge y \wedge z \\
&g(x, y) = x \wedge y \\
&\text{Comp}_j(A, B) = \text{Prox}_j(A, B)
\end{aligned}$$

Using this, equations (7) and (8) become

$$\begin{aligned}
c'_{ij} &= c_{ij} \wedge u_j(A_{ij}) \wedge \text{rel}(i) \\
\text{Supp}_{ij}(A) &= u_j(A) \wedge \text{Prox}_j(A, A_{ij}) \wedge c'_{ij}.
\end{aligned}$$

This will be used in the following example. Let us only mention here that more on aggregation operators used in information fusion frameworks can be found in [2].

5. EXAMPLE

We are going to illustrate our method on a hypothetical example. Assume that we are looking for the driver of a car in which there were 4 persons. Three witnesses saw the driver and gave some information on his age, height, hair color and sex. Anyway, we know for certain the driver is young and has brown hair. Credibilities of sources vary as one of them is a child, the second was not in a good position to see the driver well, and the third has vision disorder. The knowledge matrix $[(c_{ij}, A_{ij})]_{1 \leq i \leq 3, 1 \leq j \leq 4}$ is given in Table 1.

The first problem we are faced with is representing linguistically given fuzzy sets in the form of functions. This is a very complex problem (see [1,5,6], for example), and we will accept here without much justification the presentation given in Figure 2 for the sets appearing in Table 1. In order to do this we need first to determine the underlying sets $X_j, 1 \leq j \leq 4$. Namely,

$$\begin{aligned}
&X_1 = \{1, 2, \ldots, 100\},\ X_2 = \{70, 71, \ldots, 220\}, \\
&X_3 = \{1, 2, \ldots, 100\},\ X_4 = \{\text{M,F}\}.
\end{aligned}$$

The set X_3 needs more explanation. Indeed, we can assume that there are 100 possible hair colors. The colors are distributed so that roughly the first 5 on the scale are considered gray, from 5 to 10 are between gray and blond, from 10 to 35 are strictly blond, from 35 to 50 are blond towards brown colors, from 50 to 70 are brown, from 70 to 85 brown towards dark, and the rest are strictly dark.

	age	height	hair	sex
s_1	(0.8, approx. 30)	(0.8,$\geq$ 150)	(0.8, brown)	(1,M)
s_2		(1,180)	(0.8,dark)	
s_3	(0.6,15-50)	(0.8,$\geq$ 170)	(0.2, gray)	(1,M)

