Part I

Intelligent Systems and Data Mining

Some Considerations in Multi-Source Data Fusion

Ronald R. Yager

Machine Intelligence Institute, Iona College, New Rochelle, NY 10801
yager@panix.com

Abstract. We introduce the data fusion problem and carefully distinguish it from a number of closely problems. Some of the considerations and knowledge that must go into the development of a multi-source data fusion algorithm are described. We discuss some features that help in expressing users requirements are also described. We provide a general framework for data fusion based on a voting like process that tries to adjudicate conflict among the data. We discuss various of compatibility relations and introduce several examples of these relationships. We consider the case in which the sources have different credibility weight. We introduce the idea of reasonableness as a means for including in the fusion process any information available other than that provided by the sources.

Key words: Data fusion, similarity, compatibility relations, conflict resolution

1 Introduction

An important aspect of data mining is the coherent merging of information from multiple sources [1, 2, 3, 4]. This problem has many manifestation ranging from data mining to information retrieval to decision making. One type of problem from this class involves the situation in which we have some variable, whose value we are interested in supplying to a user, and we have multiple sources providing data values for this variable. Before we proceed we want to carefully distinguish our particular problem from some closely related problems that are also important in data mining. We first introduce some useful notation. Let Y be some class of objects. By an attribute A we mean some feature or property that can be associated with the elements in the set Y. If Y is a set of people then examples of attributes are age, height, income and mother's name. Attributes are closely related to the column headings used in a table in a relational data base [3]. Typically an attribute has a domain X in which the values of the attribute can lie. If Y is an element from Y we denote the value of the attribute A for object Y as $A[y]$. We refer to $A[y]$ as a variable. Thus if John is a member of Y the Age [John] is a variable. The value of the

Ronald R. Yager: *Some Considerations in Multi-Source Data Fusion*, Studies in Computational Intelligence (SCI) **5**, 3–22 (2005)
www.springerlink.com

variable $A[y]$ is generally a **unique** element from the domain X. If $A[y]$ takes on the value x we denote this as $A[y] = x$. One problem commonly occurring in data mining is the following. We have the value of an attribute for a number of elements in the class $Y, (A[y_1] = x_1, A[y_2] = x_2, A[y_3] = x_3, \ldots, A[y_q] = x_q)$ and we are interested in finding a value $x* \in x$ as a representative or summary value of this data. We note since each of the $A[y_k]$ is different variables there is no inherent conflict in the fact the values associated with these variables are different. We emphasize that the summarizing value $x*$ is not associated with any specific object in the class Y. It is a value associated with a conceptual variable. At best we can consider $x*$ the value of a variable $A[Y]$. We shall refer to this problem of attaining $x*$ as the **data summarization problem**. A typical example of this would if Y are the collection of people in a city neighbor and A is the attribute salary. Here then we are interested in getting a representative value of the salary of the people in the neighborhood. The main problem we are interested in here, while closely related, is different. Here again we have some attribute A. However instead of being concerned with the class Y we are focusing on one object from this class y_q and we are interested in the value of the variable $A[y_q]$. For example if A is the attribute age and y_q is Osama bin Laden then our interest is in determining Osama bin Laden's age. In our problem of concern the data consists of $(A[y_q] = x_1, A[y_q] = x_2, A[y_q] = x_2, \ldots, A[y_q] = x_n)$. Here we have a number of observations provided by different sources on the value of the variable $A[y_q]$ and we are interested in using this to obtain "a value of the variable $A[y_q]$." We shall call this the **data fusion problem**. While closely related there exists differences. One difference between these problems is that in the fusion problem we are seeking the value of the attribute of a real object rather than the attribute value of some conceptual object. If our attribute is the number of children then determining then the summarizing value over a community is 2.6 may not be a problem, however if we are interested in the number of children that bin Laden has, 2.6 may be inappropriate. Another distinction between these two situations relates to the idea of conflict. In the first situation since $A[y_1]$ and $A[y_2]$ are different variables the fact that $x_1 \neq x_2$ is not a conflict. On the other hand in the second situation, the data fusion problem, since all observations in our data set are about the same variable $A[y_q]$ the fact that $x_a \neq x_b$ can be seen as constituting a conflict. One implication of this relates to the issue of combining values. For example consider the situation in which A is the attribute salary in trying to find the representative (summarizing) value of salaries within a community averaging two salaries such as \$5,000,000 and \$10,000 poses no conceptual dilemma. On the hand if these values are said by different sources to be the salary of some specific individual averaging them would be questionable.

Another problem very closely related to our problem is the following. Again let A be some attribute, y_q be some object and let $A[y_q]$ be a variable whose value we are trying to ascertain. However in this problem $A[y_q]$ is some variable whose value has not yet been determined. Examples of this would be

tomorrow's opening price for Microsoft stock or the location of the next terrorist attack or how many nuclear devices North Korea will have in two years. Here our collection of data $(A[y_q] = x_1, A[y_q] = x_2, A[y_q] = x_2, \ldots, A[y_q] = x_n)$ is such that $A[y_q] = x_j$ indicates the jth source or experts conjecture as to the value of $A[y_q]$. Here we are interested in using this data to predict the value of the future variable $A[y_q]$. While formally almost the same as our problem we believe the indeterminate nature of the future variable introduces some aspects which can effect the mechanism we use to fuse the individual data. For example our tolerance for conflict between $A[y_q] = x_1$ and $A[y_q] = x_2$ where $x_1 \neq x_2$ may become greater. This greater tolerance may be a result of the fact that each source may be basing their predictions on different assumptions about the future world.

Let us now focus on our problem the multi-source data fusion problem. The process of data fusion is initiated by a users request to our sources of information for information about the value of the variable $A[y_q]$. In the following instead using $A[y_q]$ to indicate our variable of interest we shall more simply refer to the variable as V. We assume the value of V lies in the set X. We assume a collection $S_1, S_2, \ldots, S_q$ of information sources. Each source provides a value which we call our data. The problem here becomes the fusion of these pieces of data to obtain a value appropriate for the user's requirements. The approaches and methodologies available for solving this problem depend upon various considerations some of which we shall outline in the following sections. In Fig. 1 we provide a schematic framework of this multi-source data fusion problem which we use as a basis for our discussion.

Our fusion engine combines the data provided by the information sources using various types of knowledge it has available to it. We emphasize that the fusion process involves use of both the data provided by the sources as well as other knowledge. This other knowledge includes both context knowledge and user requirements.

