What If We Use Different "And"-Operations in the Same Expert System

Mahdokht Afravi and Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
mafravi@miners.utep.edu, vladik@utep.edu

***Abstract*—In expert systems, we often face a problem of estimating the expert's degree of confidence in a composite statement $A \,\&\, B$ based on the known expert's degrees of confidence $a = d(A)$ and $b = d(B)$ in individual statements A and B. The corresponding estimate $f_\&(a, b)$ is sometimes called an "and"-operation. Traditional fuzzy logic assumes that the same "and"-operation is applied to all pairs of statements. In this case, it is reasonable to justify that the "and"-operation be associative; such "and"-operations are known as t-norms. In practice, however, in different areas, different "and"-operations provide a good description of expert reasoning. As a result, when we combine expert knowledge from different areas into a single expert system, it is reasonable to use different "and"-operations to combine different statements. In this case, associativity is no longer a natural requirement. We show, however, that in such situations, under some reasonable conditions, associativity of each "and"-operation can still be deduced. Thus, in this case, we can still use associative t-norms.**

I. "And"-Operations Why and How: A Brief Reminder

Need for expert systems. In many practical situations, we reply on expert knowledge:

- we ask medical experts to help cure patients,
- we ask human expert in piloting to pilot planes, etc.

Ideally, everyone should have access to the top experts: top experts in medicine should cure all the patients, top pilots should pilot every plane, etc. However, there are very few best experts, and it is not realistic to expect these top experts to satisfy all the demands.

It is therefore desirable to describe the knowledge of the top experts inside a computer, so that other experts can use this knowledge. This descriptions are known as *expert systems.*

Need for degrees of certainty. Experts are usually not 100% certain about their statements. For example, a medical expert may indicate some visible signs of a heart attack, but the experts understand that there is no way, based only on the visible signs, to know with absolute certainty whether a patient is experiencing a heart attack.

To make adequate decisions, the expert system must not only store the experts' statement, it must also adequately store the experts' degrees of certainty in these statements; see, e.g., [12], [18], [26].

How degrees of certainty are usually represented. In the absence of uncertainty, an expert either knows that a given statement is true, or knows that this statement is false. In the computer, "true" is usually represented as 1, and "false" as 0.

It is therefore reasonable to describe intermediate degrees of certainty by numbers intermediate between 0 and 1.

Need for "and"-operations. One of the main objectives of an expert system is to help decision maker make decisions. Decisions are rarely based on a single expert statement; usually, two or more statements are used to argue for the proper decision.

For example, in the medical case, what we want is, given the symptoms (and test results, if available), to come up with an appropriate cure. However, medical rules rarely go from symptoms directly to cure. Usually, some rules describe a diagnosis based on the symptoms, and other rules describe a cure based on the diagnosis. So, to decide on an appropriate cure based on given symptoms, we must use at least two rules: a rule describing the diagnosis, and a rule selecting a cure based on the diagnosis.

It is desirable not just to make a recommendation, but also to estimate the degree of our certainty in this recommendation. For a recommendation based on several statements, we are certain in this recommendation if we are certain in all the statements used in deriving this recommendation. Thus, the degree to which we are confident is a given recommendation is the degree to which all these statements hold, i.e., to which the first statement holds *and* the second statement holds, etc.

Thus, not only we need to know the degrees to which each statement hold, we also need to know the degrees to which each possible "and"-combination of these statement hold. Ideally, we should elicit, from the experts, the degrees to which each such combination holds. However, this is not practically possible: for n statements, we can have $2^n - (n+1)$ possible combinations, so even for a reasonable value $n \approx 100$, we have an astronomical number of combinations.

