

A Design of Incremental Granular Network for Software Data Modeling

Keun-Chang Kwak

Department of Control, Instrumentation and Robot Engineering, Chosun University, Gwangju, Korea (South)
Email: kwak@chosun.ac.kr

Received September 6th, 2010; revised September 27th, 2010; accepted October 5th, 2010.

ABSTRACT

In this paper, we propose an incremental method of Granular Networks (GN) to construct conceptual and computational platform of Granular Computing (GrC). The essence of this network is to describe the associations between information granules including fuzzy sets formed both in the input and output spaces. The context within which such relationships are being formed is established by the system developer. Here information granules are built using Context-driven Fuzzy Clustering (CFC). This clustering develops clusters by preserving the homogeneity of the clustered patterns associated with the input and output space. The experimental results on well-known software module of Medical Imaging System (MIS) revealed that the incremental granular network showed a good performance in comparison to other previous literature.

Keywords: Incremental Granular Network, Granular Computing, Information Granules, Context-Based Fuzzy Clustering

1. Introduction

Granular Computing (GrC) is a general computation theory for effectively using information granules such as classes, clusters, subsets, groups, and intervals to build an efficient computational model for complex applications with huge amounts of data, information and knowledge [1]. Furthermore, granular computing forms a unified conceptual and computing platform. Yet, it directly benefits from the already existing and well-established concepts of information granules formed in the setting of sets and interval theory, fuzzy sets, rough sets, and shadowed sets [2].

In order to form the conceptual and computing platform of granular computing, we introduce granular network with two types that directly uses the fundamental idea of fuzzy clustering. Based on this network, we also develop and design an incremental granular network that combines linear regression and local granular network [3]. First, we build a standard regression model which could be treated as a preliminary construct capturing the linear part of the data and in this way forming a backbone of the entire construct. Next, all modeling discrepancies are compensated by a collection of rules that become attached to the regions of the input space in which the error becomes localized. Here the network is de-

signed by the use of fuzzy granulation realized via context-based fuzzy clustering [4]. This clustering technique builds information granules in the form of fuzzy sets and develops clusters by preserving the homogeneity of the clustered patterns associated with the input and output space. The effectiveness of this clustering has been demonstrated on Linguistic Models (LM) [5,6], Radial Basis Function Neural Networks (RBFNNs) [7], and incremental models [8]. These models represented a nonlinear and complex characteristic more effectively than conventional models based on context-free clustering. This paper is organized as follows. Section 2 describes the architecture of granular network with two types and mechanism of context-based fuzzy clustering. In Section 3, we present the design of incremental granular network. This network is applied to software module of well-known Medical Imaging System (MIS) [9] in Section 4. Finally, conclusions are given in Section 5.

2. Granular Network

Let us firstly recall the mechanism of context-based fuzzy clustering. This clustering as an interesting variant of the fuzzy c-means is realized via individual contexts. Each context has clearly defined semantics that can be interpreted as a *large* negative error, *medium* negative error, etc. Consider a certain fixed context W_j described

by some membership function. The data point in the output space is then associated with the corresponding membership value. Let us introduce a family of the partition matrices induced by the l -th context and denote it by $U(W_l)$

$$U(W_l) = \left\{ u_{ik} \in [0,1] \mid \sum_{i=1}^c u_{ik} = w_{lk} \forall k \text{ and } 0 < \sum_{k=1}^N u_{ik} < N \right\} \quad (1)$$

where w_{lk} denotes a membership value of the k -th datum implied by the l -th context. The underlying objective function is as follows

$$Q = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^m \| \mathbf{x}_k - \mathbf{v}_i \|^2 \quad (2)$$

where \mathbf{v}_i denotes the i -th prototype. The Q is minimized under the constraints imposed by (1) as follows

$$\text{Min } Q \text{ subject to } U(W_l), l = 1, 2, \dots, p \quad (3)$$

The minimization of Q is realized by iteratively updating the values of the partition matrix and the cluster centers. The successive updates of the partition matrix are completed as follows

$$u_{ik} = \frac{w_{lk}}{\sum_{j=1}^c \left(\frac{\| \mathbf{x}_k - \mathbf{v}_i \|^2}{\| \mathbf{x}_k - \mathbf{v}_j \|^2} \right)^{\frac{2}{m-1}}} \quad (4)$$

where $i = 1, 2, \dots, c, k = 1, 2, \dots, N$

Note that u_{ik} means the partition matrix induced by the l -th context. The prototypes are determined as

$$\mathbf{v}_i = \frac{\sum_{k=1}^N u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^N u_{ik}^m} \quad (5)$$

We assume that the fuzzification factor m is 2.0. In the design of the granular network, we consider the contexts to be described by triangular membership functions being equally distributed in the error space \mathbf{E} with the 1/2 overlap occurring between two successive fuzzy sets. **Figure 1** visualizes the example of a blueprint of the incremental granular network for $p = 3$ and $c = 2$.

