

AFRL-IF-RS-TR-2005-38
Final Technical Report
February 2005



ON METHODS FOR HIGHER ORDER INFORMATION FUSION

IONA College

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

**AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK**

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2005-38 has been reviewed and is approved for publication

APPROVED: /s/

JOHN J. SALERNO, JR.
Project Engineer

FOR THE DIRECTOR: /s/

JOSEPH CAMERA, Chief
Information & Intelligence Exploitation Division
Information Directorate

REPORT DOCUMENTATION PAGE			<i>Form Approved</i> OMB No. 074-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE FEBRUARY 2005	3. REPORT TYPE AND DATES COVERED Final Oct 01 – Sep 04	
4. TITLE AND SUBTITLE ON METHODS FOR HIGHER ORDER INFORMATION FUSION			5. FUNDING NUMBERS C - F30602-01-1-0593 PE - 62702F PR - FF8B TA - II WU - 08	
6. AUTHOR(S) Ronald R. Yager				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Machine Intelligence Institute IONA College 715 North Avenue New Rochelle New York 10801			8. PERFORMING ORGANIZATION REPORT NUMBER N/A	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory/IFEA 525 Brooks Road Rome New York 13441-4505			10. SPONSORING / MONITORING AGENCY REPORT NUMBER AFRL-IF-RS-TR-2005-38	
11. SUPPLEMENTARY NOTES AFRL Project Engineer: John J. Salerno, Jr./IFEA/(315) 330-3667/ John.Salerno@rl.af.mil				
12a. DISTRIBUTION / AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.				12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 Words) The primary focus of this effort was on the development of theory useful for the intelligent fusion of information and its subsequent use in situation awareness and assessment. This work was focused on the representation of complex and imprecise relationships between objects as well as facility for representing and manipulating various kinds of uncertainties.				
14. SUBJECT TERMS Indicator and Warnings, Situation Awareness, Fuzzy Sets, Fusion				15. NUMBER OF PAGES 103
				16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

TABLE OF CONTENTS

SUMMARY.....	1
PART I: A FRAMEWORK FOR MULTI-SOURCE DATA FUSION.....	3
1. INTRODUCTION	3
2. SOME CONSIDERATIONS IN DATA FUSION.....	3
3. CHARACTERIZING USER REQUIREMENTS.....	9
4. BASIC CONCEPTUAL FRAMEWORK FOR DATA FUSION.....	11
5. SOME COMMON COMPATIBILITY RELATIONSHIPS	13
6. SOME FEATURES ON THE FUNCTION F	16
7. ON MONOTONICITY IN ORDINAL SPACES.....	17
8. CREDIBILITY WEIGHTED SOURCES	19
9. ON THE INCLUSION OF REASONABLENESS	21
10. GRANULAR OBJECTS AS FUSED VALUE	24
11. MULTIPLE GRANULAR OBJECTS AS FUSED VALUES.....	29
12. REFERENCES PART I.....	30
PART II. FUZZY METHODS IN WEB QUESTION ANSWERING SYSTEMS	32
1. INTRODUCTION	32
2. BASIC CONCEPTS OF APPROXIMATE REASONING.....	34
3. SEMANTICS OF AR THE KNOWLEDGE REPRESENTATION	39
4. TRANSLATION OF INFORMATION	42
5. INFERENCE AND KNOWLEDGE MANIPULATION	44
6. PROTOFORMS.....	47
7. KNOWLEDGE TREES	50

8. NON-MONOTONIC POSSIBILISTIC PROPOSITIONS.....	53
9. DEFUZZIFICATION	60
10. DEMPSTER-SHAFFER GRANULES FOR KNOWLEDGE REPRESENTATION	62
11. REFERENCES PART II	65
PART III. UNCERTAINTY MANAGEMENT FOR INTELLIGENCE ANALYSIS	68
1. INTRODUCTION	68
2. VARIABLES AND QUESTION ANSWERING	68
3. BASIC KNOWLEDGE REPRESENTATION USING FUZZY SETS	70
4. ON THE MEASURES OF POSSIBILITY AND CERTAINTY	73
5. HEDGING ON OUR DATA.....	78
6. MULTI-SOURCE INFORMATION FUSION	80
7. MULTIPLE FUSED VALUES FROM MULTI-SOURCE DATA.....	84
8. FUSING PROBABILISTIC AND POSSIBILISTIC DATA.....	86
9. ALTERNATIVE MEASURES OF CERTAINTY	91
10. REFERENCES PART III.....	96

List of Figures

FIGURE 1 SCHEMATIC OF DATA FUSION	4
FIGURE 2. ILLUSTRATION OF SATURATION	18
FIGURE 3A. EFFECT OF SATURATION ON MONOTONICITY	18
FIGURE 3B. EFFECT OF SATURATION ON MONOTONICITY	18
FIGURE 4. TYPICAL KNOWLEDGE TREE	52
FIGURE 5. SUBTREE FROM NON-MONOTONIC PROPOSITIONS.....	60
FIGURE 6. DEFUZZIFICATION SUB TREE	62
FIGURE 7. TASK OF ANSWERING A QUESTION	69
FIGURE 8. DIFFERENT RELATIONS BETWEEN KNOWLEDGE AND QUESTION	73
FIGURE 9. QUANTIFIERS PARAMETERIZED BY β	94

SUMMARY

In part one we look at a framework for the multi-source data fusion process. Some of the considerations and information that must go into the development of a multi-source data fusion algorithm are described. Features that play a role in expressing users' requirements are also described. We provide a framework for data fusion based on a voting like process that tries to adjudicate conflict among the data. We discuss the idea of a compatibility relationship and introduce several important examples of these relationships. We show that our formulation results in some bounding conditions on the fused value implying that the fusion process has the nature of a mean type aggregation. Situations in which the sources have different credibility weights are considered. We present a concept of reasonableness as a means for including in the fusion process any information available other than that provided by the sources. We consider the situation when we allow our fused values to be granular objects such as linguistic terms or subsets.

In part two we consider web based question-answering systems. We point out that question-answering systems differ from other information seeking applications, such as search engines, by having a deduction capability or an ability to answer questions by a synthesis of information residing in different parts of its knowledge base. This capability requires appropriate representation of various types of human knowledge, rules for locally manipulating this knowledge and a framework for providing a global plan for appropriately mobilizing the information in the knowledge base to address the question posed. We suggest some tools to provide these capabilities. We describe how the fuzzy set based theory of approximate reasoning can aid in the process of representing knowledge. We discuss how protoforms can be used to aid in deduction and the local manipulation of knowledge. The idea of a knowledge tree is introduced to provide a global framework for mobilizing the knowledge base in response to a query. We look at some types of commonsense and default knowledge. This requires us to address the complexity of the non-monotonicity that these types of knowledge often display. We also briefly discuss the role that Dempster-Shafer structures can play in representing knowledge.

Considerable concern has arisen regarding the quality of intelligence analysis. This has

been in large part motivated by the task, prior to the Iraq war, of determining whether Iraq had weapons mass destruction. One problem that made this analysis difficult was the uncertainty in much of the information available to the intelligence analysts. In part three we introduce some tools that can be of use to intelligence analysts for representing and processing uncertain information. We make considerable use of technologies based on fuzzy sets and related disciplines such as approximate reasoning.

PART I: A Framework for Multi-Source Data Fusion

1. Introduction

With the rapid growth of the Internet and other electronic sources of information capture, the problem of the coherent merging of information from multiple sources has become an important issue. This problem has many manifestation ranging from data mining to information retrieval to multi-sensor fusion [1-4]. Here we consider one type of problem from this class. We shall consider the situation in which we have some attribute variable, whose value we are interested in supplying to a user. In our situation we have multiple sources providing data values for this variable as input to our system. We call this a data fusion problem. Even this limited example of multi-source information fusion is a very expansive problem. Here we shall restrict ourselves to the discussion of some issues related to this problem.

2. Some Considerations in Data Fusion

The process of data fusion is initiated by a request for information about an attribute variable to our sources of information. Let V be a attribute variable whose value lies in the set X , called the inverse of V . We assume a collection $S_1, S_2, \dots, S_q$ 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 fused 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. In figure 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.

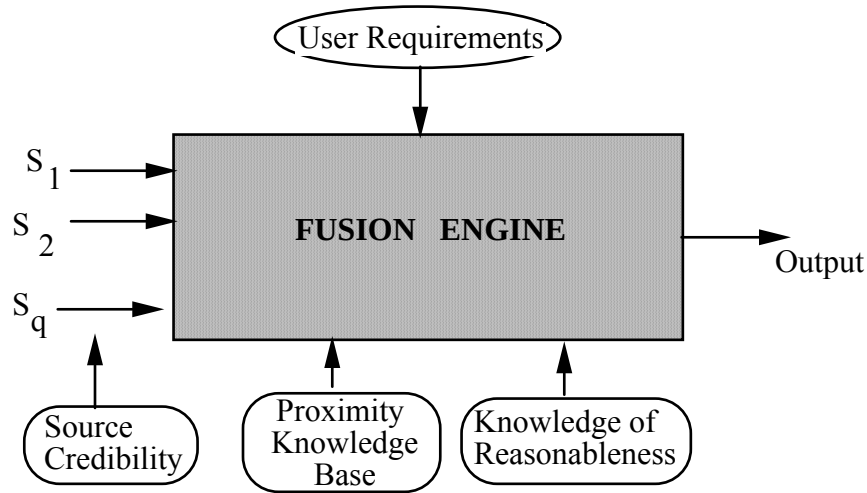


Figure 1 Schematic of Data Fusion

Let us begin to look at some of the considerations that effect the mechanism that is 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, in 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 the fuzzy measure [5, 6] and the Dempster-Shafer belief functions [7, 8] provide two general frameworks for representing uncertainty information. In the following unless otherwise specified, we shall assume the information provided by a source is a specific value in the space X .

Another consideration is the inclusion of source credibility information in the process. Source credibility is a user generated or sanctioned knowledge base. It associates with the data provided by the source a weight indicating its credibility. This credibility weight must be drawn from a scale that can be either ordinal or numeric. That is, there must be some ordering of the credibility values. The process of assigning the credibility to the data reported by a source can 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 variables involved. For example, one source may be very reliable with information about ages while not very good about information about a person's income. Even more sophisticated

distinctions can be made, for example, a source could be good about information on high income people but bad about low income people.

As we noted the information about credibility must be at least ordered. It may or may not be expressed using a well defined bounded scale. For example, the credibility information may be expressed by ordering of the sources themselves. Alternatively the credibility of a source may be expressed in terms of the assignment of a value drawn from a well defined scale having an upper and lower element. Generally in the case when the credibility is selected from such a scale the assignment of the upper value to a source indicates that the data should be given full weight. An assignment of the lowest value generally means don't use the value, this implies the value should have no or little influence in the fusion process. In some ways the association of a less than complete credibility with a source is closely related to a kind of uncertainty assigned to the value provided by the source.

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 value. This is somewhat akin to the relationship between false and complementation in logic.

Central to the problem of data fusion is the issue of conflict and its resolution. The proximity knowledge base and the reasonableness knowledge base play important roles in the handling of this issue.

One obvious form of conflict arises when we have multiple readings of a variable that may not be the same or even compatible. For example, one source may say Osema Bin Laden is 35 years old, another may say he is 45 and another may say he is 55. 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 can incur even when we only have a single reading for a variable. This can occur when a reported reading conflicts with what we know to be the case, what is reasonable. Assume when searching for the age of Osema Bin Laden one of the sources report that he is eighty years old. This conflicts with what we know to be reasonable.

This is information that 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 is what we have indicated as the knowledge of reasonableness in figure 1. This type of a priori context or domain knowledge can take many forms and can be represented in different ways.

As an illustration of one method of handling this type of domain knowledge is where we assume our knowledge base is 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 . Typically, T is the unit interval $I = [0, 1]$. In this case $R(x) = 1$ 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 $T = \{t_1, \dots, t_n\}$ which has a largest and smallest element, 0 and 1. R can be viewed as a fuzzy subset of reasonable values.

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 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, $g(t, r) \geq f(\hat{t}, \hat{r})$ if $t \geq \hat{t}$ and $r \geq \hat{r}$. 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 can be associated with this operation is symmetry, $g(t, r) = g(r, t)$. Although we may or may not 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 $g(t, r) \leq \text{Min}[t, r]$.

Relationships conveying information about the congeniality¹ 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 common example is the use of least squared based methods. The use of linguistic concepts and other granulation techniques are based on these relationships [9, 10]. Clustering operations require these relationships. These relationships are related to equivalence relationships and metrics. Central to these relationships are the properties of reflexivity (a value is congenial to its self) and symmetry.

For our purposes we shall find the concept of a **proximity relationship** [11] useful in discussing data fusion. Formally, a proximity relationship on a space X is a mapping $\text{Prox}: X \times X \rightarrow T$ having the properties: (1). $\text{Prox}(x, x) = 1$ (**reflexive**) and (2). $\text{Prox}(y, x) = \text{Prox}(x, y)$ (**symmetric**). Here T is ordered space having largest and smallest elements denoted 1 and 0. Often T is the unit interval. Intuitively the value $\text{Prox}(x, y)$ is some measure of degree to

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

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 [12, 13]. Formally, a similarity relationship on a space X is a mapping $\text{Sim}: X \times X \rightarrow T$ having the properties

- | | |
|---|-------------------|
| 1) $\text{Sim}(x, x) = 1$ | reflexive |
| 2) $\text{Sim}(x, y) = \text{Sim}(y, x)$ | symmetric |
| 3) $\text{Sim}(x, z) \geq \text{Sim}(x, y) \wedge \text{Sim}(y, z)$ | transitive |

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 has its value taken from a numeric scale then the bases of the proximity relationship is the difference $|x - y|$. However the mapping of $|x - y|$ into $\text{Prox}(x, y)$ may be highly non-linear.

For attributes assuming non-numeric values a relationship of proximity is based on relevant features associated with the elements in the variables universe. Here we can envision a variable having multiple appropriate 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 countries 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 or matrimonial** nature of the values of the variable V . This characteristic plays an important role in the types of operations that are available to use during the fusion process. It determines what types of aggregations we can perform. This concept is closely related to the idea of scale and measurement, a topic studied extensively in the field of mathematical psychology. For our purposes we shall distinguish between three types of variables with respect to this characteristic. The first type of variable is what we shall call *celibate*. The word *nominal* can be used here. 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 this is completely inappropriate. More formally these are variables whose universe has no mathematical structure. These variables only allow comparison as to whether they are equal.

A more structured type of variable is an ordinal variable. For these types of variables there exists some kind of meaningful ordering of the members of the universe. An example of this is a variable corresponding to size that has as its universe {small, medium, large}. For these variables some kind of compositional process is allowable, 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 a large body of compositional operators.

3. Characterizing 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 output 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 closely related issue relates to the level of granulation. As described by Zadeh [14] 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 variable. In providing information we must satisfy the user's level of granularity for the task at hand. 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 we use a vocabulary of linguistic terms to express solutions. An example is where using a term such as "cold" as the value for the temperature would be acceptable.

Another issue related to the form of the output is whether all output values presented to the user are required to be values that appear in the input or can we blend output 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. In addition, we should note 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 discussed earlier input values 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. Basic Conceptual Framework for 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 V having an underlying universe X . Consider the situation in which we have a collection of q assessments of the variable, inputs to our fusion engine. Each assessment is information supplied by one of our sources. Let a_i be the value provided by the i^{th} source. Our desire here is to fuse these values to obtain some value $\tilde{a} \in X$ as the fused value. We denote this as $\tilde{a} = \text{Agg}(a_1, \dots, a_n)$. 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 $\mathbf{a}$ be a proposed solution, some element from X . Each source can be seen as "voting" whether to accept this solution. Let us denote $\text{Sup}_i(\mathbf{a})$ as the support for solution $\mathbf{a}$ from source i . We then need some process of combining the support for $\mathbf{a}$ from each of the sources. We let

$$\text{Sup}(\mathbf{a}) = F(\text{Sup}_1(\mathbf{a}), \text{Sup}_2(\mathbf{a}), \dots, \text{Sup}_q(\mathbf{a}))$$

be the total support for $\mathbf{a}$. Thus F is some function that combines the support from each of the sources. The aggregated value $\tilde{a}$ is then obtained as the value $a \in X$ that maximizes $\text{Sup}(\mathbf{a})$. Thus $\tilde{a}$ is such that $\text{Sup}(\tilde{a}) = \text{Max}_{a \in X}[\text{Sup}(\mathbf{a})]$.

One natural property associated with F is that the more support from the individual sources the more overall support for $\mathbf{a}$. Formally if $\mathbf{a}$ and $\mathbf{b}$ are two values and if $\text{Sup}_i(\mathbf{a}) \geq \text{Sup}_i(\mathbf{b})$ for all i then $\text{Sup}(\mathbf{a}) \geq \text{Sup}(\mathbf{b})$. This requires that F be a monotonic function, $F(x_1, x_2, \dots, x_q) \geq F(y_1, y_2, \dots, y_q)$ 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 $F(x_1, \dots, x_q) > F(y_1, \dots, y_q)$.

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 $t_1, \dots, t_q$ and $\hat{t}_1, \dots, \hat{t}_q$ are two sets of arguments of F , $\text{Sup}_i(\mathbf{a}) = t_i$ and $\text{Sup}_i(\hat{\mathbf{a}}) = \hat{t}_i$. Let perm indicate a permutation of the arguments, where $\text{perm}(i)$ is the index of the i^{th} 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

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

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

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 in the fuzzy set literature [11, 15, 16]. Here then we shall assume a relationship Comp , called the compatibility relationship, which has at least the properties of a proximity relationship. Thus $\text{Comp}: X \times X \rightarrow T$ 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 $a_i = \mathbf{a}$ for all i . In this case $\text{Comp}(\mathbf{a}, a_i) = 1$ for all i . From this it follows that for any $b \in X$ $\text{Comp}(\mathbf{a}, a_i) \leq \text{Comp}(\mathbf{b}, a_i)$ hence $\text{Sup}_i(\mathbf{a}) \geq \text{Sup}_i(\mathbf{b})$ for all $\mathbf{b}$ thus $\text{Sup}(\mathbf{a}) \geq \text{Sup}(\mathbf{b})$ for all $\mathbf{a}$. Thus there can never be a better solution than $\mathbf{a}$. 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.

We now introduce the idea of a solution set and the related idea of minimal solution set. We say that a subset G of X is a solution set if all $\mathbf{a}$ s.t. $\text{Sup}(\mathbf{a}) = \text{Max}_{a \in X}[\text{Sup}(a)]$ are contained in G . We shall say that a subset H of X is a **minimal solution** set if there always exists one element $\mathbf{a} \in H$ s.t. $\text{Sup}(\mathbf{a}) = \text{Max}_{a \in X}[\text{Sup}(a)]$. Thus a minimal solution set is a set in which we can always find an acceptable fused value.

5. Some Common Compatibility Relationships

We now look at some very important special examples of compatibility relationships and show that in these cases there exists some minimal solution sets which are easily definable in terms of the data to be fused. These minimal solution sets can be seen as boundaries on the subset of elements of X in which we look to find a solution. These boundaries seem to reflect a general idea that the aggregation procedure has the nature of a mean (averaging) operator.

The first situation we consider is a very strong compatibility requirement. Here 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.

Let B be the subset of X containing the input data values, $B = \{b \mid \text{s.t. } \exists a_i = b\}$. Let d be some element not in B , $d \in X - B$. In this case $\text{Comp}(d, a_i) = 0$ for a_i . Let b be some element from B , $b \in B$. Here there exists at least one a_i s.t. $\text{Comp}(b, a_i) = 1$. Thus we see that $\text{Comp}(b, a_j) \geq \text{Comp}(d, a_j)$ for all j and there exists at least one i such that $\text{Comp}(b, a_i) > \text{Comp}(d, a_i)$. Hence $\text{Sup}_i(b) \geq \text{Sup}_i(d)$ for all i and at least one $\text{Sup}_i(b) > \text{Sup}_i(d)$. From this we see that $\text{Sup}(b) \geq \text{Sup}(d)$. Furthermore if F is strictly monotonic then $\text{Sup}(b) > \text{Sup}(d)$. Since $\tilde{a}$ is the element which has the largest Sup then an implication of this is that the aggregated value should be an element from the collection of input values, in particular $\tilde{a} \in B$. Here B is a minimal solution set. This can be seen as a kind of boundedness condition on the aggregated value, $\tilde{a}$

must be a value in the set B of input values. This is of course a very natural value. We note that idempotency is also assured here - in the case of idempotency all the elements $a_i = a$ and hence $B = \{a\}$.

We next show that if Comp is any equivalence relationship then B still provides a minimal solution set-no solution can be better than some element in B . We recall if Comp is an equivalence relationship then for each $x \in X$ there exists a subset E_x of X , called the equivalence of x , such that $\text{Comp}(x, y) = 1$ for all $y \in E_x$ and for any $z \notin E_x$ we have $\text{Comp}(x, z) = 0$. Let $B^* = \bigcup_{x \in B} E_x$, the union of all equivalence class corresponding to elements in B . If $z \in X - B^*$ then $\text{Comp}(z, a_i) = 0$ for all i . Since for any $b \in B^*$ there always exists some a_i such that $\text{Comp}(b, a_i) = 1$ hence $\text{Sup}(b) \geq \text{Sup}(z)$. Thus there never exists an element $z \in X - B^*$ better than any element in B^* . Furthermore if F is strictly monotonic then any element in B^* is always better than any $z \in X - B^*$. Thus we need never look for a solution in $X - B^*$. Furthermore we see the following. Let $a \in B^*$ but $a \notin B$, thus a is in the equivalence class of some element in B . Let $a \in E_b$ where $b \in B$. We see that for any a_i , $\text{Comp}(b, a_i) = \text{Comp}(a, a_i)$ hence $\text{Sup}_i(b) = \text{Sup}_i(a)$. Thus there exists no element better than some element in B . Then B is a minimal solution. Thus we can bound our search for solution by B .

We now turn to another type of compatibility relationship - the situation in which there exists some linear ordering on the space X that underlies the compatibility relation. In particular we let L be a linear ordering on X . We shall use the notation $x \underset{L}{>} y$ to indicate that x is higher² in the ordering than y . We now require that our compatibility relation, in addition to being reflexive and symmetric, be such that the closer the elements are in the ordering L the more compatible they are. Thus we have that if $x \underset{L}{>} y \underset{L}{>} z$ then $\text{Comp}(x, y) \geq \text{Comp}(x, z)$. We say this connection between ordering and compatibility is strict if $x \underset{L}{>} y \underset{L}{>} z$ implies $\text{Comp}(x, y) > \text{Comp}(x, z)$.

We now show that in this situation there also exists a boundedness condition. Let $a^* = \mathbf{Max}_L[a_i]$, a^* is the largest element of the input values with respect to the underlying ordering $\underset{L}{>}$. Let $a_* = \mathbf{Min}_L[a_i]$ the smallest element in B with respect to the ordering. Then we

² At times we shall find it more natural to use the term larger in place of higher.

shall see that we can always find a solution $\tilde{a}$ satisfying

$$a_* \underset{L}{>} \tilde{a} \underset{L}{>} a^*$$

We see this as follows. Assume $\tilde{b} \underset{L}{<} a_*$, $\tilde{b}$ is below a_* in the ordering L , then for any input a_i we have $a_i \underset{L}{>} a_* \underset{L}{>} \tilde{b}$ hence $\text{Comp}(a_i, a_*) \geq \text{Comp}(a_i, \tilde{b})$. From this it follows that $\text{Sup}(a_*) \geq \text{Sup}(\tilde{b})$. Thus a_* is at least as good as any smaller $\tilde{b}$. Similarly, if $\tilde{d} \underset{L}{>} a^*$ then for any input data a_i we have $a_i \underset{L}{>} a^* \underset{L}{>} \tilde{d}$ therefore for all a_i we have $\text{Comp}(a_i, a^*) \geq \text{Comp}(a_i, \tilde{d})$. From this we get that $\text{Sup}(a^*) \geq \text{Sup}(\tilde{d})$. Thus we see the subset of X lying between a_* and a^* , with respect to the ordering L , provides a minimal solution set. This is the usual boundary condition associated with mean aggregation operators. Thus we are beginning to see that Agg must be a mean-like operator.

Let us now consider the situation in which our underlying compatibility relationship is a similarity relationship [12]. That is, we have the additional property that for all x, y and z

$$\text{Comp}(x, z) \geq \text{Max}_y [\text{Comp}(x, y) \wedge \text{Comp}(y, z)].$$

An important feature of any similarity relationship is the following [12]. For every $\alpha \in T$ there exists a partitioning of X into classes E_j^α , $X = \bigcup_j E_j^\alpha$ and $E_i^\alpha \cap E_j^\alpha = \emptyset$ for $i \neq j$, such that for all

x and $y \in E_j^\alpha$ we have $\text{Comp}(x, y) \geq \alpha$ and for all $z \notin E_j^\alpha$ and $x \in E_j^\alpha$ we have $\text{Comp}(x, z) < \alpha$.

Let B be our collection of observations $B = \bigcup_i \{a_i\}$. Let α be the minimal similarity between any elements in B , $\alpha = \text{Min}_{a_i, a_j \in B} [\text{Comp}(a_i, a_j)]$. Let the collection of subsets E_j^α be the α -level partitioning of X . This exists one member of this partitioning containing B . Let $E_{j^*}^\alpha$ be the member of the partitioning class containing B . Here then $B \subseteq E_{j^*}^\alpha$ and $E_j^\alpha \cap B = \emptyset$ for all $j \neq j^*$. We shall now show that $E_{j^*}^\alpha$ provides a minimal solution set. We see this as follows. If $z \in E_{j^*}^\alpha$ then $\text{Comp}(z, y) \geq \alpha$ for all $y \in E_{j^*}^\alpha$. Since $B \subseteq E_{j^*}^\alpha$ then $\text{Comp}(z, a_i) \geq \alpha$ for all input. Assume $x \notin E_{j^*}^\alpha$ then $\text{Comp}(x, y) < \alpha$ for all $y \in E_{j^*}^\alpha$. Since $B \subseteq E_{j^*}^\alpha$ then $\text{Comp}(x, a_i) < \alpha$ for all i . From this it follows $\text{Sup}_i(z) \geq \text{Sup}_i(x)$ and hence $\text{Sup}(z) \geq \text{Sup}(x)$. Thus we see a element not in $E_{j^*}^\alpha$ can't score better then an element in $E_{j^*}^\alpha$, thus $E_{j^*}^\alpha$ provides a minimal solution set when our Comp relationship is a similarity relationship.

