An Entropy-based Uncertainty Measure for Developing Granular Models

Muhammad Zaiyad Muda and George Panoutsos
Department of Automatic Control and Systems Engineering
The University of Sheffield
Sheffield, UK
mzmuda1@sheffield.ac.uk

Abstract— **There are two main ways to construct Fuzzy Logic rule-based models: using expert knowledge and using data mining methods. One of the most important aspects of Granular Computing (GrC) is to discover and extract knowledge from raw data in the form of information granules. The knowledge gained from the GrC, the information granules, can be used in constructing the linguistic rule-bases of a Fuzzy-Logic based system. Algorithms for iterative data granulation in the literature, so far, do not account for data uncertainty during the granulation process. In this paper, the uncertainty during the data granulation process is captured using the fundamental concept in information theory, entropy. In the proposed GrC algorithm, data granules are defined as information objects, hence the entropy measure being used in this research work is to capture the uncertainty in the data vectors resulting from the merging of the information granules. The entropy-based uncertainty measure is used to guide the iterative granulation process, hence promoting the formation of new granules with less uncertainty. The enhanced information granules are then being translated into a Fuzzy Logic inference system. The effectiveness of the proposed approach is demonstrated using established datasets.**

Keywords-Granular models; Fuzzy Logic; Information Theory; Granular Computing

I. INTRODUCTION

One of advantages of Fuzzy Logic (FL) based systems is interpretability due to their linguistic nature. The interpretability feature comes from the linguistic interpretation in the form of if-then rules that is close to human-like thinking. However, this approach presents some challenges because the expert knowledge to establish the rule-base system is not always available. To overcome this problem, knowledge discovery from data can be utilized to establish the FL parameters, and data clustering is one effective way to implement this [1].

Many clustering methods have been implemented in the literature to construct fuzzy rule-bases, such as fuzzy c-means [1] and hierarchical clustering [2]. Data clustering is a tool that categorizes similar data into groups, and one of the most commonly used techniques is hierarchical clustering. A granular computing (GrC) algorithm proposed by Bargiela and Pedrycz (2002), also known as granular clustering, resembles the concept of agglomerative hierarchical clustering [3]. However, there are significant differences between these algorithms. In GrC there is strong linkage between original data and the information granules. This is because every granule comprises of sub-granules [4] with each one originating directly from the data. In addition, the compatibility measure in GrC is a very important tool that can be used as guidance to terminate the clustering process. The decrement of compatibility measure can be visualized throughout the iterative granulation process, indicating the dissimilarity between granules towards the end of granulation process.

One of the most important aspects in GrC is the concept of information granulation. It is implemented through data organization and data comprehension [3]. Other than expert knowledge, FL models can be developed using data mining technique and this is where GrC plays its important role, to discover and extract knowledge by forming the information granules from the data. GrC mimics human reasoning in gathering objects that have similarities. A FL model can be formed using GrC in a transparent manner [4]. Transparency here refers to the link between data and information granules in the knowledge extraction process, and the use of this knowledge to build the linguistic rule-bases, from which we get certain decision or prediction (via some inference mechanism).

However, in complex real-world problems, information is always related with uncertainty. According to Klir (1995), uncertainty is resulted from information deficiency, i.e. incomplete, imprecise, vague or contradictory [5]. Most database systems are capable of handling ideal data, while the data that we encounter everyday are exposed to uncertainty. Therefore, there is a need to incorporate the data uncertainty in GrC, more particularly in the data granulation process. In this paper, the uncertainty during the data granulation process is captured using the concept in information theory, entropy.

In the case of uncertain data, there is a tendency of two dissimilar granules to be merged. This is because in most clustering algorithm, the geometrical distance is used as the similarity measure [6]. This will affect the distinguishability of the initial granular framework, the estimation of the FL parameters, and hence, the accuracy of the prediction or classification. In the presence of outliers, information granule cannot represent the actual data [7]. This will deter the interpretability of FL rule-base, and lead to inaccurate representation of the system under investigation. As the consequence, the FL rules being formed will be contradictory and inconsistent. Even though Fuzzy Logic is well known for its capability in modeling uncertainty, it is important to highlight that Fuzzy Logic mainly handles the kind of uncertainty termed as vagueness where the information is dealing with fuzzy quantifiers, e.g., most, many, few, almost all, etc. [8]. The novelty of this work is that for the first time in the literature, information theory is used as an assisting metric during iterative data granulation in order to construct Fuzzy-Logic rule-bases. Entropy is used to measure the uncertainty during the granulation process, hence constructs higher quality information granules. The result shows an enhanced performance as compared with a previous modelling framework based on the same datasets.

