"At Least k out of n" under Fuzzy Uncertainty: Efficient Algorithm for General "And"-Operations

Julia Liepold[1,2], Olga Kosheleva[3], Vladik Kreinovich[4], and Klaus-Peter Adlassnig[1,5]
[1]Medexter Healthcare, Borschkegasse 7/5, 1090 Vienna, Austria, jl@medexter.com
[2]Institute for Logic and Computation, Technical University of Vienna, 1040 Vienna, Austria
[3,4]Departments of [3]Teacher Education and [4]Computer Science
University of Texas at El Paso,
500 W. University, El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu
[5]Medical University of Vienna, Center for Medical Data Science, Institute of Artificial Intelligence, Spitalgasse 23, 1090 Vienna, Austria,
klaus-peter.adlassnig@meduniwien.ac.at

1. “At least k out of n” is practically useful

- In medicine, often a diagnosis is based on the condition that a certain number of symptoms are present.
- Usually, it is as least k symptoms out of n possible symptoms, for some k.
- Let us give two examples.
- “A patient is classified as high risk for septic shock if they show at least four out of the following signs:
  - hypotension, tachycardia, fever, leukocytosis,
  - altered mental status, elevated serum lactate levels.”

2. "At least k out of n" is practically useful (cont-d)

- "A patient is classified as low risk for septic shock if they show at most two out of the following signs:
  - hypotension, tachycardia, fever, leukocytosis,
  - altered mental status, elevated serum lactate levels."
- This means that when there are at least three signs, the risk is medium or high.

3. Need to take fuzzy-type uncertainty into account

- In principle, one can use "yes"-"no" (crisp) definitions of the symptoms.
- For example, we can say that a body temperature of 38 C or higher is fever, while anything smaller than 38 is not a fever.
- However, from the commonsense viewpoint, it does not make sense to call 38.0 a fever and 37.9 not a fever.
- The difference between the two values is close to the accuracy of the thermometer.

4. Need to take fuzzy-type uncertainty into account (cont-d)

- From this viewpoint, it makes more sense to talk about to what degree the patient has a fever:
  - temperature much smaller than 38 means that this is definitely not a fever,
  - temperature much larger than 38 means that this is definitely a fever, but
  - temperature close to 38 mean that we have a fever to some degree.
- The technique to deal with such degrees is what is usually called *fuzzy logic*.
- In these terms, what we need is to take into account fuzzy uncertainty.

5. How to estimate to what extent is the at-least-k-out-of-n condition satisfied: a natural idea

- When each statement is either true or false, it is straightforward to decide when we have at least k out of n symptoms.
- However:
  - when we have symptoms satisfied to certain degrees $m_1, \ldots, m_n$,
  - the at-least-k-out-of-n condition is only satisfied to some degree.
- How can we compute this degree?
- In fuzzy techniques, a usual way to assign a degree to a complex statement is:
  - to represent this statement in terms of the basic logical connectives – "and", "or", and "not" – and then
  - to use fuzzy analogs of these connectives.

6. How to estimate to what extent is the at-least-k-out-of-n condition satisfied: a natural idea (cont-d)

- These analogs include:
  - "and"-operations $f_\&(a,b)$ (also known as t-norms),
  - "or"-operations $f_\vee(a,b)$ (also known as t-conorms), and
  - negation operations $f_\neg(a)$.
- For example, "at least k out of n" means that the set of symptoms can be any set S with at least k elements.
- For each of these sets, the above approach leads to the formula

$$m(S)=f_\&(m_{i_1},\ldots,n_{i_\ell},f_\neg(m_{j_1}),\ldots,f_\neg(m_{j_p})).$$

- Here $i_1,\ldots,i_\ell$ are all elements of the set S, while j_1, …, j_p are all symptoms that do not belong to the set S.

7. How to estimate to what extent is the at-least-k-out-of-n condition satisfied: a natural idea (cont-d)

- Then, the degree m to which the condition is satisfied:
  - can be computed as $f_\vee(m(S_1), m(S_2), \ldots)$,
  - where S_1, S_2, etc. are all the subsets of the set $\{1, 2, \ldots, n\}$ that have at least k elements.

