Measures of Specificity Used in the Principle of Justifiable Granularity: A Theoretical Explanation of Empirically Optimal Selections

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

***Abstract*—To process huge amounts of data, one possibility is to combine some data points into granules, and then process the resulting granules. For each group of data points, if we try to include all data points into a granule, the resulting granule often becomes too wide and thus rather useless; on the other case, if the granule is too narrow, it includes only a few of the corresponding point – and is, thus, also rather useless. The need for the trade-off between coverage and specificity is formalized as the *principle of justified granularity*. The specific form of this principle depends on the selection of a measure of specificity. Empirical analysis has show that exponential and power law measures of specificity are the most adequate. In this paper, we show that natural symmetries explain this empirically observed efficiency.**

I. Formulation of the Problem

Granular computing: a brief reminder. In many practical situations, it is difficult to deal with the whole amount of data:

- it may be that we have too much data, so it is not feasible to apply the usual data processing algorithms to the data as a whole; this is the situation known as *big data*; see, e.g., [9];
- it may be that while in principle, it is possible to eventually process all the data points, this would take longer time than we have – e.g., when we need to make a decision right away;
- it may also be that we want to use our intuition to better process the data, and to use our intuition, we need to present the data in presentable form.

There may be other cases when we have too much data. To deal with such cases, a natural idea is compress the original data into a smaller set.

The overall amount of available data can be estimated by multiplying the overall number of data points by the average amount of bits in each data point. In general, each data point does not carry too much information, so the main way to decrease the overall amount of information is to decrease the number of data points. Of course, we could simply take a sample from the original data set, but that would deprive us of all the information provided by the un-used data points.

A much better idea is to each each new "data point" correspond to several original ones. This "combined" data point is known as a *granule*, and the resulting technique is known as *granular computing*. The general idea of granular computing can be traced to Lotfi Zadeh [21]; for latest developments, see, e.g., [16], [17].

There are many possible types of granules. For example, instead of several numerical values:

- we can consider *intervals* that contain all – or at least most – of the data points; see, e.g., [5], [9], [12];
- we can consider *fuzzy sets*, that describe not only which values are possible, and also to what degree different data points are possible; see, e.g., [1], [6], [10], [14], [20];
- we can consider *type-2 fuzzy* or probabilistic granules; see, e.g., [10], [11]
- we can consider *rough sets*, etc.

How to combine data points into a granule: towards the principle of justifiable granularity. Once we have selected a group of data points that we want to compress into a granule, the question is which granule to select based on these data points.

- If we try to include all data points into a granule, the resulting granule often becomes too wide and thus rather useless.
- On the other case, if the granule is too narrow, it includes only a few of the corresponding point – and is, thus, also rather useless.

We thus need to achieve a trade-off between coverage and specificity.

According to decision theory (see, e.g., [3], [4], [7], [13], [18]), decisions of a rational decision maker can be described as optimizing the expected value of a special function – called *utility function* $u(s)$ – that describes the corresponding preference. In other words, if after making a selection a, we get situations $s_1, \ldots, s_n$ with probabilities $p_1(a), \ldots, p_n(a)$, then we should make a selection for which the expected value

$$p_1(a) \cdot u(s_1) + \ldots + p_n(a) \cdot u(s_n)$$

attains its largest possible value.

One can easily check that if we replace the utility function $u(a)$ by a re-scaled one

$$u_1(s) = k \cdot u(s) + \ell, \tag{1}$$

then we get the same order between selections. Vice versa, if two utility functions $u(s)$ and $u_1(s)$ always lead to the same decisions, then these two functions are linearly related, i.e., there exist constants $k > 0$ and ℓ for which the formula (1) holds for all situations s. In this sense, utility is similar to physical quantities like time or temperature, whole numerical values can change if we select:

- a different measuring unit and/or
- a different starting point.

In our case, hen we replace several data points, we lose information, so in this case, the utility is negative. In our problem, we have two situations.

For some points, we replace these points with a granule.

