Towards an (Even More) Natural Probabilistic Interpretation of Fuzzy Transforms (and of Fuzzy Modeling)

Irina Perfilieva[1] and Vladik Kreinovich[2]
[1]Institute for Research and Applications of Fuzzy Modeling
University of Ostrava
701 03, Ostrava Czech Republic, Irina.Perfilieva@osu.cz
[2]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA, vladikutep.edu

Abstract

In many practical applications, it turns out to be useful to use the notion of *fuzzy transform*: once we have functions $A_1(x) \geq 0$, ..., $A_n \geq 0$, with $\sum_{i=1}^{n} A_i(x) = 1$, we can then represent each function $f(x)$ by the coefficients $F_i = \dfrac{\int f(x) \cdot A_i(x)\, dx}{\int A_i(x)\, dx}$. Once we know the coefficients F_i, we can (approximately) reconstruct the original function $f(x)$ as $\sum_{i=1}^{n} F_i \cdot A_i(x)$. The original motivation for this transformation came from fuzzy modeling, but the transformation itself is a purely mathematical transformation. Thus, the empirical successes of this transformation suggest that this transformation can be also interpreted in more traditional (non-fuzzy) mathematics as well.

Such an interpretation is presented in this paper. Specifically, we show that the 2002 probabilistic interpretation of fuzzy modeling by Sánchez et al. can be modified into a natural probabilistic explanation of fuzzy transform formulas.

Keywords: F-transform, fuzzy modeling, probabilistic interpretation, fuzzy control

1 Introduction: Fuzzy Transform and the Need for Its Probabilistic Interpretation

Fuzzy transform: a definition. The notion of a fuzzy transform (*F-transform*, for short) turned out to be very useful in many application areas

such as image compression, solving differential equations under initial uncertainty, etc.; see, e.g., [12, 13] and references therein.

Generally speaking, the F-transform of function f is a vector with weighted local mean values of f as components. The first step in the definition of the F-transform of $f : X \longrightarrow \mathbb{R}$ is a selection of a *fuzzy partition* of universal set X (e.g, a bounded interval $[a, b]$ on $\mathbb{R}$) by a finite set of *basic functions*

$$A_1(x) \geq 0, \dots, A_n(x) \geq 0,$$

which are continuous and satisfy the condition: $\sum_{i=1}^{n} A_i(x) = 1$. Basic functions are called *membership functions* of respective fuzzy sets, or, alternatively, *granules*, information pieces, etc. Their choice reflects the type of uncertainty which is related to the knowledge of x.

Once the basic functions are selected, we define the F-transform of a continuous function $f : X \longrightarrow \mathbb{R}$ as a vector $(F_1, \dots, F_n)$ where

$$F_i \stackrel{\text{def}}{=} \frac{\int f(x) \cdot A_i(x)\, dx}{\int A_i(x)\, dx}. \tag{1}$$

F-transform satisfies the following properties [12, 13]:

- $y = F_i$ minimizes $\int_a^b (f(x) - y)^2 A_i(x) dx$,

- for a twice continuously differentiable function f, $F_i = f(x_i) + O(h_i^2)$, where h_i is the length of the support of A_i.

F-transform is used in applications as a "skeleton model" of f. This model provides a compressed image if f is an image [3], values of a trend if f is a time series [14], a numeric model if f is used in numeric computations (integration, differentiation) [15], etc.

Once we know the F-transform components F_i, we can (approximately) reconstruct the original function f as

$$\overline{f}(x) = \sum_{i=1}^{n} F_i \cdot A_i(x). \tag{2}$$

In [12], the formula (2) is called the *F-transform inversion formula*. The formula (2) represents a continuous function that approximates f. Under certain reasonable conditions, a sequence of functions represented by (2) uniformly converges to f (see [12] for more details).

Example. Let us give an example of the F-transform of x^2 on the domain $[0, 1]$ with respect to $A_1, \dots, A_5$. For simplicity, we assume that basic functions $A_1, \dots, A_5$ are of triangular shape and constitute a uniform fuzzy partition of

$[0, 1]$. Their analytical representation is as follows:

$$A_1(x) = \begin{cases} 1 - 4x, & \text{if } x \in [0, 0.25], \\ 0, & \text{otherwise,} \end{cases} \quad A_2(x) = \begin{cases} 4x, & \text{if } x \in [0, 0.25], \\ 2 - 4x, & \text{if } x \in [0.25, 0.5], \\ 0, & \text{otherwise,} \end{cases}$$

$$A_3(x) = \begin{cases} 4x - 1, & \text{if } x \in [0.25, 0.5], \\ 3 - 4x, & \text{if } x \in [0.5, 0.75], \\ 0, & \text{otherwise,} \end{cases} \quad A_4(x) = \begin{cases} 4x - 2, & \text{if } x \in [0.5, 0.75], \\ 4 - 4x, & \text{if } x \in [0.75, 1.0], \\ 0, & \text{otherwise,} \end{cases}$$

$$A_5(x) = \begin{cases} 4x - 3, & \text{if } x \in [0.75, 1.0], \\ 0, & \text{otherwise.} \end{cases}$$

