Measurement of Employee Productivity using Cluster Analysis of BehavioralIntegrity

AnupamaVishwasGajbhiye , JaiprakashKashinathSirur

[1]*avgajbhiye@yahoo.co.in*, [2]*jksirur@gmail.com*
Department of Computer Science
J.J.T.University, Chudela,Jhunjhunu,Rajasthan

Abstract: *In the area of Human Resource Development there are two known methods:(1)In the first method organizations identify people who can do the job and then find out how their knowledge and skills are different from people who have just been recruited.(2) In the second method there is a need to identify people who can do the job well and perform better.*

Neither of the above two methods solve the problems of getting good people into jobs so that they do well in terms of Integrity. It is therefore important to investigate the relationship of the Productivity of the employee measured using a test of General Intelligence of the employee and the Behavioral Integrity of employee.

In the present work we study the Comparative Performance of individual employee Scores of their : (i) General Intelligence(GI) and (ii) Behavioral Integrity(BI) by using the Synthetic Data of ten Grade-A employees working in a University.

In the proposed work Cluster analysis is performed with two objectives to investigate: (1) The Interconnectedness and Closeness amongst the employee Clusters.(2) The Cophenetic Correlation coefficient used to determine whether General Intelligence(GI) of the employee is related to the Behavioral Integrity(BI).

Keywords: *General Intelligence (GI) ,Behavioral Integrity (BI), Cluster Analysis*

I. INTRODUCTION

Human Resources and Personnel Planning and Development is a module in the Enterprise Resource Management which is concerned with the development of employees. The purpose of personnel planning is to determine who will be required and when[1], specially to avoid violence and organizational rivalry at work .Also when important personal traits of trustworthiness, increased concentration and honor of the organization in the context of profit is required.

Motivation: The defamation of the University by the Employees in all the GradeCategories of – Grade A, B and C are increasing at an alarming rate which has provided and impetus to work upon the statistical and computational aspect of measuring the General Intelligence (GI) of the employees and their Behavioral Integrity(BI)-which is the quality of being honest and hold moral uprightness[2]. For measuring the productivity a test of the employees is conducted. This new model attempts to final the Cluster Analysis of the employee- BI Score and their GI Score.

Cluster Analysis consists of different methods for finding structure within complex bodies of data[3].In a typical example one has a sample of data units (subjects, persons, cases) each described by scores on selected variables (attributes, characteristic or measurements) in order to logically group them together and identify the natural association amongst them while the clusters are independent and are relatively distinct from each other.

II. RELATED WORK

The research work related to the study is presented in the chronological order:

Fisherfirst conducted a study of 150 Iris flowers of 3 types and implemented the use of multiple measurements in taxonomic problems[4].In this work Fisher initiated the concept of analyzing Clusters in flowers and grouped them using the petal length, sepal length, petal width and sepal width of iris flower. In a similar study Ward performed the hierarchical cluster grouping and used it to optimize an objective function [5].Edwards invented a method for investigating the relationships of points in the multidimensional space. By using this analysis of variance technique the points are divided into the two most compact clusters and the process is repeated sequentially so that a "tree" diagram is generated[6].In the area of Multivariate Research , MacQueen proposed some methods which are used for Classification and Analysis of Multivariate Observations[7].

Englemann and Hartigan proposed a method of Percentage Points of a test used for different clusters [8].Pattern Clustering study was conducted by Wolfe by using the multivariate mixture analysis [9]. The concept of Likelihood Ratio Criteria in clustering methods was proposed by Scott and Symons[10].The work related to probability theory of cluster analysis was presented by Ling[11].Gordon and Henderson proposed an algorithm for finding Euclidean sum of squares Classifications[12].Mezzich and Solomon presented the quantitative approach in behavior science and performed clustering procedures and the design of an empirical evaluation of quantitative taxonomic methods. They conducted the cluster analysis study of Iris specimens, treatment environments, archetypal psychiatric patients, ethnic populations and evaluated quantitative taxonomic methods. Mezzich studied comparative performance of quantitative taxonomic methods across databases [13].

III. PROBLEM DEFINITION

In the present work we study the Comparative Performance of individual employee Scores of their : (i) General Intelligence(GI) and (ii) Behavioral Integrity(BI) by using the Synthetic Data of ten Grade-A employees working in a University.

IV. PROPOSED WORK

In the proposed work Cluster analysis is performed with two objectives to investigate: (1) The Interconnectedness and Closeness amongst the employee Clusters.(2) The Cophenetic Correlation coefficient used to determine whether General Intelligence(GI) of the employee is related to the Behavioral Integrity(BI).

Objectives of the proposed work are as follows:

1.To find the correlation between BI and GI.

2.To visualize theDendrogram.

3.To find the Interconnectedness of clusters.

4.To find the Closeness amongst the clusters.

V.METHODOLOGY

Data used: Synthetic data is used in the study which holds no personal information to protect the privacy and confidentiality of a set of data pertaining to Grade-A employees so that it cannot be traced back to any individual employee. In the present work we have used synthetic data instead of realistic data related to 10 employees.

