

Software Project Schedule Management Using Machine Learning & Data Mining

Wang Wei, Muhammad Ehsan Rana

Abstract— With the rapid advancement in science and technology, the development speed has surpassed peoples' anticipation. However with the advent of this increasingly mature technology, small and medium-sized companies are still utilizing management methods using traditional development project schedule management tools for software development. These conventional management software have limitations and cannot meet the needs of modern software requirements. Artificial intelligence is the current trend of technology development. Many industries in the society involve artificial intelligence technologies, such as manufacturing and service industries. This paper focuses on the use of artificial intelligence based techniques such as machine learning and deep data mining in software project schedule management to improve the efficiency and reduce the cost of software development to a significant extent.

Index Terms— Software Project Schedule Management, Bigdata, Machine Learning (ML), Data Mining, Artificial Intelligence

1 INTRODUCTION

WITH the continuous popularization and development of information technology and intelligent technology, software engineering has gradually become a pillar industry to promote the rapid development of the national economy. The rise of information technology companies around the world marks a new stage in the development of the software industry. The progress control in software project management is based on the scientific evaluation and effective control of manpower, resources, schedule, risk, etc., and the corresponding software development plan is formulated. The existing resources are rationally allocated to ensure that the minimum investment can create the largest Profits have become a top priority in the day-to-day management of software companies. For small and medium-sized enterprises, the traditional pure labor method is also used in the software project schedule management, which not only makes the management method limitation increase the labor cost of software development. This paper will propose the use of machine learning and data mining. A technical approach to assist developers in project progress management.

2 SOFTWARE PROJECT SCHEDULE MANAGEMENT

The progress is the work plan date table specified for the executed activities and milestones. the project schedule management is also called project time management and time management. It refers to the progress of the work in each stage and the final completion of the project during the project implementation process. The management of the term is to ensure the necessary management process for the project to be completed on time. project schedule management is one of the important measures to ensure the project is completed on time and rationally arrange the supply of resources and save engineering costs.

Project schedule management is to use the corresponding method to analyze the activities and interrelationships between the projects, estimate the time required for each subproject, and arrange the subprojects and the reasonable time within the time limit specified by the project. Control the start and end time of the subproject. PSM (Project schedule management) can be summarized as the following six main parts: Activity definition: The parent project is divided into multiple subprojects, and each subproject must have deliverables. Detailed definition of the specific activities of the subproject [1]. Activity sequencing: Determine the dependencies between each subproject and finally generate a document form. Activity resource estimating: Estimate the quantity and type of resources, equipment, and other resources required for each subproject [2]. Activity duration estimating: Estimate the time at which a single subproject is completed. Schedule development: The analysis is performed first and then the project schedule is developed for the sequence of activities, time and resource requirements [3]. Schedule control: Supervise the status of project activities, control changes to project schedules, and finally ensure that projects are completed within the required timeframe.

3 FACTORS AFFECTING THE PROGRESS OF SOFTWARE DEVELOPMENT PROJECTS

Efficiently control the progress, and then analyze the factors that affect the progress one by one and take necessary precautions through pre-judgment. Ensure that the actual progress is minimally offset from the planned progress and proactively control the project. There are many factors that influence software project development, technical factors, environmental factors, funding factors, human factors, and so on. The human factor is the most important factor in the implementation of software development projects, and the technical factors are in other words human factors. The main problem of software development project schedule control management is still reflected in the consideration of various factors. There are generally the following situations:

3.1 80-20 PRINCIPLE AND OVERLY OPTIMISTIC SCHEDULE CONTROL

The 80-20 principle means that 80% of the work in software project schedule management can be completed in 20% of the time, while the remaining 20% of the work takes 80% of the time. 80% of the work that can be completed in 20% of the

- Afreen Iqbal, Asia Pacific University of Technology & Innovation, Kuala Lumpur, Malaysia. E-mail: afreenvoh@gmail.com
- Muhammad Ehsan Rana, Asia Pacific University of Technology & Innovation, Kuala Lumpur, Malaysia. E-mail: muhd_ehsanrana@apu.edu.my

time is not necessarily in the early stages of the project, they may be distributed at all stages of the project, but usually the remaining 20% of the work is generally distributed in the later stages of the project [4]. Therefore, software development will give people a feeling of "fast progress" after entering the coding stage, which makes project managers, project team members, users and senior leaders have an overly optimistic estimate. Some leaders saw that the software was delivered to the user, and they landed on a stone "together," and at the same time they might withdraw some human resources that were deemed unnecessary. However, in many cases, this is a last resort measure to deal with unreasonable delivery deadlines. The result is a delay in the post work, and if the software is not mature, it will have a bad impact on the user [5].

