Simulation Model for Project Duration Estimation with Parameters Extraction Method from Accumulated Data

Kazuo HIEKATA[a,1] , Rujia WANG[a] and Taiga MITSUYUKI[b]
[a] *Graduate School of Frontier Sciences, The University of Tokyo, Japan*
[b] *Faculty of Engineering, Yokohama National University, Japan*

Abstract. A framework combining project simulation model with extracted uncertainty parameters from past data is proposed in this paper. In the existing simulation models which take uncertainty elements into consideration, a large number of input parameters are requested. However, most of these parameters are difficult or time-consuming to state. The aim of this paper is to provide a possible solution proposing a new simulation model based on extracted parameters and setting up a protocol to extract uncertainty parameters from past log data. The output of the simulation model will be the project duration, giving feedback to the design of human resource. The proposed protocol includes a definition of necessary past data and how to calculate minimum work amount, delay probability and rework probability. On the other hand, the proposed simulation model includes basic model describing project structures such as task dependency and resource skills, delay model describing variation of task work amount, and rework model describing transition among different tasks. Besides, in the case study, we test the program of the proposed framework constructed of the parameter extracting protocol and the simulation model. After that, we apply the framework on a project introduced in existing research. Two human resource strategies are considered of, on the basis of the duration estimation results. The conclusion of this study is that the proposed framework is able to conduct duration estimation and support the decision-making process around human resource at the early stage of a project.

Keywords. Duration Estimation, Simulation, Uncertainty, Parameter Extraction

Introduction

In large-scale projects which generally exist nowadays, how to design a workflow and allocate human resources properly is an important and complicated task even for the most experienced managers. Since the project itself is a complex system, small changes may lead to considerable influence on project result. Therefore, there is always high risk in the execution of large projects. In such complex systems like large projects, one of the critical reasons of the complexity is generally called as "uncertainty". As the premise of further discussion, a definition of uncertainty is usually given. For example, Earl et al. [1] defines project uncertainty into four categories, including known uncertainties, unknown uncertainties, uncertainties in the data and uncertainties in the description, while Browning et al. [2] lists uncertainty in project development as intentional iteration,

---

[1] Corresponding Author, Email: hiekata@edu.k.u-tokyo.ac.jp.

unintentional iterations, activity set completeness, activity flexibility, etc. The application of simulation plays an important role in managing uncertainty. In order to simulate the projects that have complex structures, various kinds of input information are usually necessary and the information about uncertainty elements is especially difficult to obtain. In most situations, uncertainty parameters are from experts' personal experience. To reduce the subjective part when modelling uncertainty, driver analysis as well as questionnaire methods have been proposed. However, existing methods often require much human work and complex analysis processes, the normal application in real projects is limited.

Considering the difficulty in decision making process when managing projects and the limitation of the existing methods, the aim of this paper is to propose a duration prediction method based on discrete-event simulation while obtaining input of uncertainty elements from past data.

1. Methodology

1.1. Overview

Figure 1 shows the overview of the proposed method. First, as the input for the whole system, resource information, the tasks that make up a project, dependency between the tasks and past data are demanded, among which the resource information is the thing under design, and the others are unchanging regarding the nature of a project. Second, a simulation model and the parameter extraction method are proposed. The simulation model is constructed of basic model, delay model and rework model, and the parameters for delay model and rework model are supposed to be extracted from the past data.
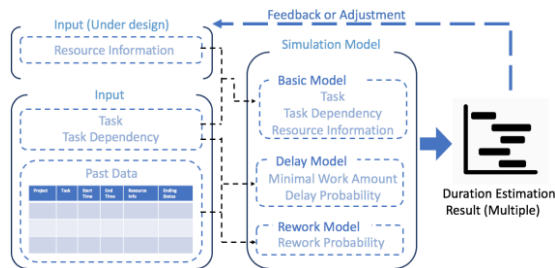

Figure 1. Overview of the proposed method

1.2. Simulation model

Tasks, task dependency and resource information are included in the basic model, while delay model includes minimal work amount as well as delay probability, and rework model consists of a series of rework probability. Following is the detailed description the simulation model.