Finally we introduce a trivial example of a proximity relationship. We shall say that a proximity relationship has a **wild card** if there exists an element $x \in X$ having the property

$\text{Prox}(x, y) = 1$ for all $y \in X$. In this case x is called a wild card. Clearly if a proximity relationship has a wild card it is always a best fused value, it provides a minimal solution set.

6. Some Features on the Function F

We described the process of determining the fused value to a data collection $\langle a_1, \dots, a_q \rangle$ 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 $\text{Sup}(a) = F(\text{Sup}_1(a), \dots, \text{Sup}_q(a))$
- 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 . The first was that we assumed that F is symmetric with respect to the arguments, the indexing of input information is not relevant. The second assumption we made about F was that it is monotonic with respect to the argument its values. An implicit assumption we made about F was an assumption of pointwiseness. Here the determination of the valuation of any solution $\tilde{a}$, $V(\tilde{a})$ depends only on $\tilde{a}$, it is independent of any of other possible solutions. This property imposes the condition of indifference to irrelevant alternatives, a requirement that plays a central role in two fundamental works on opinion aggregation Arrow's impossibility theorem [17] and Nash's bargaining problem [18]. Specifically this condition assures us that the inclusion or removal of possible solutions, elements from X doesn't affect the relationship between other solutions. Essentially this assures that if $\text{Sup}(x) \geq \text{Sup}(y)$ then this will not change if we add or remove another element z from the space X .

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 [19]. Assume that we have a data set $\langle a_1, \dots, a_q \rangle$ 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} for which $a_{q+1} = \tilde{a}$, 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 solutions and let $\tilde{c}_i = \text{Comp}(\tilde{a}, a_i)$ and $\hat{c}_i = \text{Comp}(\hat{a}, a_i)$. We note that if $a_{q+1} = \tilde{a}$ then $\tilde{c}_{q+1} = \text{Comp}(\tilde{a}, a_{q+1}) \geq \tilde{c}_i$ for all $i = 1$ to

q and $\tilde{c}_{q+1} \geq \hat{c}_i$ for $i = 1$ to $q + 1$. Using this we can more formally express our additional requirement on F . If

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

and if $\tilde{c}_{q+1} \geq \tilde{c}_j$ for $j = 1$ to q and $\tilde{c}_{q+1} \geq \hat{c}_j$ for $j = 1$ to $q + 1$ then we require that

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

Our work so far assumed a very general formulation for F . 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, $F(x_1, x_2, \dots, x_q) = \sum_{i=1}^q x_i$.

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.

What becomes clear here is the fused value depends very strongly on the compatibility relationship. Let us consider the special situation in which our variable takes as its value numbers and the compatibility between elements is directly related to the distance between the elements. Here then $\text{Comp}(x, y) = 1 - \frac{D(x, y)}{D_{\max}}$ where $D_{\max}$ is the largest distance. Using this compatibility relationship we get $\text{Sup}(a) = \sum_{i=1}^q \frac{D_{\max} - D(a, a_i)}{D_{\max}}$. The fused value, $\tilde{a}$, is the value that maximizes $\text{Sup}(a)$. Here this is the value $\tilde{a}$ that minimizes $\sum_{i=1}^q D(a, a_i)$. If $D(a, a_i)$ is taken as the square difference, $D(a, a_i) = (a - a_i)^2$, then it is easy to show that $\tilde{a} = \frac{1}{q} \sum_{i=1}^q a_i$, is simply the average.

Thus we see using the average of the data values is a special case under particular assumptions for F and Comp . We should point out that this assumption about the compatibility relationship may not always be the appropriate assumption and hence the use of the average may not always be appropriate.

7. On Monotonicity in Ordinal Spaces

We again return to the case in which the data is drawn from a set which has an ordering that is used to generate the proximity relationship. In the preceding sections we showed that in this case our fused value $\tilde{a}$ must be bounded $\text{Min}_i[a_i] \leq \tilde{a} \leq \text{Max}_i[a_i]$. Let us now look a condition

of monotonicity of aggregated value **usually** associated with this environment: if $\hat{a}_j \geq a_j$ for all j then this monotonicity property requires that $\text{Agg}(\hat{a}_1, \dots, \hat{a}_q) \geq \text{Agg}(a_1, \dots, a_q)$. We shall see that our formulation doesn't impose this condition. This is as it should be in a general formulation. For this generality allows us to include "mode-like" type fusion methods which are known not to be monotonic in the sense described above.

Assume the data set has been indexed such that $a_1 \leq a_2 \leq \dots \leq a_q$. Let its fused value be $\tilde{a}$. Let $\hat{a}_j$ be another collection of q values such that $\hat{a}_j = a_j$ for all $j \neq k$ and having $\hat{a}_k > a_k$. Thus all the elements in $\hat{a}_j$ are equal to those in a_j except the k^{th} . Data value monotonicity requires us to show that if $\text{Agg}(\hat{a}_1, \dots, \hat{a}_q) = \hat{a}$ then $\hat{a}$ can **not** be less than $\tilde{a}$. Let us see what is needed to assure this requirement.

It appears that one feature that leads to difficulty in attaining this type of monotonicity are proximity relationships that manifest **saturation**. We refer to figure 2 to understand this idea.

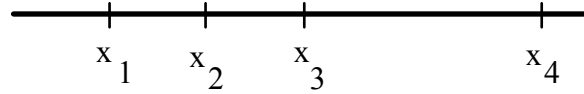


Figure 2. Illustration of Saturation

If $\text{Prox}(x_1, x_3) < \text{Prox}(x_2, x_3)$ and $\text{Prox}(x_1, x_4) = \text{Prox}(x_2, x_4)$ then we have saturation. Here x_4 has moved so far away from x_1 and x_2 that their proximity to x_4 has become about the same. To illustrate how this saturation effect interferes with the attainment of monotonicity consider the following illustration (figure 3a).

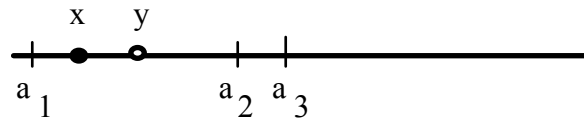


Figure 3a. Effect of Saturation on Monotonicity

In the above let a_1 , a_2 and a_3 be three pieces of data and let x and y be two proposed fused values. Assume here that the F and the Prox relationship are such that y is a better solution than x . This is essentially based on the fact that y is closer to the data points a_2 and a_3 . Now consider figure 3b.

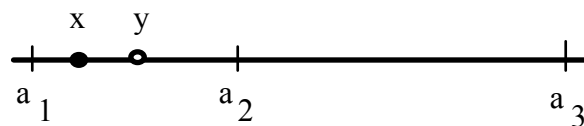


Figure 3b. Effect of Saturation on Monotonicity

In figure 3b we have increased the value of a_3 . If this increase is of such a nature that it

eliminates any distinction between the compatibility of a_3 with x or y then comparison of x and y as possible fused values is simply based on there relationships with a_1 and a_2 . Since x is closer to a_1 then y is to a_2 the possibility arises for x to be a better choice for fused value the y . Thus we see that in the face of this saturation, an increase in data value can result in a decrease in fused value.

Partitioning type equivalence relationships are particular examples of this type of saturating proximity relationships. Let X be an ordered space $x_1 > x_2 > \dots > x_n$. Assume Prox is an equivalence relationship in which A_j are the equivalence classes, $X = \bigcup_i A_j$. Let $x_1 < x_{k_1} < x_{k_3}$. Assume x_1 and $x_{k_1} \in A_1$ and $x_{k_3} \in A_3$. Then $\text{Prox}(x_1, x_{k_1}) = 1$ and $\text{Prox}(x_1, x_{k_3}) = 0$, $\text{Prox}(x_{k_1}, x_{k_3}) = 0$. Assume now x_{k_1} increase to $x_{k_2} \in A_2$. Now we $\text{Prox}(x_1, x_{k_2}) = 0$, $\text{Prox}(x_{k_2}, x_{k_3}) = 0$ and $\text{Prox}(x_{k_3}, x_{k_3}) = 0$.

As we already noted this lack of universally requiring monotonicity with respect to data values allows us to include different types of fusion techniques in our framework. A notable example of a non-monotonic type fusion operator is the mode. Consider the case where we have observations 10, 10, 10, 15, 15, 15, 15. The mode is 15. Assume a data set 10, 10, 10, 15, 15, 20, 20. We have increased two of the fifteens to 20. Our mode becomes 10.

8. 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 (our input are 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 $\text{Sup}(a) = F(\text{Sup}_1(a), \dots, \text{Sup}_q(a))$ where $\text{Sup}_i(a)$ is the support for 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 a 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 $((w_1, a_1), \dots, (w_q, a_q))$ has as its fused value $\tilde{a}$ then the data $((w_1, a_1), \dots, (w_q, a_q), (w_{q+1}, a_{q+1}))$ 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.

We now show that the boundary conditions also hold in this case where the sources have weights. First let our situation be one in which our Comp relationship is strict, it is such that $\text{Comp}(a, b) = 1$ if $a = b$ and $\text{Comp}(a, b) = 0$ if $a \neq b$. Again let $B = \bigcup_i \{a_i\}$, the set of all the values provided by the sources. If $a \notin B$ then for all i we have $\text{Comp}(a, a_i) = 0$ and from the above requirements we get $\text{Sup}_i(a) = 0$. Let $b \in B$ then $\text{Comp}(b, a_i) = 1$ for all i such that $a_i = b$. Hence $\text{Comp}(b, a_i) \geq \text{Comp}(a, a_i)$ for all i and therefore $\text{Sup}_i(b) \geq \text{Sup}_i(a)$ for all i and hence $\text{Sup}(b) \geq \text{Sup}(a)$. Therefore we can always find the fused solution in B .

In a similar way we can show that when the proximity relationship is based on an ordering and the input data have weights it is the case the fused value must lie between the largest and smallest input values. The justification of this is based on the monotonicity of g with respect to the compatibility. In particular if $\text{Min}_i[a_i] = a_1$ and if $a < a_1$ then $\text{Comp}(a, a_j) \leq \text{Comp}(a_1, a_j)$ for all j hence $g(w_j, \text{Comp}(a, a_j)) \leq g(w_j, \text{Comp}(a_1, a_j))$ and hence $\text{Sup}_j(a) \leq \text{Sup}_j(a_1)$ for all j . In a

similar way we can shown that if $\text{Max}_j[a_i] = a_q$ and if $a > a_q$ then $\text{Sup}_j(a) \leq \text{Sup}_j(a_q)$ for all j .

9. On the Inclusion of Reasonableness

In an earlier part we introduced the idea of reasonableness and indicated its importance in the data fusion process. At a meta level we mean to use this idea to introduce any information, exclusive of the data provided by the sources, we may have about the value of variable of interest. The information we have about reasonableness will affect the 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 and hence its importance in the fusion process reduced. 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 representation and inclusion of reasonableness in the data fusion process is complex and beyond our immediate aim as well as beyond our complete understanding at this time. In many ways the issue of reasonableness goes to the very heart of intelligence. We leave a more expansive discussion of this concept to some future work. Here we shall focus on the representation of some very specific type of knowledge and its role in introducing a consideration of reasonableness in the fusion process.

Information about the reasonableness of values of a variable of interest can be either pointed or diffuse. By pointed we mean information specifically about the object while diffuse information is about objects of a class in which our object of interest lies. Generally, pointed information has a possibilistic nature, while diffuse information is probabilistic. Here we consider the situation in which our information about reasonableness is pointed and captured by a fuzzy subset, a mapping $R: X \rightarrow T$ and thus has a possibilistic nature. Here for any $x \in X$, $R(x)$ indicates the reasonableness (or possibility) that x is a solution of variable of interest. For example, if our interest is to obtain a person's age and before soliciting data from external sources we know that the person is *young* then we can capture this information with R and thus constrain the values that are reasonable [20, 21]. In the above we assume T is a linear ordering having maximal and minimal elements, usually denoted 1 and 0.

Let us see how we can include this information into our data fusion process. Assume the data provided by source i is denoted a_i and w_i is the credibility assigned to source i . We shall assume these credibilities are measured on the same scale as the reasonableness. 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 consideration in the fusion as is the case for a piece of data coming from a source with low credibility whatever the reasonableness of its data. 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 $\langle (u_1, a_1), \dots, (u_q, a_q) \rangle$.

As in the preceding, the support for a proposed fused value $\mathbf{a}$, should be a function of its support from input data:

$$\text{Sup}(\mathbf{a}) = F(\text{Sup}_1(\mathbf{a}), \dots, \text{Sup}_q(\mathbf{a}))$$

The support provided from source i , for solution $\mathbf{a}$, 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, $\text{Comp}(\mathbf{a}, a_i)$. In addition, we must also include information about the reasonableness of the proposed solution $\mathbf{a}$. For a solution $\mathbf{a}$ to get support from source i it should be compatible with the data a_i and compatible with what we consider to be reasonable, $\text{Comp}(\mathbf{a}, R)$; we let $\text{Comp}_i(\mathbf{a}) = \text{Comp}(\mathbf{a}, a_i) \wedge \text{Comp}(\mathbf{a}, R)$. Furthermore, $\text{Comp}(\mathbf{a}, R) = R(\mathbf{a})$ and $\text{Comp}_i(\mathbf{a}) = \text{Comp}(\mathbf{a}, a_i) \wedge R(\mathbf{a})$. 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 $\text{Sup}_i(\mathbf{a}) = g(u_i, \text{Comp}_i(\mathbf{a}))$. Substituting our values we get:

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

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 $\text{Sup}_i(\mathbf{a}) = 0$. In the case in which we interpret g as implementing an *anding* and using the Min operator as our *and* we get $\text{Sup}_i(\mathbf{a}) = w_i \wedge R(a_i) \wedge R(\mathbf{a}) \wedge \text{Comp}(\mathbf{a}, a_i)$. 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.

Let us see how the introduction of reasonableness affects our results about boundedness and minimal solution sets.

Consider the case in which our underlying proximity relationship is very precise, $\text{Comp}(x, y) = 1$ iff $x = y$ and $\text{Comp}(x, y) = 0$ $x \neq y$. Let B be the set of input values and let $\hat{B}$ be the subset of B such that $b \in \hat{B}$ if $R(b) \neq 0$. If $a \notin \hat{B}$ then $\text{Comp}(a, a_i) = 0$ for all a_i and hence $\text{Sup}_i(a) = 0$ for all i . Let $d \in B - \hat{B}$, here $R(d) = 0$ and again we get that $\text{Sup}_i(d) = 0$ 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 $\text{Sup}_j(b) > 0$. 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.

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 $y \in E_i$, $\text{Prox}(y, a_i) = 1$. Let $E = \bigcup_i E_i$, the union of all equivalence classes that have input value. If $a \notin E$ then $\text{Prox}(a, a_i) = 0$ for all i . From this we see that if $a \notin E$ then $\text{Sup}_i(a) = 0$ 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 $R(d_i) = \text{Max}_{x \in E_i}(R(x))$. 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 , $\text{Comp}(e_i, a_j) = \text{Comp}(d_i, a_j)$. Since $R(d_i) \geq R(e_i)$ we see that $\text{Sup}_j(d_i) \geq \text{Sup}_j(e_i)$ 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 $D = \bigcup_i D_i$. Furthermore, if x and $y \in D_i$ then $R(x) = R(y)$ and $\text{Comp}(x, z) = \text{Comp}(y, z)$ for all z . Hence $\text{Sup}_i(x) = \text{Sup}_i(y)$ and $\text{Sup}(x) = \text{Sup}(y)$. The result is that we can consider any element in D_i . Thus all we need consider is the set $\tilde{D} = \bigcup_i \{\tilde{d}_i\}$ 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 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 L . Let H be the set of x_j so that $x^* \underset{L}{>} x_j \underset{L}{>} x_*$. In the preceding we showed that we can always find a fused value element 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 $\text{Sup}_i(a) = g(u_i, \text{Comp}_i(a))$ where g is monotonic in both the arguments, $u_i = w_i \wedge R(a_i)$ and $\text{Comp}_i(a) = R(a) \wedge \text{Comp}(a, a_i)$. We shall now show that here we can have an element $a \notin H$ in which $\text{Sup}_i(a) \geq \text{Sup}_i(b)$ 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: $H = \{x \mid x_L^* > x > x_*\}$ where let $\alpha_* = R(x_*)$ and let $\alpha^* = R(x^*)$. Let $H^* = \{x \mid x > x_L^* \text{ and } R(x) > R(x^*)\}$ and $H_* = \{x \mid x > x_* \text{ and } R(x) > R(x_*)\}$. 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 $x > x_L^*$ 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 $x > x_L^*$ having $R(x) \leq R(x^*)$. Using similar arguments we can eliminate $x > x_*$ which have $R(x) \leq R(x_*)$.

10. Granular Objects as Fused Value

In the preceding sections we considered the situation in which we were required to find, as the fused value, some solution that was an element of the set X . We now look at the situation in which we allow our solution to be some subset of X . The use of subsets as our fused value is an example of what Zadeh [9] calls granulation. For simplicity we initially will not include any considerations of reasonableness and credibility.

Again assume V can take its value in X . Let our data be the collection a_i for $i = 1$ to q . These are the values we desire to fuse. Here we assume the existence of a proximity relationship

Prox on $X \times X$ where $\text{Prox}(x, y)$ takes its value in an ordered space T which for simplicity in the following we shall assume to be the unit interval.

Let A be any subset of X . Using the approach introduced in the preceding we define the support for A as the fused value

$$\text{Sup}(A) = F(\text{Sup}_1(A), \text{Sup}_2(A), \dots, \text{Sup}_q(A))$$

As in the preceding we then select as our fused value the subset A having largest value for Sup .

Here $\text{Sup}_i(A)$ is the support for A from data a_i it is the compatibility of a_i with A which we obtain as

$$\text{Sup}_i(A) = \text{Max}_{y \in A} [\text{Comp}(y, a_i)],$$

the maximal compatibility of a_i with any element in A . We note that if A is a fuzzy subset we can express this as

$$\text{Sup}_i(A) = \text{Max}_{y \in X} [A(y) \wedge \text{Comp}(y, a_i)]$$

Recalling that $\text{Comp}(y, a_i)$ is the support for y from a_i we can express $\text{Sup}_i(A)$ as

$$\text{Sup}_i(A) = \text{Max}_{y \in X} [A(y) \wedge \text{Sup}_i(y)]$$

The extension of this situation to include considerations of credibility of the data, w_i and a reasonableness function over X is rather straightforward. Here we recall

$$\text{Sup}_i(y) = g(u_i, R(y) \wedge \text{Comp}(y, a_i))$$

where $u_i = w_i \wedge R(a_i)$. Using this we can get

$$\text{Sup}_i(A) = \text{Max}_{y \in X} [A(y) \wedge g(u_i, R(y) \wedge \text{Comp}(y, a_i))]$$

If we further assume that g is implemented using an *anding*,

$$g(u_i, R(y) \wedge \text{Comp}(y, a_i)) = u_i \wedge R(y) \wedge \text{Comp}(y, a_i)$$

we get that

$$\text{Sup}_i(A) = u_i \wedge \text{Max}_{y \in X} [A(y) \wedge R(y) \wedge \text{Comp}(y, a_i)]$$

In the following discussion we shall initially neglect considerations of credibility and reasonableness, we assume $w_i = 1$ for all i and $R(x) = 1$ for all x . In this case $\text{Sup}_i(A) = \text{Max}_{y \in X} [A(y) \wedge \text{Comp}(y, a_i)]$.

Usually when we allow subsets as the fused value there are some constraints on which subsets we can use. First however, let us consider the situation in which we have no explicitly stated restriction on which subsets we can use as the fused value. Let B be the subset of all the

data, $B = \{a_1, \dots, a_q\}$. Here $\text{Sup}_i(B) = \text{Max}_{y \in B} [\text{Comp}(y, a_i)]$ and since $a_i \in B$ then $\text{Sup}_i(B) = 1$ for all i . From this it follows that for any other subset A of X , $\text{Sup}_i(B) \geq \text{Sup}_i(A)$ hence $\text{Sup}(B) \geq \text{Sup}(A)$. Thus, if we have no restrictions on which subsets we can use as the fused value, the best solution is always the subset consisting of all the input data values. We note that even in the case of introducing differing credibilities associated with the sources the set B is still the best answer. This however may not be the case when we consider reasonableness.

Let us now turn to the more usual situation in which there are some constraints on which subsets we allow as the fused value. Sometimes these constraints are explicit, other times they may implicit.

A number of methods can be described for introducing meaningful constraints on the allowable subsets. One approach is to supply the system with a collection of subsets of X from which it must select the fused value. We call this the case of a **user supplied vocabulary**. In this case if $A = \{A_1, \dots, A_m\}$ is the user supplied vocabulary we then select as our fused value subset $A^* \in A$ such that $\text{Sup}(A^*) = \text{Max}_{A_j \in A} [\text{Sup}(A_j)]$. We pick the term (subset) in the vocabulary with the largest support. Here considerable use can be made of fuzzy set theory and Zadeh's related idea of computing with words [22] to provide a mechanism for representing linguistic concepts in terms of sets. A prototypical example of this situation is one in which the A_j are a collection of fuzzy subsets, corresponding to linguistic terms related to the variable V . For example, if V is age then these could be terms like *old*, *young*, *middle age*.

Another, and more general, approach to restricting the subsets available as outputs of the fusion process is to use a measure on space of subsets of X to indicate our constraints. In this case our measure is a mapping $\mu: 2^X \rightarrow [0, 1]$. Here for any subset of A of X , $\mu(A)$ indicates the degree to which it is acceptable to the user to provide A as the fused value.³ We can denote this measure as the **Client Acceptability Measure (CAM)**.

With the availability of such a measure we can proceed in the following manner. We calculate the support for A by source i as

³We could define μ over the set of fuzzy subsets $\mu: I^X \rightarrow [0, 1]$.

$$\text{Sup}_i(A) = \mu(A) \wedge \text{Max}_y[A(y) \wedge \text{Sup}_i(y)]$$

where as in the preceding, $\text{Sup}_i(y) = g(w_i \wedge R(a_i), \text{Comp}(y, a_i) \wedge R(y))$. Thus $\mu(A)$ bounds the support available from any source.

If we neglect the reasonableness, assume $R(x) = 1$ for all $x \in X$, and assume $w_i = 1$ then

$$\text{Sup}_i(A) = \mu(A) \wedge \text{Max}_y[A(y) \wedge \text{Comp}(y, a_i)]$$

Let us look at some natural features of a measure μ which is used to convey the acceptability of considering a subset A as the fused value. One characteristic of a set that may be important in determining its appropriateness as a fused value is its size or cardinality. Since the fewer the number of elements in a subset, the more informative (useful) it is as a fused value, it is natural to consider smaller sets more acceptable than larger sets. This observation is reflected in the general feature that $\mu(A) \geq \mu(D)$ if $A \subset D$. A related property of this measure is that any subset consisting of a singleton should be completely acceptable, $\mu(\{x\}) = 1$.

In applying these CAM's in the manner described above we are essentially trying to reflect some criteria or requirements that a user has with respect to how they will use the fused value. Often these criteria reflect some operational or cognitive need of the user. There are two attributes associated with a subset which can be used to help in the expression of these criteria within a CAM.

In general, when we allow subsets as fused values, we prefer them to contain elements that are consistent (similar) rather than a collection of diverse values. In order to capture this feature of granularization we can make use of the proximity relationship. Specifically, some indication of the internal compatibility of the elements in A can be used to capture this consideration in the determination of $\mu(A)$. To express this notion we can suggest using as the internal compatibility of the subset A , the formulation $\text{I-Comp}(A) = \text{Min}_{x,y \in A} [\text{Comp}(x, y)]$. Thus, here we take the

smallest compatibility of any two elements in A as its internal compatibility. Thus, $\text{I-Comp}(A)$ can be used to help in the formulation of $\mu(A)$ to aid in capturing this notion of consistency. In the most basic application of this, we can define $\mu(A) = \text{I-Comp}(A)$. We note that if $A \subset D$ then $\text{I-Comp}(A) \geq \text{I-Comp}(D)$. If A is a singleton set, $A = \{x\}$, since $\text{Comp}(x, x) = 1$ for all x then $\text{I-Comp}(A) = 1$.

In the above we implicitly assumed that A was a crisp set. In the case where A is fuzzy we can define

$$\text{I-Comp}(A) = \min_{x,y \in A} [(\bar{A}(x) \wedge \bar{A}(y)) \vee \text{Comp}(x, y)]$$

where $\bar{A}(x) = 1 - A(x)$. Other more sophisticated definitions of the notion of internal compatibility of a set can be obtained with the use of soft computing technologies such as fuzzy modeling [23]

The second feature of a subset that can be used in the formulation of the CAM to help reflect the user desires is the cardinality of a subset. As we have noted this can help capture the fact that users typically prefer smaller sets to bigger sets. Here we shall not pursue this topic but only indicate that considerable use can be made of Zadeh's fuzzy set based idea of computing with words to relate information about the cardinality of a set A and its value $\mu(A)$. For example, we can capture a user's desire that the fusion set contain **only a few** elements. In this case we can, using Zadeh's idea of linguistic quantities [24], represent **only a few** as a fuzzy subset Q of non-negative integers in which $Q(1) = 1$ and $Q(x) \geq Q(y)$ if $x < y$. Then for any A , $Q(\text{Card } A)$ can be used to indicate $\mu(A)$.

The construction of $\mu(A)$ must take into account the preferences of the user as well as the structure of the underlying proximity relationship. Consider the situation in which Prox is an equivalence relations. In this case we see $\text{I-Comp}(A) = 0$ if there exists x and $y \in A$ from different equivalence classes. Thus $\text{I-Comp}(A) = 1$ iff A is contained in an equivalence class, otherwise it is zero. In this situation it would appear that the construction of μ using $\text{I-Comp}(A)$ is not appropriate. Here, when the proximity is an equivalence relationship it is best to construct μ based on some function of the cardinality of A .