- The probability P of this replacement can be naturally computed as the proportion of data points that fit into the corresponding granules. This proportion depends on the size of the granule: depends on the size ε of the granule: $P = P(\varepsilon)$: the larger the size, the higher the proportion.
- The utility of this replacement also depends on the size ε of the granule: $u = u(\varepsilon)$: the larger the size, the smaller the utility.

Other points do not fit into the granule and are, thus, simply dismissed (or at least processed in a more complex way).

- The probability of this dismissal (or alternative processing) is, clearly, the remaining probability $1 - P(\varepsilon)$. item Let us denote the utility of this dismissal (or alternative processing) by u_0.

According to decision making, we thus need to select the size ε that maximizes the expected utility

$$P(\varepsilon) \cdot u(\varepsilon) + (1 - P(\varepsilon)) \cdot u_0.$$

This expression can be equivalently rewritten as

$$P(\varepsilon) \cdot S(\varepsilon) + u_0, \tag{2}$$

where we denoted

$$S(\varepsilon) \stackrel{\text{def}}{=} u(\varepsilon) - u_0.$$

The objective function (2) can be further simplified if we take into account that subtracting the same value u_0 from all the values does not change the order and thus, does not change the optimal selection.

Thus, we need to select the value ε for which the product

$$P(\varepsilon) \cdot S(\varepsilon) \tag{3}$$

takes the largest possible value. This ideas has indeed been used to select an appropriate granule:

- the probability $P(\varepsilon)$ that describes what proportion of the data points is covered by the granule is known as the *coverage*;
- the expression $S(\varepsilon)$ – that describes how specific is the granule – is known as *measure of specificity*, and
- the idea of maximizing the product (3) is known as the *Principle of Justified Granularity.*

Which specificity functions work best? The specific selection of the granule size depends on+ the selection of the measure of specificity.

Empirical analysis has shown that, out of several measures of specificity that have been tested, the most adequate results are obtained we when use the following two measures of specificity:

- the exponential measure of specificity

  $$S(\varepsilon) = \text{const} \cdot \exp(-c \cdot \varepsilon),$$

  for some constant c, and
- the power law measure of specificity

  $$S(\varepsilon) = \text{const} \cdot (1 - c \cdot \varepsilon)^{\xi},$$

  for some constants c and ξ.

What we do in this paper. In this paper, we provide a theoretical explanation for this empirical choice. Namely, we show that this choice follows form natural symmetries.

Comment. By definition, the measure of specificity $S(\varepsilon)$ differs from the utility function $u(\varepsilon)$ only by an additive constant u_0. Since, as we have mentioned, the utility function is defined modulo an additive constant ℓ anyway, we can as well talk about selecting an appropriate utility function.

II. Shift-Invariance: First Natural Symmetry

Shift-invariance: formulation of the first natural symmetry. The data points come from measurements (or from expert estimates). Measurements are never absolutely accurate. Thus, the measured values are, in general, somewhat different from the actual (unknown) values of the corresponding quantity.

We usually take the measurement uncertainty into account. However, often, there is an additional source of error that we did not think about. What if there is indeed such additional source of error, of size ε_0? In this case, when a granule of size ε includes all appropriate measurement results, for this granule to include the actual values, we must increase the granule size to $\varepsilon + \varepsilon_0$.

It is reasonable to require that the relative quality of different granules not change if we take this unknown uncertainty into account. In other words, it is reasonable to require that selections based on the shifted utility

$$u_1(\varepsilon) \stackrel{\text{def}}{=} u(\varepsilon + \varepsilon_0)$$

lead to the same selection as selections based on the original utility.

Analysis of the problem. We have already mentioned that when two different utility functions lead to the same selections, we must have

$$u_1(\varepsilon) = k \cdot u(\varepsilon) + \ell$$

for some $k > 0$ and ℓ.

The coefficients k and ℓ, of course, depend on the shift ε_0. Thus, we conclude that for every ε, there exists the values $k(\varepsilon_0)$ and $\ell(\varepsilon_0)$ for which, for all possible values $\varepsilon > 0$ and ε_0, we have

$$u(\varepsilon + \varepsilon_0) = k(\varepsilon_0) \cdot u(\varepsilon) + \ell(\varepsilon_0). \tag{4}$$

Additional natural requirement: smoothness. It is also reasonable to require that when we change the granule size a little bit, the utility will also change a little bit. In mathematical terms, this means that the desired utility function $u(\varepsilon)$ should be smooth, i.e., differentiable.