1.2.1 Assumption regarding rework

The assumption regarding rework states the foundation for the parameters in delay and rework model. In this assumption, each time a task is executed, the occurrence time of the task is counted. The rule of counting the occurrence time using a simple example is

shown in Figure 2. On the basis of this assumption, parameters of a task, except for task dependency, are defined for each occurrence time.
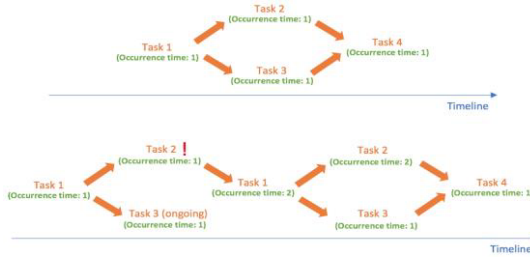

Figure 2. Assumption regarding rework

1.2.2 Basic Model

The concept of basic model is similar to the previous research conducted by Mitsuyuki et al. [3]. A project is considered as an aggregate of multiple tasks releated with each other and resource information is also defined. The task dependency is abstracted as finish-to-start only. Resources are grouped as multiple resource teams, having the abilities to execute certain tasks with different progress.

1.2.3 Delay Model

Delay model defines the duration of tasks using two kinds of parameters, minimal work amount and delay probability, shown in Equation (1) and (2). Equation (1) defines the minimal work amount for task i at occurrence time o, while Equation (2) defines the delay probability for task i to have a delay which additional work amount is *wa* at the end of occurrence time *o*. The judgement of delay will be carried out once a task reach its minimal work amount.

$$mwa_{i,o} > 0, i = 1,2,\dots,N, o \in \mathbb{N} \tag{1}$$

$$dp_{i,o,wa} \in (0,1], i = 1,2,\dots,N, o = 1,2,\dots,O, wa \in \mathbb{N} \tag{2}$$

1.2.4 Rework Model

A series of rework probability, defined in Equation (3), makes up the rework model for a task at certain occurrence time. In Equation (3), a rework probability is the probability that a failure happens in task i_1 and rework is necessary from task i_2 when the progress of task i_1 is equal to *pro*. Task progress is defined in Equation (4), indicating the quotient of actual work amount on simulation (*awa*) and minimal work amount. If rework happens, the progress of task i_1 and task i_2 will return to 0, and occurrence time and remaining work amount will be updated according to the assumption regarding rework.

$$rp_{i_1,oc,pro,i_2} > 0, i_1 = 1,2,\dots,N, i_2 = 1,2,\dots,i_1 - 1, i_1 + 1, \dots, N, o \in \mathbb{N} \tag{3}$$

$$pro_i(t) = floor(10 \times awa_i(t)/mwa_{i,oc})/10 \qquad (4)$$

1.2.5 Simulation Flow

A flowchart of simulation is shown in Figure 3. Except for the step 6~9, the simulation flow is referred to the previous research by Mitsuyuki et al. [3]. The details of each steps are described below.
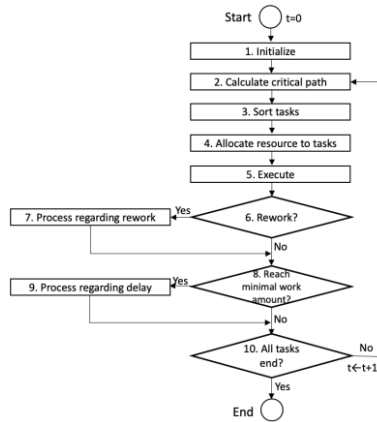
Figure 3. Simulation flowchart

1. Reading in the input, initializing all the elements and setting time to 0.
2. Calculating critical path, getting latest start time and earliest start time.
3. Sorting the tasks that satisfies task dependency and priorify a task having smaller difference between its latest start time and earliest start time.
4. Finding available resources that are not working at the current time, and allocate the resources to the sorted tasks.
5. Executing working tasks by subtracting the remaining work amount and adding the actual work amount with the summation of allocated resource ability.
6. Calculating task progress and doing the judgement of rework.
7. Processing rework as described in rework model.
8. For all the tasks whose progress reach 1, doing the judgement of delay.
9. Processing delay as described in delay model.
10. If all the tasks end, finishing simulation with an output of gantt chart and total duration. If not, time will be updated and keep looping from step 2.