8. Limitations of the above natural idea

- In the above natural-idea approach:
  - to find the desired degree m,
  - we need to compute $m(S)$ for a large number of sets – close to 2^n.
- For large n, this number becomes astronomical.
- Because of this limitation, it is desirable to come up with an efficient way to define and compute the desired degree.

9. What is known and remaining problem

- We want our degree to reflect the practice of the corresponding discipline – e.g., of the corresponding branch of medicine.
- In general, reasoning in different branch of knowledge is best described by different "and"-operations.
- This was first discovered when the first expert systems appeared.
- Historically the first expert system was the system MYCIN that was focused on a certain class of blood diseases.
- Designers of this system spent a lot of time and efforts:
  - searching for the "and"-operation
  - that provides the most accurate description of the reasoning of the medical doctors dealing with these diseases.
- At first, they were under the impression that they found universal laws of human reasoning.

10. What is known and remaining problem (cont-d)

- However:
  - when they tried to apply the same "and"-operation to a different application area – geophysics,
  - they found out that their degrees for "and"-statements were very different from what geophysicists',
  - and thus, that a different "and"-operation is needed for geophysical applications.
- This difference in "and"-operations make perfect sense.
- In medicine, one needs to be very cautious, and to prescribe some cure only if we are reasonable sure that it will help.
- Medical mistakes can be deadly.

11. What is known and remaining problem (cont-d)

- In contrast, in mining applications of geophysics:
  - if a company waits too long for a perfect conformation that there is oil in a field,
  - it may lose to competitors.
- Too much caution can ruin a company.
- An efficient algorithm for estimating the desired degree m is known for the case when the "and"-operation is $f_\&(a,b)=\max(a+b-1,0)$.
- We want to take care of all possible applications where the "at least k out of n" idea is used.
- So, we need to extend the above-mentioned feasible algorithm for a *specific* "and"-operation to the case of *general* "and"-operations.
- In this talk, we solve this problem by providing an efficient algorithm for general "and"-operations.

12. How can we describe a general "and"-operation: a brief reminder

- It is known that any continuous "and"-operation can be approximated:
  - with any given accuracy $\varepsilon > 0$,
  - by a *strictly Archimedean* "and"-operation, i.e., operation of the type $f_{\&}(a, b) = f^{-1}(f(a) \cdot f(b))$.
- Here, $f(a)$ is a continuous strictly increasing function $f(a)$.
- And $f^{-1}(a)$ means the inverse function, i.e., the function for which $f^{-1}(b) = a$ if and only if $f(a) = b$.
- Based on finitely many experiments with real experts, we can only determine the "and"-operation with some accuracy.

13. How can we describe a general "and"-operation: a brief reminder (cont-d)

- This means that we can have a strictly Archimedean "and"-operation that is in perfect accordance with the experimental data.
- Thus, to describe actual expert reasoning, we can always use strictly Archimedean "and"-operations.

14. Our first idea

- The use of strictly Archimedean "and"-operation means, in effect, that:
  - if we re-scale degrees, i.e., replace each degree m_i with the degree $m'_i = f(m_i)$,
  - then we can simply use multiplication as "and", and
  - at the end of computations, we need to transform the resulting degree m' back to the original scale by taking $m = f^{-1}(m')$.
- So this is our idea of how to estimate the desired degree m under a general strictly Archimedean "and"-operation:
  - first, we transform the original degrees m_i into re-scaled ones $m'_i = f(m_i)$,
  - then, we use the values m'_i to perform multiplication-based estimation of the degree m, and
  - after that, compute the desired estimate m as $f^{-1}(m)$.

15. Our second idea: how to perform multiplication-based computations

- We have reduced the problem for a general "and"-operation to its specific case, when the "and"-operation is simply multiplication.
- So, to solve the general problem, it is sufficient to solve it for the case of the product "and"-operation $f_{\&}(a,b) = a \cdot b$.
- For this operation, the "and"-formula is similar to the probabilistic case, where:
  - for two independent events,
  - the probability that both will occur is equal to the product $a \cdot b$ of the probabilities a and b of each of the two events.
- Probability theory exists for many centuries, many algorithms have been designed for it.

