Computing and Informatics, Vol. 42, 2023, 943–964, doi: 10.31577/cai_2023_4_943

DYNAMIC MATCHING ALGORITHM OF HUMAN RESOURCE ALLOCATION BASED ON BIG DATA MINING

Yuping Yan

Guangdong Power Grid Co., Ltd.
Guangzhou 510000, China

Peiyao Xu*, Jianyong Wang

Guangdong Electric Power Information Technology Co., Ltd.
Guangzhou 510030, China
e-mail: xpy784574@163.com

Abstract. In order to ensure the dynamic matching effect of human resources allocation and improve the accuracy and efficiency of dynamic matching of human resources allocation, a dynamic matching algorithm of human resources allocation based on big data mining is studied. Analyze the meaning and function of big data mining, and explain the common analysis principles of big data mining. The information entropy is selected as the basis for measuring human resource allocation, the human resource allocation is extracted, and the similarity of human resource allocation is calculated using the Huasdorff similarity method based on time interpolation. According to the Apriori algorithm and FP-Growth classification algorithm, the human resource allocation is classified and mined, and the K-Means clustering algorithm is used to realize the dynamic matching of human resource allocation. The experimental results show that the proposed algorithm has better dynamic matching effect of human resources allocation, and can effectively improve the accuracy and efficiency of dynamic matching of human resources allocation.

Keywords: Big data mining, apriori algorithm, FP-growth classification algorithm, human resource allocation, K-means clustering algorithm, Huasdorff similarity method, dynamic matching

* Corresponding author

1 INTRODUCTION

With the integration of the world economy and China's entry into the WTO, the competition for talents will become more intense. Human resource is the most valuable and important resource among various resources of an enterprise, and it is the "first resource" for enterprise development. The combination and application of other resources in an enterprise must be promoted by human resources [1]. Across successive periods of the Industrial Revolution, organized labor, theory of management, cognitive economics, or social interactions, the idea of HRM has developed. As a result, the phrase Personnel Management has increasingly given way to the idea of HRM. However, the accumulation of human resources alone is not enough for an enterprise, and human resources must be allocated effectively and reasonably in order to maximize its benefits. The goals of human resource administration are to find qualified candidates, engage people, educate individuals, but then assist people in enhancing their productivity to enable them to accomplish great things and contribute to the company objectives. The allocation of enterprise human resources is to arrange all kinds of talents who meet the needs of enterprise development in the required positions in a timely and reasonable manner through assessment, selection, employment and training [2, 3]. To effectively utilise ability, increase the capabilities of employees, integrate it with other financial resources, and maximise the production of additional social and economic advantages for the business. Allocating human resources serves as both the beginning and the culmination of HR management. The types of HR allocation are workforce professional, Human Resource department trainee, HR professional, Human Resources supervisor. Its ultimate goal is to match individuals and positions, realize dynamic matching of human resource allocation, and improve the overall efficiency of the organization. The important element that decides whether the organization can develop consistently, steadily, and quickly is the amount of human resource allocation advantage, which has a direct impact on the prudent utilization of other assets or the entire allocating profit of the company. A continual flow of data and activity is HRM. The HRM element would suffer greatly from inaction. Keeping continuously mindful of how workers are performing, whether successfully there are doing everything, or how people think towards doing their duties is thus a crucial component of HRM. Therefore, the human resources management department must do a good job in the research of dynamic matching of human resources allocation according to the actual situation of the unit and the needs of the work tasks.

At present, scholars in related fields have carried out research on the dynamic matching of human resource allocation. [4] proposed a dynamic modeling algorithm for human resource allocation in construction projects. The distribution of human resources throughout execution of the project, as an efficient managed service or development element, can significantly affect program project in terms of quality and timeline. As a result, the assignment method can be made better by calculating labour requirements and comparing various human resource allocation plans.

Labor distribution is laborious due to the dynamic, complex relationships and feedback that exist within the project. A dynamic model for efficient labor allocation using a system dynamics approach is proposed. Using the model proposed in the current study enables the project to accurately estimate labor requirements and their efficient allocation, allowing for the necessary planning for the timely supply and distribution of project labor, both before and throughout project implementation. [5] proposed a business process human resource allocation algorithm based on team fault lines. This paper introduces the team fault line into the problem of human resource allocation. The resource characteristics are first analyzed from the demographic perspective and business process, and then the key characteristics are selected and the corresponding weights are determined using the information value. Secondly, qualitatively identify the team fault line according to the human resources clustering results, and quantitatively measure the strength and distance of the team fault line. Utilizing layered perspectives, basic or collective predictive algorithms are created. Following that, the assignment theory and procedure were created. The rationality and effectiveness of the method are evaluated through a real scenario, which can effectively allocate human resources and optimize business processes. Nevertheless, the aforementioned methods still struggle with a poor matched impact, low accuracy, or lack of efficiency.

Aiming at the above problems, a dynamic matching algorithm for human resource allocation based on big data mining is studied. The information entropy is used to extract the human resource configuration, and the Huasdorff similarity method based on time interpolation is used to calculate the similarity of human resource configuration. According to Apriori algorithm and FP-Growth classification algorithm, human resource allocation is classified and mined, and K-Means clustering algorithm is used to realize dynamic matching of human resource allocation. The dynamic matching effect of the algorithm is good, and it can dynamically match the accuracy and efficiency.

2 BIG DATA MINING TECHNOLOGY

2.1 Significance and Role of Big Data Mining

In today's world, no matter in the fields of economic operation, engineering construction, medical treatment, scientific research and invention or human resource management, a large amount of data is generated every day, and these data are very meaningful for obtaining valuable new discoveries. Through in-depth analysis of these data, people can better grasp people's needs and make accurate decisions. However, due to the large scale, complex content, and many data attributes involved in these data, traditional conventional methods have been unable to analyze these data in a timely and effective manner. Data analysis facilitates information discovery from unstructured, original data. Data mining methods enable information to be extracted from database systems as well as the centralized data or document store. However, big data mining technology can abstract some implicit and useful information from it, and provide decision-makers with more accurate decision-making basis by discovering the correlation or law between different data.

The role of big data mining mainly has two aspects: one is to carry out task prediction, that is, to predict the value of a specific attribute according to some data attributes, so as to achieve the purpose of pre-judgment. The second is to describe the task, that is, to summarize the characteristics of some potential connections in the data through some unique patterns (such as correlation, trend, clustering, anomaly, etc.), so as to explore some characteristics, so as to obtain regularity [6, 7, 8].