We point out that the description of a user supplied vocabulary can be made with the use of a CAM μ in which all words in the vocabulary have $\mu(A) = 1$ and all those not in the vocabulary have $\mu(A) = 0$. In some situations we may describe our desired fused sets using both a user supplied vocabulary as well as criteria based on $\text{I-Comp}(A)$ and/or the cardinality of A . An important example of this arises in the case where our proximity relationship is based on a linear ordering L on X . Here we may require that our fused subsets be **intervals**. We recall that A is

an interval if there exists some x and $z \in X$ such that $A = \{y | x \underset{L}{>} y \underset{L}{>} z\}$. In this case we may also associate with each interval a value $\mu_2(A)$ which depends on $I\text{-Comp}(A)$ and a value $\mu_3(A)$ which depends on $Q(\text{Card } A)$. Then we use

$$\mu(A) = \mu_1(A) \wedge \mu_2(A) \wedge \mu_3(A)$$

where $\mu_1(A) = 1$ if A is an interval and $\mu_1(A) = 0$ if A is not an interval.

11. Multiple Granular Objects as Fused Values

In the following we shall make some other observation regarding the presentation of fused information. Initially we addressed the problem of selecting an element from X as our fused value. We then considered the situation in which we allowed our fused value to be some subset of X . In this case our fused value is essentially multiple elements from X . Here we introduced a CAM, measure μ on X , to reflect the users desires with respect to acceptability of different subsets

We will now take this one step further by considering the situation in which we allow as the multiple subsets of X as the fused value. For example, multiple intervals or multiple user vocabulary words. Here a basic consideration is how many subsets can we use. In order to do this we must introduce an additional measure η . Let C be a set whose elements are subsets of X , if $\mathbf{P}$ is the power set of X , C is a subset of $\mathbf{P}$. We define $\eta: \mathbf{P} \rightarrow [0, 1]$, where for any C , $\eta(C) \sim Q(\text{Card}(C))$. That is $\eta(C)$ depends on the cardinality of C , the number of subsets of X in C . Thus if q is any integer $Q(q)$ indicates the acceptability of providing q subsets. Here we require $Q(1) = 1$ and generally we expect $Q(i) \geq Q(j)$ if $i < j$.

Using this we can now express the support of any subset C of the power set as

$$\text{Sup}(C) = F(\text{Sup}_i(C), \dots, \text{Sup}_i(C))$$

Let us determine the support for C from i . We can express this as:

$$\text{Sup}_i(C) = \eta(C) \wedge \text{Max}_{A \in C} [\text{Sup}_i(A)]$$

Thus $\text{Sup}_i(C)$ depends upon the support from source i for all the subsets of X that are in C . Since we have already provided a way for determining $\text{Sup}_i(A)$ we are able to determine $\text{Sup}_i(C)$. We note that if $\eta(C) = 0$ for $\text{Card}(C) > 1$ we then have effectively reduced this to a case where we

only allow one subset.

12. References Part I

- [1]. Waltz, E. and Llinas, J., Multisensor Data Fusion, Artech House: Norwood, MA, 1990.
- [2]. Frakes, W. B. and Baeza-Yates, R., Information Retrieval: Data Structures and Algorithms, Prentice-Hall: Upper Saddle River, NJ, 1992.
- [3]. Berry, M. J. A. and Linoff, G., Data Mining Techniques, John Wiley & Sons: New York, 1997.
- [4]. Baeza-Yates, R. and Ribeiro-Neto, B., Modern Information Retrieval, Addison Wesley: Reading, MA, 1999.
- [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]. Yager, R. R., "Uncertainty representation using fuzzy measures," IEEE Transaction on Systems, Man and Cybernetics 32, 13-20, 2002.
- [7]. Shafer, G., A Mathematical Theory of Evidence, Princeton University Press: Princeton, N.J., 1976.
- [8]. Yager, R. R., Kacprzyk, J. and Fedrizzi, M., Advances in the Dempster-Shafer Theory of Evidence, John Wiley & Sons: New York, 1994.
- [9]. 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.
- [10]. Lin, T. S., Yao, Y. Y. and Zadeh, L. A., Data Mining, Rough Sets and Granular Computing, Physica-Verlag: Heidelberg, 2002.
- [11]. Kaufmann, A., Introduction to the Theory of Fuzzy Subsets: Volume I, Academic Press: New York, 1975.
- [12]. Zadeh, L. A., "Similarity relations and fuzzy orderings," Information Sciences 3, 177-200, 1971.
- [13]. 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.

[14]. 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.

[15]. Sheno, S. and Melton, A., "Proximity relations in fuzzy relational databases," *Fuzzy Sets and Systems* 31, 287-298, 1989.

[16]. Bouchon-Meunier, B., Rifqi, M. and Bothorol, S., "Towards general measures of comparison of objects," *Fuzzy Sets and Systems* 84, 143-153, 1996.

[17]. Arrow, K. J., *Social Choice and Individual Values*, John Wiley & Sons: New York, 1951.

[18]. Nash, J. F., "The bargaining problem," *Econometrica* 18, 155-162, 1950.

[19]. Yager, R. R. and Rybalov, A., "Noncommutative self-identity aggregation," *Fuzzy Sets and Systems* 85, 73-82, 1997.

[20]. Zadeh, L. A., "Calculus of fuzzy restrictions," in *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, edited by Zadeh, L. A., Fu, K. S., Tanaka, K. and Shimura, M., Academic Press: New York, 1-40, 1975.

[21]. 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.

[22]. Zadeh, L. A., "Fuzzy logic = computing with words," *IEEE Transactions on Fuzzy Systems* 4, 103-111, 1996.

[23]. Yager, R. R. and Filev, D. P., *Essentials of Fuzzy Modeling and Control*, John Wiley: New York, 1994.

[24]. Zadeh, L. A., "A computational approach to fuzzy quantifiers in natural languages," *Computing and Mathematics with Applications* 9, 149-184, 1983.

PART II. FUZZY METHODS IN WEB QUESTION ANSWERING SYSTEMS

1. Introduction

The wide spread availability of the internet has generated considerable interest in electronic sources of information. Various types of software applications have been developed to support this interest. We can generically refer to these as **Applications Useful for Seeking Information**. (AUSIN). One widely used class of these AUSIN is search engines. These applications retrieve pointers to pages or files based on their matching some key words specified by the user. Another widely used class of AUSIN is data base systems. In some ways database applications and search engines are opposite extremes of these AUSIN. Database applications give precise information however, they only work in highly structured environments. The search engines while only giving imprecise responses in the form of pointers are capable of functioning in highly unstructured almost chaotic environments. However, neither of these two applications typically have a reasoning capacity.

Another class of AUSIN is question answering systems [1-6]. An important dimension along which question answering systems differ from search engines and most databases is in their reasoning ability [6]. A fundamental characteristic of question answering systems is its ability to reason over its information base. This facility leads to a fundamental difference in the nature of their response to queries. In particular the response from a search engine is a pointer to a document (file or web page) resident in its library⁴. The database responds essentially by providing some value already resident in the database. A question answering system, because of its reasoning capacity, can construct new knowledge that is not resident in its knowledge base in response to a query. While currently less pervasive than either of the other information seeking systems, question answering systems with reasoning ability have the potential of being much

⁴We use the word library very broadly in that it could mean a subspace of the internet

more useful with regards to there responding to the real desires of the human user.

Intelligent question-answering systems generally require a gathering of information relevant to a question to be answered and then a processing of (reasoning over) this information. The processing often requires some appropriate representation of the information to be used. The general chaotic nature of the Internet makes the dynamic unrestrained gathering and subsequent representation of information very difficult. It often requires levels of understanding well beyond our current capability for computational intelligence⁵. The manipulation of appropriately gathered and represented knowledge is more within our current grasp.

Our point of departure in this work will be the availability of a domain knowledge base that has been built, with the aid of possible human interaction, using knowledge obtained from the internet as well as local domain knowledge. We do not preclude the system from dynamically interacting with appropriate Internet sources in response to queries. We can envision knowledge bases of this type being available within many organizations.

While the types of questions envisioned to be processed by future question answering systems is wide and varied in this work, we restrict ourselves to questions related to finding the value of some attribute variable using the information in the knowledge base.

Our interest in this work is on the development of tools to help in the representation of various types of knowledge as well as on the development of a framework and mechanism for reasoning and manipulating the knowledge in response to users questions.

Our approach will be in the spirit of Zadeh's paradigm of computing with words [7, 8]. In providing a schema for representing and locally manipulating knowledge we shall rely heavily on the fuzzy set based theory of approximate reasoning [9] and particularly the idea of protoforms [10, 11]. Considerable use will be made of knowledge trees [12, 13] to provide a global framework for structuring and directing the process of answering a question posed to the system. As we shall subsequently see a knowledge tree takes the contents of the knowledge base and structures it in a manner to answer the question presented. These trees can be seen as a mobilization of the knowledge base to address a particular task. At a meta level the knowledge

⁵This is to say nothing about the important problem of verifying the quality of the information.

tree provides a global plan for solving the problem, the answering of the question. It is worth noting, although we shall not pursue it here, a knowledge tree can be used to help point directions for going outside our knowledge base for information to improve our answer.

2. Basic Concepts of Approximate Reasoning

The primary elements in an Approximate Reasoning (AR) representation of a knowledge base are a collection of attribute variables, V_j for $j = 1$ to n , called the atomic variables. These attribute variables are the objects of interest in the current context. It is information about the value of these variables and relationships between them that constitutes the knowledge of interest. Associated with each variable V_j is a set X_j , indicating the allowable values for the variable. The terms base set, domain and universe of discourse are used to indicate the set X_j . A joint variable is any tuple of one or more distinct atomic variables; V_3 , (V_2, V_5) and (V_1, V_2, V_6) are examples of joint variables. Associated with any joint variable is a base set consisting of the Cartesian product of the domains of the individual variables making up the joint variable. Thus, if a joint variable has q components, its domain is a set of q -tuples. It is implicitly assumed, unless otherwise stated, that the variables of interest can assume only one value in its base set (see [14] for a discussion of different variable types).

A categorical proposition (or statement) in theory of approximate reasoning is of the form V is A , here V is a joint variable and A is a fuzzy subset⁶ of the domain of V . A proposition involving only one variable is called a canonical proposition while those involving two or more variables are called relational or joint propositions. A large variety of types of knowledge can be represented by propositions of the above form.

A proposition in AR is viewed as imposing constraints on the possible values of the variables involved. For example, if A is a crisp subset then the meaning of the proposition V_1 is A is to indicate that the value for the variable V_1 is restricted to be a member of the set A ,

⁶We shall follow the convention of expressing a fuzzy subset A of X as $\bigcup_{x \in X} \{\frac{A(x)}{x}\}$.

that is elements in A are the only possible values for V_1 . The use of fuzzy subsets provides for a grading of this idea. If B is a fuzzy subset, then one meaning of the proposition V_1 is B is that for any $x \in X_1$, $B(x)$ indicates the possibility that $V_1 = x$. If $V = (V_1, V_2, V_3)$ is a joint variable then the meaning of the proposition V is M is that for any $(x_1, x_2, x_3) \in X_1 \times X_2 \times X_3$, $M(x_1, x_2, x_3)$ is the possibility that $V_1 = x_1$, $V_2 = x_2$, and $V_3 = x_3$. Thus, we see that the knowledge represented by the propositions is essentially contained in the fuzzy subsets associated with the propositions.

We shall say a proposition, V is M , is normal (consistent) if $\text{Max}_x[M(x)] = 1$, then there exists at least one element in the base set of V that has possibility one. We shall call a proposition **subnormal** if $\text{Max}_x[M(x)] < 1$. We call a proposition V is M a tautology if $M(x) = 1$ for all x in the domain of V . It should be noted that a tautology existentially provides no restriction of the value of a variable and thus induces no new information other than that the variable must be in its base set.

When using the AR framework we are generally involved in one of three tasks. The first task, translation or knowledge representation deals with the process of taking knowledge normally expressed in natural language and converting it into an appropriate representation within the framework of AR. We construct our knowledge base via this translation process. The second task is extracting information from the knowledge base. This task is called the inference process. It involves the manipulation of propositions to obtain other propositions. The third task is that of retranslation, where we are interested in converting the propositions obtained via the inference process into statements of natural language. Here we shall mainly focus on the second task, inference, and to a lesser extent on the translation task. In [8] we have addressed the retranslation problem.

The basic operations used for knowledge manipulation within the AR framework are conjoin, containment and negation. We now look at these operations.

The conjoin operation provides the system with the facility for combining or fusing information. In the translation process it allows us to represent knowledge involving multiple variables and allows us to construct joint variables from atomic variables. The inference process uses this operation to combine individual pieces of knowledge. The operation of conjoin is a

generalization and unification of the set operations of conjunction and cartesian product. It is also closely related to the join operation used in the databases.

Definition: Assume V_a and V_b are two joint variables on the universes X and Y respectively. Let V_a is D and V_b is E be two propositions. Their conjoin (conjunction) denoted V_a is $D \times V_b$ is E is the proposition V is F . Here V is a joint variable consisting of the union of the atomic variables making up V_a and V_b . F is a fuzzy subset of the domain of V , Z , such that for each $z \in Z$ we have

$$F(z) = D(x) \wedge E(y)$$

where x is the portion of z corresponding to the domain X and y is the portion corresponding to the domain Y . The operator $\wedge$ is the Min, although more generally it can be a t-norm [15].

In the case when the two variables being conjoined are the same the conjoin operation reduces to the usual intersection of fuzzy sets:

$$V_a \text{ is } D \times V_a \text{ is } E = V_a \text{ is } F \text{ where } F = D \cap E$$

In this situation we shall sometimes find it convenient to use $\cap$ in place of $\times$. In the case when the two joint variables being conjoined have no common variables this operation reduces to the cartesian product: $V_a \text{ is } D \times V_b \text{ is } E = (V_a, V_b) \text{ is } F \text{ where } F = D \times E$.

We now define a special conjoin operation. This operation plays a role in making propositions which are not necessarily about the same variable be about the same variable.

Definition: Assume V_a and V_b are two joint variables such that V_b contains all the variables that are in V_a . The **cylindrical extension** of the proposition V_a is F to the proposition V_b is F° is defined by

$$V_b \text{ is } F^\circ = V_a \text{ is } F \times V_1 \text{ is } X_1 \times V_2 \text{ is } X_2 \times \dots \times V_q \text{ is } X_q$$

where $V_1, V_2, \dots, V_q$ are the atomic variables in V_b that are not in V_a and the X_i 's are the base sets of these variables.

Note: We can equivalently express this cylindrical extension of V_a is F to V_b by

$$V_b \text{ is } F^\circ = V_a \text{ is } M \times V_b \text{ is } X$$

where X is the domain of the variable V_b .

The membership function F° can be obtained from the membership function of F by setting $F^\circ(y) = F(x)$ where x is the tuple in the base set of V_a that corresponds to the portion of y

in this subspace.

We now turn to the second basic operation used in approximate reasoning, containment. At a very fundamental level this operation provides us with a facility for ordering propositions in approximate reasoning.

Definition: Assume V_a is D and V_b is E are two propositions, we say that V_a is D contains V_b is F denoted V_b is $E \subseteq V_a$ is D if

$$F^\circ(z) \geq D^\circ(z) \text{ for all } z$$

where F° and D° are the cylindrical extensions of F and D to the base set of variable V , the union of the atomic variables in V_a and V_b .

Note 1: If V_a and V_b are the same variables then this is the definition of fuzzy containment suggested by Zadeh [16].

Note 2: It can be shown that if P_1 and P_2 are two propositions and $P_3 = P_1 \times P_2$ then

$$P_3 \subseteq P_1 \text{ and } P_3 \subseteq P_2.$$

Containment allows us to define equivalence of propositions.

Definition: Two propositions P_1 and P_2 are equivalent if $P_1 \subseteq P_2$ and $P_2 \subseteq P_1$, we shall denote this as $P_1 = P_2$

Note: Equivalent propositions are interchangeable under conjoin, assume P_1 and P_2 are equivalent, $P_1 = P_2$, and let P_4 be any other proposition then $P_1 \times P_4 = P_2 \times P_4$

The third basic operation in AR is negation. While the other two operations involve more than one proposition involves only one proposition.

Definitions: Assume V is A is a proposition where the domain is X . The negation of V is A , denoted **not** (V is A) is the proposition V is $\bar{A}$, $\bar{A}(x) = 1 - A(x)$ for each x .

An important relationship exists between negation and containment

Theorem: If V_a is $A \subseteq V_b$ is B then V_b is $\bar{B} \subseteq V_a$ is $\bar{A}$

Proof: If V_a is $A \subseteq V_b$ is B then for all z in the extensions $A^\circ(z) \leq B^\circ(z)$, this implies $\bar{A}^\circ(z) \leq \bar{B}^\circ(z)$ and the result follows.

We shall here introduce an operation which is based upon a combination of conjoin and negation.

Definition: Assume V_a is A and V_b is B are two propositions. The operation $\perp$ is such that

$$V_a \text{ is } A \perp V_b \text{ is } B = \text{not } (V_a \text{ is } \bar{A} \times V_b \text{ is } \bar{B}) = \overline{(V_a \text{ is } \bar{A} \times V_b \text{ is } \bar{B})}.$$

Note: $V_a \text{ is } A \perp V_b \text{ is } B$ can be expressed as $V \text{ is } D$ where V is the union of the atomic variables in the joint variables V_a and V_b and D is a fuzzy subset on the universe of V such that

$$D(z) = A(x) \vee B(y),$$

where $\vee$ is the Max operator, x is the component of the tuple z corresponding to the domain of V_a and y is a component of the tuple z corresponding to the domain of V_b .

Note: If V_a and V_b are the same the operator $\perp$ is the union. With this in mind we shall use $\cup$ synonymously with $\perp$.

The preceding operations are essentially logical operations. We now introduce another operation that will play an important role in the process of question-answering is the Zadeh extension principle [14, 16-18].

Assume X and Y are two crisp subsets and let f be a mapping from X into Y , $f: X \rightarrow Y$. Here for each $x \in X$, we have $f(x) = y \in Y$. Assume $A = \bigcup_{x \in X} \{\frac{A(x)}{x}\}$ is a fuzzy subset of X

using the Zadeh extension principle, we can define $f(A)$ as a fuzzy subset of Y such that

$$f(A) = \bigcup_{x \in X} \{\frac{A(x)}{f(x)}\}$$

We see the Zadeh extension principle allows us to extend operation on elements of a set to act on fuzzy subsets

As a simple illustration assume $X = \{1, 2, 3\}$ and $Y = \{a, b, c, d, e\}$. Let $f: X \rightarrow Y$ be defined as $f(1) = a$, $f(2) = e$ and $f(3) = b$. Let $A = \{\frac{1}{1}, \frac{0.7}{2}, \frac{0.4}{3}\}$ then $f(A) = \{\frac{1}{a}, \frac{0.7}{e}, \frac{0.4}{b}\}$.

Note 1: It can be shown that if we denote $B = f(A)$ then $B(y) = \text{Max}_{\substack{\text{all } x, \text{ s.t.} \\ f(x) = y}} [A(x)]$

Note 2: Often the extension principle is used when the spaces X and Y are the same. For example if f is $y = 2x$.

The Zadeh extension principle can be generalized. Let $X_1, X_2, \dots, X_n$ and Y be a collection of sets. Assume f is a mapping $f: X_1 \times X_2 \times \dots \times X_n \rightarrow Y$. That is for each tuple $(x_1, \dots, x_n) \in X_1 \times X_2 \times \dots \times X_n$ we have $f((x_1, \dots, x_n)) = y \in Y$. Let A_i be a subset of X_i . The extension principle allows for the evaluation of $f(A_1, \dots, A_n)$. Denoting $F(A_1, \dots, A_n) = B$ where

B is a fuzzy subset of Y we have $B = \bigcup_{x_1, \dots, x_n} \left\{ \frac{\text{Min}[A_i(x_i)]}{f(x_1, \dots, x_n)} \right\}$ thus we have

$$B(y) = \text{Max}_{\substack{\text{all } x_1, \dots, x_n \text{ s.t.} \\ f(x_1, \dots, x_n) = y}} [A_1(x) \wedge A_1(x) \wedge \dots \wedge A_n(x_n)]$$

This form of the extension principal allows us to define arithmetic operations on fuzzy numbers. Assume R is the set of real numbers. Addition is a mapping from $R \times R$ into R. That is for each $x, y \in R$ we have $f(x, y) = z$ where $x + y = z$. If A and B are two fuzzy numbers their sum $C = A + B$ is a fuzzy number so that $C = \bigcup_{\text{all } (x, y)} \left\{ \frac{A(x) \wedge B(y)}{x + y} \right\}$. Here we see for any $z \in R$

$$C(z) = \text{Max}_{\substack{\text{all } (x, y) \text{ s.t.} \\ x + y = z}} [A(x) \wedge B(y)]$$

More generally if $\perp$ is any binary arithmetic operation then with A and B fuzzy numbers we have $A \perp B = C$ where C is also a fuzzy number such that $C = \bigcup_{\text{all } (x, y)} \left\{ \frac{A(x) \wedge B(y)}{x \perp y} \right\}$.

Equivalently for all $z \in R$ we have $C(z) = \text{Max}_{\substack{\text{all } (x, y) \text{ s.t.} \\ x \perp y = z}} [A(x) \wedge B(y)]$. Here $\perp$ can be addition,

subtraction, multiple, division, exponentiation. It also can be the maximum or minimum of two numbers.

We shall provide another form of the extension principle involving fuzzy subsets. In anticipation of providing this form of the extension principle we introduce the idea of level sets. Let A be a fuzzy subset of X the α level set of A, denote A_α , is a crisp subset of X defined as $A_\alpha = \{x/ A(x) \geq \alpha\}$. Again let X at Y be two crisp subset and assume $G: 2^X \rightarrow Y$, it is a mapping from subsets of X, into Y. We can extend G to map fuzzy subsets of X. In particular. $G(A) = \bigcup_{\alpha \in [0, 1]} \left\{ \frac{\alpha}{G(A_\alpha)} \right\}$. Here $G(A)$ is a fuzzy subset of Y

3. Semantics of AR the Knowledge Representation

An important task in the use of the AR framework in question-answering systems is the translation of our knowledge from statements in a natural language into propositions in AR. Here we first provide some basic understanding of the semantics of our formal representation. The most basic translation rule is the assignment of a value to an atomic variable. An example of this occurs if we have the information that John is young. We first represent the concept young as a fuzzy subset A and then assign it to constrain the value of the variable V, Johns age. Using this

representation we are saying that there is some uncertainty with regard to our knowledge of the value of V and $A(x)$ is the possibility that $V = x$.

Some specific cases of A are worth noting. If $A = \{x^*\}$, a singleton, then V is A conveys the fact that x^* is the exact value of V . This is equivalent to the ordinary statement $V = x^*$. Another special case is when A is a crisp set. In this case, the statement V is A is effectively indicating that V is an element in A , $V \in A$. Another important example of this is when $A = X$. Here we are saying that A can be any value in X . This corresponds to the case when we have no knowledge about V other than its domain. It is clear that this corresponds to a least restrictive constraint and can always be assumed.

We note that if $\text{Max}_X[A(x)] < 1$ then there is no element in X having a possibility of one of being the value of V . It is an indication of some conflict between our knowledge and the assumption that V must assume its value in X . The extreme case is when $A = \emptyset$, here no element has any possibility of being the value of X . The information out of which the knowledge base is constructed is usually assumed to involve fuzzy subsets that are normal, $\text{Max}_X[A(x)] = 1$. The conflicts generally arise as a result of the process of combining pieces of knowledge.

Consider the statement V is A , where for ease of explanation we assume A is crisp. As we indicated, this statement is saying that our knowledge of V is that its value lies in the subset A . If we know that the value of V lies in the subset A then if $A \subseteq B$ we also know that the value of V lies in B . For example, if we know that John is in his twenties then we can infer that he is over fifteen years old. This can be seen as a kind of basic inference process. We call this process **entailment**. While we are able to go from smaller subsets to larger subsets with certainty we can't go the other way around, from larger subsets to smaller subsets with the same certainty. For example, knowing John is in his twenties doesn't allow us to conclude that John is between 21 and 23.

Consider now that we have two pieces of knowledge about the same variable; V is A and V is B . Again for simplicity of discussion we shall assume A and B are crisp subsets. The first statement indicates that the possible value for V lies in the set A and the second statement says that the possible value for V lies in the set B . Consider the situation where one statement says John is between 10 and 20 and the other says that he is over 15. In this case we can conclude that

he is between 15 and 20. More generally in the case of multiple pieces of information we take the conjunction; that is, we conclude V is D where $D = A \cap B$. As we shall subsequently see the processes of conjunction and entailment form the basis of reasoning.

As we have indicated, statements of the form V is A allows for a representation which can capture uncertainty in our information. A number of measures have been introduced to help quantify different aspects of the uncertainty associated with these propositions. One of these is the concept of specificity of a fuzzy subset. In [19] Yager describes its properties and a number of possible formulations for quantifying it. Assume A is a fuzzy subset of X and let x^* be an element with maximal membership grade in A , $A(x^*) = \text{Max}_X[A(x)]$, and let $\text{Aver}_{x^* \notin X}[A(x)]$ be the average of all membership grades in A excluding x^* , then we define the specificity of A as $\text{Sp}(A) = A(x^*) - \text{Aver}_{x^* \notin X}[A(x)]$. We see some properties of this measure: **1.** $\text{Sp}(A) = 1$ if A consists of exactly one element, $A = \{x\}$. **2.** $\text{Sp}(\emptyset) = \text{Sp}(X) = 0$. **3.** If A and B are two normal fuzzy subsets of X , then $\text{Sp}(A) \geq \text{Sp}(B)$ if $A \subset B$.

One application of the concept of specificity is that it provides a measure of the amount of information contained in a proposition in AR. Thus if $\mathbf{P}$, is a proposition expressible as V is A then the amount of information contained in $\mathbf{P}$, $\text{Inf}(\mathbf{P}) = \text{Sp}(A)$. An important tool in AR is the principle of Minimal Specificity [20]. This states that if we have a number of different possible representations of some piece of knowledge, then we should choose the one with the minimal specificity. We see this as being akin to the principle of maximal entropy in probability theory. Another related use of specificity is that a property of a valid inference is that it never increases specificity. Thus, if we infer P_2 from P_1 , $P_1 \mapsto P_2$, then it must be the case that $\text{Sp}(P_1) \geq \text{Sp}(P_2)$. That is inference doesn't allow us to increase information.

We now turn to another measure of uncertainty. Assume V is A and V is B are two propositions in AR. We define the possibility of V is A given V is A , denoted $\text{Poss}[V \text{ is } A | V \text{ is } B]$ as $\text{Max}_X[F(x)]$ where $F = A \cap B$. Possibility measures the degree to which two propositions have some solution in common.

