ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES
Knowledge to Innovation

Online - 2455-3891
Print - 0974-2441

Special Issue (April)

Full Proceeding Paper

EMPLOYEE PERFORMANCE APPRAISAL SYSTEM BASED ON RANKING AND REVIEWS

ASHISH MODI, SHARATH KUMAR J, MURALIDHAR A

Department of , School of Computing Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India. Email: vijayalakshmi.av@vit.ac.in

Received: 23 January 2017, Revised and Accepted: 03 March 2017

ABSTRACT

Objective: In many organizations, employee data have to be maintained and utilized for many purposes. Here, in this paper, we are going to use such data to calculate an employee's performance.

Methods: This employee data may be converted into useful information using data mining techniques such as K-means and decisions tree. K-means is used to find the rank of the employee means that the employee may come under in his criteria. Decision tree is used to find the review of an employee means that the employee needs improvement or he/she meets expectation.

Results: This algorithm when utilized can identify the top employee who can be considered for appraisal or the eligible candidates for promotion. Hence, these algorithms such as K-mean and decision tree that help to find best employees for any association and help us to take a good decision in less time.

Conclusion: There are various factors which should be considered and are limited to this algorithm, so human intervention is required to consider those factors. However, ranking and appraisal are seen in many companies, and this algorithm will definitely identify the potential candidates.

Keywords: Employee performance, Clustering, Decision tree, K-means, Employee performance, Data mining, Euclidean distance.

 DOI: http://dx.doi.org/10.22159/ajpcr.2017.v10s1.1069

INTRODUCTION

Industrial sector has a collection of employee data. This data may be converted into useful information for taking a good decision. In every industrial company, there are number of employee data such as position, salary, and attendance. These data may be quality wise analyzed using some mining technique. The performance of employee shows a very important role to get a good job in any company. Performance plays a role on the basis of working experience of the employee. The very important element that is used to evaluate an employee performance in a company is his position and his experience. The other elements such as salary, projects, knowledge, and skills also play an important role to employment opportunity to an employee. Employee should not evaluate only on his experience and position that he achieved. Hence, a grouping of employees is based on the performance of the employee. The research has been based on k-mean clustering and decision tree algorithm for categorizing employees in different clusters. Here, k-mean is used to find the cluster of the employee, and this may help us to find the rank of the employee means either it comes under criteria or not. Then, decision tree is used here by the purpose of finding the review of the employee. Decision tree helps us to find the result of the predefined dataset means, and it gives us to feedback or review of the employee.

LITERATURE SURVEY

The paper describes about Silhouette method in K-mean clustering. Silhouette refers to a method of clarification and justification within clusters of data [1,2]. This methodology provides a graphical representation, how all values lie within the cluster. K-mean algorithm splits the number of observation into K clusters [3], where each observation can belong to the cluster with the nearest mean [4]. To use K, silhouette measures are used. The cluster can be ranked or classified to get an enhanced student performance [5,6]. This algorithm may help to get a student performance result in a ranking manner [7].

The paper deals with decision tree algorithm [8]. Algorithm classifies map data into predefined group of class means data that may be classified according to their gain entropy [8,9]. Then, higher weighted class goes on the root and then just lower than higher class further classified into another class [10,11]. When the classification comes to the last class, then it divided into single value that is the feedback of employee [12]. The result of the employee is based on employee's performance.

PROPOSED WORK

Clustering technique (K-means)

Cluster research or clustering is the way of group a set of entities in which those entities in similar group (cluster) are more same (similar) to each other than other clusters (groups). Euclidean distance is also used to find the clusters of the K-means algorithm. To find Euclidean distance between observations, first take initial cluster centroid, cluster 1 and cluster 2 make them to centroid and calculate. Based on the calculation (Euclidean distance), every observation assigns to one cluster - calculating minimum distance.

$$\text{Euclidean distance} = \sqrt{(X_H - H_1)^2 + (X_W - W_1)^2}$$

Where
X_H: Observation value of variable height
H_1: Centroid value of cluster 1 for variable height
X_W: Observation value of variable weight
W_1: Centroid value of cluster 1 for variable weight

Clustering technique (decision tree)

Decision tree is a way that is used as a hierarchy such as a map of choice and available issues, along with their results, ability amount, and convenience. This is the root, which we are using to show this approach.