1.3. Parameter Extraction

Parameters for delay and rework model are to extracted from a certain form of past data. This chapter will start from the definition of past data and introduce the extraction procedures in detail.

1.3.1 Data definition

Table 1 shows an example for formulated data necessary for the proposed extraction method. Project name, task name, start date/time, end date/time, resource ability when the task was executed and status information are demanded. Among these items, resource

capacity is for calculating a task's work amount and the status information is necessary for judge if a tasks is interrupted due to rework or not. 3 kinds of status information are listed as "Normal", "Rework" (along with the previous task that rework started) and "Suspend", which are supposed to be concluded before inputted into the extraction program.

Table 1. Data record template with sample data.

Project	Task	Start	End	Resource	Status
project 1	task 2	2018/01/01:0900	2018/01/03:1700	3 work amount/h	Rework to task 1

1.3.2 Preprocessing

Preprocessing deals with the conflicts in data caused by recording mistakes, simplifies complex task dependency in real project. Two steps of preprocessing is included:

1. Preprocessing regarding task dependency: pairwisely compare the data, if a downstream task with dependency relationship started before the upstream tasks end, the end time of upstream tasks will be considered as the start time of the downstream task to keep pre defined finish-to-start relationship.
2. Preprocessing regarding status information: work amount and iteration of each task is extracted in this process. If the status information of a data record is "Suspend", this recorda will be integrated to the next one with the same project and task name.

1.3.3 Parameter extraction for delay model

At the procedure of extracting parameters for delay model, the minimal work amount is to be counted at first. For each iteration of each task, the program searches the data record with the minimal work amount, whose status is "Normal". After that, additional work amount will be calculated by subtracting the minimal work amount from recorded work amount, and the additional work amount will be rounded down. Finally, delay probabilities will be calculated by dividing the number of delayed log by the number of all logs, for each occurrence, task, and additional work amount.

1.3.4 Parameter extraction for rework model

For each data record, the mount of progress is calculated. Basically, the progress is calculated following Equation (4) by substitute actual work amount by the work amount in data. The progress is calculated for any data record whose status is "Rework", no matter the progress is over 1 or not. Since the status "Rework" is stated to be recorded along with the name of a previous task, rework probabilities are calculated, dividing the number of data whose status is "Rework" at certain progress by the number of data whose progress is more than the progress.

2. Case study

Two case studies are conducted in this chapter. Case study 1 is for verification and validation of the proposed parameter extraction method and simulation model using a

virtual project, while case study 2 is a demonstration case showing the usability of the system in decision making process.

2.1. Case study 1: verification and validation

A small project is adopted in case study 1. A simulator[2] generates project data and we consider the data as "past data". Our proposed method calculates and sets the model parameters based on "past data." In this case, the the accurate model parameters can be identified by analytical approach and the both results for model paramters are compared to confirm the consistency.

2.1.1 Project background and past data

The project of case study 1 has 3 tasks and the task dependency shown in Figure 4, while the resource information recorded in past data is also shown in the same figure. As for the input data, Figure 5 is considered as the accumlation of "past data" in the gantt chart format. The reults includes rework.
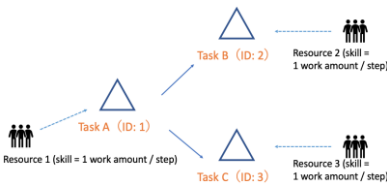

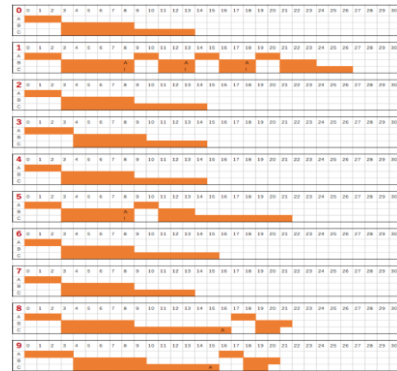**Figure 4.** Tasks and resources in case study 1.

