7-2016

Why Ragin's Fuzzy Techniques Lead to Successful Social Science Applications: An Explanation

Olga Kosheleva
The University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Comments:
Technical Report: UTEP-CS-16-52

Why Ragin's Fuzzy Techniques Lead to Successful Social Science Applications: An Explanation

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract

To find the relation between two concepts, social scientists traditionally look for correlations between the numerical quantities describing these concepts. Sometimes this help, but sometimes, while we are clear that there is a relation, statistical analysis does not show any correlation. Charles Ragin has shown that often, in such situations, we *can* find statistically significant correlation between the *degrees* to which experts estimate the corresponding concepts to be applicable to given situations. In this paper, we provide a simple explanation for this empirical success.

1 Ragin's Approach to Social Research: A Brief Description and Need for Justification

Typical social science questions. Based on several observations and/or on some theoretical ideas, a social science researcher formulates a hypothesis that some property A imply some other property B: e.g., that rich countries are usually democracies, or that the socio-economic status of parents affects the success of their children in school.

Once the hypothesis is formulated, it needs to be checked against all available data. In some cases, as a result of this check, the hypothesis is validated. In other cases, a detailed data analysis shows that, contrary to anecdotal evidence and/or theoretical ideas, there is no causal relation between the corresponding phenomena A and B.

How to check the proposed hypothesis: a traditional correlation-based approach. Traditionally, the hypothesis is checked in the following way. First, we come up with reasonable numerical quantities a and b that describe to what extent properties A and B are satisfied.

For example, to describe how rich a country is, it is reasonable to use the average (or median) income a. To describe how democratic a country is, it makes sense to use the democracy index provided by the Economist Intelligence Unit; this index is, in effect, the sum of the weights corresponding to different aspects of democracy.

Then, researchers look for a correlation between the corresponding quantities a and b.

The traditional approach does not always work. In some cases, the data analysis shows that there is a correlation between the corresponding quantities a and b. In these cases, the original hypothesis is validated.

In some other cases, however, the correlation data is inconclusive: the data seems to show that A implies B, but the correlation analysis does not lead to a statistically significantly positive correlation coefficient.

Ragin's approach: a brief description. To check hypotheses in situations in which the traditional correlation-based approach does not work, C. C. Ragin proposed, instead of using the numerical values a and b, to use the *degrees* $\mu_A(x)$ and $\mu_B(x)$ to which the experts believe that the corresponding cases x satisfy the properties A and B; see, e.g., [7, 8, 9].

For example, instead of using the mean or average income a, Ragin proposed to use the degree $\mu_A(x)$ to which the experts believe that the country x is rich.

Then, instead of looking for a correlation between the numerical quantities a and b, we look for a correlation (or other measure of dependence) relating the degrees $\mu_A(c)$ and $\mu_B(c)$.

Important comment. It is important to emphasize that while the use of degrees is the main idea behind Ragin's approach, this approach goes beyond this main idea. For example, Ragin proposes different measures of relation beyond the usual correlation.

Ragin's approach: successes. Ragin's approach has led to many interesting results in social research; see, e.g., [7, 8, 9].

Ragin's approach: need for a justification. While Ragin's approach is empirically successful, it is semi-heuristic, many aspects of this approach lack a convincing justification.

For example, Ragin bases his approach on the main ideas of *fuzzy logic* (see, e.g., [2, 6, 12]), an approach where similar semi-heuristic degree-based approach is used to describe and exploit expert knowledge in engineering problems.

While the general degree-based idea is very reasonable, there are many different numerical implementations of this fuzzy idea, and it is not clear which of these implementations to use.

What we do in this paper. In this paper, we provide a theoretical explanation of the main Ragin's idea.

2 Justification of Ragin's Approach

Why not correlation. The main problem with correlation is that it only works when the dependence between the corresponding quantities a and b is linear; see, e.g., [11]. In social sciences, most relations are highly non-linear, and the usual correlation is not well-suited to capture such non-linear dependencies.

Our idea: degrees are, in effect, subjective probabilities. There are many possible interpretations of degrees, and many possible ways to elicit such degrees. One of the main approaches is to interpret them as *(subjective) probabilities* $\mu_A(x) = P(A|x)$ and $\mu_B(x) = P(B|x)$ that an object x satisfies properties A and B; see, e.g., [1, 3, 4, 5, 10].

