IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved**, **Volume 11, Issue 1 , 2022**

REUSABILITY PREDICTION IN AN EDUCATIONAL ENVIRONMENT

Dr. N Murali Krishna
Professor
Department of Computer Science and Engineering
VIGNAN Institute of Technology & Science, Hyderabad, India
India E-mail: muralinamana@gmail.com

Abstract

A great deal of research over the past several years has been devoted to the development of methodologies to create reusable software components but the successfulness of the reuse program is still in its naïve stage. In this paper a novel method for estimating the reusability across an educational domain is presented. The data collected is clustered and the K-means algorithm is applied to the clustered data which is of the matrix form containing the binary data. The relevant data is extracted based on the test data. This method results in identifying the K-mean variants and following the reusability activities for which the reusability factors can be predicted.

Keywords-Reusability, Clustering,K-means

1. Introduction

This study explains the basic scenario of predicting the reusability across an educational environment where considering the vast increase of the educational environment across the globe with the simultaneous increase of technology for maintaining the data related to the system

In the traditional system of college environment similar courses are being initiated in different regulations and with different titles. The usage of data for different educational systems has been a problem of concern since redundant data is being used across the educational groups. In order to minimize the redundancy and to maximize the

retrievals a scheme is to be developed varying the most frequently repeated items can be integrated and when needed the data can be dumped. The mechanism decreases not only storage area but also maximizes computational efforts .In the case of software environment this mechanism is very useful as the cost of the code and time minimizes by considering reusability applications.

In this project a new model for estimating the reusability of the core structures in various educational colleges have been presented. In this approach to predict reusability components the various department profiles, core structures, modules were studied and the commonalities have been identified.

K-means algorithm is utilized to segment the modules and reusability component is calculated using method prescribed by Basili**[4].**The main advantage of the proposed model is that as the number of organizations increases we can create the methodologies for reuse and even component libraries. In the traditional methods of the educational methods these methodologies does not exists and every data is to be created.

The paper is organized as follows, in section-2,the reusability issues are presented,section-3 deals with clustering the data ,using K-mean algorithm, in section-4 ,the methodology is presented and in section -5,the experimental data is given with experimental data ,finally in section -6 conclusions are given . To demonstrate our model we have to collect the data of the engineering colleges in India using the internet.

2. Reusability:

Reusability is considered as the basic concept of software engineering playing an important role in the software development and also occupies a significant area in research and practice related to the software engineering domain as a means to reduce the development costs, time and improved quality .It allows new programs to be assembled quickly from the existing components by building a library of frequently used components.

Two approaches for reuse are 1) develop the reusable code from the scratch (or) 2) identify and extract the reusable code from already developed code.

Generally the cost of developing the software can be saved by identifying and extracting components from already developed systems or legacy systems[1]. The concept of reusability is universally accepted basing on the fact that "A product will work properly if it has already worked before", but the issue of how to identify the reusable components has remained unexplored. After reusing there is a need to evaluate the quality of potentially reusable piece of software. The contribution of metrics to overall quality of the software need to be recognized[2]-[3] but how these collectively determine the reusability of software components is at its early stage.The reusability ranking of software is generated by identifying the reusability components from the database and grouping the relevant data in the form of a dissimilarity matrix where the relevant items are clubbed and a value of 1 is assigned and if the data are irrelevant value of 0 is assigned and a matrix is generated consisting of 0's and 1's based on the relevancy .This matrix is called training data in order to identify reusability component, the test data is also converted in the form of dissimilarity matrix containing 0's and 1's and this data is compared with training data for further analysis.

2.1 Reusability activities

The reusability activities[**5**] that are common across phases are

1. Studying the problem and finding the available solutions to the problem and develop a reusable plan.
2. Identifying a solution structure by following a plan.
3. Using structure for the next phase
4. Evaluating the end product

The reuse activity in the current project is divided into major steps as follows

1. Developing a database from 5 different universities and identifying the available solutions
2. Identifying the structure, for solving the problem using the reusable plan
3. Reconfiguring the solution structure for the next phases
4. Evaluating the product

The major task under the first step is to understand the problem domain and identify the reusability components and find the reusability components and find the alternative ways of solving these reusability components. In the next step is to develop a structure that is well suited for the problem following the reusable plan stated in the first phase. various attributes (groups)that are to be identified as the reusable components and methodologies to be evaluated to solve this components. Once the structure and the reusability components are identified we have to optimize the data such that the reusable components can be carried to the next phase that is when we consider an educational domain, in Andhra at the time of engineering admissions students are required to choose the colleges of interest through online web services .

