Knowledge Push Method Based on Clustering Algorithm and DSM Matrix

Ding Wang, Hongfei Zhan*, Junhe Yu and Rui Wang
Faculty of Mechanical Engineering & Mechanics, Ningbo University, Ningbo, China
Email: zhanhongfei@nbu.edu.cn

Abstract. In the process of business execution of traditional enterprises, traditional knowledge push is difficult to provide accurate knowledge support for business execution decision-making due to the complicated business problems. Therefore, this paper proposes a knowledge-push model for business execution. Firstly, the business problems in the business process are clustered through the hierarchical clustering algorithm to ensure the cohesion of the business problems after clustering. Then DSM matrix is used to represent the coupling of clustered business problems. Finally, knowledge push is carried out on the divided business problems to provide accurate knowledge support for business executives to make business problems decisions.

Keywords. business execution; knowledge push; clustering algorithm; DSM matrix.

1. Introduction
Each enterprise concentrates on its own core competitive advantage development [1]. As an important part of core competitiveness, knowledge management are widely used in business execution. However, knowledge management cannot provide effective knowledge support for business executives because of the disconnection between business process and knowledge management and the complexity of business problem in business process execution.

The combination of knowledge management and business process has become a hot topic of knowledge management, scholars in China and abroad have conducted corresponding research on it. Liu et al. established the database of business process cases, then used case-based reasoning to mine knowledge of corresponding topics, and recommended it to designers with similar problem [2]. Wang et al. proposed a knowledge correlation network including three similarities of content, context and task to improve the richness of knowledge recommendation for business problems [3]. Wei et al. conducted relevant knowledge mining on product modular design [4]. Yu et al. pushed patent knowledge to the target through deep learning [5].

The existing knowledge management for business execution focuses on how to mine the knowledge involved in business problem, but ignores the influence of business problem on knowledge [6]. There is high coupling between some business issues and the subject matter of the business issues themselves is vague, which leads to mutual restriction and interference between business issues. In the process of business execution, the inability to accurately provide knowledge assistance for different business problems causes a great waste of resources. This paper uses hierarchical clustering algorithm and DSM matrix to divide business problems. Use knowledge push for the divided business problems to establish a knowledge-assisted model. Provide accurate knowledge assistance for business problem decision-making.
2. Knowledge Push Model Framework

The main appeal of enterprises attaching to knowledge management is to provide auxiliary support for business problem that arise in the execution of business processes, and to improve the efficiency and accuracy of problem solving [7]. In this paper, the knowledge push framework between business process execution process, business problem allocation, business problem handling, knowledge resources and business personnel is established as shown in figure 1.

![Image of Knowledge Push Model Framework](image)

**Figure 1. Knowledge push framework.**

The system framework is mainly composed of five parts: business process layer, business problem layer, knowledge push layer, business personnel layer and knowledge resource library. The business process layer mainly analyses the specific business, which refers to the business cases in the knowledge resource base to determine the execution process of specific business activities. After the business process is decomposed, the business problems to be solved are obtained and sent to the business problem layer. The business problem layer performs cluster analysis on the underlying business problem, clusters similar business problem, and conducts coupling analysis on them through the DSM matrix, and aggregates them into business problem with high cohesion and low coupling, and Send to the knowledge push layer. After the knowledge push layer receives the specific business problem, it conducts text mining on the relevant knowledge in the knowledge resource database, and calculates the matching degree of the mined text, and pushes it to the business staff layer when the matching degree threshold is reached. The business staff layer records the identity, business experience and specific capabilities of each business staff. By clarifying the specific capabilities required to solve the corresponding business problem, the business personnel layer automatically assigns business problem to specific business personnel. After the execution of the business process is completed, take the business process as a specific case and update it into the knowledge resource database to provide case assistance for the
execution of similar business activities in the future, thus forming a complete knowledge assistance model.

3. Division of Business Problem

In the process of business execution, the relationship between knowledge is too complex because of the large scope of business problem and the wide range of knowledge involved. According to the requirements of high cohesion, low coupling and high parallelism for the division of business problem, hierarchical clustering algorithm is used to aggregate business problem involving similar knowledge, so that business personnel can refer to the specific knowledge of a topic for processing [8]. DSM matrix is used to express the business problems after clustering. According to the results of DSM matrix, business personnel aggregate the business problems with high coupling degree to prevent mutual interference between business problems and maximize the value of knowledge resources.

3.1. Hierarchical Clustering of Business Problems

There are a large number of business problems in the process of business execution. Different business problems present hierarchical characteristics, and the relationship between them is too complex, which will lead to mutual interference in the process of business execution. In this paper, according to the hierarchical characteristics of business problems, there are similarities between multiple underlying business problems. By clustering the underlying business problems, multiple similar underlying business problems can be aggregated to maximize the utility of business personnel and improve the execution efficiency of business process [9].

In this paper, hierarchical clustering algorithm is used for processing, hierarchical clustering algorithm is a kind of clustering algorithm, through the similarity between the underlying business problem to aggregate into a multi-level clustering feature tree. Hierarchical clustering algorithm can show different clustering situations at different levels [10]. In this paper, hierarchical clustering tree is aggregated from bottom to top.