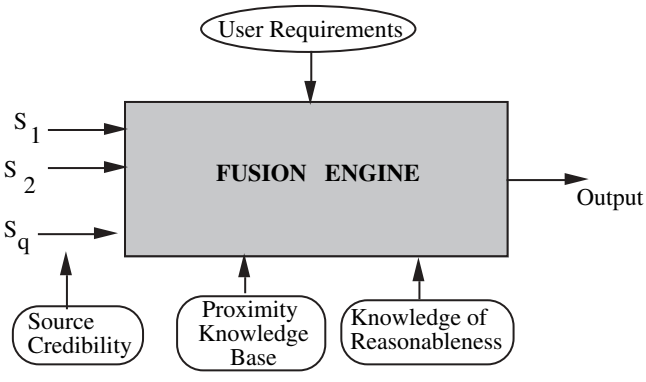

Fig. 1. Schematic of Data Fusion

2 Considerations in Data Fusion

Here we discuss some considerations that effect the mechanism used by the fusion engine. One important consideration in the implementation of the fusion process is related to the form, with respect to its certainty, with which the source provides its information. Consider the problem of trying to determine the age of John. The most certain situation is when a source reports a value that is a member of X, John's age is 23. Alternatively the reported value can include some uncertainty. It could be a linguistic value such as John is "young." It could involve a probabilistic expression of the knowledge. Other forms of uncertainty can be associated with the information provided. We note that fuzzy measures [5, 6] and Dempster-Shafer belief functions [7, 8] provide two general frameworks for representing uncertainty information. Here we shall assume the information provided by a source is a specific value in the space X.

An important of the fusion process is the inclusion of source credibility information. Source credibility is a user generated or sanctioned knowledge base. It associates with the data provided by a source a weight indicating its credibility. The mechanism of assignment of credibility weight to the data reported by a source can be involve various degrees of sophistication. For example, degrees of credibility can be assigned globally to each of the sources. Alternatively source credibility can be dependent upon the type of variable involved. For example, one source may be very reliable with information about ages while not very good with information about a person's income. Even more sophisticated distinctions can be made, for example, a source could be good with information about high income people but bad about income of low people.

The information about source credibility must be at least ordered. It may or may not be expressed using a well defined bounded scale. Generally when the credibility is selected from a well defined bounded scale the assignment of the highest value to a source indicates give the data full weight. The assignment of the lowest value on the scale generally means don't use it. This implies the information should have no influence in the fusion process.

There exists an interesting special situation, with respect to credibility where some sources may be considered as disinformative or misleading. Here the lowest value on the credibility scale can be used to correspond to some idea of taking the "opposite" of the value provided by the source rather than assuming the data provided is of no use. This somewhat akin to the relationship between false and complementation in logic. This situation may require the use of a bipolar scale [9, 10]. Such a scale is divided into two regions separated by a neutral element. Generally the type of operations performed using values from these bipolar depend on from portion of the scale which it was drawn.

Central to the multi-source data fusion problem is the issue of conflict and its resolution. The proximity and reasonableness knowledge bases shown in Fig. 1 play important roles in the handling of this issue.

One form of conflict arises when we have multiple values of a variable which are not the same or even compatible. For example one source may say the age of Osama Bin Laden is 25 another may say he is 45 and another may say he is 85. We shall refer to this as data conflict. As we shall subsequently see the proximity knowledge base plays an important role in issues related to the adjudication of this kind of conflict.

There exists another kind of conflict, one that can occur even when we only have a single reading for a variable. This happens when a sources reported value conflicts with what we know to be the case, what is reasonable. For example, if in searching for the age of Osama Bin Laden, one of the sources reports that he is eighty years old. This conflicts with what we know to be reasonable. This is information which we consider to have a higher priority than any information provided by any of the sources. In this case our action is clear: we discount this observation. We shall call this a context conflict, it relates to a conflict with information available to the fusion process external to the data provided by the sources. The repository of this higher priority information what we have indicated as the knowledge of reasonableness in Fig. 1. This type of a priori context or domain knowledge can take many forms and be represented in different ways.

As an illustration of one method of handling this type of domain knowledge we shall assume our reasonableness knowledge base in the form of a mapping over the domain of V. More specifically a mapping $R : X \rightarrow T$ called the **reasonableness mapping**. We allow this to capture the information we have, external to the data, about the possibilities of the different values in X being the actual value of V. Thus for any $x \in X, R(x)$ indicates the degree of reasonableness of x. T can be the unit interval $I = [0, 1]$ where $R(x) = 1$ indicates that x is a completely reasonable value while $R(x) = 0$ means x is completely unreasonable. More generally T can be an ordered set $T = \{t_1, \ldots, t_n]$. We should point out that the information contained in the reasonableness knowledge base can come from a number of modes. It can be directly related to object of interest. For example from picture of bin Laden in a newspaper dated 1980, given that we are now in 2004, it would clearly be unreasonable to assume that he is less than 24. Historical observations of human life expectancy would make it unreasonable to assume that bin Laden is over 120 years old. Commonsense knowledge applied to recent pictures of him can also provide information regarding the idea reasonableness regarding bin Laden's age. In human agents their use of a knowledge of reasonableness plays fundamental role in distinguishing high performers from lesser. With this in mind it is noted that the need for tools for simply developing and applying these types of reasonableness knowledge bases is paramount.

The reasonableness mapping R provides for the inclusion of information about the context in which we are performing the fusion process. Any data

provided by a source should be acceptable given our external knowledge about the situation. The use of the reasonableness type of relationship clearly provides a very useful vehicle for including intelligence in the process.

In the data fusion process, this knowledge of reasonableness often interacts with the source credibility in an operation which we shall call reasonableness qualification. A typical application of this is described in the following. Assume we have a source that provides a data value a_i and it has credibility t_i. Here we use the mapping R to inject the reasonableness, $R(a_i)$, associated with the value a_i and then use it to modify t_i to give us z_i, the support for data value a_i that came from source S_i. The process of obtaining z_i from t_i and $R(a_i)$ is denoted $z_i = g(t_i, R(a_i))$, and is called **reasonableness qualification**. In the following we shall suppress the indices and denote this operator as $z = g(t, r)$ where $r = R(a)$. For simplicity we shall assume t and r are from the same scale.

Let us indicate some of the properties that should be associated with this operation. A first property universally required of this operation is monotonicity, $g(t_1, r_1) \geq g(t_2, r_2)$ if $t_1 \geq t_2$ and $r_1 \geq r_2$. A second property that is required is that if either t or r is zero, the lowest value on the scale, then $g(t, r) = 0$. Thus if we have no confidence in the source or the value it provides is not reasonable, then the support is zero. Another property that may be associated with this operation is symmetry, $g(t, r) = g(r, t)$. Although we may necessarily require this of all manifestations of the operation.

The essential semantic interpretation of this operation is one of saying that in order to support a value we desire it to be reasonable and emanating from a source in which we have confidence. This essentially indicates this operation is an "anding" of the two requirements. Under this situation a natural condition to impose is the $g(t, r) \leq \text{Min}[t, r]$. More generally we can use a t-norm [11] for g. Thus we can have $g(t, r) = \text{Min}[t, r]$ or using the product t-norm $g(t, r) = tr$.

Relationships conveying information about the congeniality[1] between values in the universe X in the context of their being the value of V play an important role in the development of data fusion systems. Generally these types of relationships convey information about the compatibility and interchangeability between elements in X and as such are fundamental to the resolution and adjudication of internal conflict. Without these relationships conflict can't be resolved. In many applications underlying congeniality relationships are implicitly assumed, a most common example is the use of least squared based methods. The use of linguistic concepts and other granulation techniques are based on these relationships [12, 13]. Clustering operations require these relationships. These relationships are related to equivalence relationships and metrics.

