XIX IMEKO World Congress
Fundamental and Applied Metrology
September 6-11, 2009, Lisbon, Portugal

PROPERTIES OF FUZZY NOMINAL SCALES

Eric Benoit

LISTIC-Université de Savoie, Annecy, France, eric.benoit@univ-savoie.fr

Abstract – Fuzzy nominal scales were introduced in order to propose a formalism to the representation of empirical quantities by fuzzy subsets of words. The purpose of this paper is to explore the possible relations and operators that can be used on measurements performed with this kind of scale. Then an extension of these scales, named fuzzy metric scale, and based on the definition of a distance between measured values is given. As examples of operators, a medium operator and a median filter are given. Finally a comparison procedure of 2 fuzzy metric scale is given.

Keywords: Fuzzy nominal scale, Measurement theory, Fuzzy subsets theory.

1. INTRODUCTION

The usual way to compare or to predict the evolution of the properties of objects of the empirical world is to manipulate their equivalent properties into a model of the empirical world also named the information world [1]. From this viewpoint, a measurement process translates the manifestation of the object property from the empirical world into a property value of the corresponding virtual object in the information world. The scale formalism defines the link between the relations on real objects of the empirical world and the relations between virtual objects of the information world [2].

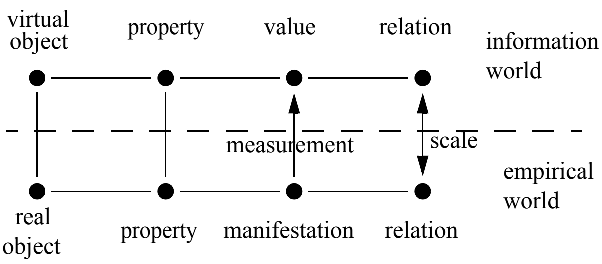

Fig 1. Schematic of the scale formalism.

Actually, it must be considered that using a scale for a measurement process restricts the set of relations that can be applied to the properties values of virtual objects. For example, values on the Beaufort wind force scale can be compared or ordered, but not added.

Fuzzy nominal scales where introduced [3][4] in order to formalize an application to the measurement process of a mechanism of description of a quantity by a fuzzy subset of symbols [5]. With these scales, values in the information world are fuzzy subsets of symbols also called in this paper *lexical fuzzy subsets* (LFS).

As for any scale, using a fuzzy nominal scale for measurement restricts the set of relations that can be used with these values. Actually, only a fuzzy equivalence relation can be used. Other relations can be deduced from the first one. For example, a basic distance had been defined from the fuzzy equivalence relation in order to perform signal processing on LFS [6].

After a short recall on the formalism of fuzzy nominal scales, the purpose of this paper is first to explore the available processing on values issued from a measurement based on a fuzzy nominal scale. Then the introduction of additional knowledge that can increase the number of available relations and the consequence on the nature of the scale is discussed. The introduction of a knowledge about distance on values will allow to define a distance operator.

2. THE FUZZY NOMINAL SCALES

First recall the previous studies relative to the fuzzy nominal scales.

The link between a manifestation and its representation in the information world is characterized by a symbolism defined by the triplet <*E*,*S*,*R*> where:

- *E* is the set of manifestations,
- *S* is the lexical set used to represent measurement results,
- *R* is a relation on *E*x*S*.

Two mappings can be extracted from this relation: The *description mapping* denoted *D* associates a subset of *S* to any item of *E*, and the *meaning mapping* denoted *M* associates a subset of *E* to any item of *S*. These two mappings are linked with the following equation.

$$\forall e \in E, \forall s \in S, e \in M(s) \Leftrightarrow s \in D(e) \qquad (1)$$

When the *R* relation is a fuzzy relation, the translation of a physical state into its linguistic representation is called a *fuzzy linguistic description mapping* or simply a *fuzzy description mapping*. It transforms an object *e* of the set of manifestations *E* into a fuzzy subset of linguistic terms called the *fuzzy description* of *x*. The dual mapping, called the *fuzzy meaning* mapping, associates a fuzzy subset of *E* to each term *s* of the lexical Set *S*. This fuzzy subset is the *fuzzy meaning* of *s*. In the paper the fuzzy subsets of linguistic terms also named *lexical fuzzy subsets* are denoted LFS.

