Fuzzy Logic Beyond Traditional "And"-Operations⋆

Vladik Kreinovich[0000−0002−1244−1650] and
Olga Kosheleva[0000−0003−2587−4209]

University of Texas at El Paso, El Paso, TX 79968, USA,
vladik@utep.edu, olgak@utep.edu

Abstract. In the traditional fuzzy logic, we can use "and"-operations (also known as t-norms) to estimate the expert's degree of confidence in a composite statement $A\,\&\,B$ based on his/her degrees of confidence $d(A)$ and $d(B)$ in the corresponding basic statements A and B. But what if we want to estimate the degree of confidence in $A\,\&\,B\,\&\,C$ in situations when, in addition to the degrees of estimate $d(A)$, $d(B)$, and $d(C)$ of the basic statements, we also know the expert's degrees of confidence in the pairs $d(A\,\&\,B)$, $d(A\,\&\,C)$, and $d(B\,\&\,C)$? Traditional "and"-operations can provide such an estimate – but only by ignoring some of the available information. In this paper, we show that, by going beyond the traditional "and"-operations, we can find a natural estimate that takes all available information into account – and thus, hopefully, leads to a more accurate estimate.

Keywords: Fuzzy logic · "And"-operations (t-norms) · Maximum entropy approach.

1 Formulation of the Problem

Why do we need fuzzy logic in the first place? A large amount of human activity has been automated, but in many areas, human expertise, human skills are still needed. We use human doctors when we are ill, we use human drivers and human pilots, etc.

Not all experts and specialists are equal, some are much better than others. In the ideal world, all diagnoses will be made by the top medical doctors, all planes should be controlled by the top pilots – but in reality, there are not that many top doctors, not that many top pilots, not that many top drivers, and it is not possible for them to serve all patients and all the planes.

It is therefore desirable to use the knowledge of the top experts to help others make better decisions – and even, if possible, to design automatic systems that

⋆ This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes). It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.

would diagnose patients, fly planes, and drive cars as well as the best human specialists.

Usually, top experts are quite willing to share their expertise, to teach others. But the problem is that when their share their expertise, they use imprecise ("fuzzy") words from natural language like "small", "medium", "large", "short", etc. This is easy to explain: many of us drive cars, but hardly anyone can express his/her driving experience in precise terms. If you ask any driver how much to brake if a car 100 meters in front slows down from 100 to 95 km/h, a natural answer is "a little bit" – while an automatic system needs to know for how many milliseconds to press the brake and with how many Newtons of force.

To describe such important knowledge in precise terms, Lotfi Zadeh came up with the idea of *fuzzy logic*; see, e.g., [1, 5, 8, 12, 13, 15]. His main observation was that, in contrast to properties like "less than 0.5 sec" which are either true or false for any given time duration, for properties like "short" the situation is different: yes, very short time durations are absolutely short, and very long time durations are absolutely not short, but for intermediate time durations, their "shortness" is only true to some degree.

In a computer, "absolutely true" is usually represented by 1, and "absolutely not true" ("false") by 0. It is therefore reasonable to characterize intermediate degrees of confidence by numbers between 0 and 1. This is exactly what Zadeh proposed to describe properties like "small": ask the expert to indicate, for each possible value x of the corresponding quantity, to what extent – on the $[0,1]$-scale – this value is small. The resulting function $\mu(x)$ assigning a degree to each value x is known as the *membership function* or, alternatively, as the *fuzzy set*.

Why we need "and"-operations (t-norms). Expert rules usually have several conditions: e.g., we can have a braking rule that describes what happens when the car is close *and* slows down a little bit, we can have a different rule that describes what happens when the car is close *and* slows down drastically.

We can ask an expert, for each possible value d of the distance, to what extent this distance is close. We can also ask the expert, for each possible value Δv of slowing down, to what extent this value can be described as "a little bit". But what we need, to implement this rule, is to know the degree to which, for two given values d and Δv, to what extent d is small *and* Δv corresponds to "a little bit".

Strictly speaking, for this, we need to ask the expert's opinion about all possible pairs of values. Often – e.g., in medical diagnostics – we need to take into account the values not of two but of a dozen or more different quantities: temperature, upper and lower blood pressure, pulse, etc. Even if we use only 3 or 4 different values of each quantity, we can have 3^{12} or 4^{12} possible combinations of values. The value 4^{12} is about 16 million, and there is no way that we can ask the expert these thousands and millions of questions.