II. ITERATIVE GRANULATION

GrC is a framework that imitates human reasoning in grouping things [9]. The outcome of information granulation are information granules that consist of meaningful knowledge representing the data space. The GrC algorithm in this paper is performed by iteratively searching for two most compatible information granules and repeating this step until acceptable level of granulation has been accomplished [3]. The extracted knowledge and granular features are then translated into linguistic rule-base of a fuzzy system. GrC resembles the strategy of agglomerative hierarchical clustering by treating every data point as a single granule at the first iteration, but different in terms of its compatibility measure. Compatibility measure, in this case, refers to how close the information granule to the other. In GrC the compatibility measure C between two granules A and B is defined as:

$$C(A,B) = Distance_{MAX} - Distance_{A,B}.\exp(-\alpha R) \qquad (1)$$

In which

$$Density\ R = \frac{C_{A,B}/Cardinality_{MAX}}{L_{A,B}/\ Length_{MAX}} \qquad (2)$$

$Distance_{MAX}$ is the sum of maximum distances in every dimension i, given by:

$$Distance_{MAX} = \sum_{i=1}^{d}(distance_i\ ) \qquad (3)$$

$Distance_{A,B}$ is the distance between granule A and B, given by:

$$Distance_{A,B} = \frac{\sum_{i=1}^{d} w_i(D_1 - D_2)}{\mathrm{d}} \qquad (4)$$

Where

$$D_1 = \max(max_{Ai}, max_{Bi}) \qquad (5)$$

$$D_2 = \min(min_{Ai}, min_{Bi}) \qquad (6)$$

w_i: weight for dimension i, max_{Ai}: maximum value of dimension i in granule A, max_{Bi}: maximum value of dimension i in granule B, min_{Ai}: minimum value of dimension i in granule A, min_{Bi}: minimum value of dimension i in granule B, α: weightage of distance and density ranges from 0 to 1, $Cardinality_{MAX}$: the maximum number of objects in the data space, $Length_{MAX}$: sum of the maximum length in every dimension, $C_{A,B}$: the sum of cardinality of granule A and B, and $L_{A,B}$: length of the granule A and B, given by:

$$L_{A,B} = \sum_{i=1}^{d}(max_{Xi} - min_{Xi}) \qquad (7)$$

where max_{Xi} and min_{Xi} are the maximum and minimum length of the resulting granule at each dimension i, respectively.

After one iteration, a compatibility matrix will be created. It consists of compatibility index for each projected merging. The two granules with the highest compatibility index are considered the most compatible, hence will be merged. The compatibility index also plays role as a monitoring tool for the whole granulation process.

Fig. 1 shows the typical plot of the evolution of the compatibility index throughout the iterative granulation process based on the well-known Iris data. Iris data contains three main categories, namely Setosa, Versicolour and Virginica of 50 instances each. The x-axis represents the number of iteration while the y-axis represents the maximum compatibility index for each iteration. The small gradient of the curve at the early stages of granulation process indicates merging of compatible granules, while merging of incompatible granules is indicated by the large gradient at the end of granulation. The gradients of the curve are shown by the dotted lines, and the intersection of these two gradient lines can be used to approximate the optimal number of clusters. An example of compatibility matrix is shown in Table I.