3.1 IMPACT OF SCOPE AND QUALITY FACTORS ON SCHEDULE

The development of software projects usually requires frequent changes. This is probably because users think that software changes only require programmers to change the code, and the cost of modification is not large. But this is not the case. Changes in the user's thoughts result in a change in requirements, and the project manager is unable to reject the user's changes for a variety of reasons. This will invisibly affect the progress of the project. If the quality of a sub-project completed is not as expected, for example, there are often loopholes, maintenance difficulties, and so on. This sub-project must be rewritten, which wastes a lot of time by wasting human resources and other inputs. On the other hand, it also delays the progress of the project. Therefore, the quality of the subproject affects the progress of the entire project, and the subproject with higher priority will affect the priority to the subproject. [6].

3.2 IMPACT OF RESOURCE AND BUDGET CHANGES ON SCHEDULE

The resources here refer to human resources. In some sub-projects, there may be insufficient team members, or members of individual sub-projects in the whole project will be called into other sub-projects or one person is responsible for multiple sub-projects at the same time. project. Another resource refers to information resources. The legal standards of each country are different. The income of citizens in each city is different. The standards of each industry are not uniform. These information resources are rarely provided by customers. Team members collect, if the information resources are not available on time, it will also affect the project's needs analysis, overall design and programming work. In addition, other resources refer to development equipment and development environment software. These resources will not affect the progress of the entire project in time. The budget can also be said to be a resource. The amount of budget that affects other resources will ultimately affect the overall progress. For example, a high-budget development environment that uses a high-performance environment will speed up the team's completion of the project and vice versa. [7].

3.3 UNDERESTIMATING THE CONDITIONS FOR THE IMPLEMENTATION OF SOFTWARE DEVELOPMENT PROJECTS

Underestimating the technical difficulty, coordination complexity, and environmental factors of software development projects can be attributed to underestimating the conditions for software development projects. Underestimating the technical difficulty of software development means that

software development project members or senior project managers often underestimate the technical difficulties of the project. Their ideas are generally very optimistic, that is, they overestimate the ability of themselves or team members. They are optimistic that the project will be carried out according to a customized plan, but this is not the case. Software development is a high-tech project in the project. There are bound to be many technical problems, so the team needs high-tech personnel. In addition, project experiments need to be considered in consideration of post-maintenance. For example, a website initially has too few visitors, but the number of people who visit for a long time will increase. Developers need to think about it in the future. Underestimating the complexity of the project coordination, underestimating the difficulties faced by the entire project team coordination. Members of the software development project team not only need to have higher personal abilities but also need to have a higher sense of team, which is like an athlete on the football field. This can greatly reduce the coordination complexity of the entire project. When a large project has many sub-projects, if there is insufficient communication between the sub-projects, this will increase the difficulty of project coordination and schedule control. Environmental factors are often underestimated by team leaders or project managers, social environment, industry environment, economic environment, user environment and so on. Underestimating these environmental factors can also be said that the collection of information is not comprehensive enough, and the entire project is still not fully prepared. [8].

3.4 PROJECT STATUS INFORMATION COLLECTION

Due to the experience or quality of the project manager, the mastery of project status information collection is insufficient, and the accuracy of timeliness accuracy is relatively poor. There are other reasons that can cause this phenomenon. Some project team members do not report their worries and do not want others to know that their work is not good. For example, the preparation of software programs may first create some superficial things. The existing interface seems to complete the task. Actually, Just a "prototype system" or demo system. Give leaders a sense of optimism. If the project manager or the management team does not check the situation in time, it will have a serious impact on the progress of the project. Of course, if there is such an atmosphere that needs to be guarded against each other at all times, managers should review it from a management perspective and from an institutional perspective and make improvements so that everyone can communicate realistically [9].

4 MACHINE LEARNING

Machine learning involves many disciplines, such as probability, statistics, algorithmic complexity theory, and so on. Machine learning is devoted to how to make computers simulate human learning behavior. Used to help the machine acquire new information knowledge, which makes the machine more perfect. Just like the artificial intelligence, face recognition, and driverless cars that we often mention now, they use machine learning algorithm technology. The ability of the learned function to apply to a new sample is called generalization capability [10].

