Departmental Technical Reports (CS)   Department of Computer Science

5-2014

Wiener's Conjecture About Transformation Groups Helps Predict Which Fuzzy Techniques Work Better

Francisco Zapata
University of Texas at El Paso, fazg74@gmail.com

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

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

Abstract—**Often, application success only comes when we select specific fuzzy techniques (t-norm, membership function, etc.) – and in different applications, different techniques are the best. How to find the best technique? Exhaustive search of all techniques is not an option: there are too many of them. We need to come up with a narrow class of promising techniques, so that trying them all is realistic. In this paper, we show that such a narrowing can be obtained from transformation groups techniques motivated by N. Wiener's conjecture – which was, in its turn, motivated by observations about human vision.**

I. Formulation of the Problem

Formulation of the problem. Often, application succeed only comes when we select specific fuzzy techniques (t-norm, membership function, etc.) – and in different applications, different techniques are the best; see, e.g., [4], [8], [14]. How to find the best technique?

Exhaustive search of all techniques is not an option: there are too many of them. We need to come up with a narrow class of promising techniques, so that trying them all is realistic.

What we do in this paper. We show that such a narrowing can be obtained from transformation groups techniques motivated by N. Wiener's conjecture – which was, in its turn, motivated by observations about human vision.

Comment. Some of these results first appeared in [5], [6], [7].

II. Wiener's Conjecture: Reminder

Observation that motivated Wiener's conjecture. The closer we are to an object, the better we can determine its shape. Experiments show that there are distinct phases in this determination.

- When the object is very far, all we see is a formless blurb – in other words, objects obtained from other by arbitrary smooth transformations cannot be distinguished.

- When the object gets closer, we can detect whether it is smooth or has sharp angles. We may see a circle as an ellipse, a square as a rhombus (diamond). At this stage, images obtained by a projective transformation are indistinguishable.

- When the object gets closer, we can detect which lines are parallel but we may not yet detect the angles. For example, we are not sure whether what we see is a rectangle or a parallelogram. This stage corresponds to affine transformation.

- Then, we have a stage of similarity transformations – when we detect the shape but cannot yet detect its size.

- Finally, when the object is close enough, we can detect both its shape and its size.

Each stage can be this described by an appropriate transformation group (see a formal description below).

Wiener's conjecture. Humans result from billions of years of evolution. So, Wiener conjectured that if there was a group intermediate between, e.g., all projective and all continuous transformations, our vision mechanism would have used it. Thus, according to the 1940s Wiener's conjecture, such intermediate groups are not possible [13].

Wiener's conjecture confirmed. In the 1960s, Wiener's conjecture was proven [2], [10].

In the 1-D case, projective transformations are simply fractionally linear, and affine are simply linear. Thus, this theorem means that any group containing all linear transformation is either the group of all fractionally-linear ones or the group of all transformations.

III. How Wiener's Conjecture Helps

How this helps: general idea. Fuzzy degrees are not uniquely determined: different elicitation techniques lead, in general, to different values. Sometimes, different scales are related by a linear transformation, sometimes by a non-linear one. In practice, we want a description with finitely many parameters, i.e., we want a finite-dimensional transformation group. Due to

the above result, all such transformations are fractionally linear. We show that this can explain why some t-norms, membership functions, etc., are empirically more successful.

Let us now describe this idea in more detail.

Different assignment procedures are in use. Intelligent systems use several different procedures for assigning numeric values that describe uncertainty of the experts' statements. The same expert's degree of uncertainty that he expresses, for example, by the expression "for sure", can lead to 0.9 if we apply one procedure, and to 0.8 if another procedure is used. Just like 1 foot and 12 inches describe the same length, but in different scales, we can say that 0.9 and 0.8 represent the same degree of certainty in two different *scales*.

Comment. Some scales are different even in the fact that they use an interval different from $[0, 1]$ to represent uncertainty. For example, the famous MYCIN system uses the interval $[-1, 1]$ [1], [9].

In some sense all scales are equal, but some are more reasonable than others. From a mathematical viewpoint, one can use any scale, but from the practical viewpoint some of them will be more reasonable to use, and some of them less reasonable. We'll consider only practically reasonable scales, and we'll try to formalize what that means.