By (1), the value of the components $F_1, \ldots, F_5$ of the F-transform are:

$$F_1 = \frac{1}{0.125} \cdot \int_0^{0.25} x^2 \cdot (1 - 4x) \, dx \approx 0.01,$$

$$F_2 = \frac{1}{0.25} \cdot \left(\int_0^{0.25} x^2 \cdot 4x \, dx + \int_{0.25}^{0.5} x^2 \cdot (2 - 4x) \, dx \right) \approx 0.08,$$

$$F_3 = \frac{1}{0.25} \cdot \left(\int_{0.25}^{0.5} x^2 \cdot (4x - 1) \, dx + \int_{0.5}^{0.75} x^2 \cdot (3 - 4x) \, dx \right) \approx 0.26,$$

$$F_4 = \frac{1}{0.25} \cdot \left(\int_{0.5}^{0.75} x^2 \cdot (4x - 2) \, dx + \int_{0.75}^{1} x^2 \cdot (4 - 4x) \, dx \right) \approx 0.58,$$

$$F_5 = \frac{1}{0.125} \cdot \int_{0.75}^{1} x^2 \cdot (4x - 3) \, dx \approx 0.85.$$

Figure 1 provides a graphical representation of the basic functions $A_1, \ldots, A_5$, of the function $f(x) = x^2$, of its F-transform components $F_1, \ldots, F_5$, and of the inverse F-transform $\overline{f}(x)$ of x^2.

F-transform: original motivation. The original motivation for F-transform came from fuzzy modeling [12, 13]. For example, in the situation corresponding to the inverse F-transform, we have n rules

$$\text{IF } x \text{ is } A_1 \text{ THEN } y = F_1;$$
$$\ldots$$
$$\text{IF } x \text{ is } A_n \text{ THEN } y = F_n.$$

These rules are Takagi-Sugeno (TSK) rules with singleton (constant) right-hand sides. For TSK rules, the value corresponding to a given input x is $\overline{f}(x) = \dfrac{\sum\limits_{i=1}^{n} F_i \cdot A_i(x)}{\sum\limits_{i=1}^{n} A_i(x)}$. Since $\sum\limits_{i=1}^{n} A_i(x) = 1$, we get the above formula (2).
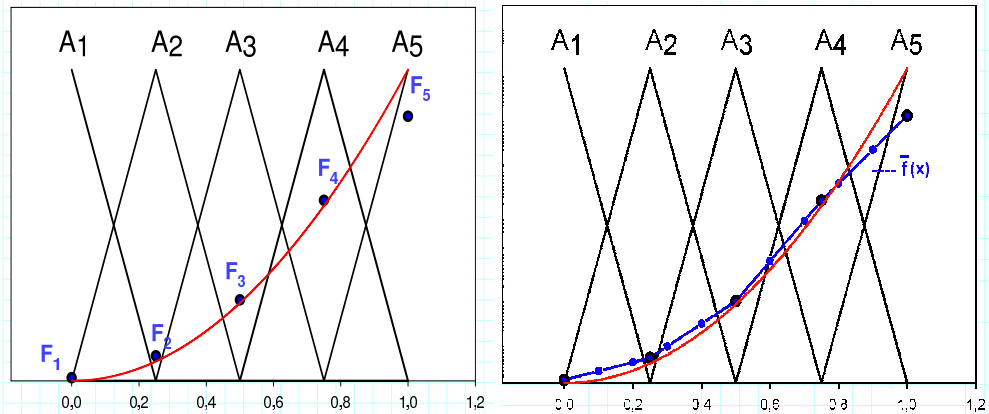
3

Figure 1: Function x^2 on $[0, 1]$ and its F-transform components $F_1, \ldots, F_5$ with respect to $A_1, \ldots, A_5$ (on the left). Function x^2 on $[0, 1]$ and its inverse F-transform $\overline{f}$ (on the right).

The purpose was to show that this type of modeling can be as useful in applications as more traditional techniques such as Fourier transform and wavelet transform. Moreover, F-transform has a potential advantage over Fourier and wavelet transforms: in contrast to the purely mathematical basic functions used in Fourier and wavelet transforms, the basic functions A_i in a fuzzy partition usually come from natural language terms like "low" or "high". (For a detailed description of fuzzy modeling, see, e.g., [5, 10].)

Just like any other tool of applied mathematics, F-transform is not a panacea. It is more successful in some problems, and in other problems, it is less successful. It is therefore desirable to combine F-transform with other mathematical tools, so as to combine relative advantages of different techniques. For combining F-transform with other mathematical tools, it is desirable to come up with a purely mathematical (non-fuzzy) interpretation for this transform.

In particular, since most mathematical data processing tools are based on probability and statistics, it is desirable to come up with a probabilistic interpretation for F-transform.

The known probabilistic interpretation of fuzzy modeling leads to a probabilistic interpretation of F-transform. We have mentioned that F-transform was originally designed as a particular case of fuzzy modeling. A seminal paper [17] provided a reasonable probabilistic model for a particular case of fuzzy modeling. Specifically, this paper shows that if we use piece-wise constant probability density functions for describing the output, then we get a particular case of a fuzzy model – the case when we use product for "and" and sum for "or". Since F-transform corresponds to exactly this type of fuzzy modeling, we thus get a probabilistic model for F-transform as well.

4

What we do in this paper. In this paper, we show that a modification of the probabilistic interpretation from [17] enables us to justify formulas of F-transform without making any additional assumptions about the probability distributions. In mathematical terms, this modification consists of using Bayes formulas – and making assumptions about *prior* distributions (a natural way to describe prior knowledge in statistics) instead of making assumptions about the *actual* distributions.