4.1. General Steps in Applying Machine Learning

Select a suitable model, which usually depends on the actual problem. For different problems and tasks, you need to choose the appropriate model. The model is a collection of functions. To judge the quality of a function, it is necessary to determine a metric, which is the Loss Function we usually say. The determination of the loss function also depends on the specific problem. For example, the regression problem generally adopts the Euclidean distance. Classification problems generally use a cross entropy cost function. Find out the "best" function, how to find the "best" one from among the many functions. This step is the biggest difficulty. It is often not an easy task to do it quickly and accurately. Commonly used methods are gradient descent algorithm, least squares method, and other tricks [5]. After learning the "best" function, you need to test it on a new sample, and only perform well on the new sample to be a "good" function.

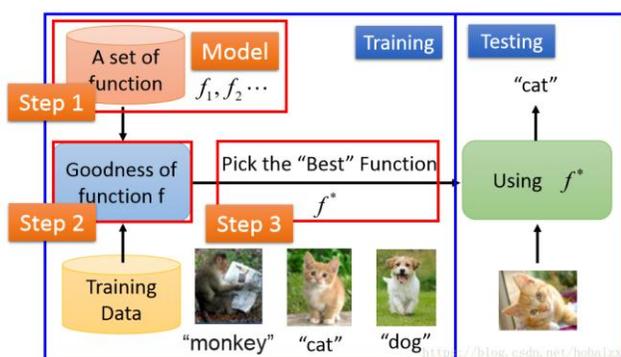


Figure 1: General Steps in Applying Machine Learning [11]

4.2. Use of Machine Learning to Improve Software Project Schedule Management

The main challenge facing the project manager is to accomplish the project's goals under the given constraints. In general, estimating the precise timeframe for project completion may be one of the toughest parts of their work. Using ML can take software project schedule management to a whole new level. There is many project management software, such as ClickUp, which are testing machine learning to predict what users might do [12]. The project schedule management used by ML technology should have the following capabilities: predict and assign tasks to the rightful team members [13]. Automatically tag users in comments that are relevant to them. Visualize notifications and updates based on their relevancy to a particular user. Predict and determine when deadlines aren't going to be met. Correct task time estimates.

5 DATA MINING

Data mining is the processing of many uncertain, ambiguous, random data, from the analysis and integration of data to find useful information and valuable data. A sophisticated data mining system contains a lot of content and must have all the information and equipment to process the information.

5.1. Data Mining Steps and Methods

In doing data mining, data should be organized and processed. There are four main methods of data mining, namely genetic algorithm, rough set method, decision tree

method and neural network method. When you have big data, you go through a series of steps and processes. The first is to judge the data problem, and then need to use a certain standard to measure, then the data is extracted and integrated, and then the data residue is removed, and this final data is your result [14].

5.2 Use of Data Mining in Software Schedule Management

The entire process of software project schedule management will be interspersed into the various departments and management processes of the software company and involve different people and technologies. Therefore, the steps to use data mining techniques in software project schedule management should include the following: First, integrate the information on each stage, each department, and the technical and personnel involved in the project, and form a data warehouse. Secondly, using data mining techniques to analyze the formed data warehouse, mining potential valuable data information, and obtaining the internal relationship between them, classifying and integrating, forming models and systems, and achieving high knowledge sharing. Finally, it is implemented by a software system. The main application process of data mining technology in software engineering management includes using effective classification, clustering and statistical analysis and prediction techniques to discover potential information, internal connections, etc. from various data warehouses, and form feedback information to guide the software. The project corresponds to the actions of the stage to achieve the purpose of improving development efficiency.

6 MACHINE LEARNING TECHNIQUES TO IMPROVE SOFTWARE PROJECT SCHEDULE MANAGEMENT

There are many kinds of learning algorithms in the field of machine learning. In order to effectively improve the efficiency of software project schedule management, we use two types of learning algorithms: instance-based learning algorithms and regression models.