16. Our second idea: how to perform multiplication-based computations (cont-d)

- So, for the multiplication "and"-operation, a natural way to provide an estimate is:
  - to view the values m'_i as probabilities and
  - to estimate the probability that at least k out of n symptoms are satisfied.
- To use this idea, we need to come up with the efficient algorithm for computing this probability.
- Of course, it is important to emphasize that the use of probabilities here is simply a mathematical trick.
- It does not mean that expert-produced degrees m_i somehow became probabilities.

17. Numerical example

- Let us illustrate our approach on a simple numerical example, in which $f_\&(a,b)=a\cdot b$. In this case, $f(m)=m$.
- Let us consider the case when $n=4$, $k=3$, $m_1=m_2=0.5$, and $m_3=m_4=0.6$.
- In this case, to find the desired degree to which at least 3 out of 4 symptoms are satisfied, we need to consider the following cases:
- We need to consider cases when exactly 3 symptoms are present:

$$m(\{1,2,3,4\}-\{1\})=m(\{2,3,4\})=(1-0.5)\cdot 0.5\cdot 0.6\cdot 0.6=$$
$$0.5\cdot 0.5\cdot 0.6\cdot 0.6=0.09;$$
$$m(\{1,2,3,4\}-\{2\})=m(\{1,3,4\})=0.5\cdot(1-0.5)\cdot 0.6\cdot 0.6=$$
$$0.5\cdot 0.5\cdot 0.6\cdot 0.6=0.09;$$
$$m(\{1,2,3,4\}-\{3\})=m(\{1,2,4\})=0.5\cdot 0.5\cdot(1-0.6)\cdot 0.6=$$
$$0.5\cdot 0.5\cdot 0.4\cdot 0.6=0.06.$$

18. Numerical example (cont-d)

- One more such case:

$$m(\{1,2,3,4\} - \{4\}) = m(\{1,2,3\}) = 0.5 \cdot 0.5 \cdot 0.6 \cdot (1 - 0.6) =$$

$$0.5 \cdot 0.5 \cdot 0.6 \cdot 0.4 = 0.06;$$

- We also need the case when all four symptoms are present:

$$m(\{1,2,3,4\}) = 0.5 \cdot 0.5 \cdot 0.6 \cdot 0.6 = 0.09.$$

- Then, the desired degree m' that we have at least 3 symptoms is $m' =$

  $m(\{2,3,4\}) + m(\{1,3,4\}) + m(\{1,2,3\}) + m(\{1,2,3\}) + m(\{1,2,3,4\}) =$

$$0.09 + 0.09 + 0.06 + 0.06 + 0.09 = 0.39.$$

19. Comment

- For small n, we could simply enumerate all possible sets of desired size.
- However, for larger n, as we have mentioned earlier:
  - this will be very computationally intensive – and
  - even, for large n, infeasible.
- So, we need to come up with a more efficient algorithm.

20. How to effectively compute the desired probability: towards an algorithm

- The probability m' that at least k out of n events occur can be described as $m' = 1 - p_0 - p_1 - p_2 - \ldots - p_{k-1}$.
- Here p_j is the probability that exactly j events happened independently.
- So, to compute m', it is sufficient to be able to estimate the values p_j.
- One can easily check that we have

$$p_j = \sum_{S:|S|=j} \left(\prod_{i \in S} m'_i \cdot \prod_{i \notin S} (1 - m'_i) \right).$$

- Here $|S|$ denotes the number of elements in the set S.

21. How to effectively compute the desired probability: towards an algorithm (cont-d)

- One can check that such sums appear as coefficients at z^j when we compute the product $P(z)$ of all n terms $(1-m'_i)+z\cdot m'_i$:

$$P(z)\stackrel{\text{def}}{=}\prod_{i=1}^{n}((1-m'_i)+z\cdot m'_i)=\sum_j p_j\cdot z^j.$$

- Indeed, let us use distributivity and represent the product of the sums as the sum of all possible products of terms m'_i and $1-m'_i$.
- We can see that the only such products that lead to a coefficient at z^j are the products that have:
  - exactly j terms m'_i and,
  - correspondingly, $n-j$ terms of the type $1-m'_i$.
- By using the above formula, we can compute the values of the polynomial $P(z)$ for different inputs z.