Since we cannot directly ask the expert about his/her degree of confidence in all possible "and"-combinations $S_1 \,\&\, S_2 \,\&\, \ldots \,\&\, S_n$, we therefore need to be able, given the expert's degrees of confidence a and b in statement A and B, to estimate his/her degree of confidence in the composite statement $A \,\&\, B$. The value of the resulting estimate – which we will denote by $f_\&(a,b)$ – is known as the "and"-operation or, for historical reason, a *t-norm*.

From the meaning of this operation, we can extract its natural properties. For example, "A and B" means the same as "B and A". It is therefore reasonable to require that our estimates for these two equivalent statements coincide, i.e., that $f_\&(a,b) = f_\&(b,a)$ for all a and b. In mathematical terms, this means that the "and"-operation should be commutative.

Similarly, since "(A and B) and C" means the same as "A and (B and C)", we can conclude that the resulting estimates should coincide, i.e., that we should have $f_\&(f_\&(a,b),c) = f_\&(a, f_\&(b,c))$ for all a, b, and c. In mathematical terms, this means that the "and"-operation should be associative.

Similar arguments explain that the "and"-operation should be monotonic, continuous, etc.

There are many possible "and"-operations. There exist many operations that satisfy all these properties. We need to select the one which best reflects the expert's reasoning.

This selection was first done for the historically first medical expert system MYCIN (see, e.g., [2]), and since then, has been done for many application areas. Interestingly, in different application areas – and sometimes even in the same application areas but for different tasks – different "and"-operations are most adequate.

Comment. The desired most adequate "and"-operation can be determined as follows:

- for several pairs of statements (A_k, B_k), we ask the experts to estimate their degrees of confidence $d(A_k)$, $d(B_k)$, and $d(A_k \,\&\, B_k)$ in statements A_k, B_k, and $A_k \,\&\, B_k$, and then
- we find a function $f_\&(a,b)$ for which $d(A_k \,\&\, B_k) \approx f_\&(d(A_k), d(B_k))$ for all k.

Why do we need to go beyond traditional "and"-operations. So far, we have considered two extreme situations. To describe such situations, let us denote possible basic statements by $S_1, \ldots, S_n$.

- In the first – ideal – situation, we know the expert's degrees of confidence in these statements $d(S_i)$ and in all possible "and"-combinations of these statements $d(S_{i_1} \,\&\, \ldots \,\&\, S_{i_k})$.
- The second – more realistic – situation is when we only know the degrees of confidence $d(S_i)$ in the basic statements. In this case, we estimate our degree of confidence in each "and"-combination $S_{i_1} \,\&\, \ldots \,\&\, S_{i_k}$ as $f_\&(d(S_{i_1}), \ldots, d(S_{i_k}))$.

The problem is that in practice, we sometimes have intermediate situations, when we know the degrees of confidence in *some* – but not all – "and"-combinations, and we are interested in estimating the expert's degree of confidence in other "and"-combinations. For example, in addition to the degrees of confidence $d(S_1)$, $d(S_2)$, and $d(S_3)$ in the three basic statements, we may know the degrees of confidence in all possible pairs $d(S_1 \,\&\, S_2)$, $d(S_1 \,\&\, S_3)$, and $d(S_2 \,\&\, S_3)$, and we want to estimate the degree of confidence $d(S_1 \,\&\, S_2 \,\&\, S_3)$ in all three of these statements.

By using the traditional "and"-operation, we can several estimates for this desired degree, e.g., $f_\&(d(S_1), d(S_2 \,\&\, S_3))$, $f_\&(d(S_1 \,\&\, S_2), d(S_3))$, etc., but they will be, in general, different – and each of them takes into account some available information while ignoring other information.

How can we take all the available information into account – and thus come up with the most adequate estimate? We cannot do this by using the traditional "and"-operations, we need to go beyond.

This is what we will do in this paper: we will show how such an estimate can be obtained.

2 Analysis of the Problem

What are subjective probabilities and how they are related to fuzzy degrees. The ultimate goal of expert's estimates is to make a decision. The diagnosis of a medical expert helps decide which treatment to select for a given patient. The decision of an expert pilot helps decide how, e.g., how to best avoid the turbulence zone. So, to solve problems related to expert estimates, it makes sense to recall how exactly these estimates are used in decision making.

Decision theory – see, e.g., [3, 6, 7, 9, 11, 14] – deals, in particular, with situations in which a decision maker is uncertain about some possible events E. Decision theory provides a natural scale for measuring this uncertainty – namely, we compare the E-related lottery

$$L(E) \stackrel{\text{def}}{=} \text{"I get \$100 if } E\text{, otherwise I get nothing"}$$

with lotteries $L(p)$ in which a person gets \$100 with some probability p.