Since we cannot elicit the degrees for all "and"-combinations directly from the experts, we must therefore estimate these degrees based on the known expert's degrees of confidence in the component statements. In other words, we need to be able, given the expert's degrees $a = d(A)$ and

$b = d(B)$ in two statements A and B, to come up with an estimate for the expert's degree of confidence in the "and"-combination $A \,\&\, B$. This estimate – depending on a and b – will be denoted by $f_\&(a, b)$; it is known as an "and"-operation.

t-norms. Usually, we assume that the same "and"-operation can be used for all possible pairs of statements (A, B). Under this assumption, we get reasonable requirements on the "and"-operation.

For example, since $A \,\&\, B$ means the same as $B \,\&\, A$, it is reasonable to require that the result of applying the "and"-operation should be the same for both "and"-combinations, i.e., we should have

$$f_\&(a, b) = f_\&(b, a)$$

for all a and b. In mathematical terms, this means that the "and"-operation should be commutative.

Similarly, since $A \,\&\, (B \,\&\, C)$ means the same as $(A \,\&\, B) \,\&\, C$, we should get the same estimate if we apply the "and"-operation to both expressions, i.e., we should have

$$f_\&(a, f_\&(b, c)) = f_\&(f_\&(a, b), c).$$

In mathematical terms, this means that the "and"-operation should be associative.

If A or B is absolutely true (e.g., if $a = d(A) = 1$), then our degree of belief in $A \,\&\, B)$ should be equal to our degree of belief in the remaining statement:

$$f_\&(1, b) = b \text{ and } f_\&(a, 1) = a.$$

If A or B is absolutely false, i.e., either $a = 0$ or $b = 0$, then, of course, the "and"-combination $A \,\&\, B$ should be false too, so we must have

$$f_\&(a, 0) = f_\&(0, b) = 0$$

for all a and b.

It is also reasonable to require that small changes in the degrees a and b only lead to small changes in the estimate, i.e., in mathematical terms, that the function $f_\&(a, b)$ be continuous.

Finally, if we increase our degree of confidence in A and/or B, this should either increase our degree of confidence in $A \,\&\, B$ or at least keep it the same. Thus, if $a \le a'$ and $b \le b'$, we should have $f_\&(a, b) \le f_\&(a', b')$. In mathematical terms, this means that the "and"-operation should be monotonic.

"And"-operations that satisfy these requirements are known as *t-norms*.

Additional requirement and Archimedean t-norms. If one of the degrees is 0 (e.g., if $a = d(A) = 0$), then increasing our degree of confidence in another statement does not change the estimate for $A \,\&\, B$: even if $b < b'$, we still have $f_\&(a, b) = f_\&(a, b')$, since both these values are equal to 0.

However, if none of the degree is equal to 0, then it is reasonable to require that an increase in degrees in A or B will increase our degree of confidence in $A \,\&\, B$: if $a > 0$ and $b < b'$, then $f_\&(a, b) < f_\&(a, b')$.

t-norms that satisfy this additional requirement are known as *Archimedean*.

Archimedean t-norms are universal approximators. Not all t-norms are Archimedean: e.g., $f_\&(a, b) = \min(a, b)$ is not an Arhimedean t-norm, but it can be proven that every t-norm can be approximated, with any given accuracy, by an Archimedean one; see, e.g., [17].

Since in practice, the degrees are known with some accuracy anyway, this universal approximation result means that without losing any generality, we can always assume that our t-norms are Archimedean.

General form an Archimedean t-norm. It is known that a general Archimedean t-norm can be obtained form

$$f_\&(a, b) = a \cdot b$$

by an appropriate re-scaling, i.e., it has the form

$$f_\&(a, b) = g^{-1}(g(a) \cdot g(b))$$

for some 1-1 continuous function $g : [0, 1] \to [0, 1]$.

Possibility of an inverse operation. When $a \le b$, then, for a multiplication t-norm $f_\&(a, b) = a \cdot b$, there exists a unique degree c for which $a = f_\&(b, c)$: namely, $c = a/b$. This inverse operation corresponds to *implication* $\supset$: $B \supset A$ is such a statement that, when we combine it with B, we get A.

