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

Kuo-Ping Lin
Department of Information Management, Lungwha University of Science and Technology, Taiwan
kplin@mail.lhu.edu.tw

Teresa L. Ju
Department of Information Management, Lungwha 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, Lungwha 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. Second, 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 \(\omega\), number of epochs to carry out \(L\), and the tolerance in stopping criterion \(\delta\), are preset. Then, in IFC-LSSVR, similar to the fuzzy c-regression models, initial memberships of data \(Y\) of crisp input and output \((i = 1, \ldots, N)\) to the clusters \(j (j = 1, \ldots, c)\) may be randomly generated and denoted as \(U(l) = [\mu_{ij}^{(l)}]_{N \times c}\) (with \(l = 0\), under the condition that \(\sum_{j} \mu_{ij}^{(l)} = 1\) \(\forall i\). Or alternatively a conventional clustering method may be used so that an initial partition of the data \(X_S\) 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>
<thead>
<tr>
<th>IFC-LSSVR algorithm</th>
<th>Input: Sammon mapping data (X_S), (c), (\omega), parameter of intuitionistic fuzzy (\alpha), parameters of LS-SVR ((C, \sigma))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output: fuzzy partition (U), prototypes (\theta)</td>
<td></td>
</tr>
<tr>
<td>1: procedure IFC-LSSVR (data (X_S), (c), (\omega), (\alpha), (C), (\sigma))</td>
<td></td>
</tr>
<tr>
<td>2: Initialize (U^{(0)}) to random fuzzy partition</td>
<td></td>
</tr>
<tr>
<td>3: repeat</td>
<td></td>
</tr>
<tr>
<td>4: update (U^{(l+1)}) by calculating the cluster centers using Eq.(9)</td>
<td></td>
</tr>
</tbody>
</table>
6: until $|U^{(n)} - U^{(n-1)}| < \delta$ is satisfied.
7: return $U^{(n)}$
8: end procedure

Particle Swarm Optimization is employed to search parameters of IFC-LSSVR which includes $c$, $\omega$, parameter of intuitionistic fuzzy $\alpha$, parameters of LS-SVR $(C, \sigma)$, 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 dataset. The IFC-LSSVR with sammon mapping can effectively discover the structure of web-based adaptive learning environments.

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

<table>
<thead>
<tr>
<th>Title</th>
<th>Method</th>
<th>Parameters</th>
<th>Classification rate</th>
<th>The best classification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web-based adaptive learning environments</td>
<td>KM</td>
<td>$c=4$</td>
<td>0.3059±0.076</td>
<td>0.4806</td>
</tr>
<tr>
<td></td>
<td>FCM</td>
<td>$c=4$, $\omega=2$</td>
<td>0.4109±0</td>
<td>0.4109</td>
</tr>
<tr>
<td></td>
<td>IFC-LSSVR with sammon mapping</td>
<td>$D=\ldots$, $c=4$, $\omega=45.8$, $C= 22.8$, $\sigma= 86.02$, $D=4$, $c=4$, $\omega= 67.83$, $\alpha= 42.8$, $D=3$, $c=4$, $\omega=4.69$, $\alpha= 2.89$, $D=2$, $c=4$, $\omega=4.69$, $\alpha= 2.89$</td>
<td>0.3682 ±0</td>
<td>0.5233</td>
</tr>
<tr>
<td></td>
<td>IFC-LSSVR with sammon mapping</td>
<td>$D=\ldots$, $c=4$, $\omega=4.69$, $\alpha= 2.89$</td>
<td>0.5233±0</td>
<td></td>
</tr>
</tbody>
</table>

—: 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 traditonal 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