A related measure is the certainty measure. The certainty of V is A given V is B , is defined as $\text{Cert}[V \text{ is } A | V \text{ is } B] = 1 - \text{Poss}[V \text{ is } \bar{A} | V \text{ is } B]$. If B is normal then $\text{Poss}[V \text{ is } A | V \text{ is } B] \geq \text{Cert}[V \text{ is } A | V \text{ is } B]$. If B is a singleton, $B = \{x^*\}$ then $\text{Poss}[V \text{ is } A | V \text{ is } B] = \text{Cert}[V \text{ is } A$

$$| V \text{ is } B] = A(x^*)$$

We note that Poss is a measure of the degree of intersection of A and B and $\text{Cert}[A|B]$ is a measure of the degree to which B is contained in A. An important interpretation of the measures of possibility and certainty are respectively as the upper and lower bounds of the truth of the statement. V_a is A given the knowledge that V_b is B.

4. Translation of Information

An important step in the construction of question-answer systems is the translation of our knowledge from statements in a natural language into propositions in AR. A study of Zadeh's [21, 22] work will provide a large collection of translation rules. Here we discuss some of these.

As we have already indicated, the basic statement involves the association of a fuzzy subset with a variable. Here we shall see how to build more complex representations using this basic building block.

Assume P is a statement that has representation as V_a is A and Q is a statement with representation as V_b is B. Using these we obtain the following translation rules.

- not (P) is represented as V_a is $\bar{A}$ where $\bar{A}$ is the negation of A, $\bar{A}(x) = 1 - A(x)$
- P and Q is represented as V_a is A $\times$ V_b is B. In the special case where $V_a = V_b$ we have V_a is $A \cap B$. In another special case where V_a and V_b are disjoint we get (V_a, V_b) is D where $D(x,y) = \text{Min}(A(x), B(y))$. Another case which will be useful is where we have partial overlap that is $V_a = (V_1, V_2)$ and $V_b = (V_1)$. If we let X and Y respectively be the domains of V_1 , and V_2 then we get (V_1, V_2) is E when $E(x,y) = \text{Min}[A(x,y), B(x)]$
- P or Q is translated as $V_a = A \cup V_b$ is B. In the special case where $V_a = V_b$ we get (V_a, V_b) is D where $D = A \cup B$, $D(x) = \text{Max}[A(x), B(x)]$. In the case when V_a and V_b are disjoint we get V_a is E where $E(x,y) = \text{Max}(A(x), B(y))$
- If P then Q is represented as V_a is $\bar{A} \cup V_b$. In the special case where V_a and V_b are disjoint we get (V_a, V_b) is E where $E(x, y) = \text{Max}[\bar{A}(x), B(y)]$

In the preceding we have discussed the representation the kinds of knowledge whose structure is very close to the class of propositions appearing in classical logic. The knowledge

that can be represented in the AR framework does not have to be restricted to these kinds of classical examples. More generally the AR paradigm of translating knowledge into statements constraining the values of the relevant variables can be applied to a large variety of knowledge. A simple illustration of this involves the knowledge that Faye and Lotfi are close in age. If V_a represents the variable Faye's age and if V_b represents the variable Lotfi's age then we represent this knowledge as (V_a, V_b) is **CLOSE**. Here **CLOSE** is a fuzzy relationship such that for each pair of ages x and y the membership grade **CLOSE**(x, y) indicates the degree to which x and y satisfy the condition of being close.

An important class of knowledge involves functional relationships. Many scientific disciplines represent their knowledge in this manner. In addition, modern technological tools such as data mining use functional relationships to express the knowledge they discover [23]. Tools are available for translating functional relationships into propositions in AR. Let us look at some of these.

Let V and U be two disjoint variables with domains X and Y . A functional relationship, $U = f(V)$ is a mapping $f: X \rightarrow Y$ such that $y = f(x)$; it assigns to every value x in X a value Y . We can represent this knowledge in the AR framework as propositions of the form (V, U) is F where F is a relationship on the space $X \times Y$ such that $F = \bigcup_{x \in X} \{(x, f(x))\}$. Thus F is the union of all

pairs $(x, F(x))$ which are solutions to $U = f(V)$.

A special type of functional relationship are those commonly known as fuzzy models or fuzzy graphs [24]. We denote this type of fuzzy relationship as $V_b = F(V_a)$ where V_a and V_b are disjoint variables with domains X and Y . In this case instead knowing exact pairs of points we have a collection of fuzzy solution points, P_i and the relationship is the disjunction of the set points. A fuzzy solution point P_i , is characterized by a pair $\langle q_i, r_i \rangle$, $q_i \equiv V_a$ is A_i and $r_i \equiv V_b$ is B_i , and is defined as the conjoin of the components of the pair, V_a is $A_i \times V_b$ is B_i and thus it is a proposition of form (V_a, V_b) is P_i . Since the relationship between these fuzzy solution points in a disjunction then the overall function relationship $F = P_1$ or P_2 or or P_n . More formally $V_b = F(V_a)$ is represented as

$$F = \langle q_1, r_1 \rangle \perp \langle q_2, r_2 \rangle \perp \dots \perp \langle q_n, r_n \rangle.$$

This results in a proposition of the form $(V_a, V_b) \text{ is } P_1 \perp P_2 \perp \dots \perp P_n$

Another feature of the AR framework is the ability to include in the representation knowledge about the certainty or more generally an importance weight of a proposition. This is accomplished through the process of certainty (or importance) qualification. A statement

$$V \text{ is } A \text{ is } \alpha \text{ certain}$$

can be expressed as the proposition $V \text{ is } B$ where $B(x) = \bar{\alpha} \vee A(x)$. Here it is assumed $\alpha \in [0, 1]$.

We see if $\alpha = 1$ then $B = A$ and if $\alpha = 0$ then $B(x) = X$, the whole space. The case of $\alpha = 0$ results in the proposition expressing no knowledge.

5. Inference and Knowledge Manipulation

We now introduce the basic rules of inference available in AR [25]. In the following we use the notation $(P_1, \dots, P_n) \mapsto Q$ to indicate that we can infer Q from the collection $(P_1, \dots, P_n)$

The first rule of inference is called the **entailment principle**. This rule says that if P_1 and P_2 are two propositions such that $P_1 \subseteq P_2$ then from P_1 we can infer P_2 . We formally express this as inference rule 1

$$\text{IR-1: } (P_1) \mapsto P_2 \text{ if } P_1 \subseteq P_2$$

One important application is the following. If A and B are fuzzy subsets over the same universe such that $A \subseteq B$ then from the statement $V \text{ is } A$ we can infer $V \text{ is } B$.

A second important application of this inference rule is related to the process of projection.

Definition: Let V_a and V_b be two variables with domains X and Z respectively. The **projection** of $V_a \text{ is } A$ on V_b , denoted $\text{Proj}_{V_b}[V_a \text{ is } A]$ is defined as $V_b \text{ is } B$ where $B(z) = \text{Max}_Q[A(x)]$.

Here Q is the subset of elements in the domain of V_a , X , which have the value as z for the atomic variables which V_a and V_b have in common⁷. (If V_a and V_b are disjoint then $B(z) = 1$ for all z).

One important use of projection is the case where $V = (V_1, \dots, V_n)$, where all V_i are

⁷Assume V_1, V_2 and V_3 are atomic variables with domains X_1, X_2 and X_3 and let $V_a = (V_1, V_2)$ and $V_b = (V_1, V_3)$. If $z = (x_1, x_3)$ then Q is the subset of $X_1 \times X_2$ consisting of all points (x_1, y) where y is any point in X_2 .

atomic variables and we desire $\text{Proj}_{V_i}[V \text{ is } A]$. Let us first consider the special case $V = (V_1, V_2)$ and we desire $\text{Proj}_{V_1}[V \text{ is } A]$. We assume domain of V_1 is X and the domain of V_2 is Y . In this case, we get $V_1 \text{ is } B$ where $B(x_i) = \text{Max}_{y \in Y}[A(x_i, y)]$. More generally, if $V = (V_1, \dots, V_n)$ and X_i is the domain of V_i then $\text{Proj}_{V_1}[V \text{ is } A] = V_1 \text{ is } B$ where

$$B(x) = \text{Max}_{(x_2, \dots, x_n)} [A(x, x_2, x_3, \dots, x_n)]$$

$B(x)$ is the maximal membership grade in A of any tuple containing x .

The usefulness of the projection operator lies in the fact [25] that if $P_1 \equiv V \text{ is } A$ and $P_2 \equiv V_a \text{ is } \text{Proj}_{V_a}[V \text{ is } A]$ then $P_1 \subseteq P_2$. Thus using the entailment principle, IR-1, given $V \text{ is } A$ we can infer $V_a \text{ is } B$ where $B = \text{Proj}_{V_a}[V \text{ is } A]$. So we see that projection allows us to infer information about a single variable from information about joint variables. This is sometimes called marginalization.

Observation: Assume $(V_a, V_b) \text{ is } R$ where R is a subset of $X \times Y$. If R is such that for each $x \in X$ there exists a $y \in Y$ such that $R(x, y) = 1$ then $\text{Proj}_{V_a}[V \text{ is } A] = B$ is such that $B(x) = 1$ for all x . Hence all we infer is $V_a \text{ is } X$ which is essentially no knowledge.

The second rule of inference relates to the combination of multiple propositions. Assume we have a collection propositions P_i for $i = 1$ to q . Since, as we have previously indicated, a proposition $V_a \text{ is } A$ places a constraint on the allowable values of the associated variable, the total effect of a collection of propositions can be seen to be the conjunction (anding) of the individual propositions. This is the bases if the second inference rule

$$\text{IR-2: } \{P_1, P_2, \dots, P_n\} \mapsto P_1 \times P_2 \times \dots \times P_n$$

We are now in a position to understand the process of reasoning, obtaining inferences from a knowledge base consisting of the collection $\{P_1, P_2, \dots, P_n\}$. The set of propositions in our knowledge base induces a combined restriction which is the conjunction of all propositions in the knowledge base, $KB \equiv P_1 \times P_2 \times \dots \times P_n$. Thus our knowledge base is also a proposition, $V \text{ is } A$, which is the conjunction of all pieces of knowledge.

The use of IR-2, the conjunction of all the individual pieces of knowledge along with IR-2, the entailment principle, provides the bases of the reasoning mechanism

Basic Inference: Proposition P is inferable from a KB of proposition $P_1, \dots, P_n$ denoted, $(P_1, \dots, P_n) \mapsto P$ if $P_1 \times P_2 \times \dots \times P_n \subseteq P$

It is important to initially emphasize that the inference requires that we must use the whole knowledge base. However, as we shall subsequently see, in some cases we need not use the whole knowledge to make inferences. This simplification results in making the inference process tractable and useful. In cases in which we must use the whole knowledge base, the process of reasoning becomes particularly onerous and often impractical.

We now introduce an important idea related to the concept of monotonicity.

Definition: We say that a proposition Q is monotonic if for any proposition $P = V$ is A it is always the case $P \times Q \subseteq P$. If Q doesn't satisfy this condition we call it a non-monotonic proposition.

All the propositions that we have introduced so far are monotonic. As we shall subsequently see non-monotonicity often arises in the representation of common sense knowledge.

In the following we shall assume all the propositions in the knowledge base are monotonic, we have a monotonic knowledge base. The assumption of monotonicity allows for a great simplification in the reasoning process. It is the basis of the type of deduction that is common in reasoning.

Let us understand the great significance of monotonicity. The assumption of monotonicity implies that if $(P_1, \dots, P_n) \mapsto P$ then $(P_1, \dots, P_n, P_{n+1}) \mapsto P$. It means that adding any proposition doesn't cause us to have to withdraw any inference. More pragmatically it means that if we infer a piece of knowledge from some subset of the knowledge base we are sure it is valid for the whole KB

More specifically the assumption of monotonicity implies the following. If $\{P_1, \dots, P_n\}$ is our knowledge base and if $\{\tilde{P}_1, \dots, \tilde{P}_q\}$ and $\{\hat{P}_1, \dots, \hat{P}_r\}$ are two subsets of the KB then.

M1. If $\{\tilde{P}_1, \dots, \tilde{P}_q\} \mapsto \tilde{P}$ then $\{P_1, \dots, P_n\} \mapsto \tilde{P}$

M2. If $\{\tilde{P}_1, \dots, \tilde{P}_q\} \mapsto \tilde{P}$, $\{\hat{P}_1, \dots, \hat{P}_r\} \mapsto \hat{P}$ and if $(\tilde{P}, \hat{P}) \mapsto P$ then $(P_1, \dots, P_n) \mapsto P$

This property allows us to use deduction and take advantage of various types of local reasoning of patterns, such as modus ponens, and proceed in a step by step fashion to obtain complex inferences. This situation allows for a localization which greatly simplifies the inference process. In the following we shall refer to a subset of our knowledge base as a local knowledge

base. The observation of the preceding is that we can make valid inferences using local knowledge bases.

6. Protoforms

Recently, Zadeh introduced the idea of protoforms [10, 11, 26]. These protoforms are local reasoning patterns based on the structure of the component propositions. A protoform consists of a collection $P_1, \dots, P_q$ of premises and a consequent Q and is of the nature $(P_1, \dots, P_q) \mapsto Q$. For the most part the justification of these protoforms is based upon application of conjunction (IR-1) and entailment (IR-2). The importance of these protoforms is that if we can find a local knowledge base (subset) of our knowledge base consisting of propositions whose structure matches the premise of a protoform, then we can infer the consequent. Furthermore, given properties M1 and M2 above, these protoforms can be independent parts of a complex deductive reasoning chain in monotonic knowledge bases.

In the following we shall identify some protoforms available in AR. Unless otherwise indicated we assume all fuzzy subsets are normal. We use the notation PF-K to indicate protoform #K. As in the preceding the premises are identified by P_j and the consequent is denoted Q .

PF-1 provides a basic protoform for inferring information about a component variable from a joint relation.

PF-1 Projection protoform

$P_1: (V_1, V_2, \dots, V_q) \text{ is } H$

$Q: V_1 \text{ is } D \text{ where } D = \text{Proj}_{V_1}((V_1, V_2, \dots, V_q) \text{ is } H)$

We recall $D(x_1) = \text{Max}_y[H(x_1, y)]$ where $y = (x_2, \dots, x_q) \in Y$ and $Y = X_2 \times X_3 \times \dots \times X_q$

PF-2 is a fundamental protoform. It provides the basic protoform for making inferences from a joint relationship and companion propositions about its components. It essentially involves a conjunction of the premises followed by a projection onto the relevant variable.

PF -2 Conjunction/Projection Protoform

$P_1: V_1 \text{ is } B_1$

P₂: V₂ is B₂

P_q: V_q is B_q

P_{q+1}: (V₁, V₂, ..., V_q) is H

Q: V_j is D (D = Proj_{V_j}((V₁, V₂, ..., V_q) is E))

Here $E(x_1, \dots, x_q) = \text{Min}[B_1(x_1) \wedge B_2(x_2) \wedge \dots \wedge B_q(x_n) \wedge H(x_1, \dots, x_q)]$. Note if any of the P₁, ..., P_q are missing these still holds we can replace B_i by X_i.

The following protoforms can be seen as important special cases of the preceding.

PF-3 Modus Ponens

P₁: IF V is A then U is F

P₂: V is B

Q: U is D where $D(y) = \text{Poss}[\bar{A}|B] \vee F(y)$ ($\text{Poss}[\bar{A}|B] = \text{Max}_x[\bar{A}(x) \wedge B(x)]$)

This PF is a special case of PF -2. Here P₁ is represented as (V, U) is H and

$$H(x, y) = \bar{A}(x) \vee F(y)$$

$$E(x, y) = (\bar{A}(x) \wedge B(x)) \vee (F(y) \wedge B(x))$$

$$D(y) = \text{Max}_x[E(x, y)] = \text{Poss}[\bar{A}|B] \vee F(y)$$

Note 1: If B is subnormal, $\text{Max}_x[B(x)] = b$ then $D(y) = \text{Poss}[\bar{A}|B] \vee (F(y) \wedge b)$

Note 2: Since $\text{Cert}[A|B] = 1 - \text{Poss}[\bar{A}|B]$ then we can express $D(y) = (1 - \text{Cert}[A|B]) \vee F(y)$

PF- 4 Weighted Modus Ponens

P₁: If V is A then U is F is α certain

P₂: V is B

Q: U is D where $D(y) = \text{Poss}[\bar{A}|B] \vee \bar{\alpha} \vee F(y)$

Here P₁ becomes (V, U) is H is α becomes (V, U) is G where $G(x, y) = H(x, y) \vee \bar{\alpha}$

Note 3: If $\alpha = 0$ then $D(y) = 1$

Note 4: If B is subnormal, $\text{Max}_x[B(x)] = b$, the $D(y) = \text{Poss}[\bar{A}|B] \vee (b \wedge (\bar{\alpha} \vee F(y)))$

PF-5 Modus Tollens

P₁: If V is A then U is F

P₂: U is C

Q: V is G where $G(x) = \bar{A}(x) \vee \text{Poss}[F|C]$

Here $E(x, y) = (\bar{A}(x) \vee F(y)) \wedge C(y) = (\bar{A}(x) \wedge C(y)) \vee (F(y) \wedge C(y))$

$$G(x) = \text{Max}_y[E(x, y)] = \overline{A}(x) \vee \text{Poss}[F|C]$$

Note 5: If $\text{Poss}[F|C] = 1$ then $G(x) = 1$ for all x and we infer no information

Note 6: If C is subnormal, $\text{Max}_y(C(y)) = c$ then $G(x) = (\overline{A}(x) \wedge c) \vee \text{Poss}(F|C) \leq c$

The above protoforms were essentially special cases of PF-2 we now turn to another special case of this.

PF-6 Extension Principle

$$P_1: U = f(V)$$

$$P_2: V \text{ is } A$$

$$Q: U \text{ is } f(A)$$

In this case $U = f(V)$ is translated as (V, U) is H where $H(x, y) = \bigcup_{x \in X} \left\{ \frac{1}{(x, f(x))} \right\}$ and as

in the preceding $G(x, y) = H(x, y) \wedge A(x)$. Here again $D(y) = \text{Max}_x[H(x, y) \wedge A(x)]$ however in this case $D(y) = \text{Max}_x[H(x, y) \wedge A(x)] = \text{Max}_{x \text{ s.t. } f(x)=y} [A(x)]$. Hence $D = f(A)$

Note: More generally if $U = f(V_1, \dots, V_q)$ and V_i is A_i for all i then we infer U is $f(A_1, \dots, A_q)$ and $f(A_1, \dots, A_q) = D$ where $D(y) = \text{Max}_{(x_1, \dots, x_q) \text{ s.t. } f(x_1, \dots, x_q)=y} [A_1(x_1) \wedge \dots \wedge A_q(x_q)]$

PF - 6b Inverse Extension

$$P_1: U = f(V)$$

$$P_2: U \text{ is } B$$

$$Q: V \text{ is } f^{-1}(B)$$

Again, $U = f(V)$ is translated as (V, U) is H where $H(x, y) = \bigcup_{x \in X} \left\{ \frac{1}{(x, f(x))} \right\}$. However

here $G(x, y) = H(x, y) \wedge B(y)$ and $D(x) = \text{Max}_y[H(x, y) \wedge B(y)]$. In this case

$$D(x) = \text{Max}_{y \text{ s.t. } f^{-1}(y)=x} [B(y)]$$

PF - 7 Fuzzy Systems Modeling

$$P_1: \text{Fuzzy Systems Model } (U, V) \text{ is } (A_1 \cap B_1) \cup (A_2 \cap B_2) \cup \dots \cup (A_n \cap B_n)$$

$$P_2: U \text{ is } E$$

$$Q: V \text{ is } G \text{ where } G(y) = \text{Max}_i[(\text{Poss}[A_i|E] \wedge B_i(y))]$$

7. Knowledge Trees

In the following we shall describe an approach to question answering systems using the concept of a knowledge tree [12, 13]. As we shall see, a knowledge tree is a structure generated from a knowledge base in response to a query. It can be seen as a mobilization of the knowledge base to address a particular task. At a meta level it provides a plan for answering the question. We shall denote our knowledge base as KB and initially we assume it consists of a collection of monotonic propositions. Associated with the KB is a collection of atomic variables $V_1, \dots, V_n$ and their associated domains, $X_1, \dots, X_n$. We view the KB as consisting of two classes of propositions. The first class which we denote as D consists of all the atomic propositions, propositions of the form $V_j \text{ is } A_j$ where A_j is a fuzzy subset of the universe of V_j . Among the propositions in D are the atomic propositions $V_i \text{ is } X_i$, these just specify the domain of V_i . The second class are those that involve joint variables, they specify relationships between variables. We denote the collection of these as J ,

We shall concern ourselves here specifically with questions that ask about the value of some atomic variable V_i . Our procedure for answering a question will be a two phase process. The first phase is a generation of a knowledge tree in response to the posing of a question. The second phase is the contraction of the tree to find the value the variable queried about. The first phase can be seen as generating the direction and plan for answering the question

Before proceeding let us clarify the tree terminology we shall use. A node n_a will be called an ancestor of the node n_b if n_a appears anywhere on the path from the initial node of the tree to n_b . In this case n_b is called a decedent of n_a . A node n_a is called the parent of node n_b if n_b emanated directly from n_a . In this case n_b is called the child of n_a . The terms immediate ancestor and immediate decedent are equivalent to parent and child.

The following is the basic procedure for the **construction** of the knowledge tree in response to a query about an atomic variable V in the case where the KB is as described in the preceding

Basic Tree Generation Algorithm

1. The posing of a question in the form

? V is

initiates the knowledge tree with a triangle of the form 

2. The appearance, anywhere in the knowledge tree, of a triangle labeled with a variable **not having itself as an ancestor** causes the emanation from the triangle of a family immediate successor nodes:

a. One corresponding to each data proposition in the knowledge base subset D involving the label of the triangle. A successor node corresponding to an element from D will be manifested by a rectangle labeled by the name of the data proposition, i.e.

$d_i: V \text{ is } A$

b. One corresponding to each proposition in J having the label of the parent triangle as one of its component variables. A successor node corresponding to an element from J is manifested by a circle labeled by the name of the proposition, i.e. 

c. If there are no appropriate elements from either a or b this causes the emanation of a rectangle labeled by the domain of the variable, $V \text{ is } X$

3. The appearance anywhere in the knowledge tree of a triangle labeled with a variable **having itself as an ancestor** causes the emanation from that node of one successor node, a rectangle labeled by variables domain, $V \text{ is } X$

4. The appearance of a node corresponding to an element from J , a labeled circle, causes the emanation from it of a family of successor triangle nodes each labeled by a variable appearing in the proposition denoted by the circle

5. The appearance of a labeled rectangle terminates that branch of the tree.

6. The construction of the knowledge tree is completed when all branches terminate in rectangles.

7. The completion of the knowledge tree generation phase initiates the evaluation phase.

In figure 4 we illustrate a typical knowledge tree.

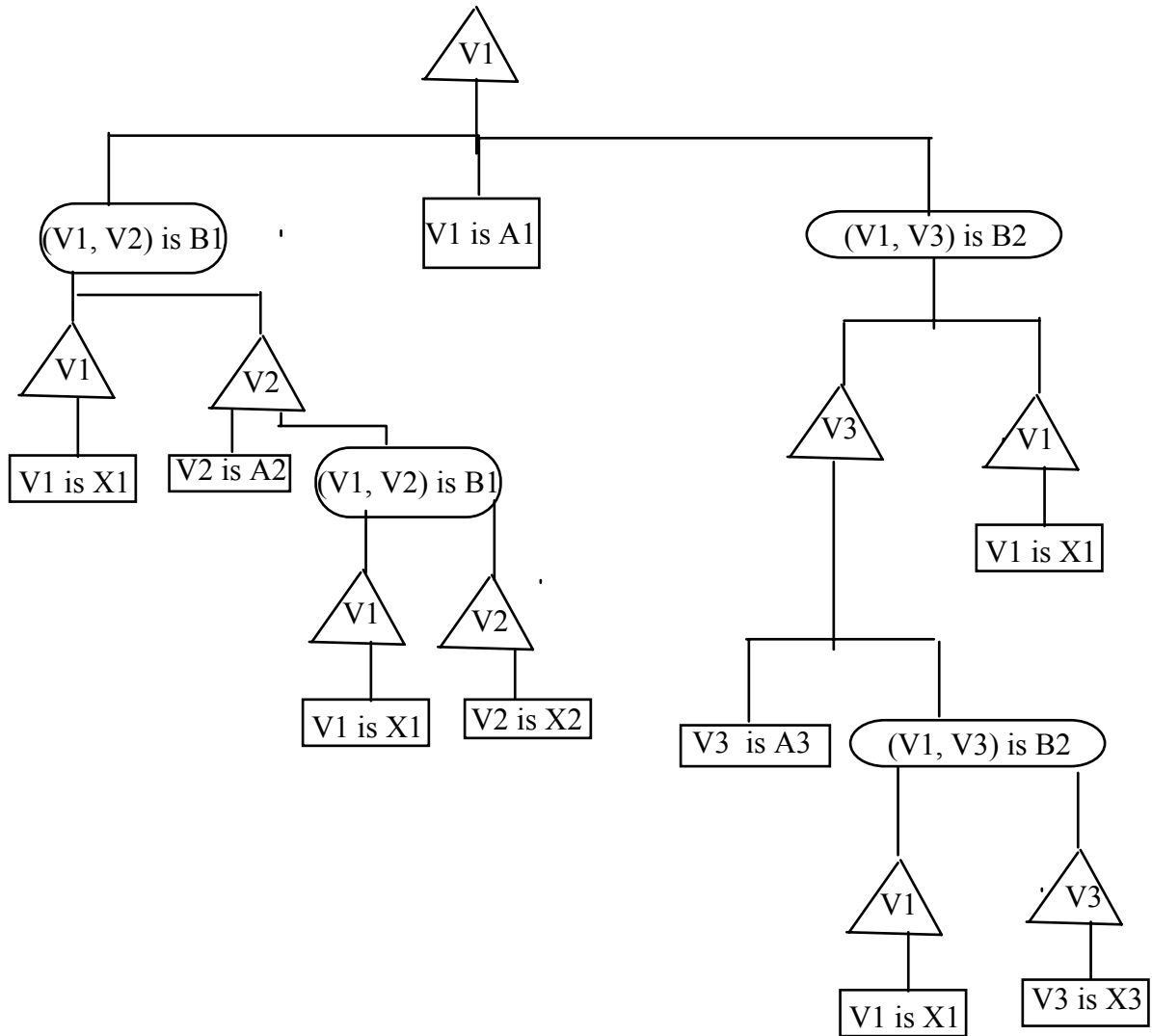


Figure 4. Typical Knowledge Tree

Once having completed the construction of the knowledge tree initiated by our query we must next apply an evaluation procedure to calculate the value of the desired variable. The evaluation algorithm is based upon the rules of approximate reasoning and makes considerable use of the protoforms introduced earlier.