Each context generates a number of induced clusters whose activation levels are afterwards summed up as shown in **Figure 2**.

Denoting those by $\xi_1, \xi_2, \dots, \xi_n$ the output of the network is granular. Assuming the triangular form of the contexts, the result is a triangular fuzzy number E as follows

$$E = W_1 \otimes \xi_1 \oplus W_2 \otimes \xi_2 \oplus \dots \oplus W_n \otimes \xi_n \quad (6)$$

We denote the algebraic operations by \otimes, \oplus to emphasize

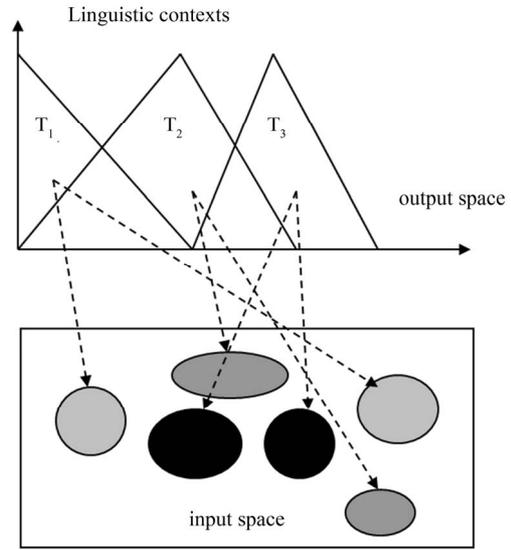


Figure 1. Concept of context-based fuzzy clustering.

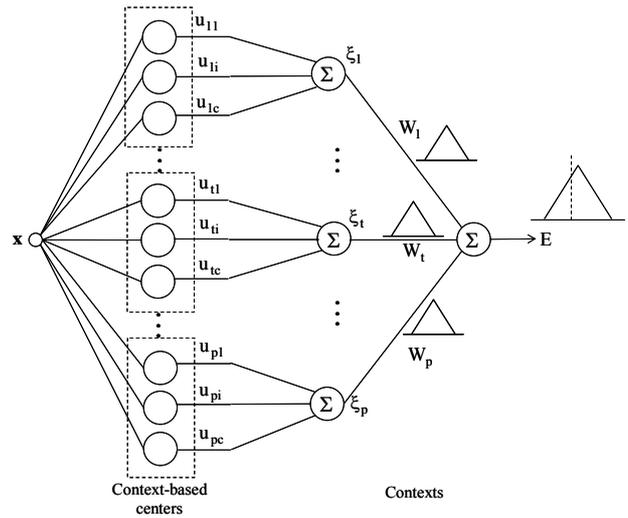


Figure 2. Architecture of the granular network (case 1).

that the underlying computing operates on a collection of fuzzy numbers. As such, E is characterized by its three parameters that are a modal value, the lower bound, and upper bound.

On the other hand, we develop the advanced granular network with detailed linguistic context as shown in **Figure 3**. The consequent part is obtained by Constrained Least Square Estimate (CLSE) method as follows

$$\min_{\theta} \| U \theta - Y \|, \text{ subject to } \min(Y) \leq \theta \leq \max(Y) \quad (7)$$

where U and Y denote the activation levels in layer 2 and the actual output, respectively. The parameter θ to be estimated is the modal values of the detailed linguistic contexts. For further details on the CLSE method, see [10].

