Representation, Uncertain Imprecision, and Dependency

Lawrence J. Mazlack
Computer Science
University of Cincinnati
Cincinnati, Ohio 45220
mazlack@uc.edu

Abstract

This paper is an initial exploration of representation, uncertain precision, and dependency as a connected concern that is not easily decomposed. Often for reasons of computational simplicity, the assumption is made that attributes are independent of each other. Similarly, whether dependency exists is classically expressed in terms of first order logic The actuality is that often the assumptions produce results that are not good representations of reality. An integrated representation of multiple, related attributes is difficult. Usually, different attributes have different ranges and linearity. Sometimes, normalization is first step in meaningfully representing different kinds of data; or, as a first step to the combination different kinds of data. Most values are not crisply known. A method of dependency representation is needed. The issue of how to represent both dependency and uncertainty needs to be resolved. In a very real sense, dependency representations and their imprecision both enriches and constrains how we approach the solutions to our problems.[1]

1. Introduction

This paper is an initial exploration of representation, uncertain precision, and dependency as a connected concern that is not easily decomposed. The goal of this paper is to consider how they may be related. How to accomplish a meaningful fusion is not suggested.

Often for reasons of computational simplicity, the assumption is made the effects of one attribute are not affected by the effects of another attribute. Similarly, dependency existence is often expressed as a binary True/False and is described in terms of first order logic.

The actuality is that these assumptions are an oversimplification. For example, say that an artisan has the intention to carve a wooden bowl of a particular shape. Whether a bowl of the desired shape is eventually created depends on: the strength of the artisan's original intention, the causal complex of how the artisan interacts with the wood, and whether the wood itself can be transformed into a bowl. All of these items have their own causal possibilities that affect the final outcome. None of these possibilities may be precisely described.

The issue of how to represent both dependency and uncertainty needs to be resolved; both separately and jointly. In a very real sense, dependency representations, including their imprecision description, both enriches and constrains how we approach the solutions to a problem.

One of the key concepts in probability theory is the notion of independence. Under independence, we can decompose a complex problem into simpler components and build a model from the components [Castillo, 1997] [Pearl, 1988]. The term *stochastic independence* refers to the standard concept of independence in probability theory [Couso, 1999]. Stochastic independence assumes a unique underlying probability distribution.

It is unlikely that stochastic probability can be applied to the concerns of this paper. *Imprecise probabilities* [Walley, 1991] may be a better choice. Among other types of belief, imprecise probabilities include fuzzy sets. There is disagreement about how to define independence when using imprecise probabilities [Campos, 1995].

The concepts of *uncertainty* and *information* are closely connected [Klir, 1999]. Uncertainty involved in any problem solving situation is a result of some information deficiency or random factor. Information about the situation can be obtained by any action that reduces the uncertainty. Information can then be described in terms of uncertainty reduction.

If we want to be able to deal with uncertainty, we need to be able to measure the uncertainty in question. This is not a straight forward process as there are many approaches to measuring both information and uncertainty. Measuring uncertainty implies the ability to assign a number or a value from some ordinal scale to a given model. For some work on ordinal measures of uncertainty see [Yao, 1995].

[1] Parts of this work were accomplished while the author was visiting at BISC, Computer Science Division, EECS Department, University of California, Berkeley

2. Normalization and dependency

Different attributes may have different ranges. One attribute may reach a different maximum than the other. One may go from minimum to maximum over a different scale, or if on the same scale, over a different range on the scale. Sometimes, normalization is a first step in meaningfully representing different kinds of data; or, as a first step to the combination different kinds of data.

Many methods assume that data relationships are independent from each other. However, many data items are, in fact, dependent. How to effectively deal with dependent data is a question that must be addressed by anyone seeking to fully explore data dependencies.

2.1 Normalization

To have an acceptable measure of *total uncertainty* normalization is one of the needed elements [Harmanec, 1999].

