The Third AFSS(1998), 573−578

New Fuzzy Concepts as a consequence of the encoding with intervals

Fatiha KARBOU
ENSEM, Dept. EG, B. P. 8118, Oasis, Route d'El Jadida, Casablanca, Maroc

Abstract

In this paper, we propose a new technique of codification. The purpose of this method is to take in consideration the natural language nuances and the fuzziness that characterizes the human reasoning. So, we warranted a means of more flexible encoding that translates as well the linguistic descriptions. Its principle is simple and intuitive. It consists simply in replacing in ambiguous cases, a unique number by an interval. The introduction of the new codification necessitates the elaboration of metric or similarity in order to compare two intervals. This comparison must take in consideration the difference of their size, the remoteness of their center and the width of their intersection. In consequence, we defined three new fuzzy concepts: "fuzzy inclusion degree", "fuzzy resemblance degree", and "fuzzy curve".

Keywords: Fuzzy sets, *fuzzy inclusion degree, fuzzy resemblance degree, fuzzy curve.*

1. Introduction

For different agricultural species, several varieties of seeds and plants are certified in Morocco. Each of these varieties is described by a great number of morphological and vegetable characters, etc. These characters constitute the basic data that compare the varieties' tests with the different varieties of the same species already exist in the official catalog [1]. The numerical codification of these characters is realized in two stages:

1. First, the character is described with a linguistic expression such as: "weak", "very weak", "medium", etc.
2. Then, each of these expressions is quantified by associating, according to UPOV's norms [1], a number of decimal system: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

However, the attribution of these expressions, based essentially on fuzzy terms, could vary from one operator to another. Indeed, what might seem "weak" for example for one person could appear "very dim" to another and inversely. So, several different encode could be obtained for the same variety. In addition, some character necessitates a description where intervenes several linguistic expressions. So, They could not be expressed only by unique encodes of decimal system. The following table illustrate such descriptions.

CHARACTER	DESCRIPTION
coloration	middle, red lightly roseate
shape	rounded, lightly flattened

In this paper, we propose a new technique of codification that permits to overcome these inconveniences. The purpose of this method is to take in consideration the natural language nuances and the fuzziness that characterizes the human reasoning, without disrespecting the UPOV's norm. So, we warranted a means of more flexible encoding that translates as well the linguistic descriptions.

2. Coding with intervals

The objective of this method is of describing the characters objectively and naturally, in taking into account the fuzziness that characterizes the words of human language. Its principle is simple and intuitive. It consists simply in replacing in ambiguous cases, a unique number by an interval The following table is given as an example.

CHARACTER	DESCRIPTION	CODE
caliber	middle, small	[3,5]
shape	lightly flattened, flattened	[1,3]

For no ambiguous descriptions, which consists simply in the presence or the absence of a character, the two Boundaries of interval coincident. This case brings us back to the classic case of encoding by numbers. So, the classic method is generalized.

In consequence, the introduction of the new codification necessitates the elaboration of metric or similarity in order to compare two intervals. This comparison must take in consideration the difference of their size, the remoteness of their center and the width of their intersection. In consequence, we defined three new fuzzy concepts in order to improve the performances of data analysis's methods: "fuzzy inclusion degree", "fuzzy resemblance degree", and "fuzzy curve".

2.1. Concept of fuzzy inclusion degree

ZADEH [2], has introduced the theory of fuzzy sets in 1965. He defined the fuzzy membership degree where an element belongs to a set neither by yes nor by no, but with a degree who is often submitted to a subjective attitude.

In this part, we propose a fuzzy inclusion degree like the fuzzy membership degree where the inclusion of an interval in another interval is neither by yes nor no, but with a degree submitted to some parameters. In addition, the best functions of fuzzy membership depend of the distance $D(x, x_0)$ [3] of an arbitrary element x to the reference element x_0. So, we must formulate necessary an adequate distance between the intervals who takes into account all their features. Then, we use the diversity of the fuzzy functions formulated in the applications of the theory of fuzzy sets.

Problem's position: Let $I_1=[a,b]$ the interval describing a given character C of reference variety V_R, and $I_2=[x,y]$, the one representing the same character C for a second variety V.