In the process of business execution, the number of business problem is relatively large, and different business problem present hierarchical characteristics, and the relationship between them is too jumbled, which will lead to mutual interference in the process of processing [11]. According to the hierarchical characteristics of business problem, multiple underlying business problem can be processed by business personnel with a certain ability to cluster the underlying business problem, so that multiple similar problem can be aggregated to maximize the utility of business personnel and improve the execution efficiency of business process.

In this paper, hierarchical clustering algorithm is used for processing, hierarchical clustering algorithm is a kind of clustering algorithm, through the similarity between the underlying business problem to aggregate into a multi-level clustering feature tree. Hierarchical clustering algorithm can show different clustering situations at different levels. In this paper, hierarchical clustering tree is aggregated from bottom to top.

First, all the underlying business problem in the business process are regarded as the initial clusters, and a threshold is set as the distance metric between the initial different clusters. The smaller the distance is, the higher the similarity is. If the distance between clusters is less than the threshold, they are merged, and each iteration forms a new cluster. The cluster is represented by three tuple $CF$ (1):

$$CF = < n, LS, SS >$$

(1)

$n$ is the number of initial clusters, $LS$ is the vector sum of all $n$ initial clusters $\sum_{i=1}^{N} \bar{X}_i$, and $SS$ is the sum of squares of $n$ initial clusters $\sum_{i=1}^{N} \bar{X}_i^2$. In a class cluster, the initial cluster can be expressed as $\{ \bar{X}_i, i=1,2,3,...,N \}$. The center points $\bar{X}_0$ between the initial clusters in the cluster are expressed as equation (2):
The radius distance in the cluster named $R$ is expressed by equation (3):

$$ R = \left( \frac{\sum_{i=1}^{N} (\bar{X}_i - \bar{X}_0)^2}{N} \right)^{\frac{1}{2}} \quad (3) $$

The average distance between clusters named $D$ can be used to represent the distance between different clusters as shown in equation (4):

$$ D = \left( \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} (X_i - X_j)^2}{N_1N_2} \right)^{\frac{1}{2}} \quad (4) $$

The clustering between clusters satisfies the linear relationship, as shown in equation (5):

$$ CF_i + CF_j = (N_i + N_j, LS_i + LS_j, SS_i + SS_j) \quad (5) $$

After the information features inside and between clusters are represented, the clustering feature tree is generated. The designer sets the important parameters of the clustering feature tree, the maximum number of clustering features $B$ of each initial cluster, the distance threshold $T$ of the initial cluster, and the maximum number of clustering features $l$ of the cluster. Designers can define the number of class clusters at the time of termination to ensure the granularity of business problem partition. At the beginning of the algorithm, the clustering feature tree is empty, and the sample points are read from the data set in turn. The distance between different initial clusters is calculated by equation (4) to determine whether it can be clustered. If it can be clustered, CF triples will be updated. If not, new clusters will be established. After several iterations, the final clustering feature tree will be formed.

### 3.2. DSM Matrix Representation Based on Coupling

DSM (design structure matrix) is called design structure matrix which represents the coupling relationship between elements in the form of matrix. DSM matrix is an $n$-dimensional square matrix. The elements of the system are arranged in the same order on the left side and above of the matrix. Because the diagonal is the relationship between the elements themselves, the diagonal itself does not need to express its coupling. If different elements $i$, $j$ have coupling, they can be expressed in the non-diagonal $i$ line $j$ column and $j$ line $i$ column. Element coupling degree is shown in figure 2:

|     | 1   | 2   | 3   | 4   | 5   | 6   |
|-----|-----|-----|-----|-----|-----|-----|
| 1   |     | 1   |     |     |     |     |
| 2   | 1   |     |     |     |     |     |
| 3   |     |     |     |     |     | 4   |
| 4   |     | 4   |     |     |     |     |
| 5   |     |     |     |     | 3   |     |
| 6   |     |     |     |     |     |     |

**Figure 2.** DSM matrix.

The DSM matrix and hierarchical clustering algorithm are used to deal with the bottom business problem partition, which improves the cohesion of business problem, reduces the coupling between business problem, and ensures that the business problem can be divided into specific topics to provide knowledge assistance.
4. Knowledge Push
In the previous chapter, cluster analysis and coupling representation of underlying business problems are carried out through DSM matrix and hierarchical clustering algorithm. In this chapter, the related topic text is mined and the text matching degree is calculated according to the business problems that have been divided. The text calculated by matching degree is structured and pushed to relevant business personnel.