2.2 Common Analysis Methods of Big Data Mining

According to the task requirements of big data mining, there are several commonly used analysis methods.

1. Predictive modeling: The model is mainly used to establish a model for the target variable by explaining the function of the variable. A quantitative logistic regression utilized for prediction evaluation is known as linear regression. Among the most basic and straightforward methods, its system analysis to demonstrate the connection among dependent variable. It mainly has two types of modeling tasks. One is classification, which is mainly used to predict discrete target variables. The second is regression, which is mainly used to predict continuous target variables.
2. Association analysis: It is mainly used to discover patterns of strongly associated features in data. The goal of association rule is to uncover intriguing connections within huge datasets. Both frequent item sets and clustering algorithms can be used to describe these fascinating interactions. A group of objects that commonly appear with others is called a frequent pattern combination. At its most basic stage, association rule extraction uses machine learning algorithms to search databases for similarities or co-occurrences in information. It detects common if-then relationships, that are the connection laws in and of itself. The patterns it finds are usually represented in the form of subsets with certain characteristics, and from large-scale data, it can extract the most interesting patterns in an efficient manner. Finding trends or correlations within a database that happen regularly is known as regular pattern mining in information gathering. This is often accomplished by searching through huge databases for things or groups of things that commonly occur simultaneously. It mainly discovers interesting associations or correlations between large transaction or relational datasets by mining frequent patterns, associations and correlations. By spotting patterns in the data, we may group objects that are highly connected collectively as well as quickly spot shared traits and correlations. Frequent pattern mining opens the door to independent research, such as grouping, categorization, or other information mining operations. The research focuses on mining frequent patterns and discovering effective association rules. Forecasting, correlation, or grouping are

the three broad subcategories into which data mining activities and structures can usually be divided. The main algorithms are Apriori algorithm, FP-growth algorithm, equivalence class transformation algorithm, etc. The Apriori algorithm uses database systems to establish frequent patterns and mine frequently occurring item sets. So soon as such item sets exist in the dataset frequently enough, it moves forward by detecting the frequently specific components or expanding those to progressively larger subsets. For frequent patterns mining, the FP-growth method is an enhanced variant of the Apriori method. It is a method for analysing sets of data to discover recurring connections or patterns.

3. Classification analysis: It is mainly used to determine which predefined target class the object belongs to. Its task is mainly to obtain an objective function f by learning, and then map each attribute set x to a predefined class label y. Classification generally has to go through two stages. The first stage is the learning stage, that is, the establishment of a classification model, which mainly uses the classification algorithm to construct a classifier through analysis and learning; the second stage is the classification stage, that is, using the previous learning. A classification model to predict the class label for the given data. The main classification methods are decision tree classification, rule-based classification, neural network, support vector machine and naive Bayes classification.
4. Cluster analysis: It is mainly used to divide data into meaningful or useful groups. Its task is to group data objects according to the information in the data that describes the objects and their relationships, and the objects within the group are similar (related) to each other. While objects in different groups are different (unrelated), the greater the similarity (correlation) within the group and the greater the difference between groups, the better the clustering. It has a wide range of applications in psychology, social sciences, biology, statistics, pattern recognition, information retrieval, machine learning and data mining. Clustering has different models, mainly hierarchical and divided, mutually exclusive, overlapping and fuzzy, complete and partial. The clusters clustered in practice can be divided into distinct separation, prototype-based, graph-based, density-based, and common property, etc., all of which are meaningful. There are three main types of clustering algorithms, namely K-Means algorithm (prototype-based, partitioned), agglomerative hierarchical clustering algorithm (graph-based), DBSCAN algorithm (density-based) and so on.
5. Anomaly detection: It mainly recognizes that the value of a special data point in the data is significantly different from the value of other data points, and has always achieved the purpose of early warning and prevention. An isolated data item that differs significantly from other values in the database is known as an outlier. It is a database abnormality that could have been brought about by a variety of mistakes in the collection, storage, or data manipulation. Because anomalies can distort general data patterns, anomaly detection systems are crucial in statistical. Outliers, also known as outliers, are points that are far away from other data points in the distribution map. Anomalous data objects

are unusual, or significantly inconsistent, so an analysis of outliers can reveal a bit of useful information. To look over thousands of transactions, classify, organize, or partition information in order to locate trends or identify theft, information extraction. In order to effectively identify unusual activity, neural nets acquire characteristics that appear suspect. It has a wide range of applications in fraud detection, intrusion detection, ecosystem imbalance detection, public health anomaly detection, and unusual case detection in healthcare systems.

2.3 Principle of Apriori Algorithm

Apriori is an algorithm for mining association rules in big data [9, 10]. Association rules are implication expressions in the form of $X \to Y$, where X and Y are disjoint item sets, namely $X \cap Y = \emptyset$. An associating policy's conviction or acceptance can be used to determine how strong it is. In contrast to Apriori, that searches the occurrences for every repetition, this method must search the dataset once. This approach is faster because it does not couple the elements. A compressed version of the information is kept in recollection. Support refers to the proportion of buying X and buying Y in the total number of item sets. The confidence level is to determine the proportion of Y buying both X and Y on the premise of buying X. The forms of support S and confidence C are defined as follows:

$$C(X \to Y) = \frac{\partial(X \cup Y)}{\partial(X)}, \tag{1}$$

$$C(X \to Y) = \frac{\partial(X \cup Y)}{\partial(X)}. \tag{2}$$

The principle of certainty, that is the proportion of the number of events that include the issue setting towards the quantity of transactions that have the antecedents, could be used to build frequent patterns after the common sets of items have been identified. The basic principle of Apriori is to generate an item set, denoted as L_1, by scanning the database for the first time and combining and decomposing. Then, through the L_1 item set, set the minimum support degree, continue to combine the items that satisfy the support degree, and obtain 2-item sets from 1-item set, denoted as L_2. Then, from 2-item sets, continue to generate 3-item sets recursively until no more frequent item sets are generated, and the whole process is over. The flow of the Apriori algorithm is as follows:

1. Set the minimum support threshold as $\min S$, and the credibility threshold as $\min C$;
2. Scan the database D, and set C_1 to represent the candidate item set, differentiate the frequent 1-item sets through $\min S$, and obtain the frequent 1-item sets L_1;
3. The candidate 2-item set C_2 is obtained after the L_1 operation, and then C_2 is divided by $\min S$ to obtain the frequent item set L_2;

4. Repeat the iteration until the largest frequent item set L_k is found to stop the iteration;
5. A strong rule equal to or exceeding $\min C$ is mined from the frequent item set L, and the rule is formulated as an association rule.

