

IMPROVING THE SOFTWARE ARCHITECTURE THROUGH FUZZY CLUSTERING TECHNIQUE.

Shaheda Akthar¹, Sk.Md.Rafi²

1.Department of Computer science and Engineering,, Sri Mittapalli college of Engineering,A.P. INDIA.

2.Department of Computer science and Engineering,, Sri Mittapalli Institute of Technology for Women,A.P. INDIA

Abstract: Software Architecture Recovery is one of the finest parts of reverse engineering. Several different techniques have been adopted in vast literature to recover Software Architecture. One of the techniques is clustering, which extracts the similar components from the software. In general the component characteristics and its state are not clear. For that, Architecture recovered from ordinary clustering will not be appropriate in it. In this paper we used a fuzzy clustering technique to make the Software Architecture recovery to be more efficient and accurate. Our experimental results have shown that Architecture recovers from fuzzy clustering is better than ordinary clustering

Keywords Software Architecture Recovery, Fuzzy Clustering, fairness index.

6.1 Introduction

Software Architecture Recovery is the process of extracting the software architecture from given large source code. Over the years many software decomposition techniques are proposed like clustering technique and pattern matching technique. Generally software recovery is part of reverse engineering process. Clustering has been used in many disciplines to support upon similar object of a system [3]. Clustering technique has been used since long time in science and engineering. Clustering techniques has been used in software engineering in different phases. Many clustering techniques have been presented in recovery process [3, 4]. But when we look into the recovered architecture only maximum 90% of original architecture. But when we observed the nature of the components in the software, their behavior is not accurate. The components does not share the common set of attribute like main classes,

sub classes, internal classes and external classes. Some of the classes and their characteristics lie between the two clusters. Components show a fuzzy characteristic in nature. Some of the components apparently does not belong to the neither of the cluster called outliers. Many clustering algorithms do not address this issue. For that I proposed a new architecture recovery by the fuzzy clustering [3]. One of the most popular fuzzy clustering algorithms is called the fuzzy C-Means (FCM) algorithm [1, 2, 5, 6]

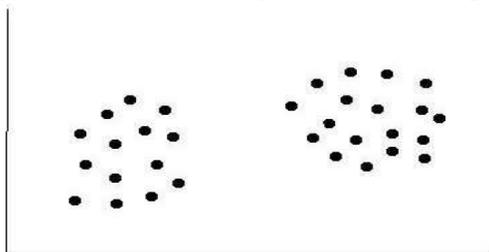


Fig 6.1 sample clusters

6.2 Clustering technique

6.2.1 General clustering technique

General clustering technique is done by identifying the number of data sets, calculating resemblance coefficients and executing the clustering method. Here dataset is a object matrix. Clusters are groups of similar data elements. Resemblance coefficient represents the degree of similarity and non similarity between the items. The main aim of clustering analysis is identify and quantification of these architecture elements. Identifying the membership and location of center of the clusters are main process in the cluster analysis. Some time data in the cluster is well packed. But due to the complex nature of the components the data may not be packed well in the clusters. Some of the elements lies outside the cluster region. Several clustering algorithms are [4] presented un weighted pair group (UPGMA). Clustering techniques are used to minimize the interconnectivity among the components.

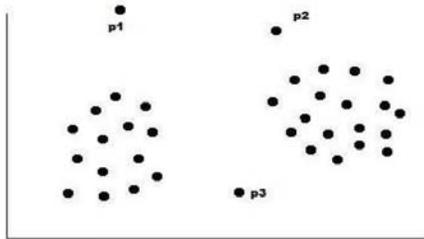


Fig 6.2 clusters with outlier points

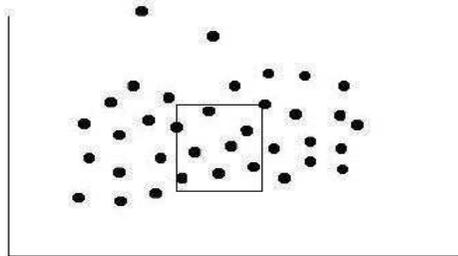


Fig 6.3 Ambiguous dat points in two clusters

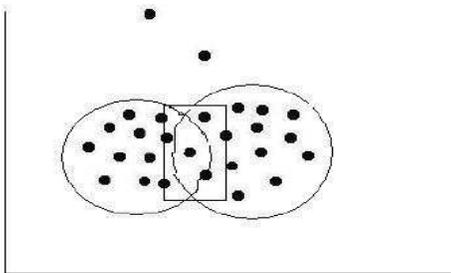


Fig 6.4 common points in two neighboring clusters

6.2.2 Fuzzy clustering

Fuzzy clustering has been used in many applications like image processing, communication devices and software engineering. In a fuzzy clustering approach data points are given partial degree of membership in multiple nearby clusters. Central point in the fuzzy clustering is always no unique partitioning of the data in a collection of clusters. In this we assign the membership value to each cluster. Sometimes this membership has been used to decide whether the data points belong to the cluster or not [3]. A fuzzy clustering provides the flexible and robust method of assigns the data points to the clusters. Each data point has associated degree of membership which specifies the strength of its placement into the cluster. Each cluster is represented as the circles. As the data moves from the center of the cluster to the diameter its degree of membership decreases and reaches to zero value. Fuzzy clustering technique gives us more realistic approach than other techniques. Here we used the fuzzy C means algorithm to find the related data items which share the common properties.

