RESEARCH AND APPLICATION OF DATA MINING TECHNOLOGY ON CONSTRUCTION PROJECT COST CONTROL SYSTEM

Ying ZHOU, Lie Yun DING
School of Civil Engineering, Huazhong Science and Technology Univ., Wuhan, Hubei, China.

Abstract: With the continuous development of the digital construction management, it is important to get the project cost information more and more precisely. This article applies Data mining technology in construction cost control system to solve the shortcomings in traditional management. It provides support in information integration among subsystems and builds cost control system based on web. The authors advance multi-dimension data model and build a custom-made applying scheme. It satisfies the managers with providing project’s cost information from all views and improves the share-out in overhead cost. Also, it represents the cost information more fully and turns from passive management to active and enhances the cost management efficiency.

Keywords: Construction Project; Cost Control system; Data Mining; Data Modeling

1 Introduction

In project management, usually compared to the schedule control, the cost control needs strengthening. The basic reason lies in the data collection and data processing are more complex in the cost analysis.

First, there are numerous participants in the construction project. Each participant manages respectively so that it is difficult to collect and track data. Meanwhile, construction enterprises lay particular emphasis on the enterprise cost, but weaker on the project cost. Daily management and project management have not been integrated. Also people do not build up an independent checking system to the single-item engineering, result in multiple project cost data mix together, and it is difficult to distinguish each project actual cost and contrast it to the budget cost. Project profit can be calculated only after work has been completed.

Second, the existing cost management mainly aims at the direct cost expenses; not enough valuing to the overhead cost management (Kim and Ballard, 2002). The overhead expense sharing rule is too simple to tell how much each item costs. Cost management mainly depends on accountancy system and it satisfies the financing but neglects the construction management, such as the rationality of the cost and the quantity cost. Such situation leads project managers can not acquire comprehensive cost information and difficult to make decisions.

The third, the inconformity project information breakdown structure make the information which comes from different management angle split into pieces (Hu 2005). Members in the same construction but from different departments always do the same work repeatedly. There are so many redundant data in the information system, and it is difficult to assure their veracity.

Aim at the above-mentioned problems, this article advance to use data mining technology on construction project cost management. It uses data dividing and data recomposing technologies to discover more information from the spot data. Also, it helps to build a coordinated working condition to realize the construction project integrated management and data sharing.

2 Data Mining Technology and Construction Management

2.1 Data mining technology

Construction project always involves many participants who come from different specialties. In the working period, it brings lots of data from different contractors and different management areas. In the enormous database, project managers can not tell what is important for them to make decisions. How to pick up the most useful information from this huge database? It is necessary to use data mining technology in construction management, but what is data mining?

Data mining, also popularly referred to as knowledge discovery in databases (KDD), is the automated or
convenient extraction of patterns representing knowledge implicitly stored in large databases, data warehouses, and other massive information repositories (Han and Kamber, 2001). Data mining does borrow from statistics, machine learning, databases, information retrieval, data visualization and other fields.

Construction field deals with a vast amount of data, which needs to be turned into something meaningful, knowledge. For example, workers in construction department which supervise the schedule record data in their own specialty but usually, the data is more meaningful than people think and is helpful to contract department. So what we need to do is building a multiple-angle data model to make data more precise and more useful to avoid missing values.

2.2 Data model

Traditional cost management object is single-dimension. People record data in their own views and data in the whole system can not be flowed fluently so that the value of data is always missed. It can not provide detailed cost information for different decision-makers. Data needs to be divided into several parts. For example, owners can select three dimensions including management fields, contractors and construction sects to analyze cost, see figure 1 below. Every data put in the system will be projected to three dimensions so that people can never lose any information in his special purpose.

![Figure 1 Three-dimensional Data Model](image1)

In actual construction process, people select different dimensions according to different demands. For instance, the management areas which are considered as one dimension in above data model can be divided into three dimensions --- cost management, schedule management and quality management if you want.

Sometimes, data is fragmented and complanate in the model. Also, different data may be combined into new datum. Figure 2 shows two data in the data model and figure 3 shows a new datum after combined.

![Figure 2 Data Dividing in Three-dimensional Data Model](image2)
However, when analysis dimensions are more than three, we cannot express the data in figure style but in table style. For discussion more conveniently, this article will only consider three-dimensional data model.