The **proximity relationship** [14, 15] is an important example of these relations. Formally a proximity relationship on a space X is a mapping Prox:

[1] We use this term to indicate relationships like proximity, similarity, equivalence or distance.

$X \times X \rightarrow T$ having the properties: 1. $\text{Prox}(x, x) = 1$ (**reflexive**) and 2. $\text{Prox}(y, x) = \text{Prox}(x, y)$ (**symmetric**). Here T is an ordered space having a largest and smallest element denoted 1 and 0. Often T is the unit interval. Intuitively the value $\text{Prox}(x, y)$ is some measure of degree to which the values x and y are compatible and non-conflicting with respect to context in which the user is seeking the value of V. The concept of metric or distance is related in an inverse way to the concept of proximity.

A closely related and stronger idea is the concept of similarity relationship as introduced by Zadeh [16, 17]. A similarity relationship on a space X is a mapping $\text{Sim}{:}X \times X \rightarrow T$ having the properties: **1)** $\text{Sim}(x, x) = 1$, **2**) $\text{Sim}(x, y) = \text{Sim}(y, x)$ & **3**) $\text{Sim}(x, z) \geq \text{Sim}(x, y) \wedge \text{Sim}(y, z)$. A similarity relationship adds the additional requirement of transitivity. Similarity relationships provide a generalization of the concept of equivalent relationships.

A fundamental distinction between proximity and similarity relationships is the following. In a proximity relationship x and y can be related and y and z can be related without having x and z being related. In a similarity relationship under the stated premise a relationship must also exist between x and z.

In situations in which V takes its value on a numeric scale then the bases of the proximity relationship is the absolute difference $|x - y|$. However the mapping of $|x - y|$ into $\text{Prox}(x, y)$ may be highly non-linear.

For variables having non-numeric values a relationship of proximity can be based on relevant features associated with the elements in the variables universe. Here we can envision a variable having multiple proximity relationships. As an example let V be the country in which John was born, its domain X is the collection of all the countries of the world. Let us see what types of proximity relationship can be introduced on X in this context. One can consider the continent in which a country lies as the basis of a proximity relationship, this would actually generate an equivalence relationship. More generally, the physical distance between the country can be the basis of a proximity relationship. The spelling of the country's name can be the basis of a proximity relationship. The primary language spoken in a country can be the basis of a proximity relationship. We can even envision notable topographic or geographic features as the basis of proximity relationships. Thus many different proximity relationships may occur. The important point here is that the association of a proximity relationship over the domain over a variable can be seen as a very creative activity. More importantly the choice of proximity relationship can play a significant role in the resolution of conflicting information.

A primary consideration that effects the process used by the fusion engine is what we shall call the compositional nature of the elements in the domain X of V. This characteristic plays an important role in determining the types of operations that are available in the fusion process. It determines what types of aggregations we can perform with the data provided by the sources. We shall distinguish between three types of variables with respect to this characteristic. The first type of variable is what we shall call celibate or nominal. These

are variables for which the composition of multiple values is meaningless. An example of this type of variable is a person's name. Here the process of combining names is completely inappropriate. Here fusion can be based on matching and counting. A next more structured type of variable is an ordinal variable. For these types of variables these exists some kind of meaningful ordering of the members of the universe. An example of this is a variable corresponding to size which has as its universe {small, medium, large}. For these variables some kind of compositional process is meaning, combining small and large to obtain medium is meaningful. Here composition operations must be based on ordering. The most structured type of variable is a numeric variable. For these variables in addition to ordering we have the availability of all the arithmetic operators. This of course allows us a great degree of freedom and we have a large body of compositional operators.

3 Expressing User Requirements

The output of any fusion process must be guided by the needs, requirements and desires of the user. In the following we shall describe some considerations and features that can be used to define or express the requirements of the user.

An important consideration in the presentation of the output of the fusion process is the users level of conflict tolerance. Conflict tolerance is related to the **multiplicity** of possible values presented to the user. Does the user desire one unique value or is it appropriate to provide him with a few solutions or is the presentation of all the multi source data appropriate?

Another different, although closely related, issue focuses on the level of granulation of the information provided to the user. As described by Zadeh [18] a granule is a collection of values drawn together by proximity of various types. Linguistic terms such as cold and old are granules corresponding to a collection of values whose proximity is based on the underlying temperature scale. In providing information we must satisfy the user's required level of granularity for the task for which he is requiring the information. Here we are not referring to the number of solutions provided but the nature of each solution object. One situation is that in which each solution presented to the user must be any element from the domain X. Another possibility is one in which we can provide, as a single solution, a subset of closely related values. Presenting ranges of values is an example of this. Another situation is where use a vocabulary of linguistic terms to express solutions. For example if the task is to determine what jacket to wear being told that it is cold is sufficient. Using $a > b$ to indicate that a has larger granularity than b if we consider providing information where somebody lives we see

country > region > state > city.> building address > floor in building > apartment on floor.

Recent interest in ontologies [19] involves many aspects related to granulation.

Another issue related to the form of the output is whether output values presented to the user are required to be values that correspond to one supplied by a source as the input or can we blend source values using techniques such as averaging to construct new values that didn't appear in the input. A closely related issue is the reasonableness of the output. For example consider the attempt to determine the number of children that John has. Assume one source says 8 and another says 7, taking the average gives us 7.5. Well, clearly it is impossible for our John to have 7.5 children. For some purposes this may be an appropriate figure. In addition we should note the that sometimes the requirement for reasonableness may be different for the output than input.

Another feature of the output revolves around the issue of qualification. Does the user desire qualifications associated with suggested values or does he prefer no qualification? As we indicated data values inputted to a fusion system often have attached values of credibility, this being due to the credibility of the source and the reasonableness of the data provided. Considerations related to the presentation of this credibility arise regarding the requirements of the user. Are we to present weights of credibility with the output or present it without these weights? In many techniques, such as weighted averaging, the credibility weight gets subsumed in the fusion process.

In most cases the fusion process should be deterministic, a given informational situation should always result in the same fused value. In some cases we may allow for a non-deterministic, random mechanism in the fusion process. For example in situations in which some adversary may have some role in effecting the information used in the fusion process we may want to use randomization to blur and confuse the influence of their information.

4 A Framework for Multi-Source Data Fusion

Here we shall provide a basic framework in which to view and implement the data fusion process. We shall see that this framework imposes a number of properties that should be satisfied by a rational data fusion technology.

Consider a variable of interest V having an underlying universe X. Assume we have as data a collection of q assessment of this variable, $\{V = a_1, V = a_2, V = a_3, \ldots, V = a_q\}$ Each assessment is information supplied by one of our sources. Let a_i be the value provided by the source S_i. Our desire here is to fuse these values to obtain some value $\tilde{a} \in X$ as the fused value. We denote this as a $\tilde{a} = \text{Agg}(a_1, \ldots, a_n)$. The issue then becomes that of obtaining the operator Agg that fuses these pieces of data. One obvious requirement of such an aggregation operator is idempotency, if all $a_i = a$ then $\tilde{a} = a$.

