Measurement of The Performance Factor: A Forward Feature Selection Approach

Prafulla Bharat Bafna

*Symbiosis Institute of Computer Studies and Research (SICSR),
Symbiosis International University (SIU), Atur Centre, Gokhale Cross Road, Model
Colony, Pune –411 016, Maharashtra, India
prafulla.bafna@sicsr.ac.in*

Abstract

In recent years Human Resource Management (HRM) innovations like performance appraisal reform have become particularly a sensitive issue. It's a technique to ensure that employees are in tune with developments in the business. It involves information gathering and decision making. There are several parameters which are considered in appraisal process. These parameters are domain specific. It means that performance factors applicable in commercial domain may not be suitable in academic domain.

Clustering based technique is proposed which extracts major performance factors amongst several by forward feature selection and experts' knowledge. It reduces time, efforts and resources.

Outcomes of cluster quality is used as decision support for the performance factors to rank the employees and clustering on complete data set for effective decision support. The data set used is related to top 100 employees with respect to their appraisal count.

Keywords: performance factor, scale, optimal, entropy, purity, SMSE, clustering

Introduction

With advances in technology, there is exponential growth in the industry sector. Appraisal is the technique to evaluate job performance of the employee.

Objective of appraisal is not only to evaluate the employee but to provide feedback, improving communication, understanding training needs, clarifying roles and responsibilities and determining how to allocate rewards. [8].To get optimal factors to evaluate performance of an employee of an academic sector, a clustering based approach is proposed in which cluster accuracy is used as an indicator to select the set of features. K-means and hierarchical algorithm are selected to confirm

optimal factors. The selection criteria are entropy, purity and Sum of Mean Square Error (SMSE).

Considering points in performance appraisal forms , top 100 employees analyzed out with respect to their appraisal count[1] . Appraisal count is nothing but the sum of all the measures assigned to each question present in employee appraisal form. Performance appraisal contains a set of questions and a measure is allocated for each of the question .So .Question weight is calculated according to the performance of the employee with respect to that particular question. For each employee 35 parameters are considered and distance matrix is formed.

Employee represents rows and scale assigned to each performance factor e represents a column Hierarchical and K-means algorithms are applied for different combination of features by considering expert's advice and forward feature selection method [7]. The selection criteria are based upon minimum entropy, maximum purity, minimum SMSE is selected. Finally13 performance factors were selected from 35. These principles can be considered as mostly focused to evaluate employees.

Background

Organizations are run and steered by people. The performance of an organization is thus dependent upon the sum total of the performance of its members. The performance of an employee is his resultant behavior on task which can be observed and evaluated. Performance can be measured by some combination of quantity. Performance Appraisal system Provides management an opportunity to recall as well as feedback to people as to how they are doing so that they can correct their mistake and acquire new skill[9].Performance appraisal and evaluation has repeatedly been a central and sensitive area of these changes, and this has been so mainly because new accountabilities [8]Generally collected data contains irrelevant or redundant attributes. Classification and clustering do not give accurate result if there are interdependent attributes. Correct feature selection is a fundamental data preprocessing step in data mining. Feature Mine algorithm contains sequence mining and classification algorithms which efficiently handles very large data sets with thousands of items and millions of records.[4] Edie Rasmussen states Cluster analysis is a technique which assigns items to groups based on a calculation of the degree of association between items and groups. Cluster analysis can be used for hierarchical algorithm. Nested data set is produced in which pairs of items or clusters are connected successively. However, the hierarchical methods are better information retrieval. The commonly used hierarchical methods, such as single link, complete link, group average link, and Ward's method, have high space and time requirements. In order to cluster the large data sets with high dimensionality there is need to have a better algorithm Examples are the minimal spanning tree algorithms for the single link method, the Voorhees algorithm for group average link, and the reciprocal nearest neighbor algorithm for Ward's method. Edie listed steps of clustering including Selecting of the attributes on which items are to be clustered , selecting appropriate clustering method, Creating the clusters or cluster hierarchies ,interpreting clusters and validating the results etc.[5]They have focused on feature selection algorithms for

classification(knowing class label) and clustering (unsupervised feature selection) where data is unlabeled. Feature selection algorithms designed with different evaluation criteria broadly fall into three categories: the filter model , the wrapper model and the hybrid model . The filter model relies on [6]