2.4 Principle of FP-Growth Algorithm

FP-Growth is a tree-based frequent item set mining algorithm. It recursively constructs a conditional FP-Tree and mines rule item sets on the conditional FP-tree [11, 12]. FP-Growth takes a divide and conquer approach, recursively transforming one problem into multiple problems. Store compressed transaction data along with important information about frequent item sets:

1. Contains a root node specified as empty, a set of item prefix subtrees and frequent item header tables as child nodes of the underlying node;
2. Each piece of data in the item header table records two parts, one part is the item set data, and the other part is the support count.

The running idea of FP-Growth algorithm is as follows:

1. For the frequent items to be mined, first construct a tree structure generated according to the transaction data list;
2. Construct the partition of the conditional pattern base for the generated tree structure. According to the item sets of the conditional pattern base, arrange the item sets that are greater than the minimum support count in the path to generate the conditional FP tree;
3. After the conditional FP tree is generated, the nodes under the previous path form a frequent item set, and the number of paths can be combined and connected with the previous path of the tree to obtain frequent items.

The FP-Growth algorithm obtains the number of items and support at the beginning of scanning the dataset. Sort the item sets in descending order of support. Then create the root node Root of the FP tree, and record it as empty. Then do the following operations for each item in the data set, and select the transaction item through the arrangement order of the item set. Then call the function inserted into the tree structure with the properly sorted transaction item table in the transaction and generate the result set.

The flow of the FP-Growth algorithm is as follows:

Step 1: Construct the FP tree:

1. Scan the database to count each item set;
2. Define the minimum support $\min S = 20\,\%$, that is, the minimum support (the minimum number of times the item appears) is 2;

3. Rearrange the item set in descending order. If there is an item less than 2, it needs to be deleted;
4. Readjust the item list in the database according to the number of item occurrences;
5. Build the FP tree: add the first record $(I2, I1, I5)$, add the second record $(I2, I4)$, and accumulate $(I2)$ when the same node appears. Add the third record $(I2, I3)$, the fourth record $(I2, I1, I4)$, the fifth record $(I1, I3)$, the sixth $(I2, I3)$, the seventh $(I1, I3)$, the eighth $(I2, I1, I3, I5)$, and the ninth $(I2, I1, I3)$, so the FP tree is established.

Step 2: Mining frequent item sets:

1. First, in the order from bottom to top, consider $I5$ first, get the conditional pattern bases $\langle(I2, I1 : 1)\rangle, \langle(I2, I1, I3 : 1)\rangle$, construct the FP tree, delete the nodes less than the support degree, form a single path and combine to get the $I5$ frequent item set $\{(I2, I5 : 2), (I1, I5 : 2), (I2, I1, I5 : 2)\}$;
2. Then consider $I4$, obtain the conditional pattern basis $\langle(I2, I1 : 1)\rangle, \langle(I2 : 1)\rangle$, construct the conditional FP tree, and obtain the $I4$ frequent item set $\{(I2, I4 : 2)\}$;
3. Then consider $I3$, and get the conditional pattern basis $\langle(I2, I1 : 2)\rangle, \langle(I2 : 2)\rangle, \langle(I1 : 2)\rangle$ to construct the conditional FP tree. Since this tree is not a single path, it is necessary to recursively mine $I3$ and consider $I3$ recursively. At this time, the conditional pattern base $\langle(I2 : 2)\rangle$ of $I1$ is obtained, that is, the conditional pattern base of $I1, I3$ is $\langle(I2 : 2)\rangle$, and the conditional FP tree is constructed to obtain the frequent item set $\{(I2, I3 : 4), (I1, I3 : 4), (I2, I1, I3 : 2)\}$ of $I3$;
4. Finally, consider $I1$, get the conditional pattern base $\langle(I2 : 4)\rangle$ to construct the conditional FP tree, and get the frequent item set $\{(I2, I1 : 4)\}$ of $I1$.

2.5 Principle of K-Means Clustering Algorithm

The K-Means algorithm is a well-known algorithm in the distance-based clustering algorithm. It uses distance as the evaluation index of similarity, and its principle is: if the vectors representing each point are close to each other in space, these points can be regarded as a class [13, 14]. That is, for a given sample set, when classifying, the sample set is divided into K classes according to the distance between samples, and the distance between points within a class is as small as possible, and the distance between classes is as large as possible. Max-min distance metric is used in K-Means grouping. Nevertheless, as the total of quadratic departures from the median is equivalent to the total of bilateral quadratic Euclidean values multiplied by the number of locations, K-Means is indirectly predicated on bilateral Euclidean similarities among sample points.

The K-Means algorithm has a basic assumption for the data that needs to be clustered: for each class, a center point can be selected so that the distance from all points in the class to the center point is less than the distance to the centers of other classes. However, in the actual classification, the obtained data often cannot directly meet such requirements, and can only be as close as possible. The reasons for such differences are often inherent in the data itself, or the data can no longer be classified.

The method of the K-Means algorithm is to select K initial reference points according to the input parameter K, and divide all the points in the data set into K classes according to the K reference points. Given that movement is subjective by design, frames of reference are crucial since they accurately express an individual's attitude. It makes sense that the more optimally placed such initially cluster centers are placed, the less repetitions of the K-Means classification methods will be required to reach an ideal set of anchor nodes or grouping participation according to the distance from such anchor nodes. The centroid of these K classes (the average of all points in the class) is used as the reference point for the next iteration, and the dataset is divided into final K classes through continuous iterative updates. The iteration makes the chosen reference points get closer and closer to the true class centroid, consequently, the clustering effect is improving.