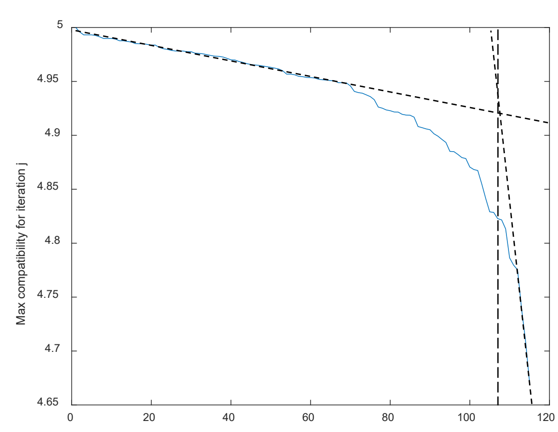


Figure 1. Maximum compatibility index

TABLE I. EXAMPLE OF COMPATIBILITY MATRIX

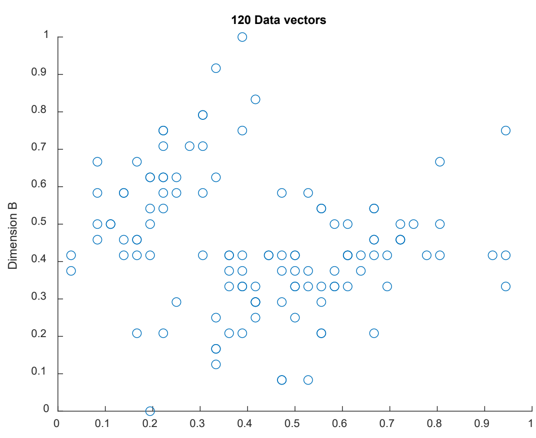
	G1	G2	G3	G4	G5
G1	-	4.41	4.48	4.2	4.15
G2	-	-	4.6	4.57	4.6
G3	-	-	-	4.61	4.53
G4	-	-	-	-	4.7
G5	-	-	-	-	-

In Table I, the two most compatible granules are granule 4 (G4) and granule 5 (G5) with compatibility of 4.7, and hence, these two granules will be merged. In the next iteration, the compatibility index will be computed from four granules: G1, G2, G3 and (G4+G5). The merging process is done iteratively until a satisfactory level of granulation is achieved.

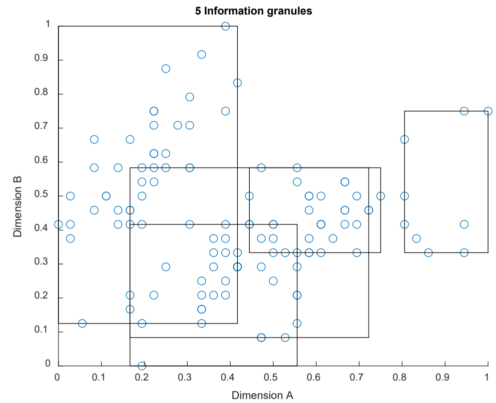
In this research all attributes are treated with same importance, hence the w_i for all attributes is one. However, this weightage can be adjusted in the case of different importance among the features.

In the first iteration, every data point is considered as one granule, meaning that the maximum and minimum boundaries are the same. Information granules are then formed in the form of hyper boxes. Since the compatibility index considers the maximum and minimum points during merging of two granules, a data point will always present at the boundaries of the hyper boxes.

Fig. 2 shows a granulation process, starting from 120 data vectors of Iris training dataset and compressed to five information granules. These five information granules are then translated into five fuzzy rules. The details are elaborated in Section III. To see the relationship between the input and output, all the four-input data are granulated together with the output, hence this approach can be considered as supervised learning. However, this approach is also applicable in unsupervised learning. Out of these five dimensions, only two dimensions are shown in Fig. 2.



(a)



(b)

Figure 2. Granulation process from (a) 120 data vectors to (b) 5 information granules

III. FROM INFORMATION GRANULES TO FUZZY RULE-BASES

Each information granule represents one fuzzy rule-base [10]. In this study Gaussian membership functions (MFs) are used. The formation of Gaussian MFs requires two important parameters, namely centre of MFs c and sigma σ . Median and standard deviation of the data in granules are used to represent c and σ. The relationship between an information granule and a FL linguistic rule is based on one-to-one basis. The FL rule-base can be written as:

$$IF\ Variable_1\ is\ A\ and\ Variable_2\ is\ B\ THEN\ ...$$

$$....IMPLICATION \quad (8)$$

Fig. 3 shows how five information granules in Fig. 2(b) are being translated into five Fuzzy MFs. These MFs represent the rule-bases for Iris data which consists of four inputs and one output.