Thus, we get an even more natural probabilistic interpretation of F-transform. Specifically:

- the paper [17] shows, in effect, that *there exists* a reasonable probabilistic interpretation of the F-transform formulas;

- however, in principle, this interpretation leaves the possibility that *there exist other* equally reasonable assumptions about the probability distributions can lead to different formulas;

- in our modified interpretation, we show that the basic probabilistic setting *uniquely determines* the F-transform formulas – without the need to make any assumptions about the probability distributions.

We also show that a similar modification can be applied to the probabilistic interpretation of general fuzzy modeling formulas.

Comment. From the mathematical viewpoint, the resulting formulas are very similar to the formulas from [17] (with the exception of the Bayes formula step). However, in our opinion, this mathematically *minor* modification leads to a *major* change in interpretation: now, to probabilistic researchers, F-transform is:

- not just *a* possible model, corresponding to one of the possible reasonable choices of probability distributions,

- but *the* model uniquely emerging from the natural probabilistic setting.

Similar conclusion can be made about the probabilistic interpretation of more general fuzzy models.

In other words, our minor modification uncovers an even deeper fundamental meaning of the probabilistic interpretation originally proposed in [17].

2 A Natural Practical Problem that Leads to F-transform

Physical setting: general discussion. Let us assume that we have a physical process that is characterized by two quantities x and z, and we know that these quantities are related by a functional dependence $z = f(x)$.

In the ideal situation of complete knowledge,

- we know the exact value of x, and

- we have the exact description of the function f.

In this case, we can get the corresponding *exact* value $z = f(x)$ of the second quantity.

In practice, we know the value x with uncertainty, i.e., several different values of x are consistent with our knowledge. We must therefore provide a reasonable *estimate* for z. Finding such an estimate will be the *first problem* with which we will be dealing. In this first problem, we assume that the function f is known *exactly*.

If this function has to be determined *empirically*, then we shall transform the empirical (often, partial) knowledge about f into a reasonable estimate for this function. This will be the *second problem* with which we will be dealing in this section.

First problem: estimating the value $f(x)$ for an imprecisely known x. If we only know one piece of information X_i about x, what is the reasonable estimate for $z = f(x)$?

Second problem: estimating the function $z = f(x)$ based on partial information about the dependence between x and z. Assume that for every information piece X_i, $1 \leq i \leq n$, we have the corresponding measured value F_i of z. Since we know only n numerical characteristics F_i of the unknown function f, we cannot exactly reconstruct this function. Instead, we need to provide a good estimate for each value $f(x)$ of this function.

3 A Natural Probabilistic Problem that Leads to the Probabilistic Interpretation of F-transform

Uncertainty in x: a general probabilistic description. Assume that we have a *model* of the estimation procedure, that enables us, given the actual value x, to compute the probability $P(X_i \,|\, x) \geq 0$ of this procedure resulting in X_i – under the condition that the actual (unknown) value of the estimated quantity is x.

To simplify formulas, we denote

$$A_i(x) \stackrel{\text{def}}{=} P(X_i \,|\, x). \tag{3}$$

Since for every x, we must have exactly one of the n possible outcomes, we thus conclude that the probabilities $P(X_1 \,|\, x)$, ..., $P(X_n \,|\, x)$ of different estimation results must add up to one, i.e., we must have

$$P(X_1 \,|\, x) + \ldots + P(X_n \,|\, x) = 1. \tag{4}$$

In the above simplified notation, this formula takes the form

$$A_1(x) + \ldots + A_n(x) = 1. \tag{5}$$

First problem: estimating the value $f(x)$ for an imprecisely known x. Let us consider the first problem. In practice, we do not know the exact value of the quantity x. Instead, we only have one of the information pieces X_i, $1 \le i \le n$. Under the assumption that we know X_i, what is the reasonable estimate for $z = f(x)$?

In terms of probability theory, we would like to find the conditional expected value $F_i \stackrel{\text{def}}{=} E[z \,|\, X_i] = E[f(x) \,|\, X_i]$ of $z = f(x)$ under the condition X_i.

By definition, this expected value is equal to

$$F_i = E[f(x) \,|\, X_i] = \int f(x) \cdot P(x \,|\, X_i)\, dx. \tag{6}$$

Thus, to compute this expected value, we must know the probabilities $P(x \,|\, X_i)$. Instead, we know the probabilities $P(X_i \,|\, x)$.

In general, the problem of reconstructing

- probabilities $P(H_x \,|\, X_i)$ of different hypotheses H_x based on the observation X_i from

- conditional probabilities $P(X_i \,|\, H_x)$ of this observation under different hypotheses H_x

is well known in probability theory; it is solved by applying the Bayes theorem. The continuous version of this theorem is

$$P(H_x \,|\, X_i) = \frac{P(X_i \,|\, H_x) \cdot P(H_x)}{\int P(X_i \,|\, H_y) \cdot P(y)\, dy}, \tag{7}$$

in which $P(H_x)$ is a prior probability of the hypothesis H_x (strictly speaking, $P(H_x \,|\, X_i)$ and $P(H_x)$ are probability densities).

In our case, different hypotheses H_x correspond to different possible values x of the quantity of interest. Thus, (7) takes the form

$$P(x \,|\, X_i) = \frac{P(X_i \,|\, x) \cdot P(x)}{\int P(X_i \,|\, y) \cdot P(y)\, dy}. \tag{8}$$

Since there is no a priori reason to prefer one value of x to the other, it is reasonable to assume that all the values x are equally probable, i.e., that all prior values $P(x)$ are equal to each other: $P(x) = P_0$.