In order to obtain acceptable forms for Agg we must conceptually look at the fusion process. At a meta level multi-source data fusion is a process in which the individual sources must agree on a solution that is acceptable to each of them, that is compatible with the data they each have provided.

Let $\boldsymbol{a}$ be a proposed solution, some element from X. Each source can be seen as "voting" whether to accept this solution. Let us denote $\text{Sup}_i(\boldsymbol{a})$ as the support for solution $\boldsymbol{a}$ from source i. We then need some process of combining the support for $\boldsymbol{a}$ from each of the sources. We let

$$\text{Sup}(\boldsymbol{a}) = F(\text{Sup}_1(\boldsymbol{a}), \text{Sup}_2(\boldsymbol{a}), \ldots, \text{Sup}_q(\boldsymbol{a}))$$

be the total support for $\boldsymbol{a}$. Thus F is some function that combines the support from each of the sources. The fused value $\tilde{a}$ is then obtained as the value $a \in X$ that maximizes $\text{Sup}(\boldsymbol{a})$. Thus $\tilde{a}$ is such that $\text{Sup}(\tilde{a}) = \text{Max}_{a \in X}[\text{Sup}(a)]$. In some situations we may not have to search through the whole space X to find an element $\tilde{a}$ having the property $\text{Sup}(\tilde{a}) = \text{Max}_{a \in X}[\text{Sup}(a)]$.

We now introduce the ideas of solution set and minimal solution set which may be useful We say that a subset G of X is a **solution set** if all $\boldsymbol{a}$ s.t. $\text{Sup}(\boldsymbol{a}) = \text{Max}_{a \in X}[\text{Sup}(a)]$ are contained in G. The determination of G is useful in describing the nature of the type of solution we can expect from a fusion process. We shall say that a subset H of X is a **minimal solution** set if there always exists one element $\boldsymbol{a} \in H$ s.t. $\text{Sup}(\boldsymbol{a}) = \text{Max}_{a \in X}[\text{Sup(a)}]$. Thus a minimal solution set is a set in which we can always find an acceptable fused value. The determination of a minimal solution set can help reduce the task of searching.

Let us consider some properties of F. One natural property associated with F is that the more support from the individual sources the more overall support for $\boldsymbol{a}$. Formally if $\boldsymbol{a}$ and $\boldsymbol{b}$ are two values and if $\text{Sup}_i(\boldsymbol{a}) \geq \text{Sup}_i(\boldsymbol{b})$ for all i then $\text{Sup}(\boldsymbol{a}) \geq \text{Sup}(\boldsymbol{b})$. This requires that F be a monotonic function, $F(x_1, x_2, \ldots, x_q) \geq F(y_1, y_2, \ldots, y_q)$ if $x_i \geq y_i$ for all i. A slightly stronger requirement is **strict monotonicity**. This requires that F be such that if $x_i \geq y_i$ for all i and there exists at least one i such that $x_i > y_i$ then $F(x_i, \ldots, x_q) > F(y_1, \ldots, y_q)$.

Another condition we can associate with F is a symmetry with respect to the arguments. That is the indexing of the arguments should not affect the answer. This symmetry implies a more expansive situation with respect to monotonicity. Assume $t_1, \ldots, t_q$ and $\hat{t}_1, \ldots, \hat{t}_q$ are two sets of arguments of F, $\text{Sup}_i(a) = t_i$ and $\text{Sup}_i(\hat{a}) = \hat{t}_i$. Let perm indicate a permutation of the arguments, where perm(i) is the index of the ith element under the permutation. Then if there exists some permutation such that $t_i \geq \hat{t}_{\text{perm}(i)}$ for all i we get

$$F(t_1, \ldots, t_q) \geq F(\hat{t}_1, \ldots, \hat{t}_q) \; .$$

Let us look further into this framework. A source's support for a solution, $\text{Sup}_i(\boldsymbol{a})$, should depend upon the degree of compatibility between the proposed solution $\boldsymbol{a}$ and the value provided by the source, a_i. Let us denote $\text{Comp}(\boldsymbol{a}, a_i)$ as this compatibility. Thus $\text{Sup}_i(\boldsymbol{a})$ is some function of the compatibility between a_i and $\boldsymbol{a}$. Furthermore, we have a monotonic type of relationship. For any two values $\boldsymbol{a}$ and $\boldsymbol{b}$ if $\text{Comp}(\boldsymbol{a}, a_i) \geq \text{Comp}(\boldsymbol{b}, a_i)$ then $\text{Sup}_i(\boldsymbol{a}) \geq \text{Sup}_i(\boldsymbol{b})$.

The compatibility between two objects in X is based upon some underlying proximity relationship. The concept of a proximity relationship, which we introduced earlier, has been studied is the fuzzy set literature [20]. Here then we shall assume a relationship Comp, called the compatibility relationship, which has at least the properties of a proximity relationship. Thus Comp: $X \times X \rightarrow T$ in which T is an ordered space with greatest and least elements denoted 1 and 0 and having the properties: 1) $\text{Comp}(x, x) = 1$ and 2) $\text{Comp}(x, y) = \text{Comp}(y, x)$. A suitable although not necessary, choice for T is the unit interval.

We see that this framework imposes an idempotency type condition on the aggregation process. Assume $a_i = \boldsymbol{a}$ for all i. In this case $\text{Comp}(\boldsymbol{a}, a_i) = 1$ for all i. From this it follows that for any $b \in X$ $\text{Comp}(\boldsymbol{a}, a_i) \leq \text{Comp}(\boldsymbol{b}, a_i)$ hence $\text{Sup}_i(\boldsymbol{a}) \geq \text{Sup}_i(\boldsymbol{b})$ for all $\boldsymbol{b}$ thus $\text{Sup}(\boldsymbol{a}) \geq \text{Sup}(\boldsymbol{b})$ for all $\boldsymbol{a}$. Thus there can never be a better solution than $\boldsymbol{a}$. Furthermore, if F is assumed strictly monotonic and Comp is such that $\text{Comp}(a, b) \neq 1$ for $a \neq b$ then we get a strict idempotency.