ISBN 978-963-88410-0-1

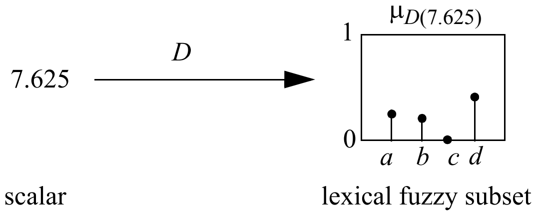

Fig 2. Example of lexical fuzzy subset (LFS).

The equation (1) is translated into:

$$\forall s \in S, \forall e \in E \qquad \mu_{M(s)}(e) = \mu_{D(e)}(s) \qquad (2)$$

Let $F(S)$, a subset of $FS(S)$ the set of all fuzzy subsets on S, be the image of E by D, i.e. be the set of all possible LFS obtained by the associated fuzzy description.

Foulloy [7] categorized the symbolisms $<E,S,R>$ into families and gave a formalism to associate to each family a fuzzy equivalence relation ~ on $F(S)$. Using R, this fuzzy equivalence relation is mapped to an other fuzzy equivalence relation ~ on E. Then $<E,S,R,\{\sim\},\{=\},\{(\sim,=)\}>$ is defined as a fuzzy nominal scale.

The symbolism family that we use in this paper is characterized by the property: the image $M(S)$ is a Ruspini partition on E:

$$\forall x \in E, \sum_{s \in S} \mu_{M(s)}(x) = 1 \qquad (3)$$

This property induces the next property on the elements of $F(S)$.

$$\forall A \in F(S), \sum_{s \in S} \mu_A(s) = 1 \qquad (4)$$

The fuzzy equivalence relation on $F(S)$ is then:

$$\forall (A, B) \in F(S)^2, \mu_{\sim}(A, B) = \sum_{s \in S} min(\mu_A(s), \mu_B(s)) \quad (5)$$

that induces the fuzzy equivalence relation on E:

$$\forall (x, y) \in E^2, \mu_{\sim}(x, y) = \sum_{s \in S} min(\mu_{M(s)}(x), \mu_{M(s)}(y)) \quad (6)$$

At this step the set of operators and relations that can be used with LFSs are directly derived from the ~ operator without any additional knowledge. As for nominal scales, the measurement process is a classification and statistical operations can be applied to the values. As for example the operation introduced to be used with nominal scales are the calculation of the indices of qualitative variation [8][9]. These indices are measures of statistical distribution of a set of sample over a set of categories. Some of them can simply be extended to be used with fuzzy nominal scales.

2.1. Building a fuzzy indice of qualitative variation

Let us present how the *index of deviation from the mode* defined in [8] is extended to fuzzy nominal scales. This index is defined by

$$MODVR = \sum_{\mathcal{C} \in \mathcal{CAT}} (card(\mathcal{C}_m) - card(\mathcal{C})) \qquad (7)$$

Where $\mathcal{CAT}$ is the set of categories, $\mathcal{C}_m$ is the modal category, i.e. the category with the larger number of samples, and *card*($\mathcal{C}$) is the number of samples in the category $\mathcal{C}$.

Let *Smpl*, a subset of F(S), be the set of samples.

Considering the case where a category $\mathcal{C}_C$ is represented by a lexical fuzzy subset C, such that the set of items that fit to a category $\mathcal{C}_C$ is defined by:

$$A \in \mathcal{C}_C \Leftrightarrow A \sim C \qquad (8)$$

That is simply expressed with membership functions as

$$\mu_{\mathcal{C}_C}(A) = \mu_{\sim}(A, C) \qquad (9)$$

The scalar cardinality of a fuzzy subset U on a set Y is defined as

$$card(U) = \sum_{y \in Y} \mu_U(x) \qquad (10)$$

Then the scalar cardinality of a category $\mathcal{C}_C$ is given by

$$card(\mathcal{C}_C) = \sum_{A \in Smpl} \mu_{\mathcal{C}_C}(A) = \sum_{A \in Smpl} \mu_{\sim}(A, C) \quad (11)$$

And the general form of the *index of deviation from the mode* (*MODVR*) is given by

$$\sum_{\mathcal{C}_C \in \mathcal{CAT}} \left(\sum_{A \in Smpl} \mu_{\sim}(A, C_m) - \sum_{A \in Smpl} \mu_{\sim}(A, C) \right) \qquad (12)$$