When $a > b$, then such an inverse operation is not defined on the interval $(0, 1]$, but, since we have assumed strict monotonicity (i.e., the Archimedean property), we can naturally extend multiplication to the whole set of possible numbers. In this case, the inverse operation a/b is always uniquely defined for non-zero degrees.

Likewise, for all other Archimedean t-norms, we can get a similar extension if we extend the function $g(a)$ to the set of all real numbers.

II. Formulation of the Problem

Need for different "and"-operations. There are many different "and"-operations. In each area, we should select the one which is the best fit for the reasoning for experts from this area.

The first experience of selecting the appropriate "and"-operation came when researchers designed the world's first expert system MYCIN that collected expertise about rare blood diseases [7]. At first, the authors of MYCIN thought that their "and"-operations reflect the general features of human reasoning. However, when they tried to apply the same formulas to geophysical experts, it turned out that expertise of geophysicists corresponds to completely different "and"-operations.

It is now well known that in different control situations, different "and"-operations are most adequate – this depends, e.g., on whether we are interested in making smooth transitions or in the fastest way to achieve the goal; see, e.g., [16], [21].

Usually, in fuzzy logic, it is still assumed that the "and"-operation is the same in each problem – while it may differ from problem to problem. However, in interdisciplinary situations, it is reasonable to use different "and"-operations to combine degrees corresponding to statements from different disciplines.

In such situations, associativity is no longer a reasonable requirement, since we may use different "and"-operations to combine A and B than when we combine B and C.

Question. So what can we conclude in such a situation?

III. Towards Solving the Problem

Inverse operations. While in the general case, it is no longer reasonable to require associativity, it *is* still reasonable to require strict monotonicity. Thus, it is still reasonable to require that each "and"-operation can be extended to a large domain so that it becomes reversible – after we exclude the degree 0. In mathematical terms, this corresponds to the following definition.

Definition 1. *A function $f : V_a \times V_b \to V_c$ is called* invertible *if the following two conditions are satisfied:*

- *for every $a \in V_a$ and for every $c \in V_c$, there exists a unique value $b \in V_b$ for which $c = f(a, b)$;*
- *for every $b \in V_b$ and for every $c \in V_c$, there exists a unique value $a \in V_a$ for which $f(a, b) = c$.*

Comment. In mathematics, functions invertible in the sense of Definition 1 are called *generalized quasigroups*; see, e.g., [4].

Discussion. Please note that, to make our results most general, we did not assume commutativity. This makes sense: While in expert systems, we normally assume that "and"-operation is commutative, a natural language "and" is not always commutative. For example, the phrase "I ate a big dinner and I felt sleepy" has a different meaning than "I felt sleepy and I ate a big dinner".

We also do not necessarily assume that the degrees of confidence from different areas are described by the same set of values. In general, these sets V_a, V_b, and V_c can be all different.

What do we have instead of associativity? Suppose that we have four different types of statements (in general, each with its own set of possible degrees V_a, V_b, V_c, and V_d). We want to use the fact that the statements $(A \,\&\, B) \,\&\, (C \,\&\, D)$ and $(A \,\&\, C) \,\&\, (B \,\&\, D)$ are equivalent. It is therefore reasonable to require that for these two statements, we get the same estimates. The difference from the case when we use a single "and"-operation is that now, in general, we have one "and"-operations $f^{ab}_\&$ to combine values from V_a and V_b, another "and"-operation $f^{ac}_\&$ to combine value from V_a and V_c, etc.

To formalize this description, we also need to have sets of degrees for each of the combinations $A \,\&\, B$, $C \,\&\, D$, $A \,\&\, C$, and $B \,\&\, D$. We will denote these sets of degrees by, correspondingly, V_{ab}, V_{bd}, V_{ac}, and V_{bd}. We also need to describe a set of value V for the whole complex statement. Thus, we arrive at the following definition.

