The Thirty-Third International
FLAIRS Conference (FLAIRS-33)

From Association to Reasoning, an Alternative to Pearls' Causal Reasoning

Usef Faghihi, [1] **Serge Robert,** [2] **Pierre Poirier,** [2] **Youssef Barkaoui**[1]
University of Québec at Trois-Rivière, [1] QC, Canada, University of Québec at Montréal, [2] QC, Canada
usef.faghihi@uqtr.ca, [1] robert.serge@uqam.ca, [2] poirier.pierre@uqam.ca, [2] youssef.barkaoui@uqtr.ca[1]

Abstract

Computer scientists use causal inference for reasoning. In causal inference, researchers are interested in finding the relationship between two observable events. In this paper, we will explore the first step towards finding causality using probabilistic fuzzy logic (PFL). We will also show that PFL is more precise than Pearl's causality model.

Introduction

Benjio, Pearl and Faghihi (Bengio, 2018; Faghihi, Fournier-Viger, & Nkambou, 2011; Pearl & Mackenzie, 2018) suggest that machines will only be intelligent once they can reason. Causal reasoning is the use of knowledge to explain what is already observed in order to predict the future. Pearl (Pearl & Mackenzie, 2018), for example, suggests the use of inferential logic, with 3 levels of causal hierarchy: 1) Association, to identify interrelated phenomena; example: what if we do X?, 2) Intervention, to predict the consequence(s) of an action; example: How does the duration of my planned life change if I became a vegetarian? 3) Counterfactual (Rubin, 1974, 1977, 1978, 1980), to reason about hypothetical situations and possible outcomes; example: Would my grandfather still be alive if he had not smoked? To introduce counterfactual capability in computers, Pearl and Mackenzie (Pearl & Mackenzie, 2018) use causal inference which represents causal rules using causal diagrams. We will explain this using the Smoke-Tar-Cancer (Figure 1) example from (Pearl & Mackenzie, 2018). Suppose we would like to study the effects of Smoking on Cancer. In Figure 1, the direct causes of Cancer are shown as Tar and Smoking Gene. At this point, we have no way to observe whether a Smoking Gene does exist. Thus, Smoking is influencing Cancer through Tar and Gene is not observable, we cannot use the mere probability methods to solve this problem.

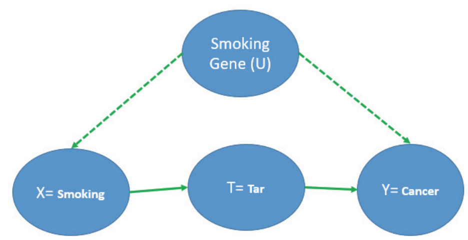

Figure 1. A causal diagram of the Smoking and Cancer example from (Pearl & Mackenzie, 2018), page 225

To do so, Pearl suggests cutting the arrow between Smoking Gene and Smoke. That is, making the Smoking (X) constant, which gives us P (y | do(x)) in Figure 2 (see above).

Furthermore, in order to solve this problem Pearl suggests the following rules:

Rule1: P (Y |do(X), Z, W) = P (Y |do(X), Z)

The probability distribution of a variable Y will not change after eliminating the variable W, if we observe that the variable W is unrelated to the variable Y.

Before explaining Rule 2, we need to explain very briefly what is a back-door criterion. In a directed acyclic graph (DAG) (i.e., Figure 1), a set of variables Z fulfills the back-door criterion if:

1) We cannot find a node in Z that is a descendent of Xi,
2) Z blocks all paths between Xi and Xj that has an arrow into Xi. Accordingly, conditional on Z, do(X) is equal to X:

Rule 2: P (Y| do(X), Z) = P (Y |X, Z)

Furthermore, suppose that X and Y are two disjoint subsets of nodes in Figure 1. Z can fulfill back-door criterion for (X, Y), if it fulfills the criterion for any pair of (X_i, X_j) where $X_i \in$ X and $X_j \in$ Y. In Figure 2, to prevent back-door effect on Smoking, the back-door arrow between Smoking Gene and Smoking is deleted.

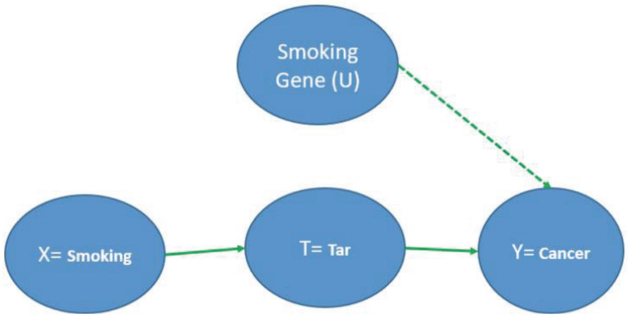

Figure 2. P(y|x) ≠ P(y|do(x), "The probability of a patient having lung cancer Y = y, given that we intervene and give the person to smoke a pack of cigarette perday (set the value of X to x) and subsequently observe what happens." (Pearl & Mackenzie, 2018)

Rule 3: P(Y |do(X))= P(Y |X)

When there is no causal path between X and Y, we can eliminate do(X) from the above formula.