Normalization can be used to place data on the same scale. For example, dogs and people live a different period of years and reach physical maturity at different ages. Linear normalization is commonly used to normalize the age range; e.g., saying one "dog year" is worth "seven human years." This is an over simplification as the relative aging between dogs and humans is not linear. (A better piecewise linear approximation of dog:man life span is 1:7 for the first 2 years of a dog's life and 1:5 thereafter.) Nor is the range uniform between species uniform; e.g., small dogs such as Miniature Poodles tend to live longer than larger dogs such as Siberian Huskies.

Similarly, a plot of dog weight (or human weight) against age (or available food) would not come up with the same maximum weights. Miniature Poodles simply never get as heavy as Siberians, or Siberians as people.

If we are comparing one pair of attributes, we can normalize both to the same maximums (say 1.0) on all vectors. However, how to do this for more than two attributes at the same time is a difficult question for many domains, including bioinformatics.

For example, if Corn independently grows as shown in *Figure 1* for varying quantities of water, fertilizer, and sun; water and fertilizer growth could be normalized as shown in *Figures 2* and *3*. (Notice that the range has been standardized to the same length where growth ceases.) If included, hours light/day could be normalized to height, but as additional light does not cause plant death, how to satisfactorily normalize on the other axis is unclear. *Figure 2* retains the relative difference in heights while *Figure 3* normalizes to a value of 1.0 for each maximum height. (Note: The corn growth graphs are only hypothetical and are here for conceptual illustration. They do not correspond to known experimental results.)

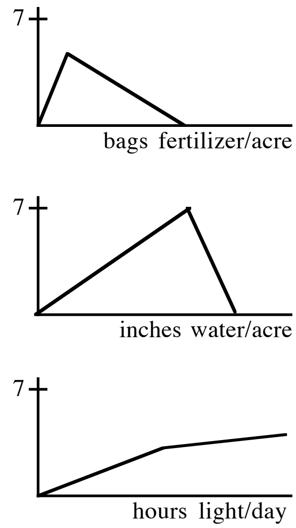

Figure 1. Corn growth dependent on independent factors.

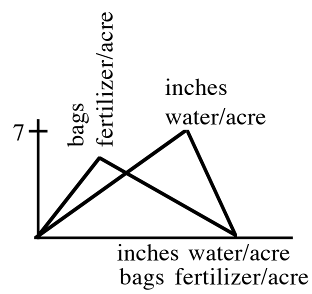

Figure 2. Growth normalized to plant death.

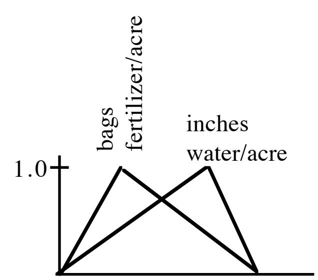

Figure 3. Growth normalized to maximum plant height and to plant death.

Simple linear normalization may work well enough when the values being normalized are linear and not of a high order of dimensionality. However, how to normalize non-linear attributes is unclear. The effects of higher dimensionality may be problematic.

2.2 Non-Independence

Many methods assume that data relationships are independent from each other. How to effectively deal with dependent data is an open question that must be addressed by anyone seeking to fully explore data dependencies.

There are several things to be worried about when considering dependence. One of them is measuring the degree of dependence. Another is the certainty of the dependence. Still another is whether the dependence changes as the attributes change; there is no reason to assume that dependency effects would be constant.

For an example of variable dependence, let us say that initially corn grows higher as it gets small amounts of both fertilizer and water. Then, as the quantities increase, the dependency is lessened. A positive dependency is shown in *Figure 4*. The result is shown in *Figure 5*.

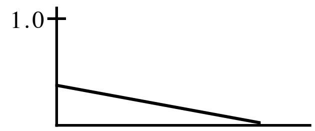

Figure 4. Positive growth dependency as a function of normalized water and fertilizer.

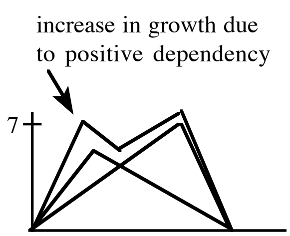

Figure 5. Effect of Figure 4 applied to Figure 2.