Definition 2. *Let V_a, V_b, V_c, V_d, V_{ab}, V_{cd}, V_{ac}, V_{bd}, and V be sets. We say that invertible operations*

$$f^{ab}_\& : V_a \times V_b \to V_{ab}, \quad f^{cd}_\& : V_c \times V_d \to V_{cd},$$
$$f^{ac}_\& : V_a \times V_c \to V_{ac}, \quad f^{bd}_\& : V_b \times V_d \to V_{bd},$$
$$f^{(ab)(cd)}_\& : V_{ab} \times V_{cd} \to V, \quad \textit{and}$$
$$f^{(ac)(bd)}_\& : V_{ac} \times V_{bd} \to V$$

satisfy the generalized associativity requirement *if for all $a \in V_a$, $b \in V_b$, $c \in V_c$, and $d \in V_b$, we have*

$$f^{(ab)(cd)}_\&(f^{ab}_\&(a, b), f^{cd}_\&(c, d)) =$$
$$f^{(ac)(bd)}_\&(f^{ac}_\&(a, c), f^{bd}_\&(b, d)).$$

Comment. In mathematical terms, this requirement is known as *generalized mediality* [4].

Groups and Abelian groups: reminder. To describe the main result, we need to recall that a set G with an associative operation $g(a, b)$ and a unit element e (for which $g(a, e) = g(e, a) = a$) is called a *group* if every element is invertible, i.e., if for every a, there exists an a' for which $g(a, a') = e$.

A group in which the operation $g(a, b)$ is commutative is known as *Abelian*.

Proposition. *For every set of invertible operations that satisfy the generalized associativity requirement, there exists an Abelian group G and 1-1 mappings*

$$r_a : V_a \to G, \quad r_b : V_b \to G,$$
$$r_c : V_c \to G, \quad r_d : V_d \to G,$$
$$r_{ab} : V_{ab} \to G, \quad r_{cd} : V_{cd} \to G,$$
$$r_{ac} : V_{ac} \to G, \quad r_{bd} : V_{bd} \to G, \quad \textit{and}$$
$$r : V \to G$$

for which, for all $a \in V_a$, $b \in V_b$, $c \in V_c$, $d \in V_d$, $v_{ab} \in V_{ab}$, $v_{cd} \in V_{cd}$, $v_{ac} \in V_{ac}$, and $v_{bd} \in V_{bd}$, we have:

$$f^{ab}_\&(a, b) = r^{-1}_{ab}(g(r_a(a), r_b(b)));$$
$$f^{cd}_\&(c, d) = r^{-1}_{cd}(g(r_c(c), r_d(d)));$$
$$f^{ac}_\&(a, c) = r^{-1}_{ac}(g(r_a(a), r_c(c)));$$
$$f^{bd}_\&(b, d) = r^{-1}_{bd}(g(r_b(b), r_d(d)));$$
$$f^{(ab)(cd)}_\&(v_{ab}, v_{cd}) = r^{-1}(g(r_{ab}(v_{ab}), r_{cd}(v_{cd})));$$
$$f^{(ac)(bd)}_\&(v_{ac}, v_{bd}) = r^{-1}(g(r_{ac}(v_{ac}), r_{bd}(v_{bd}))).$$

Proof. The proof of this statement is, in effect, contained in [2], [3], [4], [23], [24], [25].

Discussion. Thus, after appropriate re-scalings r_i, all the "and"-operations reduce to associative operation $g(a, b)$.

Conclusion. So, even if we have several different "and"-operations, and we can no longer directly justify associativity, associativity can still still be deduced from the natural generalized associativity requirement.

IV. Possible Application to Copulas: An Argument for Using Associative Copulas

What is a copula. Similar "and"-operations are used in the probabilistic case. Specifically, a 1-D probability distribution of a random variable X can be described by its *cumulative distribution function* (cdf) $F_X(x) \stackrel{\text{def}}{=} \text{Prob}(X \le x)$.

