Indian Journal of Science and Technology, Vol 9(45), DOI: 10.17485/ijst/2016/v9i45/106369, December 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Validation of Software Component Selection Algorithms based on Clustering

Jagdeep Kaur[1]* and Pradeep Tomar[2]

[1]CSE and IT Department.,TheNorthCap University, Gurgaon - 122017, Haryana, India; jagdeep_kaur82@rediffmail.com
[2]Department of Computer Science and Engineering, School of ICT, Gautam Buddha University, Greater Noida - 201038, Uttar Pardesh, India; parry.tomar@gmail.com

Abstract

Objectives: To search the components that provides desired functionality, from the finite set of component set by the use of software component selection process. The selection process helps in choosing the optimal set of components from the third party repository. **Methods/Statistical Analysis**:In order to select the optimal set of component having multiple attributes, clustering is found to most suitable technique as revealed in the literature. This paper presents the validation of clustering based algorithms used for software component selection. It mainly covers the fuzzy-c means clustering and subtractive clustering. It also includes the earlier software component selection techniques proposed by the authors, hybrid XOR based clustering technique and fuzzy relation based fuzzy clustering. **Findings:**The Fuzzy c-means technique requires the need of mentioning the number of clusters centers in advanceand the radii of the cluster in case of Subtractive clustering are required. The disadvantage of Hybrid XOR based clustering is its dependency on subjective judgment of the developer. The FREFCOSCO algorithm has eliminated the usage of similarity index. It is able to deal with multi-attributes component and can generate the optimal set of components. **Application/Improvements:**The algorithms are validated on a set of components taken from an online repository. The improvement in the FREFCOSCO algorithm can be done by using an appropriate validity mechanism.

Keywords: Fuzzy C Means, Fuzzy Clustering, Hybrid XOR and Fuzzy Relation Based Fuzzy Clustering, Software Component Selection Algorithm, Subtractive Clustering, Validation of Algorithm

1. Introduction

Software is a collection of computer programs. The program consists of a set of instructions for the hardware to execute. The programs are developed using programming languages. Developing the programs over the years using engineering principles has transformed programming into a full-fledged engineering discipline. Since then, programming is referred as software engineering. But the software crisis arises due to high cost of software development and delay in project delivery. So, the Information Technology industry professionals are concerned about these issues. They can be resolved by using the concept of software reusability. The Component Based Software Engineering (CBSE) is a processes that emphasizes the design and assembly of automated systems by reusable software components.The major part in CBSE is classification and retrieval of software components from the repositories. From the repositories the components are selected for further use in Component-based development. In components selection, a number of software components selected from a subset of components or from components repository in such a way that their composition satisfies a set of objectives.There are numerous ways for component selection using computational intelligent techniques primarily by using fuzzy logic. Clustering is another way to group the similar types of components together having common features.The paper is organized as follows: In section two a review of computational intelligence based algorithms in software component selection

*Author for correspondence

is presented. The section three discusses the four major algorithms: fuzzy c-means, subtractive clustering, Hybrid XOR based clustering and Fuzzy relation based fuzzy clustering.

Many methods/techniques/algorithms have been proposed for software component selection based on clustering. A software component selection using analytic network process based on quality criteria like effectiveness, efficiency, satisfaction, safety and usability is proposed[1]. Because of dependencies among the components this analytic network process based approach was used. A coarse grain classification[2] based on fuzzy subtractive clustering is proposed. Subsequently, it's correctness is checked by precision and recall method and if satisfactory results are found the fine grain classification is made. Another component selection process[3] based on some quality attributes of the components is proposed. Firstly, metrics are made to quantify the attributes of the components. By using these metrics, fuzzy clustering is applied to select the best candidate. Further, in a formal approach for component based assessment[4] is proposed. The main components of this model are: the assessment domain, objectives, formal definitions of metrics and measurement result analysis method. Here system entities are identified, and then properties of these entities are identified using formal specifications. Lastly a problem is identified while interpreting the assessment results so fuzzy clustering analysis is proposed to place a component in more than one cluster and hence reducing the rigidity of threshold values of metrics. While reviewing the fuzzy relation based cluster analysis we found many researchers have established this process. According to an advanced fuzzy cluster analysis algorithm[5] is proposed based on fuzzy equivalence relation. It was further demonstrated that the fuzzy equivalence matrix can generate the fuzzy clustering.Another attempt[6] is to use the fuzzy cluster analysis based on fuzzy relations using max-t similarity relations. Another methodology[7] is based on clustering algorithm using fuzzy relation to produce different partition trees with different levels according to different t norms. In a fuzzy clustering technique[8] for data mining of records based on their importance is presented. It is realized that most of the previous work done for software component selection based on fuzzy clustering is either using the default values of acceptance and rejection ratios of MATLAB built-in fuzzy clustering methods or the number of clusters generated are less as compared to actual number of clusters that can be generated. Recent survey[9] presented different algorithms based on model, hierarchy, partitioning, grid and density. Over the past few years attempts have been made to help the application developer for finding a suitable component from the repository for reuse.