This definition is strongly simplified when each category is represented by a singleton $\{c\}$. Indeed, the relation (5) between a LFS and a singleton is:

$$\forall A \in F(S), \forall b \in S, \mu_{\sim}(A, \{b\}) = \mu_A(b) \qquad (13)$$

Then the index *MODVR* is given by

$$\begin{aligned} MODVR &= \sum_{\mathcal{C}_{\{c\}} \in \mathcal{CAT}} \left(\sum_{A \in Smpl} \mu_A(c_m) - \sum_{A \in Smpl} \mu_A(c) \right) \\ &= \sum_{\mathcal{C}_{\{c\}} \in \mathcal{CAT}} \left(\sum_{A \in Smpl} \mu_A(c_m) - \mu_A(c) \right) \end{aligned} \qquad (14)$$

This example shows the possibility to extend statistical operators originally defined for nominal scales to fuzzy nominal scales. This extension based on the scalar cardinality stays simple and extensions based on fuzzy cardinality need to be studied.

In order to extend the set of available operators and relations, external knowledge must be used to create new operators. The simplest way is to apply the rule based aggregation mechanism commonly use in fuzzy subset theory.

3. RULE BASED OPERATORS

A general way to define a new operator is to use a rule based approach. The knowledge about the operator is included into a set of rules. One representation of the rules for an operator $C = o(A, B)$ is:

$$\text{if } A \text{ is } a_i \text{ and } B \text{ is } b_i \text{ then } C \text{ is } c_i \quad (15)$$

where

- i indices the rules,
- a_i, b_i, c_i are respectively elements of S_1, S_2, S_3,
- A, B, C are respectively elements of $FS(S_1)$, $FS(S_2)$, $FR(S_3)$,
- $FR(S)$ is the subset of $F(S)$ that verifies an extension of (4):

$$\forall A \in FR(S), \sum_{s \in S} \mu_A(s) = 1 \quad (16)$$

The problem of the C computation is similar to the aggregation of complementary measurements given in [10]. We impose that $S_1 = S_2 = S$. The membership of each element s to the LFS C is given by

$$\forall s \in S_3, \mu_C(s) = \sum_{\{i | c_i = s\}} \mu_A(a_i) \cdot \mu_B(b_i) \quad (17)$$

This simple approach is enough to change the nature of the scale. For example an ordering relation can be defined with a rule set that links the initial lexical set $\{a,b,c\}$ with a set $\{lower, higher\}$ throw this typical rule set:

Table 1. Rule set for comparison operator

o	*a*	*b*	*c*
a	*lower*	*higher*	*higher*
b	*lower*	*lower*	*higher*
c	*lower*	*lower*	*lower*

The ordering relation on E can be deduced from (2) and (17):

$$\forall (x, y) \in E^2$$
$$\mu_{\text{lower}}(x, y) = \sum_{\{i | c_i = lower\}} \mu_{D(x)}(a_i) \cdot \mu_{D(y)}(b_i) \quad (18)$$

In this example we shown that a rule set that defines an ordering relation on S induces also an ordering relation on $FR(S)$ associated to an ordering relation on E. This means that the definition of a relation on the values of virtual object properties adds a knowledge about the relation on manifestation. Two possible conclusions can be made:

- If a ordering relation on the manifestation really exists then this scale is an ordinal scale.
- If no ordering relation on manifestation can be exposed then the scale including this relation is not consistent, and the knowledge included into the rule set is questionable.

In this other example, the definition of this medium operator come from an external knowledge about relation between elements of S. Defining this operator means including this knowledge into the scale then modifying its nature: The operator on S is used to create the equivalent operator on $FR(S)$ that is transposed as an operator on E.

Table 2. rule set for medium operator

	a	*b*	*c*
a	*a*	*b*	*b*
b	*b*	*b*	*b*
c	*b*	*b*	*c*

Like for the preceding example, if the medium operation on manifestation makes no sense, the medium operator on the measured values is meaningless and is not part of the set of operators allowed by the scale.