Let the Equation (3) q dimensional data set

$$W = \{w_i \mid w_i \in R^q, i = 1, 2, \dots, N\} \tag{3}$$

be aggregated into K classes α_1, α_2, $\dots$, α_K, and their centroids are in turn z_1, z_2, $\dots$, z_K, where:

$$z_i = \left(\frac{1}{m_i}\right) \sum_{w \in \alpha_K} w. \tag{4}$$

In Equation (4), m_i is the number of data points in class α_i. The quality of the clustering effect is represented by the objective function E:

$$E = \sum_{i=1}^{k} \sum_{j=1}^{m_i} b_{ij}\left(w_j, z_i\right). \tag{5}$$

In Equation (5), $b_{ij}\left(w_j, z_i\right)$ is the Euclidean distance between w_j and z_i. The objective function E is actually the sum of the distances between each data point and the centroid of the class, so the smaller the E value, the more compact the class is. Therefore, the algorithm can seek a good clustering method by continuously optimizing the value of E. When E takes a minimum value, the corresponding clustering method is the optimal method.

The steps of the K-Means algorithm are as follows:

Step 1: randomly select K initial reference points z_1, z_2, $\dots$, z_K from W;

Step 2: Use z_1, z_2, $\dots$, z_K as a reference point to divide W. The division is based on the following principles: If $b_{ij}\left(w_i, z_j\right) < b_{in}\left(w_i, z_n\right)$, among them, $n = 1, 2, \dots, K$; $j = 1, 2, \dots, K$; $j \neq n$; $i = 1, 2, \dots, N$, then divide w_i into class α_j;

Step 3: According to Equation (4), recalculate the centroid z_1^*, z_2^*, …, z_K^* of the class;

Step 4: If $z_i^* = z_i$ is true for any $i \in \{1, 2, \dots, K\}$, the algorithm ends, and the current z_1^*, z_2^*, …, z_K^* represents the final class; Otherwise, let $z_i^* = z_i$ go back to step 2 for execution.

In order to prevent the infinite loop from occurring because the termination condition in step 4 cannot be satisfied, a maximum number of iterations is usually set in the algorithm, or a fixed threshold β is set. The algorithm is considered to end when there is $|z_i^* - z_i| < \beta$ for all z_i. The iterative process of the K-Means algorithm is shown in Figure 1.

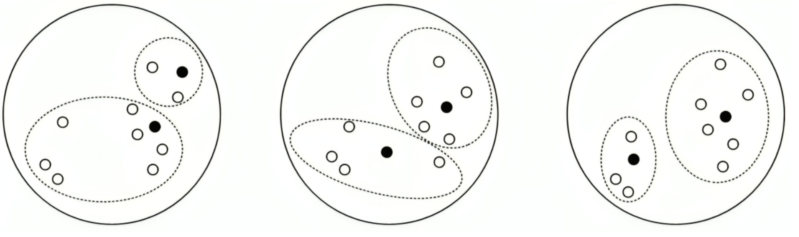


Figure 1. Iterative process of K-Means algorithm

The time complexity of the K-Means algorithm is $O(\chi KN)$. Where χ is number of algorithm iterations, K is the number of categories, and N is the number of data points in the dataset. The size of k, that must always be provided in order to conduct any grouping investigation, determines how the K-Means method operates. Ultimately, grouping with various input parameters will yield various outcomes. Many factors can affect the performance of the K-Means algorithm, including the number of cluster categories K, the choice of initial reference points, and the type of input data. Therefore, it is necessary to correctly determine the initial reference point according to the specific application, and select a similarity measurement strategy that is consistent with the data type, so as to ensure that the K-Means algorithm obtains better performance.

3 DYNAMIC MATCHING ALGORITHM OF HUMAN RESOURCE ALLOCATION

3.1 Extracting Human Resource Configuration

In order to effectively realize the dynamic matching of human resource allocation and further reduce the amount of data in the matching process, information entropy is selected as the basis for measuring human resource allocation, and an appropriate amount of human resource allocation is obtained for subsequent analysis and processing of the similarity of the allocation. Data minimization can save collection expenses while also increasing storing speed and economy. Feature extraction employs a variety of techniques to lessen the amount of information that is kept on the device.

Using information entropy as an evaluation parameter, the human resource allocation at this time has more characteristic information for human resource allocation analysisUtilizing HR information enables us to develop a company or its individual divisions in the ways that we desire it to expand. Our firm will expand responsibly and employ a content staff with the support of thorough HR dataset obtained or a solid method of marketing. In particular, information entropy can also be used to measure the amount of information about human resource allocation. It is believed here that when the amount of information on human resource allocation is more, that is, the greater the number of different human resource allocations, it has more reference value in the analysis of human resource allocation. Whenever all possible benefits are taken into account, data entropy is typically defined as the average quantity of data that an occurrence conveys.

Human resource allocation is divided by time and there is

$$R = \{SubR_1, SubR_2, \dots, SubR_5\}.$$

Take one of the sub-human resources allocation $SubR_a$ $(1 \leq a \leq S)$ as an example in Equation (6):

$$\mathrm{SubR_a} = \{y_{k+1}, y_{k+2}, \dots, y_{k+r}\}. \tag{6}$$

There are r individual human resource configuration points in this sub-human resource configuration. Among them, different human resource allocation values are $U = \{u_1, u_2, \dots, u_e\}$, the number of occurrences of each human resource allocation value is $USum$. The calculation formula is as follows:

$$P_i = \frac{U\,\mathrm{Sum}_i}{\mathrm{sum}_j}\ (1 \leq i \leq \theta). \tag{7}$$

In Equation (7), $USum$ is the quantity of the human resource configuration in the sub-human resource configuration, sum_j is the total number of human resource configuration points in the sub-human resource configuration, and e is the number of non-repetitive human resource configuration in the subhuman resource configuration. The degree of surprise (or uncertainty) associated with the quantity of a stochastic process or the result of a randomized operation is measured by entropy values, often known as Shannon's entropy. Its importance in the tree structure comes from the fact that it enables us to calculate the diversity or impurities of the target attribute. Joint Entropy and Conditional Probability are the two kinds. The information entropy of the human resource allocation can be obtained by the frequency of the above human resource allocation points, as shown in Equation (8):

$$H_j = -\sum_{i=1}^{e} P_i \log_2 P_i. \tag{8}$$

When each human resource configuration has the corresponding information entropy, extract the information entropy $H > 0$, and a total of H_s human resource configurations. And according to the size of the information entropy, the extraction percentage δ of human resource allocation is given, and the total number of human resource allocation H_n to be extracted can be calculated and retained. The relevant formula is as follows in Equation (9):

$$H_n = H_s \times \delta. \tag{9}$$

The human resource configuration with the quantity H_n extracted from the above is sorted according to the time period number and retained.