To create a decision tree, used these following rules:

- Choose one instance of practice example divide it into available instances so that the practice example is scattered in some small-small sets.

- After this method is applied for all instance of practice example. The knob of entire practice example refers to equal hierarchy, and no left instance will be done for another dividation then break diving the hierarchy.

Decision tree is further classified into two parts or nodes as follows:

- Leaf node - it shows the appraisal of the instance.
- Decision node - it defines many analysis for single instance appraisal distinct branch with one decision tree for every available outcome of the analysis.

For develop tree, we want to analysis the info of every instance. Choose an instance with very high value.

Consider two classes X and Y

- Take the sample Z which carry X values of class X and Y values of class Y.
- For finding that, Z is depend to X and Y is consider as follows:

$$I(X,Y)=\frac{X}{X+Y}*\log_2\left(\frac{X}{X+Y}\right)-\left(\frac{Y}{X+Y}\right)*\log_2\left(\frac{Y}{X+Y}\right)$$

Let that with the help of instance V, Z can be divided into $\{Z_1,Z_2,Z_3,Z_4.....Z_n\}$

Suppose, Z_i holds X_i of X and Y_i of Y, the conventional info wants to divide the attribute in the hierarchy is as follows:

$$E(V)=\sum\left(\frac{(X_i+Y_i)}{X+Y}\right)*I(X_i,Y_i)$$

i = 1

Entropy

Conventional info wants to initialize a random class attribute in Z in the minimal, least-path code.

Measure info gained, that is, gain (V). Calculate the dividation of entropy gained by the hierarchy. Select the higher dividation hierarchy.

Gain(V) = I(X,Y)-E(V)

To find the distance in k-means algorithm, we use Euclidean distance. Euclidean distance between observation and cluster centroid 1 and 2 is calculated. After calculating Euclidean distance, each observation is assigned to 1 of the cluster based on its minimum distance.

PERFORMANCE ANALYSIS

K-means (Euclidean distance)

$$\sqrt{(35000-35000)^2+(26-26)^2+(5-5)^2}=0$$
$$\sqrt{(35000-47000)^2+(26-25)^2+(5-3)^2}=12000.00021$$
$$\sqrt{(35000-52000)^2+(26-28)^2+(5-3)^2}=17000.00024$$
$$\sqrt{(35000-28000)^2+(26-32)^2+(5-8)^2}=7000.003214$$
$$\sqrt{(35000-50000)^2+(26-23)^2+(5-2)^2}=15000.0006$$
$$\sqrt{(47000-35000)^2+(25-26)^2+(3-5)^2}=12000.00021$$
$$\sqrt{(47000-47000)^2+(25-25)^2+(3-3)^2}=0$$
$$\sqrt{(47000-52000)^2+(25-28)^2+(3-3)^2}=5000.0009$$
$$\sqrt{(47000-28000)^2+(25-32)^2+(3-8)^2}=19000.00195$$
$$\sqrt{(47000-50000)^2+(25-23)^2+(3-2)^2}=3000.000833$$

Apply the formula of Euclidean distance in Table 1 to find the value of the each data set. Then, we get values for dataset.

In Table 2, after getting the values of dataset, then assign each row to a cluster according to their values. Hence, in previous to find cluster through K-means using Euclidean distance technique, we first considered two centroids then calculate the cluster value of first two rows. Then, after this, we update the cluster again, and again, by taking the mean of the centroid. However, in the modified algorithm, there is no need to update centroid again, and again, fix the value of centroid then calculate the values of the cluster. It remains same, but it decreases the calculation time. If column one gets the highest values, then assigns that cluster value of that row is 1 otherwise gives it to cluster 2.