4. FUZZY METRIC SCALES

Some physical properties like colour or smell are usually measured with nominal scales or fuzzy nominal scales, but some experts, i.e. anybody, are able to give a distance between manifestation. For example yellow is close to orange and very far from blue, or robusta is close to moka, quite close to cacao and very far from mint. Supposing that the concept of distance makes sense on such manifestations, a scale that represents a distance on manifestations with a distance on measured values is possible.

In order to increase the set of available relations, we propose to add a knowledge expressed by the link between the lexical set and a numerical set throw a distance operator d. For each fuzzy nominal scale, a distance on $F(S)$ can be deduced from the relation ~ and from the distance d on S [11]. This distance denoted d_{tp} and named transportation distance [12] is computed as solution for a mass transportation problem where the masses are membership degrees, sources and destinations are items of the lexical set and the unit cost from a source to a destination is given by the distance d on S.

It verifies the following properties:

- The first property is the singleton coincidence: $d_{tp}(\{a\},\{b\}) = d(a,b)$
- The continuity property.
- The precision property that imposes that the distance between two LFSs must be a positive real number.
- The consistency property that is usually verified by distances on crisp subsets

As defining d_{tp} on $F(S)$ is equivalent to defining d_{tp} on E, the scale now includes a new link between the empirical world and the information world and is no more a fuzzy nominal scale and not yet an ordinal scale. To avoid confusion, the name *fuzzy metric scale* is used in this paper:

$$<E,S,R,\{\sim, d_{tp}\},\{=, d\},\{(\sim,=), (d_{tp}, d)\}> \quad (19)$$

This new kind of scale is more suitable for signal processing and allows to define useful operators as for example the medium operator presented below.

4.1. The medium operator

A consequence of the existence of this distance operator is the possibility to compare LFSs. For example, a medium operator can be defined as:

$$m(A,B) = \{C \mid d_{tp}(A,C) = d_{tp}(C,B) = d_{tp}(A,B) / 2\} \quad (20)$$

Such operator gives a set of LFS obtained with the mass transportation process: The computation of the solution of the mass transportation problem used to compute the distance $d_{tp}(A,B)$ gives a set of shipments where each shipment is on the form (source, destination, quantity). The distance is then given by the total cost of shipments:

$$d(A, B) = \sum_{(src_i, dst_j, q_{ij}) \in smt} q_{ij} \cdot c_{ij} \quad (21)$$

where smt is the set of shipments, src_i is the source, dst_j is the destination, q_{ij} is the quantity of transported material and c_{ij} is the unit cost.

The elements of $m(A,B)$ are computed by applying the same set of shipments to A, but with a total cost divided by 2.

This operator gives a set of values instead of single one. It can be remarked that a particular LFS belonging to $m(A,B)$ is given by dividing the quantity by 2 on each shipment, but there is no reason to prefer this one instead of another element of $m(A,B)$.

This medium operator starts a new family of operators that can be used on manifestations measured with fuzzy metric scales like colour or smell. The temptation to perform signal processing on values obtained throw such scales is high but most of usual signal processing algorithms use operators that are banned for theses values. The only algorithms that can be used are the one based on distances like for example the Vector Median Filter.

4.2. Vector median filter

The definition of the vector median filters is based on distance between samples [13][14]. This operator is then an interesting candidate to create a filter on LFSs given by a fuzzy metric scale. The definition of this filter is recall below.

Let X^n be an n-dimensional space.

Let $W=\{x_1, ..., x_n\}$ be a set, denoted *window*, of elements of X^n.

$$\forall i \in [1, n], D_i = \sum_{j=1}^{n} d(x_i, x_j) \quad (22)$$

The output of the vector median filter denoted *VMF* is defined by

$$VMF(W) = x_m \quad \text{such that} \quad D_m = min_i D_i \quad (23)$$

Where d is a distance on X^n.

The definition of a median filter on LFSs is then proposed:

Let $W=\{a_1, ..., a_n\}$ be a set, denoted *window*, of elements of $F(S)$

$$D_i = \sum_{j=1}^{n} d_{tp}(a_i, a_j) \quad (24)$$

The output of the median filter denoted *MF* is defined by

$$MF(W) = a_m \quad \text{such that} \quad D_m = min_i D_i \quad (25)$$

Such filter is a first step for a new field of signal processing where the processed values are lexical fuzzy subsets given throw a fuzzy metric scale.