3.2 Calculate the Similarity of Human Resource Allocation

On the basis of the above-mentioned extraction of human resource configuration, the dynamic matching accuracy of human resource configuration is high. Therefore, using the Huasdgrff similarity method based on time interpolation, the similarity of human resource allocation is calculated. Training personnel and coordinating their goals for personal growth with the overarching objectives of the organisation or company are the main concerns of developing human resources. The administration of human resources places more of an emphasis on conformity, insurance, pay, and labour rights.

In order to improve the dynamic matching accuracy of human resource allocation, a time interpolation method is introduced to supplement human resource allocation. Non-human assets are material goods or items that are present without the presence of individuals. Humans are able to view, interact with, and use them. Material wealth are another name for non-human elements. Vehicles, clinics, schools, libraries, playgrounds, gasoline, laptops, magazines, calendars, flowers, and cash are a few instances. From the target human resource configuration R_A and the matching human resource configuration R_B, under each human resource configuration time period, two non-empty human resource configurations $SubR_A$ and $SubR_B$ are extracted. Retain the time when the $SubR_A$ non-empty sub-HR configuration appears. For segments in the chronology whose frequency deviates from the sequencing parameters, time approximation is used. When transferring the schedule in its entirety to a frequency other than the series parameters, temporal compression is used. The possible positions of $SubR_B$ non-kongzi human resource allocation under these times are calculated by the $SubR_B$ non-kongzi human resource allocation function, that is, the interpolation point. Similarly, keep the time when the $SubR_B$ non-empty sub-human resource configuration appears, and calculate the position where the $SubR_A$ non-empty sub-human resource configuration appears under these times. The extracted human resource configuration needs to complete the above interpolation operation.

If an opponent selects a location within one of the different pairs, from which one should subsequently move to the second setting, the Hausdorff duration is the greatest length that can be required of someone. It represents the largest possible distance between a location in one set and its nearest neighbour in the opposite set. Using the Hausdorff distance method [15], the human resource allocation after time interpolation is matched with the target human resource allocation by similarity. One can gain the following skills through working as a HRM expert: understanding of regional, state wide, or national employment regulations; knowledge of the Office Software Suite and personnel software solutions, understanding of experienced experts and planning tools in the field. According to the concept, the operational plan and HR systems must be handled in a manner that is in line with the corporate objectives. Add up the Haysdorff distances obtained between each corresponding human resource configuration, and obtain the mean value as the final distance between human resource configurations. The specific formula is as follows:

$$H_k\left(SubR_A, SubR_B\right) = \max\left(h_k\left(SubR_A, SubR_B\right)\right), \left(h_k\left(SubR_B, SubR_A\right)\right). \quad (10)$$

In Equation (10), $H_k\left(SubR_A, SubR_B\right)$ is the Hausdorff distance between $SubR_A$ and $SubR_B$ of nonempty human resource allocation, $h_k\left(SubR_A, SubR_B\right)$ is the one-way Hausdorff distance from $SubR_A$ to $SubR_B, h_k\left(SubR_B, SubR_A\right)$ is the one-way Hausdorff distance from $SubR_B$ to $SubR_A$. Through the above process, the similarity of human resource configuration is calculated accordingly. The closest locations will possess the smallest distances whenever evaluating through length; however, the closest locations will display the greatest similarities when comparing by resemblance.

3.3 Classification and Mining of Human Resource Allocation

The implemented to achieve of human resource distribution is separated into four intervals following the comparability of human resource allocation calculation in order to enable categorization and extraction of personal allocation of resources

$$u_1 : 0, 5), u_2 : 5, 10), u_3 : 10, 15), u_4 : 10, +\infty).$$

Finding the economic indicators among the provided datasets is made easier by doing this. To calculate the derivative of a variable for an intermediary variable of the independence variable, this procedure is usually required. A daemon in the local scheduler records time slice information for human resource configuration. The human resource configuration transaction set is shown in Table 1.

In Table 1, (d_1, d_2, d_3) represents the time slice information of human resource configuration, u_{ij} represents the time slice interval of human resource configuration i is j, and the time slice without computing tasks is marked as u_{i1}, where i is the human resource configuration number.

According to the Apriori algorithm and the FP-Growth classification algorithm, the transactions in Table 1 are mined, and the minimum support count $\min S = 2$

ID	d_1	d_2	d_3	User ID
1	u_{11}	u_{22}	u_{31}	c_1
2	u_{12}	u_{21}	u_{32}	c_2
3	u_{14}	u_{21}	u_{31}	c_1
4	u_{11}	u_{21}	u_{31}	c_1
5	u_{12}	u_{22}	u_{33}	c_2
6	u_{13}	u_{22}	u_{34}	c_3
7	u_{12}	u_{22}	u_{32}	c_2

Table 1. Human resource configuration transaction set

is set, and the frequent pattern is

$$\{\langle u_{11}, u_{31}, c_1\rangle, \langle u_{12}, u_{22}, c_2\rangle, \langle u_{12}, u_{32}, c_2\rangle, \langle u_{21}, u_{31}, c_1\rangle\}.$$

Set the minimum confidence min $C = 70\,\%$ to generate the classification rules as Equations (11) and (12):

$$u_{11} \wedge u_{31} \Rightarrow c_1, u_{11} \wedge c_1 \Rightarrow u_{31}, u_{11} \Rightarrow u_{31} \wedge c_1, u_{21} \wedge u_{31} \Rightarrow c_1, u_{21} \wedge c_1 \Rightarrow u_{31}, \quad (11)$$

$$u_{12} \wedge u_{22} \Rightarrow c_2, u_{22} \wedge c_2 \Rightarrow u_{12}, u_{12} \wedge u_{32} \Rightarrow c_2, u_{32} \wedge c_2 \Rightarrow u_{12}, u_{32} \Rightarrow c_2 \wedge u_{12}. \quad (12)$$

According to the correspondence between u_{ij} and d_i, it is concluded that the time slice of c_1 using d_1 and d_2 is u_1 or the time slice using d_2 and d_3 is $u_1 \cdot c_2$ uses d_1 and d_3 time slices for u_2 or d_1 and d_2 time slices for u_2.