Using the three rules explained briefly above, Pearl solves the problem demonstrated in Figure 1. For more details, readers are encouraged to consult (Pearl & Mackenzie, 2018).

Pearl's inferential logic uses a probabilistic approach. Probability theories are associated with events, and these events can occur or not. Furthermore, the probability deals with the uncertainty inherent to our knowledge of facts and whether they are true or false. There is no gradient possible about the reality of the facts. For instance, to find out whether smoking will cause cancer, using probability theories, we must compare people who smoke all the time or who never smoke. That is, one needs to calculate the conditional probability of having cancer given the presence or absence of smoking. However, without additional knowledge of causal structure, probability theories cannot generalize hypothetical scenarios and interventions. Furthermore, there might be hidden confounders[1]. Perhaps a gene corresponding to cancer causes a person to smoke (Figure 1). Having the gene cancer, the person is more likely to Smoke and have Cancer. The mere conditional probability cannot say whether two quantities are causally related.

Pearl's approach to causation cannot answer the following problem: given that you smoke **a little**, what is the probability that you have cancer **to a certain degree**?

Furthermore, Pearl uses Direct Acyclic Graphs (DAG) to construct causal models. However, we cannot use DAG to create real-life problems. For instance, in a network, the nodes can send and receive data mutually.

[1]" *In statistics, a confounder (also confounding variable, confounding factor, or lurking variable) is a variable that influences both the dependent variable and independent variable, causing a spurious association. Confounding is a causal concept, and as such, cannot be described in terms of correlations or associations*" Wikipedia

An alternative to what Pearl is suggesting is probabilistic fuzzy logic (Yager & Zadeh, 2012). In fuzzy logic, a set of elements always belongs to a degree of membership which fits into an interval between [0,1]. Probabilistic-fuzzy logic (PFL) processes sources of uncertainty which are: 1) randomness, or 2) probabilistic uncertainty, and fuzziness. By representing "*degrees of being*" as well as degrees of certainty (e.g., probabilities of being a little or a lot toxic, etc.), PFL can both manage the uncertainty of our knowledge and the uncertainties inherent in the world's complexity (Yager & Zadeh, 2012).

Probabilistic-Fuzzy Logic (PFL)

What we propose is to keep Pearl's association, intervention and counterfactual, but with fuzzification of each of these steps. 1) Fuzzy association: for instance, events A and B can each have a qualitative value such as "*a lot of A*", "a *little bit of B*", "*a lot of A and B at the same time*", and so on. Then, we will make quantitative fuzzifications of these qualitative values, relying on the judgment of the experts in the domain, assigning values to the variables, so that event A might have a fuzzy value of .5, or .8 and so on. Then, the fuzzy dependency between fuzzy event A and fuzzy event B will be calculated using fuzzy implication rules such as, for example, the Lukasiewicz implication rule. This way, the associations between events will be fuzzified (Yager & Zadeh, 2012). 2) Fuzzy intervention rule can be built as min (1, 1-A+B), which is Lukasiewicz implication rule and which corresponds to a fuzzy version of Pearl's do(x): in other words, what will happen to fuzzy B if we apply fuzzy C to the relation between A and B? For example, given that a lot of A produces a little bit of B, what happens to B when we add an important amount of C to A? 3) Fuzzy counterfactual rule can be calculated as a fuzzy Lukaziewicz (Yager & Zadeh, 2012) conjunction of not A and B, that is min (1-A, B), which corresponds to the occurrence of B from a cause that is different from A. That is, B can occur but without the occurrence of A, an alternative cause like D would have done the causal job required for the occurrence of B.

This "not-A and B" situation is more technically called the dual operator of the conditional *"If A, then B"*. This way, alternative causes get fuzzified by fuzzy duals in fuzzy causation. So, the system resulting from these operations will be causal chains in which every node and every segment of dependency between these nodes will be fuzzified (Yager & Zadeh, 2012). After the establishment of this fuzzy system, it can be made probabilistic using Bayes rule, as Pearl does in (Pearl & Mackenzie, 2018). Applying Bayes rule to each of our fuzzy conditional relations will make a system more flexible than Pearl's. It will result in a Probability Fuzzy Logic (PFL) system in which any new event can be taken into account and will provide learning. Learning occurs by modifications to any of the constituents of the system. When the system converges to satisfying conclusions on a given case, it will defuzzify these conclusions in order to provide crisp decisions for actions.