Now, we are ready for our first result.

Proposition 1. *If $u(\varepsilon)$ is a differentiable function that satisfies the equation (4) for some $k(\varepsilon_0)$ and $\ell(\varepsilon_0)$, then*

$$u(\varepsilon) = \text{const} \cdot \exp(-c \cdot \varepsilon)$$

or

$$u(\varepsilon) = \text{const} \cdot (1 - c \cdot \varepsilon)$$

for some constant c.

Discussion. Thus, we have justified:

- the exponential measure of specificity and
- a specific case of the power law measure of specificity corresponding to $\xi = 1$.

Proof.

1°. We assumed that the utility function $u(\varepsilon)$ is differentiable. Let us prove that in this case, the auxiliary functions $k(\varepsilon_0)$ and $\ell(\varepsilon_0)$ are also differentiable.

Indeed, if we pick two different values $\varepsilon = \varepsilon_1$ and $\varepsilon = \varepsilon_2 \neq \varepsilon_1$, then the formula (4) takes the following form:

$$u(\varepsilon_1 + \varepsilon_0) = k(\varepsilon_0) \cdot u(\varepsilon_1) + \ell(\varepsilon_0);$$

$$u(\varepsilon_2 + \varepsilon_0) = k(\varepsilon_0) \cdot u(\varepsilon_2) + \ell(\varepsilon_0).$$

Thus. we have a system of two linear equations for the two unknowns $k(\varepsilon_0)$ and $\ell(\varepsilon_0)$.

By the Cramer's rule, the solution to this system is a rational – hence differentiable – function of the coefficient and free terms. Since the function $u(\varepsilon)$ is differentiable, all these coefficients and free terms are also differentiable. Thus, we can conclude that the functions $k(\varepsilon_0)$ and $\ell(\varepsilon_0)$ are differentiable.

2°. Now that we know that all three functions $u(\varepsilon)$, $k(\varepsilon_0)$, and $\ell(\varepsilon_0)$ and differentiable, we can use this fact to come up with a differential equation – an equation that will be easier to solve than the original functional equation (4).

For this purpose, let us differentiate both sides of the equation (4) with respect to ε_0. As a result, we get the following expression, where, as usual, $f'(x)$ denotes the derivative of the function $f(x)$:

$$u'(\varepsilon + \varepsilon_0) = k'(\varepsilon_0) \cdot u(\varepsilon) + \ell'(\varepsilon_0).$$

Substituting $\varepsilon_0 = 0$ into this formula, we get

$$u'(\varepsilon) = k_0 \cdot u(\varepsilon) + \ell_0, \tag{5}$$

where we denoted

$$k_0 \stackrel{\text{def}}{=} k'(0) \text{ and } \ell_0 \stackrel{\text{def}}{=} \ell'(0).$$

Since $u' = \dfrac{du}{d\varepsilon}$, we can rewrite the resulting differential equation as

$$\frac{du}{d\varepsilon} = k_0 \cdot u + \ell_0. \tag{6}$$

3°. Let us now solve the corresponding differential equation.

For this purpose, let us separate the variables. We can do it if we:

- multiply both sides of the equation (6) by $d\varepsilon$ and
- divide both sides of this equation by $k_0 \cdot u + \ell_0$.

As a result, we get the following formula:

$$\frac{du}{k_0 \cdot u + \ell_0} = d\varepsilon. \tag{7}$$

Here, we have two options:

- the first option is that $k_0 = 0$;
- the second option is that $k_0 \neq 0$.

Let us consider these two options one by one.

3.1°. Let us first consider the case when $k_0 = 0$. In this case, integrating both sides of the equation (7), we get:

$$\frac{u}{\ell_0} = \varepsilon + C,$$

where C is an integration constant, i.e.,

$$\frac{u}{\ell_0} = C \cdot (1 - c \cdot \varepsilon),$$

where we denoted

$$c \stackrel{\text{def}}{=} -\frac{1}{C}.$$

Multiplying both sides of this formula by ℓ_0, we get

$$u(\varepsilon) = C_1 \cdot (1 - c \cdot \varepsilon),$$

where we denoted $C_1 \stackrel{\text{def}}{=} \ell_0 \cdot C$.