A data gathering systems could also be categorized according to the types of information mining, expertise mineable, methodologies used, and systems adopted. At the same time, classification and mining can be carried out according to the configuration of human resources, and classification rules based on configuration can be generated, and the classification rules can be obtained as follows:

$$u_2 \wedge c_1 \Rightarrow u_1, u_1 \wedge c_2 \Rightarrow u_2, u_2 \wedge c_2 \Rightarrow u_1, c_2 \Rightarrow u_1 \wedge u_2. \quad (13)$$

That is, if c_1 uses the u_2 task request, c_1 will also use the u_1 task. c_2 uses both types of time slices u_1 and u_2 when requesting human resource allocation.

3.4 Realize Dynamic Matching of Human Resource Allocation

After classifying and mining human resource allocation, K-Means clustering algorithm is used to realize dynamic matching of human resource allocation. Suppose there are g configurations in human resources, which are $p_1, p_2, \ldots, p_g$, respectively, each configuration has h_i task requests, and the time length of task requests is denoted as U_{ij}. Among them, i is the human resources configuration number, j is the task number, and $1 \leq i \leq g, 1 \leq j \leq h_i$. From g, h_i and U_{ij}, it can be concluded that the total length of time slice U_s for all human resource allocation and all task requests is expressed by Equation (14):

$$U_s = U_{11} + U_{12} + \cdots + U_{1h_1} + \cdots + U_{g1} + U_{g2} + \cdots + U_{gh_g}. \tag{14}$$

Suppose there are k departments in the enterprise, which are responsible for providing corresponding human resource allocation, which are $l_1, l_2, \ldots, l_k$ respectively. Each department contains V_i personnel, and each personnel includes ε isomorphic human resource allocations. The total number of human resource allocations B_s in the entire enterprise is in Equation (15):

$$B_s = \varepsilon \times (V_1 + V_2 + \cdots + V_k). \tag{15}$$

Through classification mining, the human resource allocation of the same department can be classified together, resulting in the following rules:

$$c_i \wedge \cdots \wedge c_j \Rightarrow l_\phi. \tag{16}$$

In Equation (16), $1 \leq i \leq k$, $1 \leq j \leq g$, $1 \leq \phi \leq k$. Based on the classification rules and mining patterns obtained by the classification mining algorithm, the K-Means clustering algorithm is used to submit the grouped tasks to the corresponding personnel, and distribute them to the personnel corresponding to the human resource configuration according to the physical location of the task. In the human resource allocation B_s in the whole enterprise, $f_r = \sum_{i=l_\phi} f_i$, that is, after any one task is completed, it means that the human resource allocation is insufficient. When $f \sum_{i=l_\phi} f_{i_{\max}}$, the corresponding personnel will execute within B_s with the closest adjacent distance, and the remaining tasks will be handed over to B_s for dynamic matching. Through the above steps, dynamic matching of human resource allocation is realized.

4 EXPERIMENTAL ANALYSIS

4.1 Experimental Environment and Data

In order to verify the effectiveness of the dynamic matching algorithm of human resource allocation based on big data mining, the experimental platform adopts Thinksad L430 loaded with Windows 7 system, the memory is 4 GB, and the CPU is Intel Core i7 with 2.9 GHz. Non-relational documents library MongoDB supports retention which is similar to JSON. The MongoDB system has complete querying capabilities, persistence, as well as a configurable schema that makes it possible to hold complex data. It also features comprehensive and user-friendly APIs. The experiment uses Java language to operate MongoDB for data processing and analysis, and uses Java and Matlab to implement the proposed algorithm. Multiple information scientific operations, such as data gathering, comprising information importing, metadata management, deep learning, scientific techniques, Natural Language Processor (NLP), or data visualisation, make extensive use of Java. The main parameter settings are shown in Table 2.

Parameter	Numerical Value
Number of departments in the enterprise	$k = 3$
Number of people in each department	$V_i = 5$
Number of processors per node The number of homogeneous human resource allocations in each person	$\varepsilon = 8$
The total number of human resource allocations in the entire enterprise	$B_s = 120$

Table 2. Main parameter settings

The algorithm of reference [4], the algorithm of reference [5] and the proposed algorithm are respectively used to compare the dynamic matching resource utilization, dynamic matching accuracy and dynamic matching time of different algorithms, so as to verify the effectiveness of the proposed algorithm.

4.2 Comparison Results of Dynamic Matching Effect of Human Resource Allocation

In order to verify the dynamic matching effect of human resource allocation of the proposed algorithm, the dynamic matching resource utilization rate is taken as the evaluation index. The higher the dynamic matching resource utilization rate is, the better the dynamic matching effect of the algorithm is. The algorithm of reference [4], the algorithm of reference [5] and the proposed algorithm are used to compare, and the comparison results of the resource utilization ratio of dynamic matching of human resource allocation of different algorithms are obtained as shown in Figure 2.

According to Figure 2, when the total number of human resource allocations in the entire enterprise reaches 120, the average human resource allocation dynamic matching resource utilization rate of the algorithm of reference [4] is 89.3 %, the average human resource allocation dynamic matching resource utilization rate of the algorithm of reference [5] is 81.2 %. The average human resource allocation dynamic matching resource utilization rate of the proposed algorithm is as high as 98.7 %. It can be seen that the dynamic matching of human resource allocation of the proposed algorithm has a higher utilization rate of resources, indicating that the dynamic matching of human resources allocation of the proposed algorithm has a better effect.

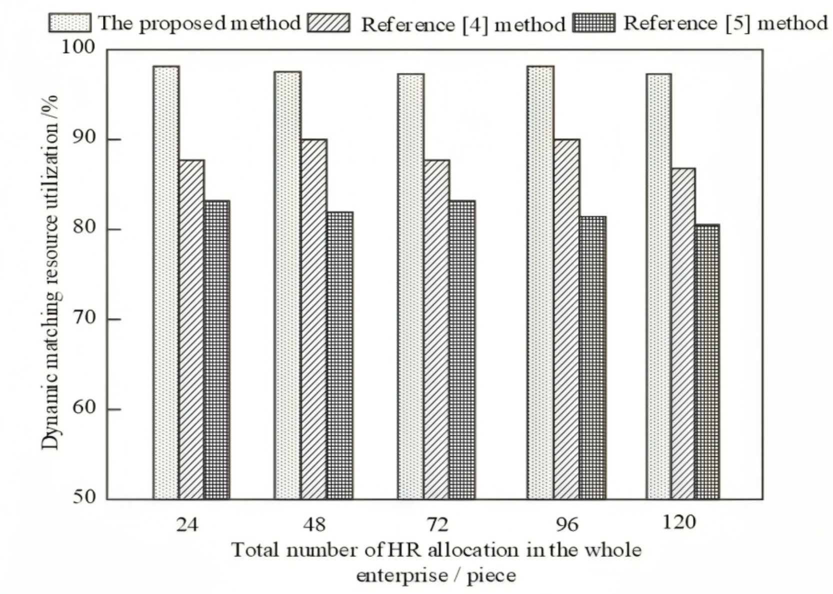


Figure 2. Comparison results of resource utilization ratio of dynamic matching of human resource allocation with different algorithms

