

A New Way to Describe Spatiotemporal Data

Mingli Song

School of Computer Science
 Communication University of China
 Beijing, China
 sml607@yahoo.com.cn

Abstract—Data analysis especially dealing with spatiotemporal data in a human-centric way becomes essential in modern engineering and science fields. In this study, we present a granular way of data analysis where the data are described by information granules defined in spatial and temporal domain. We show that the data can be effectively described through a collection of Cartesian products of information granules and their convolution. The quality of the codebooks (vocabulary) is evaluated by data granulation-degranulation. Finally a synthetic example is given to illustrate our algorithm.

Keywords—granular descriptor; spatiotemporal data; information granules; Cartesian; granulation-degranulation

I. INTRODUCTION

In the literature, we can find a great deal of methodological and algorithmic pursuits as well as a wave of application developments of fuzzy sets. People use them to express fuzzy concepts on various kinds of data. In this paper, we concentrate on spatiotemporal data analysis by proposing a new granular descriptor. This descriptor is realized by the operations of Cartesian and convolution on fuzzy sets. To evaluate the quality of the predefined fuzzy sets, a process of granulation-degranulation is adopted which at the same time serves as a tool of optimization of fuzzy sets.

Spatiotemporal data become visible in numerous applications, such as sensors, oil wells and so on. There is not a comprehensive spatiotemporal data analysis with information granules yet in literature. We can use some selected methods to define some fuzzy sets on spatial data and temporal data respectively. And then execute operations of those fuzzy sets and obtain granular descriptors. The descriptor shows a degree of consistency. In the descriptor formula, there is an operator named t norm and s norm which have been studied a lot by many researchers.

In what follows, the definition of the granular descriptor is introduced. The process of granulation-degranulation is discussed realized with seven t norms and s norms. A Synthetic data example shows the effectiveness of our method. This paper also outlines some research topics worthy pursuing in the future.

II. A GRANULAR DESCRIPTOR OF SPATIOTEMPORAL DATA

In this paper, we define a granular description (descriptor) to evaluate the consistency of information granules used to represent spatial data and the associated temporal data. The

formula is realized by performing Cartesian and a convolution of the elements of vocabularies (information granules). This idea is quite commonly seen in signal processing as well as the technology of fuzzy sets [1, 2] that is:

$$g_{ijl} = (A_i \times B_j) \circ C_l \quad (1)$$

Assume that $i = 1, 2, \dots, n, j = 1, 2, \dots, m, l = 1, 2, \dots, r$. If we express the fuzzy sets of the codebooks in terms of membership functions A_i, B_j , and C_l , we obtain the values of g_{ijl} as a result of the max-t composition operator of those fuzzy sets. Please refer to (2)(3) for one dimensional spatial data and (4)(5) for two dimensional case.

$$g_{il} = \max_{x_k, z_k} (A_i(x_k) t C_l(z_k)) \quad (2)$$

or

$$g_{il} = \min_{x_k, z_k} ((1-A_i(x_k)) s C_l(z_k)) \quad (3)$$

$$g_{ijl} = \max_{x_k, y_k, z_k} (A_i(x_k) t B_j(y_k) t C_l(z_k)) \quad (4)$$

or

$$g_{ijl} = \min_{x_k, y_k, z_k} (1-A_i(x_k) t B_j(y_k) s C_l(z_k)) \quad (5)$$

The symbols “t” and “s” represent t norm and s norm [3] respectively. x_k is the x coordinate of the kth one dimensional spatial data; (x_k, y_k) are x and y coordinates of the kth two dimensional data. z_k is a time series of the kth data. g_{ijl} (g_{il}) is a number between 0 and 1 and is a measure of consistency of codebooks of spatial and temporal data.

Obviously, give different granular codebooks (vocabulary), we arrive at different description or view of the same data. The elements of g_{ijl} being arranged into a single fuzzy relation $G = [g_{ijl}]$ can be treated as a granular signature of the spatial-temporal pair data realized for a certain collection of generic information granules **A**, **B** and **C** [4]. It should be noted that G provides an important way of compression: the original data are reduced to granular data of $n \times m \times r$ dimensionality. The performance of codebooks can be ordered linearly depending upon the values of $G = [g_{ijl}]$; higher values of its entries indicate that the corresponding entries are more consistent.

In fact, the formulas (2) and (4) supply an upper bound of the consistency; whereas the formulas (3) and (5) offer lower bounds. Thus we can view an interval field which makes the result more flexible.

III. THE PROCESS OF GRANULATION-DEGRANULATION: A WAY TO EVALUATE QUALITY OF VOCABULARY

In the previous section we introduced a method to calculate a measure of granular description of spatiotemporal data. Those spatiotemporal data are firstly described by some codebooks (vocabulary). Hence, how to evaluate the vocabulary becomes one of our tasks in this paper. Given the descriptor G and the codebooks A and B, we can reconstruct C. The degranulation problem associates with a solution to the system of relational equation (1) in which fuzzy sets A_i , B_j as well as the entries of the fuzzy relation $G = [g_{ij}]$ are given while the fuzzy relation C_1 is to be determined. The solution comes as the maximal fuzzy relation C_{ul} or C_{ll} .