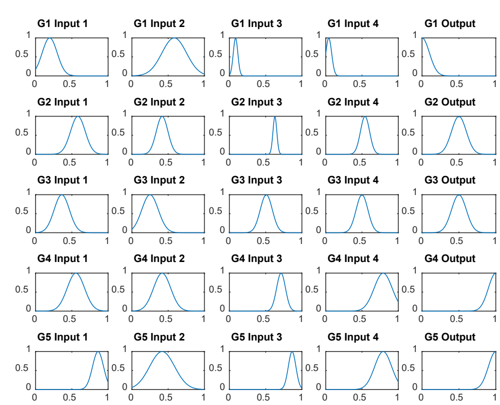


Figure 3. Converting information granules to fuzzy membership functions / rule-bases.

IV. USE OF INFORMATION THEORY TO MEASURE UNCERTAINTY

Information theory is one of the theories that concerns with quantifying uncertainty. One of the most important concepts in information theory is Shannon's entropy [11], that can be formulated as:

$$H(X) = -\sum_{x \epsilon X} p(x) log_2 p(x) \quad (9)$$

where $p(x)$ is the probability of occurrence of an event x and X is a random variable. In information theory, the dimension or feature is described as a random variable.

In GrC, the compatibility measure is used to determine the most compatible granules to be merged. However, there is uncertainty occurs throughout the granulation process. For example, in the case of outliers, the new formed granule is not capable to represent the actual data space. Besides, there is a tendency to create unnecessary overlapping. This will result in information granules with high uncertainty and hence, producing indistinguishable fuzzy rule-bases.

To overcome this issue, a new framework of iterative granulation is proposed, considering the uncertainty measure. In information theory, the measure of uncertainty in a random variable is described by Shannon's entropy in equation in (9). The concept of entropy is always used to represent disorder, chaos, or unpredictability in a dataset [12]. Hence, entropy in the proposed GrC compatibility measure characterizes the reluctance of granules to be merged. In other words, the target of the proposed method is to find a grouping of objects such that uncertainty is smallest.

As mentioned, entropy is a measure of uncertainty in a random variable. This is, however, not applicable for a random vector. Let X_i be the random variable, and $X = [X_1, X_2, X_3, \dots X_d]$ is considered the random vector. The entropy $H(X)$ is defined using the chain rule in [13] as:

$$H(X) = H(X_1, X_2, \dots X_d) = \sum_{i=1}^{d} H(X_i | X_{i-1}, \dots X_1) \quad (10)$$

For example, in the case of two random variables, the entropy of a random vector is given by:

$$H(X) = H(X_1, X_2) = H(X_1) + H(X_2 | X_1) \quad (11)$$

The use of geometrical distance in (1) is sensitive to outliers. Therefore, the entropy provides additional information to capture the differences between uncertain objects with different distributions. The example in Fig. 4 illustrates the entropy of Granule 1 and Granule 2 if object (b_1, c_2) is being merged. For simplicity, a two-dimensional data is used in this example. The object (b_1, c_2) is potentially being merged with Granule 1, or Granule 2. Even though it has similarity with both Granules (elements b_1 and c_2), but it is more likely to merge with Granule 2, since the entropy is lower than in Granule 1.

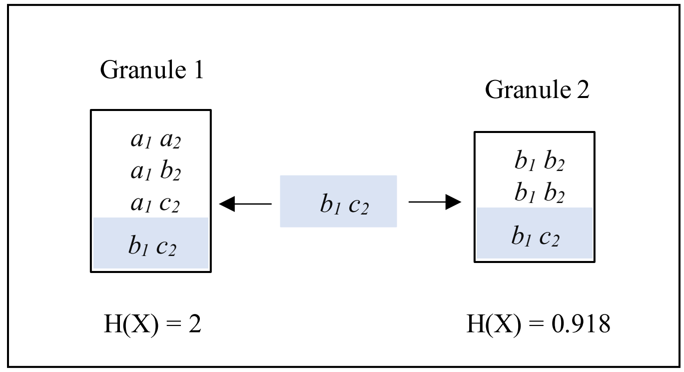


Figure 4. Entropy in granules.