2. A Validation of Fuzzy Clustering Based Software Component Selection Algorithms

This section covers the four main algorithms based on clustering, namely: Fuzzy c-means, Subtractive Clustering, Hybrid XOR based Clusteringand Fuzzy relation based clustering (FREFCOSCO). Out of all these, three are fuzzy clustering based and one is purely hard clustering based.

2.1 Fuzzy c-means Algorithm

When there are chances that the data point can belong to more than one cluster; fuzzy c-means is used. The main advantage of this algorithm is that it is capable of assigning membership to data points that belong to two or more clusters. It is used for image segmentation with spatial information[10]. The component dataset is collected from online repository. Figure 1 is plotted for twenty components where the download rating is plotted against the bestseller rating. Here, the number of clusters are specified to be five.The red points indicate the components and the black dots are the cluster centers. The main disadvantage of this approach is one need to pre specify the number of clusters.

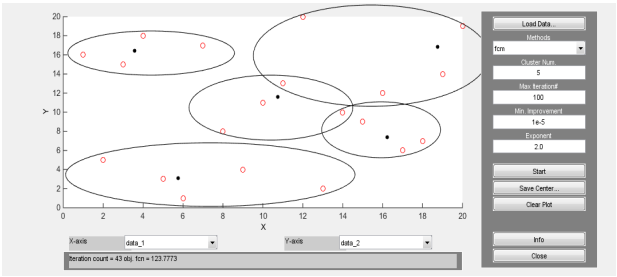

Figure 1. Kaur :The Fuzzy c-means Algorithm.

2.2 Subtractive Clustering

In this technique each data point is considered as a prospective cluster center. It measures the possibility that

each data point would term as the cluster center. This decision is established on the fact that concentration of surrounding data points is high. The basic steps of the algorithm are:

1. The first prospective cluster center is selected.
2. Based on the value of radii, all the data points surrounding the first cluster center are removed so that the next cluster centers are detected.
3. These steps are repeated until all the data points is within radii of a cluster center.

The previous dataset of twenty components are used for subtractive clustering. The clustering result is displayed in Figure 2.Here, three parameters play important role. The squash factors determine the neighborhood of a cluster center, so as to reject the potential for outlying points to be considered as a part of that cluster. The accept ratio represents only data points that have a very strong potential for being cluster center. The reject ratio sets the potential below which a data point is rejected as a cluster center. The same twenty components are used for subtractive clustering. It was observed that the subtractive clustering produced steady results on repeated iterations as compared to the fuzzy c-means.

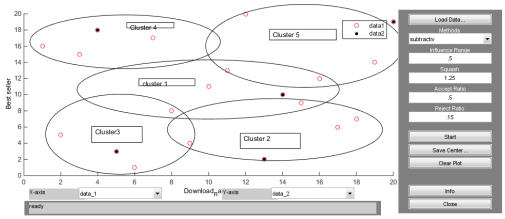

Figure 2. Kaur: Subtractive clustering on components.

```
function [out] =hybridxorFunc(d1,d2)
  col1=size(d1,2);
  col2=size(d2,2);
if col1~=col2
     error('the component pair dont have equal features');
end
  out=zeros(1,col1);
for i=1:col1
if d1(1,i)==0 && d2(1,i)==0
        out(1,i)=0;
elseif d1(1,i)==0 && d2(1,i)==1
        out(1,i)=0;
elseif d1(1,i)==1 && d2(1,i)==0
        out(1,i)=0;
elseif d1(1,i)==1 && d2(1,i)==1;
        out(1,i)=1;
end
end
```

Figure 3. Kaur: Hybrid XOR script in MATLA

2.3 Hybrid XOR based Clustering

Another type of clustering is the hybrid XOR based clustering[11]. The clustering approach based on this similarity function is used for finding extent of likeness between component clusters. A similarity matrix of the order n-1 by n is constructed for given n components. The hybrid XOR, implemented in MATLAB, used to find the similarity of two components is shown in Figure 3. The appropriate feature vector representation of the component set is made. This is executed for set of thirteen searching and sorting components.