Table 1

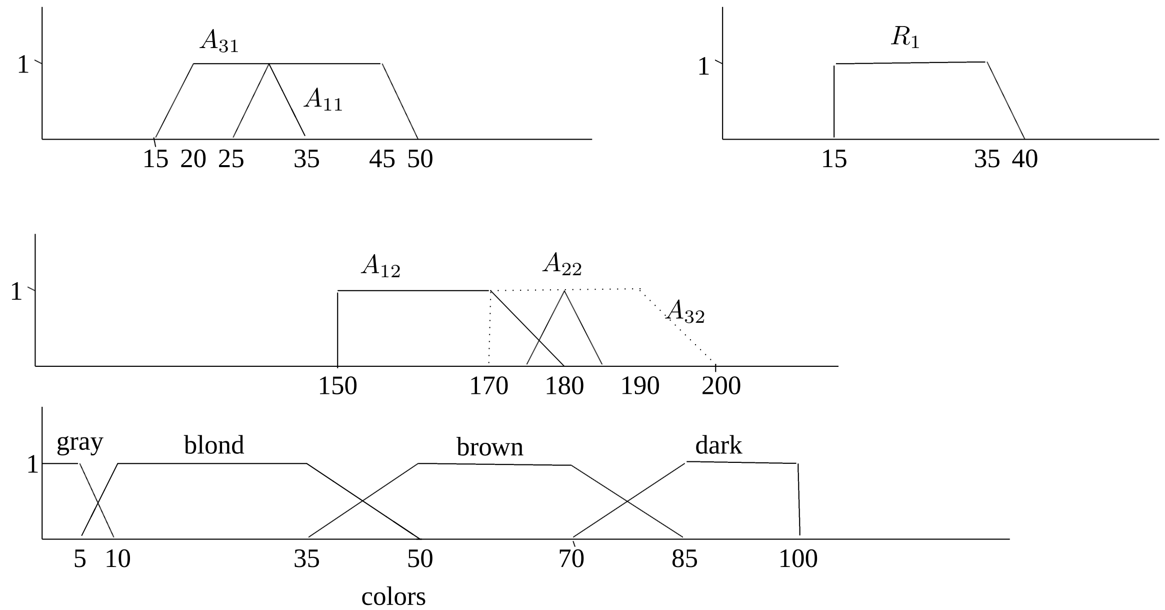

Figure 2

Proximity relations are given by

$$\text{Prox}_1(x,y) = \begin{cases} 1-|x-y|/5, & |x-y| \leq 5 \\ 0, & \text{otherwise} \end{cases}$$
$$\text{Prox}_2(x,y) = \begin{cases} 1-|x-y|/10, & |x-y| \leq 10 \\ 0, & \text{otherwise} \end{cases}$$
$$\begin{aligned} \text{Prox}_3(x,y) = & \ (\text{gray}(x) \wedge \text{gray}(y)) \vee \\ \vee & \ (\text{blond}(x) \wedge \text{blond}(y)) \vee \\ \vee & \ (\text{brown}(x) \wedge \text{brown}(y)) \vee \\ \vee & \ (\text{dark}(x) \wedge \text{dark}(y)) \end{aligned}$$
$$\text{Prox}_4(x,y) = \begin{cases} 0, & x \neq y \\ 1, & x = y \end{cases}$$

All the measures are defined in the same way

$$\mu_j(A) = 1 - \frac{\int_{X_j} A(x)dx}{\int_{X_j} dx}. \tag{12}$$

Before analyzing alternatives, new credibilities c'_{ij} are computed. It is easy to see that $r_1(A_{11}) = r_1(A_{31}) = r_3(A_{13}) = 1$, $r_3(A_{23}) = 0.4667$, $r_3(A_{33}) = 0$ and all the rest is 1 by default. Since $c_{33} = 0.2$ is rather small, then $r_3(A_{33}) = 0$ will only imply $c'_{33} = 0$, but still rel$(3) = 1$. Knowing this, we can discard the information on the place 33 from further consideration. Hence, all the sources have reliability 1. The measures of A_{ij} sets, according to (12), are:

$$\begin{array}{ll} \mu_1(A_{11}) = 0.95 & \mu_2(A_{12}) = 0.8333 \\ \mu_3(A_{13}) = 0.65 & \mu_4(A_{14}) = 1 \\ \mu_2(A_{22}) = 0.9667 & \mu_3(A_{23}) = 0.775 \\ \mu_1(A_{31}) = 0.65 & \mu_2(A_{32}) = 0.8333 \\ \mu_4(A_{34}) = 1 & \end{array}$$

Using (3) we get

$$\begin{array}{ll} u_1(A_{11}) = 0.95 & u_2(A_{12}) = 0.8333 \\ u_3(A_{13}) = 0.65 & u_4(A_{14}) = 1 \\ u_2(A_{22}) = 0.9667 & u_3(A_{23}) = 0.4667 \\ u_1(A_{31}) = 0.65 & u_2(A_{32}) = 0.8333 \\ u_4(A_{34}) = 1 & \end{array}$$

This gives new credibilities by (7)

$$\begin{array}{llll} c'_{11} = 0.8 & c'_{12} = 0.8 & c'_{13} = 0.65 & c'_{14} = 1 \\ & c'_{22} = 0.9667 & c'_{23} = 0.4667 & \\ c'_{31} = 0.6 & c'_{32} = 0.8 & c'_{33} = 0 & c'_{34} = 1 \end{array}$$

Credibilities in positions 13, 22 and 23 became lower for different reasons: the set A_{13} contains rather small information comparing c_{13}, the same can be said about A_{22} since we have considered 180 as fuzzy number and had very high credibility $c_{22} = 1$, and A_{33} has conflict with the fact that hair is brown.

We are proceeding by analyzing alternatives. The set X consists of 4 persons that were in the car, and here are their data

(a) John is 33, tall 185 cm, has brown hair;
(b) Mary is 30 and tall 170;
(c) Peter is more than 25 and has dark hair;
(d) Bob is 37, tall 180, has brown hair.