Evaluation Algorithm

1. Any subtree consisting of a triangular node and a family of children which are all rectangular nodes can be replaced by a rectangular node whose label is $V \text{ is } H$, $\boxed{V \text{ is } H}$ Here V is the label of the triangular node and H is the conjunction of the values of the children of this triangular node. (It is an application of the conjunction protoform)
2. Any subtree consisting of a circular node, a relational proposition, and a family of

children all of which are rectangular nodes can be replaced by a rectangular node whose label is V_1 is G where V_1 is the label of the triangular node immediately preceding the circular node. Typically this is accomplished by application of PF-2, the Conjunction/Projection Protoform. However, more generally this may involve a two step process. The first step is the application of the protoform associated with the circular node and its children. The second step is an application of a projection onto V_1 of the proposition resulting from step 1.

3. The evaluation phase terminates when there exists one rectangular node whose label is the desired variable with its associated its value as the answer to the question.

The knowledge tree forms a useful framework for formulating the question answering process. As we see it takes a knowledge base and restructures the information in it to address the question being asked. We briefly point out one feature of the knowledge tree approach which, while we shall not currently pursue, is a very useful property. In particular the knowledge tree can point us in a direction to seek additional information not in our KB that may help us in getting a good answer. A simple illustration of this is the following. Assume we are interested in finding the value of V_1 . Assume in our KB we have a relation (V_1, V_2) is G . This relation will appear as a decedent of our query to find V_1 . Let us further assume we have no knowledge of V_2 . Thus this branch will terminate with a rectangle V_2 is X_2 . If as a result of our working on the whole tree we get an answer for V_1 that is not sufficiently informative for the user we see that this branch provides a potential path for improving our answer. It is telling us if you are able to get better information about V_2 you can possibly improve you answer for V_1 . At meta level this is essentially pointing in some direction for improving the quality.

8. Non-Monotonic Possibilistic Propositions

In order to take full advantage of the wide array of different types of information available we must extend our capability to allow the inclusion in the knowledge base of the types of commonsense knowledge that play a substantial role in human reasoning [6, 11]. Examples of this

are "generally it is difficult to find parking in New York City," "rainy weather usually results in more traffic congestions" and "a cup of coffee costs about a dollar." This commonsense knowledge often is not the kind of categorical knowledge that can be simply represented by the type of AR proposition we have thus far considered. It is often tentative and context dependent. This means it can often be involved in inferences that must be withdrawn in face of gaining additional knowledge. This situation is formally referred to as non-monotonicity [27, 28]. In the preceding we assumed all the knowledge to be categorical, of a monotonic type, this allowed us to use deductive reasoning techniques. Here we shall begin to consider the handling of types of propositions that can display a non-monotonicity.

As we noted non-monotonic propositions often come from including commonsense knowledge [29]. One important class of these propositions are of a default type, they provide values which we can associate with a variable in a situations in which our other available information about the variable doesn't provide a sufficiently precise value. More generally, commonsense knowledge is frequently used to help in situations where we don't have enough direct information about the situation. An important aspect non-monotonic propositions is that they are submissive to other categorical information. For example, assume that we are interested in John's age and we know that he is in high school. As commonsense knowledge we know that typically high school students are teenagers. Using this piece of commonsense knowledge our best guess is to conclude that John is a teenager. However assume now that we have the additional knowledge that John is in his thirties. In this situation we want to the default based inference that John is a teenager to defer to the categorical knowledge, that John is in his thirties. Implicit in this situation is some idea of a prioritization of knowledge. In particular the default knowledge has a lower priority then categorical knowledge.

Another use of default propositions are as a simplifying tool. Here they are used to allow us to make inferences without requiring us to obtain some antecedent pieces of knowledge with complete certainty. In this use the possibility of some antecedent being true, it has not been falsified, is sufficient grounds for making an inference. This use often acts to simplify our inference process. An illustrative example of this is a rule about crossing streets: "if the light is red and the driver is sober then assume they will stop." Often when using this we can't

realistically ascertain with complete certainty that a driver is sober, we usually settle for less certainty, he doesn't manifest any drunkenness. Here we can cross without having to test each driver.

The inclusion of non-monotonic propositions can add considerable complexity to the reasoning process. This complexity results from the fact that if P_1 is non-monotonic and $(P_1, P_2) \mapsto Q$ it may be the case that using the whole knowledge base, $(P_1, P_2, \dots, P_n)$, doesn't justify the inferring of Q . Thus, with use of non-monotonic knowledge, all local inferences may not be valid. A deeper understanding of the situation can benefit from the introduction of the concept of a "tail" associated with a non-monotonic proposition. Let $P = (P_1, P_2, \dots, P_n)$ be our knowledge base and let P^* be a proposition in P . The tail T_{P^*} associated with P^* is a subset of P consisting of the minimal number of propositions such that any proposition inferred from P^* along with its tail is also inferable from the whole knowledge base⁸. The point being that protoforms involving non-monotonic knowledge must include all members of its tail⁹. This can lead to very complex protoforms that may involve many propositions and is contextually dependent upon what propositions exist in the KB. We observe that categorical propositions have empty tails.

An important class of commonsense knowledge can be represented by propositions that have a possibilistic component. As we shall subsequently see this allows us to make inferences using the fact that some antecedent is possibly true rather than certainly true. This softening of the requirements brings with it the potential that we are wrong and hence introduces non-monotonicity. In the following we shall look at some basic protoforms associated with propositions having a possibilistic component.

In order to get an understanding of an paradigm that plays a fundamental role in many protoforms associated with possibilistic non-monotonic reasoning patterns we review the classical

⁸We note that this is true for any subset that contains the tail, without introducing any other non-monotonic propositions not in the tail.

⁹Here we are assuming a constant knowledge base. In dynamic knowledge bases the addition of new knowledge can be seen as opening the possibility that the original tail used in making an inference has been increased as a result of getting new information.

modus ponens protoform. In modus ponens we have $(P_1, P_2) \mapsto Q$ where

P1: If V is A then U is F

P2: V is B

Q: U is D where $D(y) = (1 - \text{Cert}[A|B]) \vee F(y)$

In this protoform we see require that we are **certain** that V is A to infer U is F

Note: It is interesting to note that we can view U is D as a kind of certainty weighted proposition:

U is F is α certain

where $\alpha = \text{Cert}[A|B] = 1 - \text{Poss}[\bar{A}|B]$

We now consider a related non-monotonic protoform

NMPF-1: Basic Non-Monotonic Possibilistic Protoform: $(P^*, P_2) \mapsto Q$ where¹⁰

P*: If V is A **is possible** then U is F

P2: V is B

Q: U is E where $E(y) = (1 - \text{Poss}[A|B]) \vee F(y)$

Here we see that we infer F if A is possible given we have B . Note in the usual modus ponens we require A is certain given we have B . Since $\text{Poss}[A|B] \geq \text{Cert}[A|B]$ ¹¹, then this is easier to satisfy.

A simple example illustrates this. Let V have domain $X = \{a, b, c, d\}$. Let $A = \{a\}$ and let $B = \{a, b\}$. In the case of models ponens we get $D(y) = (1 - \text{Cert}[A|B]) \vee F(y)$. Since $\text{Cert}[A|B] = 1 - \text{Poss}[\bar{A}|B] = 1 - \text{Max}_X[\bar{A}(x) \wedge B(x)] = 0$ then $D(y) = (1 - \text{Cert}[A|B]) \vee F(y) = 1$ and we can't infer anything. In the case of P* we get $E(y) = (1 - \text{Poss}[A|B]) \vee F(y)$. Since $\text{Poss}[A|B] = \text{Max}_X[A(x) \wedge B(x)] = 1$ then $E(y) = F(y)$. Thus here we infer U is F

However we also can easily illustrate the non-monotonicity associated with the use of P*. Assume that we get the additional piece of information P3: V is G where $G = \{b, c\}$. Here now the tail associated with P* consists of both P2 and P3. In this case we must use (P^*, P_2, P_3) and our knowledge about V is that V is H where $H = G \cap B = \{b\}$. In this case $\text{Poss}[A|H] = 0$, hence

¹⁰We are implicitly assuming here that P_2 is the tail of P*

¹¹This holds for normal sets

$E(y) = 1 - 0 \vee F(y) = 1$, $E = X$, and thus we can't infer anything.

A slightly more general version of the above protoform can be described. Let P^* be as in the preceding, If V is A **is possible** then U is F . Let T_{P^*} be the tail of P^* . Assume $T_{P^*} \mapsto V$ is H then $(P^*, T_{P^*}) \mapsto U$ is E where $E(y) = (1 - \text{Poss}[A|H]) \vee F(y)$. Let us formally denote this protoform.

NMPF-1b: General Non-Monotonic Possibilistic Protoform: $(P^*, T_{P^*}) \mapsto Q$ where

P^* : If V is A **is possible** then U is F

T_{P^*} : $\mapsto V$ is H

Q : U is E where $E(y) = (1 - \text{Poss}[A|H]) \vee F(y)$

An issue here is the determination of the tail of P^* . In this case, it is easily obtained by querying the knowledge base for the value of V . Thus, in this case the appearance of P^* in a knowledge tree would generate a triangle labeled by V . This triangle would in turn generate a whole subtree which essentially constitutes T_{P^*} . One fundamental difference between this situation with P^* and that of the usual modus ponens with $P1$ is the following. While the appearance of either of them in a knowledge tree will generate a trailing subtree initiated by a triangle labeled with V , the modus ponens situation will allow us to prune this subtree and still obtain a resulting inference for U that is a valid inference while the pruning of the subtree in the case of P^* can lead to invalid inferences.

An important class of non-monotonic possibilistic protoforms relates to reasoning with rules that supply default or typical values. In reasoning with these types of propositions we must often distinguish between what we know about a variable before we apply the proposition and value of the variable after we apply the protoform. Here our prior knowledge about the variable is important in that it effects how a protoform works. This knowledge is generally obtained from propositions with a higher priority and a stronger certainty, then the protoform being implemented. The following protoform relates to the reasoning with default or typical values

NMPF-2: Typical Value Protoform

P^* : Typically (V is A)

$P2$: V is B (Prior Knowledge)

Q : V is D where $D(x) = B(x) \wedge (A(x) \vee [1 - \text{Poss}(A|B)])$

In this case we represent the knowledge typically (V is A) as the possibilistic proposition.

If V is A is **possible** (with what we already know about V) then V is A

Here what we already know is that V is B. We see that we are taking as D the conjunction of B and A is credible given B this gives us

$$D = B \cap (A \text{ is credible given } B)$$

A slightly more general version of the above protoform can be described. Let P^* be as in the preceding, Typically (V is A) Let T_{P^*} be the tail of P^* . Assume $T_{P^*} \mapsto V \text{ is } H$ then $(P^*, T_{P^*}) \mapsto V \text{ is } E$ where $E(x) = H(x) \wedge (A(x) \vee (1 - \text{Poss}(A|H)))$. We shall formally denote this protoform as **NMPF-2b: Typical Value Protoform**

Another closely related protoform is

NMPF-3: General Typical Value Protoform

P^* : If V is C and U is A **is possible** then U is A

$P2$: V is G

T_{P^*} : $\mapsto U \text{ is } B$

$P3$: $U^* \text{ is } B$

Q : U is D where $D(y) = B(x) \wedge (A(x) \vee (1 - \text{Cert}[C|G]) \vee (1 - \text{Poss}[A|B]))$

We now turn to the formulation of knowledge bases having default or other types of non-monotonic knowledge. As we already noticed when working with these types of proposition, the idea of priority of proposition plays an important role. This prioritization of propositions must be incorporated in the knowledge base. We note that the ordinary categorical type knowledge has the highest priority.

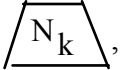
In constructing our knowledge base we shall add an additional class to J and D consisting of our possibilistic non-monotonic propositions. We shall denote this category as N. Furthermore we shall assume that the propositions in N are ordered with respect to their priority, there is a hierarchy of these propositions. To avoid unnecessary complexities at this junction we assume this is a linear ordering. Thus N is a collection of non-monotonic propositions where

$$N = N_q > N_{q-1} > \dots > N_2 > N_1.$$

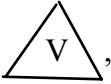
Here the bigger index then the higher the priority. This hierarchy is a reflection of the fact that the higher priority, the sooner we want to consider that information in the reasoning process.

Thus, if N_i and N_j are such that $i > j$ then we want to use N_i before we use N_j . The categorical propositions in D and J are assumed of higher priority than any element in N and as such should be considered before any element in N . We note prior consideration in the tree evaluation process means an appearance **lower** in the knowledge tree.

As is the case with the elements in J and D the propositions in N have associated variables and generally function by enforcing constraints on these variables.

The inclusion of the component N in our knowledge base effects the process of generating the knowledge tree described earlier. In the following we shall use a labeled trapezoid, , to indicate the proposition N_k in N .

We now provide a modified version of step 2 in the tree generation process to account for the inclusion of non-monotonic elements in N .

The appearance in the knowledge tree of a triangle labeled with a variable, , not having itself as an ancestor causes the emanation from the triangle of a tail consisting of .

i. A **serial** emanation of labeled trapezoids corresponding to all propositions in N having V as one of its components. This series begins with the **lowest** priority proposition and proceeds in increasing order of priority. Furthermore, emanating from each trapezoid is a special dashed triangle¹², , labeled by V as well as a regular triangle labeled by any other variable associated with the preceding trapezoid.

ii. The end of the chain of trapezoids causes the emanation from its trailing labeled dashed triangle of a parallel family of immediate successor nodes.

(a). One corresponding to each proposition in the knowledge base subset D involving the label of the dashed triangle. These are labeled rectangles.

(b). One corresponding to each proposition in J having the label of the parent dashed triangle as one of its component variables. These are labeled rectangles as described earlier. Finally if there are no appropriate elements from either a or b this causes the emanation of a

¹²We use a dashed triangle so as to indicate that this is not a new appearance of the variable but part of the current tail.

rectangle labeled by the domain of the variable, $\boxed{V \text{ is } X}$.

Figure 5 we provide an illustration of a sub-branch of a knowledge tree generated using this revised method

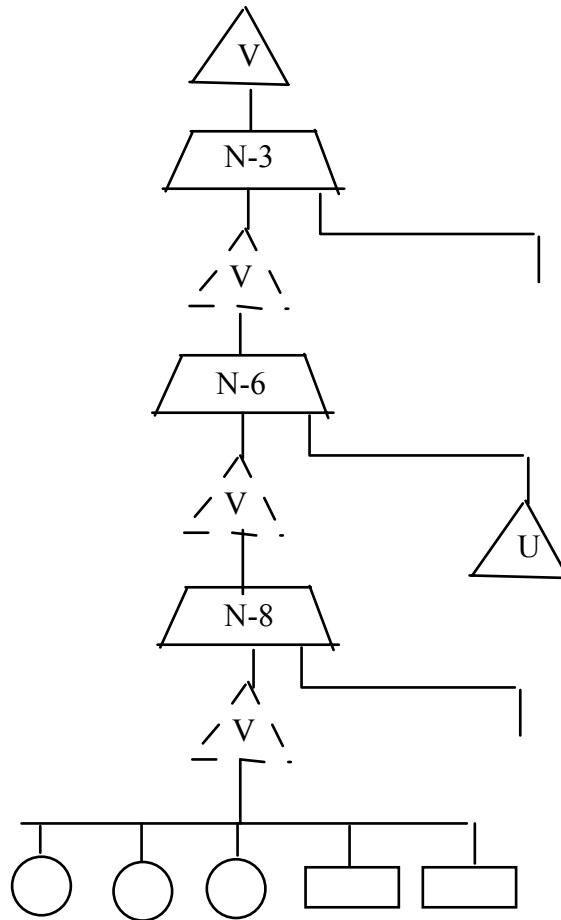


Figure 5. Subtree from Non-monotonic Propositions

The evaluation step is only modified by an instruction for handling trapezoidal nodes. We add the following item.

3. Any subtree consisting of a trapezoidal node and a family of children which are all rectangular nodes, can be replaced by a rectangular node whose label is determined by the preceding triangle and whose value is determined by the protoform associated with the proposition denoted by trapezoid.

9. Defuzzification

Let V be a variable taking its value in the domain X . In the preceding we considered the problem of finding the value of this variable from a knowledge base. In answering this question

we allowed our answer to be a fuzzy subset. More generally we allowed for an answer that contains some uncertainty. In some cases we may be required to provide an answer that is exactly one value from the set X . An example of this occurs in the widely used technology of fuzzy systems modeling [24, 30]. If the available knowledge about V allows us only to infer that V is A , where A is a fuzzy subset we must provide a process to obtain a single element from X using this information. Here shall refer to this step as defuzzification, Defuzz. Other closely related terms for this process are choice, selection and decision.

The process of defuzzification has been studied in the literature [31] and there exist a large number of methods for implementing this operation. In its most general sense the protoform associated defuzzification (choice) is an example of a non-monotonic operation. Let us look at this operation. In the following we provide a basic protoform of this operation

Protoform: For Defuzz(V is A)

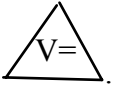
P1: V is A

Q: $V = x_q$ where $x_q = \text{Max}_{x \in X}[A(x)]$.

We note other forms of Defuzz are possible. For example if X is the real line then Filev and Yager [32] defined the general BADD defuzzification of V is A as $V = a$ where $a =$

$$x_j \frac{A(x_j)^r}{\sum_{x_k} A(x_k)^r} \text{ with } r > 0. \text{ We see if } r \rightarrow \infty \text{ we get the Max operator.}$$

Let us see the non-monotonicity implicit in this operation. Assume we infer V is A from our current knowledge base. Using defuzzification we get $\text{DEFUZZ}(V \text{ is } A) \Rightarrow V = a$. Assume we get more knowledge and subsequently obtain V is $\tilde{A}$. In this case we get $\text{DEFUZZ}(V \text{ is } \tilde{A}) \Rightarrow V = \tilde{a}$. For the defuzzification process to be monotonic we must have $\tilde{a} = a$, a situation which can not be guaranteed. At a meta level, the reason for the non-monotonicity associated with the defuzzification operation is that in choosing a precise element we are assuming information, above and beyond what we have a right to do. We are essentially using an operation related to the principle of maximal entropy which is non-monotonic.

In using our knowledge tree approach we can include these types of exact questions in the following manner. We initiate our knowledge tree with a special a triangle . The

occurrence of this equal triangle causes the emanation of the Defuzz protoform as shown in figure 6. The appearance of the ordinary triangle induces a process of finding V.

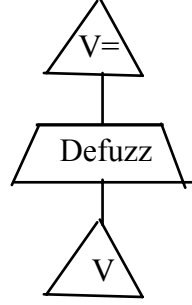


Figure 6. Defuzzification sub tree

10. Dempster-Shafer Granules for Knowledge Representation

Dempster-Shafer belief structures [33-35] provide a useful class of knowledge representation tools. For our purposes of representing and manipulating knowledge within question answering systems we shall find it convenient to use a closely related framework that we shall refer to as Dempster-Shafer granules, or simply D-S granules. A D-S granule consists of two components. The first is a collection of n categorical proposition of the type previously discussed:

$$g_1: V \text{ is } A_1, \quad g_2: V \text{ is } A_2, \dots, g_n: V \text{ is } A_n$$

Here V is a joint variable and A_i is a fuzzy subset on its domain X . We shall refer to this collection as the **body** of the granule. Thus each component in the body has a structure of a proposition from AR.

In addition, associated with a D-S granule is a mapping m that associates with each proposition g_i in the body, a value m_i such that: 1. $m_i \in [0, 1]$ and 2. $\sum_{i=1}^n m_i = 1$. We refer to

this as the qualifier of the granule.

A very special example of a D-S granule is one in which $n = 1$. In this case the body consists of a single proposition $g_1: V \text{ is } A$ and $m_1 = 1$. This is the same as an ordinary proposition.

A closely related and slightly more general structure is one in which we allow the values m_i to be fuzzy numbers $\tilde{m}_i$. These are closely related to what Zadeh calls Possibility–

Probability granules. [26]

In the following we shall describe some of the types of information that can be represented by D-S granules. One situation is when our knowledge about a variable has a random component. For example, if V is an atomic variable taking values in $X = \{x_1, \dots, x_n\}$ and we know that $\text{Prob}(x_j) = p_j$, we can represent it as a D-S granule with components $g_j: V \text{ is } \{x_j\}$ and $m_j = p_j$

More imprecise knowledge such as that the probability of subset B of X is at least α can be expressed using a D-S granule where $g_1: V \text{ is } B \ \& \ m_1 = \alpha$, $g_2: V \text{ is } X \ \& \ m_2 = 1 - \alpha$.

The case where we know that the probability of B is exactly α can be represented with the D-S granule $g_1: V \text{ is } B \ \& \ m_1 = \alpha$, $g_2: V \text{ is } \bar{B} \ \& \ m_2 = 1 - \alpha$.

Quantified propositions can be expressed using D-S structures. A simple example of this is "Most high income Americans live in urban areas." Let us see how D-S granules help us represent this type of knowledge. Let V_Z and U_Z be joint variables and let Q be a non-decreasing type quantifier such as **at least α** . Consider the proposition

For $Q \ z \in Z$ if V_Z is A then U_Z is B .

(For most people if Income is HIGH then Residence is URBAN AREA)

We can represent this by a collection of D-S granules such that for each $z \in Z$ we have

$$\begin{array}{ll} g_1: \text{If } V_Z \text{ is } A \text{ then } U_Z \text{ is } B & m_1 = \alpha \\ g_2: \text{If } V_Z \text{ is } A \text{ then } U_Z \text{ is } X & m_2 = 1 - \alpha \end{array}$$

here X is the domain of U

The meta protoform for manipulating D-S granules makes considerable use of the fact that the body of the D-S granule is made up of proposition of the type we have already dealt with.

Assume M_1 is a D-S granule with body consisting of propositions g_i for $i = 1$ to n and associated weights m_i . Let M_2 be another independent D-S granule with body consisting propositions $\tilde{g}_j$ for $j = 1$ to $\tilde{n}$ and associated weights $\tilde{m}_j$. The protoform for combining these two D-S granules denoted $\text{PF}(M_1, M_2)$ is

P1: M_1

P2: M_2

Q: M

Here M is a D-S granule whose body consists of the $n\tilde{n}$ propositions $h_{ij} = \text{PF}[g_i, \tilde{g}_j]$ and $\text{PF}[g_i, \tilde{g}_j]$

is the protoform associated with the components. In addition, associated with each object h_{ij} in M is a weight $m_i \tilde{m}_j$. We emphasize the autonomy of the operations performed on the components of the body of the granules, the g_i and $\tilde{g}_j$, from the operations performed on the qualification weights, the m_i and $\tilde{m}_j$, we refer to this as Body-Qualification autonomy, B-Q autonomy.

We illustrate this protoform with a simple example of modus ponens. Assume the granule M_1 is made up of $g_1: V \text{ is } B_1$ with $m_1 = \alpha$, and $g_2: V \text{ is } B_2$ with $m_2 = 1 - \alpha$. Let M_2 be made up of $\tilde{g}_1: \text{If } V \text{ is } A_1 \text{ then } U \text{ is } F$ with $\tilde{m}_1 = \beta$ and $\tilde{g}_2: \text{If } V \text{ is } \bar{A}_1 \text{ then } U \text{ is } X$ with $\tilde{m}_2 = 1 - \beta$

Here the internal operation $PF[g_i, \tilde{g}_j]$ is simply modus ponens. We recall that this protoform is $P1: V \text{ is } B, P2: \text{if } V \text{ is } A \text{ then } U \text{ is } F \Rightarrow U \text{ is } D$ where $D(y) = \text{Poss}[A|B] \vee F(y)$

Using this on M_1 and M_2 we get M such that

$$h_{11}: D_{11} = \text{Poss} [\bar{A}_1|B_1] \vee F(y) \quad \text{with weight } \beta\alpha$$

$$h_{21}: D_{21} = \text{Poss} [\bar{A}_1|B_2] \vee F(y) \quad \text{with weight } (1 - \alpha)\beta$$

$$h_{12}: D_{12} = X \quad \text{with weight } 1 - \beta$$

The process of reasoning with tree like structures that contain knowledge represented by D-S granules can get complex and raises some very fundamental issues notable among these are those related to independence and the lack of idempotency when using the same knowledge more than once in a tree. Rather than superficially touching on this very complex and important topic we refer the reader to the work of Shafer, Shenoy and their colleagues [36, 37] on reasoning with D-S granules in tree like structures.

11. References Part II

- [1]. Chaudhri, V. and Fikes, R., Question Answering Systems, AAAI Press: Menlo Park, 1999.
- [2]. Clark, P., Thompson, J. and Porter, B., "A knowledge based approach to question-answering," in Question Answering Systems, edited by Chaudhri, V. and Fikes, R., AAAI Press: Menlo Park, 43-51, 1999.
- [3]. Maybury, M., New Directions in Question Answering: Papers from the 2003 AAAI Spring Symposium, AAAI Press: Menlo Park, 2003.
- [4]. McGuinness, D. L., "Knowledge representation for question answering," in New Directions in Question Answering: Papers from the 2003 AAAI Spring Symposium, edited by Maybury, M., AAAI Press: Menlo Park, 75-77, 2003.
- [5]. Chu-Carroll, J., Ferrucci, D., Pragger, J. and Welty, C., "Hybridization in question answering systems," in New Directions in Question Answering: Papers from the 2003 AAAI Spring Symposium, edited by Maybury, M., AAAI Press: Menlo Park, 116-121, 2003.
- [6]. Zadeh, L. A., "Web intelligence and world knowledge-The concept of Web IQ (WIQ)," Presentation Abstract Berkeley Institute in Soft Computing (BISC), University of California, 2004.
- [7]. Zadeh, L. A., "Fuzzy logic = computing with words," IEEE Transactions on Fuzzy Systems 4, 103-111, 1996.
- [8]. Yager, R. R., "On the retranslation process in Zadeh's paradigm of computing with words," IEEE Transactions on Systems, Man and Cybernetics: Part B 34, 1184-1195, 2004.
- [9]. Zadeh, L. A., "A theory of approximate reasoning," in Machine Intelligence, Vol. 9, edited by Hayes, J., Michie, D. and Mikulich, L. I., Halstead Press: New York, 149-194, 1979.
- [10]. Zadeh, L. A., "Precisiated natural language (PNL)," Report Berkeley Initiative on Soft Computing (BISC) University of California, Berkeley, 2003.
- [11]. Zadeh, L. A., "Web intelligence and fuzzy logic-the concept of Web IIQ (WIQ)," [Http://wi-consortium.org/pdf/zadah-wiq.pdf](http://wi-consortium.org/pdf/zadah-wiq.pdf), 2003.
- [12]. Yager, R. R., "Querying knowledge base systems with linguistic information via knowledge trees," Int. J. of Man-Machine Studies 19, 73-95, 1983.
- [13]. Yager, R. R., "Knowledge trees in complex knowledge bases," Fuzzy Sets and Systems 15,

45-64, 1985.