A 2-D distribution of a random vector (X, Y) can be similarly described by its 2-D cdf

$$F_{XY}(x, y) = \text{Prob}(X \le x \,\&\, Y \le y).$$

It turns out that we can always describe $F_{XY}(x, y)$ as

$$F_{XY}(x, y) = C_{XY}(F_X(x), F_Y(y))$$

for an appropriate function $C_{XY} : [0, 1] \times [0, 1] \to [0, 1]$ known as a *copula*; see, e.g., [15], [20].

For a joint distribution of several random variables X, Y, $\ldots$, Z, we can similarly write

$$F_{XY\ldots Z}(x, y, \ldots, z) \stackrel{\text{def}}{=}$$

$$\text{Prob}(X \le x \,\&\, Y \le y \,\&\, \ldots \,\&\, Z \le z) =$$

$$C_{XY\ldots Z}(F_X(x), F_Y(y), \ldots, F_Z(z))$$

for an appropriate multi-D copula $C_{XY\ldots Z}$.

Vine copulas. When we have many ($n \gg 1$) random variables, then to exactly describe their joint distribution, we need to describe a general function of n variables. Even if we use two values for each variable, we get 2^n combinations, which for large n can be astronomically large. Thus, a reasonable idea is to approximate the multi-D distribution.

A reasonable way to approximate is to use 2-D copulas. For example, to describe a joint distribution of three variables X, Y, and Z, we first describe the joint distribution of X and Y as $F_{XY}(x, y) = C_{XY}(F_X(x), F_Y(y))$, and then use an appropriate copula $C_{XY,Z}$ to combine it with $F_Z(z)$:

$$F_{XYZ}(x, y, z) \approx C_{XY,Z}(F_{XY}(x, y), F_Z(z)) =$$

$$C_{XY,Z}(C_{XY}(F_X(x), F_Y(y), F_Z(z)).$$

Such an approximation, when copulas are applied to one another like a vine, are known as *vine copulas*; see, e.g., [1], [5], [6], [8], [9], [10], [11], [13], [14], [19], [22].

Natural analogue of associativity. It is reasonable to require that the result of the vine copula approximation should not depend on the order in which we combine the variables. In particular, for four random variables X, Y, Z, and T, we should get the same result in the following two situations:

- if we first combine X with Y, Z and T, and then combine the two results; or
- if we first combine X with Z, Y with T, and then combine the two results.

Thus, we require that for all possible real numbers x, y, z, and t, we get

$$C_{XY,ZT}(C_{XY}(F_X(x), F_Y(y)), C_{ZT}(F_Z(z), F_T(t))) =$$

$$C_{XZ,YT}(C_{XZ}(F_X(x), F_Z(z)), C_{YT}(F_Y(y), F_T(t))).$$

If we denote $a = F_X(x)$, $b = F_Y(y)$, $c = F_Z(z)$, and $d = F_T(t)$, we conclude that for every a, b, c, and d, we have

$$C_{XY,ZT}(C_{XY}(a, b), C_{ZT}(c, d)) =$$

$$C_{XZ,YT}(C_{XZ}(a, c), C_{YT}(b, d)).$$

This is exactly our generalized associativity requirement. Thus, if we assume that the copulas are invertible, we conclude that they can be *re-scaled to associative operations* – in the sense of the above Theorem.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

References

[1] K. Aas, C. Czado, A. Frigessi, and H. Bakken, "Pair-copula constructions of multiple dependence", *Insurance: Mathematics and Economics*, 2009, Vol. 44, pp. 182–198.

[2] J. Aczél, V. D. Belousov, and M. Hosszú, "Generalized associativity and bisymmetry on quasigroups", *Acta Math. Acad. Sci, Hungar.*, 1960, Vol. 11, pp. 127–136.

[3] J. Aczeél, "Quasigroup-net-nomograms", *Advances in Mathematics*, 1965, Vol. 1, pp. 483–450.

[4] J. Aczél and J. Dhombres, *Functional Equations in Several Variables*, Camridge University Press, Cambridge, UK, 2008.

