

To propose a new algorithm for improving software architecture with clustering algorithm

Jyoti Godara¹ and Shilpa Sharma²

¹Department of Computer Science Engineering, Lovely Professional University, India
Email: jyoti.poonia6@gmail.com

²Department of Computer Science Engineering, Lovely Professional University, India
Email: shilpa.sharma15@yahoo.com

Abstract: The size and difficulty of business potential software systems are continually rising. This means that the task of supervision of a large software project is becoming even more demanding, particularly in light of high turnover of skilled human resources. Software clustering approaches can help with the task of understanding large, complicated software systems by automatically decomposing them into smaller, easier-to-manage subsystems. In this paper we recognize significant study instructions in the area of software clustering that require further attention in order to develop more effective and efficient clustering methodologies for software engineering. To that end, we first there the state of the art in software clustering research. We consider the methods of clustering that have received the most attention from the research community and outline their strengths and weaknesses. Our paper defines every stages of a clustering algorithm separately. We also present the most important approaches for evaluating the value of software clustering.

Keywords: Software Architecture, Clustering, K-Mean Clustering

1. Introduction

Software engineering is the branch of study and application of engineering to the software which initially design development of the same software by using various technique or updating, maintenance of the software. Software engineering deals with the all kind of software production, design to coding, software accuracy and deals with the complexity of any software system. Software industry is moving very fast in the current scenario. Even big industries spend large some of amount on their software engineer for the software development. Software Architecture is professional who design software or other computer application, or who prepare plans and strategy who upgrade it from time to time or according or user requirement. In other words software architecture is a process to defining a proposed solution that fulfills the entire technical and operational requirement offered to the user [1].

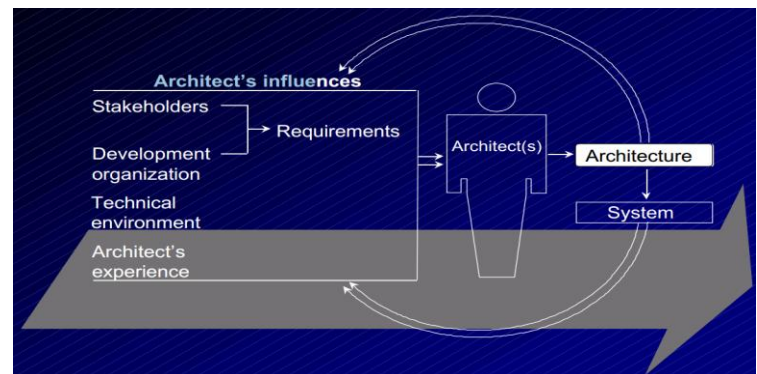


Figure1. Software Architecture [11]

2. Clustering

The basic idea behind clustering is to create the groups together for similar object or for similar purposes [9]. In other word clustering may be define as a portioning of data contain into subset or in the small size cluster. Clustering wide use of algorithms are (K-Mean, C-Mean etc.)

2.1 Cluster Analysis

Cluster analysis may be define as a creating group of object in such a way that the object in a group will be related to one other and un related to the object in another group. We can't say that cluster analysis is not restricting to one specific

algorithm, but its main motive to solve the task. It can be also achieved by using algorithm but in different manner^[3]. Clustering analysis used in various felids that is biology, trade and business. In each of the field clustering analysis has is on importance. In broader view we can say that clustering analysis is not limited to computer team as it is contributing to other felids also.

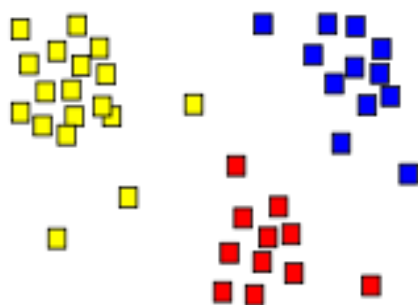


Figure2. Cluster analysis^[10]

2.2 Types of Clustering

Cluster: It is said to be “Collection of data objects”. The two types of similarities of clustering’s:-

- a. **Intra class similarity** - Objects are similar to objects in same cluster
- b. **Interclass dissimilarity** - Objects are dissimilar to objects in other clusters

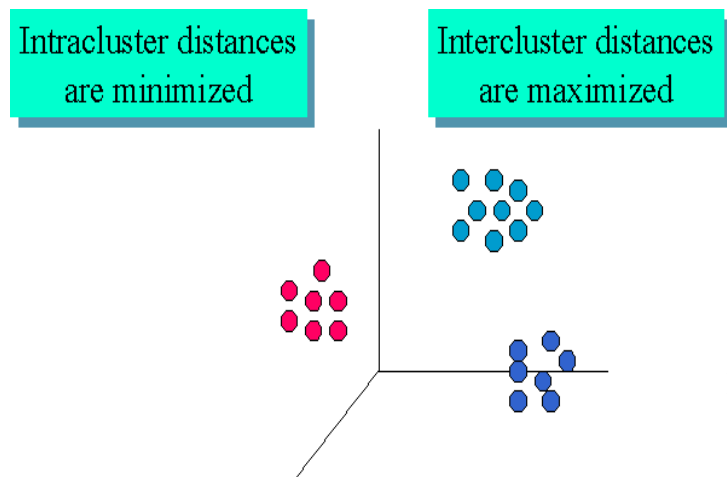


Figure3. Type of clustering similarity^[10]