We emphasize the word "subjective" to distinguish these probabilities – that describe expert opinion – from "objective" probabilities (like the probability that a fair coin falls tails) that describe the frequency with which certain events happen in the real world.

This idea indeed explains Ragin's approach. In probabilistic terms, what does it mean that A positively affects B? For example, what does it means that rich countries are usually democracies?

In probabilistic terms, it means that under the condition A, the conditional probability $P(B|A)$ of the property B is higher than the conditional probability $P(\neg B|A)$ that the property B will not be satisfied.

In terms of the corresponding conditional probabilities, if we know the probability $P(A|x)$ that a given object x satisfies the property A, then the probability that this object satisfies the property B can be computed by using the formula of total probability:

$$P(B|x) = P(B|A) \cdot P(A|x) + P(B|\neg A) \cdot P(\neg A|x).$$

Here, $P(\neg A|x) = 1 - P(A|x)$, so the above formula takes the form

$$P(B|x) = P(B|A) \cdot P(A|x) + P(B|\neg A) \cdot (1 - P(A|x)) =$$

$$P(B|\neg A) + (P(B|A) - P(B|\neg A)) \cdot P(A|x).$$

So, for the corresponding subjective probabilities $\mu_A(x) = P(A|x)$ and $\mu_B(x) = P(B|x)$, we indeed have a linear dependence:

$$\mu_B(x) = c_0 + c_1 \cdot \mu_A(x),$$

where $c_0 \stackrel{\text{def}}{=} P(B|\neg A)$ and $c_1 \stackrel{\text{def}}{=} P(A|B) - P(A|\neg B)$. Since the usual correlation is designed to capture linear dependencies, not surprisingly,

- in situations when the dependence between the quantities a and b is strongly non-linear and correlation analysis does not detect the dependence between a and b,
- the same correlation method finds the dependence between the degrees $\mu_A(x)$ and $\mu_B(x)$.

So, this idea indeed explains the empirical success of Ragin's approach.

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.

The authors are thankful to all the participants of the 2016 IEEE World Congress on Computational Intelligence (Vancouver, Canada, July 24–29, 2016) for valuable suggestions.

References

[1] P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.

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

[3] D. Luce, *Inividual Choice Behavior: A Theoretical Analysis*, Dover, New York, 2005.

[4] R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

[5] H. T. Nguyen, O. Kosheleva, and V. Kreinovich, "Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction", *International Journal of Intelligent Systems*, 2009, Vol. 24, No. 1, pp. 27–47.

[6] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.

[7] C. C. Ragin, *Fuzzy-Set Social Science*, University of Chicago Press, Chicago, Illinois, 2000.

[8] C. C. Ragin, *Redesigning Social Inquiry: Fuzzy Sets and Beyond*, University of Chicago Press, Chicago, Illinois, 2008.

[9] C. C. Ragin, "Fuzzy sets and social research", *Proceedings of the IEEE World Conference on Computational Intelligence*, Vancouver, Canada, July 24–29, 2016.

[10] H. Raiffa, *Decision Analysis*, Addison-Wesley, Reading, Massachusetts, 1970.

[11] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

[12] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.

A An Auxiliary Idea

One more idea of Ragin: one more limitation of the traditional correlation techniques. As we have mentioned, Ragin's approach goes beyond the main idea of using degrees.

One of the additional ideas is that when we say that rich countries are usually democracies, we mean exactly this, we do not necessarily mean that poor countries are usually not democracies.

On the other hand, a correlation between the corresponding quantities a and b means that $b \approx c_0 + c_1 \cdot a$ for appropriate coefficients c_0 and c_1 – which implies not only that rich countries are democracies, but also that poor countries are not democracies.

In situations when the corresponding opposite statement is actually not true, it is not surprising that the traditional correlation techniques cannot validate the original hypothesis.

How can we modify the traditional correlation techniques. In view of the above limitation, Ragin proposes to replace the traditional correlation techniques with an appropriate fuzzy approach which is free of this limitation.

However, the fact that we explained his main idea in probabilistic terms makes us think that maybe this other idea of his can also be described in probabilistic terms.