In this scenario, the concept of reusability plays a vital role where the various universities ,inside Andhra are provided and the various commonalities in each of these universities are identified by decomposing it into 4 groups such as i)courses being offered ii)evaluation system iii)syllabus and iv) subjects. The student has to give the set of alternatives of the choices for his admissions. In this scenario we built a framework of reusability where we try to group the colleges that have the commonalities with respect to above 4 groups .since we have identified the commonalities within these groups the student is benefitted such that he can have an idea about the courses that he is going to study and various similar options that can choose ,for opting in the web counseling .

The developed system will be carried out into the next phase and will be very beneficiary to the student committee. The amount of reusability can be calculated using the model proposed by Basili**[4].**

3. Clustering

The goal of clustering also known as cluster analysis is to discover the groupings of a set of objects ,points or patterns .Webster [Merriam-Webster Online Dictionary ,2008] defined cluster analysis as “A statistical classification technique for discovering whether the individuals of a population falls into different groups by making quantitative comparisons of multiple characteristics”.The operational definition can be stated as : Given a representation of n objects ,find k groups based on the measure of similarity such that the similarities between objects in the same group are high while the similarities between the objects in the different groups are low that is on the criteria of” maximizing the inter cluster similarities and minimizing the intra cluster similarities”.

Clustering can be used for following purposes

i) Underlying structure: to gain insight into the data and identify the salient features

ii) Natural classification: to identify the degree of similarity among the objects

iii) Compression: a method for organizing and summarizing through prototypes.

In this step, k-Means clustering algorithm is used for partitioning the data into different level of reusability value based on the structural metric values as k-means is the well known approach that classify data into different k groups where K is a positive integer. Grouping of data is done on the basis of minimizing sum of squares of distances between data and their cluster centroid.

3.1 K-means algorithm:

Let $X=\{x_i\}$, i =1,…n be the set of n dimensional points to be clustered into a set of K clusters ,$C=\{c_{k,}$ k=1…..K). partition is done using K-means algorithm such that squared error between the empirical mean of a cluster and the points in the cluster is minimized.

Let the mean of the cluster c_k be μ_k. The squared error between μ_k and the points in the cluster c_k is defined as

$$J\,(c_k) = \sum_{xi \in ck} \left\| x_i - \mu_k \right\|^2 \quad . \ldots\ldots\ldots. (1)$$

The goal of K-means is to minimize the sum of the squared error over all K clusters,

$$J(C) = \sum_{K=1}^{K} \sum_{xi \in ck} \left\| x_i - \mu_k \right\|^2 \quad \ldots\ldots\ldots .(2)$$

K-means start with an initial partition with K clusters and assign patterns to clusters so as to reduce the squared error .since the squared error always decrease with an increase in the number of clusters K (with J(C)=0 when K= n), it can be minimized for a fixed number of clusters.

The steps of K-means algorithm are as follows :

1. Select an initial partition with K clusters; repeat steps 2 and 3 until cluster membership stabilizes.
2. Generate a new partition by assigning each pattern to its closest cluster center.
3. Compute new cluster centers.

3.2 Parameters of K-means

The K-means algorithm requires three user specified parameters: number of clusters K, cluster initialization and distance metric .K can be chose using a number of heuristics when no perfect mathematical criterion exists as proposed by [5] **[Tibshirani et.al].**K-means is runs independently for different values of K and the partition that appears most meaningful to the domain expert is selected.

4) Methodology

The project serves the purpose of the user searching for the information regarding the courses that are offered in that particular institution and selects the appropriate one depending on the availability .The search criterion is based on the coincidence of the courses in various colleges for which the person is intended to pick .Our model is beneficial for the person who prefers to join the engineering courses after the completion of the basic education ,by

checking the availability of courses in their respective colleges ,if the desired course is not available in a particular college he can choose the other college which has the similar course there by referring the commonalities in the searching criteria .This commonality is compared with the concept of reusability referring to various aspects of core structures and evaluation criterion to be the reusable components. The database is generated by pooling the relative information from the engineering colleges in India using the internet. In this database we have generated 4 groups from 5 universities as i)courses being offered ii)evaluation system iii)syllabus and iv) subjects in each branch. The data that is common is considered and a matrix is generated. If the data is common 1 is assigned or else 0 is assigned there by a binary matrix is created, called the dissimilarity matrix. Now, basing on the query the data is segmented into clusters ,so that query is within each of these 5 groups .now if query processed is within this group ,we say that it is a reusable component if it matches else other .