2.3 Applications of Clustering

- a) **Data Mining:** - Clustering is often one of the first steps in data mining analysis. It identifies groups of related records that can be used as a starting point for exploring

further relationships. This technique supports the development of population segmentation models, such as demographic-based customer segmentation^[6]. Additional analyses using standard analytical and other data mining techniques can determine the characteristics of these segments with respect to some desired outcome. For example, the buying habits of multiple population segments might be compared to determine which segments to target for a new sales campaign.

- b) **Search Engines:** - Where information retrieval system is works in the web documents on it. The document source is said to be the documents of the web page. The query is said to be the search engine^[7]. Using cluster the documents are classified based on the query in the information retrieval system. The ranked documents represent the relevant details present in the documents which are relevant to the search of the query. It is the mining of the data in the web page in the database websites.
- c) **Text mining:** - Text mining consists of extraction information from hidden patterns in large text-data collections^[8]. The query is given in the system were the given query is been founded by using the search navigation system. Where the documents based on query search is been given here in the diagram. Where is been extracted using name extractor. From the authorization list the ranking details are viewed on it.

3. Clustering Techniques

3.1 Fuzzy C-Means

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. 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. 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 [5]. 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 [9]. Here we used the fuzzy C means algorithm to find the related data items which share the common properties.

3.2 K-Means Clustering

Simply the K Means is an algorithm to classify or to group your objects based on attributes/features into K number of group. K is positive integer number [4]. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus, the purpose of K-mean clustering is to classify the data.

3.3 Hierarchical Clustering

Hierarchical clustering (also called hierarchical cluster analysis or HCA) is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types [2]:

a) Agglomerative: This is a "bottom up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

b) Divisive: This is a "top down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

4. Advantages of Clustering

- a) Easy to repair
- b) Protect the shore
- c) Function after minor damage
- d) Don't disturb the shoreline

5. Disadvantages of Clustering

- a) Expensive
- b) Ugly
- c) Ruin the surfing tourism industry
- d) Can be displaced easily by waves

e) Don't absorb all energy due to gaps between the breakwaters

6. Conclusion

It is imperative to make technology decisions at the good time with good techniques and for the good logics. For Batter business suggests good people with proper supporting tools so they can develop very effective products. When it comes to establish the software, that time handle difficult language problem head-on is one constraint for today's creative manager. When combined with alternative software engineering applications, a effective language decision can support the cost-effective software systems development that, in turn, it arrange beneficial and effective, good support of business. In this paper presented the state of development and the evaluation methodologies of software clustering. We also describe the most valuable research challenges for this valuable research area. It should be also feasible that while the most valuable advances have already taken place, there are still many different paths for more research which will more effective for software engineers far and wide.

REFERENCE

- [1] Lingming Zhang, Ji Zhou, Dan Hao ,Lu Zhang, Hong Mei" *Prioritizing JUnit Test Cases in Absence of Coverage Information*" IEEE 2009.
- [2] Lior Rokach, Oded Maimon " *Clustering Method*" Tel-Aviv University DATA MINING AND KNOWLEDGE DISCOVERY
- [3] Yorku.ca PeterTryfos," *Clustering Analysis*"1997.Thisversionprinted:14-3-2001
- [4] Leo Buttou ,Yoshua Bengio" *Covergence properties of K-Means Algorithm*" AT&T Bell Labs, Holmdel, NJ 07733
- [5] Richard J. Hathaway " Fuzzy c-Means Clustering of Incomplete Data" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 31, NO. 5, OCTOBER 2001
- [6] G. Pour, " *Component-Based Software Development Approach: New Opportunities and*

Challenges,” Proceedings Technology of Object-Oriented Languages, 1998. TOOLS 26., pp. 375-383.

[7] Ji-Rong Wen, Jian-Yun Nie “Clustering User Queries of a Search Engine” 10, May 1-5, 2001, Hong Kong. ACM 1-58113-348-0/01/0005

[8] Michael W. Berry and Malu Castellanos, Editors” Survey of Text Mining:Clustering, Classification, and Retrieval, Second Edition September 30, 2007

[9] Shaheda Akthar¹ , Sk.Md.Rafi , “*Improving the Software Architecture through Fuzzy Clustering Technique*” Vol 1 No 1 54-57

[10] [wikipedia.org/wiki/ Cluster_analysis](http://wikipedia.org/wiki/Cluster_analysis)

[11] matei/EECE417/BASS/ch01lev1sec1