Hence, the proposed compatibility measure includes the entropy H, and the weight w in the equation as follows:

$$C(A, B) = Distance_{MAX} - (wH + Distance_{A,B}.\exp(-\alpha R) \quad (12)$$

In the case study, the equation in (12) is used to compute the compatibility index. It can be expected that the compatibility in GrC with uncertainty is less than the conventional GrC. This indicates that the merging process in the case with an entropy measure is more competitive because it will find the granules that not only with shortest distance, but also with minimum uncertainty.

V. CASE STUDY AND SIMULATION RESULTS

Fig. 5 shows how the maximum compatibility for each iteration in GrC with entropy measure changes. With entropy measure, the compatibility decreases because the merging process now becomes more competitive. The iterative granulation process starts to be selective in choosing the best pair to be merged, not only distance wise, but also to produce granule with minimum uncertainty. The performance of the proposed algorithm is evaluated with Iris, wine and glass datasets and the results are shown in Table II.

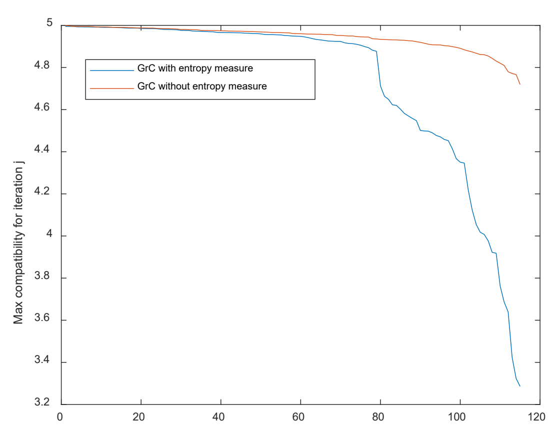


Figure 5. Comparison between compatibility index of GrC and GrC with entropy measure

TABLE II. AVERAGE PERFORMANCE OF 10 TRIALS OF GRC, GRC WITH ENTROPY, K-MEANS AND FUZZY C-MEANS

Dataset	Iris		Wine		Glass	
	RMSE	**Accuracy (%)**	**RMSE**	**Accuracy (%)**	**RMSE**	**Accuracy (%)**
GrC	0.1354	91	0.1233	91.33	0.1968	69.77
GrC with entropy measure	0.0957	96.33	0.0964	95.33	0.2065	72.09
K-means	0.1032	94.33	0.1055	92.67	0.2293	67.44
Fuzzy c-means	0.1257	96	0.088	95	0.2173	65.17

The datasets are divided into training and testing data with the ratio set to 80:20. All features are normalized to the interval [0,1] to bring all variables to the same range. Ten experiments are conducted for each algorithm, and their performance is evaluated by measuring the accuracy of the classification and the root mean square error (RMSE). The accuracy is calculated based on the percentage of correct classifications out of the total number of predictions, while RMSE represents the model error.

For simplicity purpose, probability mass functions are used to represent the probability of a random variable in the case of GrC with entropy measure. The weight w used for Iris, wine and glass are 0.2, 0.1 and 0.3, respectively. Ideally, the value of w is used in the interval [0,1]. For the glass dataset, the data needs to be pre-processed by using a bootstrapping method due to the imbalance among the six classes.

From Table II, it can be viewed that in general, GrC with entropy measure outperforms the conventional GrC, k-means and fuzzy c-means. It achieves highest accuracy for all datasets and lowest RMSE for Iris. This is due to the penalizing the uncertainty during the iterative granulation process, and hence, producing low uncertainty granules. The accuracy of prediction using GrC with entropy measure for Iris, wine and glass are 96.33%, 95.33% and 72.09%, respectively, which are acceptable as benchmarked with other works in literature using the same datasets, such as [14].

CONCLUSION

In this paper, a data capture and modelling framework based on Granular Computing and information theory is proposed. The fundamental concept of information theory is implemented to model the uncertainty during the data granulation. In the presence of outliers, entropy appears as a significant tool to identify uncertainties. The entropy represents the hesitation of two granules to be merged and potentially guide the granulation process into merging the granules with minimum uncertainty in order to produce high quality information granules.