When this probability is equal to 1, i.e., when the person gets \$100 unconditionally, then clearly the lottery $L(1)$ is better; we will denote this situation by $L(E) < L(1)$. On the other hand, if the probability p is equal to 0, then the person does not get anything at all, so clearly the lottery $L(E)$ in which there is a change to get something is better: $L(0) < L(E)$.

As we increase the probability p from 0 and continue comparing, at some probability level p_0, we will switch from $L(p) < L(E)$ to $L(E) < L(p)$. This threshold value p_0 is known as the *subjective probability* $ps(E)$ of the event E.

Both degree of confidence and subjective probability describe our degree of belief that the event will happen – i.e., that the corresponding statement is true. If in two situations, we have the same degree of belief, it is reasonable to expect that we have the same subjective probabilities and the same degrees of confidence. In mathematical terms, this means that the degree of confidence uniquely determines the subjective probability, i.e., that $ps(E) = f(d(E))$ for some monotonic function $f(d)$.

How can we determine the corresponding function $f(d)$**?** If we know the degrees of confidence a and b in statements A and B, then we estimate the degree of confidence in $A \,\&\, B$ as $f_\&(a,b)$.

What if we only know the subjective probabilities $ps(A)$ and $ps(B)$ and we want to estimate the subjective probability $ps(A \,\&\, B)$? In principle, we have several different probability measures with different values of $ps(A \,\&\, B)$. Which of these values should we choose?

The usual approach in probability theory is to take into account that different alternative have different uncertainty – as measured, e.g., by *entropy* – the average number of binary ("yes"-"no") questions that we need to ask to fully determine the situation. In general, if have N alternatives with probabilities $P_1,\ldots,P_N$, then entropy is equal to $S = -\sum\limits_{i=1}^{N} P_i \cdot \log_2(P_i)$; see, e.g., [4, 11]. For two statements A and B, we have 4 possible alternatives: $A \,\&\, B$, $A \,\&\, \neg B$, $\neg A \,\&\, B$, and $\neg A \,\&\, \neg B$. Once we know the probabilities $p(A)$, $p(B)$, and $p(A \,\&\, B)$, we can determine the probabilities P_i of all these 4 events: indeed, we already know the probability $P_1 = p(A \,\&\, B)$, we can then determine

$$P_2 = p(A \,\&\, \neg B) = p(A) - p(A \,\&\, B), \quad P_3 = p(\neg A \,\&\, B) = p(B) - p(A \,\&\, B),$$

$$\text{and } P_4 = p(\neg A \,\&\, \neg B) = 1 - p(A \,\&\, B) - p(A, \&\, \neg B) = p(\neg A \,\&\, B).$$

For different values of $p(A \,\&\, B)$, we get, in general, different values of the entropy: some are smaller, some are larger. The only thing that we know about this uncertainty is that it is in some interval $[\underline{S}, \overline{S}]$. We can guarantee that the average number of binary questions does not exceed $\overline{S}$.

If we select a value $p(A \,\&\, B)$ for which $S < \overline{S}$, we then artificially add certainty which is not there, we kind of cheating by pretending that we have less uncertainty than possible. To avoid such cheating, it makes sense to select the value $p(A \,\&\, B)$ for which $S = \overline{S}$, i.e., for which entropy is the largest possible. This idea is known as the *maximum entropy approach*.

For the above case, as one can show, this approach leads to $p(A \,\&\, B) = p(A) \cdot p(B)$. In particular, for subjective probabilities, we get $ps(A \,\&\, B) = ps(A) \cdot ps(B)$. Taking into account that $ps(A) = f(d(A)) = f(a)$, $ps(B) = f(d(B)) = f(b)$, and $ps(A \,\&\, B) = f(d(A \,\&\, B)) = f(f_\&(a,b))$, we conclude that $f(f_\&(a,b)) = f(a) \cdot f(b)$, i.e., equivalently, that

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

Here $f^{-1}(p)$ denotes the inverse function: $f^{-1}(p)$ is the value d for which $f(d) = p$.

So, once we empirically determine the "and"-operation $f_\&(a,b)$, we can then determine the corresponding function $f(d)$ as the one for which (1) holds.

Is not the formula (1) an additional restriction on possible "and"-operations? Can such a function $f(d)$ be found for all possible "and"-operations? From the purely mathematical viewpoint, the formula (1) is indeed a limitation: e.g., a popular "and"-operation $f_\&(a,b) = \min(a,b)$ cannot be represented in this form.