Cluster analysis is used perform the natural divisions in data. The hypothesis is stated as follows:

H_0 : Behavioral Integrity(BI) is not related to General Intelligence(GI).

H_A: BI is positively related to GI.

The cophenetic correlation coefficient is used to test the hypothesis. This score is based on the GI- test which is used to measure the Productivity of an employee based on a Rule – Base.

The following Rule-Base is used to evaluate the GI measurement.

Table 1: Rule – Base for finding GI measure

S.No.	GI Score	Grade	Numeric Value
1	>=85(Outstanding)	A	5
2	84-70(Very Good)	B	4
3	69-60(Good)	C	3
4	61-40(Average)	D	2
5	<40(Below Average)	E	1

Table 2: Total score based GI secured by an individual

Employee	Grade -5	Grade -4	Grade -3	Grade -2	Grade -1	Score
E-1	1	-	-	-	-	5
E-2	-	1	-	-	-	4
E-3	-	-	1	-	-	3
E-4	-	-	-	-	1	1
E-5	1	-	-	-	-	5
E-6	-	-	-	1	-	2
E-7	-	-	-	1	-	2
E-8	-	-	-	1	-	2
E-9	-	-	-	-	1	1
E-10	-	-	-	1	-	2

The BI score is obtained by using the Precept-score which is compared using the Precept: Rule-base. The precept score is based on the rule-base by using the employee's natural precept attributes based on his / her Behavioral habits.

Table 3: Rule-Base for BI

S.No.	Precept	Absence=0 and Presence = 1
1	P1:Abstinence from killing	0/1
2	P2: Abstinence from taking what is not given	0/1
3	P3: Abstinence from sexual misconduct	0/1
4	P4: Abstinence from false speech	0/1
5	P5: Abstinence from intoxicating drinks and drugs	0/1

By using the precept measurements a Behavioral – integrity (BI) Score Matrix is generated.

Table 4 : BI- Score Matrix

Emp	P1	P2	P3	P4	P5	Total Score
E1	1	1	1	1	1	5
E2	1	1	0	1	1	4
E3	0	0	1	1	0	2
E4	1	0	0	0	0	1
E5	1	1	1	1	1	5
E6	1	1	1	0	0	3
E7	0	1	1	0	0	2
E8	0	0	0	0	1	1
E9	1	0	0	0	0	1
E10	1	1	1	0	0	3

Matlab®Software is used to define the two dimensional matrix. The computational processing is performed using the following steps:

Step 1 : A matrix called X is generated with 10 – objects treated as 10 – employees. The values of BI and GI are inputted in the 10-objects.The matrix is defined as follows:

X= [5 5; 4 4; 2 3; 1 1; 5 5; 3 2; 2 2; 1 2; 1 1; 32]

Step 2 : The pair-wise distance is computed for individual employee.

Y = pdist(X)

Step 3: The squareform is generated by using the function:

squareform(Y)

Step 4 : The linkage between the clusters is computed.

 Z = linkage (Y)

Step 5 : We then find the cophenetic correlation coefficient which is used to verify the dissimilarity amongst the two variables Z and Y using the function cophenet.

c= cophene (Z, Y).We obtain the value of c = 0.9024

Step 6: Compute :

1) Pairwise distance between the object based on the 'cityblock' metric and
2) Create agglomerative hierarchical cluster tree by using un-weighted average distance (UPGMA) – by using 'average' method of cluster computation.

Y = pdist (X, 'cityblock');

Z = linkage (Y, 'average');

c= cophenet (Z, Y)

We obtain c = 0.9131 as the result .

Step7: Verification of consistency co-efficient is done.

We compute the in-consistency co-efficient.

I = inconsistent (Z)

Table 5: Values of inconsistency co–efficient

Mean	Standard Deviation	No. Of Links	Inconsistency Co-efficient
0	0	1	0
0	0	1	0

0	0	1	0
0.5000	0.7071	2	0.7071
0.5000	0.7071	2	0.7071
0.5000	0.7071	2	0.7071
1.3333	0.4714	2	0.7071
1.6944	0.7087	3	1.0190
3.1230	2.5506	3	1.1093

Step 8: - To find the Natural Divisions in Data.
 If we lower the inconsistency co-efficient threshold to 0.8 and we use the function cluster which is used to find the natural divisions in data
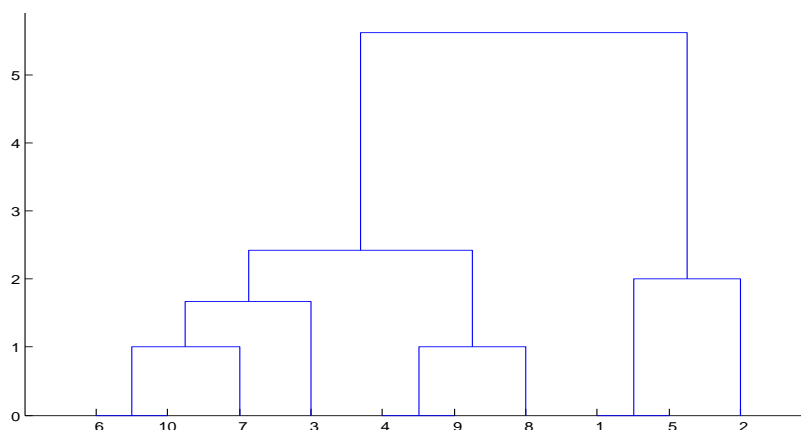T= cluster (Z ,'cutoff' , 0.8) .