The establishment of knowledge push framework is to effectively organize knowledge resources in the process of business personnel dealing with business problems. Due to the large number of semi-structured and unstructured text data in text resources on the network, it is often impossible for business personnel obtain target information from the text. Therefore, this article establishes a framework to represent the knowledge in multiple dimensions as shown in table 1.

| Knowledge Source | Knowledge Sources                  |
|------------------|-----------------------------------|
| Definition       | Knowledge Connotation             |
| Axiom            | Knowledge Principle               |
| Background       | Knowledge Background              |
| Effect           | Knowledge Effect                  |
| Application      | Knowledge Use Scenarios           |

The specific steps of knowledge push are as follows:
- Business personnel put forward specific knowledge requirements through business problem, which are vectorized.
- Clean the semi-structured and unstructured text data crawled by Python to remove all symbols (numbers, punctuation, letters).
- Jieba segmentation is performed on the cleaned text data, and stop words (such as “de”, “Ya”, “this”) are filtered to get standardized text.
- The standard text is extracted by word2vec algorithm and expressed by vector.
- Taking the vectorized knowledge demand as the condition, the semantic similarity between the text data and it is calculated, sorted according to the similarity, and the similarity threshold is set to get the knowledge matching result.
- The knowledge above the threshold is filled into the knowledge representation framework, represented and pushed to the business personnel.
- The business personnel judge whether they meet the knowledge requirements. If they meet the requirements, it ends. If they do not, they return to step 1 and ask the business personnel to update the knowledge requirements.

In the process of knowledge push, the key step is whether the knowledge push text meets the knowledge needs of business personnel and ensures the effectiveness of knowledge push. In this paper, the knowledge requirements of business personnel for business problem are submitted in the form of natural language text. The cosine similarity based on word vector is used to measure the similarity between two texts. If it exceeds the threshold, the effectiveness of knowledge push text is determined.

The user’s knowledge service requirement text is segmented into words, and then word2vec algorithm is used to represent it as a word vector to express the characteristics of the knowledge service requirement text from the actual semantic level. The vector represents the knowledge service demand text and knowledge push text, which can be calculated by cosine similarity, as shown in equation (6):

$$\cos(X, Y) = \frac{X \cdot Y}{\|X\| \|Y\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}}$$  (6)
The value range of $\cos(X,Y)$ is $[-1,1]$. The closer the cosine value is to 1, the more similar the text matching is. The threshold value is set for text matching. When the cosine value exceeds the threshold value, the text matching is proved to be passed, and the knowledge push text is provided to the business personnel for accurate knowledge assistance.

5. Conclusion
In order to solve the problem that the decision-making of traditional business execution lacks accurate knowledge push, this paper establishes a knowledge push model. Aiming at the business problem, hierarchical clustering algorithm is used to cluster the business problem, and the business problem after clustering are represented by coupling, which improves the cohesion of business problem and reduces the coupling between business problem. Then the knowledge of the divided business problem is pushed, and the text matching degree is calculated by cosine similarity. Compared with other traditional business push models, the knowledge push model in this paper has the advantages of high accuracy and low interference.

Acknowledgments
I would like to extend my gratitude to all those who have offered support in writing this thesis from National Key R&D Program of China (2019YFB1707101, 2019YFB1707103) and the Zhejiang Provincial Public Welfare Technology Application Research Project (LGG20E050010).

References
[1] Dang D P, Chen C X, Li H C, Yan R G, Guo Z X and Wang X J 2021 Deep knowledge-aware framework for web service recommendation The Journal of Supercomputing https://doi.org/10.1007/s11227-021-03832-2.
[2] Liu H T and Zhao W D 2017 Process-oriented knowledge recommendation by mining knowledge patterns Computer Integrated Manufacturing System (02) 396-403.
[3] Wang L K, Jiang Z H and Li X Y 2021 Topic diversity knowledge recommendation method in engineering filed Computer Integrated Manufacturing System 1-21.
[4] Wei J C 2018 An Investigation to Module Knowledge Mining and a Modular Design Methodology for Products (Northwestern Polytechnical University) PhD Dissertation pp 37-57.
[5] Yu X 2020 Research on Patent Knowledge Pushing Method Based on Deep Learning (Zhejiang University) Master’s Thesis pp 25-50.
[6] Xu Z G and Dang Y Z 2021 Solution knowledge mining and recommendation for quality problem-solving Computers & Industrial Engineering (159) 107313.
[7] Wang X, Liu X, Liu J, Chen X M and Wu H 2021 A novel knowledge graph embedding based API recommendation method for Mashup development World Wide Web: Internet and Web Information Systems 24 869-894.
[8] Wang T C, Hua Y and Wu Y 2020 Extension knowledge push model for automobile engine design Journal of South China University of Technology (Natural Science Edition) (02) 107-115.
[9] Wang Y Q, Dong L Y, Zhang H, Ma X T, Li Y L and Sun M H 2020 An enhanced multi-modal recommendation based on alternate training with knowledge graph representation IEEE Access 8 213012-213026.
[10] Cena F, Console L and Vernero F 2021 Logical foundations of knowledge-based recommender systems: A unifying spectrum of alternatives Information Sciences 546 60-73.
[11] Yang Z S and Cheng J Y 2021 Recommendation algorithm based on knowledge graph to propagate user preference International Journal of Computational Intelligence Systems 14 1564-1576.