



Granulation of Temporal Data: A Global View on Time Series

Andrzej Bargiela

*School of Computing
and Mathematics*

*The Nottingham Trent University
Nottingham NG1 4BU, UK*

Witold Pedrycz

*ECERC
University of Alberta, Edmonton,
Canada T6G 2G7
&
Systems Research Institute, Polish
Academy of Sciences
01-447 Warsaw, Poland*

Abstract

In this paper we discuss the issue of granular representation of time series. The critical concern is the ability to capture the essential features of the time series in the abstract granular representation of it. The discussion uses a set-theoretical framework of fuzzy sets and employs the Fuzzy C-means algorithm for the evaluation of the information granules obtained in various ways.

Keywords: time-domain granulation, phase-space granulation, information density, FCM

1 Introductory comments

The purpose of information granulation is the organisation of detailed numerical data into abstract knowledge. In the case of temporal data, an essential characteristic is the time ordering of individual data items. This time ordering allows the analysis of relationships that occur between the various dimensions of the data stream.

In this paper we take a global view onto temporal data and attempt to uncover the relationships that might exist in such data. The challenge of analysing temporal data is stemming from the necessity of a synchronised analysis of concurrent data streams. In order to meet this challenge we propose to use granulation as an information abstraction mechanism that enables focusing on essential features and ignoring the irrelevant detail present in data. The quality of such an abstraction (granulation) is clearly related to the ability to retain the essence of the original data. We assess this quality by comparing the FCM prototypes that have been obtained with 3 different granulation schemes: time-domain granulation, phase-space granulation and information density based granulation.

To illustrate the operation of granulation algorithms we will apply them to a synthetic 2dimensional time series presented in Figure 1.

The time series contains four distinct data patterns; two of which are described by 100 data points and two others by 20 and 10 data points respectively.

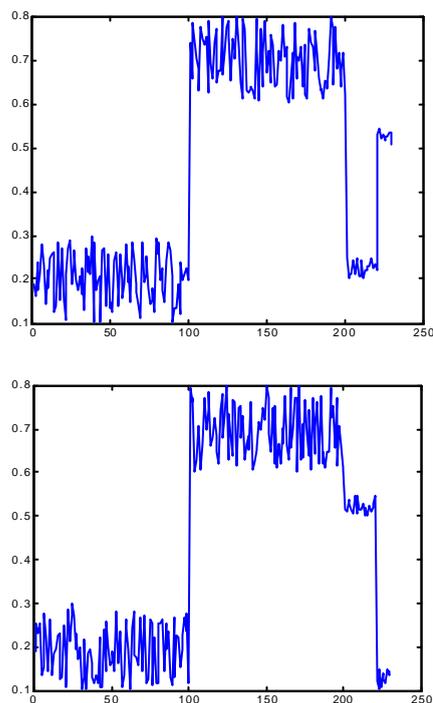


Figure 1. 2-dimensional synthetic time series.

2 Time-domain granulation

An intuitive approach to granulation of time series is to define a 'window of observation' and to evaluate an appropriate granular representative within each segment

of time series. This can be formalised as a representation of the time series as a fuzzy set and the use of possibility and necessity measure to derive characterisation of individual information granules. The attractiveness of this formalization derives from the sound theoretical foundations of the possibility measure and fuzzy relational equations. In particular, one can solve the problem of reconstruction of the fuzzy set (time series) in terms of the maximal feasible solution.

Similarly, the use of necessity measure on the fuzzy set of interest makes it possible to find the minimal feasible solution to the reconstruction problem.

Figure 2 illustrates the simple time-domain granulation of time series. Such a granulation can be seen as a mapping of the original data set $\mathbf{X} = \{x^1, x^2, \dots, x^N\}$ onto a set of intervals $\mathbf{I} = \{I^1, I^2, \dots, I^G\}$ where N is a number of elements in the time series and G is the number of granules.