[14]. 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.

[15]. Klement, E. P., Mesiar, R. and Pap, E., *Triangular Norms*, Kluwer Academic Publishers: Dordrecht, 2000.

[16]. Zadeh, L. A., "Fuzzy sets," *Information and Control* 8, 338-353, 1965.

[17]. Yager, R. R., "A characterization of the extension principle," *Fuzzy Sets and Systems* 18, 205-217, 1986.

[18]. Xu, Y., Kerre, E. E., Ruan, D. and Song, Z., "Fuzzy reasoning based on the extension principle," *International Journal of Intelligent Systems* 16, 469-495, 2001.

[19]. Yager, R. R., "On measures of specificity," in *Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration with Applications*, edited by Kaynak, O., Zadeh, L. A., Turksen, B. and Rudas, I. J., Springer-Verlag: Berlin, 94-113, 1998.

[20]. Dubois, D. and Prade, H., "The principle of minimum specificity as a basis for evidential reasoning," *Uncertainty in Knowledge-Based Systems*, Bouchon, B. & Yager R.R., (Eds.), Springer-Verlag: Berlin, 75 - 84, 1987.

[21]. 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.

[22]. Klir, G. J. and Yuan, B., *Fuzzy Sets, Fuzzy Logic and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh*, World Scientific: Singapore, 1996.

[23]. Dunham, M., *Data Mining*, Prentice Hall: Upper Saddle River, NJ, 2003.

[24]. Yager, R. R. and Filev, D. P., *Essentials of Fuzzy Modeling and Control*, John Wiley: New York, 1994.

[25]. Yager, R. R., "Approximate reasoning as a basis for computing with words," in *Computing with Words in Information/Intelligent Systems* 1, edited by Zadeh, L. A. and Kacprzyk, J., Springer-Verlag: Heidelberg, 50-77, 1999.

[26]. Zadeh, L. A., "Toward a perception-based theory of probabilistic reasoning with imprecise probabilities," *Journal of Statistical Planning and Inference* 105, 233-264, 2002.

- [27]. Reiter, R., "Nonmonotonic reasoning," in Annual Reviews of Computer Science vol 2, Palo Alto: Annual Reviews, 147-186, 1987.
- [28]. Ginsberg, M. L., Readings in Nonmonotonic Reasoning, Morgan Kaufmann: Los Altos, CA, 1987.
- [29]. McCarthy, J., "Applications of circumscription to formalizing common sense knowledge," Artificial Intelligence 28, 89-116, 1986.
- [30]. Ross, T. J., Fuzzy Logic with Engineering Applications, McGraw Hill: New York, 1995.
- [31]. Yager, R. R. and Filev, D. P., "On the issue of defuzzification and selection based on a fuzzy set," Fuzzy Sets and Systems 55, 255-272, 1993.
- [32]. Filev, D. and Yager, R. R., "A generalized defuzzification method under BAD distributions," International Journal of Intelligent Systems 6, 687-697, 1991.
- [33]. Dempster, A. P., "Upper and lower probabilities induced by a multi-valued mapping," Ann. of Mathematical Statistics 38, 325-339, 1967.
- [34]. Shafer, G., A Mathematical Theory of Evidence, Princeton University Press: Princeton, N.J., 1976.
- [35]. Smets, P. and Kennes, R., "The transferable belief model," Artificial Intelligence 66, 191-234, 1994.
- [36]. Shenoy, P. P. and Shafer, G., "Propagating belief functions with local computations," IEEE Expert Vol 1, No 3, 43-52, 1986.
- [37]. Shafer, G., Shenoy, P. P. and Melloui, K., "Propagating belief functions in qualitative Markov trees," International Journal of Approximate Reasoning 1, 349-400, 1987.

PART III. UNCERTAINTY MANAGEMENT FOR INTELLIGENCE ANALYSIS

1. Introduction

In his report to Congress [1], David A. Kay, who led the US government's efforts to find evidence of Iraq's illicit weapons programs, reported that the current intelligence systems dealing with weapons of mass destruction are increasingly based on limited information. In light of this situation, he indicated that modern intelligence analysis systems need a way for an analyst to say, "I don't have enough information to make a judgment," a capacity that he felt the current intelligence systems do not possess. Central to attaining this capability is the ability to deal with uncertain and imprecise information. We believe that fuzzy logic with its focus on uncertainty can help. It has the ability to simultaneously exploit both precise formal measurements of the type obtained from state of the art electronic and mechanical monitoring devices as well the type of imprecise information obtained from human sources which is often perception based and expressed in linguistic terms. Here, we begin to look at the possibilities of using fuzzy logic [2] and related soft computing technologies to provide the tools necessary to supply this capability to intelligence analysts. As we shall subsequently see, the dual measures of possibility and certainty [3] provide a useful way of formalizing the concept of not knowing with certainty.

2. Variables and Question Answering

By a variable we shall mean an attribute associated with some specific object. Thus, if V is a variable then $V \equiv \text{attribute (object)}$. John's age and the number of nuclear devices possessed by North Korea are examples of variables. In the first case, the attribute is age and the object is John. In the second case, the attribute is the quantity of nuclear devices and the object is North Korea. Typically with a variable, we assume it has a domain, X , consisting of the set of possible values. In many situations a task of great interest is the answering of some question about a variable. For example, is John over 65? Another closely related task is that of making a decision in which knowledge about a variable is central to the decision. For example, a bartender deciding

whether to serve John a drink must ascertain that his age is at least 21.

We emphasize the distinction between the task of finding the value of a variable and that of answering a question about a variable. Clearly, although knowing the value of a variable can help in answering a question, it is not always necessary. That is, there can to some uncertainty and still we can answer a question about a variable with certainty.

In order to be able to answer a question about the value of a variable, we must draw upon all our sources of information about the variable. Figure 7 illustrates this situation. The information provided by these sources may be related to the variable of interest in a number of different ways. It may be information directly about the value of the variable of interest, an observation on the age of John. An example of this is a birth certificate. It may be about the attribute without specific reference to John. Human beings typically live no more than about 85 years. It may be information about the value of another attribute associated with John, "the color of John's hair is grey." It may be information relating the variable of interest to other attributes or variables, "John is five years younger than Mary." Furthermore, each of these pieces of information may have different degrees of credibility. In addition, the information from the sources may be obtained from precise measurement or may be based upon perceptions and observations. It may be expressed formally or in linguistic terms.

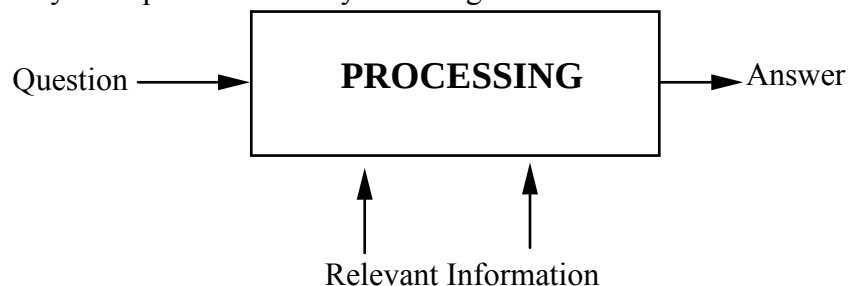


Figure 7. Task of answering a question

The process of answering the question about the attribute involves combining this information. In some cases, this process may involve a fusing of the available information to obtain an effective value for the variable. The answer to the question of interest is obtained with respect this fused information. In some cases, the answer to the question may be obtained using a process that doesn't depend upon obtaining an effective value.

3. Basic Knowledge Representation Using Fuzzy Sets

Among the central tasks involved in providing an answer to a question is the representation of the relevant information in a manner that allows formal manipulation. The representational language should be rich enough to allow the modeling of different types of information. Fuzzy subsets provide the basis for a very expressive framework for the representation of a wide body of knowledge. This knowledge can be either precise or imprecise. It can be used to represent knowledge expressed using linguistic values. Here, we shall briefly discuss this representational capability, however, we note the extensive literature on this subject and specifically refer the reader to the work of Zadeh under his paradigm of computing with words [4, 5] and the related theory of approximate reasoning [6-10].

Within the framework provided by fuzzy sets knowledge about the value of a variable V is expressed using a statement V is A where A is a fuzzy subset of the domain X . The use of this type of representation can be seen as a generalization of the idea of imposing a constraint on the value of V , such as saying that V lies in the subset B , when B is a crisp subset of X . An example of this is saying John's age is between 25 and 35. The use of fuzzy subsets allows for a grading of this concept of V lying in the set B . Hence, the statement V is A manifests a constraint on the value of the variable V . The assignment of a fuzzy subset A to the variable induces a possibility distribution on X such that $A(x)$ indicates the possibility that x is the value of V .

These types of fuzzy assignments can arise in, although are not restricted to, situations in which the information about the value of the variable is initially expressed in linguistic terms. An example of this would be the observation that John is middle aged. In this case, the fuzzy subset is the representation of the linguistic term middle-age. Here, the definition of the fuzzy subset A is such that for $x \in X$ the membership grade $A(x)$ is the compatibility of the age x with the concept being represented, middle-age. We should note that while the use of a crisp subset allows for a representation of uncertainty of value fuzzy subsets allows for a more sophisticated representation. They allow for more than just a simple distinction between those values that are possible and those that are impossible - it allows a grading of possibility

We note that, in the case where $A = \{x\}$, then the statement V is A is equivalent to saying that $V = x$. Another special case is when $A = X$. Here, the statement V is X is equivalent to

saying that we don't know. If B is some crisp subset of X , then the statement V is B is equivalent to saying the value of V lies in B . The situation when $A = \emptyset$, the null set, corresponds to the case where we are saying our knowledge is that V is not in X . This situation indicates a complete conflict with our assumption that V must take its value in X . More generally, if A is such that $\text{Max}_X A(x) < 1$ then we have some degree of conflict with the assumption that V has X as its domain. We shall say a fuzzy subset is normal if there exists at least one $x \in X$ so that $A(x) = 1$. If $\text{Max}_X A(x) < 1$ we say A is subnormal.

Consider the situation where we have the knowledge that V lies in B , V is B , where B is the crisp subset X . From this, we can naturally infer that V lies in E where $B \subseteq E$, here E is any set containing B . Thus, knowing that John's age is between 25 and 35 allows us to infer that John's age is between 10 and 50. In the fuzzy framework that generalizes to what is called **the entailment principle** [11]. This principle states that, from the knowledge that V is A , we can infer V is F where $A \subseteq F$. We recall that for fuzzy subsets $A \subseteq F$ if $A(x) \leq F(x)$ for all x .

Clearly, the knowledge that V is contained in $[25, 35]$ is more informative less uncertain, then the knowledge that V is contained in $[10, 50]$. Furthermore the statement that V is 25 is even more informative than either of the preceding, as it contains no uncertainty. In [12, 13, 14], we introduced the concept of specificity to measure the amount of information contained in a fuzzy proportion V is A . Specificity is inversely related to the idea of uncertainty, the more specific the more certain our knowledge.

Definition¹³: Assume A is a fuzzy subset over X . Let x^* be such that $A(x^*) = \text{Max}_X[A(x)]$, it is an element having the maximal membership grade in A . Let $\hat{A}$ be the average membership grade of A over the space $X - \{x^*\}$, it is the average over all elements except x^* . The specificity of A , denoted $\text{Sp}(A)$ is defined as $\text{Sp}(A) = A(x^*) - \hat{A}$, it is the difference between the highest membership grade and the average of all the other elements.

Note 1: If more than one element attains the highest membership grade then all except one of

¹³While a number of different formal definitions have been suggested we shall this one to be useful for our purposes as it simply captures the basic idea of the concept specificity

these are used to find the average.

Note 2: We note the specificity of the statement V is A is equal to $Sp(A)$. Thus we use the terms $Sp(V \text{ is } A)$ interchangeably with $Sp(A)$

We can observe some properties of $Sp(A)$:

1. It lies in unit interval. $0 \leq Sp(A) \leq 1$.
2. $Sp(A) = 1$ iff there exists one element x^* such that $A(x^*) = 1$ and all other elements have $A(x) = 0$.
3. If $A(x) = c$ for all x , then $Sp(A) = 0$.
4. If A and B are two normal fuzzy subsets, they have one element with membership grade 1 and $A \supseteq B$ then $Sp(B) \geq Sp(A)$. Thus containment in the case of normality means an increase of specificity.

Note: Essentially specificity measures the degree to which V is A points to one and only one element as the value of V .

As we shall subsequently see, the measure of specificity can play an important role in the processing of information. Consider the statement V is A where A is a normal fuzzy subset, that is there exists at least one element that has full possibility of having the value of V . We earlier noted that if B_1 is such that $B_1 \subset A$ as well as remaining normal then V is B_1 provides more information about the value of V than the original statement V is A . Essentially in this case we introduced some degree of clarity, we reduced the uncertainty by reducing the possibility of some elements while still leaving the possibility of finding a solution. On the other hand, if $B_2 \supset A$ then V is B_2 provides less information than V is A . In this case, we have reduced our certainty because we have added more possibilities. A third situation is where we have V is B_3 but with $B_3 \subset A$ but with B_3 subnormal $\text{Max}_X[B_3(x)] < 1$. We don't have a solution completely compatible with the assumption that V lies in X . In this case, we can possibly have less information than the original statement V is A , $Sp(B_3) \leq Sp(A)$. More generally, a reduction of specificity (certainty) in our knowledge can come about from two sources, one being increased possibility and the other being an increase in conflict with the assumption that its value lies in the given domain?

4. On the Measures of Possibility and Certainty

As we earlier noted, a task of great interest is the answering of a question about some variable. Here, we shall concern ourselves with this issue when our knowledge about the variable as well as the question is represented using the preceding representation. Given the knowledge that V is A our task is the determination of the validity of the statement V is B .

In order to build our intuition, we shall initially consider the case in which the sets A and B are crisp sets. There are two situations regarding our knowledge of A . In the first, we have no uncertainty regarding our knowledge of V , $V = x_1$, here $A = \{x_1\}$. In this situation, we can very clearly answer our question about the truth of the statement V is B . If $x_1 \in B$ then the answer is yes, if $x_1 \notin B$ then the answer is no. This exact information with respect to the value of V leads to precise answers.

The second case is where A is not a singleton, there exists some uncertainty about the value of V . This is the more typical situation, as noted by Kay [1], in intelligence analysis. The uncertainty associated with the knowledge that V is A makes the clear determination of whether another statement V is B is true or false not always attainable. Using figure 8 can help us understand the situation when A is uncertain.

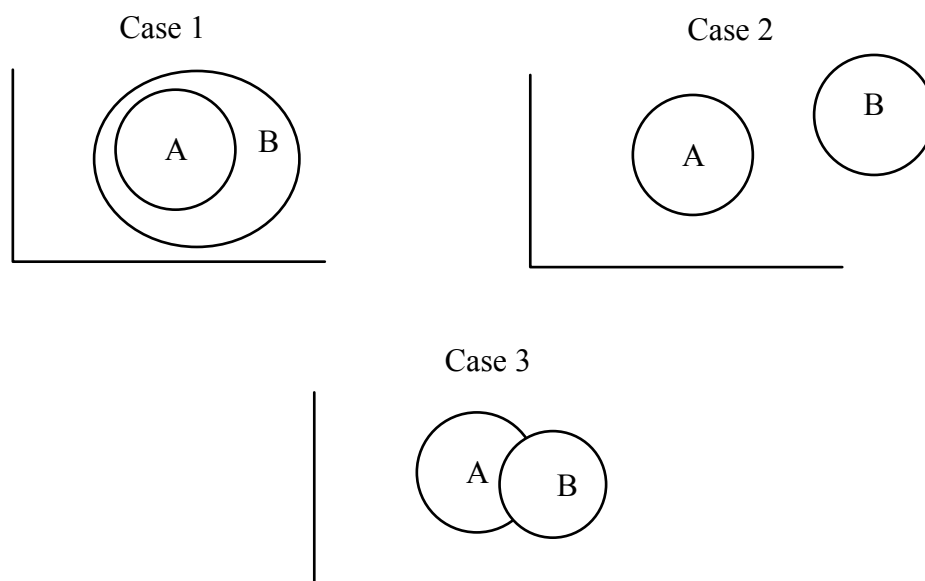


Figure 8. Different relations between knowledge and question

We see in case 1 knowing that V is A assures us that V is B is valid. In case 2, knowing that V is A assures us that V is B is not true. Finally, in case 3, we can't tell. Thus we observe

from this crisp environment that we have the following rules regarding the determination of truth of the statement V is B given V is A :

If $A \subseteq B$ then the answer is yes

If $A \cap B = \emptyset$ then the answer is no

If $A \cap B \neq \emptyset$ and $A \not\subseteq B$ then the answer is I don't know

Thus the attainment of a clear answer to questions in the face of uncertainty in our knowledge is not always attainable. We note this situation holds even in the special case when B is a singleton. We see that asking if $V = 30$ if we only know that $V \in [25, 40]$ can't be answered yes or no, the appropriate answer is *I don't know*.

In the fuzzy set environment more sophisticated tools are needed to address this problem. Two measures have been introduced by Zadeh [3] to help. These are the measures of possibility and certainty. We note that Dubois and Prade [15, 16] refer to the measure of certainty as the measure of necessity. In the following, we shall, unless otherwise stated, assume A and B are normal

The possibility that V is B given V is A is denoted by $\text{Poss}[V \text{ is } B / V \text{ is } A]$ and is defined as $\text{Poss}[V \text{ is } B / V \text{ is } A] = \text{Max}_x[D(x)]$ where $D(x) = \text{Min}[A(x), B(x)]$. Thus

$$\text{Poss}[V \text{ is } B / V \text{ is } A] = \text{Max}_x[A(x) \wedge B(x)]$$

Since $D = A \cap B$, we see that $\text{Poss}[V \text{ is } B / V \text{ is } A]$ is the maximum degree of intersection between A and B .

The second measure introduced by Zadeh is the measure of certainty. We define this as

$$\text{Cert}[V \text{ is } B / V \text{ is } A] = 1 - \text{Poss}[V \text{ is not } B / V \text{ is } A] = 1 - \text{Max}_x[A(x) \wedge \overline{B}(x)]$$

With some manipulation we attain

$$\text{Cert}[V \text{ is } B / V \text{ is } A] = \text{Min}_x[\overline{A}(x) \vee B(x)]$$

We observe that $\text{Cert}[V \text{ is } B / V \text{ is } A]$ is indicating the degree to which A is contained in B . That is if A is contained in B the knowledge that V is in A assures us that it is in B .

These measures of possibility and certainty can be seen as respectively providing upper and lower (optimistic and pessimistic) bounds, on the answer to the question of whether V is B is true given we know that V is A .

We note that if A is a normal fuzzy subset, there exists an x^* such that $A(x^*) = 1$, then

$\text{Poss}[V \text{ is } B / V \text{ is } A] \geq B(x^*)$ and $\text{Cert}[V \text{ is } B / V \text{ is } A] \leq B(x^*)$. Thus we see when A is normal, we have $\text{Cert}[V \text{ is } B / V \text{ is } A] \leq \text{Poss}[V \text{ is } B / V \text{ is } A]$.

Let us look at these measures for some special cases of A and B . We first consider the case where A and B are crisp. In this case, $\text{Cert}[B/A]$ and $\text{Poss}[B/A]$ must be either one or zero. We see that if $\text{Cert}[B/A] = 1$ then $\text{Poss}[B/A] = 1$ and this corresponds to the case where $V \text{ is } B$ is true. If $\text{Poss}[B/A] = 0$ then $\text{Cert}[B/A] = 0$ and we know that $V \text{ is } B$ is false. If $\text{Cert}[B/A] = 0$ while $\text{Poss}[B/A] = 1$ then we are in the situation in which the answer is unknown.

Consider the situation where B is a crisp subset and A can be fuzzy. Here we have that

$$\text{Poss}[V \text{ is } B / V \text{ is } A] = \text{Max}_{x \in B} [A(x)]$$

$$\text{Cert}[V \text{ is } B / V \text{ is } A] = \text{Min}_{x \notin B} [\bar{A}(x)] = 1 - \text{Max}_{x \notin B} [A(x)]$$

An important special case of this is where $B = \{x^*\}$, here we are interested in determining whether V is equal to some particular value. In this case we see that $\text{Poss}[V \text{ is } B / V \text{ is } A] = A(x^*)$ and $\text{Cert}[V \text{ is } B / V \text{ is } A] = \text{Min}_{x \neq x^*} [\bar{A}(x)] = 1 - \text{Max}_{x \neq x^*} [A(x)]$. The certainty is the negation of largest possibility of value not equal to x^* . We also observe that if $A(x^*) \neq 1$ then we must have $\text{Cert}[V \text{ is } x^* / V \text{ is } A] = 0$. This follows since with normal sets there exists some element $x_1 \neq x^*$ with $A(x_1) = 1$ and hence $1 - \text{Max}_{x \neq x^*} [A(x)] = 0$.

We also observe in the case where $X = \{x_1, x_2\}$, if we ask is $V = x_1$, we see that $\text{Cert}[V \text{ is } x_1 / V \text{ is } A] = 1 - A(x_2)$. It is simply the negation of the possibility of the other element.

Consider now the special case where A is a crisp set. Here

$$\text{Poss}[V \text{ is } B / V \text{ is } A] = \text{Max}_{x \in A} [B(x)]$$

$$\text{Cert}[V \text{ is } B / V \text{ is } A] = \text{Min}_x [\bar{A}(x) \vee B(x)] = \text{Min}_{x \in A} [B(x)].$$

If additionally we assume that $A = \{x_1\}$, the value of V is exactly known, then

$$\text{Poss}[V \text{ is } B / V \text{ is } A] = B(x_1) \text{ and } \text{Cert}[V \text{ is } B / V \text{ is } A] = B(x_1)$$

then $B(x_1)$ is the validity of the statement that $V \text{ is } B$.

Some clarification may be useful here. What we have shown is that generally when our information about a variable, $V \text{ is } A$, has some uncertainty the answer to any question about the truth of the statement $V \text{ is } B$ lies in some interval. Thus if A is not a singleton the truth of $V \text{ is } B$ lies in the interval $[l, u]$ where l is the certainty of $V \text{ is } B$ and u is the possibility that $V \text{ is } B$. Here $[l, u]$ is a subset of the unit interval. On the other hand if A is a singleton then the truth of $V \text{ is } B$

is a precise value b , in the unit interval. Additionally in the case when A is a singleton we have that B is a crisp set then b equals one or zero. The important point here is that there are two manifestations. One being as a result of our lack of certainty regarding the knowledge of A , it is granular and it is not a singleton, generally this results in an interval for our truth value. The second issue is related to a lack of crispness. The sets involved are fuzzy, this generally introduces aspects of multi-values logic, l and b are not necessarily one or zero but can be anywhere in unit interval.

In cases where the decision process requires a more precise determination of the validity of the proposition V is B then provided by the interval $[u, l]$ we must provide some means around this difficulty. However, we must emphasize that the actual processing of the information about the variable V has left us with some uncertainty. In some cases, we may be able to draw upon techniques from decision-making under uncertainty [17] to help make decisions in this kind of environment. First we want to make a clear distinction between the analyst, such as an intelligence analyst, and what we shall call the executive. It is the executive who makes the decision using as some of his input the information provided by the analyst. While it is not our purpose here to go into great detail about the executive task of decision making, since we are more interested in the analysis task, we shall make a few comments.

In making a decision, such as whether we should preemptively strike an adversary, in addition to the information provided by the analyst about the state of V which may be uncertain, such whether they have weapons of mass destruction, an executive generally draws upon two other types of information [18]. The first type of information is related to the costs or payoffs associated with the choice of an action and the possible states of the uncertain variable V . Formally this is often expressed using a payoff matrix. The second type of information is related to what we call the decision makers' attitudinal character [19]. This component of the decision process has an extremely subjective nature. It is here people can have strong differences of opinion, which are purely value and preference driven. Thus one executive in the face of an uncertainty regarding the relevant variable may decide to act in a way that defends against the worst possibility, the so-called Max-Min decision maker [20]. Given an appropriate use of this with respect to the available knowledge of the possible outcomes this type of decision cannot be

said to be right or wrong. The point we want to make here is that uncertainty in our knowledge provides space for the inclusion of subjective choices by the executive making the decision. A simple example of this may involve preparing for a party in which we are not sure whether 20 or 500 people are coming. Clearly if we prepare for 500 and only 20 show up then we wasted a lot of money. On the other hand if we prepare for 20 and 500 show up we have some embarrassment. The choice of how many people to prepare for is based on the subjective preferences of the party giver, the executive, with regard to being embarrassed or wasting money, there is no right or wrong decision.

In the preceding, we assumed normality with respect to all the sets involved; all sets were assumed to have at least one element with a membership grade 1. Here, we shall make some comments about the situation with respect to sub-normality, $\text{Max}_x[A(x)] < 1$. First, we note sub-normality is generally a reflection of some conflict. Sub-normality usually arises from the combination of information from different sources when there is some conflict between the observations of the sources. A second way it can arise is when the information provided by an individual source is in conflict with the assumption about the domain of a variable. This type of situation occurs less frequently. Thus we shall assume that our primary information supplied by the individual sources is normal.