Substituting $P(x) = P_0$ into the formula (8) and dividing both the numerator and the denominator by the common factor P_0, we get the expression

$$P(x \,|\, X_i) = \frac{P(X_i \,|\, x)}{\int P(X_i \,|\, y)\, dy}.$$

Substituting this expression into formula (6) (and renaming the variable in the denominator), we get

$$F_i = E[f(x) \,|\, X_i] = \frac{\int f(x) \cdot P(X_i \,|\, x)\, dx}{\int P(X_i \,|\, x)\, dx}.$$

In terms of the simplified notation (3), we thus get

$$F_i = E[f(x) \,|\, X_i] = \frac{\int f(x) \cdot A_i(x)\,dx}{\int A_i(x)\,dx}, \tag{9}$$

i.e., exactly the formula (1) corresponding to F-transform.

Second problem: estimating the function $z = f(x)$ based on partial information about the dependence between x and z. In some practical situations, we do not know the exact expression for the function $f(x)$. Instead, we must estimate $f(x)$ from the empirical data, i.e., from the previous results of simultaneous measuring x and z.

In each such measurement, the only information that we get about x is one of the values $X_1, \ldots, X_n$. For each case when the information about x is X_i, we have one or several values z.

Ideally, we should have a large number of values z corresponding to each x-measurement result X_i. Based on these values z, we should then be able to reconstruct the conditional distribution of z under the condition of X_i. Based on these conditional distributions, we should be able to reconstruct the values $f(x)$ for all x.

In practice, however, we have only a few values z corresponding to each x-measurement result X_i. In this case, at best, instead of the entire conditional probability distribution, we can only reconstruct a single parameter – the conditional mean $F_i = E[z \,|\, X_i]$. Since we only know n characteristics F_i of the unknown function $f(x)$, we cannot exactly reconstruct this function. Instead, we need to describe a good estimates for each value $f(x)$ of this function.

Similarly to the first problem, we take the mean as a reasonable estimate. Thus, in the above practical setting, the problem of estimating the function $f(x)$ takes the following form:

- for every i, we know the conditional mean $F_i = E[f(x) \,|\, X_i]$;

- based on these conditional means, for every x, we want to estimate the mean value $\overline{f}(x) \stackrel{\text{def}}{=} E[z \,|\, x]$.

For this problem, the formula of full probability leads to the following result:

$$E[z \,|\, x] = \sum_{i=1}^{n} E[z \,|\, X_i] \cdot P(X_i \,|\, x). \tag{10}$$

By using the notations $\overline{f}(x)$ for $E[z \,|\, x]$, F_i for $E[z \,|\, X_i]$, and $A_i(x)$ for $P(X_i \,|\, x)$, we can transform the formula (10) into the form

$$\overline{f}(x) = \sum_{i=1}^{n} F_i \cdot A_i(x), \tag{11}$$

i.e., exactly the F-transform inversion formula (2).

Conclusion. The above (minor) modification of a probability model from [17] uniquely determined both basic formulas (1) and (2) related to F-transform.

Relation with the random set interpretation of fuzzy sets. It is worth mentioning that the probabilistic interpretation from [17] is related to the random set interpretation of fuzzy sets (see, e.g., [6]).

In this interpretation, the meaning of an imprecise (fuzzy) term like "small" is based on the following idea. The fact that the term is imprecise means that for the same value x, some people will say that this value is small, while other people will say that this value is not small. To take this imprecision into account, we can store, for each person, a set of all the values that this person considers small.

Since there is no prior reason to prefer the opinion of one of these folks, we consider their opinions equally reasonable. We can then take the ratio $\mu_{\mathrm{small}}(x)$ of people who consider x to be small as a reasonable measure of smallness. (This is actually one of the standard ways to construct a membership function corresponding to a certain term.)

We can describe this ratio in probabilistic terms if we assume that all the persons are equally probable. In these terms, the value $\mu_{\mathrm{small}}(x)$ can be interpreted as the probability $P(\mathrm{small}\,|\,x)$ that a randomly selected person would consider x to be small.

This interpretation of the membership function $A_i(x)$ as the conditional probability $P(X_i\,|\,x)$ is exactly what we used in our probabilistic interpretation of F-transform.

Terminological comment. For completeness, let us explain why the above interpretation is called the random sets interpretation.

For crisp (well-defined) properties, each property can be described by the set of all the values that satisfy this property.

For each imprecise property like "small", instead of a *single* set describing all the values that satisfy this property, we have *several* sets describing the opinions of several persons. We consider the opinions of all these persons to be equally valid, so each of N persons has the exact same probability $1/N$ of being correct. In this case, we have different sets, each occurring with probability $1/N$.

In mathematical terms, we can describe this situation by saying that we have a probability distribution on the class of all possible sets. In probability theory, such a distribution is called a *random set* – similarly to the fact that a probability distribution on the class of all possible numbers is called a *random number*.

9

4 Discussion

Let us discuss what are the consequences of the above results for the meaning and usage of F-transforms[1]. To start this discussion, let us recall why F-transforms were proposed in the first place.

Need for F-transforms and the resulting main advantage of F-transforms: reminder. One of the main objectives of F-transform is to approximate general functions by functions from a selected finite-parametric family. This is well-known mathematical problem, and many successful techniques have been developed for solving this problem. For example, we can expand the original function by a polynomial, and then use the first few terms in this expansion as the desired approximation. We can also use transforms such as Fourier transform or wavelet transform, and keep only the first few terms in the corresponding expansion as the desired approximation.