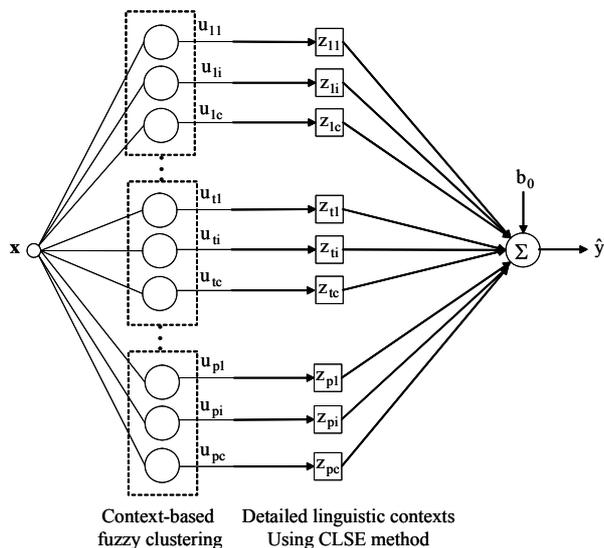


Figure 3. Architecture of the granular network (case 2)

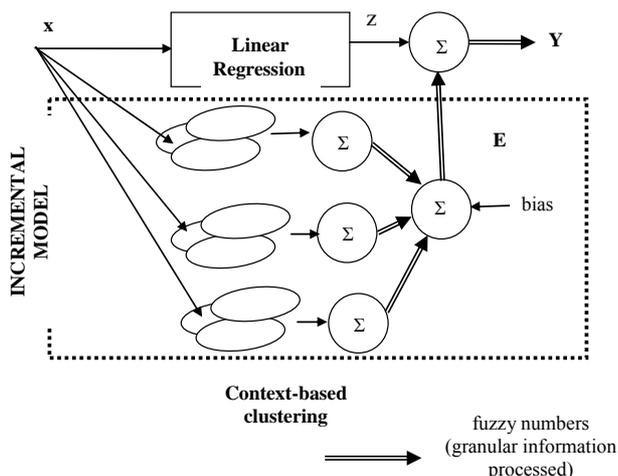


Figure 4. Overall flow of incremental granular network

3. Design of Incremental Granular Network

The main design process of the incremental granular network is shown in **Figure 4** showing how the two functional modules operate. Firstly, we decide upon the granularity of information to be used in the development of the model such as the number of contexts and the number of clusters formed for each context. The design procedure of incremental granular network is as follows [8].

[Step 1] Design of a linear regression in the input and output space, $z = L(\mathbf{x}; \mathbf{b})$ with \mathbf{b} denoting a vector of the regression hyperplane, $\mathbf{b} = [\mathbf{a} \ a_0]^T$. On the basis of the original data set formed is a collection of input-error pairs, (\mathbf{x}_k, e_k) where $e_k =$

target- $L(\mathbf{x}_k, \mathbf{a})$.

[Step 2] Construction of the collection of contexts in the space of error of the regression model E_1, E_2, \dots, E_p . The distribution of these fuzzy sets is optimized through the use of fuzzy equalization while the fuzzy sets are characterized by triangular membership functions with a 0.5 overlap between neighboring fuzzy sets.

[Step 3] Context-based fuzzy clustering completed in the input space and induced by the individual fuzzy sets of context. For “p” contexts and “c” clusters per context, obtained are $c \cdot p$ clusters.

[Step 4] Summation of the activation levels of the clusters induced by the corresponding contexts and their overall aggregation through weighting by fuzzy sets of the context leading to the triangular fuzzy number of output, $E = F(\mathbf{x}; E_1, E_2, \dots, E_p)$ where F denotes the overall transformation realized by the incremental granular network. Furthermore note that we eliminated eventual systematic shift of the results by adding a numeric bias term.

[Step 5] The result of the incremental granular network is then combined with the output of the linear part. The result is a shifted triangular number $Y, Y = z \oplus E$.

4. Experimental Results

In order to evaluate the performance of the incremental granular network for data modeling in software engineering, we applied to well-known Medical Imaging System (MIS) subset of 390 software modules written in Pascal and FORTRAN [9]. These modules consist of approximately 40,000 lines of code. We use 11 system input variables such as, LOC, CL, TChar, TComm, MChar, DChar, N, N_h , N_F , $V(G)$, and BW, The output variable to be predicted is “Changes”. The training and testing data set are randomly selected by 60%-40%.

The experiments are performed by 10 runs. The training data set is used for model construction, while the test set is used for model validation. Thus, the resultant model is not biased toward the training data set and it is likely to have a better generalization capacity to new data. We obtained the best case ($m = 3.0, p = c = 6$), while varying the number of cluster ($2 \leq p \leq 6$) and fuzzification factor ($m = 1.5, 2.0, 2.5, 3.0$).

