How to Combine Probabilistic and Fuzzy Uncertainty: Theoretical Explanation of Clustering-Related Empirical Result

László Szilágyi[a,b], **Olga Kosheleva**[c], and **Vladik Kreinovich**[d]
[a]Comput. Inteligence Research Group, Sapientia Univ., Tg. Mureş, Romania, `lalo@ms.sapiemtia.ro`
[b]Physiological Controls Research, Óbuda Univ., Budapest, Hungary, `szilagyi.laszlo@uni-obuda.hu`
[c]Teacher Education, Univ. of Texas at El Paso, El Paso, TX 79968, USA, `olgak@utep.edu`
[d]Computer Science, Univ. of Texas at El Paso, El Paso, TX 79968, USA, `vladik@utep.edu`

Abstract

In contrast to crisp clustering techniques that assign each object to a class, fuzzy clustering algorithms assign, to each object and to each class, a degree to which this object belongs to this class. In the most widely used fuzzy clustering algorithm – fuzzy c-means – for each object, degrees corresponding to different classes add up to 1. From this viewpoint, these degrees act as probabilities. There exist alternative fuzzy-based clustering techniques in which, in line with the general idea of the fuzzy set, the largest of the degrees is equal to 1. In some practical situations, the probability-type fuzzy clustering works better; in other situations, the more fuzzy-type technique leads to a more adequate clustering. It is therefore desirable to combine the two techniques, so that one of them will cover the situations where the other method does not work so well. Such combination methods have indeed been proposed. An empirical comparison has shown that out of all these combined methods, the most effective one is the method in which we the use the product of probability and fuzzy degree. In this paper, we provide a theoretical explanation for this empirical result.

Keywords: fuzzy sets, clustering, fuzzy clustering, probabilistic clustering

1 Formulation of the Problem

Probability-inspired approach to fuzzy clustering. The most widely used fuzzy-based clustering techniques – fuzzy c-means – assigns, to each object i and to each class k, the degree p_{ik} to which this object belongs to the class; see, e.g., [2]. If we want to select the cluster that most probably contains the object i, we should select the cluster k for which the value p_{ik} is the largest.

For each object, these degrees add up to 1:

$$\sum_k p_{ik} = 1.$$

Because of this fact, one of the natural interpretation of this degree is that is estimates the probability that the object i is in class k – indeed, such probabilities should add up to 1, if we assume, as it is usually done, that each object actually belongs to one of the clusters.

For simplicity, in the following text, we will call such clustering methods *probabilistic*, and the vlues p_{ik} *probabilities*.

Alternative more fuzzy-type approaches to fuzzy clustering. There exist other fuzzy-based approaches to clustering for which the resulting values μ_{ik} are more in line with the general fuzzy logic ideas (see, e.g., [1, 3, 5, 6, 7, 13]): namely, for which the constraint on the corresponding values μ_{ik} has a fuzzy-type form

$$\max_k \mu_{ik} = 1;$$

see, e.g., [4].

Need to combine probabilistic and fuzzy schemes. It turns out that both probabilistic and fuzzy schemes capture some aspects of clustering that is not well captured by the other scheme:

- for some objects i, selecting the cluster with the largest value k of the probability p_{ik} leads to a more adequate clustering, while
- for some other objects i, selecting the cluster with the largest value k of the fuzzy degree μ_{ik} leads to a more adequate clustering.

It is therefore desirable to combine the two methods, so as to combine the advantages of both methods. A natural idea is to make the selection of a cluster based on some combination $f(p_{ik}, \mu_{ik})$ of the values produced by these two methods.

Which combination method works best? Several different combination functions $f(p, \mu)$ were proposed; see, e.g., [8, 10, 11, 12]. An empirical comparison of different function showed that in many situation, the most adequate clustering comes from using the product $f(p, \mu) = p \cdot \mu$; see, e.g., [10, 11, 12].

What we do in this paper. In this paper, we provide a theoretical explanation for this empirical fact. Namely, we explain why the product function works the best, by showing that the product is the only combination function that satisfied the corresponding reasonable properties.

