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
El Paso, TX 79968, USA
mafravi@miners.utep.edu
vladik@utep.edu

1. Need for Expert Systems

- We often rely on expert knowledge; e.g.:
  - we ask medical experts to help cure patients,
  - we ask human expert in piloting to pilot planes.
- 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.
- So, 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.
- Then other experts can use this knowledge.
- This descriptions are known as *expert systems.*

2. 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
  - but experts cannot tell with absolute certainty whether a patient is experiencing a heart attack.
- The expert system must store the experts' degrees of certainty in different statements.
- In the computer, "absolutely true" is usually represented by 1, and "absolutely false" by 0.
- Thus, intermediate degrees of certainty are usually described by numbers between 0 and 1.

3. 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, we want is, given the symptoms, 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 test results), and
  - other rules describe a cure based on the diagnosis.

4. Need for "And"-Operations (cont-d)

- 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, in addition to a recommendation r, to also estimate our degree of certainty in r.
- 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:
  - the first statement holds *and*
  - the second statement holds, etc.

5. Need for "And"-Operations (cont-d)

- So, we 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.
- We cannot elicit the degrees for all "and"-combinations directly from the experts.
- We must therefore estimate these degrees based on the known degrees of confidence in individual statements.

6. "And"-Operations and t-Norms

- 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.
- 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 known as *t-norms*.
- For example, $A \,\&\, B$ means the same as $B \,\&\, A$.
- It is thus reasonable to require that $f_\&(a, b) = f_\&(b, a)$.

7. t-Norms (cont-d)

- Similarly, $A \,\&\, (B \,\&\, C)$ means the same as $(A \,\&\, B) \,\&\, C$, so 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.
- Also:
  - if we increase our degree of confidence in A and/or B,
  - this should either increase our degree of confidence in $A \,\&\, B$.
- So, the "and"-operation should be *monotonic*:

$$\text{if } a \le a' \text{ and } b \le b', \text{ then } f_{\&}(a, b) \le f_{\&}(a', b').$$

8. Archimedean t-Norms

- If $a = d(A) = 0$, then increasing our degree of confidence in B does not change the estimate for $A \,\&\, B$:

$$b < b', \text{ but } f_\&(a, b) = f_\&(a, b') = 0.$$

- However, if $a > 0$, then it's reasonable to require that in b increases confidence in $A \,\&\, B$:

$$\text{if } a > 0 \text{ and } b < b', \text{ then } f_\&(a, b) < f_\&(a, b').$$

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

- Not all t-norms are Archimedean: e.g., $f_\&(a, b) = \min(a, b)$ is not an Arhimedean t-norm.

9. Archimedean t-Norms (cont-d)

- However, it can be proven that every t-norm can be approximated,
  - with any given accuracy,
  - by an Archimedean one.
- In practice, the degrees are known with some accuracy anyway.
- Thus, without losing any generality, we can always assume that our t-norms are Archimedean.
- A general Archimedean t-norm can be obtained from $f_\&(a, b) = a \cdot b$ by a re-scaling:

$$f_\&(a, b) = g^{-1}(g(a) \cdot g(b)) \text{ for some 1-1 cont. } g : [0, 1] \to [0, 1].$$

- When $a \le b$, then, for $f_\&(a, b) = a \cdot b$, there exists a unique c for which $a = f_\&(b, c)$: namely, $c = a/b$.

10. Inverse Operations

- The 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 $(0, 1]$.
- However, we can naturally extend multiplication to all 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.

11. Formulation of the Problem

- 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.
- This started with the world's first expert system MYCIN (on rare blood diseases).
- At first, MYCIN's authors thought that their "and"-operations are general.
- However, it turns out that geophysicists use 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.

12. Formulation of the Problem (cont-d)

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

13. Towards Solving the Problem

- In the general case, 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.
- *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* $f(a, b) = c$*.*
- In mathematics, functions invertible in the sense of this Definition are called *generalized quasigroups*.