5) Experimentation:

In order to evaluate our model we have to generate a database with 5 universities and identify the core groups such as i)courses being offered ii)evaluation system iii)syllabus and iv) subjects. The data is verified across 5 universities with respect to the 4 core groups and a matrix called dissimilarity matrix is generated. The data that is common and which can be reused is enumerated as 1 otherwise 0.Now, in order to make this data beneficiary to student we have trained data and applied the clustering algorithm such that the data pertaining to each course is at one corner.Inorder to test the data student makes a query and the query is compared with that of the clustered data and the result is obtained. The methodology is evaluated by considering the heuristics presented by Basili**[4].**The reusability concepts that are covered are specified in section-2.1.

U B	GU	AU	NU	OU	SVU
IT	1	1	1	1	1
CSE	1	1	1	1	1
EEE	1	1	1	1	1
ECE	1	1	1	1	1
EIE	1	1	0	1	1
ME	1	1	1	1	1
IE	1	0	0	0	0
A_OE	0	0	0	0	1
A_EE	0	0	1	0	1
CE	1	1	1	1	1
C_HE	0	1	1	0	1
BT	1	1	1	1	0
BI	0	1	1	1	0
BME	0	0	1	1	0
BPHRMCY	1	0	0	0	0
EM	1	0	0	0	0

Table 1-screenshots of the input data.

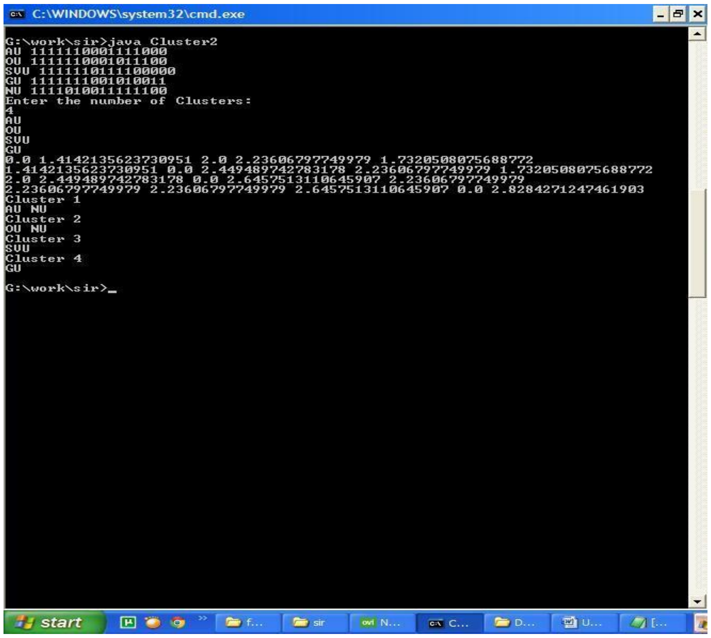
The relevant screenshots of input data and output extracted are presented in table-1 and table-2 .

To evaluate our method, the test data is considered (the query to be posed) and is converted in the form of a matrix(dissimilarity).The data is clustered using K-means clustering ,based on the binary data as mentioned by Tao Li**[6]-[7].**The relevant data is clustered and the various data elements in each clusters ,depending on the query data, the distance is computed and basing on the nearest distance the relevancy is established.

6. Conclusion

In this paper a novel methodology for software reuse architecture is presented for education domain. The data is clustered using K-means algorithm and the dissimilarity matrix is constructed and from the clustered binary data based on the test data, the relevant data is extracted.