The clusters that are generated after its application are as follows:

1. The first cluster consists of (C1, C2, C3, C4, C5, C6, C9).
2. The second cluster is (C7, C8, C10).
3. The third cluster is (C11, C12, C13).

2.4 FREFCOSCO (Fuzzy Relation based Fuzzy Clustering Of Software Components)

Recently, the authors have proposed another fuzzy clustering technique based on fuzzy relations. The algorithm is initialized with a matrix of components; the features of the components are both numerical and categorical types. Firstly normalization of the data, in terms of converting the categorical data to numerical one and assigning appropriate labels to the features is done. Further, the fuzzy relation matrix is formed by comparing the attributes of two components. The second step consists of finding the fuzzy transitive closure so that it can be used as a similarity measure. The third step will generate the partitions from the matrix obtained in step 2. The entries in the partition matrix can be obtained by using set similar to property. The fourth step, generates the final fuzzy clusters as the components are repeated in many clusters. Theimplementation of the fuzzy relation based fuzzy clustering algorithm is represented in Figure 4.

Running this algorithm for twenty components, will generate thirty three smaller clusters, from there the application developer can choose the most suitable component. Taking the best component from the cluster is based on silhouette coefficient and this can be explored further. So, the components with maximum cohesion within the cluster and minimum coupling outside the cluster are chosen.

Figure 4. Kaur: FREFCOSCO Algorithm.

3. Conclusion

The main issue in fuzzy c means and subtractive clustering is its inability to deal with multi-attributes of the components at a time. The simple fuzzy clustering technique requires the need of mentioning the number of cluster centers beforehand, in case of Fuzzy c-means and the radii of the cluster in case of Subtractive clustering. The main demerit in Hybrid XOR based clustering is its dependence on subjective judgment of application developer. The last algorithm as proposed by the authors need to validate the components by appropriate validity mechanism. It has eliminated the usage of similarity index. One is usage of silhouette coefficient. The community of researchers is still struggling with these issues in order to achieve an optimal fuzzy clustering algorithm for fuzzy clustering.

4. References

1. Nazir S, Anwar S, Khan SA, Shahzad S, Ali M, Amin R, Cosmas J. Software Component Selection based on Quality Criteria using the Analytic Network Process. Abstract and Applied Analysis. 2014 Dec; 2014:1–12
2. Nakkrasae S, Sophatsathit P, Edwards JW. Fuzzy Subtractive Clustering based Indexing Approach for Software Components Classification. International Journal of Computer and Information Science. 2004 Mar; 5(1):63–72.
3. Serban C, Vescan A, Pop HF. A New Component Selection Algorithm Based on Metrics and Fuzzy Clustering Analysis. Proceedings of International Conference on Hybrid Artificial Intelligence Systems, Berlin Heidelberg. Springer; 2009. p.621–8.
4. SerbanC,Vescan A, Pop HF. A Formal Model for Component–Based System Assessment. Proceedings of 2nd International Conference on Computational Intelligence, Modelling and Simulation (CIMSiM), Indonesia. 2010; p.261–6.
5. Quanming Z, Jianhua H. An Advanced Fuzzy Cluster Analysis Algorithm and the Applications based on Fuzzy Equivalence Relation. Proceedings of 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA), Canada. 2013. p.616–9.
6. Yang MS, Shih HM. Cluster Analysis based on Fuzzy Relations. Fuzzy Sets and Systems. 2001 Jun; 120(2):197–212.
7. Guh YY, Yang MS, Po RW, Lee ES. Establishing Performance Evaluation Structures by Fuzzy Relation-Based Cluster Analysis. Computersand Mathematics with Applications. 2008 Jul;56(2):572–82.
8. Bataineh K, Naji M, Saqer M. A Comparison Study between various Fuzzy Clustering Algorithms. Jordan Journal of Mechanical and Industrial Engineering. 2011 Aug; 5(4):335–43.
9. Sajana T, Rani CMS, Narayana KV. A Survey on Clustering Techniques for Big Data Mining. Indian Journal of Science and Technology. 2016; 9(3):1–12.
10. Krishna TVS, Babu AY, Rao A. Robust Fuzzy C-Means Cluster Algorithm through Energy Minimization for Image Segmentation. Indian Journal of Science and Technology. 2016;9(22):1–12.
11. Kaur J, Tomar P. Four Tier Architecture for Software Component Selection Process using Clustering. International Journal of Software Engineering and Application Technology.2015;1(2-4):155–71.