14. Towards Solving the Problem (cont-d)

- Please note that, to make our results most general, we did not assume commutativity:
  - while in expert systems, we normally assume that "and"-operation is commutative,
  - a natural language "and" is not always commutative.
- For example, "I ate a big dinner and I felt sleepy" is different from "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.

15. What Do We Have Instead of Associativity?

- Suppose that we have four different types of statements.
- In general, each type has its own set of possible degrees V_a, V_b, V_c, and V_d.
- We want to use the equivalence of the statements $(A \,\&\, B) \,\&\, (C \,\&\, D)$ and $(A \,\&\, C) \,\&\, (B \,\&\, D)$.
- 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.

16. Instead of Associativity (cont-d)

- To formalize this description, we also need to have sets of degrees for each of the combinations $A \& B$, etc.
- 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.

17. Main Definition

- *Let V_a, V_b, V_c, V_d, V_{ab}, V_{cd}, V_{ac}, V_{bd}, and V be sets.*
- *Let us consider invertible operations:*

$$f^{ab}_{\&} : V_a \times V_b \to V_{ab}, \quad f^{cd}_{\&} : V_c \times V_d \to V_{cd},$$

$$f^{ac}_{\&} : V_a \times V_c \to V_{ac}, \quad f^{bd}_{\&} : V_b \times V_d \to V_{bd},$$

$$f^{(ab)(cd)}_{\&} : V_{ab} \times V_{cd} \to V, \quad \textit{and}$$

$$f^{(ac)(bd)}_{\&} : V_{ac} \times V_{bd} \to V$$

- *We say that these operations* satisfy the generalized associativity requirement *if for all $a \in V_a$, $b \in V_b$, ...:*

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

- *Comment:* In mathematical terms, this requirement is known as *generalized mediality.*

18. 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*.

19. Main Result

For every set of invertible operations that satisfy the generalized associativity requirement:

- *there exists an Abelian group G and 1-1 mappings*

$$r_a : V_a \to G, \quad r_b : V_b \to G, \quad r_c : V_c \to G, \quad r_d : V_d \to G,$$

$$r_{ab} : V_{ab} \to G, \quad r_{cd} : V_{cd} \to G, \quad r_{ac} : V_{ac} \to G,$$

$$r_{bd} : V_{bd} \to G, \quad r : V \to G$$

- *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:*

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

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

$$f^{(ab)(cd)}_{\&}(v_{ab}, v_{cd}) = r^{-1}(g(r_{ab}(v_{ab}), r_{cd}(v_{cd})));$$

$$f^{(ac)(bd)}_{\&}(v_{ac}, v_{bd}) = r^{-1}(g(r_{ac}(v_{ac}), r_{bd}(v_{bd}))).$$

20. Discussion

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

21. Possible Application to Copulas

- Similar "and"-operations are used for probabilities.
- A 1-D probability distribution 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, for an appropriate function $C_{XY} : [0,1] \times [0,1] \to [0,1]$ (known as a *copula*) we have

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

22. Copulas (cont-d)

- For several random variables X, Y, $\ldots$, Z, we have:

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

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

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

- To describe a joint distribution of n variables, we need a function of n variables.
- Even if we use two values for each variable, we get 2^n combinations.
- For large n, this is astronomically large.
- Thus, a reasonable idea is to approximate the multi-D distribution.
- A reasonable way to approximate is to use 2-D copulas.

23. Vine 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*.
- It is reasonable to require that the result should not depend on the combination order.

24. Vine Copulas (cont-d)

- In particular, for four random variables X, Y, Z, and T, we should get the same result:
  - 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 $a = F_X(x)$, $b = F_Y(y)$, $c = F_Z(z)$, $d = F_T(t)$, then for all a, b, c, and d:

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

25. Copulas: Conclusion

- We have argued that the following equality is true:

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

- 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.

26. 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 from Prudential Foundation.