The main idea of this framework is to avoid having granules with high disorder in the data distribution. This prevents the granules from having outliers that may affect the distinguishability of the granules and hence, enhance the interpretability of FL rule-bases. The proposed framework is successfully tested with three datasets that deal with classification problem – Iris, wine and glass, and significant improvement can be observed in terms of the accuracy of the prediction and reducing the error.

ACKNOWLEDGMENT

This research is sponsored by Universiti Teknologi MARA, Ministry of Education, Malaysia and The University of Sheffield, UK.

REFERENCES

[1] S.S.S. Ahmad, “Fuzzy modelling through granular computing”, Ph.D. thesis, Dept. of Electrical and Computer Eng., Univ. of Alberta, Alberta, Canada, 2012.

[2] J. Kerr-Wilson, “Fuzzy rule-based systems: design, analysis, and applications”, Ph.D. thesis, Dept. of Electrical and Computer Eng., Univ. of Alberta, Alberta, Canada, 2019.

[3] W. Pedrycz and A. Bargiela “Granular clustering: a granular signature of data," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 32, no. 2, pp. 212-224, Apr. 2002. [Online]. doi: 10.1109/3477.990878.

[4] G. Panoutsos and M. Mahfouf, “A neural-fuzzy modelling framework based on granular computing: Concepts and applications," Fuzzy Sets and Systems, vol. 161, no. 21, pp. 2808-2830, 2010. [Online]. doi: 10.1016/j.fss.2010.06.004.

[5] G. J. Klir and B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications. New Jersey, USA: Prentice-Hall, 1995.

[6] B. Jiang, J. Pei, Y. Tao and X. Lin, "Clustering uncertain data based on probability distribution similarity," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 4, pp. 751-765, Apr. 2013. [Online]. doi: 10.1109/TKDE.2011.221.

[7] A. Baraka, G. Panoutsos, M. Mahfouf and S. Cater, "A Shannon entropy-based conflict measure for enhancing granular computing-based information processing," in 2014 IEEE International Conference on Granular Computing (GrC), Noboribetsu, Japan, pp.13-18. [Online]. doi: 10.1109/GRC.2014.6982799.

[8] L. A. Zadeh, “The role of fuzzy logic in the management of uncertainty in expert systems,” Fuzzy Sets and Systems, vol. 11, no. 1-3, pp. 199–227, 1983. [Online]. doi: 10.1016/S0165-0114(83)80081-5.

[9] A. Bargiela and W. Pedrycz, Granular computing : an introduction. New York, USA: Springer US, 2003.

[10] A. R. Solis and G. Panoutsos, "Granular computing neural-fuzzy modelling: A neutrosophic approach," Applied Soft Computing Journal, vol. 13, no. 9, pp. 4010–4021, Sept. 2013. [Online]. doi: 10.1016/j.asoc.2012.09.002.

[11] C. E. Shannon, "A mathematical theory of communication," in The Bell System Technical Journal, vol. 27, no. 3, pp. 379-423, July 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.

[12] H. Sun, R. Chen, Y.Qin and S. Wang, "Holo-entropy based categorical data hierarchical clustering," Informatica, vol. 28, no. 2, pp. 303–328, Oct. 2017. [Online]. doi: 10.15388/Informatica.2017.131.

[13] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York, USA: John Wiley & Sons, 1991.

[14] A. R. Solis, P. Melin, M. Hernandez and G. Panoutsos, " General type-2 radial basis function neural network: a data-driven fuzzy model," IEEE Transactions on Fuzzy Systems, vol. 27, no. 2, pp. 333–347, Feb. 2019. [Online]. doi: https://doi.org/10.1109/TFUZZ.2018.2858740.

AUTHORS' BACKGROUND

Your Name	Title*	Research Field	Personal website
Muhammad Zaiyad Muda	PhD candidate	Interpretable machine learning, computational intelligence, Granular Computing	https://www.researchgate.net/profile/Muhammad_Muda3
George Panoutsos	Full Professor (Professor of Computational Intelligence)	Interpretable machine learning, computational intelligence, Granular Computing	https://www.sheffield.ac.uk/acse/department/people/academic/george-panoutsos