Figure 5 and **Figure 6** show the contexts (Case 1) and consequent parameters (Case 2) obtained from linear regression error, respectively. **Figure 7** shows the prediction performance of incremental granular networks. **Figure 8** visualizes the distribution of clusters and some input data. **Table 1** lists the experimental comparison on RMSE (root mean square error). In the design of LM, we

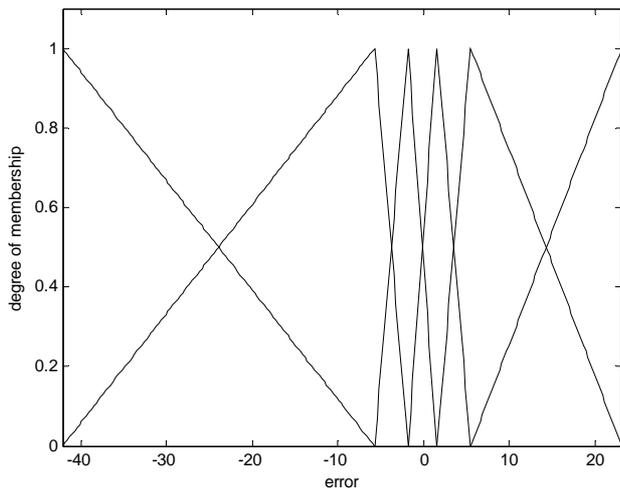


Figure 5. Contexts obtained from linear regression error (Case 1).

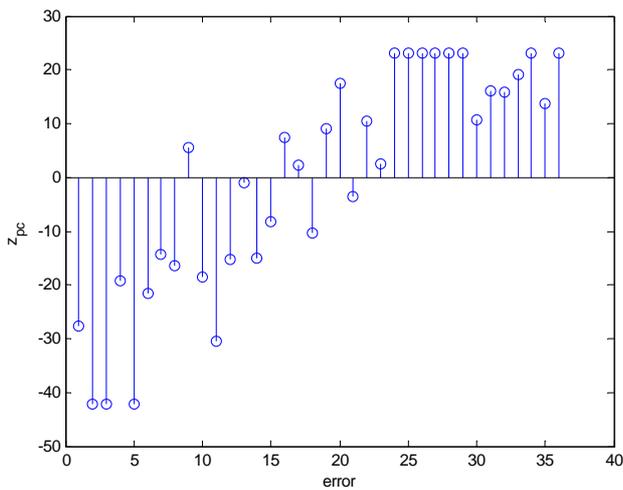


Figure 6 Consequent parameters (Case 2).

used six contexts and six clusters in each context for context-based fuzzy clustering. Although the LM has a structured knowledge representation in the form of fuzzy if-then rules, it lacked the adaptability to deal with non-linear model. Moreover, we constructed the RBFN based on six contexts and six clusters in the same manner. Here learning rate is 0.0001 and the number of epoch is 1000. As listed in **Table 1**, we can recognize that the proposed method (IGN with two cases) showed a good performance in comparison to linguistic model and RBFNNs based on context-based fuzzy clustering.

5. Conclusions

We presented the design of the incremental granular network for software data of medical imaging system. This network is adopted a construct of a linear regression as a first-principle global model, refine it through a series

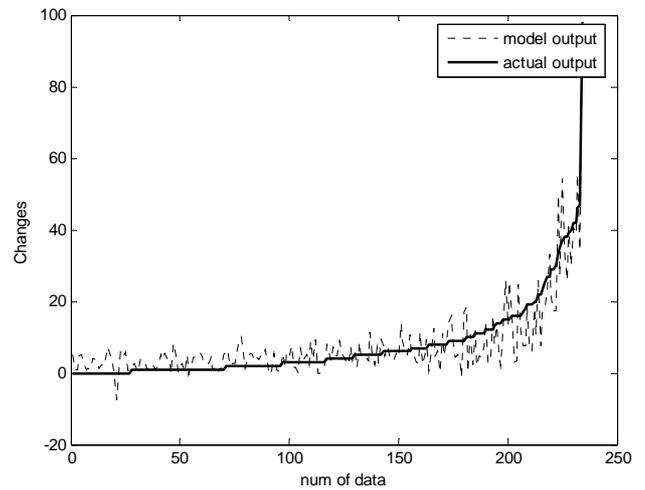


Figure 7. Predication performance for MIS data.