2 What Are the Reasonable Properties of the Combination Function?

Two possible viewpoints. We have two different approaches: probabilistic and fuzzy. Depending on what approach we start with, we can view the transition to the combined technique in two different ways:

- if we start with the probabilistic approach, we can view the use of fuzzy degrees as a correction of the original probabilistic estimate p_{ik};
- alternatively, if we start with the fuzzy approach, we can view the use of probabilities as a correction to the original fuzzy estimates μ_{ik}.

It turns out that each viewpoint leads to its own reasonable requirement on the combination function.

Main idea behind these requirements. Often, in the beginning, we have very little information. So, to be on the safe side, we consider many possible classes (clusters) to which the object can belong. Later on, we often gain additional knowledge that allows us to limit the possible choices to a smaller subset of classes.

When we limit ourselves to a smaller group of clusters, this changes the corresponding probability and/or fuzzy values. A reasonable idea is to require that the corrected probability/fuzzy values should also be similarly re-scaled.

Let us show what this idea leads to for both viewpoints.

Let us first consider the probabilistic viewpoint. For each object i and for each cluster k, we had the original probability p_{ik} that the object i belongs to the cluster k. When we gain the additional knowledge that allows us to limit the set of possible clusters to a smaller set S, this means that instead of the original probabilities p_{ik}, we now consider *conditional* probabilities – p'_{ik} under the condition that k belongs to S. By definition of conditional probability (see, e.g., [9]), we have

$$p'_{ik} = \frac{p_{ik}}{p(S)},$$

where $p(S)$ is the original probability of the set S.

Thus, from the probabilistic viewpoint, restricting the set of clusters means multiplying all the probability values by some constant $c > 0$:

$$p \mapsto c \cdot p.$$

Resulting requirement. The main idea behind our requirements is that:

- if we re-scale the original probabilities p,
- then this should leads to a similar re-scaling of the corrected probabilities $f(p, \mu)$.

Thus, we arrive at the following definition.

Definition 1. *We say that a function $f(p, \mu)$ is* reasonable from the probabilistic viewpoint *if for all possible values of p, μ, and c, we have*

$$f(c \cdot p, \mu) = c \cdot f(p, \mu). \quad (1)$$

Let us now consider the fuzzy viewpoint. In order to properly analyze the situation from the fuzzy viewpoint, let us recall where the fuzzy degrees come from.

Some popular expositions of fuzzy logic provide a simplified description that we simply ask the experts to mark, on a scale from 0 to 1, the degree to which the corresponding property is satisfied. This simplified description adequately describes situations when we have natural objects with degree 1 – or at least there are objects for which this degree is as close to 1 as possible. For example:

- for "small", clearly 0 is absolutely small; for this value, all the experts will agree that this value is small;
- for "large", clearly, the larger the value, the more confident we are that this value is large – and as this value tends to infinity, the corresponding degree tends to 1; again, all the experts will agree on this.

In such situations, the sometimes-forgotten requirement – that the largest of the degrees should be 1 – is automatically satisfied.

However, there are other situations in which this requirement is *not* automatically satisfied. For example, when we consider properties like "medium", it is highly improbable that there will be a value for which all the experts will agree that this value is absolutely medium. In such situations, the determination of membership degrees goes through two stages:

- first, we ask the experts to estimate, for each possible input k, the corresponding degree $d_k \in [0,1]$;
- then, we *normalize* these degrees by dividing them by the largest, and get the new values

$$\mu_k = \frac{d_k}{\max\limits_j d_j}$$

  for which the desired condition $\max\limits_k \mu_k = 1$ is satisfied.

From this viewpoint, if we delete the class k_0 that originally had the largest degree of confidence, then we need to again re-scale, to make sure that the largest of the degrees is still 1. This re-scaling means that we multiply all the values μ_k by the same factor:

$$\mu \mapsto c \cdot \mu$$

for some constant $c > 0$.

Resulting requirement. As we have mentioned, the main idea behind our requirements is that:

- if we re-scale the original degrees μ,
- then this should leads to a similar re-scaling of the corrected degrees $f(p,\mu)$.

Thus, we arrive at the following definition.

Definition 2. *We say that a function $f(p,\mu)$ is* reasonable from the fuzzy viewpoint *if for all possible values of p, μ, and c, we have*

$$f(p, c\cdot\mu) = c \cdot f(p,\mu). \quad (2)$$

Now, we are ready to present our main result.

