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

László Szilágyi\textsuperscript{a,b}, Olga Kosheleva\textsuperscript{c}, and Vladik Kreinovich\textsuperscript{d}

\textsuperscript{a}Computational Intelligence Research Group, Sapientia University Tg. Mureș, Romania, lalo@ms.sapientia.ro
\textsuperscript{b}Physiological Controls Research, Óbuda University Budapest, Hungary, szilagyi.laszlo@uni-obuda.hu
\textsuperscript{c,d}Departments of \textsuperscript{b}Teacher Education and \textsuperscript{c}Computer Science University of Texas at El Paso, El Paso, Texas 79968, USA olgak@utep.edu, vladik@utep.edu
1. Probability-inspired approach to fuzzy clustering

- The most widely used fuzzy-based clustering technique is fuzzy c-means.

- It assigns, to each object $i$ and to each class $k$, the degree $p_{ik}$ to which this object belongs to the class;

  - 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 probability that $i$ is in $k$.

- 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.
2. Alternative more fuzzy-type approaches to fuzzy clustering

- There exist other fuzzy-based approaches to clustering, where the constraint on the corresponding values $\mu_{ik}$ has a fuzzy-type form

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

- 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.
- For some other objects $i$, selecting the cluster with the largest value $k$ of the fuzzy degree $\mu_{ik}$ 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.
3. **Empirical result**

- Several different combination functions $f(p, \mu)$ were proposed.
- An empirical comparison of different function showed that in many situation, the most adequate clustering comes from using the product $f(p, \mu) = p \cdot \mu$.
- In this talk, we provide a theoretical explanation for this empirical fact.
4. What Are the Reasonable Properties of the Combination Function?

- 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 $p_{ik}$.
- If we start with the fuzzy approach, we can view the use of probabilities as a correction to the original fuzzy estimates $\mu_{ik}$.
- It turns out that each viewpoint leads to its own reasonable requirement on the combination function.
- 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.
5. What Are the Reasonable Properties of the Combination Function (cont-d)

- 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.
6. Let us first consider the probabilistic viewpoint

- For each object $i$ and for each cluster $k$, we had the original probability $p_{ik}$ that $i$ belongs to $k$.

- Additional knowledge may allows us to limit the set of possible clusters to a smaller set $S$; then:
  
  - 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, $p'_{ik} = \frac{p_{ik}}{p(S)}$, 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 $c > 0$:

  \[ p \mapsto c \cdot p. \]
7. Probabilistic viewpoint (cont-d)

- 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 $f(p, \mu)$.

- We say that a function $f(p, \mu)$ is reasonable from the probabilistic viewpoint if for all possible values of $p$, $\mu$, and $c$, we have
  \[
f(c \cdot p, \mu) = c \cdot f(p, \mu).
  \]
8. Let us now consider the fuzzy viewpoint

- Let us recall where the fuzzy degrees come from:
  - 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.

- Thus, we get the new values $\mu_k = \frac{d_k}{\max_j d_j}$.

- So, 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 $\mu_k$ by the same factor: $\mu \mapsto c \cdot \mu$ for some constant $c > 0$. 
9. Fuzzy viewpoint (cont-d)

- The main idea behind our requirements is that:
  - if we re-scale the original degrees $\mu$,
  - then this should lead to a similar re-scaling of the corrected degrees $f(p, \mu)$.

- We say that a function $f(p, \mu)$ is reasonable from the fuzzy viewpoint if for all possible values of $p$, $\mu$, and $c$, $f(p, c \cdot \mu) = c \cdot f(p, \mu)$. 
10. Main Result

- For a function $f(p, \mu)$ that maps two non-negative numbers to a non-negative number, the following two conditions are equivalent:
  - 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 $a > 0$.

- From the practical viewpoint, the factor $a$ is irrelevant.

- Indeed, 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 its largest value.

- Also, if we normalize the values of the combination function:
  - we get the same result, whether we start with the values $a \cdot p_{ik} \cdot \mu_{ik}$
  - or with the values $p_{ik} \cdot \mu_{ik}$.
11. Proof of Proposition

- 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.

- Vice versa, let us assume that the function $f(p, \mu)$ is reasonable both from viewpoints.

- 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

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

- Similarly, the fact that the function $f(p, \mu)$ is reasonable from the fuzzy viewpoint implies that $f(p, \mu) = f(p, \mu \cdot 1) = \mu \cdot f(p, 1)$.

- Substituting the expression for $f(p, 1)$ into the formula for $f(p, \mu)$, 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 $a = f(1, 1)$. 
12. Acknowledgments

This work was supported in part by:

- National Science Foundation grants 1623190, HRD-1834620, HRD-2034030, and EAR-2225395;
- AT&T Fellowship in Information Technology;
- program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
- a grant from the Hungarian National Research, Development and Innovation Office (NRDI).