6.1. Instance-Based Learning Algorithms

Instance-based learning is modeled by samples or cases of training data that are also considered necessary for modeling. This type of model usually builds a sample database, compares the new data with the data in the database, and finds the best match and makes predictions in this way. In other words, when making predictions, such algorithms generally use a similarity criterion, which is more predictive than the similarity between the predicted sample and the original sample. Therefore, the instance-based approach is also known as the winner-take-all and memory-based learning [15]. This should be the simplest algorithm in machine learning algorithms. It is not like other algorithms need to establish general inference formulas based on samples, but directly through the stored data sets for classification or regression learning to get results [16]. Its disadvantage is that the storage space is very demanding. The space required depends directly on the number of instances, and it needs to be compared with known instances when forecasting, and the running time will be relatively slow. Common example-based learning algorithms include: K-proximity algorithm (KNN) Learning Vector Quantization Algorithm (LVQ) Self-Organizing Mapping Algorithm (SOM) Local Weighted Learning Algorithm (LWL)

6.2. Regression Models

Regression analysis is a predictive model technique for studying the relationship between independent variables and dependent variables [17]. These techniques are applied to predict the time series model and find the relationship between variables. Regression analysis is also a commonly used statistical method that is incorporated into the field of machine learning through statistical machine learning. "Regression" can refer to both an algorithm and a problem, so it is easy to confuse when referring to it. In fact, return is a process [18]. Commonly used regression algorithms include: Ordinary Least Squares Regression (OLSR) Linear regression Step by step Multivariate Adaptive Regression Spline Method (MARS) Locally estimated smooth scatter plot (LOESS) Logistic regression (It is a classification algorithm)

7 MICROSOFT PROJECT SERVER: AN EXISTING PROJECT SCHEDULE MANAGEMENT SOLUTION

There are numerous project management software present in the market that provide convenient functions for project managers. One of the commonly used project management server solution is Microsoft Project Server. Its main strengths are explained below: Supports real-time access to Office Project Server 2007, direct access to critical business project information via the Web, shared work results and collaborative tasks, and flexibility, making it easy to modify and update tasks. [19]. Office Project Server is handy in assessing requirements, allowing customers to effectively deploy resources and create future plans for the organization. The right people can be placed in the best location to complete the project, and the project manager can track utilization, availability and work schedules through a timeline. Microsoft Project Server uses Project Web Access to launch the application. By carefully examining project data using suitable optimization techniques, it is much easier for project managers to comprehend organization's tasks carried out in each phase of project life cycle for better decision making [20].

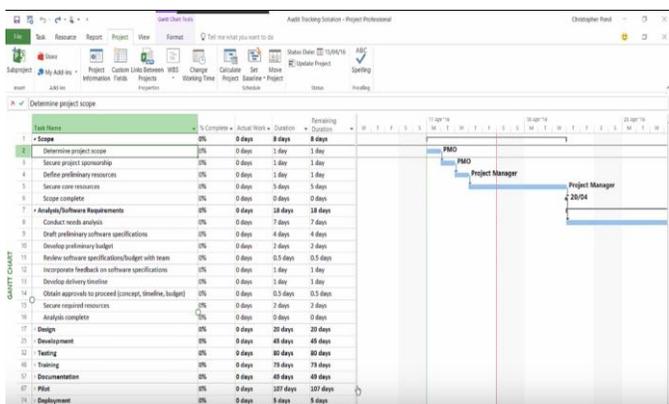


Figure 2: Microsoft Project Server Interface

The following are the shortcomings of Microsoft Project Server: Microsoft Project Server is a general-purpose project management software, it lacks industry characteristics. There are certain flaws in the estimation and measurement of IT projects, and the difficulty and limitations are greater when integrating with other software engineering tools. It only provides managers with administrative convenience, but it is not enough to attract software company decision makers to

choose this software, because its function does not bring excellent changes to software project schedule management. Microsoft Project Server, being built using the conventional technologies, do not use automated features provided by Machine learning and data mining to improve accuracy of schedule estimates and minimize the error factors. Table 1 below summarizes the strengths and weaknesses of Microsoft Project Server:

Table 1 : Strengths and Weaknesses of Microsoft Project Server

Strengths	Weaknesses
Easily collaborate and coordinate projects	Lack of industry characteristics and low integration with software engineering tools
Effectively manage resources	Only provides management features and no recommendations
Able to manage the entire project life cycle	Do not use automated features provided by Machine learning and data mining

8 OVERALL RECOMMENDATIONS

In order to improve the efficiency of small and medium-sized software companies in software project schedule management, it is a good choice to apply advanced machine learning models and data mining techniques to software project schedule management. This will not only conform to the trend of technological development, but will also improve production efficiency. The schedule management software incorporating these two technologies will not only improve the accuracy of project schedule estimates, minimize the error factors in the project, but will also improve in project success rate and customer satisfaction. With the advancement of computer technology, it is believed that in the near future, software project schedule management will be completed by machines instead of humans.