4.3 Comparison Results of Dynamic Matching Accuracy of Human Resource Allocation

On this basis, the dynamic matching accuracy of human resource allocation of the proposed algorithm is further verified, and the dynamic matching accuracy is used as the evaluation index. The higher the dynamic matching accuracy, the higher the dynamic matching accuracy of the algorithm's human resource allocation. The algorithm of reference [4], the algorithm of reference [5] and the proposed algorithm are used to compare, and the comparison results of the dynamic matching accuracy of human resource allocation of different algorithms are shown in Figure 3.

According to Figure 3, when the total number of human resource allocations in the entire enterprise reaches 120, the average dynamic matching accuracy of human resource allocation of the algorithm of reference [4] is 88.5 %, the average human resource allocation dynamic matching accuracy of the algorithm of reference [5] is 80.1 %. The average dynamic matching accuracy rate of human resource allocation of the proposed algorithm is as high as 95.8 %. It can be seen that the dynamic matching accuracy of human resource allocation of the proposed algorithm is high, indicating that the dynamic matching accuracy of human resource allocation of the proposed algorithm is high.

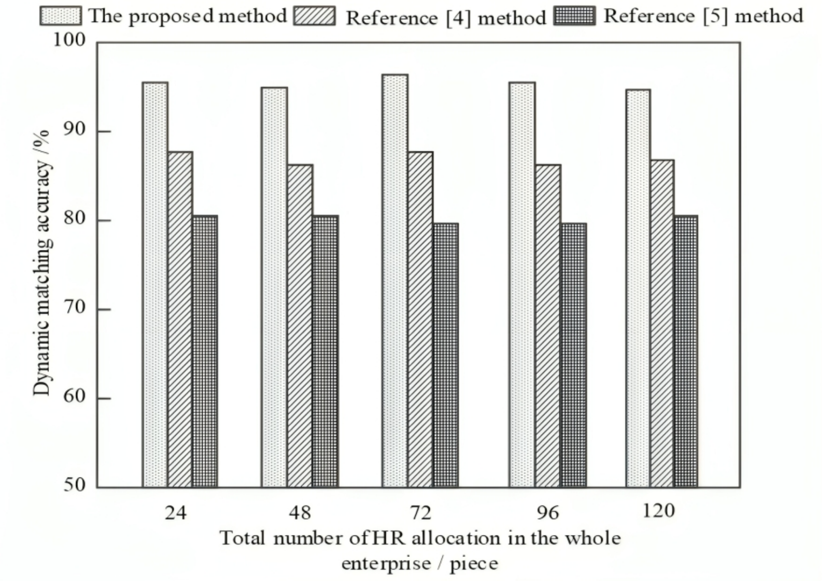


Figure 3. Comparison results of dynamic matching accuracy of human resource allocation of different algorithms

4.4 Comparison Results of Dynamic Matching Efficiency of Human Resource Allocation

The dynamic matching efficiency of human resource allocation of the proposed algorithm is further verified, and the dynamic matching time is used as the evaluation index. The process is check out the present personnel, establish a transition strategy, make a decision on how to expand assets going forward, make plans for employee training, besides conduct a gap assessment. The shorter the dynamic matching time is, the higher the dynamic matching efficiency of human resource allocation of the algorithm is. The algorithm of reference [4], the algorithm of reference [5] and the proposed algorithm are used to compare, and the comparison results of the dynamic matching time of human resource allocation of different algorithms are shown in Table 3. According to the concept, the administrative architecture or HR department must be handled in a manner that is in line with the corporate objectives. The approach is employed to make it easier to fulfil the organization's objectives in the areas of production, profitability, or effectiveness.

According to the data in Table 3, with the increase of the total number of human resource allocation in the whole enterprise, the dynamic matching time of human resource allocation of different algorithms gradually increases. When the total number of human resource allocations in the entire enterprise reaches 120,

Total Number of HR Allocation in the Whole Enterprise [Piece]	The Proposed Algorithm [s]	The Algorithm of Reference [4] [s]	The Algorithm of Reference [5] [s]
24	0.8	2.8	4.9
48	1.9	4.6	6.2
72	2.6	6.9	8.9
96	3.5	8.8	10.2
120	5.1	10.2	12.6

Table 3. Comparison results of dynamic matching time of human resource allocation with different algorithms

the dynamic matching time of human resource allocation of the algorithm of reference [4] is 10.2 s, and the dynamic matching time of the algorithm of reference [5] is 12.6 s. The dynamic matching time of human resource allocation of the proposed algorithm is only 5.1 s. It can be seen that the dynamic matching time of human resource allocation of the proposed algorithm is short, indicating that the dynamic matching efficiency of human resource allocation of the proposed algorithm is high.

5 CONCLUSION

This paper studies the dynamic matching algorithm of human resource allocation based on big data mining, and adopts the relevant algorithms of big data mining to realize the dynamic matching of human resource allocation. The algorithm has better dynamic matching effect of human resource allocation, and can effectively improve the accuracy and efficiency of dynamic matching of human resource allocation. However, the algorithm is only used for simulation experiments of a small amount of human resource allocation, and does not consider the situation of massive human resource allocation. Therefore, in the following research, further research on the allocation of massive human resources is needed.

6 DECLARATIONS

Funding: No funds, grants were received by any of the authors.

Conflict of interest: There is no conflict of interest among the authors.

Data availability: All data generated or analysed during this study are included in the manuscript.

Code availability: Not applicable.

Authors' contributions: All authors contributed to the design and methodology of this study, the assessment of the outcomes and the writing of the manuscript.

REFERENCES

[1] Monyei, E. F.—Agbaeze, K. E.—Isichei, E. E.: Organisational Paranoia and Employees' Commitment: Mediating Effect of Human Resources Policies. International Journal of Scientific and Technology Research, Vol. 9, 2020, pp. 5172–5185.