5 Compatibility Relationships

What is important to emphasize here is that by basing our fusion process on the idea of the compatibility relationship we can handle, **in a unified manner,** the fusion of variables whose values are drawn from sets (universes) having widely different properties. Consider the variables John's age and John's city of residence. These variables take their values from sets of a completely different nature. Age is drawn from a purely mathematical set possessing all the structure that this affords, we can add or subtract or multiply elements. The city of residence has none of these properties. Its universe is of a completely different nature. What is also important to emphasize is that in order to use this approach on a variable V we must be able to obtain an appropriate context sensitive compatibility relation over its domain X. It is in this process of obtaining the compatibility relationship that we make use of the nature, the features and properties, of the elements in X. The construction of the compatibility relationship is often an extremely subjective task and greatly effects the end result. While in the numeric variables the basic feature used to form $\text{Comp}(a, b)$ is related to the difference $|a - b|$ this may be very complicated. For example the compatibility between salaries of 20 million and 30 million may be greater then the compatibility between salaries of 30 thousand and 50 thousand. While in the case numeric variables where the only feature of the elements in the domain useful for constructing the compatibility relationship is the numeric value in the case of other variables such as the country of residence the elements in the domain X have a number of features that can be used as the basis of an underlying compatibility relationship. This leads to the possibility of having multiple available compatibility relationships in our fusion process. While in the remainder of our work we shall assume the fusion

process is based on one well defined compatibility relationship we would like to describe one generalization related to the situation of having the availability of multiple compatibility relations over the domain of the variable of interest. Earlier we indicated that the fused value is $\tilde{a}$ such $\text{Sup}(\tilde{a}) = \text{Max}_{a \in X}[\text{Sup}(a)]$. In the case of multiple possible compatibility relations C_k for $k = 1$ to m then if we let $\text{Sup}(a)/k$ indicate the Sup for a under compatibility relation C_k the process of obtaining the fused value may involve finding $\tilde{a}$ and compatibility relation C_k* such that $\text{Sup}(\tilde{a})/k* = \text{Max}_k[\text{Max}_{a \in X}[\text{Sup}(a)/k]]$.

At a formal level compatibility relations are mathematical structures that well studied and characterized. We now look at some very important special examples of compatibility relationships. We particularly focus on the properties of the solution sets that can be associated with relations. This helps us understand the nature of the fused values we may obtain. In the following discussion we shall let B be the set of all the values provided by the sources, $B = \{a_j \mid V = a_j \text{ for some source}\}$.

First we consider a very strict compatibility relation. We assume $\text{Comp}(a, b) = 1$ if $a = b$ and $\text{Comp}(a, b) = 0$ if $a \neq b$. This is a very special kind of equivalence relationship, elements are only equivalent to themselves. It can be shown under the condition of monotonicity of F the minimal solution set is the set B. This means the fused value for this type of compatibility relation must be one the data points provided by the sources.

Consider now the case where Comp is an equivalence relationship, $\text{Comp}(a, b) \in \{0, 1\}$ and $\text{Comp}(a, a) = 1$, $\text{Comp}(a, b) = \text{Comp}(b, a)$ and if $\text{Comp}(a, b) = 1$ and $\text{Comp}(b, c) = 1$ the $\text{Comp}(a, c) = 1$. It can be shown [21] in this case that B also provides a minimal solution set, no solution can be better than some element in B.

We turn to another type of compatibility relationship, one in which there exists some **linear ordering** on the space X which underlies the compatibility relation. Let $\mathtt{L}$ be a linear ordering on X where $x \underset{\mathtt{L}}{\geq} y$ indicates that x is larger than y in the ordering. Let Comp be a compatibility relationship on X which in addition to being reflexive and symmetric is such that the closer two elements are in the ordering $\mathtt{L}$ the more compatible they are. More formally we assume that if $x \underset{\mathtt{L}}{\geq} y \underset{\mathtt{L}}{\geq} z$ then $\text{Comp}(x, y) \geq \text{Comp}(x, z)$. We say this connection between ordering and compatibility is strict if $x \underset{\mathtt{L}}{\geq} y \underset{\mathtt{L}}{\geq} z$ implies $\text{Comp}(x, y) > \text{Comp}(x, z)$. Again let B be the set of data values provided by the sources. Let a^* be the largest element in B with respect to the underlying ordering $\underset{\mathtt{L}}{\geq}$ and let a_* be the smallest element in B with respect to the ordering. It can be shown that the subset H of H where $H = \{a \mid a_* \underset{\mathtt{L}}{\leq} a \underset{\mathtt{L}}{\leq} a^*\}$ is a minimal solution set. Thus under this type of compatibility relationship only requiring only that F is monotonic leads to the situation which our fused value will be found in the "interval of X" bounded by a_* and a^*. This is a very interesting and deep result. Essentially this is telling us that if we view the process of obtaining the fused value as an aggregation of the data, $a = \text{Agg}(a_1, a_2, \ldots, a_q\}$ then Agg is a mean like operation.

6 Additional Requirement on F

We described the process of determining the fused value to a data collection $\langle a_1, \ldots, a_q \rangle$ as to be conceptually implemented by the following process:

(1) For any $a \in X$ obtain $\text{Sup}_i(a) = \text{Comp}(a, a_i)$
(2) Evaluate $\text{Sup}(a) = F(\text{Sup}_i(a), \ldots, \text{Sup}_q(a))$
(3) Select as fused value the $\tilde{a}$ such that $\text{Sup}(\tilde{a}) = \text{Max}_{a \in X}[\text{Sup}(a)]$

We explicitly made two assumptions about the function F, we assumed that F was symmetric the indexing of input information is not relevant and F is monotonic. An implicit assumption we made about F was an assumption of pointwiseness.

There exists another property we want to associate with F, it is closely related to the idea of self-identity discussed by Yager and Rybalov [22]. Assume that we have a data set $\langle a_1, \ldots, a_q \rangle$ and using our procedure we find that $\tilde{a}$ is the best solution $\text{Sup}(\tilde{a}) \geq \text{Sup}(x)$ for all x in X. Assume now that we are provided an additional piece of data a_{q+1} such that $a_{q+1} = \tilde{a}$, the new data suggests $\tilde{a}$ as its value. Then clearly $\tilde{a}$ should still be the best solution. We shall formalize this requirement. In the following we let $\tilde{a}$ and $\hat{a}$ be two possible solutions and let $\tilde{c}_i = \text{Comp}(\hat{a}, a_i)$ and $\hat{c}_i = \text{Comp}(\hat{a}, a_i)$. We note that if $a_{q+1} = \tilde{a}$ then $\tilde{c}_{q+1} \geq \hat{c}_{q+1}$ since

$$\begin{aligned} \tilde{c}_{q+1} &= \text{Comp}(\tilde{a}, a_{q+1}) = \text{Comp}(\tilde{a}, \tilde{a}) = 1 \geq \text{Comp}(\hat{a}, \tilde{a}) \\ &\geq \text{Comp}(\hat{a}, a_{q+1}) = \hat{c}_{q+1} \end{aligned}$$

Using this we can more formally express our additional requirement on F. If

$$F(\tilde{c}_1, \ldots, \tilde{c}_q) \geq F(\hat{c}_i, \ldots, \hat{c}_q)$$

and if $\tilde{c}_{q+1} \geq \hat{c}_{q+1}$ then we require that

$$F(\tilde{c}_1, \ldots, \tilde{c}_q, \tilde{c}_{q+1}) \geq F(\hat{c}_i, \ldots, \hat{c}_q, \hat{c}_{q+1}) \; .$$

We note that this last condition is not exactly a standard monotonicity condition. We call this property stepwise monotonicity. We now have specified four conditions on F: pointwise, monotonicity, symmetry and stepwise monotonicity.