Data Collection

Data is. Collected from 1oo employees having topmost appraisal count [1] and 35 performance factors are considered for each employee. Each question has a measure associated with it K-means [5] plot was obtained for above matrix. Publications represent X-axis and conducting sessions represents Y axis. Hierarchical algorithm [5] was applied on entire data and dendrogram was obtained.
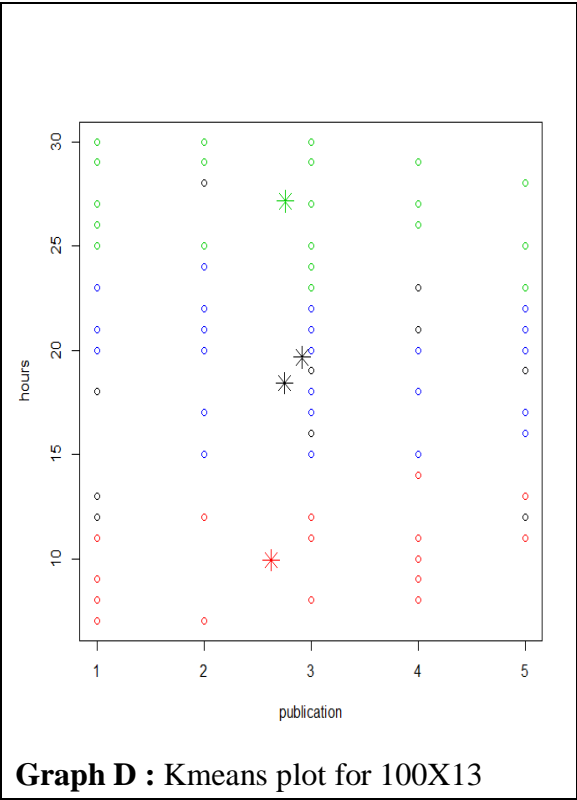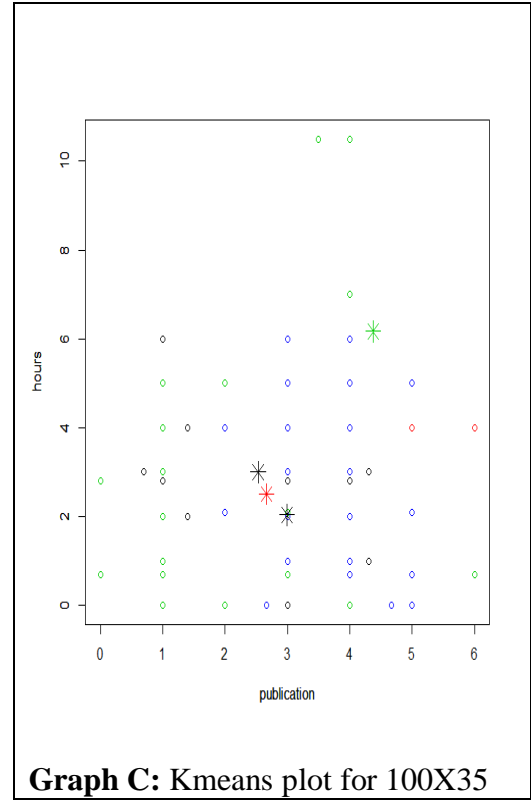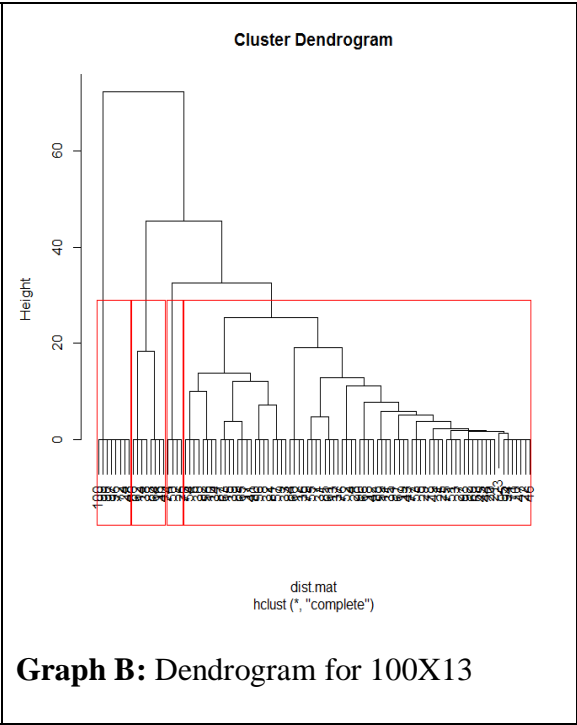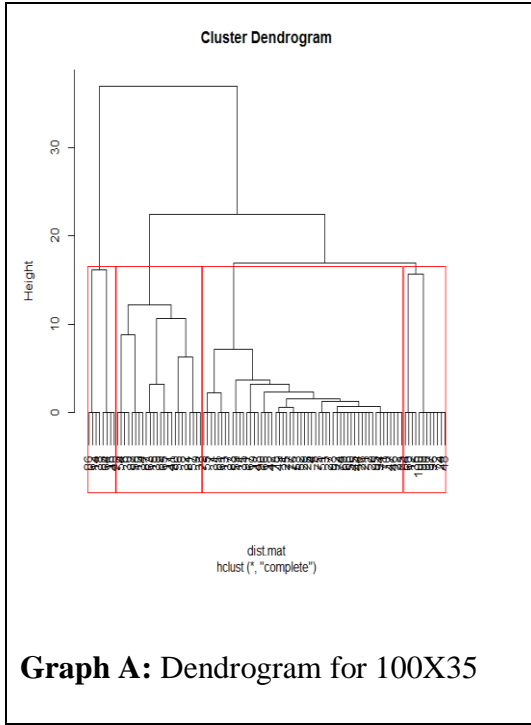
Selection of Optimal factors