2.3 A case study of data mining

Let us see an example of how to discover the data connection from OLAP — a very important method in data mining. Suppose in a rail station construction, we select tasks whose support rates is greater than 0.5 which presents the probability of each task. When their support rate is no less than 0.4, it represents the selected tasks are more important than others. Of course, we can modify 0.4 to any other number. In this case, we select 4 tasks whose support rates are from 0.6 to 0.8 and unit price are from $50 to $200. Let us see more details in table 1 below.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Unit Price</th>
<th>Support Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task1</td>
<td>$100</td>
<td>0.8</td>
</tr>
<tr>
<td>Task2</td>
<td>$200</td>
<td>0.8</td>
</tr>
<tr>
<td>Task3</td>
<td>$120</td>
<td>0.6</td>
</tr>
<tr>
<td>Task4</td>
<td>$50</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Analyze two tasks together. Also, we omit the combination whose support rate is less than 0.4. For example, the probability of the concurrence of task1 and task2 is below 0.4. We can conclude that this combination is not important enough, so after the first process, it is omitted. See table 2 below.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Total Price</th>
<th>Support Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>{Task1,Task3}</td>
<td>$220</td>
<td>0.4</td>
</tr>
<tr>
<td>{Task2,Task3}</td>
<td>$320</td>
<td>0.6</td>
</tr>
<tr>
<td>{Task1,Task4}</td>
<td>$150</td>
<td>0.4</td>
</tr>
</tbody>
</table>

In table 2, it tells 6 rules:

Rule 1: Task1 $\rightarrow$ Task3, S = 0.4, C = 0.4/0.8 = 50%
Rule 2: Task3 $\rightarrow$ Task1, S = 0.4, C = 0.4/0.6 = 67%
Rule 3: Task2 $\rightarrow$ Task3, S = 0.6, C = 0.6/0.8 = 75%
Rule 4: Task3 $\rightarrow$ Task2, S = 0.6, C = 0.6/0.6 = 100%
Rule 5: Task1 → Task4, S = 0.4, C = 0.4/0.8 = 50%
Rule 6: Task4 → Task1, S = 0.4, C = 0.4/0.6 = 67%
S means Support rate and C means Confidence.

Rule 1 means the probability of task1 is 0.4 and when task1 finishes, task3 has 50% chances to begin. In Rule 4, we can see when task3 has begun, task2 will surely occur (the percentage of task2 occurs is 100%). Also, it tells when doing task3, we should prepare $320 not $120 to ensure the schedule of the construction because probably, task2 will take place soon.

This method can describe cost information using schedule information. Well worth mentioning, task1, task2, task3, and task4 are not related by working procedure. It foretells the cost curve by data mining and data analysis only.

Suppose the working procedure is defined as table 3 shows, we can get further conclusion.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Predecessors</th>
<th>Successors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task1</td>
<td></td>
<td>Task4, Task2</td>
</tr>
<tr>
<td>Task2</td>
<td>Task1</td>
<td></td>
</tr>
<tr>
<td>Task3</td>
<td>Task4</td>
<td></td>
</tr>
<tr>
<td>Task4</td>
<td>Task1</td>
<td>Task3(SS)</td>
</tr>
</tbody>
</table>

Table 3 Working procedure

When task4 occurs, its successor—task3 will begin soon. Also from the data analysis, we infer that task2 will probably occur (Rule 4), so we can conclude that when task4 begins, the money we have to prepare is not just $50, but $50 + $120 + $200 = $370. It tells us the financial situation may be not as good as we consider.

However, it is important to realize the reliability of the conclusion is mostly depending on mass data. If stylebook is not great enough, the conclusion may be misleading.

3 The Scheme Cost Control System Based On Data Mining

3.1 Storage structure

The cost control system based on data mining uses OLAP and data warehouse technologies. OLAP projects original data to different dimensions (analysis angle) to provide more telling information. The storage structure is shown in figure 4.

Figure 4 Storage Structure
3.2 System structure

This system includes data source, data marts, data warehouse, OLAP server and end tools. The system structure is shown in figure 5.

![Figure 5 System Structure](image)

The cost control system based on data mining regards construction project as a whole system by dint of data warehouse and management technologies. It divides mass original data to utilize their value to the project cost control. This system can integrate mass disperse data with different structure into one data warehouse. Then, end users can build kinds of query and analysis for themselves according to different goals by end tools.

3.3 Network Structure

Ahmad and Ahmed (2001) noted the construction industry is highly fragmented. The majority of construction projects involve several participants such as owners, contractors, subcontractors and supervisors and so on. Also, they are in different places. It is necessary to build the cost control system by network. The network structure is shown figure 6.

![Figure 6 System Network Structure](image)

4 Conclusions and Recommendations

Construction management pays attention to multiple goals such as fast schedule, low cost and good quality, not single one. Usually, managers pursue one target and meanwhile hamper other targets. Authors advance a new method which use data mining technology in construction management to break traditional MIS (Management Information System) structure. It provides custom-made style and discovers potential value of original data.

As global competitiveness increases, so will the expectation of higher level of construction digital management. Cost control system based on data mining technology provides multiple angles to observe cost
information.
However, this method requires several conditions:
· Practitioners need collect more data or they may get misleading conclusions.
· Managers should be aware of their targets more clearly and its relations to the data dimension defining.
· Information system should be upgraded to make maximum use of just in time techniques by providing instantaneous information to all involved parties (Forbes and Ahmed, 2003).
· Each member of the construction should share their data equally to ensure system reliability.

References

Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, Higher Education Press, 2001, 4
Yong-Woo Kim, Glenn Ballard, Case Study-Overhead Costs Analysis, Proceedings IGLC-10, Gramado, Brazil, Aug. 2002, 8