INTUITIONISTIC FUZZY C-LEAST SQUARES SUPPORT VECTOR REGRESSION WITH SAMMON MAPPING CLUSTERING ALGORITHM

Kuo-PingLin
Department of Information Management, Lunghwa University of Science and Technology, Taiwan
kplin@mail.lhu.edu.tw

Teresa L. Ju
Department of Information Management, Lunghwa University of Science and Technology, Taiwan
tju@mail.lhu.edu.tw

Ping-Feng Pai
Department of Information Management, National Chi Nan University, Taiwan
paipf@ncnu.edu.tw

Chih-Hung Kuo
Department of Information Management, Lunghwa University of Science and Technology, Taiwan
Kuo771003@gmail.com

Abstract

This study proposes a novel Intuitionistic fuzzy c-least squares support vector regression (IFC-LSSVR) with sammon mapping clustering algorithm. The proposed clustering algorithm can obtain the advantages of intuitionistic fuzzy sets, LSSVR, and sammon mapping in actual clustering problems. Moreover, IFC-LSSVR with sammon mapping adopts particle swarm optimization (PSO) to search optimal parameters. Experiments on web-based adaptive learning environments data set, which is to provide enough or suitable knowledge for students/users, show that the proposed IFC-LSSVR with sammon mapping is more efficient than conventional algorithms such as the k-means (KM) and fuzzy c-means (FCM) clustering algorithm, in standard measurement indexes.

Keywords: Intuitionistic fuzzy sets, least squares support vector regression, sammon mapping, web-based adaptive learning environments

1. Introduction

Fuzzy c-regression model (FCRM) has been reported (Hathaway & Bezdek, 1993), and is an alternative in clustering algorithm. The FCRM employs the conventional regression and fuzzy clustering, where data may be assumed to be generated from the regression models and the data may fit several or all models to varying degrees (i.e., the membership grades). Sun (2006) proposed fuzzy probability c-regression based on least squares support vector machine. This study found least squares support vector machine/ least square support vector regression (LS-SVR) can discriminate the multiple regression models with a fuzzy partition of data set while fitting perfectly these models and can overcome the problem of initialization that often makes termination occurring at local minima. The LS-SVR (Van Gestel et al., 2001) was proposed; it improves the traditional support vector machine which attempts to minimize sum square errors (SSEs) of training data sets while simultaneously minimizing the margin error. Many researches had extended and applied the LS-SVR to various field (Pai et al., 2014; Hung & Lin, 2013; Lin, 2013; Lin & Pai, 2013; Peng & Wang, 2009; Suykens et al., 2002).

The purpose of this study is improved FCRM clustering algorithm. Sun (2006) used least squares support vector machine in FCRM. Lin (2015) proposed evolutionary kernel Intuitionistic fuzzy c-means clustering algorithm which used Intuitionistic fuzzy technique to improve kernel fuzzy c-means clustering algorithm. Therefore, this study proposes IFC-LSSVR with sammon mapping. The sammon mapping can reduce dimensional space. The proposed clustering algorithm can obtain the advantages of intuitionistic fuzzy sets, LSSVR, and sammon mapping in actual clustering problems. The rest of this paper is organized as follows. Section 2 introduces intuitionistic fuzzy c-least-squares support vector regression with Sammon mapping. Section 3 shows the experimental results of IFC-LSSVR with sammon mapping in Web-based adaptive learning environments data set. Finally, the research draws conclusions and makes suggestions for further research in section 4.

2. Intuitionistic fuzzy *c*-least-squares support vector regression with Sammon mapping

The FCRM were introduced by Hathaway & Bezdek (1993) and perform conventional regression analyses with the fuzzy c-mean's clustering. This study extends the fuzzy c-regression models and develops the IFC-LSSVR with sammon mapping. Picture 1 shows the flowchart of the proposed IFC-LSSVR with sammon mapping. Firstly, the study adopts sammon mapping to reduce dimensional space which can simplify the variety in dataset. Secondly, Intuitionistic fuzzy c-least-squares support vector regression is employed to cluster dataset which has been reduce dimension of raw data. In the meanwhile IFC-LSSVR with sammon mapping is employed particle swarm optimization (PSO) to search optimal parameter of IFC-LSSVR with sammon mapping. Finally, the results of clustering can be obtained to decision-makers.