Fig.1 Dendrogram generated using the average method

VI. CONCLUSION

1.The output indicates that the employees E-3, E-6, E-7 and E-10 are placed in Cluster-1;E-1, E-2 and E-5 are placed in Cluster -2 and E-4, E-8 & E-9 are placed in Cluster-3.However, arbitrary clusters many also be specified using the cluster Function.
2. The value of cophenetic correlation coefficient indicates that the two variables BI and GI are positively correlated.
3.Thecophenetic correlation coefficient shows that by using different distance metrics and different clustering methods a better cluster tree can be created using the agglomerative hierarchical linkage.
4.Thedendrogram exhibits the natural data divisions and represents the maximum number of clusters with linkage interconnections.

VII.FURTHER SCOPE

The same application may be extended to form a Data Cube and different cluster evaluation criteria may be correlated with the number of Clusters.

REFERENCES

[1]. ASAP World Consultancy and Jonathan Blain, Special Edition using SAP R/3(Second Edition, PHI Publication, New Delhi,1998).
[2]. Stuart B. Flexner ,The Random House College Dictionary,1988.
[3]. Anderberg M.R. ,Cluster Analysis for Applications(New York: Academic Press,1973).
[4]. Fisher, R. A.,"The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics,7, 1936 ,pp.179-188.
[5]. Ward, J.H.," Hierarchical grouping to optimize an objective function'', Journal of the American Statistical Association, 58, 1963 236-244.
[6]. Edwards, A.W.F. and Cavalli-Sforza, L.L., "A method for cluster analysis," Biometrics, 21, 1965, 362-375.
[7]. MacQueen, J.B., "Some Methods for Classification and Analysis of Multivariate Observations," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1,1967, 281-297.
[8]. Englemann, L. and Hartigan, J.A., "Percentage Points of a Test for Clusters," Journal of the American Statistical Association", 64, 1969,1647-1648.
[9]. Wolfe J.H., "Pattern Clustering by Multivariate Mixture Analysis'', Multivariate Behavioral Research, 5,1970, 329-350.
[10]. Scott, A.J. and Symons, M.J., "Clustering Methods Based on Likelihood Ratio Criteria", Biometrics, 27, 1971,387-397.
[11]. Ling, R.F. (1973),"A Probability Theory of Cluster Analysis," Journal of the American Statistical Association, 68,1973, 159-169.
[12]. Gordon, A.D. and Henderson, J.T., "An Algorithm for Euclidean Sum of Squares Classification," Biometrics, 33, 1977, 355-362.

[13]. Mezzich, J.E. and Solomon, H., Taxonomy and Behavioral Science(New York: Academic Press,1980).

[14]. Duran, B.S. and Odell, P.L. (1974), Cluster Analysis, New York: Spring- Verlag.

[15]. Everitt, B.S. (1980) Cluster Analysis, 2nd ed., London: Henineman Educational Books Ltd.

[16]. Hartigan, J.A. (1975), Clustering Algorithms, New York: John Wiley & Sons.

[17]. Hartigan, J.A. (1978), "Asymptotic Distributions for Clustering Criteria," Annals of Statistics, 6, 117-131

[18]. Hubert, l. (1974), "Approximate Evaluation Techniques for the Single-link and Complete-link Hierarchical Clustering Procedures, " Journal of the American Statistical association, 69, 698-704.

[19]. Hubert, L.J. and Baker, F.B. (1977), "An Empirical Comparison of Baseline Models for Goodness-of-Fit in r- Diameter Hierarchical Clustering," in Classification and Clustering, Ed. J. Van Ryzin, New York: Academic Press.

[20]. Milligan, G.W. and Cooper, M.C. (1983), "An Examination of Procedures for Determining the Number of Clusters in a Data Set," College of Administrative Science Working Paper Series 83-51, Columbus: The Ohio State University.

[21]. Wolfe, J.H. (1978), " Comparative Cluster Analysis of Pattetns of Vacational, " Multivariate Behavioral Research, 13

[22]. Wong M.S. (1982), " Asymptotic properties of Bhivariate k-means Clusters, Communication in Statistic Theory and Methods , 11, 1155-1171

[23]. Wong M. A., and Lane, T (1983) " A k th Nearest Neighbor Clustering Procedure," Journal of the Royal Statistical society, Series B, in Press.