Table 3 shows the assignment of the cluster with attributes that show the rank of employees.

for i = 1...n

for i = 1...n

$s = s+([a_{ij}]-[a_{1j}])^2$

$t = t+([a_{ij}]-[a_{2j}])^2$

$p_{i1}=\sqrt{s}$

$p_{i2}=\sqrt{s}$

If $p_{i1}>p_{i2}$ then assignment = 1

else assignment = 2

Decision tree

Table 4 contains the employee's dataset for decision tree. The result of "proper" is 7 and result of "improper" is 3, there are two different values. Hence, these two values, Class X = 8 and Y = 2, calculate the expectation info and info entropy.

$$I(X,Y)=\frac{-7}{10}*\log_2\left(\frac{7}{10}\right)-\left(\frac{3}{10}\right)*\log_2\left(\frac{3}{10}\right)=0.8812 \text{ bits}$$

Table 1: Dataset collected

Id	Salary	Age	Experience
1	35000	26	5
2	47000	25	3
3	52000	28	3
4	28000	32	8
5	50000	23	2

Table 2: Clustered data

Id	Cluster 1	Cluster 2	Assignment
1	0	12000.00021	2
2	12000.00021	0	1
3	17000.00024	5000.0009	1
4	7000.003214	19000.00195	2
5	15000.0006	3000.000833	1

Table 3: Assign cluster to dataset

Id	Salary	Age	Experience	Assignment
1	35000	26	5	2
2	47000	25	3	1
3	52000	28	3	1
4	28000	32	8	2
5	50000	23	2	1

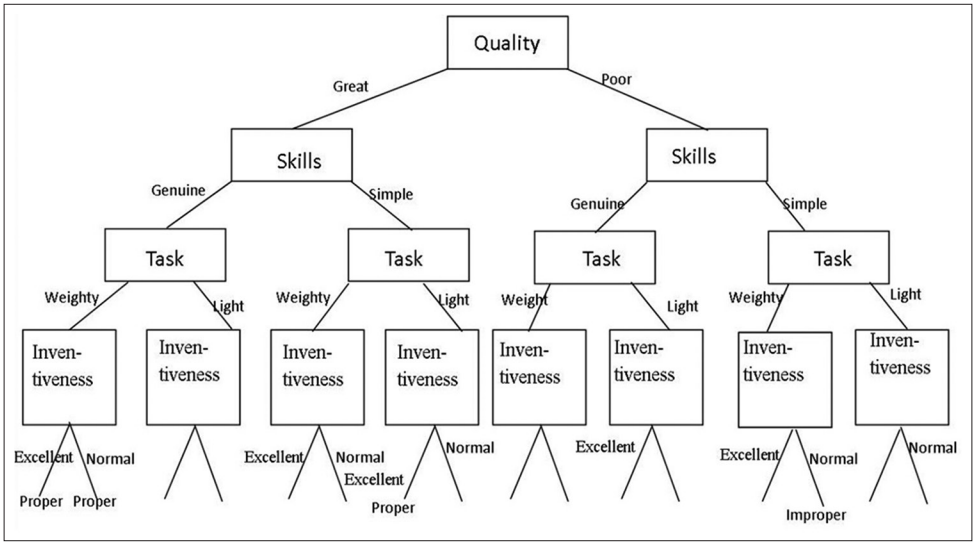

Fig. 1: Decision tree

Table 4: Dataset for decision tree

Work task	Skills	Inventiveness	Quality	Result
Weighty	Genuine	Excellent	Great	Proper
Weighty	Simple	Excellent	Great	Proper
Weighty	Genuine	Normal	Great	Proper
Weighty	Genuine	Normal	Great	Proper
Weighty	Genuine	Normal	Poor	Improper
Light	Genuine	Excellent	Poor	Improper
Light	Genuine	Excellent	Poor	Proper
Light	Simple	Excellent	Great	Proper
Light	Genuine	Excellent	Poor	Proper
Light	Simple	Normal	Poor	Improper

$$E(\text{work task})=\frac{5}{10}*I(4,1)+\frac{5}{10}*I(3,2)=0.8464 \text{ bits}$$

$$E(\text{quality})=\frac{5}{10}*I(2,3)=0.4885 \text{ bits}$$

$$E(\text{skills})=\frac{7}{10}*I(5,2)+\frac{3}{10}*I(2,1)=0.7810 \text{bits}$$

$$E(\text{inventiveness})=\frac{6}{10}*I(5,1)+\frac{4}{10}*I(2,2)=0.87074 \text{ bits}$$

Hence, calculate info gain:

Gain (work task) = 0.8812-0.8464 = 0.0348 bits

Gain (quality) = 0.8812-0.7812 = 0.1 bits

Gain (initiative) = 0.8812-0.87074 = 0.0104 bits

The maximum gain information is quality so select quality as root.