In formal reasoning systems, based on logic, the appearance of conflicting statements results in a situation in which we can infer anything, we conclude that everything is true. Our system has a similar property. Assume V is A and $A = \emptyset$ then for any statement V is B we have $\text{Cert}[V \text{ is } B / V \text{ is } A] = \text{Min}_x[\tilde{A}(x)] \vee B(x) = 1$. Thus, in the face of complete conflict, everything is certain. On the other hand, with $A = \emptyset$, $\text{Poss}[V \text{ is } B / V \text{ is } A] = \text{Max}[A(x) \wedge B(x)] = 0$. Thus nothing is possible but everything is certain. In order to avoid this difficulty of having the certainty greater than the possibility we shall use as our definition of certainty

$$\text{Cert}[V \text{ is } B / V \text{ is } A] = (\text{Min}_x[\tilde{A}(x)] \vee B(x)) \wedge (\text{Max}_x[A(x)])$$

If A is normal, this just is the definition for certainty we previously used, $\text{Cert}[V \text{ is } B / V \text{ is } A] = \text{Min}_x[\tilde{A}(x)] \vee B(x)$. If $A = \emptyset$ we get $\text{Cert}[V \text{ is } B / V \text{ is } A] = 0$. More generally using this definition we always get $\text{Cert}[V \text{ is } B / V \text{ is } A] \leq \text{Poss}[V \text{ is } B / V \text{ is } A]$

We make one further comment with respect to normality. Previously, we defined the

entailment principle as saying from V is A we can infer V is B where $A \subseteq B$. This must be modified to say that B must satisfy the $\text{Max}_X[B(x)] \leq \text{Max}_X[A(x)]$. Thus if A is normal no additional restriction exists on B . On the other hand if $\text{Max}_X[A(x)] = a$ then any statement V is B inferred from this must satisfy both $A \subseteq B$ and $\text{Max}_X[B(x)] \leq a$. The inferred set B can't be more possible then the original set A

5. Hedging on our Data

In the preceding, we introduced V is A as a structure for representing uncertain knowledge where A is a fuzzy subset of the domain X of V . We indicated that this generalized the idea of knowing that V lies in some subset. More generally this formulation imposes some constraint on the value that V can assume. One question we considered was determining whether the proposition V is x^* is valid given the knowledge V is A . We showed that, with uncertainty in our knowledge about V , the best we could do was to put some bounds on the truth of the hypothesis that V is x^* . In particular $\text{Poss}[V \text{ is } x^* / V \text{ is } A] = A(x^*)$ provided an upper bound and $\text{Cert}[V \text{ is } x^* / V \text{ is } A] = \text{Min}_{x \neq x^*} [\bar{A}(x)] = 1 - \text{Max}_{x \neq x^*} [A(x)]$ provided a lower bound. If we let $B = \{x^*\}$ then **not** x^* is $\bar{B} = X - \{x^*\}$ and we see that $\text{Cert}[V \text{ is } x^* / V \text{ is } A] = 1 - \text{Poss}[V \text{ is } \underline{\text{not}} x^* / V \text{ is } A]$.

We now consider the situation where we want to hedge on the knowledge that V is A . We let $\alpha \in [0, 1]$ indicate the degree of confidence we attribute to the proposition V is A , that is our knowledge V is A is α -certain. In [21] it was suggested that one can express this hedged knowledge as a proposition V is F where $F(x) = \text{Max}[A(x), \bar{\alpha}] = A(x) \vee \bar{\alpha}$. Since $\bar{\alpha} = 1 - \alpha$ we see if $\alpha = 1$ then $F(x) = A(x)$ and we get our original unhedged proposition. If $\alpha = 0$ then $\bar{\alpha} = 1$ and $F(x) = 1$ for all x . Here, our statement V is F effectively carries no information. Essentially this hedging loosens the constraint on the variable V .

In the following we shall let A^* denote the fuzzy set such that $A^*(x) = 1$ if $x = x^*$ and $A^*(x) = 0$ if $x \neq x^*$.

Let us see what happens to our measures of possibility and certainty in this hedged situation

$$\text{Poss}[V \text{ is } x^* / V \text{ is } A \text{ is } \alpha\text{-cert}] = \text{Max}_x[(A(x) \vee \bar{\alpha}) \wedge A^*(x)] = A(x^*) \vee \bar{\alpha}$$

In the case of certainty we have

$$\text{Cert}[V \text{ is } x^* / V \text{ is } A \text{ is } \alpha\text{-cert}] = 1 - \text{Max}_{x \neq x^*}[F(x)] = \text{Min}_{x \neq x^*}[\bar{F}(x)].$$

Since $F(x) = \bar{\alpha} \vee A(x)$ then $\text{Cert}[V \text{ is } x^* / V \text{ is } A \text{ is } \alpha\text{-cert}] = 1 - \text{Max}_{x \neq x^*}[\bar{\alpha} \vee A(x)] = 1 - \bar{\alpha} \vee \text{Max}_{x \neq x^*}[A(x)] = \alpha \wedge \text{Min}_{x \neq x^*}[\bar{A}(x)] = \text{Min}[\alpha, \text{Min}_{x \neq x^*}[\bar{A}(x)]]$. More intuitively we see that $\text{Cert}[V \text{ is } x^* / V \text{ is } A \text{ is } \alpha\text{-cert}] = \text{Min}[\alpha, \text{Cert}[V \text{ is } x^* / V \text{ is } A]]$, it is the smaller of α and the certainty of the unhedged situation.

In anticipation of what we shall do in the following, we shall refer to these as optimistic and pessimistic measures

$$\mathbf{Opt}(V \text{ is } x^* / V \text{ is } A \text{ is } \alpha\text{-cert}) = \text{Poss}[V \text{ is } x^* / V \text{ is } A \text{ is } \alpha\text{-cert}] = A(x^*) \vee \bar{\alpha}$$

$$\begin{aligned} \mathbf{Pess}(V \text{ is } x^* / V \text{ is } A \text{ is } \alpha\text{-cert}) &= \text{Cert}[V \text{ is } x^* / V \text{ is } A \text{ is } \alpha\text{-cert}] \\ &= \alpha \wedge \text{Min}_{x \neq x^*}[\bar{A}(x)] \\ &= \alpha \wedge (1 - \text{Poss}[V \text{ is not } x^* / V \text{ is } A]) \end{aligned}$$

We now consider an alternative method for representing a certainty quantified statement using the Dempster-Shafer belief structure [22]. Here we represent the statement $V \text{ is } A \text{ is } \alpha\text{-cert}$ by the proposition $V \text{ is } m$ where m is a D-S belief structure with two focal elements, $B_1 = A$ and $B_2 = X$ having $m(B_1) = \alpha$ and $m(B_2) = 1 - \alpha$. In this framework, we use the plausibility and belief measure to obtain our optimistic and pessimistic bounds on the validity of the statement $V \text{ is } x^*$. We recall the plausibility and belief measures are respectively the expected possibility and expected certainty.

$$\begin{aligned} \text{Pl}[V \text{ is } x^* / V \text{ is } m] &= \sum_{i=1}^2 m(B_i) \text{Poss}[V \text{ is } x^* / V \text{ is } B_i] \\ &= \alpha \text{Poss}[V \text{ is } x^* / V \text{ is } A] + \bar{\alpha} \text{Poss}[V \text{ is } x^* / V \text{ is } X] \\ &= \alpha A(x^*) + \bar{\alpha} = 1 - \alpha \bar{A}(x^*) \\ \text{Bel}[V \text{ is } x^* / V \text{ is } m] &= \sum_{i=1}^2 m(B_i) \text{Cert}[V \text{ is } x^* / V \text{ is } B_i] \\ &= \alpha \text{Cert}[V \text{ is } x^* / V \text{ is } A] + \bar{\alpha} \text{Cert}[V \text{ is } x^* / V \text{ is } X] \\ &= \alpha \text{Min}_{x \neq x^*}[\bar{A}(x)] + \bar{\alpha} 0 = \alpha \text{Min}_{x \neq x^*}[\bar{A}(x)] \end{aligned}$$

We observe that the more generally pessimistic measures can be generalized using a t-norm [23]. Thus if T is any t-norm then

$$\mathbf{Pess}[V \text{ is } x^* / V \text{ is } A \text{ is } \alpha\text{-cert}] = T[\alpha, \text{Min}_{x \neq x^*}[\bar{A}(x)]].$$

The optimistic measure can be generalized using a t-conorm [23]. Thus if S is any t-conorm then

$$\mathbf{Opt}[V \text{ is } x^* / V \text{ is } A \text{ is } \alpha\text{-cert}] = S(\bar{\alpha}, A(x^*))$$

We shall not here investigate the issues involved of selecting among these possibilities

6. Multi-Source Information Fusion

We now turn to the issue of aggregation of information from multiple sources, **Multi-Source Information Fusion (M-SIF)**.

If $V \text{ is } A$ and $V \text{ is } B$ are two pieces of information then their conjunction (fusion) is $V \text{ is } D$ where $D = A \cap B$, that is $D(x) = \text{Min}[A(x), B(x)]$. More generally, if $V \text{ is } A_i$, for $i = 1$ to q , are a collection of propositions¹⁴ from multiple sources then their conjunction is $V \text{ is } D$ where $D = \bigcap_{i=1}^q A_i$ here $D(x) = \text{Min}_i[A_i(x)]$. We observe one fundamental feature of this conjunction process. For all x , $D(x) \leq A_i(x)$ that is $D \subseteq A_i$. More generally, if $D = \bigcap_{i=1}^q A_i$ and $E = D \cap A_{q+1}$ then $E \subseteq D$, $E(x) \leq D(x)$ for all x . Thus we see the more information we get the smaller the fuzzy subsets.

In using multiple sources of information, usually, our objective is to increase the amount of information we have about the variable of interest. We desire to increase the specificity. We observe that if $D, \bigcap_{i=1}^q A_i$, is normal then $\text{Sp}(D) \geq \text{Sp}(A_i)$ for any i and we have we have gained information. Thus, here if the information is not conflicting then fusing the information supplied by the multiple sources is a process, which can't decrease the information we have from any of the individual sources. Normally, in this case, D usually is more informative than any of the individual sources

However, if some of the source information are conflicting, this may result in a situation in which D is subnormal, $\text{Max}_x D(x) \leq 1$. In this case, the fusion of the sources may provide us with

¹⁴Here unless otherwise stated we shall assume the A_i are normal, have at least one element with membership grade

a situation in which we are more confused and our informativeness, specificity, has decreased. Generally it is difficult dealing with situations in which we have conflicting source information. One approach to addressing this situation is not to use all the information. That is, we selectively choose which information to use and fuse. This requires adjudicating between the information supplied by the different sources. Often the choice of the appropriate manner of adjudication requires the use of subjective considerations on the part of the person ultimately responsible for fusing the information. In the following, we shall suggest one approach to addressing this problem. We should note that other approaches are possible.

As we shall subsequently see, this process generally involves a tradeoff between selecting a subset of the available information that is not conflicting and yet large enough to provide a credible fusion of the available information. The technique we shall suggest will make use of the concept a credibility measure to help in this process.

Let P_i denote V is A_i , a piece of data about the variable V . We refer to the collection of these as $P = \{P_1, \dots, P_q\}$. We associate with P a measure $\mu: 2^P \rightarrow [0, 1]$ such that for each subset B of P , $\mu(B)$ indicates the credibility of using as our fused knowledge the conjunction of the data in B . We shall call μ the credibility measure. We can associate with μ some basic properties: $\mu(\emptyset) = 0$ and $\mu(P) = 1$. Additionally μ must be monotonic, if $B_1 \subseteq B_2$ then $\mu(B_2) \geq \mu(B_1)$.

Assume B is a subset of P . Let $D_B = \bigcap_{P_i \in B} A_i$, it is the fusion of the knowledge in B .

We observe using the subset B leads to the statement V is D_B . However, any statement obtained by using only the information in B only has a credibility of $\mu(B)$.

In order to determine the quality of the knowledge obtained by using the subset B we must consider two criteria. One criteria is that the knowledge provided by fusing the data in B is informative and the other criteria is that B is credible. The degree of satisfaction to the criteria of informativeness, $\text{Inf}(B)$, can be obtained using the measure of specificity, thus $\text{Inf}(B) = \text{Sp}(D_B)$. We recall $\text{Sp}(D_B) = D_B(x^*) - \text{Ave}_{X - \{x^*\}}(D_B)$ where x^* is any element having maximal membership grade in D_B . The credibility of using the subset B , $\text{Cred}(B)$, can be measured by $\mu(B)$. Since our measure of quality is an *anding* of these criteria we can define the measure of the quality of the result obtained using the subset B as $\text{Qual}(B) = \text{Inf}(B) \text{Cred}(B)$, thus $\text{Qual}(B) = \text{Sp}(D_B) \mu(B)$.

An interesting alternative view of our measure $\text{Qual}(B)$ can be obtained. In the preceding, we indicated that a statement such as V is D_B is $\mu(B)$ -cert as being translated into V is F where $F(x) = D_B(x) \vee (1 - \mu(B))$. We note that $\vee$ is an example of a t-conorm. It is the Max t-conorm, $S(a, b) = \text{Max}(a, b) = a \vee b$. Let us consider the use of another t-conorm. A particularly interesting one is $S(a, b) = a + b - ab$, this is called the bounded sum [24]. If we use this instead of the Max we get $F_B(x) = D_B(x) + (1 - \mu(B)) - (D_B(x)(1 - \mu(B)))$. After a little algebra we get

$$F_B(x) = (1 - \mu(B)) + \mu(B)D_B(x)$$

We note that it is monotonic with respect to $D_B(x)$, if $D_B(x_1) \geq D_B(x_2)$ then $F_B(x_1) \geq F_B(x_2)$.

Consider now the measure of specificity. We first recall $\text{Sp}(D_B) = D_B(x^*) - \text{Ave}_{X-\{x^*\}}(D_B)$

Consider now $\text{Sp}(F_B)$ here since x^* still provides the largest membership in F_B then

$$\begin{aligned} \text{Sp}(F_B) &= F_B(x^*) - \text{Ave}_{X-\{x^*\}}(F_B) \\ \text{Sp}(F_B) &= ((1 - \mu(B)) + \mu(B)D_B(x^*)) - \text{Ave}_{x \neq x^*} [(1 - \mu(B)) + \mu(B)D_B(x)] \end{aligned}$$

The nature of Ave is such that

$$\text{Ave}_{x \neq x^*} [(1 - \mu(B)) + \mu(B)D_B(x)] = (1 - \mu(B)) + \mu(B) \text{Ave}_{X-\{x^*\}}(D_B)$$

Thus here

$$\begin{aligned} \text{Sp}(F_B) &= ((1 - \mu(B)) + \mu(B)D_B(x^*)) - (1 - \mu(B)) - \mu(B) \text{Ave}_{X-\{x^*\}}(D_B) \\ \text{Sp}(F_B) &= \mu(B) (D_B(x^*) - \text{Ave}_{X-\{x^*\}}(D_B)) \\ \text{Sp}(F_B) &= \alpha \text{Sp}(D_B) \end{aligned}$$

Thus using this definition for certainty qualification leads to a very nice result for the relationship between the specificities of D_B and F_B . This can be of great use in finding the best solution to the fusion problem.

We make some observations about this process of multi source fusion. First, observe that if the whole collection of data P is such that D_P is normal then for all B since $D_P \subseteq D_B$ then D_B is also normal. Hence, in this case, $\text{Sp}(D_P) \geq \text{Sp}(D_B)$. Furthermore, since $\mu(P) = 1 \geq \mu(D_B)$ then

$$\text{Qual}(P) = \text{Sp}(D_P) \geq \mu(D_B) \text{Sp}(D_B) \geq \text{Qual}(B).$$

Thus, in the case where the fusion of the data from all the sources doesn't induce any conflict the most informative thing to do is to use fusion of all the data in P .

More generally, we make the following observation.

Observation: If B_1 is a subset of P such that D_{B_1} is normal then for all subsets B_2 of P such that $B_2 \subset B_1$ then $\text{Qual}(B_1) \geq \text{Qual}(B_2)$.

Justification: Since $B_2 \subseteq B_1$ then $D_{B_1} \subseteq D_{B_2}$ and $\mu(B_1) \geq \mu(B_2)$. Since both are normal it follows that $\text{Sp}(D_{B_1}) \geq \text{Sp}(D_{B_2})$ and hence $\text{Qual}(B_1) \geq \text{Qual}(B_2)$.

Definition: We shall call a subset B where D_B is a normal a non-conflicting subset. Furthermore, we call a subset B **maximally non-conflicting** if B is non-conflicting and the addition of any other piece of data to B results in sub-normal fusion.

Observation: Any subset B of data containing a maximally non-conflicting subset can't provide the best fusion.

We shall now consider some examples of the credibility measure. One special class of credibility measure are those we call cardinality based measures. For these measures no distinction is made between the credibility of the different pieces of data, $\mu(B)$ just depends on how many pieces of data are in B , the cardinality of B . We can define a cardinality-based measure using a function $h: [0, 1] \rightarrow [0, 1]$ that satisfies $h(0) = 0$, $h(1) = 1$ and is monotonic, $h(r_1) \geq h(r_2)$ if $r_1 > r_2$. Using h we can define $\mu(B) = h\left(\frac{|B|}{|P|}\right)$. These types of functions are often obtained as a representation of some linguistic quantifier such as *most*, "*at least about half*".

Another class of credibility measures are those that are completely additive. Here we associate with each piece of data P_i , a value $\alpha_i \in [0, 1]$ and assume $\sum_{i=1}^q \alpha_i = 1$. In this case
$$\mu(B) = \sum_{j \in B} \alpha_j.$$

Let G_k , $k = 1$ to g , be a collection of subsets of P that provides a partition of P . One example of credibility measure μ using this is one where $\mu(B) = 1$ if B contains at least piece of data from each of the G_k and $\mu(B) = 0$ otherwise. Closely related to this is a measure in which we associate with each G_k a nonnegative value g_k and define
$$\mu(B) = \sum_{k=1}^g g_k \frac{|B \cap G_k|}{|B|}.$$
 Here we also assume the g_k sum to one.

Another type credibility measure is one that contains a crucial piece of data. We say that P_j is crucial if $\mu(B) = 0$ if $P_j \notin B$.

Another interesting example of credibility measure is the following. Let B_1 be a subset of P . Consider a measure such that $\mu(B) = 0$ if $B_1 \cap B \neq \emptyset$ and $B_1 \subseteq B$. This measure, which we call a balanced measure, requires that if we include any data from B_1 in our fusion we must use all the data in B_1 .

Let us summarize the procedure we suggested for providing a user with quality fusion of the data in the collection P . The first step is to calculate the subset B^* of P with the highest quality conjunction of its component data. That is we find B^* such that $\text{Qual}(B^*) = \text{Max}_{B \subseteq P} [\text{Qual}(B)]$ where $\text{Qual}(B) = \text{Sp}(D_B) \mu(B)$. Having found this subset B^* we indicate to the client that V is D_{B^*} is the result of our multi-source data fusion and that the credibility of this information is $\mu(B^*)$.

7. Multiple Fused Values from Multi-Source Data

In some situations, the presentation of a single fused value may not be sufficient or appropriate. Here we shall suggest a process that will allow us to provide multiple fused values over the data set P .

Our point of departure is again a collection of multi-source data $P = \{P_1, \dots, P_q\}$. Each piece of data P_j being of the form V is A_j where A_j is a fuzzy subset of the domain of V , X . In addition, we have a credibility measure $\mu: 2^P \rightarrow [0, 1]$ where $\mu(B)$ is the degree credibility assigned to a fusion using the data in the subset B of P .

In the preceding, we defined a process for obtaining an optimal subset B_1 and which provided a fused value V is D_{B_1} with credibility $\mu(B_1)$. Here $D_{B_1} = \bigcap_{j, P_j \in B_1} A_j$. This approach finds the subset of data B_1 such that that $\text{Qual}(B_1) = \mu(B_1) \text{Sp}(D_{B_1}) = \text{Max}_{B_1 \subseteq P} [\mu(B) \text{Sp}(D_B)]$. We shall refer to this process as **Qual-Fuse**(P, μ). Thus **Qual-Fuse**(P, μ) returns B_1 which enables the determination of D_{B_1} and $\mu(B_1)$.

In the following, we shall suggest a procedure which allows the for generation of multiple fusions from the pair (P, μ) . For notational convenience in the following we shall find it convenient to denote the fuzzy subsets A_j as A_j^1 , thus our data is still $P = \{P_1, \dots, P_q\}$ where P_j

corresponds to the observation V is A_j^1 . μ is still a credibility measure over P .

The basic algorithm of our procedure is as follows.

1. Initialize our system with P , μ and set $i = 1$.
2. Apply Qual-Fuse(P , μ) this returns B_1 and D_{B_1} and V is D_{B_1} with credibility $\mu(B_1)$.
3. Revise each of the P_j to V is A_j^2 where $A_j^2 = A_j^1 - D_{B_1}$. That is we remove the subset D_{B_1} from the subset A_j^1 . We recall $A_j^1 - D_{B_1} = A_j^1 \setminus D_{B_1}$ and therefore $A_j^2(x) = \min[A_j^1(x), 1 - D_{B_1}(x)]$
4. Set $i = 2$
5. Let $P = [P_1, \dots, P_1]$ with P_j such that V_j is A_j^i
6. Apply Qual-Fuse(P , μ). This returns B_i and the statement V is D_{B_i} with credibility $\mu(B_i)$.
Here $D_{B_i} = \bigcap_{j, P_j \in B_i} A_j^i$
7. Additional fusion desired ? No - stop, Yes - continue
8. Set $i = i + 1$
9. Calculate $A_j^i = A_j^{i-1} - D_{B_{i-1}}$
10. Go to step 5.

The final result of this process is a collection of fused values of the form

V is D_{B_1} with credibility $\mu(B_1)$

V is D_{B_2} with credibility $\mu(B_2)$

V is D_{B_k} with credibility $\mu(B_k)$

The key idea we suggested here is the removal of the already presented fused value from the data remaining to be used to fuse. This is very much in the spirit of the Mountain Clustering method [25, 26]. This removal process tends to result in a situation where the D_{B_j} are disjoint .

An interesting issue, one which we shall not investigate in detail here, is when to stop the process of providing additional fusions. In its simplest form, this can be just based on an input from the user, for example how fused values they want. A more computationally based approach could be one in which we stop when the quality of the next proposed fusion falls below some level.

8. Fusing Probabilistic and Possibilistic Data

An important issue in the field of data fusion concerns itself with the combination of two pieces of information where one is expressed in terms of a fuzzy subset (possibility distribution) and the other is expressed in terms of a probability distribution [27, 28]. Here we shall introduce some ideas related to this problem.

Let G be an attribute that is associated with some class of objects Z . Let X be the domain in which this attribute takes its value. Our interest here is on the determination of the value of the attribute G for some specific entity, z^* , from this class. Thus we are interested in the determination of the value of variable $G(z^*)$. We shall denote this variable as G_* .

Consider a piece of data about G_* such a G_* is A where A is a fuzzy subset of X . Let's look at this data more carefully. First, we see it is directly about the variable of interest. That is, it is a statement about the attribute for the object of interest. Often, this information is a result of some linguistically expressed observation such as "The bomb thrower was young." As noted by Zadeh [29] this statement puts some constraint on the possible values of the variable G_* . It generalizes the idea of having a more crisp statement such as "The age of the bomb thrower was between 18 and 25." It of course reflects some uncertainty with respect to the sources observation. In the situation in which A is assumed normal, this uncertainty can be measured by the cardinality subset A , $\sum_X A(x)$. In the case where we must deal with subnormality a more sophisticated measure such as $Un(A) = 1 - Sp(A) = 1 - (Max(A) - Ave(A))$ should be used. We see if $Max(A) = 1$ then $Un(A) = Ave(A)$ which is essentially $\sum_X A(x)$.

Let us now turn to the situation in which we have additional probabilistic information consisting of a probability distribution P over the space X where $P(x_i)$ is the probability associated with the attribute value x_i . In order to find a basis for fusing these two pieces of information, the possibility distribution A and the probability distribution P , we shall take advantage of a view proposed by Coletti and Scozzafava [30]. In [30] the authors suggested that an element's membership grade in a fuzzy, $A(x_i)$, can be viewed as the conditional probability of A given x_i , $P(A/x_i) = A(x_i)$. Having this allows us to use Bayes' rule to generate the fused information. Let $P(x/A, P)$ indicate the probability of x given are two pieces of knowledge. In particular, $P(x/A, P)$

$$= \frac{P(A/x)}{P(A)} P(x). \text{ Using } P(A/x) = A(x) \text{ we have } P(x/A, P) = \frac{A(x)}{P(A)} P(x). \text{ Furthermore, since } P(A) = \sum_{i=1}^n P(\frac{A}{x_i}) P(x_i) \text{ then } P(A) \text{ can be expressed } \sum_{i=1}^n A(x_i) P(x_i). \text{ Using this, we get } P(x/A, P) = \frac{A(x) P(x)}{\sum_{i=1}^n A(x_i) \cdot P(x_i)}. \text{ At times we shall find it convenient to express this as } P(x/A, P) = \frac{A(x)}{\sum_{i=1}^n A(x_i) \cdot \frac{P(x_i)}{P(x)}}$$

Thus the result of fusing these two pieces of data is a probability distribution with respect to the value of G_* . Using the notation suggested by Zadeh in [29] we can express this as $G_* \text{ is } \mathbf{R}$ where $\mathbf{R}$ indicates a probability distribution on X such that $P(x/A, P)$, as defined above, is the probability that G_* assumes the value x . The fact that this is the case is not surprising since the knowledge in the possibility distribution is actually saying that the value of the variable G_* lies in a set, A . So we are actually finding the probability of x conditioned on the knowledge that G_* lies in a set.

Let us look at this for some special cases to see if it is consistent with our intuition. First, consider the case where $P(x_i) = \frac{1}{n}$. Here, the probability distribution is essentially providing no information. In this case, we have $\frac{P(x_i)}{P(x)} = 1$ for all x_i and hence $P(x/A, P) = \frac{A(x)}{\sum_{i=1}^n A(x_i)}$. Thus

here we obtain $P(x/A, P)$ as simply a normalization of the possible distribution.

Consider now the case in which $A(x_i) = 1$ for all x_i . Here the possible distribution is providing no information. In the case $P(x/A, P) = \frac{P(x)}{\sum_i P(x_i)} = P(x)$. We get back the original probability distribution.