Each of the alternatives is analyzed separately.
(a) Clearly, all reasonabilities are 1, and hence u_j functions are equal to μ_j, where, by (12),

$$\mu_1(33) = 1 \qquad \mu_2(185) = 1$$
$$\mu_3(\text{brown}) = 0.65 \qquad \mu_4(\text{M}) = 1$$

and hence $\mu(\text{John}) = 3.65$. Further, $\text{Prox}_1(33, A_{11}) = 0.6$ and it is achieved for 32 since $\text{Prox}_1(33, 32) = 0.8$ and $A_{11}(32) = 0.6$. Similarly, using (5), the rest of proximity relations are calculated

$$\text{Prox}_1(33, A_{11}) = 0.6 \qquad \text{Prox}_1(33, A_{31}) = 1$$
$$\text{Prox}_2(185, A_{12}) = 0.2 \qquad \text{Prox}_2(185, A_{22}) = 0.6$$
$$\text{Prox}_2(185, A_{32}) = 1$$
$$\text{Prox}_3(\text{brown}, A_{13}) = 1 \qquad \text{Prox}_3(\text{brown}, A_{23}) = 0.4667$$
$$\text{Prox}_4(\text{M}, A_{14}) = 1 \qquad \text{Prox}_4(\text{M}, A_{34}) = 1$$

As $r(\text{John}) \wedge P(\text{John}) \neq 0$ we can continue using the same reasonabilities. This gives by (8)

$$\text{Supp}_{11}(33) = 0.6 \qquad \text{Supp}_{12}(185) = 0.2$$
$$\text{Supp}_{13}(\text{brown}) = 0.65 \qquad \text{Supp}_{14}(\text{M}) = 1$$
$$\text{Supp}_{22}(185) = 0.6 \qquad \text{Supp}_{23}(\text{brown}) = 0.4667$$
$$\text{Supp}_{31}(33) = 0.6 \qquad \text{Supp}_{32}(185) = 0.8$$
$$\text{Supp}_{34}(\text{M}) = 1$$

Computing Supp_j as average of Supp_{ij} for all i where there is relevant information gives

$$\text{Supp}_1(33) = 0.6 \qquad \text{Supp}_2(185) = 0.5333$$
$$\text{Supp}_3(\text{brown}) = 0.5584 \qquad \text{Supp}_4(\text{M}) = 1$$

and hence, by (1) and (11),

$$\text{Supp}(\text{John}) = 0.6729 \text{ and } \text{acc}(\text{John}) = 0.1844.$$

(b) When considering Mary, information on the sex of the driver given by the sources excludes Mary as the driver. Anyway, let us see how this can be obtained using our method. Namely, $\text{Prox}_4(\text{F}, A_{14}) = \text{Prox}_4(\text{F}, A_{34}) = 0$, and hence $P(\text{Mary}) = 0$. This further implies that $r'_j(\pi_j(\text{Mary})) = 0$, and so $\text{Supp}_j(\pi_j(\text{Mary})) = 0$, for every j, what finally gives $\text{Supp}(\text{Mary}) = 0$.
(c) Peter is 'more than 25'. This will be fuzzy set A given in Figure 3.

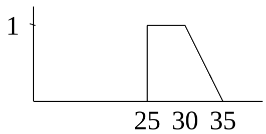

Figure 3

Measures are now

$$\mu_1(A) = 0.925 \qquad \mu_2(X_2) = 0$$
$$\mu_3(\text{dark}) = 0.775 \qquad \mu_4(\text{M}) = 1$$

and so $\mu(\text{Peter}) = 2.7$. Reasonabilities are

$$r_1(A) = 1 \qquad r_2(X_2) = 1$$
$$r_3(\text{dark}) = 0.4667 \qquad r_4(\text{M}) = 1$$

Thus,

$$u_1(A) = 0.925 \qquad u_2(X_2) = 0$$
$$u_3(\text{dark}) = 0.4667 \qquad u_4(\text{M}) = 1$$

From $u_2(X_2)$ we get $\text{Supp}_2(X_2) = 0$. Similarly to the case (a), $\text{Supp}_4(\text{M}) = 1$. Let us compute now $\text{Supp}_1(A)$ and $\text{Supp}_3(\text{dark})$. Proximities are

$$\text{Prox}_1(A, A_{11}) = 1 \qquad \text{Prox}_1(A, A_{31}) = 1$$
$$\text{Prox}_3(\text{dark}, A_{13}) = 0.4667 \qquad \text{Prox}_3(\text{dark}, A_{23}) = 1$$