Then, split the tree into just minimum gain of quality and then do where every node is node split. It shows in Fig. 1.

Decision tree is also creating by if else command in any programing language.

IF (quality = "great" and skills = "genuine" and task = "weighty" and inventiveness = "excellent"), then result = "proper"

ELSE IF (quality = "great" and skills = "genuine" and task = "light" and inventiveness = "normal"), then result = "proper"

ELSE IF (quality = "great" and skills = "simple" and task = "weighty" and inventiveness = "normal"), then result = "proper"

ELSE IF (quality = "poor" and skills = "genuine" and task = "light" and inventiveness = "excellent"),

then result = "improper"

ELSE IF (quality = "poor" and skills = "simple" and task = "weighty" and inventiveness = "normal"), then result = "improper"

CONCLUSION

We have presented in this work, a new approach for the K-mean and decision tree, that helps to find best employees for any association. A present performance analysis is required to recommendation for excellence salary changes and grade or in-grade change salary increase. This may also help the manager to check the employee performance and those employees want extra attention for decreasing falling ratio for taking the strict action right time. Decision tree method is used on the previous year data of employee performance. This data mining approach helps the institutes, companies, and anywhere, where the need of employee and this ranking and review technique will help to the manager to find best employees in minimal time when he has so many number of employee data or big dataset. Hence, here, we describe about raking and reviews that help us to take a good decision in less time.

REFERENCES

1. Kumar SA, Vijayalakshmi MN. Mining of student academic evaluation records in higher education. In: Recent Advances in Computing and Software Systems (RACSS), 2012 International Conference on IEEE; 2012. p. 67-70.
2. Geng X, Luo L. Multilabel ranking with inconsistent rankers. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2014. p. 3742-7.
3. Bouhmala N. How good is the Euclidean distance metric for the clustering problem. In: Advanced Applied Informatics (IIAI-AAI), 2016 5th IIAI International Congress on IEEE; 2016. p. 312-5.
4. Esteves RM, Hacker T, Rong C. Competitive k-means, a new accurate and distributed k-means algorithm for large datasets. In: Cloud Computing Technology and Science (Cloud Com), 2013 IEEE 5th International Conference on IEEE. Vol. 1; 2013. p. 17-24.
5. Kumar KM, Reddy AR. A fast K-means clustering using prototypes for initial cluster center selection. In: Intelligent Systems and Control (ISCO), 2015 IEEE 9th International Conference on IEEE; 2015. p. 1-4.
6. Poteraş CM, Mocanu ML. Evaluation of an optimized K-means algorithm based on real data. In: Computer Science and Information Systems (Fed CSIS), 2016 Federated Conference on IEEE; 2016. p. 831-5.

7. Kotalwar R, Gandhi S, Chavan R. Data mining: Evaluating performance of employee's using classification algorithm based on decision tree. Eng Sci Technol Int J 2014;4:29-35.
8. Yang Y, Chen W. Taiga: Performance optimization of the C4. 5 decision tree construction algorithm. Tsinghua Sci Technol 2016;21(4):415-25.
9. Guleria P, Thakur N, Sood M. Predicting student performance using decision tree classifiers and information gain. In: Parallel, Distributed and Grid Computing (PDGC), 2014 International Conference on IEEE; 2014. p. 126-9.
10. Vaidya J, Shafiq B, Fan W, Mehmood D, Lorenzi D. A random decision tree framework for privacy-preserving data mining. IEEE Trans Dependable Secure Comput 2014;11(5):399-411.
11. Lin C, Du X, Jiang X, Wang D. An efficient and effective performance estimation method for DSE. In: VLSI Design, Automation and Test (VLSI-DAT), 2016 International Symposium on IEEE; 2016. pp. 1-4.
12. Chen Q, Gong Z. Data mining modelling of employee engagement for it enterprises based on decision tree algorithm. In: Information Management, Innovation Management and Industrial Engineering (ICIII), 2013 6th International Conference on IEEE. Vol. 2; 2013. p. 305-8.