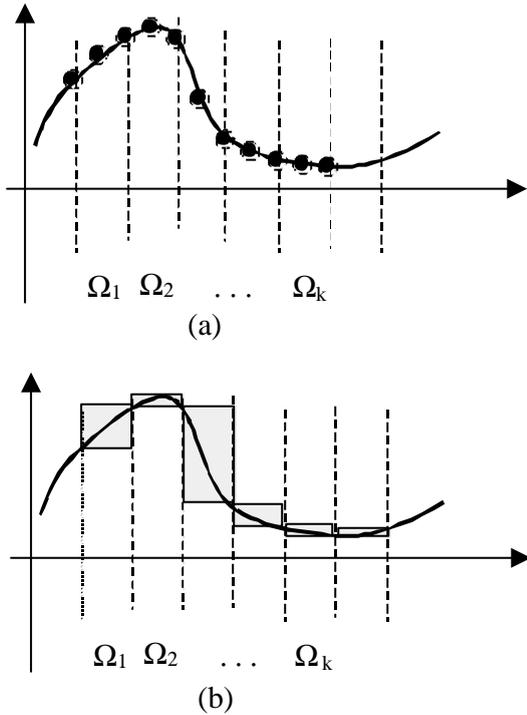


Figure 2. A simple time-domain granulation. a) time series; b) information granules formed with granulation window $\omega_0=3$.

Individual granules I^k are described as follows

$$I^k = \left(\min_{j \in \Omega_k} (x^j), \max_{j \in \Omega_k} (x^j) \right) \quad (1)$$

where

$$\Omega_k = \{i : w_o(k-1) + 1 \leq i \leq w_o k\} \quad (2)$$

and $k=1, 2, \dots, G$, $w_o G \leq N$, ω_0 is a granulation window. The expressions (1)(2) are easily generalized to multi-dimensional time series by applying the min- and max- operations to all coordinates of the original data. In which case we have

$$I^k = \left(\begin{array}{c} \min_{j \in \Omega_k} (x_1^j), \max_{j \in \Omega_k} (x_1^j), \min_{j \in \Omega_k} (x_2^j), \max_{j \in \Omega_k} (x_2^j), \\ \dots, \min_{j \in \Omega_k} (x_n^j), \max_{j \in \Omega_k} (x_n^j) \end{array} \right) \quad (3)$$

It is clear that the mapping of \mathbf{X} onto \mathbf{I} involves the increase of dimensionality of the pattern space from \mathbf{R}^n to \mathbf{R}^{2n} , where n is a dimension of patterns. The set of intervals \mathbf{I} represents now a granulated information from the original time series. In order to assess the effectiveness of such a granulation we select granulation window $\omega_0=2, 3, 5, 6, 8$, and 25 and perform the granulation of time series from Figure 1.

One thing that is immediately obvious with these granulations is that the results are extremely sensitive to the choice of the granulation window ω_0 . If the window of observation is well aligned with the boundaries of significant changes in the time series, as is the case for $\omega_0=2$ and $\omega_0=5$ ($100, 20$ and 10 divide by 2 and 5), the resulting granulation gives a good abstraction of the original data. However, in a more typical case, when the window of observation includes data that belongs to two different data groupings, the time-domain granulation generates large, unrepresentative granules.

A refinement of the simple time series granulation approach has been proposed by Das (Das, 1998). The extended method considers fixed-length subsequences of the series rather than just individual data items. The sub-sequences are represented as data-points in the augmented input space that has dimension defined by the length of the sub-sequences (ω_0). Sub-sequences that have similar 'shape' are represented as nearby points in the augmented space and can be clustered using some appropriately defined distance function.

The approach generalizes to multi-dimensional time series. In this case the sub-sequences q^j are formed by patterns $\hat{x} \in \mathbf{R}^n$, where n is a dimension of individual patterns. They are therefore elements of $(\omega_0)^n$ - dimensional space, where ω_0 is a number of patterns formed from ω_0 sequences in each dimension (typically $\omega_0=4$ for $\omega_0=3$). The clusters of sub-sequences are therefore hyperboxes in $\mathbf{R}^{2(\omega_0)^n}$. Unfortunately, the exponential increase of the dimensionality of the 'shape-space' makes this approach impractical for many real-life situations.

3 Phase-space granulation

An alternative approach to capturing the nature of sub-sequences of time series, that avoids undue augmentation of the dimension of the input space, is the phase-space granulation. We characterize the ‘shape’ of sub-sequences by a range of gradient angles between the first and every other pattern in the sub-sequence as shown in Figure 3.