k-means and hierarchical algorithms are applied on domain specific enhanced factors by forward Feature selection. approach.[7]. Graph A,B represents dendrogram produced by hierarchical algorithm using Euclidean distance for 100 employees with 35 and 13 factors respectively Graph C,D represents k- means plot for 100 employees with 35 and 13 performance factors respectively.

We can easily detect optimal factors from Table A , having minimum SMSE and entropy also with maximum purity. From 35 standard website principles, principles selected are publications, Faculty Development Program, Corporate, lectures, collaboration, syllabus, feedback, innovation, evaluation, practices, strategies, pair-share, hours

Table A: Optimal factors

No of factors	SMSE (sum mean square error)	entropy	purity
10	9.65	.32	.64
13	**3.25**	**.092**	**.89**
19	13.2	.29	.79
25	18	.36	.60
28	20	.16	.71
35	>22	.46	.63

Graph A: Dendrogram for 100X35

Graph B: Dendrogram for 100X13

Graph C: Kmeans plot for 100X35

Graph D : Kmeans plot for 100X13

Conclusion

The overall grade of the employee according to the appraisal is actually sum of many quality attributes. Forward feature selection method allows extracting crucial performance factors. Focusing these performance factors in an academic domain saves time, cost and efforts spent on decision making. Above approach of measuring performance factor using forward feature selection is productive with respect to maintain the large data.

References

[1]   Prafulla Bafna,Shailaja Shirwaikar ,Human Computer Interaction-paradigms, process, practices, Advances in computer vision and information technology, part 20,pg 989-998,2009.

[2]   Sanjeet Singh, Gagan Deep Sharma, Harman Deep Chima A Study of Effect of Performance Appraithe Organization and the Employe**e**

[3]   Lesh, N. MERL, Zaki, M. J., Scalable feature mining for sequential data, Intelligent Systems and their Applications, IEEE (Volume:15, Issue: 2), 2000

[4]   CHAPTER 16: CLUSTERING ALGORITHMS, Edie Rasmussen, University of Pittsburgh

[5]   Huan Liu and Lei Yu, Toward Integrating Feature Selection Algorithms for Classification and Clustering, IEEE Transactions on Knowledge and Data Engineering, v.5 n.6, p.914-925,2004

[6]   Prafulla Bafna[1], Pravin Metkewar[2*] and Shailaja Shirwaikar[3#],"Novel Clustering approach for Feature Selection , American International Journal of Research in Science, Technology, Engineering & Mathematics" (ISSN (Print): 2328-3491, ISSN (Online): 2328-3580, ISSN (CD-ROM): 2328-362) September-November 2014 issue 8 volume

[7]   Prafulla Bafna et al ,COMPARATIVE ANALYSIS OF APRIORI AND IMPROVED APRIORI ALGORITHM, International Journal of Emerging Technologies in Computational ,and Applied Sciences (IJETCAS), Issue 7 Volume 2 , 135-143,2014

[8]   Alan Jenkins ,Performance Appraisal Research: A Critical Review of Work on "The Social Context and Politics of Appraisal",ESSEC,March 2005.

[9]   Dr Kanchan Bhatia, Prashant Jain A study of performance appraisal and organizational effectiveness in terms of individual and organiational basis. A comparative study of BSNL and AIRTEL, International Journal on Arts, Management and Humanities,pp-63-72,2012

[10]   Prafulla Bafna1, Hema Gaikwad, A hybrid approach to measure design improvement factor of website., American International Journal of Research in Science, Technology, Engineering & Mathematics" (ISSN (Print): 2328-3491, ISSN (Online): 2328-3580, ISSN (CD-ROM): 2328-362) December-2014 to February-2015.

[11] A novel clustering approach to select optimal usability principles for educational websites, Prafulla Bafna, International Journal of Software and Web Sciences , ISSN (Print): 2279-0063, ISSN (Online): 2279-0071, Issue 11, December-2014 to February-2015 Sciences(IJSWS).