Thus, in the case of $k_0 = 0$, we get a linear measure of specificity.

3.2°. Let us now consider the case when $k_0 \neq 0$. In this case, for a new variable

$$v \stackrel{\text{def}}{=} u + \frac{\ell_0}{k_0},$$

we have $dv = du$ and

$$k_0 \cdot u + \ell_0 = k_0 \cdot v.$$

Thus, the formula (7) takes a simplified form

$$\frac{dv}{k_0 \cdot v} = d\varepsilon.$$

Integrating both sides of this formula, we get

$$\frac{1}{k_0} \cdot \ln(v) = \varepsilon + C,$$

where C is an integration constant. Multiplying both sides of this formula by k_0, we get

$$\ln(v) = k_0 \cdot \varepsilon + C_1,$$

where $C_1 \stackrel{\text{def}}{=} k_0 \cdot C$. By applying exp to both sides, we get

$$v(\varepsilon) = \exp(\ln(v)) = \exp(k_0 \cdot \varepsilon + C_1) = C_2 \cdot \exp(k_0 \cdot \varepsilon),$$

where $C_2 \stackrel{\text{def}}{=} \exp(C_1)$.

Thus for

$$u = v - \frac{\ell_0}{k_0},$$

we get

$$u(\varepsilon) = C_2 \cdot \exp(k_0 \cdot \varepsilon) + \text{const}.$$

So, in the case when $k_0 \neq 0$, we get the exponential measure of specificity.

The proposition is proven.

III. Scale-Invariance: Second Natural Symmetry

Scale-invariance: formulation of the first natural symmetry. The size of the granule is measured in the same units as the values forming this granule. For example, if the granule contains values of length, then the size – i.e., the accuracy of representing a value by the granule – is also measured by units of length.

As we have mentioned earlier, the numerical values of a physical quantity depend on the choice of a measuring unit. If we replace the original unit by a new unit which is λ times smaller, then all the numerical values are multiplied by λ.

For example, if we replace meters by centimeters, all numerical values are multiplied by 100: 2 m becomes 200 cm.

When we change the units, the values ε are replaced by new values $\lambda \cdot \varepsilon$. It therefore seems reasonable to require that the relative quality of different measures of specificity not change if we simply change the measuring unit.

In other words, it seems to reasonable to require that the utility function $u_1(\varepsilon) \stackrel{\text{def}}{=} u(\lambda \cdot \varepsilon)$ be equivalent to the original utility function.

Full scale-invariance is rarely possible. By using the above-mentioned criterion for the equivalence of two utility functions, we conclude that we should have

$$u(\lambda \cdot \varepsilon) = k(\lambda) \cdot u(\varepsilon) + \ell(\lambda), \tag{8}$$

for some functions $k(\lambda)$ and $\ell(\lambda)$ depending on λ.

We already know that, due to shift-invariance, the utility function is either exponential or linear. While linear function satisfies the equation (8), the exponential function does not.

Thus:

- if we require both shift- and scale-invariance, we end up with only linear measures of specificity, and
- we know that empirically, sometimes non-linear measures of specificity work better.

So, we cannot require both shift- and scale-invariance. What can we do?

IV. Let Us Combine Shift- and Scale-Invariance

Why such a combination makes sense. Combining several different invariances makes perfect sense. For example, in the Ohm's Law $V = I \cdot R$ that relates voltage, current, and resistance:

- if we simply change the unit for current,
- the law stops working.

For the formula to remain valid:

- for each change of the unit for measuring current,
- we also need to appropriately change the unit for measuring voltage.

In general, such a situation is typical in physics:

- when a formula is not invariant with respect to one class of transformation,
- it usually means that for each transformation from this class, there is an appropriate transformation from some related class
- so that if we apply both transformations at the same time, we get the same formula as before;

see, e.g., [2].