[2] Xu, J.—Wang, B.—Min, G.: Research on Human Resource Allocation Model Based on SOM Neural Network. Research Anthology on Human Resource Practices for the Modern Workforce, IGI Global, 2022, pp. 513–525, doi: 10.4018/978-1-6684-3873-2.ch027.

[3] Kieling, E. J.—Rodrigues, F. C.—Filippetto, A.—Barbosa, J.: Smartalloc: A Model Based on Machine Learning for Human Resource Allocation in Projects. Proceedings of the 25th Brazillian Symposium on Multimedia and the Web (WebMedia '19), 2019, pp. 365–368, doi: 10.1145/3323503.3360643.

[4] Dabirian, S.—Abbaspour, S.—Khanzadi, M.—Ahmadi, M.: Dynamic Modelling of Human Resource Allocation in Construction Projects. International Journal of Construction Management, Vol. 22, 2022, No. 2, pp. 182–191, doi: 10.1080/15623599.2019.1616411.

[5] Zhao, W.—Pu, S.—Jiang, D.: A Human Resource Allocation Method for Business Processes Using Team Faultlines. Applied Intelligence, Vol. 50, 2020, No. 9, pp. 2887–2900, doi: 10.1007/s10489-020-01686-4.

[6] Chamikara, M. A. P.—Bertok, P.—Liu, D.—Camtepe, S.—Khalil, I.: Efficient Privacy Preservation of Big Data for Accurate Data Mining. Information Sciences, Vol. 527, 2020, pp. 420–443, doi: 10.1016/j.ins.2019.05.053.

[7] Fernandez-Basso, C.—Ruiz, M. D.—Martin-Bautista, M. J.: A Fuzzy Mining Approach for Energy Efficiency in a Big Data Framework. IEEE Transactions on Fuzzy Systems, Vol. 28, 2020, No. 11, pp. 2747–2758, doi: 10.1109/TFUZZ.2020.2992180.

[8] Fernandez-Basso, C.—Francisco-Agra, A. J.—Martin-Bautista, M. J.—Ruiz, M. D.: Finding Tendencies in Streaming Data Using Big Data Frequent Itemset Mining. Knowledge-Based Systems, Vol. 163, 2019, pp. 666–674, doi: 10.1016/j.knosys.2018.09.026.

[9] Zhang, R.—Chen, W.—Hsu, T. C.—Yang, H.—Chung, Y. C.: ANG: A Combination of Apriori and Graph Computing Techniques for Frequent Itemsets Mining. The Journal of Supercomputing, Vol. 75, 2019, No. 2, pp. 646–661, doi: 10.1007/s11227-017-2049-z.

[10] Dharshinni, N. P.: Analysis of Definite Integral Material Topics for Improve Student Learning Using Apriori Algorithm. Journal of Informatics and Telecommunication Engineering, Vol. 4, 2021, No. 2, pp. 294–300, doi: 10.31289/jite.v4i2.4316 (in Indonesian).

[11] Mahrousa, Z.—Alchawafa, D. M.—Kazzaz, H.: Frequent Itemset Mining Based on Development of FP-Growth Algorithm and Use MapReduce Technique. Association of Arab Universities Journal of Engineering Sciences, Vol. 28, 2021, No. 1, pp. 83–98, doi: 10.33261/jaaru.2021.28.1.008.

[12] Firmansyah, F.—Yulianto, A.: Market Basket Analysis for Books Sales Promotion Using FP Growth Algorithm, Case Study: Gramedia Matraman Jakarta. Journal of Informatics and Telecommunication Engineering, Vol. 4, 2021, No. 2, pp. 383–392, doi: 10.31289/jite.v4i2.4539 (in Indonesian).

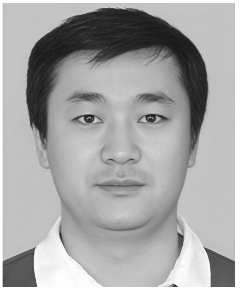
[13] Huyan, J.—Li, W.—Tighe, S.—Deng, R.—Yan, S.: Illumination Compensation Model with K-Means Algorithm for Detection of Pavement Surface Cracks with Shadow. Journal of Computing in Civil Engineering, Vol. 34, 2020, No. 1, Art. No. 04019049, doi: 10.1061/(ASCE)CP.1943-5487.0000869.

[14] Geng, D. Z.—Xu, Q.: High-Dimensional Mixed Attribute Data Mining Method Based on K-Means Clustering Algorithm. Computer Simulation, Vol. 38, 2021, No. 2, pp. 308–312 (in Chinese).

[15] Kraft, D.: Computing the Hausdorff Distance of Two Sets from Their Distance Functions. International Journal of Computational Geometry and Applications, Vol. 30, 2020, No. 1, pp. 19–49, doi: 10.1142/S0218195920500028.

Yuping Yan is a Senior Engineer. In 2007, he graduated from the Sun Yat Sen University, majoring in computer science. In 2010, he graduated from the Sun Yat Sen University with a master's degree in computer software engineering. His research field involves data mining. He has published 24 academic articles, participated in 145 scientific research projects and obtained 16 invention patents in the field of digital grid. He has won 45 awards such as the Company's Science and Technology Progress Award and Management Innovation Award. He is now working in the Guangdong Power Grid Co., Ltd., responsible for the company's digital transformation construction.

Peiyao Xu is an Intermediate Engineer. She graduated from the Guangdong University of Technology with a bachelor's degree in 2010, majoring in computer science and technology. Her research fields involve information system operation and maintenance and digital construction. She has published 2 academic articles, has participated in 2 scientific research projects, won 1 invention patent in the field of digital grid, and won 1 science and technology progress award of the company. She is now working in the China Southern Power Grid Digital Enterprise Technology (Guangdong) Co., Ltd., responsible for the company's digital construction.



Jianyong Wang is a Senior Engineer. He graduated from the China University of Mining and Technology in 2004, majoring in computer science and technology. In 2015, he graduated from the Wuhan University with a master's degree in software engineering. His research fields involve power informatization and software development. He has published 14 academic articles, has participated in 50 scientific research projects and obtained 10 invention patents in the field of digital power grid. He has won 10 awards such as the Company's Science and Technology Progress Award, Technology Improvement Award and Management Innovation Award. He is currently working in the China Southern Power Grid Digital Enterprise Technology (Guangdong) Co., Ltd., responsible for the company's digital transformation construction.