We must describe transformations between the scales. Since we are not restricting ourselves to some specific procedure of assigning a numeric value to uncertainty, we can thus allow values from different scales. If we want to combine them, we must be able to transform them all to one scale. So we must be able to describe the transformations between reasonable scales (*"rescalings"*).

How to describe transformations between reasonable scales. The class F of reasonable transformations of degrees of uncertainty must satisfy the following properties:

First, if a function $x \to f(x)$ is a reasonable transformation from a scale A to some scale B, and a function $y \to g(y)$ is a reasonable transformation from B into some other scale C, then it is reasonable to demand that the transformation $x \to g(f(x))$ from A to C is also a reasonable transformation. In other words, the class F of all reasonable transformations must be closed under composition.

Second, if $x \to f(x)$ is a reasonable transformation from a scale A to scale B, then the inverse function is a reasonable transformation from B to A.

Comment. Thus, the family F must contain the inverse of every function that belongs to it, and the composition of every two functions from F. In mathematical terms, it means that F must be a *transformation group*.

Finally, if the description of a rescaling is too long, it is unnatural to call it reasonable. Therefore, we will assume that the elements of F can be described by fixing the values of n parameters (for some small n). In mathematics, the notion of a group whose elements are continuously depending on finitely many parameters is formalized as the notion of a (connected)

Lie group. So we conclude that reasonable rescalings form a connected Lie group.

Examples of reasonable rescaling transformations. In addition to these general demands, we have some examples of rescalings that are evidently reasonable; see, e.g., [7].

One of the natural methods to assign a truth value $t(S)$ to a statement S is to ask several experts and take

$$t(S) = \frac{N(S)}{N},$$

where N is the number of all experts asked and $N(S)$ is the number of those who believe in S. If all the experts believe in S, then this value is 1 (= 100%), if half of them believe in S, then $t(S) = 0.5$ (50%), etc.

Knowledge engineers want the system to include the knowledge of the entire scientific community, so they ask as many experts as possible. But asking too many experts leads to the following negative phenomenon: when the opinion of the most respected professors, Nobel-prize winners, etc., is known, some less self-confident experts will not be brave enough to express their own opinions, so they will either say nothing or follow the opinion of the majority.

How does their presence influence the resulting uncertainty value? Let N denote the initial number of experts, $N(S)$ the number of those of them who believe in S, and M the number of shy experts added. Initially

$$t(S) = \frac{N(S)}{N}.$$

After we add M experts who do not answer anything when asked about S, the number of experts who believe in S is still $N(S)$, but the total number of experts is bigger $(M + N)$. So the new value of the uncertainty ratio is

$$t' = \frac{N(S)}{N + M} = ct,$$

where

$$c \stackrel{\text{def}}{=} \frac{N}{M + N}.$$

When we add experts who give the same answers as the majority of N renowned experts, then, for the case when $t(S) > \dfrac{1}{2}$, we get $N(S) + M$ experts saying that S is true, so the new uncertainty value is

$$t' = \frac{N(S) + M}{N + M} = \frac{N \cdot t(S) + M}{N + M}.$$

If we add M "silent" experts and M' "conformists" (who vote as the majority), then we get a transformation

$$t \to \frac{N \cdot t + M'}{N + M + M'}.$$

In all these cases the transformation from an old scale $t(S)$ to a new scale t' is a linear function $t \to a \cdot t + b$ for some constants a and b; in the most general case

$$a = \frac{N}{N + M + M'}$$

and

$$b = \frac{M'}{N + M + M'}.$$

Now we are ready to formulate precise definitions.

Definition 1.

- *By a* rescaling *we mean a strictly increasing continuous function f that is defined on an interval $[a, b]$ of real numbers.*

- *Suppose that some set F of rescalings satisfies the following properties is a connected Lie group which contains, for all non-negative integers N, M, and M', a transformation*

$$t \to \frac{N \cdot t + M'}{N + M + M'}.$$

  Then elements of this set F will be called reasonable transformations.