Let us now consider the issue of providing some formulations for F that manifest the conditions we require. Before we do this we must address the measurement of compatibility. In our work so far we have assumed a very general formulation for this measurement. We have defined Comp: $X \times X \rightarrow T$ in which T is an ordered space with greatest and least elements denoted 1 and 0. Let us consider the situation in which T has only an ordering. In this case one form for F is that of a Max operator. Thus $F(t_1, t_2, \ldots, t_q) = \text{Max}_i[C_i]$ satisfies all the conditions required. We also note that the Min operator satisfies our conditions.

If we consider the situation in which the compatibility relation takes its values in the unit interval, [0, 1] one formulation for F that meets all our required conditions is the sum or totaling function, $F(x_1, x_2, \ldots x_q) = \sum_{i=1}^{q} x_i$. Using this we get $\text{Sup}(a) = \sum_{i=1}^{q} \text{Sup}_i(a) = \sum_{i=1}^{q} \text{Comp}(a, a_i)$. Thus our fused value is the element that maximizes the sum of its compatibilities with the input.

7 Credibility Weighted Sources

In the preceding we have implicitly assumed all the data had the same credibility. Here we shall consider the situation in which each data has a credibility weight w_i. Thus now our input is q pairs of (w_i, a_i). We also note that the weight w_i must be drawn from a scale that has at least an ordering. In addition we assume this scale has minimal and maximal elements denoted 0 and 1.

Again in this situation for any $a \in X$ we calculate $\text{Sup}(a) = F(\text{Sup}_1(a), \ldots, \text{Sup}_q(a))$ where $\text{Sup}_i(a)$ is the support for $\mathbf{a}$ from the data supplied by source $i, (w_i, a_i)$. However in this case, $\text{Sup}_i(a)$ depends upon two components. The first being the compatibility of a with a_i, $\text{Comp}(a, a_i)$ and the second being the weight or strength of credibility source i. Thus in this case

$$\text{Sup}_i(a) = g(w_i, \text{Comp}(a, a_i))$$

Ideally we desire that both w_i and $\text{Comp}(a, a_i)$ be drawn from the same scale, which has at least an ordering. For the following discussion we shall not implicitly make this assumption. However, we shall find it convenient to use 0 and 1 to indicate the least and greatest element on each of the scales. We now specify the properties that are required of the function g. A first property we require of g is monotonicity with respect to both of the arguments: $g(x, y) \geq g(z, y)$ if $x > z$ and $g(x, y) \geq g(x, w)$ if $y > w$. Secondly we assume that zero credibility or zero compatibility results in zero support: $g(x, 0) = g(0, y) = 0$ for all x and y. We see that g has the character of an "and" type operator. In particular at a semantic level we see that we are essentially saying is "source i provides support for solution if the source is credible and the solution is compatible with the sources data".

With this we see that $g(1, 1) = 1$ and $g(x, y) \neq 0$ if $x \neq 0$ and $y \neq 0$. We must make one further observation about this process with respect to source credibility. Any source that has zero credibility should in no way effect the decision process. Thus if $((w_1, a_1), \ldots, (w_q, a_q))$ has as its fused value $\tilde{a}$ then the data $((w_1, a_1), \ldots, (w_q, a_q), (w_{q+1}, a_{q+1}))$ where $w_{q+1} = 0$ should also have the same result. With this understanding we can discard any source with zero credibility. In the following we shall assume unless otherwise stated all sources have non-zero credibility.

8 Including Reasonableness

In an early part we introduced the idea of a **R**easonableness **K**nowledge **B**ase (RKB) and indicated its importance in the data fusion process. Formally we use this structure to introduce into the fusion process any information we have about the value of the variable exclusive of the data provided by the sources. The information in the reasonableness knowledge base will affect our proposed fusion process in at least two ways. First it will interact with the data provided by the sources. In particular, the weight (credibility) associated with a source providing an unreasonable input value should be diminished. This results in our giving the data less importance in the fusion process. Secondly some mechanism should be included in the fusion process to block unreasonable values from being provided as the fused value.

A complete discussion of the issues related to the construction of the RKB and those related to formal methods for the interaction of the RKB with the data fusion process is complex and beyond our immediate aim as well as well being beyond our complete understanding at this time. In many ways the issue of reasonableness goes to the very heart of intelligence. Here we shall focus on the representation of a specific type of knowledge effecting what are reasonable values for a variable and suggest a method for introducing this in the fusion process.

We shall distinguish between two types of information about the value of a variable with the terms intimate and collective knowledge. Before making this distinction we recall a variable V is formally denoted as $A(y)$ where A is an attribute and y is a specific object. For example if the variable is *John's age* then age is the attribute and John is the object. By intimate knowledge we mean information directly about the variable whose value we are trying obtain. Knowing that John was born after Viet Nam war or that Mary lives in Montana are examples of intimate knowledge. By collective knowledge we mean information about the value of the attribute for a class of objects in which our object of interest lies. Knowing that Singaporeans typically are college graduates is collective knowledge while knowing that Min-Sze has a PhD is intimate knowledge. Generally intimate knowledge has a possibilistic nature while collective knowledge has a probabilistic nature. (The preceding statement is an example of collective knowledge). Another type of knowledge related to reasonableness is what has been called default (commonsense) knowledge [23, 24]. This knowledge is such that while we have not been given intimate knowledge that xyz is the value of a variable we can act as if this is the case unless we have some overriding intimate knowledge saying that this is not the case. One view of default knowledge is that it is collective knowledge that is so pervasively true from a pragmatic point of view it is more economical to act as if it is categorical, holds for all objects, and deal with exceptions as they are pointed out.

Here we consider only the situation in which our knowledge about reasonableness is intimate and can be expressed by fuzzy subset, a mapping

$R : X \rightarrow T$. As pointed out by Zadeh [25] this kind of knowledge induces a constraint on the values of the variable and has a possibilistic nature [26]. Here for any $x \in X, R(x)$ indicates the reasonableness (or possibility) that x is the value of the variable V. For example, if our interest is to obtain John's age and before soliciting data from external sources we know from our personal interview that John is *young* then we can capture this information using the fuzzy subset R corresponding to and thus constrain the values that are reasonable.

Let us see how we can include this information into our data fusion process. In the following we assume that T is a linear ordering having maximal and minimal elements, usually denoted 1 and 0. Assume the data provided by source i is denoted a_i and w_i is the credibility assigned to source i. We assume these credibilities are measured on the same scale as the reasonableness, T. In the fusion process the importance weight, u_i, assigned to the data a_i should be a function of the credibility of the source, w_i, **and** the reasonableness of the data, $R(a_i)$. An unreasonable value, whatever the credibility of the source, should not be given much significance in the fusion. Similarly a piece of data coming from a source with low credibility, whatever the reasonableness of its value, should not be given much significance in the fusion. Using the Min to implement this "anding" we obtain $u_i = \text{Min}[R(a_i), w_i]$ as the importance weight assigned to the data a_i coming from this source. In this environment the information that goes to the fusion mechanism is the collection $\langle (u_1, a_1), \ldots, (u_q, a_q) \rangle$.