Let us apply this idea to our case.

Resulting formulation. Since we cannot require that the utility function be invariant with respect to arbitrary re-scaling, let us combine it with shift-invariance the same way as it is done in physics – the same way as have just described. Namely, we require that for every λ, there exists a value $\varepsilon(\lambda)$ for which:

- the re-scaled utility function $u(\lambda \cdot \varepsilon)$ is equivalent to a
- correspondingly shifted one $u(\varepsilon + \varepsilon_0(\lambda))$.

As we have mentioned, equivalence means that

$$u(\lambda \cdot \varepsilon) = k(\lambda) \cdot u(\varepsilon + \varepsilon_0(\lambda)) + \ell(\lambda). \tag{9}$$

Proposition 2. *Let $u(\varepsilon)$, $\varepsilon_0(\lambda)$, $k(\lambda)$ and $\ell(\lambda)$ be differentiable functions for which the equality (9) is satisfied for all λ and ε. Then, either*

$$u(\varepsilon) = C \cdot (1 - c \cdot \varepsilon)^{\xi} + \text{const}$$

or

$$u(\varepsilon) = C \cdot \ln(1 - c \cdot \varepsilon) + \text{const},$$

for some constants C and c.

Discussion. This result explains the efficiency of the power law measure of specificity.

Comment. One can easily check that the logarithmic expression is actually the limit case of the power law expression when ξ tends to 0.

Proof.

1°. Similarly to the proof of Proposition 1, let us reduce the difficult-to-solve functional equation (9) to an easier-to-solve differential equation.

For this purpose, let us differentiate both side of the formula (9) by λ. As a result, we get the following formula:

$$\varepsilon \cdot u'(\lambda \cdot \varepsilon) =$$

$$k'(\lambda) \cdot u(\varepsilon + \varepsilon_0(\lambda)) + k(\lambda) \cdot u'(\varepsilon + \varepsilon_0(\lambda)) \cdot \varepsilon_0'(\lambda) + \ell'(\lambda).$$

Substituting $\lambda = 1$ into this formula, and taking into account that for $\lambda = 1$, there is no change and thus, $\varepsilon_0(1) = 0$, $k(1) = 1$, and $\ell(1) = 0$, we get:

$$\varepsilon \cdot u'(\varepsilon) = k_0 \cdot u(\varepsilon) + m_0 \cdot u'(\varepsilon) + \ell_0,$$

where we denoted

$$k_0 \stackrel{\text{def}}{=} k'(1), \quad m_0 \stackrel{\text{def}}{=} \varepsilon_0'(1), \text{ and } \ell_0 \stackrel{\text{def}}{=} \ell'(1).$$

This formula can be rewritten as

$$\varepsilon \cdot \frac{du}{d\varepsilon} = k_0 \cdot u + m_0 \cdot \frac{du}{d\varepsilon} + \ell_0.$$

2°. Let us now solve this differential equation.

Moving the terms proportional to u' to the left-hand side, we conclude that

$$(\varepsilon - m_0) \cdot \frac{du}{d\varepsilon} = k_0 \cdot u + \ell_0.$$

Now, we can separate the variable. We can do it if we:

- multiply both sides by $d\varepsilon$,
- divide both sides by $\varepsilon - m_0$, and
- divide both side by $k_0 \cdot u + \ell_0$.

As a result, we get the following equation:

$$\frac{du}{k_0 \cdot u + \ell} = \frac{d\varepsilon}{\varepsilon - m_0}. \tag{11}$$

Similarly to the proof of Proposition 1, let us consider two possible cases:

- case when $k_0 = 0$, and
- case when $k_0 \neq 0$.

2.1°. If $k_0 = 0$, then integrating both sides of the formula (11) and taking into account that $d(\varepsilon - m_0) = d(\varepsilon)$, we conclude that

$$\frac{u}{\ell_0} = \ln(m_0 - \varepsilon) + C,$$

for some integration constant C. Thus. we have

$$u(\varepsilon) = \ell_0 \cdot \ln(m_0 - \varepsilon) + C_1,$$

where $C_1 \stackrel{\text{def}}{=} \ell_0 \cdot C$.