Proposition 1. *Every reasonable transformation $f(x)$ is fractionally linear, i.e., has the form*

$$f(x) = \frac{a \cdot x + b}{c \cdot x + d}$$

for some real numbers a, b, c, and d.

Proof. One can easily prove that the finite-dimensional group F contains all linear transformations. According to the theorem that proves Wiener's conjecture this means that F coincides either with the group of all linear transformations, or with the group of all fractional-linear transformations. In both cases, all transformations from the set F are fractionally linear. The proposition is proven.

Normalizations. A special case of a transformation is a so-called *normalization*. The idea is as follows. Suppose that we are interested in the possibility of different alternatives. We used two methods (or two groups of experts) to estimate the degree of their possibility and got two results: $\mu(a)$ is the result of the first method, and $\mu'(a)$ is the result of the second one. Since we could have used different procedures of representing uncertainty to get these two sets of estimates (i.e., different scales), it makes no sense to compare the values directly. Even if for some a we have $\mu(a) > \mu'(a)$, it is still possible that they represent the same degree of certainty, but converted to different numbers by different procedures. What is independent is the ordering of the values.

If $\mu(a) > \mu(b)$ in some scale, this means that the degree of belief in a is bigger than the degree of belief in b. So the choice of the most "probable" alternative a (i.e., the one, for which the degree $\mu(a)$ is the largest) does not depend on what scale we used. Therefore, in order to compare the two sets of values we *normalize* them, i.e., reduce them both to a scale in which the maximum value $\max_{a \in A} \mu(a)$ (where A is the set of all alternatives) has some prescribed value (usually 1). Usually, in such cases, there exists an alternative a, about which we are absolutely sure that it is impossible in our situation (i.e., $\mu(a) = 0$). It is natural to demand that this value 0 should

remain the same after the "normalization" transformation. So we arrive at the following definition.

Definition 2. *By a* normalization *we mean a reasonable transformation $f(x)$, for which $f(0) = 0$.*

Proposition 2. *Every normalization has the form*

$$f(x) = \frac{k \cdot x}{1 + d \cdot x}.$$

Proof. This follows directly from Proposition 1.

Comment. The most widely used linear normalization $f(x) = k \cdot x$ – resulting in $\mu'(x) = \dfrac{\mu(x)}{\max\limits_{y} \mu(y)}$ – is a particular case of this class of reasonable normalizations.

Selecting membership functions. Suppose that we have a fuzzy notion like "small". Then for $x = 0$, and maybe for extremely small values of some physical quantity x, we are sure that it is a small value; for some sufficiently big x we are absolutely sure that it is not small. However, for intermediate values x we are uncertain whether x is small or not. The bigger the value x, the less we are certain that this value is small. In this case there are two ways to represent our uncertainty: first, we can use a general tool that translates our uncertainty into a number (the value of a membership function $\mu(x)$), and second, we can use this very value x because the bigger the x, the bigger our uncertainty.

Of course, this is true not for all possible x, but only for those x that lie in the "gray zone", between the values that are definitely small (where $\mu(x) = 1$) and those that are definitely not small (where $\mu(x) = 0$). So on every such zone we have two different scales to express uncertainty: x and $\mu(x)$, and therefore the transformation between them (i.e., a function μ) must be a transformation between two reasonable scales, i.e., in our terms, a *reasonable transformation*.

Since we already know that reasonable transformations are fractionally linear, we therefore conclude that is reasonable to restrict ourselves to piecewise fractionally linear membership functions.

Indeed, triangular and trapezoid memberships functions and, more generally, piece-wise linear membership functions – an important particular class of such functions – are the most widely (and most effectively) used in applications of fuzzy techniques.

t-norms and t-conorms. When we communicate, we often do not explicitly pronounce all the assumptions that we make; these assumptions are implicitly assumed. For example, when we ask a specialist in energetics about the perspectives of nuclear and hydro power stations, we implicitly assume that the present-day physics is correct, although we are quite aware of the fact that new physical theories can appear that will lead to new sources of energy.

In view of that there are two ways to ask questions to the experts:

- either we do not mention this implicit knowledge at all,

- or we can explicitly tell him that we are interested in his opinion, and we are not assuming that this implicit knowledge is true.