There is no reason to assume that the effects of dependency would be either piece-wise linear or constant. Dependency distributions could be multi-modal and asymmetric. A complicating factor is uncertainty: At what point does the uncertainty become greater than the effects of dependency? How to handle this is an unclear. In part, it depends on whether quantitative or qualitative descriptions are used.

3. Certainty confidence

A description of the certainty of a value is a common part of statistical analysis when we know the distribution of the data. This is essentially a degree of confidence description of the certainty of a value (or set of values).

Some form of manipulatable confidence metric is required. We need to understand the quality limits of our answers. We need to control the imprecisenss of our data and the functional imprecisness lest it overwhelms our derived answers. Knowing the critical parts of a solution that adversely affect our certainty can direct our focus to improving our confidence in focused areas. At the same time, we also need to know when our answers have been overwhelmed by imprecision and are thus not usable.

3.1 Metrics

Two classic, well-established classes of metrics are used to measure uncertainty are: *Hartley* and *Shannon metrics*. The Hartley measure applies to uncertainty formalized in terms of crisp possibility theory. The Shannon measure applies to uncertainty formalized in terms of precise probability theory. A basic discussion of evidence theory along this line can be found in Klir [1998].

Using the classical uncertainty metrics as a basis, a search for their counterparts in evidence theory and possibility theory have been pursued by numerous researches. So far, only Hartley-like metrics have been discouvered and fully justified (intuitively or axiomatically). Efforts to construct Shannon-like counterparts so far have failed to meet various required conditions, such as sub-additivity.

It is well understood that situations described in terms of evidence theory involve, in general, two types of uncertainty. However, only *nonspecificity,* is well understood; It is a Hartley-like metric. The second type of uncertainty in evidence theory, the one obtained by generalizing the uncertainty quantified in probability theory by the Shannon measure is still not well understood.

3.2 Data dispersion

Many functionally expressed data values are approximations. Typically, we can think of a particular number, say the age of a student as a given value bounded by confidence limits that may or may not be symmetric. (There are many isomorphic terms for "confidence.")

Most work on confidence descriptions has been done for unimodal, symmetric distributions. If faced with empirical data, we usually have a cloud of data; a single line representation should be understood as something drawn through it for summarization purposes. A common certainty measure is standard deviation; e.g., an average has a value of *A* with a standard deviation of *D*. Standard deviations can be used as an expression of confidence in a value.

The need is to pick a single point in the universe of discourse as a reasonable representation of the set. When we are concerned with non-stochastic belief functions, we have to find a way to deal with the *non-specifity* [Higashi, 1983] or *specifity* [Yager, 1983] of data. It was later generalized by Dubois [1995].

Most attempts at measuring uncertainty in belief functions divided the problem into two parts: non-specifity and conflict. Non-specifity is the inability to distinguish which of several possible alternatives is the true one in a particular situation. Conflict is present whenever is inconsistency or disagreement the evidence or information.

We need to worry about certainty because the uncertainty of the data can eventually overwhelm the information available. Recognizing when uncertainty will overwhelm a result's quality under the combination of multiple representations is an open question.

3.3 Soft computing

The guiding principal of the implementation of soft computing is the exploitation of the tolerance for imprecision, uncertainty, and partial truth to achieve tractability, robustness, and low solution cost [Tagarev, 1999]. Zadeh [1994] considers the main components of soft computing to be fuzzy logic, rough sets, neural network theory, and probabilistic reasoning; the later part including parts of evolutionary computation, parts of learning theory, belief networks and chaos theory. Of these, the first components are primarily concerned with imprecision of data and information; the last, with uncertainty.

Fuzzy logic systems are comprised of rules. Quite often, the knowledge that is used to construct these rules is uncertain. This leads to rules whose antecedents and/or consequents are uncertain. This produces uncertain antecedent or consequent membership functions. Type-1 fuzzy sets are unable to directly handle such rule uncertainties. Type-2 fuzzy sets can be useful in circumstances where it is difficult to determine an exact membership function for a fuzzy set.