Here,

$$m_0 - \varepsilon = m_0 \cdot (1 - c \cdot \varepsilon),$$

where

$$c \stackrel{\text{def}}{=} \frac{1}{m_0}.$$

Thus,

$$\ln(m_0 - \varepsilon) = \ln(m_0 \cdot (1 - c \cdot \varepsilon)) = \ln(m_0) + \ln(1 - c \cdot \varepsilon).$$

Hence, the above formula takes the form

$$u(\varepsilon) = \ell_0 \cdot \ln(1 - c \cdot \varepsilon) + C_2,$$

where

$$C_2 \stackrel{\text{def}}{=} C_1 + \ell_0 \cdot \ln(m_0).$$

So, in the case of $k_0 = 0$, we get the logarithmic measure of specificity.

3.2°. Let us now consider the remaining case $k_0 \neq 0$. In this case, similarly to the proof of Proposition 1, we can introduce a new variable

$$v = u + \frac{\ell_0}{k_0}$$

for which the equation (11) takes the form

$$\frac{dv}{k_0 \cdot v} = \frac{d\varepsilon}{\varepsilon - m_0}.$$

Integrating both parts of this equation, we get

$$\frac{1}{k_0} \cdot \ln(v) = \ln(m_0 - \varepsilon) + C,$$

where C is an integration constant, i.e.,

$$\frac{u}{k_0} = (1 - c \cdot \varepsilon) + C',$$

where

$$c \stackrel{\text{def}}{=} \frac{1}{m_0} \text{ and } C' = C + \ln(m_0).$$

Multiplying both sides of this equation by k_0, we conclude that

$$\ln(v) = k_0 \cdot \ln(1 - c \cdot \varepsilon) + C_1,$$

where $C_1 \stackrel{\text{def}}{=} k_0 \cdot C'$. Applying exp to both sides, and taking into account that

$$\exp(k_0 \cdot \ln(x)) = (\exp(\ln(x)))^{k_0} = x^{k_0},$$

we conclude that

$$v = C_2 \cdot (1 - c \cdot \varepsilon)^{k_0},$$

where $C_2 \stackrel{\text{def}}{=} \exp(C_1)$. Thus,

$$u(\varepsilon) = C_2 \cdot (1 - c \cdot \varepsilon)^{k_0} + \text{const}.$$

Thus, in the case of $k_0 \neq 0$, we get the power law measure of specificity.

The proposition is proven.

ACKNOWLEDGMENTS

This work was supported in part by the US National Science Foundation grant HRD-1242122.

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] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics, Addison Wesley*, Boston, Massachusetts, 2005.

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

[4] P. C. Fishburn, *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, Maryland, 1988.

[5] L. Jaulin, M. Kiefer, O. Dicrit, and E. Walter, *Applied Interval Analysis*, Springer, London, 2001.

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

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

[8] G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017.

[9] V. Mayer-Schönberger and K. Cukier, *Big Data: The Essential Guide to Work, Life, and Learning in the Age of Insight*, John Murray Publ., London, UK, 2017.

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

[11] J. M. Mendel and D. Wu, *Perceptual Computing: Aiding People in Making Subjective Judgments*, IEEE Press and Wiley, New York, 2010.

[12] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.

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

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

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

[16] W. Pedrycz, *Granular Computing: Analysis and Design of Intelligent Systems*, CRC Press/Francis Taylor, Boca Raton, Florida, 2013.

[17] W. Pedrycz, A. Skowron, and V. Kreinovich (eds.), *Handbook on Granular Computing*, Wiley, Chichester, UK, 2008.

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

[19] D. Wang, W. Pedrycz, and Z. Liu, "Design of granular interval-valued information granules with the use of the principle of justifiable granularity and their applications to system modeling of higher type", *Soft Computing*, 2016, Vol. 20, pp. 2119–2134.

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

[21] L. A. Zadeh, "Some reflections on soft computing, granular computing, and their role in the conception, design, and utilization of information/intelligent systems", *Soft Computing*, 1998, Vol. 2, No. 1, pp. 23–25.