Then:

- in the first case, the expert will most likely tell us his/her opinion in the assumption that the implicit knowledge is true, while

- in the second case, the expert will try to take into consideration the possibility that some of these implicit statements can turn out to be false.

For example, when we ask a doctor about the chances that a patient will quickly recover from a depression, and we know that he is treated by psychoanalysis, then the doctor may say something like 90%. This estimate can be different from 100% not because this doctor knows of some cases when this treatment did not help, but because he/she may have some doubts in the whole psychoanalysis methodology, and the lacking 10% represent his doubts. If we then ask this same question differently, stressing that we want to know the doctor's estimate of a patient's chances for recovery irrespective of whether the treatment that he receives now is appropriate, the doctor may give a bigger number (for example, 99%), because he/she knows of no cases when a treatment did not help, either by really treating the disease, or by a placebo effect.

Thus we have two different procedures to assign truth values to the same uncertain statement S of the same expert.

- The first procedure, when we do not mention the implicit knowledge B in our question, actually represents the expert's degree of belief in $S\&B$.

- The second procedure, when we especially mention to the expert that we want his estimate of S irrespective of whether B is true or not, we get the degree of belief in S itself.

So if in both cases we use numbers to represent the degree of uncertainty, then for the first procedure we get $t_1(S) = t(S\&B)$, and for the second procedure we get $t_2(S)$.

The transformation from $t_2(S)$ to $t_1(S)$ is an evident example of a transformation between reasonable scales for representing uncertainty, so it should belong to the class of reasonable transformations. In fuzzy techniques, we estimate $t(S\&B)$ as $f_\&(t(S), t(B))$, where $f_\&(a, b)$ is an appropriate "and"-operation (t-norm). Thus, the transformation from $t_2(S)$ to $t_1(S)$ takes the form $a \to f_\&(a, b)$, where by b we denoted $t(B)$. So our conclusion is that this transformation must be reasonable.

The value $t(B)$ can be arbitrary, therefore we can conclude that this transformation must be reasonable for every b.

Comment. It is important to take into account that a reasonable transformation must be strictly monotone, but the transformation $x \to f_\&(x, b)$ is not necessarily strictly increasing. For example, a statement S can be so highly unreliable that although $t(S)$ is positive but small, $S\&B$ is absolutely false

($t(S\&B) = 0$). This is quite possible, but in this case $0 < t(S)$ and $f_\&(0, b) = f_\&(t(S), b) = 0$. Another example is when S is so highly reliable, that our degree of belief in $S\&B$ equals to the degree of belief in B, i.e., $t(S) < 1$, but $f_\&(t(S), b) = b = f_\&(1, b)$.

So we can demand that a transformation $x \to f_\&(x, b)$ is reasonable only on the intervals where the value of this function is different from 0 and b.

So far we considered implicit knowledge that can in principle turn out to be false (i.e., crudely speaking, the pessimistic part of the implicit knowledge). But the implicit knowledge can also include an optimistic part. For example, when we form a knowledge base about energetics, we have in mind that maybe scientists will find some ecologically pure and economically cheap way to use the solar energy. Then every question that we ask the experts can have two interpretations: either we ask them whether this or that prediction is true assuming the existing technological level, or we implicitly admit the possibility of such an optimistic breakthrough. Here we also have two scales $t_1(S)$ and

$$t_2(S) = t_1(S \vee B) = f_\vee(t(S), t(B)),$$

where B is this implicit optimistic possibility and $f_\vee(a, b)$ is the corresponding "or"-operation (t-conorm). So, we can conclude that the function $g(x) = f_\vee(x, b)$ must be a reasonable rescaling for all x, for which $b < g(x) < 1$.

Definition 3.