3 Main Result

Proposition. *For a function $f(p,\mu)$ that maps two non-negative numbers to a non-negative number, the following two conditions are equivalent to each other:*

- *the function is reasonable both from the probabilistic viewpoint and from the fuzzy viewpoint,*
- *the function $f(p,\mu)$ has the form*

$$f(p,\mu) = a \cdot p \cdot \mu$$

  for some $a > 0$.

Discussion. At first glance, it may seem that this proposition *almost* explains the above-mentioned empirical fact – that out of all possible combination functions $f(p,\mu)$, the most adequate is the product function $f(p,\mu) = p\cdot\mu$. We said "almost" since the functions described by our proposition may contain an additional positive factor a. However, from the practical viewpoint, this factor does not change anything, for the following two reasons:

- First, multiplication by a positive constant does not change the order. Thus, the cluster k with the largest value of the expression $a \cdot p_{ik} \cdot \mu_{ik}$ is exactly the same cluster for which the product $p_{ik}\cdot\mu_{ik}$ attains it largest value.
- Second, if we normalize the values of the combination function – whether by dividing by the sum, to get probabilities or by dividing by the largest value, to get fuzzy degrees – we get the same result, whether we start with the values $a\cdot p_{ik}\cdot\mu_{ik}$ or with the values $p_{ik}\cdot\mu_{ik}$.

Proof of Proposition.

1°. It is easy to check that for each $a > 0$, the function $f(p,\mu) = a\cdot p\cdot\mu$ is reasonable both from the probabilistic and from the fuzzy viewpoints.

2°. Vice versa, let us assume that the function $f(p,\mu)$ is reasonable both from the probabilistic and from the fuzzy viewpoints. Under this assumption, let us prove that this function has the desired form.

Indeed, for each value p, the fact that the function $f(p,\mu)$ is reasonable from the probabilistic viewpoint implies that

$$f(p,1) = f(p\cdot 1, 1) = p\cdot f(1,1). \quad (3)$$

Similarly, the fact that the function $f(p,\mu)$ is reasonable from the fuzzy viewpoint implies that

$$f(p,\mu) = f(p,\mu\cdot 1) = \mu\cdot f(p,1). \quad (4)$$

Substituting the expression (3) for $f(p,1)$ into the formula (4), we conclude that

$$f(p,\mu) = \mu\cdot(p\cdot f(1,1)) = f(1,1)\cdot p\cdot\mu.$$

This is exactly the desired expression, with $a = f(1,1)$.

Acknowledgments

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), HRD-1834620 and HRD-2034030 (CAHSI Includes), EAR-2225395, and by the AT&T Fellowship in Information Technology.

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, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

References

[1] R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.

[2] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: the fuzzy c-means clustering algorithm", *Computers and Geosciences*, 1984, Vol. 10,No. 2, pp. 191–203.

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

[4] R. Krishnapuram and J. Keller, "A possibilistic approach to clustering", *IEEE Transactions on Fuzzy Systems*, 1993, Vol. 1, No. 2, pp. 98–110.

[5] J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.

[6] H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.

[7] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.

[8] N. Pal, K. Pal, J. Keller, and J. Bezdek, "A possibilistic fuzzy c-means clustering algorithm", *IEEE Transactions on Fuzzy Systems*, 2005, Vol. 13, No. 4, pp. 517–530.

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

[10] L. Szilágyi, "Fuzzy-probabilistic product partition: a novel robust approach to c-means clustering", In: V. Torra, Y. Narakawa, J. Yin, and J. Long (eds.), *Modeling Decisions for Artifical Intelligence MDAI 2011*, Springer Lecture Notes in Computer Science, Vol. 6829, Berlin, Heidelberg, 2011, pp. 150–161.

[11] L. Szilágyi, "Robust shperical shell clustering using fuzzy-probabilistic product partition", *International Journal of Intelligent Systems*, 2013, Vol. 28, pp. 524–539.

[12] L. Szilágyi, Z. R. Varga, and S. M. Szilágyi, "Application of the fuzzy-possibilistic product partition in elliptic shell clustering", In: T. Vicenç et al. (eds.), *Modeling Decisions for Artifical Intelligence MDAI 2014*, Springer Lecture Notes in Artifical Intelligence, Vol. 8825, Berlin, Heidelberg, 2011, pp. 158–169.

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