In terms of the membership values of the fuzzy relation, we obtain:

If C_1 is generated by (2)

$$C_{ul} = A_i \phi g_{ij} = \min_{ij} [A_i(x) \phi g_{ij}] \quad (6)$$

If C_1 is generated by (3)

$$C_{ll} = (1-A_i) \phi g_{ij} = \max_{ij} [(1-A_i(x)) \phi g_{ij}] \quad (7)$$

The pseudo-residuation operators associated with given t-norm and t-conorm are defined respectively in the following form

$$a \phi b = \sup \{v \in [0, 1] \mid a \vee v \leq b\} \quad (8)$$

$$a \phi b = \inf \{v \in [0, 1] \mid a \vee v \geq b\} \quad (9)$$

For two dimensional case:

If C_1 is generated by (4)

$$C_{ul} = (A_i \vee B_j) \phi g_{ij} = \min_{ij} [(A_i(x) \vee B_j(y)) \phi g_{ij}] \quad (10)$$

If C_1 is generated by (5)

$$C_{ll} = ((1-A_i) \wedge (1-B_j)) \phi g_{ij} = \max_{ij} [(1-A_i(x) \wedge (1-B_j(y))) \phi g_{ij}] \quad (11)$$

We can observe that $C_{ll} \subseteq C_1 \subseteq C_{ul}$. The quality of granulation-degranulation is expressed by comparing how much the reconstructed C_1 coincides with the original C_1 . The performance index can be defined as follows

$$Q = \frac{\sum_{k=1}^N \|C_{ll}^k - C_{ul}^k\|}{N} \quad (12)$$

The value of Q is minimized to obtain the best combination of fuzzy sets A_i and B_j . Due to the size of the solution space, generally an evolutionary method is considered here to compute the result.

To validate our algorithm and compare the performance of different operators, we study many papers to look for some well-known t norms and s norms [5, 6, 7, 8, 9]. Here we list seven t norms and the corresponding pseudo-residuation operators in the following.

1. Gödel: $atb = \min(a,b)$
2. Product: $atb = ab$
3. Lukasiewicz: $atb = \max(0, a+b-1)$
4. Weber: $atb = \max(0, (a+b-1+pab)/(1+p))$, $p > -1$, $p \neq 1$
5. Schweizer: $atb = (\max(0, ap+bp-1))/p$, $p > 0$, $p \neq 2$
6. Hamacher: $atb = ab/(2-a-b+ab)$
7. Frank: $atb = \log p(1+(px-1)(py-1)/(p-1))$, $p > 0$, $p \neq 1$, $p \neq 2$

The associated pseudo-residuation operators are:

1. $a \phi b = b$ if $a > b$
1 if $a \leq b$
2. $a \phi b = b/a$ if $a > b$
1 if $a \leq b$
3. $a \phi b = b+1-a$
4. $a \phi b = (2b-a+1)/(1+a)$
5. $a \phi b = (1+b^2-a^2)/2$
6. $a \phi b = ab/(a-b+ab)$ if $a > b/(1+b)$
1 if $a \leq b/(1+b)$
7. $a \phi b = \log_2 ((2b-1)/(2a-1)+1)$

The s norms of the corresponding t norms are:

1. Gödel: $asb = \max(a,b)$
2. Probabilistic sum: $asb = a+b-ab$
3. Lukasiewicz: $asb = \min(1, a+b)$
4. Weber: $asb = \min(1, ((1+p)(a+b)-pab)/(p+1))$, $p > -1$, $p \neq 1$
5. Yager: $asb = \min(1, (ap+bp)/p)$, $p > 0$, $p \neq 2$
6. Einstein: $asb = a+b/(1+ab)$
7. Frank: $asb = 1 - \log p(1+(p1-a-1)(p1-b-1)/(p-1))$, $p > 0$, $p \neq 1$, $p \neq 2$

And the pseudo-residuation operators are:

1. $a \phi b = 0$ if $a \geq b$
b if $a < b$
2. $a \phi b = 0$ if $a \geq b$
 $(b-a)/(1-a)$ if $a < b$
3. $a \phi b = 0$ if $a \geq b$
b-a if $a < b$
4. $a \phi b = 0$ if $a \geq b$
 $(2b-2a)/(2-a)$ if $a < b$
5. $a \phi b = 0$ if $a \geq b$
 $(b^2-a^2)/2$ if $a < b$
6. $a \phi b = 0$ if $a \geq b$
 $(b-a)/(1-ab)$ if $a < b$
7. $a \phi b = 1 - \log_2 (1+(21-b-1)/(21-a-1))$

IV. AN EXAMPLE

We choose a function y to illustrate our algorithm. Here y can be seemed as C_1 and $l = 1$ in this case (in Figure 1).