- *We say that an "and"-operation is* reasonable *if for every real number b from the interval (0,1) the function $g(x) = f_\&(x, b)$ is a reasonable rescaling for all x for which $0 < f_\&(x, b) < b$.*

- *We say that an "or"-operation is* reasonable *if the function $g(x) = f_\vee(a, b)$ is a reasonable rescaling for all x, for which $b < f_\vee(x, b) < 1$.*

Proposition 3. [5] *An "and"-operation is reasonable if and only if it coincides with one of the following operations:*

1) $f_\&(a, b) = \min(a, b)$;

2) $f_\&(a, b) = \dfrac{a \cdot b}{k + (1 - k) \cdot (a + b - a \cdot b)}$

   for some constant $k \geq 0$;

3) $f_\&(a, b) = \max\left(0, \dfrac{\ell \cdot a \cdot b + (\ell - 1) \cdot (a + b - 1)}{k + (k - 1) \cdot (a + b - a \cdot b)}\right)$

   for some constants k and ℓ;

4) $f_\&(a, b) = \min(a, b)$ *if at least one of the values a, b is greater than or equal to some fixed constant $a_\&$, else*

   $$f_\&(a, b) = a_\& \cdot g\left(\widetilde{a}, \widetilde{b}\right),$$

   where

   $$\widetilde{a} \stackrel{\text{def}}{=} \frac{a}{a_\&}, \quad \widetilde{b} \stackrel{\text{def}}{=} \frac{b}{a_\&},$$

   and $g(a, b)$ is equal one of the functions 2), 3).

Proposition 4. [5] *An "or"-operation is reasonable if and only if it coincides with one of the following operations:*

1) $f_\vee(a,b) = \max(a,b);$

2) $f_\vee(a,b) = \dfrac{a+b+(k-1)\cdot a\cdot b}{1+k\cdot a\cdot b}$

   for some constant $k \geq 0$;

3) $f_\vee(a,b) = \min\left(\dfrac{a+b+k\cdot a\cdot b}{1+\ell\cdot a\cdot b}, 1\right)$

   for some constants k and ℓ;

4) $f_\vee(a,b) = \max(a,b)$, *if one of the values a,b is smaller or equal than some fixed constant $a_\vee$, else*

$$f_\vee(a,b) = a_\vee + (1 - a_\vee)\cdot g\left(\widetilde{a},\widetilde{b}\right),$$

   where

$$\widetilde{a} \stackrel{\text{def}}{=} \frac{a - a_\vee}{1 - a_\vee}, \quad \widetilde{b} \stackrel{\text{def}}{=} \frac{b - a_\vee}{1 - a_\vee},$$

   and $g(a,b)$ is equal to one of the functions 2), 3).

Comment. The resulting list of reasonable "and"- and "or"-operations includes (directly or indirectly) most "and"- and "or"-operations that were actually used. This list of reasonable operations includes the original operations of fuzzy logic, algebraic (probabilistic) operations, bold operations, Hamacher operations, operations of MYCIN for positive certainty factors (for negative factors MYCIN uses non-monotonic operations, so our results are not applicable). These results are also in good accordance with the experimental results that show that among associative operations $\min$, $\max$, and algebraic sum and product are the best fit for human reasoning [3].

In addition to the well known operations we got some additional ones. Their interpretation is straightforward: Suppose we have many highly unreliable evidences E_1, E_2, ..., E_n in favor of some hypothesis H. This means that $E_1 \vee E_2 \vee E_3 \vee ... \vee E_n$ implies H, and the truth values $t(E_i)$ are small. For simplicity let's consider the case when the truth values $t(E_i)$ are equal to each other: $t(E_i) = a$ for some a that is much smaller than 1 (denoted $a \ll 1$). In this case, the truth value of H must be greater than or equal to the truth value of $E_1 \vee E_2 \vee E_3 \vee \ldots \vee E_n$.

- If we use $f_\vee(a,b) = a + b - a\cdot b$, then for sufficiently big n, the truth value of H can be arbitrarily big, which contradicts our intuition (because our confidence in the experimental confirmation is greater than in case where we have arbitrarily many confirming guesses).

- If we use $f_\vee(a,b) = \max(a,b)$, then we have another contradiction with intuition according to which two independent confirmations of some hypothesis are always better (i.e., formally, $f_\vee(a,b) > a$ and $f_\vee(a,b) > b$ for sufficiently big a,b).

The additional operations help in avoiding both contradictions: for a, b sufficiently big we have $f_\vee(a,b)$ greater than both a, b, and if a is sufficiently small ($< a_\vee$) then regardless on how many confirmations with this truth value we have, the resulting truth value can never be close to absolute certainty (i.e., to the value 1).