2.1 Intuitionistic fuzzy c-least-squares support vector regression

The parameters of IFC-LSSVR, including the number of clusters to classify (c), parameter in updating the cluster memberships (ω), number of epochs to carry out (L), and the tolerance in stopping criterion (δ), are preset. Then, in IFC-LSSVR, similar to the fuzzy c-regression models, initial memberships of data Y of crisp input and output (i = 1, …, N) to the clusters j (j = 1, …, c) may be randomly generated and denoted as $U(l) = [\mu_{ij}^{(l)}]_{N\times c}$ (with l = 0), under the condition that $\sum_{\forall j} \mu_{ij}^{(l)} = 1$ $\forall i$. Or alternatively a conventional clustering method may be used so that an initial partition of the data XS_i with a crisp partition into c clusters may be obtained. The remainder procedure of the IFC-LSSVR models can be performed as follows Table 1 with the initial value of l setting equal to 0.

Table 1: IFC-LSSVR with sammon mapping

IFC-LSSVR **algorithm**. Input: Sammon mapping data XS_i, c, ω, parameter of intuitionistic fuzzy α, parameters of LS-SVR (C, σ)
Output: fuzzy partition U, prototypes $\boldsymbol{\theta}$
1: **procedure** IFC-LSSVR (data XS, c, ω, α, C, σ)
2: Initialize $U^{(l=0)}$ to random fuzzy partition
3: **repeat**
4: update $U^{(l+1)}$ by calculating the cluster centers using Eq.(9)

6: **until** $\left\| U^{(l+1)} - U^{(l)} \right\| < \delta$ is satisfied.

7: **return** $U^{(l+1)}$

8: **end procedure**

Particle Swarm Optimization is employed to search parameters of IFC-LSSVR which includes *c*, ω, parameter of intuitionistic fuzzy α, parameters of LS-SVR (C, σ), respectively.

3. Web-based adaptive learning environments data set

A web-based adaptive learning environments were executed with various machine learning data sets using standard KM, FCM, and the proposed IFC-LSSVR with sammon mapping. The web-based adaptive applications can adopt user modelling technique. Researchers use the model to customize subjects according to their knowledge. It may provide enough or suitable knowledge for students. Therefore, the adaptation of web content is the basic requirement of web-based adaptive applications. In the study, web-based adaptive learning environments data set can refer to Kahraman et al. (2013) which proposed an intuitive knowledge classifier for web-based adaptive learning environments dataset.

Table 2 shows the results of the web-based adaptive learning environments data sets by various clustering methods. IFC-LSSVR with sammon mapping did obtain a higher classification rate than other algorithms. This means the proposed method has outperform other techniques in web-based adaptive learning environments data set. The IFC-LSSVR with sammon mapping can ecffectively discover the structure of web-based adaptive learning environments.

Table 2: Clustering results for web-based adaptive learning environments data sets.

Title	Method	Parameters	Classification rate	The best classification rate
Web-based adaptive learning environments	KM	c=4	0.3059±0.076	0.4806
	FCM	c=4, ω=2	0.4109±0	0.4109
	IFC-LSSVR with sammon mapping	D=—, c=4, ω= 45.8, α= 42.8, C= 36.5, σ= 86.02	0.3682 ±0	**0.5233**
		D=4, c=4, ω= 67.83, α= 18.67, C= 81.20, σ= 63.75	0.4845±0	
		D=3, c=4, ω=4.69, σ=3.92 ,α= 2.89	0.5233±0	
		D=2, c=4, ω=4.69, σ=3.92 ,α= 2.89	0.5233±0	