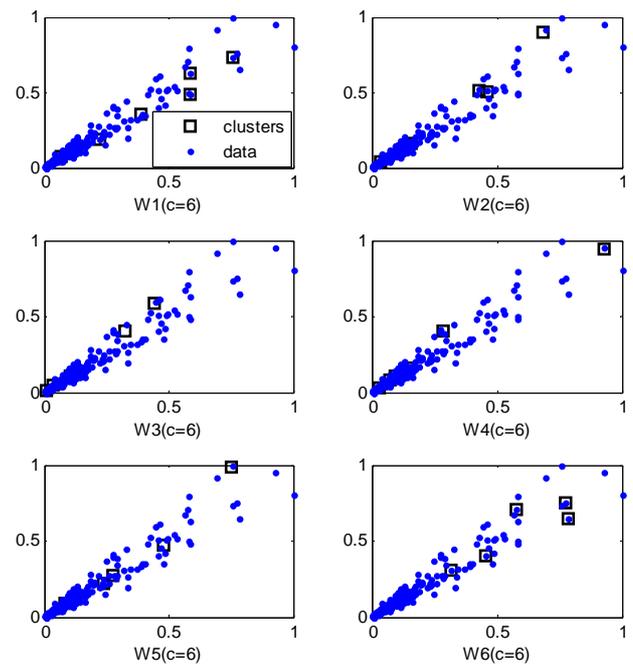


Figure 8. Distribution of clusters and input data (DChar, N).

Table 1. Performance comparison.

Methods	Prediction Performance	
	Train_RMSE	Check_RMSE
LM [4]	6.266	7.981
RBFN [6]	6.631	7.772
IGN(Case1)	4.626	6.624
IGN(Case2)	3.770	6.532

of local fuzzy rules that capture remaining and more localized nonlinearities of the system. More schematically, we could articulate the essence of the resulting incremental granular network by stressing the existence of the

two essential modeling structures that are combined linear regression and local granular network. The experimental results revealed that the incremental granular network outperformed the previous works. The granular networks used in this paper can be applied to intelligent data analysis, nonlinear system modeling, adaptive hypermedia, e-commerce, and intelligent interfaces.

REFERENCES

- [1] W. Pedrycz, A. Skowron and V. Kreinovich, "Handbook of Granular Computing," John Wiley & Sons, Hoboken, 2008.
- [2] W. Pedrycz and F. Gomide, "Fuzzy Systems Engineering: Toward Human-Centric Computing," Wiley-Interscience, Hoboken, 2007.
- [3] M. Y. Lee and K. C. Kwak, "An Incremental Granular Network for Data Modeling in Software Engineering," 2010 4th International Conference on New Trends in Information Science and Service Science (NISS), Gyeongju, Korea, May 2010, pp. 495-498.
- [4] W. Pedrycz, "Conditional Fuzzy C-Means," *Pattern Recognition Letters*, Vol. 17, No. 6, May 1996, pp. 625-632.
- [5] W. Pedrycz and A. V. Vasilakos, "Linguistic Models and Linguistic Modeling," *IEEE Transactions on Systems, Man and Cybernetics-Part C*, Vol. 29, No. 6, 1999, pp. 745-757.
- [6] W. Pedrycz and K. C. Kwak, "Linguistic Models as Framework of User-Centric System Modeling," *IEEE Transactions on Systems, Man and Cybernetics-Part A*, Vol. 36, No. 4, 2006, pp. 727-745.
- [7] W. Pedrycz, "Conditional Fuzzy Clustering in the Design of Radial Basis Function Neural Networks," *IEEE Transactions on Neural Networks*, Vol. 9, No. 4, 1999, pp. 745-757.
- [8] W. Pedrycz and K. C. Kwak, "The Development of Incremental Models," *IEEE Transactions on Fuzzy Systems*, Vol. 15, No. 3, 2007, pp. 507-518.
- [9] S. K. Oh, W. Pedrycz and B. J. Park, "Self-Organizing Neurofuzzy Networks in Modeling Software Data," *Fuzzy Sets and Systems*, Vol. 145, No. 1, July 2004, pp. 165-181.
- [10] J. Abonyi, R. Babuska and F. Szeifert, "Fuzzy Modeling with Multivariate Membership Functions: Gray-Box Identification and Control Design," *IEEE Transactions on Systems, Man and Cybernetics-Part B*, Vol. 31, No. 5, 2001, pp. 755-767.