All existing approximation techniques take a function $f(x)$ and approximate this function. In situations in which the only information that we have about the desired dependence $y = f(x)$ are the values of y measured for several values of x, this is the only thing we can do. However, in practice, we often have additional expert knowledge about the dependence $y = f(x)$. It is therefore desirable to take this understanding into account when we approximate a function.

The expert knowledge is often imprecise ("fuzzy"), i.e., formulated in terms of imprecise expert rules. A natural way to describe imprecise rules is to use fuzzy logic and fuzzy modeling, and, as we shown, the fuzzy modeling approach naturally leads to F-transforms.

The ability to take into account expert knowledge is thus the main advantage of F-transforms, the main reason reason why F-transform has led to many successful applications.

The probabilistic interpretation of F-transform leads to an additional advantage of F-transform in comparison with other approximation techniques. The above probabilistic interpretation of F-transforms shows that each component F_i of an F-transform can be interpreted as a mean value $E[f(x) \,|\, X_i]$ of the approximating function $f(x)$ under the condition that the unknown value x is consistent with the measurement result X_i. It is well known that in probability theory, the mean value can be alternatively described as the value z that minimizes the mean square difference between this value and the actual value $f(x)$, i.e., that minimizes the expression $E\left[(f(x) - z)^2 \,|\, X_i\right]$. Thus, the above relation provides an additional advantage of F-transforms in comparison with other approximation tools:

- F-transforms not only reflect expert knowledge,

---

[1]The authors are greatly thankful to the anonymous referees who proposed the main ideas of this discussion.

- F-transforms also provide a solution which is *optimal* (in a well-defined reasonable sense).

Gauging the accuracy of the resulting approximation. We have shown that each component $z = F_i$ of the F-transform provides the approximation to $f(x)$ which is the most accurate. The next natural question is: how accurate is it? In other words, what is the corresponding mean square difference $\sigma^2 = E\left[(f(x) - z)^2 \,|\, X_i\right]$? It turns out that the answer to this question can also be provided in terms of F-transforms.

Namely, as it is known, for $z = F_i = E[f(x) \,|\, X_i]$, we have

$$\sigma^2 = E\left[(f(x) - z)^2 \,|\, X_i\right] = E\left[f^2(x) \,|\, X_i\right] - \left(E\left[(f(x) - z)^2 \,|\, X_i\right]\right)^2,$$

i.e., $\sigma^2 = E\left[f^2(x) \,|\, X_i\right] - F_i^2$. The expression $E\left[f^2(x) \,|\, X_i\right]$ can also be described in terms of F-transforms. Indeed, our result about the relation between F-transform and conditional expected value applies to all possible functions, including the square $f^2(x)$ of the original function $f(x)$. Thus, each value $E\left[f^2(x) \,|\, X_i\right]$ is equal to the i-th component S_i of the F-transform of this square.

So, we arrive at the following conclusion. If we only know that x is consistent with the measurement result X_i, then:

- a reasonable approximation for $f(x)$ is the value F_i: the i-th component of the F-transform, and

- the root mean square accuracy σ of this approximation is determined by the formula $\sigma^2 = S_i - F_i^2$, where S_i is the i-th component of the F-transform of the function $f^2(x)$.

Similarly, for the second problem – reconstructing $f(x)$ when we know only know finitely many values corresponding to different i – the mean square accuracy of the corresponding approximation of the actual (unknown) function $f(x)$ by its inverse F-transform $\overline{f}(x)$ is equal to

$$\sigma^2(x) \overset{\text{def}}{=} E\left[\left(f(x) - \overline{f}(x)\right)^2 \,|\, x\right] = E\left[f^2(x) \,|\, x\right] - \left(\overline{f}(x)\right)^2.$$

The first term $E\left[f^2(x) \,|\, x\right]$ in this difference is equal to the inverse F-transform

$$\overline{f^2}(x) = \sum_{i=1}^n S_i \cdot A_i(x),$$

where the values $S_1, \ldots, S_n$ form an F-transform of the squared function $f^2(x)$.

Thus, we arrive at the following conclusion:

- If we only know the values $F_1, \ldots, F_n$ of the F-transform of the actual (unknown) dependence $f(x)$, then, as a reasonable approximation to $f(x)$, we can take the inverse F-transform $\overline{f}(x) = \sum_{i=1}^n F_i \cdot A_i(x)$.

11

- If, in addition to the values F_i, we also know the F-transform $S_1, \ldots, S_n$ of the square $f^2(x)$, then we can estimate the root means square accuracy $\sigma(x) = \sqrt{E\left[\left(f(x) - \overline{f}(x)\right)^2 \mid x\right]}$ by using the formula

$$\sigma^2(x) = \overline{f^2}(x) - \left(\overline{f}(x)\right)^2,$$

where $\overline{f^2}(x) = \sum\limits_{i=1}^{n} S_i \cdot A_i(x)$ is the inverse F-transform of the squared function.

5 A Similar Modification of a Probabilistic Interpretation Is Possible For Mamdani-Style Fuzzy Modeling (and Fuzzy Control)

From F-transform to fuzzy modeling. Let us show that the above modification of a probabilistic interpretation from [17] can be extended from F-transform to a more general case of Mamdani-type fuzzy modeling and fuzzy control.