The defuzzified output of a type-1 fuzzy logic system may be viewed as analogous to computing the mean of a probability density function (they are not the same). Just as a variance provides a measure of dispersion about the mean, and is almost always used to capture more about probabilistic uncertainty. A type-2 fuzzy logic system provides a measure of dispersion. A measure of dispersion is needed in system design that include linguistic uncertainties that translate into rule uncertainties.

Regardless of the precise mechanism or model by which data dispersions are handled, practical systems require a methodology to handle them; perhaps soft computing, because of its history of practical responses to computational difficulties.

3.4 Representation

Representing data dispersion is an interesting issue. When the data is first captured, it can be shown as a data cloud. When there are only relatively few items, a collection of points is sufficient. When there are many points, either a grey scale or coloured density image can be used. There are many pieces of software that can be used for data visualization. However, data visualization is not a great help in using dispersed data in ensuing computation.

In order to computationally manipulate dispersed data, it is best to capture the data in a functional representation. For example, simplistically, we might represent a population's *weight* versus *age* relationship an average with a standard deviation.

If the density corresponds to a uni-modal function, its description can be readily captured in a polynomial function. How to use uni-modal dispersion functions is well explored. However, handling non-unimodal data dispersions is a less developed area. In other words, expressing and manipulating data confidence measures may be difficult if the data does not correspond to a uniform distribution.

When imprecise data are combined, intuitively we can see that imprecise results will also be obtained. At some point, the derived imprecisness overwhelms the result's quality. How to functionally represent and use a knowledge of imprecision is still to be resolved.

4. Dependence

Dependence is a second order effect. It can have a non-uniform distribution. Dependence distributions are unlikely to be crisp.

There are a number of different independence definitions for different kinds of data; and, different kinds of imprecision representations. It can be argued that they are needed for different kinds of problems [Couso, 1999]. Comparisons of different kinds have been given by Campos [1995], and Walley [1991].

Causality is a strong form of dependency and is of great interest. Recognizing causal relationships is more difficult than recognizing dependence. Whether causal dependence can be expressed in the same way as simple dependence is unclear. How to apply imprecision to dependent values in the light of non-uniform, non-precise dependence distributions is a difficult, open question.

4.1 Classical dependence

4.1.1 Statistical independence

Statistical dependence is often confused with causality. Such reasoning is not correct. Two events E_1, E_2 may be statistical dependent because both have a common cause E_0. But this does not mean that E_1 is the cause of E_2.

For example, lack of rain (E_0) may cause my rose bush to die (E_1) as well as that of my neighbor (E_2). This does not mean that the dying of my rose has caused the dying of my neighbor's rose, or conversely. However, the two events E_1, E_2 are statistically dependent.

The general definition of statistical dependence is:

Let *A, B* be two random variables that can take on values in the domains $\{a_1,a_2,...,a_i\}$ and $\{b_1,b_2,...,b_j\}$ respectively. Then a_i is said to be statistically independent of *B* iff

$$\text{prob}(a_i|b_j) = \text{prob}(a_i) \text{ for all } b_j \text{ and for all } a_i.$$

The formula

$$\text{prob}(a_i|b_j) = \text{prob}(a_i)\ \text{prob}(b_j)$$

describes the joint probability of a_i AND b_j when *A* and *B* are independent random variables. Then follows the law of compound probabilities

$$\text{prob}(a_i,b_j) = \text{prob}(a_i)\ \text{prob}(b_j|a_i)$$

In the absence of causality, this is a symmetric measure. Namely,

$$\text{prob}(a_i,b_j) = \text{prob}(b_j,a_i)$$

4.1.2 Causality vs statistical dependence*:*

A causality relationship between two events E_1 and E_2 will always give rise to a certain degree of statistical dependence between them. The converse is not true. A statistical dependence between two events may, but need not, indicate a causality relationship between them. We can tell if there is a positive correlation if

$$\text{prob}(a_i,b_j) > \text{prob}(a_i)\ \text{prob}(b_j)$$

However, all this tells us that it is an interesting relationship. It does not tell us if there is a causal relationship.