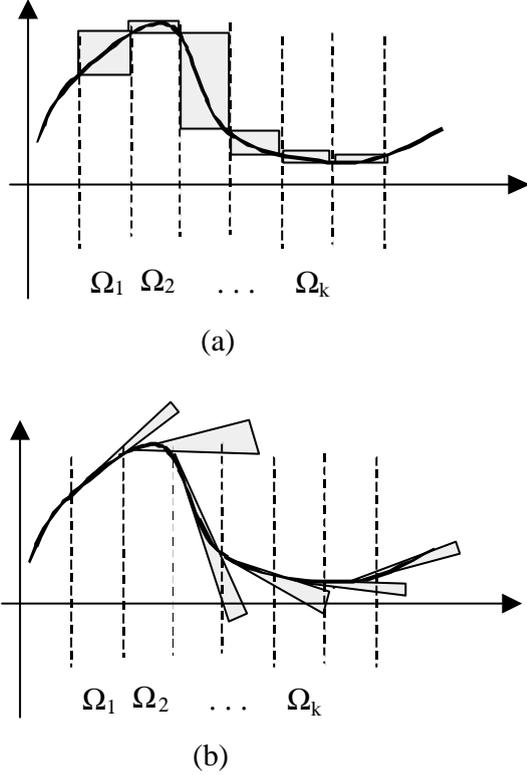


Figure 3. Phase-space granulation: a) intervals of time series values, b) intervals of gradient angles in the granulation window.

This results in an interval (hyperbox) description of ‘shape’ that is fully compatible with the interval (hyperbox) description of the time series values, as defined in (1) - (3). The advantage of this granulation is that subsequent clustering does not imply any further increase of the input space dimension.

Since the intervals of time series values in each granulation window are already contained within $[0, 1]$ range, only the intervals of gradient angles need to be normalized from $[-\pi/2, \pi/2]$ to $[0, 1]$.

We can formalize the phase-space granulation as a mapping of the original data set $\mathbf{X} = \{x^1, x^2, \dots, x^N\}$ onto a set of hyperboxes $\mathbf{H} = \{H^1, H^2, \dots, H^G\}$ where N is a number of elements in the time series and G is the number of granules. A hyperbox \tilde{H} is formed as a

Cartesian product of two intervals; $H^k = I^k \times J^k$, where I^k is an interval of time series values and J^k is an interval of gradient angles in the k -th granulation window.

$$I^k = \left(\min_{j \in \Omega_k} (x^j), \max_{j \in \Omega_k} (x^j) \right) \quad (4)$$

$$J^k = \left(\min_{j \neq l, j, l \in \Omega_k} (norm(grad(x^j, x^l))), \max_{j \neq l, j, l \in \Omega_k} (norm(grad(x^j, x^l))) \right) \quad (5)$$

where $grad(\cdot)$ is an angle-valued gradient function and $norm(\cdot)$ is a normalization function. The granulation window Ω_k is defined as in (2) and the generalization of the granulation to a multi-dimensional time series is analogous to (3). In this general case, the dimension of the input space is $4n$ (where n is a dimension of x^k) and the subsequent clustering of hyperboxes H^k does not imply any further increase of the dimension of the pattern space ($H^k \in \mathbf{R}^{4n}$). It is worth emphasizing that increasing the width of the granulation window ω_b reduces the number of granules to N/ω_b while maintaining the dimensionality of the input space. Consequently the computational complexity of the subsequent FCM is reduced by a factor $(\omega_b)^2$. This is in a sharp contrast to the granulation proposed in (Das, 1998) where the increase of the width of the granulation window reduces the number of input patterns to N/ω_b , but it increases the dimensionality of the input space by a factor ω_b thus increasing the computational complexity of FCM also by a factor ω_b .

Tests performed on the synthetic time series data indicate that while the inclusion of the gradient of the time series goes some way towards filtering out unrepresentative granules, but the quality of the resulting granules is broadly comparable to the results obtained with simple time-domain granulation. We quantify that statement by FCM clustering of the resulting granules in Section 5.

4 Information density based granulation

Building interval-valued granules arises as a compromise between two evidently conflicting requirements

- i) the interval should "embrace" as many elements of $\{x_j: j \in \Omega_k\}$ as possible (to be a sound representation of the window of observation)
- ii) the interval should be highly specific. This translates into the requirement of a minimal length of this interval (set).