Let $L_1=(b-a)$, the width of I_1, and $L_2=(y-x)$, the one of I_2, O_1 the center of I_1, and O_2 the one of I_2.
The similarity between VR and V, relatively to the character C, requires a measure of the inclusion degree of I_2 in I_1. We have two cases:
Case 1: $L_1=L_2$ (the intervals have the same size):
In this case, the Euclidean distance of two centers was sufficient. Indeed, the distance of I_1 to I_2 is accordingly smaller than O_1 brings together of O_2. So, the size of the intersection between the two intervals increases. Its minimal value, equal to 0 obtained when the centers O_1 and O_2 of two intervals coincident, therefore, where the size of the intersection is maximal.
Case 2: $L_1 \neq L_2$ (the intervals have different sizes):
In this case, the Euclidean distance of the two centers was not sufficient anymore. Indeed, all the intervals centered in O_2 will have the same distance to I_1 and therefore, the same degree of inclusion. It is incorrect because the intervals are different.

Solution's description: We propose the following fuzzy inclusion degree:

$$D_{inclusion}(I_1, I_2)=\frac{1}{1+D_{int}(I_1,I_2)^k}$$
(2.1.1)
or

$$D_{inclusion}(I_1, I_2)=e^{-k D_{int}(I_1,I_2)}$$
(2.1.2)

With: $D_{int}(I_1, I_2)$ is the distance that measures the dissimilarity between the intervals. It is a combination of the following three functions:

$$D_{int}(I_1,I_2)=f_L(L_1,L_2)(f_d(L_1,L_2)+d(O_1,O_2))$$
(2.1.3)

$$d(O_1,O_2)=(O_1-O_2)^2 \qquad (2.1.4)$$

$$f_d(L_1,L_2)=1-e^{-k(L_1-L_2)^2} \qquad (2.1.5)$$

$$f_L(L_1,L_2)=1+\log(1+(L_1-L_2)^2) \qquad (2.1.6)$$

We have:
(2.1.4) : measure the remoteness of the centers,
(2.1.5) : measure the degree of the width of the intersection,
(2.1.6) : measure the width of the difference of the sizes.

Remark: 1. These three functions are positive, symmetrical and verify the following condition:

$$f_d(x+y) \leq f_d(x)+f_d(y) \qquad (3.1.7)$$

This last condition would warrant the triangular inequality. So, we can consider one distance.
2. If L1= L2, This case brings us back to the Euclidean distance between two centers. So D_{int} generalizes the classic distance.

Case of slotted intervals: In the examples illustrated in the figure (2.1) we present some slotted intervals with different sizes in order to illustrate the result of fuzzy inclusion degree. We use the equation (2.1.2) for k= 0.2.
* Case 1: Let [10, 12] the reference interval centered on 11, and[x-1, x+1] the arbitrary intervals, when x varies in the middle of 1 and 19. We have the following results:
- If x=11, the maximal value of the curve of fuzzy inclusion degree is 1, when the two intervals with the same size coincident.
- If x≠11, the more x gets far from 11, the more the value of the curve (1) decreases.
* Case 2: Let [10-1, 12+1] the reference interval with the same center than the first case, and the same arbitrary intervals. We have the following results:
- If x= 11, the maximal value of the curve (2) of fuzzy inclusion degree is severely inferior to the maximal value of the curve (1) when the two centers coincident. Because, the width of the intersection of the second case is smaller than the first case.
- If x≠ 11, the more x gets far from 11, the more the value of the curve (2) decreases.
* Case i: We consider [10- (i-1), 12+ (i-1)] as a reference interval with the same center than the first cases, and the same arbitrary intervals. We have the following results:
- For x= 11, the maximal value of the curve (i) of fuzzy inclusion degree is severely inferior to the maximal value of the curve (i-1) when the two centers

coincident, because the width of the intersection in this case is smaller than those in the (i-1) firsts cases.
- For $x \neq 11$, the more x gets far from 11, the more the value of the curve (i) decreases.

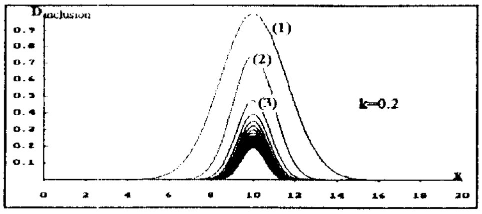

Figure 2.1. The curves of fuzzy inclusion degree

So, fuzzy inclusion degree of I_1 in I_2 gets as greater as I_1 becomes closer to I_2. Its maximal value (≤ 1) is reached when the centers O1 and O2 of the two intervals coincide. It is equal to 1 when the two intervals are exactly identical, and more the difference of the sizes is important, more this value is decreases.

The influence of k: In this part, we illustrate the influence of k on the aspect of curve of fuzzy inclusion degree for both the equations (2.1.1) and (2.1.2).
We have the reference interval [a, b]= [0, 2] centered at the beginning at 1. Then, this center baffled for each curve by increasing the value of b.