Comment. In this section, we concentrate on Mamdani's approach since F-transform can be viewed as a particular case of this approach, and since for Mamdani's approach, a probabilistic interpretation is possible [17]. Please note that while Mamdani's approach was historically the first, at present, there are many different approaches to fuzzy modeling and fuzzy control; we mention some of them in this chapter, but there are many others; see, e.g., [7, 8, 9]. How to best interpret these other approaches in probabilistic terms – and whether such an interpretation is at all possible – is an interesting open question.

For example, an interesting question is how to interpret type-2 approaches to fuzzy modeling and fuzzy control; see, e.g., [1, 2, 4, 20]; maybe via interval-valued probabilities?

Mamdani's approach to fuzzy modeling and fuzzy control: a brief reminder. In Mamdani's approach, we start with rules like

"if x is small, then u should be medium",

and then use membership functions for "small" and "medium" to transform these rules into an exact control strategy.

In general, we have rules

"if x has a property A_i then u has the property B_i" $(1 \leq i \leq n)$,

with known membership functions $A_i(x)$ and $B_i(u)$ for the corresponding properties. Mamdani's methodology is based on saying that for each input x, the value u is a reasonable value of control if and only if one of the above n rules is applicable, i.e.,

- either the first rule is applicable, i.e., x satisfies the property A_1 and u satisfies the property B_1,

- or the second rule is applicable, i.e., x satisfies the property A_2 and u satisfies the property B_2,

- ...

- or the n-th rule is applicable, i.e., x satisfies the property A_n and u satisfies the property B_n.

Once we select functions $f_\&(a, b)$ and $f_\vee(a, b)$ to represent "and" and "or" (these functions are called *t-norm* and *t-conorm*), we can thus describe the degree of our belief $\mu_x(u)$ that u is reasonable (for a given input x) as

$$\mu_x(u) = f_\vee(f_\&(A_1(x), B_1(u)), \ldots, f_\&(A_n(x), B_n(u))). \qquad (12)$$

In particular, if we select $f_\&(a, b) = a \cdot b$ and $f_\vee(a, b) = \min(a + b, 1)$ (and if the added values do not go beyond 1), we get

$$\mu_x(u) = \sum_{i=1}^{n} A_i(x) \cdot B_i(u). \qquad (13)$$

Once we know this membership function, we can find the appropriate value of u by using the so-called *centroid defuzzification*:

$$\overline{u}(x) = \frac{\int u \cdot \mu_x(u) \, du}{\int \mu_x(u) \, du}. \qquad (14)$$

A natural probabilistic analog of Mamdani's approach to fuzzy modeling. In [17], it was shown that in a probabilistic setting, we get formulas similar to Mamdani rules corresponding to $f_\&(a, b) = a \cdot b$ and $f_\vee(a, b) = \min(a + b, 1)$ – if we assume a uniform distribution on the outputs. Let us show that by using Bayes formula, we can avoid this additional assumption, and thus, make the resulting probabilistic analog of Mamdani's fuzzy modeling even more natural.

Similarly to the above probabilistic interpretation of F-transform, let us assume that we have possible pieces of information $X_1, \ldots, X_n$ about the quantity x, and that for each piece of information, we also know the corresponding probability $P(X_i \mid x)$ which we will be denoted by $A_i(x)$.

Similarly, let us assume that we have possible pieces of information $U_1, \ldots, U_m$ about u, and we know the corresponding probabilities $P(U_i \mid u)$ which we will denote by $B_i(u)$.

We know that u depends on x, but we do not know the exact dependence. Instead, for each information X_i about x, we know the corresponding information U_j about the corresponding u.

Since we did not select any specific order for the informations U_i, we can select the value corresponding to X_1 as U_1, the value corresponding to X_2 by U_2, etc. Under this selection, the available information simply means that if x is

described by the piece of information X_i, then the corresponding u is described by the piece of information U_i.

Our objective is, given these rules and given a new value x, to find a good estimate for the appropriate u.

Due to the formula of full probability, the conditional probability density $P(u \mid x)$ of u under the condition x has the form

$$P(u \mid x) = \sum_{i=1}^{n} P(u \mid U_i) \cdot P(X_i \mid x). \tag{15}$$

We know the probabilities $P(X_i \mid x) = A_i(x)$. The probability densities $P(u \mid U_i)$ can be determined by using the Bayes theorem – similarly to the F-transform case – as

$$P(u \mid U_i) = \frac{P(U_i \mid u)}{\int P(U_i \mid y)\, dy}, \tag{16}$$

i.e., in terms of the values $B_i(u)$, as

$$P(u \mid U_i) = \frac{B_i(u)}{\int B_i(y)\, dy}. \tag{17}$$

Substituting the formula (17) and the expression (3) into the formula (15) (and changing the multiplication order), we get the formula

$$P(u \mid x) = \sum_{i=1}^{n} A_i(x) \cdot \frac{B_i(u)}{\int B_i(y)\, dy}. \tag{18}$$

Once we know these probabilities, we can produce the mean $\overline{u}$ as a reasonable estimate for u:

$$\overline{u}(x) = \frac{\int u \cdot P(u \mid x)\, du}{\int P(u \mid x)\, du}. \tag{19}$$

These are exactly the formulas derived in [17] from the additional assumption of a piece-wise constant output distribution. Thus, our (minor) modification of [17] indeed uniquely determines the corresponding probabilistic analog of Mamdani's formulas.