Following this reasoning, it is reasonable to suggest that association rules developed as the result of link analysis might be considered causal; if only because of a time sequence is involved. In some applications, such as communication fault analysis Hatonen [1996], causality is assumed. In other potential applications, such as market basket analysis[1], the strength of time sequence causality is less apparent. For example, if someone buys milk on day_1 and dish soap on day_2, is there a causal relationship? Perhaps, some strength of implication function could be developed.

Some forms of experimental marketing could be used. However, how widely it might be applied is unclear. For example, a food store could carry milk ($E_{1,m=1}$) one month and not carry dish soap. The second month the store could carry dish soap ($E_{2,m=2}$) and not milk. On the third month, it could carry both milk and dish soap ($E_{1,m=3}$) ($E_{2,m=3}$). That would determine both the independent and joint probabilities (setting aside seasonality issues). Then, if

$$\text{prob}(E_{1,m=3})\ \text{prob}(E_{2,m=3}) > \text{prob}(E_{1,m=1})\ \text{prob}(E_{2,m=2})$$

there would be some evidence of a causal relationship.

[1]Time sequence link analysis can be applied to market basket analysis when the customers can be recognized. For example through the use of customer "loyalty" cards in supermarkets or "cookies" in e-commerce.

4.2 Representation

In the most focused case, recognizing and describing dependencies is similar to recognizing causal relationships. How to do this is often accomplished by some form of a directed graph. Weights indicating causal or dependency strength can be attached to an edge.

Directed graph representations are intuitively appealing because of their descriptive power. The down side is that directed graph dependency representations are computationally expensive. How to integrate them with confidence functions is not well understood.

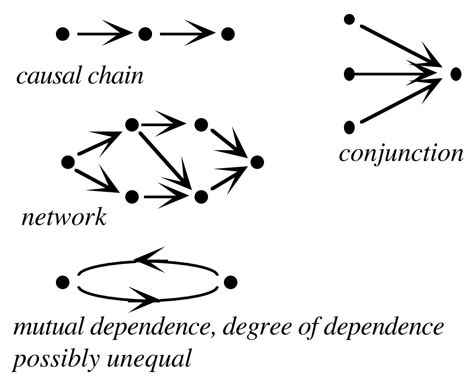

Figure 6. Directed graphs indicating dependency or causality.

Some authors have suggested that *sometimes* it is possible to recognize causal relations through developing acyclic graphs [Scheines, 1994]. Pearl [1991] and Spirtes [1993] make the claim that it is possible to infer causal relationships between two variables from associations found in observational (non-experimental) data without substantial domain knowledge. Spirtes claim that directed acyclic graphs can be used if (a) the sample size is large and (b) the distribution of random values is faithful to the causal graph. Robins [1999] argues that their argument is incorrect. Without going deeply into their debate, it would appear that part of the problem is with crispness. Possibly fuzzy functions would help. Only experimentation can tell. Lastly, Scheines [1994] only claims that in *some* situations will it be possible to determine causality.

Developing directed acyclic graphs is computationally expensive. The amount of work increases geometrically with the number of attributes. Full Bayesian networks, which are based on directed acyclic graphs are also similarly complex in that their complexity is at least exponential to the number of attributes [Heckerman, 1997]. To reduce the amount of work, the data can be sampled and sets of directed acyclic graphs constructed for each sample. The task then becomes finding the "best" graph set [Spirtes, 2000]. Work is progressing in this area; but has at yet to result in a robust solution.

It would seem that it would be reasonable to experimentally test to see if acyclic graphs can be usefully used to test causality or dependency in large data sets. Perhaps, possible experiments might also include different levels of granularity. Possibly, granularity might be increased by either concept hierarchies or clustering.

5. Epilogue

This paper was an initial exploration of representation, uncertain precision, and dependency. It considered the interrelationship between these issues. This paper offered speculations, not solutions.