As $r(\text{Peter}) \wedge P(\text{Peter}) \neq 0$ we can proceed using the same reasonabilities. This gives

$$\text{Supp}_{11}(A) = 0.8 \qquad \text{Supp}_{31}(A) = 0.6$$
$$\text{Supp}_{13}(\text{dark}) = 0.4667 \qquad \text{Supp}_{23}(\text{dark}) = 0.4667$$

and so

$$\text{Supp}_1(A) = 0.7 \qquad \text{Supp}_3(\text{dark}) = 0.4667$$

what finally implies

$$\text{Supp}(\text{Peter}) = 0.5417 \text{ and } \text{acc}(\text{Peter}) = 0.2026.$$

(d) Bob is 37, tall 180 with brown hair. We need to compute here only $\text{Supp}_1(37)$ and $\text{Supp}_2(180)$. The rest is already done in (a). Measures are $\mu_1(37) = \mu_2(180) = 1$, and thus $u_1(37) = r_1(37) = 0.6$ and $u_2(180) = 1$. Proximities are

$$\text{Prox}_1(37, A_{11}) = 0.2 \qquad \text{Prox}_1(37, A_{31}) = 1$$
$$\text{Prox}_2(180, A_{12}) = 0.5 \qquad \text{Prox}_2(180, A_{22}) = 1$$
$$\text{Prox}_2(180, A_{32}) = 1$$

It is obvious that there is no need to modify reasonabilities. Hence,

$$\text{Supp}_{11}(37) = 0.2 \qquad \text{Supp}_{31}(37) = 0.6$$
$$\text{Supp}_{12}(180) = 0.5 \qquad \text{Supp}_{22}(180) = 0.9667$$
$$\text{Supp}_{32}(180) = 0.8$$

and so

$$\text{Supp}_1(37) = 0.4 \qquad \text{Supp}_2(180) = 0.7556$$

what finally implies

$$\text{Supp}(\text{Bob}) = 0.6785 \text{ and } \text{acc}(\text{Bob}) = 0.1859.$$

Let us compare all reasonable alternatives. In Table 2 their measures, acceptabilities, the overall and separate supports are listed.

All alternatives seem to be quite good choices. There is equal amount of information about John and Bob, and Bob seems to be a slightly better alternative. The reason could be that the available information on height of the

	μ	Supp	acc	Supp_1	Supp_2	Supp_3	Supp_4	$\text{Supp}(134)$
John	3.65	0.6729	0.1844	0.6	0.5333	0.5584	1	0.7195
Peter	2.7	0.5417	0.2026	0.7	0	0.4677	1	0.7222
Bob	3.65	0.6785	0.1859	0.4	0.7556	0.5584	1	0.6528

Table 2

driver is more reliable than the one on age. Lack of information on Peter's height prevents us from any final decision since he has the greatest acceptability, as shown in the third column, and thus he seems to be the most acceptable alternative according to the available information. This can be also seen from the last column, where the information on height of the other two candidates is excluded, and hence measures of information for all the three of them are rather equal. Therefore, the final decision can not be made before the information on Peter's height becomes available.

6. CONCLUSION

The work is a contribution to a multi-source multi-attribute, or multi-objective, data fusion. In the case there is precise information on every attribute of every alternative, it results in a unique set of complete solutions. Otherwise, it enables comparing alternatives on the basis of available information, and thus can be considered as a tool for decision support. It is based on Yager's framework for multi-source data fusion. The novelty of this work is in dealing with multiple attributes and using fuzzy sets for representing information.

References

[1] C. Carlsson and R. Fullér, Multiobjective linguistic optimization, *Fuzzy Sets and Systems* 115 (2000), 5–10.

[2] D. Dubois and H. Prade, On the use of aggregation operations in information fusion processes, *Fuzzy Sets and Systems* 142 (2004), 143–161.

[3] R. R. Yager, A framework for multi-source data fusion, *Information Sciences* 163 (2004), 175–200.

[4] R. R. Yager, Uncertainity modeling and decision support, *Reliability Engineering and System Safety* 85 (2004), 341–354.

[5] L. A. Zadeh, Fuzzy logic = computing with words, *IEEE Transactions on Fuzzy Systems* 4 (1996), 103–111.

[6] H.-J. Zimmermann, *Fuzzy sets, decision making, and expert systems*, Kluwer Academic Publishers, 1987.