In Mamdani-type setting, fuzzy and probabilistic formulas are, in general, different. It is worth mentioning that

- while in F-transform, the probabilistic and fuzzy derivations lead to exactly the same formulas,

- in the general fuzzy modeling case, as mentioned in [17], the formulas are somewhat different:

  - while the formula (19) is exactly the same as (14), with $P(u \mid x)$ instead of $\mu_x(u)$;
  - the formula (18) is slightly different from Mamdani's formula (13) – by the integral in the denominator.

14

Cases when fuzzy and probabilistic formulas coincide. For F-transform (and, more generally, in all the cases when the value $\int B_i(y)\,dy$ is the same for all i), this additional denominator simply divides all the values $P(u\,|\,x)$ by the constant. This constant appears both in the numerator and in the denominator of the formula (18) and thus, it does not affect the resulting value $\overline{u}(x)$.

Another case when the fuzzy and probabilistic formulas coincide is the case of the Takagi-Sugeno (TSK) approach; see, e.g., [7]. This equivalence is, in effect, proven in [17]. In the TSK approach, rules have the type

"if x has a property A_i then $u = f_i(x)$" $(1 \leq i \leq n)$,

for known functions $f_i(x)$. In the probabilistic setting, we assume that under piece of information U_i, we must take $u = f_i(x)$. Thus, for a given input x, we select $f_i(x)$ with probability $P(X_i\,|\,x) = A_i(x)$, where $\sum\limits_{i=1}^{n} A_i(x) = 1$. The resulting mean $\overline{u}(x)$ is thus equal to $\sum\limits_{i=1}^{n} A_i(x) \cdot f_i(x)$. For the case when $\sum\limits_{i=1}^{n} A_i(x) = 1$, this is exactly the TSK formula.

Comparison between fuzzy and probabilistic modeling. For Mamdani-type situations when fuzzy and probabilistic formulas are different, the comparison of the corresponding probabilistic and fuzzy rules is done, in detail, in [17].

Let us add three more situations to this comparison, situations that are naturally related to our modified derivation.

Case when probabilistic control is better. When the values $\int B_i(y)\,dy$ are different, probabilistic control and fuzzy control lead, in general, to a different value $\overline{u}$. We will show, on an example originally proposed by R. Yager, that in this case, the result of the probabilistic control is closer to common sense that the result of Mamdani's control.

Indeed, let us consider the situation in which we have two rules:

- the first rule is a more general rule saying that if x is small, then u should be small;

- the second rule is a very specific rule, saying that if x is very close to 0.11, then u should be very close to 0.15.

Intuitively, if we have a value x for which a very specific rule is applicable, e.g., the value $x = 0.11$, then this specific rule should have a priority over the general rule. However, since the width of the membership function $B_2(u)$ is small, the corresponding term in (13) will practically not affect the resulting estimate (14).

In contrast, in the probabilistic control, the effect of $B_2(u)$ is normalized by, crudely speaking, the total width of the corresponding function $B_2(u)$. Thus, even the most specific rules will have – as desired – the significant influence on the result (19).

Comment. It should be mentioned that the problem with specific rules occurs only in Mamdani's approach to fuzzy control. In the alternative *logical* approach, this problem does not appear; see, e.g., [11].

Another case when probabilistic control is better. The probabilistic interpretation enables us to naturally consider more general situations in which the rules are themselves probabilistic, i.e., when, for each i and j, we know the conditional *probability* $P(U_i \mid X_j)$ that if x has the property X_j, then u has the property U_i.

In other words, instead of the original rules

"if x has the property X_i, then u has the property U_i",

we now have rules

"if x has the property X_j, then u has the property U_i
with probability $P(U_i \mid X_j)$".

Indeed, in this case, due to the formula of full probability, the conditional probability density $P(u \mid x)$ of z under the condition x has the form

$$P(u \mid x) = \sum_{i=1}^{n} \sum_{j=1}^{n} P(u \mid U_i) \cdot P(U_i \mid X_j) \cdot P(X_j \mid x). \tag{20}$$

Here, we know the original probabilities $P(U_i \mid X_j)$ and the probabilities $P(X_i \mid x) = A_i(x)$. The probability densities $P(u \mid U_i)$ can be determined by using the Bayes theorem as an expression (17). Substituting the formula (17) and the expression $P(X_i \mid x) = A_i(x)$ into the formula (20) (and changing the multiplication order), we get the formula

$$P(u \mid x) = \sum_{i=1}^{n} \sum_{j=1}^{n} P(U_i \mid X_j) \cdot A_j(x) \cdot \frac{B_i(u)}{\int B_i(y) \, dy}. \tag{21}$$

Once we know these probabilities, we can produce the mean $\overline{u}$ by using the formula (19).

In some cases, fuzzy control is better. We have shown that in some situations, probabilistic control is better than the original Mamdani's fuzzy control. However, in other situations, the fuzzy control is better. Let us give two examples.

Case when Mamdani's formulas are better. The above probabilistic formulas only work for the case when $\sum_{i=1}^{n} A_i(x) = 1$ – i.e., in the probabilistic terms, when the properties A_i are mutually exclusive. In practice, we may have non-exclusive properties, in which case we may have $\sum_{i=1}^{n} A_i(x) > 1$.