Consider the case where A corresponds to some crisp subset B of X . That is $A(x_i) = 1$ for $x_i \in B$. In this case $P(x/A, P) = \frac{P(x)}{P(x_i)}$. This is the classic case of conditional probability.

One issue that must be addressed is conflicting information. Consider the case where we have $A(x_1) = 1$ and $A(x_j) = 0$ for all other x_j and where $P(x_1) = 0$. In this case, we see that $P(x_i)A(x_i) = 0$ for all x_i and our aggregation leads to a kind of indeterminism. Here, we essentially must decide, do we believe the possibility distribution which says the answer is definitely x_1 or do we believe the probability distribution which says the answer is definitely not x_i .

Another form of conflict can be seen in the following case. Let $A = \{\frac{1}{x_1}, \frac{0.1}{x_2}, \frac{0}{x_3}\}$ and let the probabilistic information be such that $P(x_1) = 0$, $P(x_2) = 0.1$ and $P(x_3) = 0.9$. In this case we obtain $P(x_1/A, P) = 0$, $P(x_2/A, P) = 1$ and $P(x_3/A, P) = 0$. This may be somewhat disturbing. Here while both pieces of information lend little support to x_2 their combination leads to its strong support

In order to address this issue of conflict we must first consider the context in which we obtain probabilistic information. We can envision two situations when we obtain probabilistic information. One of these is in a frequent spirit and the other is of a subjective kind.

One situation where we have probabilistic information is where the probability distribution is a reflection of some observation about the attribute G over the objects in the class Z . Thus here $P(x_j)$ is the probability that an object in Z has value for attribute G equal to x_j . For example, if $x_j = 26$, then $P(26)$ is the probability that "a" bomb thrower is 26. The point we want to emphasize here is that this information is not directly about the entity of interest z^* . It is not information about our variable of interest G_{x^*} , $G(x^*)$. Although it is useful and valuable information, it is not directly about the object of interest. The important observation here is that the information contained in this type of probabilistic information is of a lower priority than the direct information contained in a statement $G_* \text{ is } A$. Thus, here there is a priority ordering with respect to our information and in the face of conflict we want to give preference to the direct information, $G_* \text{ is } A$.

The use of a probabilistic representation can also occur in the case in which the source is providing information directly about the attribute value for the object of interest. Consider the situation where the source has some uncertainty with regard to the actual value of the variable G_* . Here, he uses the probability framework to express his perception of the uncertainty. He is saying that my feeling about the uncertainty associated with the value of G_* is similar to that of a random experiment in which $P(x_i)$ is the probability that $G_* = x_i$. Again, in this situation, the information provided by the source is also less direct than that provided by the observation that $G_* \text{ is } A$.

The overall point we want to make here is that often the information provided using a probabilistic representation has a lesser priority than that provided using the fuzzy representation. This is not to say that fuzzy sets are better than probability but only that the type of information

represented by a probability distribution is less directly relevant

This distinction in the priority of the two different kinds of information allows us to provide a reformulation of the aggregation of these two kinds of information to allow for an intelligent adjudication of conflicts. As a first step in this process, we shall turn to the issue of measuring the conflict or conversely the consistency between a probability distribution and a possibility distribution.

Let $\Pi: X \rightarrow [0, 1]$ be a possibility distribution over the X , thus $\Pi(x_i)$ indicates the possibility of x_i . Here, we shall assume this is normal, there exists some x^* such that $\Pi(x^*) = 1$. Let $P: X \rightarrow [0, 1]$ be a probability distribution over X . $P(x_i)$ indicates the probability of x_i . The probability distribution has the added requirement that $\sum_i P(x_i) = 1$. Let $p^* = \text{Max}_i[P(x_i)]$ it is the maximal probability associated with P . We can observe that $\frac{1}{n} \leq p^* \leq 1$, where n is the cardinality of X . It is well-known that the negation of the Shannon entropy, $\sum_i P(x_i) \ln[P(x_i)]$, provides a measure of information content of a probability distribution. What is worth pointing out is the $\text{Max}_i(P(x_i))$ provides an alternative measure of this information content [31, 32]. While Shannon measure has some properties that make it preferred, especially when we consider multiple distributions, in the case when we are focusing on one probability distribution, $\text{Max}_i(P(x_i))$ provides a simple and acceptable measure of the information content of a probability distribution.

We now introduce a measure called the consistency of Π and P

$$\text{Consist}(\Pi, P) = \text{Max}_i[\Pi(x_i) \wedge \tilde{P}(x_i)] \text{ where } \tilde{P}(x_i) = \frac{P(x_i)}{p^*}$$

We observe that if P is such that if $P(x_i) = \frac{1}{n}$ for all x_i then $p^* = \frac{1}{n}$ and $\tilde{P}(x_i) = 1$ for all x_i . In this case both $\Pi(x^*) = 1$ and $\tilde{P}(x^*) = 1$ and hence $\text{Consist}(\Pi, P) = 1$. Thus, the situation when P has maximal uncertainty it is consistent with any possibility distribution. On the other hand we see that if $P(x_1) = 1$ and $\Pi(x_1) = 0$ then $\text{Consist}(\Pi, P) = 0$ they are in complete conflict. In the case where $X = \{x_1, x_2, x_3\}$ and $\Pi(x_1) = 1$, $\Pi(x_2) = 0.1$ and $\Pi(x_3) = 0$ while $P(x_1) = 0$, $P(x_2) = 0.1$ and $P(x_3) = 0.9$ we get $\tilde{P}(x_1) = 0$, $\tilde{P}(x_2) = 0.11$ and $\tilde{P}(x_3) = 1$ and hence $\text{Consist}(\Pi, P) = 0.1$

We now consider the modification of the procedure for aggregating possibility and probability distributions which uses this measure of consistency to aid in the adjudication of conflicting information.

In the preceding, we defined the aggregation of V is A and the probability distribution P as inducing a probability distribution where $P(x/A, P) = \frac{A(x) P(x)}{\sum_{i=1}^n A(x_i) P(x_i)}$.

We now provide a modification of this to account for conflicts between the input distributions. As we shall see, this will give a priority to the information V is A .

Letting $\alpha = \text{Consist}(P, A)$ we define

$$P(x/A, P) = \frac{A(x)[\alpha P(x) + \bar{\alpha} \frac{1}{n}]}{\sum_{j=1}^n A(x_j)[\alpha P(x_j) + \bar{\alpha} \frac{1}{n}]}$$

Let us see how this works. If the two sources are consistent, $\alpha = 1$, then $P(x/A, P) = \frac{A(x)P(x)}{\sum_{j=1}^n A(x_j)P(x_j)}$ and we get our original formulation. If the two pieces of

information are completely conflicting, $\alpha = 0$ we get $P(x/A, P) = \frac{A(x) \frac{1}{n}}{\sum_{j=1}^n A(x_j) \frac{1}{n}} = \frac{A(x)}{\sum_{j=1}^n A(x_j)}$. Here

we completely discount the information contained in the probability distribution P and simply obtain $P(x/A, P)$ as a normalization of A .

Here, we shall refer to $F(\alpha, P_j) = \alpha P(x_j) + \bar{\alpha} \frac{1}{n}$ as the probability transform and refer to $\lambda(x_j) = F(\alpha, P(x_j))$ as the transformed probabilities. We see that in the face of conflict the transformed probabilities move toward $\frac{1}{n}$

We further observe that if $A(x_j) = 0$, then $P(x/A, P) = 0$.

Example: Assume $X = \{x_1, x_2, x_3\}$, $A = \{\frac{1}{x_1}, \frac{0.1}{x_2}, \frac{0}{x_3}\}$ and $P(x_1) = 0$, $P(x_2) = 0.1$ and $P(x_3) = 0.9$. Here we get $\tilde{P}(x_1) = 0$, $\tilde{P}(x_2) = 0.11$ and $\tilde{P}(x_3) = 1$ and hence $\text{Consist}(\Pi, P) = 0.1$. In this case the transformed probabilities are:

$$\lambda(x_1) = (0.9) \frac{1}{3} = 0.3$$

$$\lambda(x_2) = (0.1)(0.1) + (0.9) \frac{1}{3} = 0.31$$

$$\lambda(x_3) = (0.1)(0.9) + (0.9) \frac{1}{3} = 0.39$$

In this case $\sum_{i=1}^3 A(x_i) \lambda(x_i) = 0.3 + 0.031 = 0.331$ and hence

$$P(x_1/A, P) = \frac{0.3}{0.331} = 0.906, P(x_2/A, P) = \frac{0.031}{0.331} = 0.094 \text{ and } P(x_3/A, P) = \frac{0}{0.331} = 0$$

We must consider one other issue here. We have implicitly assumed that the possibility distribution is normal, $\text{Max}_j(x_j) = 1$. If this is not the case some problems can arise. Since $\text{Consist}(A, P) = \text{Max}_j[A(x_j) \wedge \frac{P(x_j)}{p^*}] \leq \text{Max}_j[A(x_j)]$ our maximal possible consistency goes down. Here the problems of reduced consistency may be an issue related to the internal conflict of the possibility distribution rather than its incompatibility with probability distribution.

It may be interesting to consider a slight modification in the case where we have $\text{Max}_j[A(x_j)] = a^* < 1$. Here, instead of the end result being a probabilistic distribution we end up with a Dempster-Shafer belief structure m . This belief structure has $n + 1$ focal elements $B_j = \{x_j\}$ for $j = 1$ to n and $B_{n+1} = X$. Furthermore for $j = 1$ to n we have $m(B_j) = a^* P(x_j/A, P)$ where the $P(x_j/A, P)$ are calculated as in the preceding, For $B_{n+1} = X$ we have $m(X) = 1 - a^*$. We shall not pursue this but leave it as a suggestion.

9. Alternative Measures of Certainty

Here we consider a more technical issue that may not be of interest to all readers. We want to look a little more deeply at the issue of deciding whether some subset B of X contains the value V given that we know that V is A which is the basis our of definition of the measure $\text{Cert}[V \text{ is } B / V \text{ is } A]$. Here, we shall, unless otherwise stated, assume A is normal.

Our definition for the measure for $\text{Cert}[V \text{ is } B / V \text{ is } A] = \text{Min}_X[\tilde{A}(x) \vee B(x)]$ is an extremely pessimistic measure. As we see in the crisp case as long as there is one element not in B that is possible, in A , it scores a value of zero. We see that if $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ then with $B = \{1\}$ or $B = \{1, 2, \dots, 9\}$ we get the same degree of certainty, zero. Here, there exists no consideration about the cardinalities of how many elements in A are not in B , except that there exists one

We observe that our definition of certainty is the degree of truth of the proposition

"All elements not in B are not possible given A ."

which we expressed as $\text{Min}_X[\tilde{A}(x) \vee B(x)]$. In the case where B is crisp this becomes

$$\text{Cert}[V \text{ is } B / V \text{ is } A] = \text{Min}_X \in \overline{B}[\tilde{A}(x)]$$

Prade and Yager [33] suggested a softening of the measure of certainty with the concept of

expectedness. In [33] they introduced the idea of **expectedness** of V is B given V is A , denoted $\text{Exp}[B/A]$ which they defined as a degree of truth of the proposition

"Most of the elements not in B are not possible given A ."

In the case where B is crisp we can express this as

$$\text{Exp}[V \text{ is } B / V \text{ is } A] = \text{Most}_{x \in \bar{B}}[\tilde{A}(x)]$$

We note that the difference between the two concepts, certainty and expectedness, is the respective uses of the terms all and most. We observe that these two terms are examples of what Zadeh called linguistic quantifiers [34]. Here, we shall suggest a parameterized formulation which leads to a generalization of these types of measures. Let Q indicate a general member of the class of regular monotonic linguistic quantifiers [35]. Using this we introduce the idea of what we shall denote as $Q\text{-Cert}[V \text{ is } B / V \text{ is } A]$ or more succinctly $Q\text{-Cert}[B/A]$. Specifically we define $Q\text{-Cert}[B / A]$ as the truth of the statement

Q of the elements not in B are not possible given A .

First, we note that as suggested by Zadeh linguistic quantifier Q can be expressed as a fuzzy subset $Q : [0, 1] \rightarrow [0, 1]$ where $Q(r)$ indicates the degree to which the proportion r satisfies the concept Q . The fact that Q is a regular monotonic linguistic quantifier requires that Q satisfy the additional three conditions: $Q(0) = 0$, $Q(1) = 1$ and $Q(x) \geq Q(y)$ if $x \geq y$

We note some special cases of Q . The first is Q^* where $Q^*(1) = 1$ and $Q^*(x) = 0$ for all $x \neq 1$. This corresponds to the linguistic quantifier *all*. The second special case is Q^* where $Q^*(0) = 0$ and $Q^*(x) = 1$ for all $x \neq 0$. This corresponds to the quantifier *any*. Another special case is Q_A where $Q_A(x) = x$. It is suggested that this models the linguistic quantifier *some*. Furthermore it is suggested that Q_A corresponds to the quantifier that implicit when no quantifier is explicitly expressed, it is a kind of default quantifier.

We shall formally define the truth of the proposition Q of the element not in B are not possible given A , $Q\text{-Cert}[V \text{ is } B / V \text{ is } A]$, using these importance weighted OWA operator [36]. We first recall this operator.

Let (c_j, d_j) be a two tuple in which c_j is called the importance and d_j is called the argument value. We recall that the OWA aggregation of a collection of these tuples guided by a quantifier Q . $\text{OWA}_Q[(c_1, d_1), (c_2, d_2), \dots, (c_n, d_n)]$, is defined as

$$OWA_Q[(c_1, d_1), (c_2, d_2), \dots, ((c_n, d_n))] = \sum_{j=1}^n w_j d_{\sigma(j)}$$

where $\sigma(j)$ is the index of the j^{th} largest of the d_i and $w_j = Q(\frac{T_j}{T}) - Q(\frac{T_j-1}{T})$ where $T_j = \sum_{i=1}^j c_{\sigma(i)}$ and $T = \sum_{i=1}^n c_i$, the sum of all importances.

In the following we shall express $Q\text{-Cert}[V \text{ is } B / V \text{ is } A]$ using this operation. For notational convenience we assume the domain of V , $X = \{x_1, \dots, x_n\}$ and $B(x_i) = b_i$ and $A(x_i) = a_i$. Using this notation $Q\text{-Cert}[V \text{ is } B / V \text{ is } A] = OWA_Q[(\bar{b}_i, \bar{a}_i)]$ where $\bar{b}_i = 1 - b_i$ and $\bar{a}_i = 1 - a_i$ thus

$$Q\text{-Cert}[V \text{ is } B / V \text{ is } A] = \sum_{j=1}^n w_j \bar{a}_{\sigma(j)}$$

where $\sigma(j)$ is the index of the j^{th} largest of the $\bar{a}_i$ and $w_j = Q(\frac{T_j}{T}) - Q(\frac{T_j-1}{T})$ where $T_j = \sum_{i=?}^j \bar{b}_{\sigma(i)}$ and $T = \sum_{k=?}^n \bar{b}_k$.

Let us consider the environment when B is crisp. Here $b_i = 1$ if $x_i \in B$ and $b_i = 0$ if $x_i \notin B$. Thus $\bar{b}_i = 0$ if $x_i \in B$ and $\bar{b}_i = 1$ if $x_i \notin B$. In this situation, we also observe if $b_{\sigma(j)} \in B$ then $\bar{b}_{\sigma(j)} = 0$ and since $T_j = \sum_{i=1}^j \bar{b}_{\sigma(i)}$ then in this case, $b_{\sigma(j)} \in B$, $T_j = T_j - 1$ and hence $w_j = 0$.

Thus we see that all terms that are in B have OWA weights equal to zero. Furthermore, for those elements not in B we have $\bar{b}_{\sigma(j)} = 1$ and T_j is the number of elements up to including the j^{th} largest $\bar{a}_i$ that are not in B . Effectively for those $x_{\sigma(j)} \notin B$ we have $T_j = 1 + T_j - 1$ and for those $x_{\sigma(j)} \in B$, we have $T_j = T_j - 1$

Thus in this case where B is crisp situation we can just consider those elements not lying in B . We shall let $\bar{n} = |\bar{B}|$ and let $\sigma_{\bar{B}}(j)$ be the index of the element having the j^{th} largest value for $\bar{a}_i$ of those lying in $\bar{B}$. Then, in this case, B is crisp, we have $Q\text{-Cert}[B/A] = \sum_{j=1}^{\bar{n}} w_j \bar{a}_{\sigma_{\bar{B}}(j)}$ where $w_j = Q(\frac{j}{\bar{n}}) - Q(\frac{j-1}{\bar{n}})$.

Let us consider the resulting formulations for some different examples of Q . If Q is Q_* then $w_{\bar{n}} = 1$ and $w_j = 0$ for all other j and $Q_*\text{-Cert}[B / A] = \min_{x_i \notin B} [\bar{A}(x_i)]$. This was our original definition of $\text{Cert}[B / A]$. If we select $Q = Q^*$, then we get $Q^*\text{Cert}[B / A] = \max_{x_i \notin B} [\bar{A}(x_i)]$. This is what Dubois and Prade called the un-guaranteed necessity. Another special

case is where $Q(x) = x$. In this case we have $w_j = \frac{1}{n}$ and $Q\text{-Cert}[V \text{ is } B / V \text{ is } A] = \frac{1}{n} \sum_{j \notin B} \bar{A}(x_j)$, it is the average of $\bar{A}(x_j)$ for those x_j not in B .

In the following, we shall look a useful family of $Q\text{-Cert}[V \text{ is } B / V \text{ is } A]$ based on a class of quantifiers parameterized by a single scalar value λ . Consider the function Q_λ shown in figure 9.

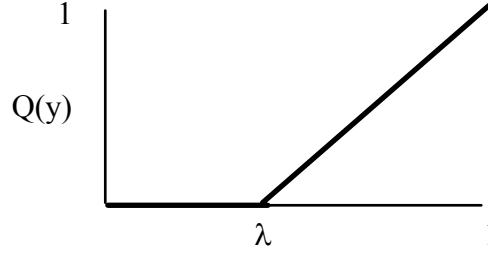


Figure 9. Quantifiers parameterized by β

We shall denote λ as the strength of necessity. We easily see that when $\lambda = 1$ we get the strongest measure $Q_\lambda\text{-Cert}[B / A] = \text{Min}_{x_i \notin B} [\bar{A}(x_i)]$. When $\lambda = 0$ then we get $Q_\lambda\text{-Cert}[B / A] = \frac{1}{n} \sum_{j \notin B} \bar{A}(x_j)$. Generally we observe as λ moves from 0 to 1, the value of $Q_\lambda\text{-Cert}[B / A]$

decreases. If we impose the additional assumption that A is also crisp we add in developing a deeper intuitive understanding of the class of formulations for uncertainty we have introduced. In the case when $\lambda = 1$ to be certain of the truth of statement $V \text{ is } B$ we require if an outcome is possible, in A , it is also in B . In the case where $\lambda = 0$ we have

$$Q_\lambda\text{-Cert}[B / A] = \frac{1}{n} \sum_{j \notin B} \bar{A}(x_j) = 1 - \frac{1}{n} \sum_{j \notin B} A(x_j).$$

Here we take average possibility of the elements not in B and subtract that from one.

In this section we have described a family of definitions for the idea of the Certainty of $V \text{ is } B$ given $V \text{ is } A$ based on the parameter Q which we denoted $Q\text{-Cert}[B / A]$. By appropriately choosing the quantifier Q we can model the formulation we want to use for our concept of certainty. One important way in which these definitions for certainty differ is with respect to their strictness. Recalling that $Q\text{-Cert}[B / A]$ is defined as the truth of the statement "Q of the element not in B are not possible given A " we that the the larger Q the stricter. In order to more formally capture this idea of strictness we can associate with any quantifier Q a value called its attitudinal

character defined as

$$A-C(Q) = \int_0^1 Q(y) dy$$

It can be shown that $A-C(Q) \in [0, 1]$. Also we can show that for $Q = Q^*$ we get $A-C(Q^*) = 0$, for $Q = Q^*$ we get $A-C(Q^*) = 1$ and for $Q(x) = x$ we get $A-C(Q) = 0.5$. Thus the smaller the value of $A-C$ the stricter our concept of certainty.

10. References Part III

- [1]. Risen, J., "Ex-inspector says C.I.A. missed disarray in Iraqi arms program," New York Times, January 27, 1, 2004.
- [2]. Zadeh, L. A., "Fuzzy sets," Information and Control 8, 338-353, 1965.
- [3]. Zadeh, L. A., "Fuzzy sets and information granularity," in Advances in Fuzzy Set Theory and Applications, edited by Gupta, M. M., Ragade, R. K. and Yager, R. R., North-Holland: Amsterdam, 3-18, 1979.
- [4]. Zadeh, L. A., "Fuzzy logic = computing with words," IEEE Transactions on Fuzzy Systems 4, 103-111, 1996.
- [5]. Zadeh, L. A. and Kacprzyk, J., Computing with Words in Information/Intelligent Systems 1, Physica-Verlag: Heidelberg, 1999.
- [6]. Zadeh, L. A., "A theory of approximate reasoning," in Machine Intelligence, Vol. 9, edited by Hayes, J., Michie, D. and Mikulich, L. I., Halstead Press: New York, 149-194, 1979.
- [7]. Dubois, D. and Prade, H., "Fuzzy sets in approximate reasoning Part I: Inference with possibility distributions," Fuzzy Sets and Systems 40, 143-202, 1991.
- [8]. Ruspini, E., "Approximate reasoning: past, present, future," Information Sciences 57-58, 297-317, 1991.
- [9]. Yager, R. R., "On the theory of approximate reasoning," Controle Automacao 4, 116-125, 1994.
- [10]. Yager, R. R., "Approximate reasoning as a basis for computing with words," in Computing with Words in Information/Intelligent Systems 1, edited by Zadeh, L. A. and Kacprzyk, J., Springer-Verlag: Heidelberg, 50-77, 1999.
- [11]. Yager, R. R., "The entailment principle for Dempster-Shafer granules," Int. J. of Intelligent Systems 1, 247-262, 1986.
- [12]. Yager, R. R., "Entropy and specificity in a mathematical theory of evidence," Int. J. of General Systems 9, 249-260, 1983.
- [13]. Yager, R. R., "Specificity measures of possibility distributions," Proceedings of the Tenth NAFIPS Meeting, U. of Missouri, Columbia, MO, 240-241, 1991.
- [14]. Yager, R. R., "On measures of specificity," in Computational Intelligence: Soft Computing

and Fuzzy-Neuro Integration with Applications, edited by Kaynak, O., Zadeh, L. A., Turksen, B. and Rudas, I. J., Springer-Verlag: Berlin, 94-113, 1998.

[15]. Dubois, D. and Prade, H., "Necessity measures and the resolution principle," IEEE Transactions on Systems, Man and Cybernetics 17, 474-478, 1987.

[16]. Dubois, D. and Prade, H., Possibility Theory : An Approach to Computerized Processing of Uncertainty, Plenum Press: New York, 1988.

[17]. Yager, R. R., "Decision making under various types of uncertainty," in Decision Making: Recent Developments and Worldwide Applications, edited by Zanakis, S. H., Doukidis, G. and Zopounidis, C., Kluwer Academic Publishers: Boston, 233-250, 2000.

[18]. Yager, R. R., "Fuzzy modeling for intelligent decision making under uncertainty," IEEE Transactions on Systems, Man and Cybernetics Part B: Cybernetics 30, 60-70, 2000.

[19]. Yager, R. R., "On the valuation of alternatives for decision making under uncertainty," International Journal of Intelligent Systems 17, 687-707, 2002.

[20]. Luce, R. D. and Raiffa, H., Games and Decisions: Introduction and Critical Survey, John Wiley & Sons: New York, 1967.

[21]. Yager, R. R., "Approximate reasoning as a basis for rule based expert systems," IEEE Trans. on Systems, Man and Cybernetics 14, 636-643, 1984.

[22]. Shafer, G., A Mathematical Theory of Evidence, Princeton University Press: Princeton, N.J., 1976.

[23]. Klement, E. P., Mesiar, R. and Pap, E., Triangular Norms, Kluwer Academic Publishers: Dordrecht, 2000.

[24]. Alsina, C., Trillas, E. and Valverde, L., "On some logical connectives for fuzzy set theory," J. Math Anal. & Appl. 93, 15-26, 1983.

[25]. Yager, R. R. and Filev, D. P., "Approximate clustering via the mountain method," IEEE Transactions on Systems, Man and Cybernetics 24, 1279-1284, 1994.

[26]. Rickard, J. T., Yager, R. R. and Miller, W., "Mountain clustering on non-uniform grids using p-trees," Technical Report# MII-2408 Machine Intelligence Institute, Iona College, New Rochelle, NY, 2004.

[27]. Yager, R. R. and Petry, F., "Fusing object information and peer information," International

Journal of Intelligent Systems 16, 1419-1444, 2001.

[28]. Yager, R. R. and Petry, F. E., "Constructing intelligent fusion procedures using fuzzy systems modeling," Proceedings of the Fourth International Conference on Information Fusion (ISIF), Montreal, Vol II, ThB3: 23-29, 2001.

[29]. Zadeh, L. A., "Toward a perception-based theory of probabilistic reasoning with imprecise probabilities," Journal of Statistical Planning and Inference 105, 233-264, 2002.

[30]. Coletti, G. and Scozzafava, R., Probabilistic Logic in a Coherent Setting, Kluwer Academic Publishers: Dordrecht, 2002.

[31]. Aczel, J. and Daroczy, Z., On Measures of Information and their Characterizations, Academic: New York, 1975.

[32]. Yager, R. R., "Measures of entropy and fuzziness related to aggregation operators," Information Sciences 82, 147-166, 1995.

[33]. Prade, H. and Yager, R. R., "Estimations of expectedness and potential surprise in possibility theory," International Journal of Uncertainty, Fuzziness and knowledge-based systems, 417-428, 1994.

[34]. Zadeh, L. A., "A computational approach to fuzzy quantifiers in natural languages," Computing and Mathematics with Applications 9, 149-184, 1983.

[35]. Yager, R. R., "Quantifiers in the formulation of multiple objective decision functions," Information Sciences 31, 107-139, 1983.

[36]. Yager, R. R., "On the inclusion of importances in OWA aggregations," in The Ordered Weighted Averaging Operators: Theory and Applications, edited by Yager, R. R. and Kacprzyk, J., Kluwer Academic Publishers: Norwell, MA, 41-59, 1997.