Often for reasons of computational simplicity, the assumption is made that attributes are independent of each other. Similarly, whether dependency exists is often expressed as a binary true or false and described in terms of first order logic The actuality is that often none of the above are good representations of reality. An initial consideration of how to describe causal strengths and how to handle the inherent dependency description imprecision were among the goals of this paper.

Most values are not crisply known. Often, how to functionally combine them is not crisp. A method of non-crisp dependency representation is needed. When imprecise data are combined, intuitively we can see that imprecise results will also be obtained. At some point, the derived imprecisness can overwhelm the result's quality.

Usually, different relationships have different ranges. One attribute may reach a different maximum than the other. One relationship may go from minimum to maximum over a different scale, or if on the same scale, over a different range. Sometimes, normalization is first step in meaningfully representing different kinds of data; or, as a first step to the combination different kinds of data.

The issue of how to represent related dependency and uncertainty needs to be resolved. In a very real sense, dependency representations and their imprecision both enriches and constrains how we approach the solutions to our problems.

Bibliography

L. de Campos, S. Moral [1995] "Independence Concepts For Convex Sets of Probabilities," *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence,* 108-115

E. Castillo, J.M. Gutierrex, A.S. Hadi [1997] **Expert Systems And Probabilistic Network Models**, Springer-Verlag, New York

I. Couso, S. Moral, P. Walley [1999] "Examples Of Independence For Imprecise Probabilities," *1st International Symposium On Imprecise Probabilities And Their Applications*, Ghent, Belgium (taken from electronic proceedings)

D. Dubois, H. Prade [1985] A Note On Measures Of Specifity For Fuzzy Sets," *International Journal Of General Systems,* 10(4) 279-283

D. Harmanec [1999] "Measures of Uncertainty And Information," *Technical Note,* http://ensmain.rug.ac.be/~ipp

D. Heckerman, C. Meek, G. Cooper [1997] A Bayesian Approach To Causal Discovery, Microsoft Technical Report MSR-TR-97-05

M. Higashi [1983] "Measures Of Uncertainty And Information Based On Possibility Distributions," *International Journal Of General Systems,* 9(1), 43-58

G. Klir, M. Wierman [1998] **Uncertainty-Based Information: Elements of Generalized Information Theory**, Physica-Verlag/Springer-Verlag, Heidelberg and New York

G. Klir [1999] "Uncertainty And Information Measures For Imprecise Probabilities: An Overview," *ISPITA,* Ghent, Belgium, July, electronic proceedings: http://decsai.ugr.es/~smc/isipta99/proc/050.html

J. Pearl [1988] **Probabilistic Reasoning With Intelligent Systems**, Morgan Kaufmann, San Mateo

J. Pearl, T. Verma [1991] "A Theory Of Inferred Causation," *Principles Of Knowledge Representation And Reasoning: Proceedings Of The Second International Conference,* J. Allen, R. Fikes, E. Sandewall, Morgan Kaufmann, 441-452

R. Robins, L. Wasserman [1999], *"On The Impossibility Of Inferring Causation From Association Without Background Knowledge,"* in **Computation, Causation, and Discovery** (eds.) C. Glymour, G.F. Cooper, AAAI Press/MIT Press, Menlo Park, 305-321

D. Heckerman, C. Meek, G. Cooper [1997] A Bayesian Approach To Causal Discovery, Microsoft Technical Report MSR-TR-97-05

T. Tagarev, P. Ivanova [1999] "Computational Intelligence In Multi-Source Data And Information Fusion," *Information & Security,* 2, 1311-1493

R. Yager [1983] "Entropy And Specificity In A Mathematical Theory Of Evidence," International Journal Of General Systems, 9(4), 249-260

Y. Yao, S. Wong, L. Wang [1995] "A Non-Numeric Approach To Uncertain Reasoning," International Journal of General Systems, 23(4), 342-359

L. Zadeh [1994] "Fuzzy Logic, Neural Networks, and Soft Computing," CACM 37(3), 77-84