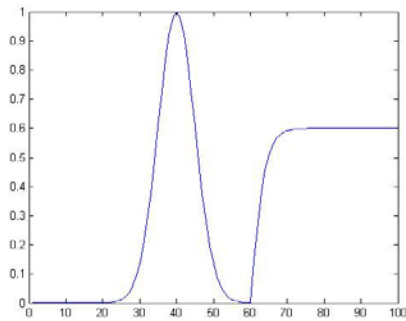


Figure 1. The original function y

We select the first five types of t norms and the corresponding s norms from the above list. Their pseudo-residuation formulas also can be found there. Three types of fuzzy sets are tried: intervals, triangular functions, asymmetric Gaussian functions. Now let's look at the results.

Table 1. The value of Q with different t norms, different types of fuzzy sets, and different numbers of fuzzy sets.

T norm no. of fuzzy sets	minimum	product	Lukasiewicz	Weber	Schweizer
intervals (3-9 fuzzy sets)	0.4055	0.4055	0.4055	0.4055	0.4055
	0.2826	0.2826	0.2826	0.2826	0.2826
	0.2043	0.2043	0.2043	0.2043	0.2043
	0.1742	0.1742	0.1742	0.1742	0.1742
	0.1617	0.1923	0.1617	0.1734	0.1753
	0.1143	0.1184	0.1548	0.1557	0.1665
triangular (3-9 fuzzy sets)	0.1591	0.1282	0.1183	0.1125	0.1164
	0.4177	0.4387	0.6033	0.5450	0.6468
	0.3490	0.3383	0.4430	0.3986	0.5416
	0.2901	0.2841	0.3245	0.3045	0.4488
	0.2193	0.1982	0.2969	0.2625	0.4225
	0.1989	0.1742	0.2774	0.2307	0.4221
Gaussian (3-9 fuzzy sets)	0.1738	0.1724	0.2709	0.2241	0.4116
	0.1625	0.1628	0.2663	0.2108	0.4041
	0.3214	0.2713	0.4337	0.3986	0.4618
	0.2399	0.1890	0.3085	0.3136	0.2986
	0.2669	0.1707	0.3053	0.1689	0.2024
	0.1756	0.1360	0.1796	0.0809	0.2126
	0.1838	0.0815	0.1393	0.1718	0.2307
	0.1873	0.0739	0.1456	0.0520	0.2451
	0.1484	0.0582	0.1509	0.0513	0.1678

We can see from the above table that the best result occurs in Weber t norm and 9 fuzzy sets. It is straight forward that the more fuzzy sets we use, the better performance we will get. However, the performance of t norms depends upon the problems at hand.

The two bounds of this best performance are shown in Figure 2. It is obvious that the two bounds are very close to the original function.

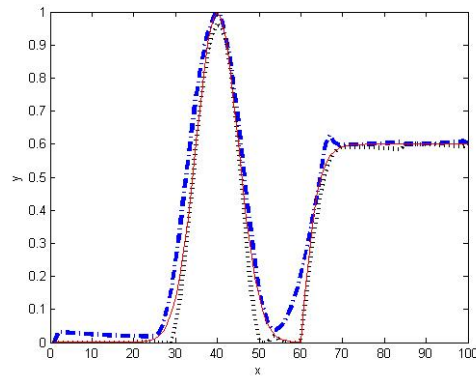


Figure 2. The original function and its two bounds

V. CONCLUSIONS

In this paper, we propose a way to develop a granular descriptor and evaluate the quality of the codebooks which are used to construct the operator. Seven popular t norms are adopted in our experiments. We can try more different t norms, different numbers of fuzzy sets, and different types of fuzzy sets in our future works.

ACKNOWLEDGMENT

This study is supported by Engineering Project of Communication University of China numbered XNG1239 and XNG1246.

REFERENCES

- [1] W. Pedrycz and F. Gomide. Fuzzy Systems Engineering: toward human-centric computing. John Wiley and Sons, 2007.
- [2] A. Bargiela and W. Pedrycz. Granular Computing: An introduction. Kluwer Academic, 2003.
- [3] E.Klement, R. Mesiar and E. Pap, Triangular Norms, Kluwer Academic Publishers, Dordrecht, 2000.
- [4] H. Nobuhara, W. Pedrycz, S. Sessa, K. Hirota, "A motion compression/reconstruction method based on max t -norm composite fuzzy relational equations", Information Sciences, vol. 176, pp. 2526-2552, 2006.
- [5] M. J. Frank, "On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$ ", Aequationes Math. vol. 19, pp. 194-226, 1979.
- [6] B. Schweizer and A. Sklar, Associative functions and statistical triangle inequalities, Publicationes Mathematicae Debrecen 8, pp. 169-186, 1961.
- [7] S. Weber, A general concept of fuzzy connectives, negations, and implications based on t -norms and t -conorms, Fuzzy Sets Syst. vol. 11, pp. 115-134, 1983.
- [8] R. R. Yager, On a general class of fuzzy connectives, Fuzzy Sets and Systems, vol. 4, pp. 235-242, 1980.
- [9] H. Hamacher, Über logische Aggregationen nicht-binär expliziter Entscheidungskriterien, Frankfurt/Main, 1978.