How dependence is determined now: a reminder. Suppose that we are looking for a linear or nonlinear regression describing how the value of the quantity b depends on the quantity a.

Whether we are looking of a linear or a nonlinear dependence, we usually have a family of functions $f(a, c)$ depending on some parameter tuple c. To determine the values of the parameters, we need to use the known observations (a_k, b_k), $1 \le k \le K$, in which we observe the values of both a and b.

Of course, knowing a does not uniquely determine the value b; even for the best values of the parameters c, the predicted values $f(a_k, c)$ are, in general, different from the actual observed values b_k.

The corresponding approximation error $b_k - f(a_k, c)$ is usually caused by the joint effect of many different difficult-to-take-into-account factors. It is known that, according to the Central Limit Theorem, the distribution of the joint effect of many small independent factors is close to Gaussian; see, e.g., [11]. It is therefore reasonable to assume that the approximation errors are normally distributed, with 0 mean and some standard deviation σ, i.e., with probability density

$$\rho_k = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(b_k - f(a_k, c))^2}{2\sigma^2}\right).$$

Different observations are themselves independent. Thus, for each value of the parameter c, the probability L to observe all these approximation errors is equal

to the product of the corresponding probabilities:

$$L = \prod_{k=1}^{k} \rho_k = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(b_k - f(a_k, c))^2}{2\sigma^2}\right).$$

It is reasonable to select the *most probable* values of the parameters c, i.e., the values for which the corresponding probability L is the largest possible. This is known as the *Maximum Likelihood* approach.

For the Gaussian distribution, maximizing the above expression for L is equivalent to minimizing

$$-\ln(L) = \text{const} + \sum_{k=1}^{K} \frac{(b_k - f(a_k, c))^2}{2\sigma^2},$$

which is equivalent to minimizing the sum

$$\sum_{k=1}^{K} (b_k - f(a_k, c))^2.$$

Thus, we arrive to the widely used *Least Squares* approach to finding the corresponding parameters.

How to appropriately modify the usual approach. In the usual approach, we assume that all approximation errors $b_k - f(a_k, c)$ are small. However, according to the above Ragin's idea, they are small only when the variable a satisfies an appropriate condition – e.g., in case of a relation between richness a and democracy b, when the income a corresponds to a rich country.

If form some k, the value a_k absolutely satisfies the corresponding property A, then the difference $b_k - f(a_k, c)$ should be small, and thus, it makes sense to describe it by the Gaussian (normal) distribution corresponding to some small standard deviation σ. On the other hand, when a_k absolutely does not satisfy the corresponding property, the difference $b_k - f(a_k, c)$ can be large. Of course, since this difference is still caused by many independent factors, it still makes sense to assume that this difference is normally distributed – but the corresponding standard deviation σ_0 is much larger: $\sigma_0 \gg \sigma$. In this case,

$$\rho_{0k} = \frac{1}{\sqrt{2\pi} \cdot \sigma_0} \cdot \exp\left(-\frac{(b_k - f(a_k, c))^2}{2\sigma_0^2}\right).$$

For many values a_k, we are not 100% sure that this value a_k satisfies the desired property. There is some probability $P(A|a_k)$ that it does, but with the remaining probability $1 - P(A|a_k)$ it does not. So, the actual probability of the current difference has the form

$$p_k = P(A|a_k) \cdot \rho_k + (1 - P(A|a_k)) \cdot \rho_{k0} =$$

$$\mu_A(a_k) \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(b_k - f(a_k, c))^2}{2\sigma^2}\right) +$$

$$(1-\mu_A(a_k))\cdot\frac{1}{\sqrt{2\pi}\cdot\sigma_0}\cdot\exp\left(-\frac{(b_k-f(a_k,c))^2}{2\sigma_0^2}\right).$$

In this case, we should select parameters c for which the product of these probabilities takes the largest possible values:

$$L=\prod_{k=1}^{K}\left(\mu_A(a_k)\cdot\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left(-\frac{(b_k-f(a_k,c))^2}{2\sigma^2}\right)+\right.$$

$$\left.(1-\mu_A(a_k))\cdot\frac{1}{\sqrt{2\pi}\cdot\sigma_0}\cdot\exp\left(-\frac{(b_k-f(a_k,c))^2}{2\sigma_0^2}\right)\right)\to\max_c .$$