As in the preceding the overall support for a proposed fused value $\boldsymbol{a}$ should be a function its support from each of the sources, $\text{Sup}(\boldsymbol{a}) = F(\text{Sup}_i(\boldsymbol{a}), \ldots, \text{Sup}_q(\boldsymbol{a}))$. The support provided from source i for solution $\boldsymbol{a}$ should depend on the importance weight u_i assigned to data supplied by source i as well as the compatibility of the data a_i and the proposed fused value, $\text{Comp}(\boldsymbol{a}, a_i)$. In addition we should also include information about the reasonableness of the proposed solution $\boldsymbol{a}$. Here then for a solution $\boldsymbol{a}$ to get support from source i it should be compatible with the data a_i and compatible with what we consider to be reasonable, $\text{Comp}(\boldsymbol{a}, R)$. Here then we let $\text{Comp}_i(\boldsymbol{a}) = \text{Comp}(\boldsymbol{a}, a_i) \wedge \text{Comp}(\boldsymbol{a}, R)$. Furthermore $\text{Comp}(\boldsymbol{a}, R) = R(\boldsymbol{a})$ hence $\text{Comp}_i(\boldsymbol{a}) = \text{Comp}(\boldsymbol{a}, a_i) \wedge R(a)$. In addition, as we have indicated, the support afforded any solution by source i should be determined in part by the importance weight assigned i. Taking these considerations into account we get $\text{Sup}_i(\boldsymbol{a}) = g(u_i, \text{Comp}_i(\boldsymbol{a}))$. Substituting our values we get

$$\text{Sup}_i(\boldsymbol{a}) = g(w_i \wedge R(a_i), \text{Comp}(\boldsymbol{a}, a_i) \wedge R(\boldsymbol{a}))$$

What is clear is that g should be monotonically increasing in both its arguments and be such that if any of the arguments are 0 then $\text{Sup}_i(\boldsymbol{a}) = 0$. In the case where we interpret g as implementing an *anding* and using the Min operator as our *and* we get $\text{Sup}_i(\boldsymbol{a}) = w_i \wedge R(a_i) \wedge R(\boldsymbol{a}) \wedge \text{Comp}(\boldsymbol{a}, a_i)$. Here we observe that the support afforded from source i to any proposed fused solution is related to the credibility of the source, the reasonableness of value

provided by the source, the reasonableness of the proposed fusion solution and the compatibility of the data and solution.

Earlier we looked at the form of solution set for the fused value under different assumptions about the underling compatibility relationship. Let us now investigate how the introduction of reasonableness affects our results about boundedness and minimal solution sets. For simplicity neglect the issue of source credibility, we assume all sources are fully credible.

Consider the case in which our underlying compatibility relationship is very strict, $\mathrm{Comp}(x, y) = 1$ iff $x = y$ and $\mathrm{Comp}(x, y) = 0 x \neq y$. Let B be the set of data values and let $\hat{B}$ be the subset of B such that $b \in \hat{B}$ if $R(b) \neq 0$, it is the set of reasonable data values. If $a \notin B$ then $\mathrm{Comp}(a, a_i) = 0$ for all a_i and hence $\mathrm{Sup}_i(a) = 0$ for all i. Let $d \in B - \hat{B}$, here $R(d) = 0$ and again we get that $\mathrm{Sup}_i(d) = 0$ for all i. On the other hand for $b \in \hat{B}$ then $R(b) \neq 0$ and $b = a_j$ for some j and hence $\mathrm{Sup}_j(b) > 0$. Thus we see that we will always find our solution in the space $\hat{B}$, the set of data values that are not completely unreasonable. Actually in this case for each $b \in \hat{B}$ its overall support is the number of sources that provided this value.

Consider now the case in which Prox is an ordinary equivalence relation. Again let $\hat{B}$ be our set of input data which have some degree of reasonableness. Let E_i be the equivalence class of a_i, for all $y \in E_i$, $\mathrm{Prox}(y, a_i) = 1$. Let $E = \underset{i}{\cup} E_i$, the union of all equivalence classes that have input value. If $a \notin E$ then $\mathrm{Prox}(a, a_i) = 0$ for all i. From this we see that if $a \notin E$ then $\mathrm{Sup}_i(a) = 0$ for all i and hence we can always find at least as good a solution in E. We can obtain a further restriction on the minimal solutions. Let $D_i \subseteq E_i$ be such that $d_i \in D_i$ if $R(d_i) = \mathrm{Max}_{x \in F_i}(R(x))$. Thus D_i is the subset of elements that are equivalent to a_i and are most reasonable. For any $d_i \in D_i$ and any $e_i \in E_i$ we have that for all input data a_j $\mathrm{Comp}(e_i, a_j) = \mathrm{Comp}(d_i, a_j)$. Since $R(d_i) \geq R(e_i)$ we see that $\mathrm{Sup}_j(d_i) \geq \mathrm{Sup}_j(e_i)$ for all j. Hence d_i is always at least as good a fused value as any element in E_i. Thus we can always find a fused solution in $D = \underset{i}{\cup} D_i$. Furthermore if x and $y \in D_i$ then $R(x) = R(y)$ and $\mathrm{Comp}(x, z) = \mathrm{Comp}(y, z)$ for all z. Hence $\mathrm{Sup}_i(x) = \mathrm{Sup}_i(y)$. Thus $\mathrm{Sup}(x) = \mathrm{Sup}(y)$. The result is that we can consider any element in D_i. Thus all we need consider is the set $\tilde{D} = \underset{i}{\cup} \{\tilde{d}_i\}$ where $\tilde{d}_i$ is any element from D_i. We note that if $a_i \in D_i$ then this is of course the preferred element.

We now consider the case where the proximity relationship is based on a linear ordering $\mathtt{L}$ over space X. Let B be the set of data values provided by the sources. Let x^* and x_* be the maximal and minimal elements in B with respect to the ordering $\mathtt{L}$. Let H be the set of x_j so that $x^* \underset{\mathtt{L}}{\geq} x_j \underset{\mathtt{L}}{\geq} x_*$. In the preceding we showed that we can always find a fused value element $\boldsymbol{a}$ in H. We now show that the introduction of reasonableness removes this property.

In the preceding we indicated that for any proposed fused value we get that $\mathrm{Sup}_i(a) = g(u_i, \mathrm{Comp}_i(a))$ where g monotonic in both the arguments, $u_i = w_i \wedge R(a_i)$ and $\mathrm{Comp}_i(a) = R(a) \wedge \mathrm{Comp}(a, a_i)$. We shall now show that here we can have an element $a \notin H$ in which $\mathrm{Sup}_i(a) \geq \mathrm{Sup}_i(b)$ for all $b \in H$.