As far as the first requirement is concerned, the cardinality of the set covering elements of Ω_k is a suitable criterion, that is

$$\text{card}(I) = \sum_{x_i \in X} c_{[a,b]}(x_k) \quad (6)$$

where $I = [a,b]$ denotes the interval we are about to construct and $\chi_{[a,b]}$ stands for its characteristic function, that is

$$\chi_{[a,b]}(x) = \begin{cases} 1, & \text{if } x \text{ is in } [a, b] \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

The specificity of the interval can be directly associated with its width (a suitably defined function of width),

$$\phi(\text{width}([a,b])) \quad (8)$$

We can therefore take a ratio of (6) and (8)

$$\sigma = \frac{\text{card}(I)}{\phi(\text{width}(I))} \quad (9)$$

and determine the interval I so that it maximizes expression (9). In this way, we cope simultaneously with maximization of cardinality (criterion ‘‘i’’) and minimization of width (criterion ‘‘ii’’). We refer to the optimization expressed by (9) as maximization of ‘information density’ of granules. This is to distinguish it from the concept of ‘data density’ that is typically represented as a ratio of cardinality of a given set over the volume of the pattern space containing this set. Consequently ‘data density’ is not defined for a single numeric data (zero volume in pattern space).

In order to achieve a dimensionally-invariant optimisation we consider the following mapping $\phi(u)$.

$$\phi(u) = \exp(\max_i (u_i)) * \exp(\max_i (u_i) - \min_j (u_j)) \quad (10)$$

where $u = [u_1 \ u_2 \ \dots \ u_n]$, $u_i = \text{width}([a_i, b_i])$ and $i, j = 1, 2, \dots, n$, is an index of the dimension of the pattern space. The first exponent function in (10) ensures that the specificity of information granules is maximized through the reduction of the maximum width of the hypercube along all dimensions in the pattern space. The second exponent in (10) ensures that the hyperboxes are as similar to hypercubes as possible. Also, the function satisfies the original boundary condition $\phi(0) = 1$, since for the point-size data $\max_i (u_i) = \min_j (u_j) = 0$, $i = 1, \dots, n$.

The optimization-based granulation of data is carried out as a one-pass simulation process:

- a) Normalize data to a unit hypercube;
- b) Initialize data structures;
- c) Calculate and store the value of ‘information density’ matrix D of size $N \times N$, where N is the cardinality of the input data set.
- d) Find a maximum entry in D ;

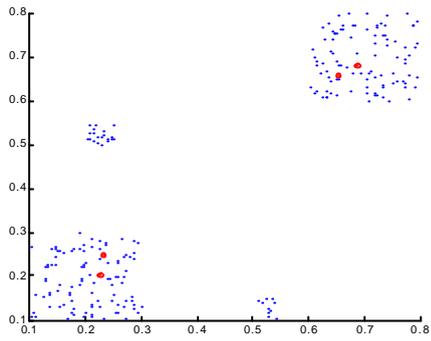
- e) If the maximum corresponds to an off-diagonal element (i -th and j -th coord):
 - merge the two granules (identified by the i -th row and j -th column)
 - update the cardinality of the resulting granule;
 - update the i -th row and column and remove the j -th row and column from D ;
 - return to d)
- f) If the maximum corresponds to a diagonal element ($i=j$):
 - copy the granule to an output list and remove the corresponding row and column from matrix D ;
 - if the size of matrix D is greater than 1, return to d), otherwise terminate.

Computational complexity of this granulation algorithm is $O(N^2)$ owing to the computations of matrix D in step c). However, unlike the clustering techniques (such as FCM), the granulation process implied by the above algorithm has an inherently local character and can be easily applied to a partitioned input data thus circumventing the high computational cost associated with large data sets. It is worth pointing out that the size of matrix D is being reduced by one row and column in each iteration thus the number of iterative steps equals $N-1$.

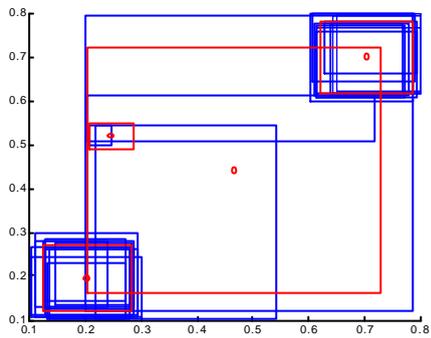
Since the algorithm maintains linear computational complexity with respect of the input space dimension (not to be confused with the complexity with respect of the cardinality of the data set which is $O(N^2)$), it is particularly suitable for processing multi-dimensional data. Also its ability to deal easily with the partitioned pattern space opens a possibility of achieving significant computational gains. Furthermore the uniform representation of data points and data granules enables application of the algorithm to the partially granulated patterns to arrive at the globally optimal granulation.