The same interpretation can be given to the additional "and"-operation. For example, although physicists understand that their formulations of physical laws can, in principle, be wrong, they are sufficiently confident in them, and even when they make long sequences of arguments, they still believe in all the results with the same confidence. This means that if a and b are greater than some crucial value $a_\&$, then our degree of belief in $a\&b$ is still greater than $a_\&$.

Negation operations. A negation operation can be defined as a function $f_\neg(x)$ which extends the usual negation from the values 0 and 1 to the whole interval $[0,1]$, i.e., as a function $f_\neg(x)$ for which $f_\neg(0) = 1$ and $f_\neg(1) = 0$.

Negation is also a transformation between reasonable scales, because we can express our uncertainty in a statement A either by our degree of belief in $t(A)$, or by our degree of belief $t(\neg A) = f_\neg(t(A))$ in its negation $\neg A$. So the operation $f_\neg(x)$ that transforms $t(A)$ into our estimate for $t(\neg A)$ is a reasonable transformation.

We already know that every reasonable transformation is fractionally-linear. By considering fractionally linear transformations which satisfy the properties $f_\neg(0) = 1$ and $f_\neg(1) = 0$, we conclude that each reasonable negation operation is described by a formula

$$f_\neg(x) = \frac{1 - x}{1 + k\cdot x}$$

for some $k > -1$.

In particular, for $k = 0$, we get the operation $f_\neg(x) = 1 - x$ which was originally proposed by Zadeh [14] and experimentally confirmed in [3]. The above more general negation operations was originally proposed by Sugeno [11], [12]; these operations are also sometimes a good fit for human reasoning [12]. So, the above result explains the present choice of negation operations.

Comment. Similar results explain which defuzzification to use; see, e.g., [5].

Wiener's conjecture also predicts which neural techniques work better. A similar result also explains why sigmoid activation functions are most successful in neural networks; see, e.g., [6], [7].

References

[1] B. G. Buchanan and E. H. Shortliffe, *Rule-Based Expert Systems. The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, Massachusetts and Menlo Park, California, 1984.

[2] V. M. Guillemin and S. Sternberg, "An algebraic model of transitive differential geometry", *Bulletin of American Mathematical Society*, 1964, Vol. 70, No. 1, pp. 16–47.

[3] H. M. Hersch and A. Caramazza, "A fuzzy-set approach to modifiers and vagueness in natural languages", *Journal of Experimental Psychology: General*, 1976, Vol. 105, pp. 254–276.

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

[5] V. Kreinovich, C. Quintana, R. Lea, O. Fuentes, A. Lokshin, S. Kumar, I. Boricheva, and L. Reznik. "What non-linearity to choose? Mathematical foundations of fuzzy control", *Proceedings of the 1992 International Conference on Fuzzy Systems and Intelligent Control*, Louisville, Kentucky, 1992, pp. 349–412.

[6] V. Kreinovich and C. Quintana, "Neural networks: what non-linearity to choose?", *Proceedings of the 4th University of New Brunswick Artificial Intelligence Workshop*, Fredericton, N.B., Canada, 1991, pp. 627–637.

[7] H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.

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

[9] E. H. Shortliffe, *Computer-Based Medical Consultation: MYCIN*, Elsevier, New York, 1976.

[10] I. M. Singer and S. Sternberg, "Infinite groups of Lie and Cartan, Part 1", *Journal d'Analyse Mathematique*, 1965, Vol. XV, pp. 1–113.

[11] M. Sugeno, *Theory of fuzzy integrals and its applications*, Ph.D. Thesis, Tokyo Institute of Technology, Tokyo, 1974.

[12] M. Sugeno, "Fuzzy measures and fuzzy integrals", In: N. M. Gupta, G. N. Saridis, and B. R. Gaines (eds.), *Fuzzy Automata and Design Processes*, Amsterdam, New York, 1977, pp. 89–102.

[13] N. Wiener, *Cybernetics, or Control and Communication in the Animal and the Machine*, 3rd edition, MIT Press, Cambridge, Massachhusetts, 1962.

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