Probabilistic Fuzzified Version of the Smoke-Cancer Problem

Roughly speaking, to solve a problem with fuzzy logic one can apply three steps 1) Fuzzification, 2) Defuzzification, 3) Application of probability methods. In order to create our model for the Smoke-Cancer problem in Figure 1, we use fuzzy logic rules explained in the previous section. The data for the following discussion comes from (Doll & Hill, 1950).

1) Fuzzification

From Figure 1, we have a fuzzification model for Smoking, Smoking Gene and Tar. The fuzzification is the process of mapping a given value by a human to a value between [0,1]. That is, every fuzzy value belongs to a membership interval. For instance, in Figure 3, there are three colored Gaussian functions in a rectangle representing the mapping of the smoking scale variable on an interval between [0,1] on the y-axis. The blue line represents the '*little smoking*' membership function. The smoking scale for *little smoking* on x-axis is between [7.5,0] which is mapped to [0,1] interval on the y-axis. The orange line represents '*often smoking*' with the smoking scale on the x-axis [2.5, 10, 17.5] which is mapped to [0,1,0] interval on the y-axis. The green line represents '*too much smoking'*. Its smoking scale on the x axis is between [12.5, 20] which is mapped to the [0,1] interval on the y axis. The same steps apply for the fuzzification of Tar and Cancer nodes. For instance, if someone has Tar with value 4, in Fuzzy logic, the Tar value can be *little* and *average* at the same time. This can be demonstrated by mapping value 4 to y-axis which gives us a minimum value of 0.38 (*little*), and a maximum value of 0.6 (*often*). So, fuzzy values can be assigned as an interval in which there is a minimum value and a maximum value. This gives us great flexibility in dealing with unknown situations. As can be noticed, FL embeds the causal diagrams that Pearl discusses in (Pearl & Mackenzie, 2018). That is, Smoking Gene values can influence directly or indirectly the fuzzy values of the Smoking or Cancer nodes or both nodes. The dual rules of PFL will result in assigning minimum value to the Smoking Gene node. Thus, this implies that it does not influence any node in Figure 1.

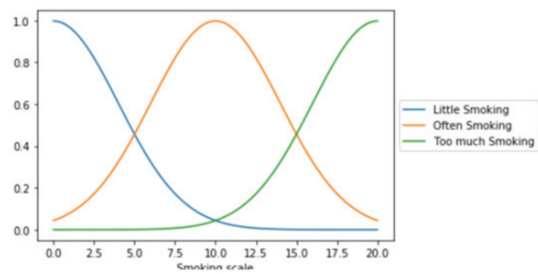

Figure 3. The fuzzification of smoking in the smoking-cancer example

averge smoked cigaretes	No patients	cancer	percentage
0	57	0	0.138350
0	32	1	0.077670
3	29	0	0.070388
3	19	1	0.046117
10	62	0	0.150485
10	71	1	0.172330
20	52	0	0.126214
20	90	1	0.218447

P(c=1|smoking=little)= 0.3958333333333333
P(c=1|smoking=average)= 0.5338345864661654
P(c=1|smoking=alot)= 0.6338028169014085

Figure 4. Doll, R. & Hill, A.B. (1950)

2) Defuzzification

The defuzzification function, takes as input a scalar value between [0,1] (from the y-axis) and its corresponding fuzzy set which, is one of the three following vectors: [1,0,0] which corresponds to '*a little cancer'*, [0,1,0] which corresponds to '*medium cancer'* and [0,0,1] which corresponds to '*much cancer'*. Using the centroid and the first maxima methods (Yager & Zadeh, 2012), FL computes the defuzzified value corresponding to each node in Smoke-Cancer example. For instance, a scalar value 0.4 on the y-axis, corresponds to the '*little cancer'*, [1,0,0] *vector* on the x-axis. According to (Doll & Hill, 1950), the centroid method will output number 19.25. In order to eliminate the back-door effect of the Smoking Gene on Smoking, our model uses **fuzzy dual (counterfactual)** rule min (1-A, B). This is because the Smoking Gene is the confounder for the nodes Smoking and Cancer in Figure 1. In order to calculate the influence of Smoking on Tar our model has a *cause_function* that takes a scalar value between [0, 20] (Figure 4) for Smoking and Tar and applies the fuzzy dual rule μ(x,y)= ($\bar{x}$,y)=(1-x,y) (explained in the Probabilistic-fuzzy logic section). In addition to Smoking Gene, we also considered pollution as another hidden element that can influence the risk of having cancer in a patient. To do so, we used the fuzzy dual rule min (1-A, B). This is implemented as the *reverse_function* in our code (see [2]) which approximates the causal effect of external unobserved variables such as pollution on cancer. For instance, if the

[2] https://github.com/joseffaghihi/FUZZY-Causal-Models

current value of Cancer in Figure 1, is 80 with the Smoking and Tar values of 8 and 6 respectively, the estimated effects of the Smoking Gene would be 6.6 and 8.4 for the pollution factor respectively.