5 Evaluation of granulation through FCM clustering

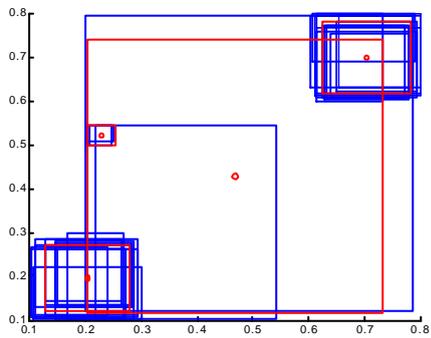
We assess the quality of granulation, discussed above, by identifying FCM-prototypes in original data and in the corresponding sets of granules. In order to do that we introduce a uniform representation of data points as hyperboxes in which the minimum and maximum hyperbox vertices coincide. With this modification the FCM returns prototypes that can be directly interpreted in the input data space. Figure 4 depicts typical results of FCM clustering of the information granules identified by methods described in Sections 2-4.



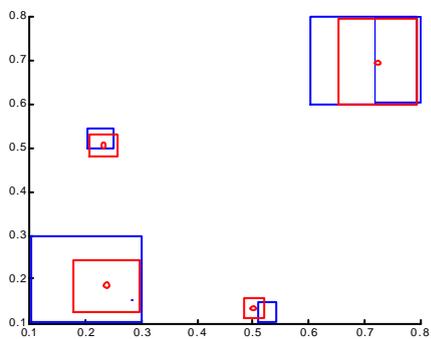
(a)



(b)



(c)



(d)

Figure 4. FCM-prototypes identified for the original time-series data a), time-domain granulated data b), phase-space granulated data c) and granules found through the optimisation of the information density.

It is clear from the results that this particular time series data having two large and two relatively small data groupings gives problem to FCM in that it results in FCM finding prototypes only inside the large data groups. In this sense the granulation that reduces the imbalance of cardinality of these data groupings produce an improvement in the sense of finding prototypes also inside a smaller data groupings. However, the mis-alignment of the granulation window with the boundaries of the significant changes in the time-series results in large information granules produced by both time-domain and phase-space granulation. This leads to large, uninformative FCM prototypes. By contrast, the granules produced by the optimisation-based granulation are much more compact and consequently the FCM prototypes identified from them are both highly specific and cover all four data groupings.

6 Conclusions

We have discussed three approaches to granulation of time-series data. The intuitive time-domain and phase-domain granulation was shown to require precise alignment of the granulation window with the significant changes in the data. If such an alignment is not performed the methods return a generally poor results. The optimisation of the information density approach results in a much-improved granulation that exhibits several desirable features:

- information granules are compact;
- small data groupings are fully taken into account;
- the local nature of optimisation allows for distributed computations (thus overcoming the quadratic computational complexity with respect to the cardinality of the data set);
- the linear computational complexity with respect of the dimensionality of the pattern space makes it suitable for highly-dimensional data;

Acknowledgements

The support from the Engineering and Physical Sciences Research Council (EPSRC, UK), the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Alberta Software Engineering Research Consortium (ASERC) is gratefully acknowledged.

References

- [1] Bargiela, A., Pedrycz, W., Recursive information granulation, *IEEE Transactions SMC-B*, 33, 1, 2003, 96-112.
- [2] Pedrycz, W., F. Gomide (1998), *An Introduction to Fuzzy Sets: Analysis and Design*, MIT Press, Cambridge, MA.
- [3] Zadeh, L.A. (1979), Fuzzy sets and information granularity, In: M.M. Gupta, R.K. Ragade, R.R. Yager, eds., *Advances in Fuzzy Set Theory and Applications*, North Holland, Amsterdam, 3-18.
- [4] Zadeh, L.A. (1996), Fuzzy logic = computing with words, *IEEE Trans. on Fuzzy Systems*, vol. 4, 2, 103-111.
- [5] Zadeh, L.A. (1997), Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, 90, 111-117.
- [6] Das, et al., (1998), Rule discovery from time series, 4th Int. Conf. On Knowledge Discovery and Data Mining, 16-22.