22. How to effectively compute the desired probability: towards an algorithm (cont-d)

- So, to find the values p_j, we need to find the coefficients of a polynomial based on its values.
- This is a known computational problem with a known solution.
- Namely, we compute $P(z)$ for the values

$$z_k = \exp\left(\mathrm{i}\cdot\frac{2\pi\cdot k}{n}\right), k = 0, \ldots, n-1, \text{ where } \mathrm{i} \stackrel{\text{def}}{=} \sqrt{-1}.$$

- The values $P(z_k)$ form the Fourier transform of the sequence $(p_0, \ldots, p_{n-1})$.
- Thus, to compute the values p_j, we can apply the inverse Fourier transform to the sequence $(P(z_0), \ldots, P(z_{n-1}))$.

23. How to effectively compute the desired probability: towards an algorithm (cont-d)

- There is a known efficient algorithm for performing the inverse Fourier transform.
- It is known as inverse Fast Fourier Transform, or iFFT, for short.
- Thus, we arrive at the following algorithm for solving our problem.

24. Resulting algorithm

- We know the degrees $m_1, \ldots, m_n$ of each of n symptoms.
- We also know the function $f(x)$ for which:
  - the "and"-operation $f^{-1}(f(a) \cdot f(b))$
  - best describes the reasoning of people from this particular application area.
- We want to estimate the degree to which, based on this information, at least k symptoms are present.
- So first, we compute the values $m'_i = f(m_i)$ for $i = 1, \ldots, n$.
- Then, for $k = 0, \ldots, n-1$, we compute the values $P(z_k)$ as described earlier.
- After that, we apply the inverse Fast Fourier Transform algorithm to the resulting sequence of values $(P(z_0), \ldots, P(z_{n-1}))$.
- This produces the values $p_0, \ldots, p_{n-1}$.

25. Resulting algorithm (cont-d)

- Then, we compute the value m' by using the above formula.
- Finally, we compute the desired value $m = f^{-1}(m')$.

26. What is the computational complexity of this algorithm

- Computing each of n terms $(1-m'_i)+z\cdot m'_i$ requires one multiplication, one subtraction, and one addition.
- So, overall we need 3 computational steps, if we count each arithmetic operation as one step.
- To compute each product $P(z_k)$, we need:
  - to compute n such terms –that will take $3n$ steps,
  - and then to compute their product – which requires n more steps.
- So, computing each value $P(z_k)$ takes $4n$ computational steps.
- In our algorithm, we need n values $P(z_k)$ corresponding to $k = 0, \ldots, n-1$.
- Thus, the overall computation of all these values takes time $n \cdot 4n = 4n^2$, which is $O(n^2)$.
- Inverse Fast Fourier Transform takes $O(n \cdot \log(n))$ steps.

27. What is the computational complexity of this algorithm (cont-d)

- Applying the formula for m' takes $k \le n$ steps, i.e., $O(n)$.
- Thus, overall, we need

$$O(n^2) + O(n \cdot \log(n)) + O(n) = O(n^2) \text{ computational steps.}$$

- So, our algorithm requires quadratic time, which is very feasible – and definitely much faster than computing 2^n terms.

28. Clinical testing

- We are currently testing our technique on numerical values coming from actual clinical practice.
- Our preliminary results show that the results of applying our approach are in good accordance with the opinion of medical doctors.
- In particular:
  - preliminary results of applying our approach to the first example – of checking whether there is a high risk of septic shock
  - are described in a to-appear medical paper.

29. Acknowledgments

This work was supported in part:

- by the US National Science Foundation grants:
  - 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science),
  - HRD-1834620 and HRD-2034030 (CAHSI Includes),
  - EAR-2225395 (Center for Collective Impact in Earthquake Science C-CIES),
- by the AT&T Fellowship in Information Technology, and
- by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Focus Program SPP 100+ 2388, Grant Nr. 501624329,