[5] T. Bedford and R. M. Cooke, "Monte Carlo simulation of vine dependent random variables for applications in uncertainty analysis", *Proceedings of European Safety and Reliability Conference ESREL'2001*, Turin, Italy, 2001.

[6] T. Bedford and R. M. Cooke, "Vines – a new graphical model for dependent random variables", *Annals of Statistics*, 2002, Vol. 30, No. 4, pp. 1031–1068.

[7] B. G. Buchanan and E. H. Shortliffe, *Rule Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, Massachusetts, 1984.

[8] C. Czado, "Pair-copula constructions of multivariate copulas", In: P. Jaworski (Ed.), *Copula Theory and Its Applications*, Lecture Notes in Statistics, Vol. 198, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 93–109.

[9] H. Joe, "Dependence comparisons of vine copulae with four or more variables", In: D. Kurowicka and H. Joe (Eds.), *Dependence Modeling: Vine Copula Handbook*, World Scientific, Singapore, 2010.

[10] H. Joe and T. Hu, "Multivariate distributions from mixtures of max-infinitely divisible distributions", *Journal of Multivariate Analysis*, 1996, Vol. 57, No. 2, pp. 240–265.

[11] H. Joe, H. Li, and A.K. Nikoloulopoulos, "Tail dependence functions and vine copulas", *Journal of Multivariate Analysis*, 2010, Vol. 101, pp. 252–270.

[12] G. Klir and B. Yuan, "Fuzzy Sets and Fuzzy Logic", Prentice Hall, Upper Saddle River, New Jersey, 1995.

[13] D. Kurowicka and R.M. Cooke, *Uncertainty Analysis with High Dimensional Dependence Modelling*, Wiley, New York, 2006.

[14] D. Kurowicka and H. Joe (Eds.), *Dependence Modeling: Vine Copula Handbook*, World Scientific, Singapore, 2010.
[15] R. B. Nelsen, *An Introduction to Copulas*, Springer, New York, 1999.
[16] H. T. Nguyen, V. Kreinovich, "Methodology of fuzzy control: an introduction", In: H. T. Nguyen and M. Sugeno (eds.), *Fuzzy Systems: Modeling and Control*, Kluwer, Boston, MA, 1998, pp. 19–62.
[17] H. T. Nguyen, V. Kreinovich, and P. Wojciechowski, "Strict Archimedean t-Norms and t-Conorms as Universal Approximators", *International Journal of Approximate Reasoning*, 1998, Vol. 18, Nos. 3–4, pp. 239–249.
[18] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.
[19] A. K. Nikoloulopoulos, "Vine copulas with asymmetric tail dependence and applications to financial return data", *Computational Statistics and Data Analysis*, 2012, Vol. 56, pp. 3659–3673.
[20] A. Sklar, "Fonctions de répartition á n dimensions et leurs marges", *Publ. Inst. Statist. Univ. Paris*, 1959, Vol. 8, pp. 229–231.
[21] M. H. Smith and V. Kreinovich. "Optimal strategy of switching reasoning methods in fuzzy control", Chapter 6 in H. T. Nguyen, M. Sugeno, R. Tong, and R. Yager (eds.), *Theoretical aspects of fuzzy control*, J. Wiley, New York, 1995, pp. 117–146.
[22] S. Sriboonchitta, O. Kosheleva, and H. T. Nguyen, "Why Are Vine Copulas So Successful in Econometrics?", *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems (IJUFKS)*, 2015, Vol. 23, Suppl. 1, pp. 133–142.
[23] M. A. Taylor, "The generalized equations of bisymmetry, associativity and transitivity on quasigroups", *Canadian Math. Bulletin*, 1972, Vol. 15, pp. 119–124.
[24] M. A. Taylor, "Certian functional equations on groupoids weaker than quasigroups", *Aequationes Mathematicae*, 1973, Vol. 9, pp. 23–29.
[25] M. A. Taylor, "On the generalized equations of associativity and bisymmetry", *Aequationes Mathematicae*, 1978, Vol. 17, pp. 154–163.
[26] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.