This implies that we can't be guaranteed of finding the fused value in H. Consider now the case in which there exists $b \in H$ for which $R(b) \leq \alpha$. In this case $\text{Sup}_i(b) = g(u_i, R(b) \wedge \text{Comp}(b, a_i)) \leq g(u_i, \alpha)$. Let $a \notin H$ be such that $R(a) > \alpha$. For this element we get $\text{Sup}_i(a) = g(u_i, R(a) \wedge \text{Comp}(a, a_i))$. If $\text{Comp}(a, a_i) > \alpha$ then $R(a) \wedge \text{Comp}(a, a_i) = \beta$ then $\beta > \alpha$ and hence $\text{Sup}_i(a) = g(u_i, \beta) \geq g(u_i, \alpha) = \text{Sup}_i(b)$ and then it is not true we can eliminate a as a solution. Thus we see that the introduction of this reasonableness allows for the possibility of solutions not bounded by the largest and smallest of input data.

An intuitive boundary condition can be found in this situation. Again let H be the subset of X bounded by our data: $H = \{x \mid x^* \underset{L}{\geq} x \underset{L}{\geq} x_*\}$ where let $\alpha^* = R(x_*)$ and let $\alpha_* = R(x^*)$. Let $H^* = \{x \mid x \underset{L}{>} x^*$ and $R(x) > R(x^*)\}$ and let $H_* = \{x \mid x \underset{L}{<} x_*$ and $R(x) > R(x_*)\}$. Here we can restrict ourselves to looking for the fused value in the set $\hat{H} = H \cup H_* \cup H^*$. We see that as follows. For any $x \underset{L}{>} x^*$ we have, since the proximity relationship is induced by the ordering, that $\text{Comp}(x, a_i) \leq \text{Comp}(x^*, a_i)$ for all data a_i. If in addition we have that $R(x) \leq R(x^*)$ then $\text{Sup}_i(x) = g(u_i, R(x) \wedge \text{Comp}(x, a_i)) \leq \text{Sup}_i(x^*) = g(u_i, R(x^*) \wedge \text{Comp}(x^*, a_i))$ for all i and hence $\text{Sup}(x) \leq \text{Sup}(x^*)$. Thus we can eliminate all $x \underset{L}{>} x^*$ having $R(x) \leq R(x^*)$. Using similar arguments we can eliminate $x \underset{L}{<} x_*$ which have $R(x) \leq R(x_*)$.

9 Conclusion

We presented a general view of the multi-source data fusion process and described some of the considerations and information that must go into the development of a data fusion algorithm. Features playing a role in expressing users requirements were also discussed. We introduced a general framework for data fusion based on a voting like process which made use of compatibility relationships. We described several important examples of compatibility relationships. We showed that our formulation resulted in specific bounding conditions on the fused value depending on the underlying compatibility relationships. We noted the existence of these bounding conditions essentially implied that the fusion process has the nature of a mean type aggregation. We presented the concept of reasonableness as a means for including in the fusion process any information available other then that provided by the sources. We considered the situation in which we allowed our fused value to be granular objects such as linguistic terms or subsets.

References

1. Berry, M. J. A. and Linoff, G., Data Mining Techniques, John Wiley & Sons: New York, 1997. 3

2. Dunham, M., Data Mining, Prentice Hall: Upper Saddle River, NJ, 2003. 3
3. Han, J. and Kamber, M., Data Mining: Concepts and Techniques, Morgan Kaufmann: San Francisco, 2001. 3
4. Mitra, S. and Acharya, T., Data Mining: Multimedia. Soft Computing and Bioinformatics, New York: Wiley, 2003. 3
5. Murofushi, T. and Sugeno, M., "Fuzzy measures and fuzzy integrals," in Fuzzy Measures and Integrals, edited by Grabisch, M., Murofushi, T. and Sugeno, M., Physica-Verlag: Heidelberg, 3–41, 2000. 6
6. Yager, R. R., "Uncertainty representation using fuzzy measures," IEEE Transaction on Systems, Man and Cybernetics 32, 13–20, 2002. 6
7. Shafer, G., A Mathematical Theory of Evidence, Princeton University Press: Princeton, N.J., 1976. 6
8. Yager, R. R., Kacprzyk, J. and Fedrizzi, M., Advances in the Dempster-Shafer Theory of Evidence, John Wiley & Sons: New York, 1994. 6
9. Yager, R. R. and Rybalov, A., "Uninorm aggregation operators," Fuzzy Sets and Systems 80, 111–120, 1996. 6
10. Yager, R. R., "Using a notion of acceptable in uncertain ordinal decision making," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10, 241–256, 2002. 6
11. Klement, E. P., Mesiar, R. and Pap, E., Triangular Norms, Kluwer Academic Publishers: Dordrecht, 2000. 8
12. Zadeh, L. A., "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," Fuzzy Sets and Systems 90, 111–127, 1997. 8
13. Lin, T. S., Yao, Y. Y. and Zadeh, L. A., Data Mining, Rough Sets and Granular Computing, Physica-Verlag: Heidelberg, 2002. 8
14. Kaufmann, A., Introduction to the Theory of Fuzzy Subsets: Volume I, Academic Press: New York, 1975. 8
15. Bouchon-Meunier, B., Rifqi, M. and Bothorol, S., "Towards general measures of comparison of objects," Fuzzy Sets and Systems 84, 143–153, 1996. 8
16. Zadeh, L. A., "Similarity relations and fuzzy orderings," Information Sciences 3, 177–200, 1971. 9
17. Yager, R. R., Ovchinnikov, S., Tong, R. and Nguyen, H., Fuzzy Sets and Applications: Selected Papers by L. A. Zadeh, John Wiley & Sons: New York, 1987. 9
18. Zadeh, L. A., "Toward a logic of perceptions based on fuzzy logic," in Discovering the World with Fuzzy Logic, edited by Novak, W. and Perfilieva, I., Physica-Verlag: Heidelberg, 4–28, 2001. 10
19. Gomez-Perez, A., Fernandez-Lopez, M. and Corcho, O., Ontological Engineering, Springer: Heidelberg, 2004. 11
20. Shenoi, S. and Melton, A., "Proximity relations in fuzzy relational databases," Fuzzy Sets and Systems 31, 287–298, 1989. 13
21. Yager, R. R., "A framework for multi-source data fusion," Information Sciences 163, 175–200, 2004. 14
22. Yager, R. R. and Rybalov, A., "Noncommutative self-identity aggregation," Fuzzy Sets and Systems 85, 73–82, 1997. 15
23. Reiter, R., "A logic for default reasoning," Artificial Intelligence 13, 81–132, 1980. 17
24. McCarthy, J., "Applications of circumscription to formalizing common sense knowledge," Artificial Intelligence 28, 89–116, 1986. 17

25. Zadeh, L. A., "Outline of a computational theory of perceptions based on computing with words," in Soft Computing and Intelligent Systems, edited by Sinha, N. K. and Gupta, M. M., Academic Press: Boston, 3–22, 1999. 18
26. Zadeh, L. A., "Fuzzy sets as a basis for a theory of possibility," Fuzzy Sets and Systems 1, 3–28, 1978. 18