The study of the influence of k allows us to define the level of bringing together of the curves (1), (2), … , in order to choose the case that arranges us.

Indeed, if we consider the curves(a) and (c) of the figures (2.2) and (2.3), we will note that the curves of the equation (2.1.2) tend to bring together when k is very small. This is logical since the exponential function increases and decreases quickly.

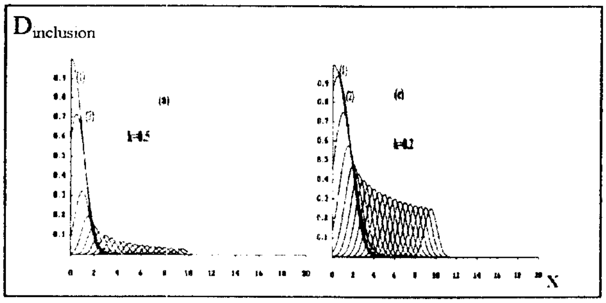

Figure 2.2. The influence of k for the equation (2.1.2)

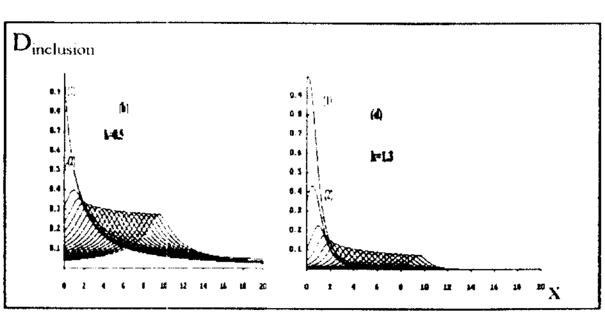

Figure 2.3. The influence of k for the equation (2.1.1)

For the two other cases (b) and (d) for the equation (2.1.1) of this figure we have the same case. The alone difference resides in the fact that the decreasing of the curves of the equation (2.1.2) is more homogeneous than the one of the equation (2.1.1). Indeed, the curves (1) and (2) of the equation (2.1 1) is nearly at the Same level for the two values of k, while that of the equation (2.1.2) has a decreasing nearly proportionally to k.

Remark: The number of curves no null decreases for the intervals of different size, if k becomes infinite. So for the equation (2.1.2) it is necessary to take k between 0 and 1, and for the equation (2.1.1) k must vary between 0 and 3.

2.2. Concept of fuzzy resemblance degree

The reasoning made in the previous paragraph concerns a unique character. Nevertheless, it permitted us simply of illustrating the concept of fuzzy inclusion degree. Really, the comparison between two varieties necessitates to take in consideration all characters. We use a fuzzy resemblance degree as we compare the resemblance of two sets and not an element and a set represented by his vector center like the case of membership degree.

Definitions: 1. The fuzzy resemblance degree D_R is the generalization of the fuzzy inclusion degree to the case where we consider a number n of intervals. Let V_R the reference variety defined by the intervals I_i, and V the variety of test defined by Ji with $1 \leq i \leq n$. The fuzzy resemblance degree is definite by the following's relationships:
- If we consider a vector with n intervals:

$$D_R(V_R,V)=\frac{1}{n}\sum_{i=1}^{n} D_{inclusion}(I_i,J_i)$$

(2.2.1)
- if we consider a fuzzy curve (see below):

$$D_R(V_R,V)=\frac{1}{n-1}\int_1^n D_{inclusion}(I_i,J_i)(x)\partial x$$

(2.2.2)
2. Let I a set of classes, and i an arbitrary element of I. A fuzzy set of fuzzy resemblance degree of I is a set of couples (i, $D_{R_A}(i)$) where $D_{R_A}(i)$ measures the fuzzy resemblance degree of the class i with all the classes of I.

Remarks: 1. we can use the same notations and the same terminology of the theory of fuzzy set.
2. If the set of departure of the function of fuzzy resemblance is reduced to vectors, we recover in this case, the fuzzy membership's degree.