However, from the practical viewpoint, there is no limitation: it is known (see, e.g., [10]) that for every "and"-operation $f_\&(a,b)$ and for every $\varepsilon > 0$, there exists an ε-close "and"-operation of the type (1). For sufficiently small $\varepsilon > 0$, ε-close operations are practically indistinguishable: hardly an expert can say that his/her degree of confidence is 0.51 and not 0.52.

Maximum entropy approach is more general than using for "and"-operations. We have mentioned that the maximum entropy approach can be used to estimate the probability $p(A\,\&\,B)$ of an "and"-statement $A\,\&\,B$ when all we know are probabilities $p(A)$ and $p(B)$ of the basic statements. However, the same maximum entropy approach can be – and is – used in many other situations when we only have partial information about the probabilities. It can be used to find any missing probability – including a missing probability of an "and"-combination.

Since there is a natural transformation $ps = f(d)$ from degrees of confidence d to probabilities ps, we can therefore find the missing degrees as follows:

- first, we transform all known degrees into probabilities;
- then, we use the Maximum Entropy approach to find the missing probabilities;
- finally, we use the inverse function $f^{-1}(p)$ to transform the newly found probabilities into degrees of confidence.

Let us describe this procedure in precise terms.

3 Resulting Procedure

Preliminary step: Version 1. For some application areas (and for the given class of problems), we have already determined the "and"-operation $f_\&(a,b)$ that most adequately describes the expert reasoning in this area.

In this case, we find a function $f(d)$ for which, for every a and b, we have $f(f_\&(a,b)) = f(a) \cdot f(b)$.

Preliminary step: Version 2. In some application areas, we have not yet determined the appropriate "and"-operation. In such cases:

- for several pairs of statements (A_k, B_k), we ask the experts to estimate their degrees of confidence $d(A_k)$, $d(B_k)$, and $d(A_k\,\&\,B_k)$ in statements A_k, B_k, and $A_k\,\&\,B_k$, and then
- we find a function $f(d)$ for which $f(d(A_k\,\&\,B_k)) \approx f(d(A_k)) \cdot f(d(B_k))$ for all k.

The corresponding problem. We have several basic statements $S_1, \ldots, S_n$. For some propositional combinations $C_1, \ldots, C_m$ of these statements, we have expert estimates $d(C_i)$ of their degree of confidence. We also have another propositional combination C for which we do not have the expert's estimate.

Based on the available information – i.e., on the values $d(C_i)$ – we want to estimate the expert's degree of confidence $d(C)$ in the statement C.

Example 1. The traditional "and"-operation corresponds to the case when $n = 2$, $m = 2$, $C_1 = S_1$, $C_2 = S_2$ and $C = S_1 \,\&\, S_2$. This is the case for which the traditional fuzzy "and"-operation provides a reasonable solution

$$d(C) \approx f_\&(d(S_1), d(S_2)) = f^{-1}(f(d(S_1), f(d(S_2))).$$

Example 2. Here is an example when we need to go beyond the traditional "and"-operation: $n = 3$, $m = 6$, $C_1 = S_1$, $C_2 = S_2$, $C_3 = S_3$, $C_4 = S_1 \,\&\, S_2$, $C_5 = S_1 \,\&\, S_3$, $C_6 = S_2 \,\&\, S_3$, and $C = S_1 \,\&\, S_2 \,\&\, S_3$. We know the values $d_i \stackrel{\text{def}}{=} d(S_i)$ and $d_{ij} \stackrel{\text{def}}{=} d(S_i \,\&\, S_j)$, and we want to estimate the degree $d \stackrel{\text{def}}{=} d(S_1 \,\&\, S_2 \,\&\, S_3)$.

Solution.

– first, we transform all known degrees $d(C_i)$ into subjective probabilities, by computing $ps(C_i) = f(d(C_i))$;
– then, among all the probability distributions with given values $ps(C_i)$, we find the one for which the entropy is the largest possible, and for this maximum-entropy distribution, we determine the (subjective) probability $ps(C)$;
– finally, we transform this probability back into degrees by computing $d(C) = f^{-1}(ps(C))$.

Comment. For n statements, to get a full probability distribution, we need to know the probability of all 2^n atomic combinations, i.e., combinations of the form $S_1^{\varepsilon_1} \,\&\, \ldots \,\&\, S_n^{\varepsilon_n}$, where ε_i is either $+$ or $-$, S_i^+ means S_i, and S_i^- means $\neg S_i$.