It is not clear how to handle this situation within the probabilistic approach. However, such situations are not a problem if we apply fuzzy control: its formulas are applicable no matter whether we satisfy the requirement $\sum_{i=1}^{n} A_i(x) = 1$ or not.

Other cases when Mamdani's formulas are better. The probabilistic interpretation is only possible when we use multiplication and addition as "and" and "or" operations $f_\&$ and $f_\vee$.

Fuzzy control does not necessarily have to use these operations, it can use different t-norms and t-conorms. It is an empirical fact that in many control situations, the use of t-norm different from the product and of the t-conorm different from the sum leads to a much better quality control – e.g., a more stable or a smoother one.

In [19], we have formulated the problem of selecting the t-norm and the t-conorm as a precise optimization problem, and for several objective functions like smoothness or stability, we gave an explicit analytical solutions to these optimization problem – specifically, we described the selection that leads to the optimal values of smoothness or stability. In many of these case, the optimal selection is indeed different from the probabilistic case of product and sum. Thus, fuzzy control methodology indeed leads to a better quality control.

6 Conclusion

The fuzzy transform (F-transform) techniques have been lately shown to be very successful in various applications, including applications where until recently, only more traditional tools like Fourier transform or wavelet transform have been applied. In many other applications, however, the traditional tools have a clear advantage. It is therefore desirable to combine F-transform with the more traditional tools, so as to combine the relative advantages of both techniques. To make this combination easier, it is desirable to interpret F-transform in traditional mathematical terms.

In this paper, we describe a modification of a probabilistic interpretation described in [17]. In this modification, the corresponding probabilistic model uniquely leads to the formulas of the F-transform. A similar modification is described in a more general situation of fuzzy modeling.

References

[1] O. Castillo, P. Melin (Eds.), Special issue on hybrid intelligence systems, Information Sciences 177 (10) (2007).

[2] O. Castillo, P. Melin (Eds.), Special issue on high order fuzzy sets, Information Sciences 179 (13) (2009).

[3] F. Di Martino, S. Sessa, V. Loia, I. Perfilieva, An image coding/decoding method based on direct and inverse fuzzy transforms, International Journal of Approximate Reasoning 48 (2008) 110–131.

[4] H. Hagras. Type-2 fuzzy logic controllers: a way forward for fuzzy systems in real world environments, In: J.M. Zurada, G.G. Yen, J. Wang (Eds.), Computational Intelligence: Research Frontiers, IEEE World Congress on Computational Intelligence, WCCI 2008, Hong Kong, China, June 1-6, 2008, Plenary/Invited Lectures, Springer Lecture Notes in Computer Science, 5050 (2008) 181–200.

[5] G. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applicatins, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[6] H.T. Nguyen, An Introduction to Random Sets, Chapman & Hall/CRC, Boca Raton, Florida, 2006.

[7] H.T. Nguyen, N.R. Prasad, C.L. Walker, E.A. Walker, A First Course in Fuzzy and Neural Control, Chapman and Hall/CRC, Boca Raton, Florida, 2003.

[8] H.T. Nguyen, M. Sugeno (Eds.), Fuzzy Systems: Modeling and Control, Kluwer, Boston, MA, 1998.

[9] H.T. Nguyen, M. Sugeno, R. Tong, R. Yager (Eds.), Theoretical aspects of fuzzy control, J. Wiley, N.Y., 1995

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

[11] V. Novak, I. Perfilieva, J. Močkoř, Mathematical Principles of Fuzzy Logic, Kluwer, Dordrecht, 1999.

[12] I. Perfilieva, Fuzzy transforms: theory and applications, Fuzzy Sets and Systems 157 (2006) 993–1023.

[13] I. Perfilieva, Fuzzy transforms: a challenge to conventional transforms, Advances in Imaging and Electron Physics 47 (2007) 137–196.

[14] I. Perfilieva, V. Novák, V. Pavliska, A. Dvořák, M. Štěpnička, Analysis and prediction of time series using fuzzy transform. In: Proceedings of WCCI 2008, IEEE International Conference on Neural Networks IJCNN'2008, Hong Kong, June 1–6, 2008, pp. 3875–3879.

[15] I. Perfilieva, H. De Meyer, B. De Baets, D. Plšková, Cauchy problem with fuzzy initial condition and its approximate solution with the help of fuzzy transform. In: Proceedings of WCCI 2008, IEEE International Conference on Fuzzy Systems FUZZ-IEEE'2008, Hong Kong, June 1–6, 2008, pp. 2285–2290.

[16] S. Rabinovich, Measurement Errors and Uncertainties: Theory and Practice, Springer Verlag, New York, 2005.

[17] L. Sánchez, J. Casillas, O. Cordón, M.J. Jesus, Some relationships between fuzzy and random set-based classifiers and models, International Journal of Approximate Reasoning 29 (2002), 175–213.

[18] D. J. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, Chapman & Hall/CRC, Boca Raton, Florida, 2004.

[19] M.H. Smith, V. Kreinovich, Optimal strategy of switching reasoning methods in fuzzy control, In: H.T. Nguyen, M. Sugeno, R. Tong, R. Yager (Eds.), Theoretical aspects of fuzzy control, J. Wiley, N.Y., 1995, pp. 117–146. 2005.

[20] R. Sepúlveda, O. Castillo, P. Melin, A.R. Díaz, O. Montiel, Experimental study of intelligent controllers under uncertainty using type-1 and type-2 fuzzy logic, Information Sciences 177 (10) (2007) 2023–2048.