Results

In this section, we used the dataset that Pearl discussed in (Doll & Hill, 1950) which contains the observation (Figure 4) of the values for the Smoking-Cancer problem in Figure 1. Figure 4, has four columns: 1) shows the observation of the average number of cigarettes the patients smoked per day, 2) the number of patients, 3) whether the patients have cancer and, 4) the results of the diagnosis.

It must be noted that in Figure 4, smoking *a lot* starts from 0.6338. As opposed to Pearl's assumption in (Pearl & Mackenzie, 2018) that Tar can be observed, in this experiment the Tar values are missing. Fuzzy logic has a solution for the missing data. We used the Gödel t-conorm rule which is μ(x, y) = max(x, y) to model the causal effect of the smoking on Tar. In Table 1, the values in the Tar column correspond to the average Smoked Cigarettes column.

Avg Smoked Cigaretes	Tar
0	0
0	0
3	0.39
3	0.39
10	0.8
10	0.8
20	1.0
20	1.0

Table 1. Tar value estimation using Gödel t-conorm rule

We then used Fuzzy association rules μ(x, y) = (x, y) = (1-x, y) explained in the previous section to compute the influence of Smoke on Tar. To compute the influence of unobserved nodes such as Smoking Gene or pollution on Smoking and Cancer in Figure 4, our model used the fuzzy dual (**counterfactual)** rule which is min (1-A, B).

To create the probabilistic fuzzy logic model from data in Figure 4, we applied probabilistic methods to the fuzzy logic model that we have created in the previous section. For instance, to compute P(cancer = c | smoking = s) in Figure 4:

P(cancer=c|smoking=s)= (P(smoking∩cancer)/P(smoking))

For instance to compute P(c=1| smoking= little) =19/(19+29) = 0.3958.

Unfortunately, this database has only 400 rows. Thus, we could not use part of it as a test for predictions.

Conclusion

In this paper, we showed the first step toward using probabilistic fuzzy logic (PFL) as an alternative to Pearl's inferential logic. To construct the causal model, we can use the Fuzzy logic (FL) fuzzification method. After the fuzzification step, one can apply the conditional probability method to fuzzy logic to make it a probabilistic fuzzy (PFL) model. Using PFL, we conclude that this method is more accurate than Pearl's one when it comes to finding possible causes.
Furthermore, Fuzzy logic deals with intervals. Calculating the probability of intervals for both outcomes and treatments is completely different from the simple probability approach. Our next step is to implement more complex causal models Pearl discussed in his book. Furthermore, we are aiming at applying our PFL causal model to other domains.

References

Bengio, Y. (2018). One of the fathers of ai is worried about its future. Artificial Intelligence / Machine Learning.

Cheng, P.-C., Rohatgi, P., Keser, C., Karger, P. A., Wagner, G. M., & Reninger, A. S. (2007). Fuzzy multi-level security: An experiment on quantified risk-adaptive access control. Paper presented at the 2007 IEEE Symposium on Security and Privacy (SP'07).

Doll, R., & Hill, A. B. (1950). Smoking and carcinoma of the lung. British medical journal, 2(4682), 739.

Faghihi, U., Fournier-Viger, P., & Nkambou, R. (2011). Implementing an efficient causal learning mechanism in a cognitive tutoring agent. Paper presented at the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems.

Pearl, J., & Mackenzie, D. (2018). The book of why: The new science of cause and effect: Basic Books.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of educational Psychology, 66(5), 688.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. Journal of educational Statistics, 2(1), 1-26.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. The Annals of statistics, 34-58.

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. Journal of the American Statistical Association, 75(371), 591-593.

Wang, L.-X., & Mendel, J. M. (1992). Generating fuzzy rules by learning from examples. IEEE Transactions on systems, man, and cybernetics, 22(6), 1414-1427.

Yager, R. R., & Zadeh, L. A. (2012). An introduction to fuzzy logic applications in intelligent systems (Vol. 165): Springer Science & Business Media.