9 CONCLUSION

The current project schedule software are mostly management centric, hence provides focus on the management functions only. By utilizing an intelligent tool, appropriate advice can be advocated to the users which can greatly reduce the difficulty of managers in decision-making and will also allow the software being built to occupy more markets. At present, small and medium-sized companies still have many defects in the schedule management of software projects. It is difficult to deliver complete projects to customers within the specified time using traditional project schedule management methods. This paper first summarizes the problems encountered in the current software project schedule management, and then proposes the use of machine learning and data mining techniques to the software project schedule management. Finally, the advantages and disadvantages of current software project schedule management software are exemplified. Using machine learning technology and data mining technology into software project management can effectively advise managers, so that managers can reasonably plan the software development cycle, and finally realize the delivery of complete projects to customers within the specified time.

REFERENCES

- [1] A. Ahmed, *Software Project Management A Process-Driven Approach*, New York: Auerbach Publications, 2016.
- [2] T. L. Young, *Successful Project Management*, Kogan Page Publishers, 2016.
- [3] F. G. K. J. O. Howard Lei, *A statistical analysis of the effects of Scrum and Kanban on software development projects*, Elsevier, 2017.
- [4] J. E. G. Yucheng Low, "GraphLab: A New Framework For Parallel Machine Learning," 2014.
- [5] D. B. C. E. G. J. H. Aapo Kyrola, "GraphLab: A New Framework For Parallel Machine Learning," 2014.
- [6] D. Hulett, *Integrated Cost-Schedule Risk Analysis*, London, 2016.
- [7] P. N. R. Mathieu Lavallée, "Why good developers write bad code: an observational case study of the impacts of organizational factors on software quality," in *ICSE '15 Proceedings of the 37th International Conference on Software Engineering*, 2015.
- [8] S. Harwood, *ERP: The Implementation Cycle*, 2017: Routledge, 2017.
- [9] T. Dingsøy, "Exploring software development at the very large-scale: a revelatory case study and research agenda for agile method adaptation," *Empirical Software Engineering*, 2017.
- [10] T. M. M. M. I. Jordan, "Machine learning: Trends, perspectives, and prospects," *Science*, pp. 25-30, 2015.
- [11] zxhohai, "CSDN.net," 5 6 2018. [Online]. Available: <https://blog.csdn.net/hohaizx/article/details/80584307>. [Accessed 15 1 2019].
- [12] S. M. J. Riana Steyn, "The Use of a Learning Management System to Facilitate Student-Driven Content Design: An Experiment," in *International Symposium on Emerging Technologies for Education, Emerging Technologies for Education*, 2017, pp. 75-94.
- [13] E. F. M. A. H. Ian H. Witten, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2016.
- [14] Asare-Frempong, J. and Jayabalan, M., 2017, September. Predicting customer response to bank direct telemarketing campaign. In *2017 International Conference on Engineering Technology and Technopreneurship (ICE2T)* (pp. 1-4). IEEE.
- [15] J. P.-R. G.-P. Aida de Haro-García, "Combining three strategies for evolutionary instance selection for instance-based learning," Elsevier, 2018.
- [16] C. Gong, T. Liu, Y. Tang, J. Yang, J. Yang and D. Tao, "A Regularization Approach for Instance-Based Superset Label Learning," IEEE, 2017.
- [17] E. G. W. M. S. William S. Cleveland, "Local Regression Models," in *Statistical Models in S*, New York, Routledge, 2017, p. 68.
- [18] J. Fox, *Applied Regression Analysis and Generalized Linear Models*, SAGE, 2015.
- [19] L. T. S. Jana Kostalova, "upport of Project Management Methods by Project Management Information System," Elsevier, 2015.
- [20] J. Stark, "Decision Engineering," in *Product Lifecycle Management*, Springer, Cham, 2015, pp. 1-29.
- [21] B. Wei, "Talking about the application and development of statistics," *Mathematical Statistics and Management*, pp. 85-97, 2015.
- [22] Y. B. G. H. Yann LeCun, "Deep learning," 2015.
- [23] J. Wang, "Research on the Application of Data Mining Technology in the Big Data Era in College Ideological and Political Work," *Media age*, pp. 23-30, 2015.