This method is very much useful and it reduces the cost and Lines of code as some of the redundant data can be used for further need.



```
G:\work\sir>java Cluster2
AU 1111110001111000
OU 1111110001011100
SVU 1111101111100000
GU 1111111001010011
NU 1111010011111100
Enter the number of Clusters:
4
AU
OU
SVU
GU
0.0 1.4142135623730951 2.0 2.23606797749979 1.7320508075688772
1.4142135623730951 0.0 2.449489742783178 2.23606797749979 1.7320508075688772
2.0 2.449489742783178 0.0 2.6457513110645907 2.23606797749979
2.23606797749979 2.23606797749979 2.6457513110645907 0.0 2.8284271247461903
Cluster 1
AU NU
Cluster 2
OU NU
Cluster 3
SVU
Cluster 4
GU

G:\work\sir>_
```

Table-2:screenshots of the extracted output .

5.1 Appendix

B-Branch, **U**-university, **GU**-GitamUniversity,**AU**-AndhraUniversity,**NU**-NagarjunaUniversity,OU-OsmaniaUniversity ,**SVU**-SriVenkateswaraUniversity.**IT**-Information Technology, **CSE**-Computer science and Engineering **, EEE**-Electrical and Electronics Engineering, **ECE**-Electrical and Communication Engineering, **EIE**-Electrical and Instrumentation Engineering, **ME**-Mechanical Engineering**,** **IE**-Industrial Engineering, **A_OE**-Automobile Engineering, **A_EE**-Aeronautical Engineering, **CE**-Civil Engineering, **C_HE**-Chemical Engineering, **BT**-Bio Technology, **BI**-Bio Informatics, **BME**-Bio Medical

Engineering**, BPHRMCY**-Bachelor of Pharmacy .**EM**-Environmental Management.

Reference

[1] G.CaldieraandV.R.Basili,"Identifying and Qualifying Reusable Software components",IEEE Computer, February 1991,pp.61-70.

[2]. .Humphrey,Managing the Software Process,SEI Series in Software Engineering ,Addison-Wesley,1989.

[3].R.S.Pressmen,SoftwareEngineering:A Practitioner'sApproach,McGraw-Hill Publications,5th edition,2005.

[4] Basili,V.R.(1989)"Software development: A Paradigm for the future", Proceedings COMPAC'89,LosAlamitos,California,IEEE CS press,1989,pp.471-485.

[5] Ajay Kumar (2012) "Measuring Software Reusability using SVM Bases Classifier Approach" International Journal of information Technology and Knowledge Management January-june 2012,Volume 5 ,No. 1,pp.205-209.

[6] Tao Li, Shengzhou Zhu (2005) "On Clustering Binary Data". SIAM, proceedings in data mining.

[7] Tao Li School of Computer Science ,Florida International University "A unified View On Clustering Binary Data", September 30,2005.

[8] Jain, A.K. Data clustering: 50 years beyond K-means. Pattern Recognition Letters.(2009), doi:10.1016/j.patrec.2009.09.011.

[9]Tibshirani,R.,Walther,G.,,& Hastie,T.2001."Estimating the number of clusters in the data set via the gap statistic." Journal of the Royal Statistical Society, 411-423.

UNIVERSITY BRANCH	GITAM UNIVERSITY	ANDHRA UNIVERSITY	NAGARJUNA UNIVERSITY	OSMANIA UNIVERSITY	SRI VENKATESWARA UNIVERSITY
	UNDER GRADUATE COURSES				
INFORMATION TECHNOLOGY	1	1	1	1	1
COMPUTER SCIENCE ENGINEERING	1	1	1	1	1
ELECTICAL& ELECTRONICS ENGINEERING	1	1	1	1	1
ELECTRICAL& COMMUNICATION ENGINEERING	1	1	1	1	1
ELECTRICAL& INSTRUMENTATION ENGINEERING	1	1	0	1	1
MECHANICAL ENGINEERING	1	1	1	1	1
INDUSTRIAL ENGINEERING	1	0	0	0	0
AUTOMOBILE ENGINEERING	0	0	0	0	1
AERONOTICAL ENGINEERING	0	0	1	0	1
CIVIL ENGINEERING	1	1	1	1	1
CHEMICAL ENGINEERING	0	1	1	0	1
BIOTECHNOLOGY	1	1	1	1	0
BIOINFORMATICS	0	1	1	1	0
BIO MEDICAL ENGINEERING	0	0	1	1	0
BACHELOR OF PHARMACY	1	0	0	0	0
ENVIRONMENTAL MANAGEMENT	1	0	0	0	0