Thus, the entropy is $-\sum\limits_{\varepsilon_1,\ldots,\varepsilon_n} ps(S_1^{\varepsilon_1} \,\&\, \ldots \,\&\, S_n^{\varepsilon_n}) \cdot \log_2(ps(S_1^{\varepsilon_1} \,\&\, \ldots \,\&\, S_n^{\varepsilon_n}))$.

Example. In the second example, first, we compute the probabilities $p_i = ps(S_i) = f(d_i)$ and $p_{ij} = ps(S_i \,\&\, S_j) = f(d_{ij})$.

Once we know the subjective probability p of the desired statement $S_1 \,\&\, S_2 \,\&\, S_3$, we can then determine the (subjective) probabilities of all 8 atomic statements:

$$ps(S_1 \,\&\, S_2 \,\&\, S_3) = p; \quad ps(S_1 \,\&\, S_2 \,\&\, \neg S_3) = p_{12} - p;$$

$$ps(S_1 \,\&\, \neg S_2 \,\&\, S_3) = p_{13} - p; \quad ps(\neg S_1 \,\&\, S_2 \,\&\, S_3) = p_{23} - p;$$

$$ps(S_1 \,\&\, \neg S_2 \,\&\, \neg S_3) = p_1 - p_{12} - p_{13} + p;$$

$$ps(\neg S_1 \,\&\, S_2 \,\&\, \neg S_3) = p_2 - p_{12} - p_{23} + p;$$

$$ps(\neg S_1 \,\&\, \neg S_2 \,\&\, S_3) = p_3 - p_{13} - p_{23} + p;$$

$$ps(\neg S_1 \,\&\, \neg S_2 \,\&\, \neg S_3) = 1 - p_1 - p_2 - p_3 + p_{12} + p_{13} + p_{23} - p.$$

The value p can be determined by maximizing the corresponding entropy.

Comment. In the simplified case when $p_1 = p_2 = p_3$ and $p_{12} = p_{13} = p_{23}$, the expression for the entropy has the form

$$-p \cdot \log_2(p) - 3(p_{ij} - p) \cdot \log_2(p_{ij} - p) - 3(p_i - 2p_{ij} + p) \cdot \log_2(p_i - 2p_{ij} + p) -$$
$$(1 - 3p + 3p_{ij} - p) \cdot \log_2(1 - 3p + 3p_{ij} - p).$$

References

1. Belohlavek, R., Dauben, J. W., Klir, G. J.: Fuzzy logic and mathematics: a historical perspective. Oxford University Press, New York (2017)
2. Buchanan, B. G., Shortliffe, E. H.: Rule-based expert systems: the mycin experiments of the stanford heuristic programming project. Addison-Wesley, Reading, Massachusetts (1984)
3. Fishburn, P. C.: Utility Theory for Decision Making. John Wiley & Sons Inc., New York (1969)
4. Jaynes, E. T., Bretthorst, G. L. (2003) Probability theory: the logic of science. Cambridge University Press, Cambridge, UK (2003)
5. Klir, G., Yuan, B.: Fuzzy sets and fuzzy logic. Prentice Hall, Upper Saddle River, New Jersey (1995)
6. Kreinovich, V.: Decision making under interval uncertainty (and beyond)", In: Guo. P., Pedrycz, W. (eds.) Human-centric decision-making models for social sciences. Springer Verlag (2014) pp. 163–193
7. Luce, R. D., Raiffa, R.: Games and Decisions: Introduction and Critical Survey. Dover, New York (1989)
8. Mendel, J. M.: Uncertain rule-based fuzzy systems: introduction and new directions. Springer, Cham, Switzerland (2017)
9. Nguyen, H. T., Kosheleva, O., Kreinovich, V.: Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction. International Journal of Intelligent Systems **24**(1), 27–47 (2009)
10. Nguyen, H. T., Kreinovich, V., Wojciechowski, P.: Strict Archimedean t-norms and t-conorms as universal approximators. International Journal of Approximate Reasoning **18**(3–4), 239–249 (1998)
11. Nguyen, H. T., Kreinovich, V., Wu, B., Xiang, G.: Computing statistics under interval and fuzzy uncertainty. Springer Verlag, Berlin, Heidelberg (2012)
12. Nguyen, H. T., Walker, C. L., Walker, E. A.: A first course in fuzzy logic. Chapman and Hall/CRC, Boca Raton, Florida (2019)
13. Novák, V., Perfilieva, I., Močkoř, J.: Mathematical principles of fuzzy logic. Kluwer, Boston, Dordrecht (1999)
14. Raiffa, H.: Decision analysis. McGraw-Hill, Columbus, Ohio (1997)
15. Zadeh, L. A.: Fuzzy sets. Information and Control **8**, 338–353 (1965)