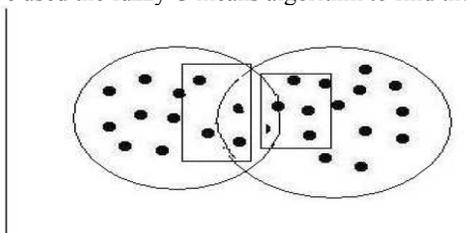


Fig 6.5 overlapping clusters

6.3 Fuzzy C means algorithm [3]

6.3.1 Technique

The original crisp methodology is modified and studied by Bezdek. He introduces a fuzzification parameter m such that m ranges from $[1, n]$ which determines degree of fuzziness in the cluster. The parameter m controls the degree of fuzziness among the clusters. When $m=1$ the clusters does not match where as if $m > 1$ the circles of clusters begins to come closer to each other. Degree of closeness to the cluster is determined by m which is calculated from the degree of closeness of the components. Following algorithm defines the fuzzy clustering technique for software architecture.

Algorithm:

- 1) N_i represents the all data points in the software $i=1,2,\dots,n$.
- 2) $S_{i,j}$ the distance of i th data point from the j th cluster.
- 3) P represents number of fuzzy clusters.
- 4) m a fuzzification parameter in the range $\{>1 \text{ and } <2\}$
- 5) C^j is the center of the fuzzy cluster
- 6) $A_j(N_i)$ a fuzzy membership qualification indicating the membership of sample N_i to the j th cluster.

6.3.2 Fuzzy clustering technique for the software architecture recovery

In this we will see the process of software architecture recovery by fuzzy clustering. At first identify the data sets present in the software. Then calculate the degree of relatedness of these components. Then apply the fuzzy c means algorithm to reconstruct the components obtained. Fuzzy c means clustering involves with two phases

- 1) Calculate the cluster centers
- 2) Assign these points to the clusters

This process is repeated until the cluster center is stabilized.

$$\sum_{j=1}^P \alpha_j(N_i) = 1 \quad i=1,2,3 \dots \dots \dots h$$

Here P is the numbers of clusters and h is the number of data points.

Now we will define the equation of assignment of data points into clusters with varying degree of member ship. By this we come to know at which cluster the data point has to be assigned. Now the following equation is needed to calculate the new cluster center value :

$$C^j = \frac{\sum_i (A_j(N_i))^m \times N_i}{\sum_i (A_j(N_i))^m} \quad j \text{ represents the center of the } j\text{th cluster, } N_i \text{ represents the } i^{\text{th}} \text{ data point and } m \text{ is the fuzzification parameter.}$$

6.4 Experimental Results

Here I have implemented a C++ program. And results are compared with ordinary clustering. Here we used the fairness index ν such that it defines as degree of closeness of recovered architecture with original architecture.

$$\nu = \frac{\text{sum of the components in recovered software architecture}}{\text{sum of the components in the original architecture}}$$

Greater the value of ν efficient is the technique is

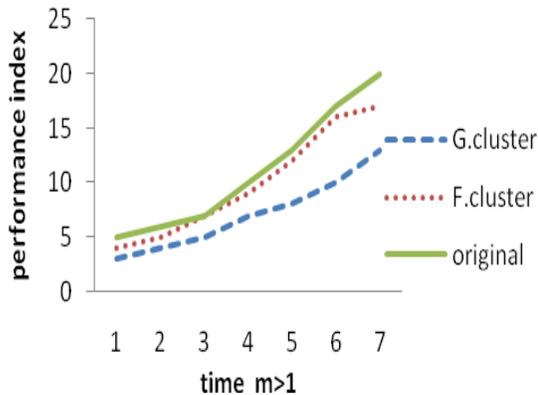


Fig 6.7 Performance Analysis

Conclusion This paper presents the applicability of Fuzzy clustering in Software architecture Recovery. The architecture which is recovered from fuzzy clustering is more appropriate and efficient than ordinary clustering techniques.

References

- [1] A.Kandel, "Fuzzy Techniques in Pattern Recognition" John Wiley & Sons, Inc, 1982[**]
- [2] Bezdek, J.C., Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, 1978[*]
- [3] Chin-Cheng Hung, Wenping Liu and Bor-Chen Kuo. A new adaptive fuzzy clustering algorithm for remotely sensed images. [14]
- [4] Markus Bauer, Mircea Trifu. Combining Clustering with Pattern Matching for Architecture Recovery of OO Systems[47]
- [5] Zhou, J. and C.C. Hung, "A Generalized Approach to Possibility Clustering", International Journal of Uncertainty, Fuzziness & Knowledge-Based Systems Vol. 15., 110 – 132, Suppl, April 2007[***]
- [6] Zhou, J., C.C. Hung, J. He and Y.Luo, "Normalized Possibilistic Clustering Algorithms", The 6th International conference on Information and Management Sciences, July 2000[****]

Shaheda Akthar received Bachelor of computer science & Master of Computer Science from Acharya Nagarjuna University, M.S (Software Systems) from BITS, Pilani, Pursuing PhD from Acharya Nagarjuna University. Presently working as Associate Professor in Sri Mittapalli College of engineering, affiliated to J.N.T.U, Kakinada. My area of interest is Software Reliability, Software Architecture Recovery, Network Security, and Software Engineering

Sk.MD.Rafi received B.Tech (comp) from Jawaharlal Nehru Technological University, M.Tech (comp) from Acharya Nagarjuna University. Pursuing PhD from Jawaharlal Nehru Technological University. Presently working as Associate Professor in Sri Mittapalli Institute of Technology for women, affiliated to J.N.T.U, Kakinada. My area of interest is Software Reliability, Software Architecture Recovery, Network Security, and Software Engineering.