2.3. Concept of fuzzy curve

In order to facilitate the task of multidimensional reasoning to agronomist experts, we have proposed to represent graphically each variety with a curve in the following manner:

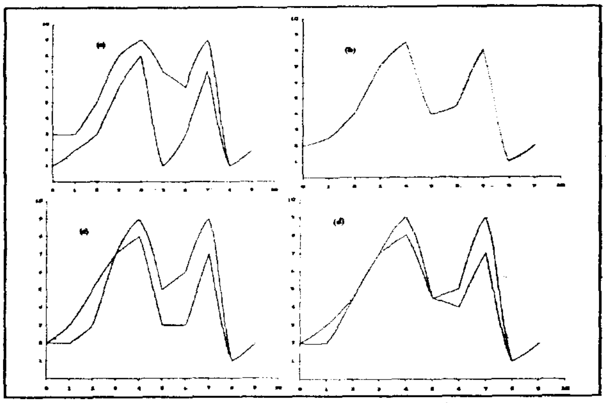

Figure 2.4. Examples of classic and fuzzy curve

The axis of X indicates the n° of the dimension, of the multidimensional space. The axis of the Y indicates the corresponding interval code. So, we make a projection of the multidimensional space on the plan. This operation is easier to imagine and to analyze.

The figure (2.4) illustrated the difference between the curves. A curve (b) is a classic one, but those of the curve (a), (c) and (d) are fuzzy. It is going to describe the best the data.

These fuzzy curves represent the best the reality. Indeed, if we represent the data with a classic curve, we lose some information. The examples of the figure (2.4) present three different fuzzy curves (a), (c) and (d), which has the same middle curve (b). These curves generated by neural networks [4][7].

3. Optimization of encoding by the densest intervals

The encoding technique by intervals introduces in the previous section facilitates pleasantly the data's description. It will serve as a basis for the creation of very superior systems in artificial intelligence. However, the fact of considering the uniform interval for describing a character could conceal some information. Indeed, we consider the following example:

The character "shape" of the SOLIDO variety (Nun 5062) of the species tomato[1], is described by the expression "lightly flattened, flattened, rounded," which the encodes is respectively 5, 3 et 7, so, by the [3,7] interval.

But, the shape's description "lightly flattened ", stake in the first position is rife than the "flattened" one, stake in the second position. This last description is rife than the "rounded" one. Thus, this last information is hidden in the interval. The following table illustrate such situations:

Character	Description	Codes	description's interval	densest Intervals
caliber	middle, small, big	5, 3, 7	[3, 7]	[5]
number of stalls	5-11 generally 6-7	6-7, 5-11	[5, 11]	[6, 7]

In this part, we propose a optimization of the previous encoding technique that permits to overcome this inconvenience. The purpose of the method is to describe objectively the data in taking in consideration all the information.

The method's principle is simple and intuitive. It's consists simply to add in ambiguous case one or several other subintervals to the interval of description. These added subintervals representing the densest parts, the minus densest, etc., according to the precision that we want. So, we can represent well the information.

In consequence, we must necessitate an adaptation of the new fuzzy concepts exposed in first section such: " fuzzy inclusion degree ", " fuzzy resemblance degree " and " fuzzy curve ".

3.1. Concept of fuzzy inclusion degree

This degree could be easily generalized in order to take in consideration the densest intervals. Indeed, let I_i=[a,b] be the interval describing a given character C for a reference variety VR, and J_i=[xi, yi], the one representing the same character C for a second variety V. And let L_i= (b_i- a_i), the width of I_i, and l_i =(yi - xi), the one of J_i, O_i the center of I_i, and G_i the one of J_i, with:

- $0 \leq i \leq n$,
- if $0 < j < i \leq n$, then I_i (respectively J_i) is denser than I_j (respectively J_j),
- I_0 and J_0 are the intervals of description who contain all the other densest subintervals.

The similarity between VR and V, relatively to the character C, necessitates the measurement of fuzzy inclusion degree of I_0 in J_0 with taking into account the other densest intervals. Let us consider the degree of inclusion between two any intervals definite in the previous section by the previous relations (2.1.1) and (2.1.2).

The Ddense inclusion degree that we suggest that takes into account the densest parts is a linear combination of the fuzzy inclusion degree for the all intervals describing a character. I.e., the interval of description and all the densest intervals. It is definite in the following manner:

$$D_{dense}(I_0,J_0)=\sum_{i=0}^{i=n}\mu_i D_{inclusion}(I_i,J_i) \quad (3.1.1)$$

under the constraint:

$$\sum_{i=0}^{n}\mu_i=1 \quad (3.1.2)$$

Where: μ_i is a number that determines the weight of the density for each interval in the resolution.

Examples of applications: We consider the following examples, where we have the same interval of description, but the densest parts are different. We use the equation (2.1.2) for K= 0.2. Also, we consider the case where we have n= 1, that is, where we have the interval of description and an other interval representing the densest part of this description's interval. We also consider the case where μ_i=0.5 for i= 0,1, i.e., when the weight of discrimination of the description's interval is equal to the one of the densest part. We have the following results:

The fuzzy inclusion degree without considering the densest parts gives 1 for all the description because the intervals of descriptions are identical. (See table 1)

The fuzzy inclusion degree with considering the densest parts gives us different values. These values depend on the position of the densest part in the description's interval. (See table 2)

So, this inclusion degree represents the best the reality.