4.3. Comparing scales

If the knowledge used to define a scale come from a human expert, the comparison of the scale compares the experts. This approach is applied to the fuzzy metric scale as defined in (19) and the comparison is performed with the d_{tp} distance operator.

The principle is to ask two experts to qualify their fuzzy metric scale on a unique set of manifestations E and a unique lexical set S with a distance d defined on it.

let $<E,S,R_1>$ and $<E,S,R_2>$ be the two symbolisms defined by the expert. The difference between the representation of experts on a given manifestation x is given by

$$d_{12}(x) = d_{tp}(D_1(x), D_2(x)) \quad (26)$$

The d_{12} mapping gives a map of expert differences on the set of manifestations. It mainly gives information on the distortions between the two experts.

5. CONCLUSION

The scales properties are important study topics for the definition of the set of available operators for a given scale. This paper shows that fuzzy statistical indices can be deduced from the unique allowed relation that can be used with values issued from fuzzy nominal scales. Defining a new operator that cannot be deduced from the set of allowed relations given by a scale is a way to add a knowledge on the relational structure of the information world. As the information world is a representation of the empirical world the scale is consistent only if the new relational structure of the information world represents an existing relational structure of the empirical world.

A knowledge on distance had been added on the information world associated with a fuzzy nominal scale in order to create a new scale named fuzzy metric scale. This scale maps a metric from the empirical world to the information world. It increases the number of relations and operators that can be used on property values. This paper gives for example a signal filter that now allows to perform signal processing on manifestation of properties like colour or smell.

REFERENCES

[1] L. Mari, "The duality of sensing and coding as a fundamental topic of information science", *World Conf. onSystemics, Cybernetics and Informatics (SCI 2001)*, vol.XVI, pp. 177–180, Orlando, USA, 2001.

[2] L. Finkelstein, "Representation by symbol systems as an extension of the concept of measurement," *Kybernetes*, Vol. 4, pp. 215-223, 1975.

[3] E. Benoit, L. Foulloy, "Towards fuzzy nominal scales", *Measurement*, ISSN 0263-2241, Vol. 34, No. 1 (Fundamental of Measurement), pp. 49-55, 2003.

[4] E. Benoit, L. Foulloy, G. Mauris, "Fuzzy approaches for measurement", *Handbook of Measuring Systems Design*, ISBN: 0-470-02143-8, Vol. 1, Eds. Peter Sydenham and Richard Thorn, John Wiley & Sons, pp. 60-67, 2005.

[5] L.A. Zadeh, "Quantitative fuzzy semantics", *Information Sciences*, Vol. 3, pp. 159-176, 1971.

[6] T. Allevard, E. Benoit, L. Foulloy, "Dynamic gesture recognition using signal processing based on fuzzy nominal scales", *Measurement*, No. 38, pp. 303-312, 2005.

[7] L. Foulloy, E. Benoit, "Building a class of fuzzy equivalence relations", F*uzzy Sets and Systems,* No. 157, pp. 1417-1437, 2006.

[8] A. R. Wilcox, "Indices of Qualitative Variation", *Oak Ridge National Laboratory research report,* N° 3445605133753, October 1967.

[9] D. A. Swanson, "A Sampling Distribution and Significance Test for Differences in Qualitative Variation", *Social Forces*, *Business Source Premier*, Vol. 55, no. 1, pp. 182-184, 1976.

[10]G. Mauris, E. Benoit, L. Foulloy, "The aggregation of complementary information via fuzzy sensors", *Measurement*, Vol. 17, No. 4, pp. 235-249, 1996.

[11]T. Allevard, E. Benoit, L. Foulloy, "The transportation distance for fuzzy descriptions of measurement", *Metrology and Measurement Systems*, Vol. XIV, No. 1/2007, pp. 25-35, 2007.

[12]S.T. Rachev, L. Rüschendorf, *Mass transportation problems*, Springer-Verlag, New-York, 1998.

[13]R. Lukac, B. Smolka, K. Martin, K. Plataniotis, and A. Venetsanopoulos, "Vector filtering for color imaging" *IEEE Signal Process. Mag.*, vol. 22, pp. 74–86, 2005.

[14]J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters" *Proc. IEEE*, vol. 78, no. 4, pp. 678–689, Apr. 1990