—: Non

4. Conclusion

This study firstly developed a IFC-LSSVR with sammon mapping and examined it using web-based adaptive learning environments data set. The results indicate that IFC-LSSVR with sammon mapping offers a promising alternative for clustering. Overall, the EKIFCM can provide more stable and better performance in classification rate. The superior performance of IFC-LSSVR with sammon mapping can be attributed to several factors. First, the sammon mapping can reduce the dimension of raw dat. Second, the LS-SVR can improve tradtional regression line in dataset. Third, this IFC-LSSVR incorporates another uncertainty, the hesitation degree that arises while defining the membership function. Hesitation degrees are used for determining precisely the membership function because the Yager class of IFSs (hesitation degrees) can help to amplify differentiation of clusters. This is done to maximize the good points in the class. In the experiments, using an Yager class of IFSs which increases the membership can yield better classification rates. Fourth, PSO mechanisms can effectively improve IFC-LSSVR performance.

For future work, other types of machine learning data set with the IFC-LSSVR with sammon mapping would be a challenging issue for study. Future studies could also consider using others data preprocessing techniques to improve the IFC-LSSVR model.

Acknowledgments

The authors would like to thank the Ministry of Science and Technology of the Republic of China, Taiwan for financially supporting this research under Contract Numbers MOST 103-2410-H-262 -010.

References

1. Burillo, P., & Bustince, H. (1996). Vague sets are intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 79, 403–405.
2. Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annual Eugenics*, 7, 179-188.
3. Hathaway, R.J., & Bezdek, J.C. (1993). Switching regression models and fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, 1, 195–204.
4. Hung, C.-C., Kulkarni, S., & Kuo, B.-C. (2011), A new weighted fuzzy c-means clustering algorithm for remotely sensed image classification. *IEEE Journal of selected topics in signal processing*, 5, 543–553.
5. Hung, K.-C., & Lin, K.-P. (2013). Long-term business cycle forecasting through a potential intuitionistic fuzzy least-squares support vector regression approach. *Information Sciences*, 224, 37–48.
6. Kahraman, H.T., Sagiroglu, S., & Colak, I. (2013). The development of intuitive knowledge classifier and the modeling of domain dependent data. *Knowledge-Based Systems*, 37, 283–295.
7. Kannan, S.R., Ramathilagam, S., & Chung, P.C. (2012). Effective fuzzy c-means clustering algorithms for data clustering problems. *Expert Systems with Applications*, 39, 6292–6300.
8. Kennedy, J. (1997). The particle swarm: Social adaptation of knowledge. In Proc. IEEE Int. Conf. Evol. Comput. (pp. 303-308), Indianapolis.
9. Li, X., Ming, S., & Lixing, D. (2010). Particle swarm optimization-based LS-SVM for building cooling load prediction. Journal of Computers, 5, 614-621.
10. Lin, K.-P. (2013). Application of least-squares support vector regression with PSO for CPU Performance Forecasting. *Advanced Materials Research*, 630, 366–371.
11. Lin, K.-P., Pai, P.-F., Lu, Y.-M., & Chang, P.-T. (2013). Revenue forecasting using a least-squares support vector regression model in a fuzzy environment. *Information Sciences*, 220, 196–209.
12. Peng, X., & Wang, Y. (2009). A normal least squares support vector machine (NLS-SVM) and its learning algorithm. Neurocomputing, 72, 3734–3741.
13. Sammon, J.W., Jr (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, C-18(5), 401–409.
14. Sun, Z. (2006). Fuzzy probability c-regression estimation based on least squares support vector machine. *Lecture Notes in Computer Science*, 4232, 874-881.
15. Van Gestel, T., Suykens, J.A.K., Baestaens, D.-E., Lambrechts, A., Lanckriet, G., Vandaele, B., De Moor, B., & Vandewalle, J. (2001). Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Transactions on Neural Networks*, 12, 809–821.
16. Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., & Vandewalle, J. (2002). *Least squares support vector machines*. Singapore: World Scientific.
17. Wang, H., & Fei, B. (2009). A modified fuzzy C-means classification method using a multiscale diffusion filtering scheme. *Medical Image Analysis*, 13, 193–202.