Table 1 Classic fuzzy inclusion

$D_{INCLUSION}$	[4,6], [4,11]	[9,11], [4,11]	[6,8], [4,11]
[4,6], [4,11]	1	1	1
[9,11], [4,11]	1	1	1
[6,8], [4,11]	1	1	1

Table 2 Densest fuzzy inclusion

D_{DENSE}	[4,6], [4,11]	[9,11], [4,11]	[6,8], [4,11]
[4,6], [4,11]	1	0.503369	0.724446
[9,11], [4,11]	0.503369	1	0.582449
[6,8], [4,11]	0.724446	0.582449	1

3.2. Fuzzy resemblance degree

The fuzzy resemblance degree will stay the same since it will be based on Ddense inclusion degree. Indeed, we define it in the following manner:

The fuzzy resemblance degree D_R is the generalization of fuzzy inclusion degree to the case where we consider n intervals. Let VR the reference variety defined by the intervals codes I_i, and V the variety of test definite by J_i with $1\leq i\leq n$. The resemblance degree is definite by:

- If we consider a vector with n intervals:

$$D_R(V_R,V)=\frac{1}{n}\sum_{i=1}^{n}D_{dense}(I_i,J_i)$$

(3.2.1)

- If we considers a fuzzy curve:

$$D_R(V_R,V)=\frac{1}{n-1}\int_1^n D_{dense}(I_i,J_i)(x)\partial x$$

(3.2.2)

Remark: The D_{dense} degree can be defined differently for all the n intervals. In deed, we can consider for I_i and J_i x densest intervals and for I_j and J_j y densest intervals. The number of the densest Intervals of each description depends of the descriptions and of the precision that we want. The value of μ_i can be chosen like a weight that affect the resolution.

3.3. Concept of fuzzy curve

The introduction of the notion of the densest intervals made the curves fuzzy as illustrated in the figure (3.1), where the shaded part represents the densest part. These figures show four curves with the same fuzzy curve of description, but the densest part are different.

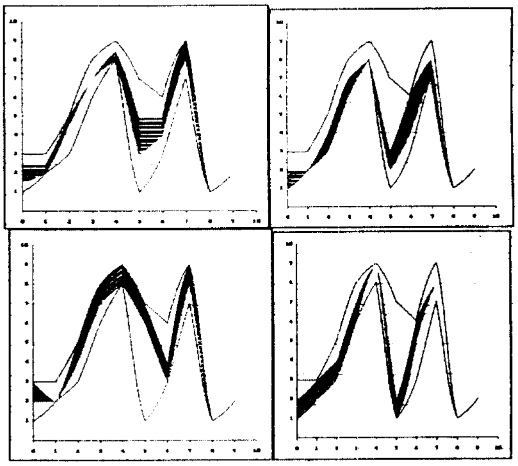

Figure 3.1: Examples of the new fuzzy curves

These curves are generated by the neural networks [4]. So, we see that these curves are different because their densest parts are different.

4. Conclusion

The results obtained with these new fuzzy concepts are encouraging. They open a domain of very hopeful research.

5. References and bibliography

[1] A. J. G. VANGASTEL
Morphological varietal description (Wheat)
The big blue book

[2] L. A. ZADEH
Fuzzy sets Information and control 8, pp 338_353, 1965
[3] R. KRISHNAPURAM, J. M. KELLER
A possibilistic approach to clustering IEEE, transaction on fuzzy system, vol.1, N° 2, may 1993
[4] F. KARBOU
Nouveaux concepts flous en analyse de données, pour le codage, la similarité et la représentation graphique des variétés agricoles CIMASI'96, pp 155-163, ENSEM, Casablanca, 14-16 Novembre (1996)
[5] J. M. KELLER, M. R. GRAY, J. A. GIVENS, A fuzzy K- Nearest Neighbor Algorithm, IEEE, Vol. SMC-15, N° 4, Juillet/Aout (1985), pp 580-585
[6] Michael P. WINDHAM, Cluster Validity for Fuzzy
Clustering Algorithm, Fuzzy sets and systems 5 (1981) 177-185, North-Holland Publishing Company
[7] YOH-HAN PAO, Adaptive pattern recognition and neural networks, Addison-Wesley, 1989.