Figure 5. Generated past data in case study 1.

2.1.2 Verification and Validation

As mentioned above, we compared the model paramters extracted from a set of past data and by analytical calculations. The parameters by analytical calculations shown in Table 2 are considerd as the reference values. To verify the parameter extraction method from past data, we generates 20 million runs of project simulation results and check the consistency with the analytical results.

Table 2. Analytical parameter extraction result.

Task	Delay model	Rework model
A	$mwa_{A,1}=3$ $mwa_{A,2}=2$ $mwa_{A,3}=2$ $mwa_{A,4}=2$ $dp_{A,1,1}=0.2$	-
B	$mwa_{B,1}=6$ $mwa_{B,2}=3$ $mwa_{B,4}=3$	$rp_{B,1,1,A}=0.2$ $rp_{B,2,1,A}=0.25$ $rp_{B,3,-1,A}=1$
C	$mwa_{C,1}=11$ $mwa_{C,2}=2$ $dp_{C,1,1}=0.3$ $dp_{C,1,2}=0.1$ $dp_{C,1,3}=0.1$ $dp_{C,1,6}=0.1$ $dp_{C,1,7}=0.1$	$rp_{C,1,1,A}=0.1$ $rp_{C,1,1,2,A}=0.333$

The graphs in Figure 6 show the diffierence between analytical probability and the frequency in simulation results for different duration output after 20 million simulations in a histogram, and the analytical probability for each duration is also shown in the same graph as a line chart. It is obvious that after 20 milion simulations, all of the differences for all durations converge to less than 0.0001, which means the results of simulation model converge to the analytical results.
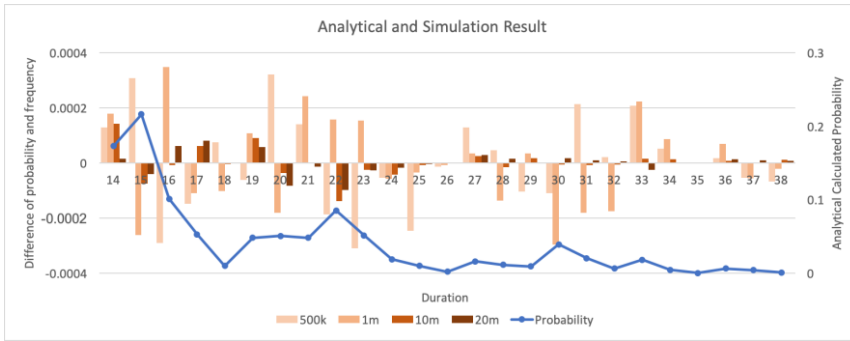

Figure 6. Comparison between analytical and simulation results.

Whereas, through looking into the simulation result, various combinations among rework and delay appear, providing emergence for unexperienced patterns of uncertainty events. As all the rework and delay probabilities are simulated in the proposed model, and there is no unexceptable work amount reduction, it is fair to use the proposed model to simulate and predict the project duration.

2.2. Case study 2: demonstration

To demonstrate how the proposed method contribute to the practical project management, a project case[2] is adopted. Similar as case study 1, since there is no accumulated past data, the generated data using the existing simulation model is used. 2 strategies for human resource assignment are evaluated by the proposed method. By comparing the simulation results for the 2 strategies, the project manager will be able to choose a more relevant strategy for the project.

2.2.1 Project background and past data

As mentioned above, the project is taken from existing literature, having 14 tasks following a standard design workflow. Detailed task information can be found in [2].

Table 3. DSM of the project in case study 2.

Task ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	-	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	-	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	-	0	0	0	0	0	0	0	0	0	0	0
4	1	0	1	-	0	0	0	0	0	0	0	0	0	0
5	1	0	1	0	-	0	0	0	0	0	0	0	0	0
6	1	0	0	0	1	-	0	0	0	0	0	0	0	0
7	1	0	0	0	0	1	-	0	0	0	0	0	0	0
8	0	0	0	0	0	1	0	-	0	0	0	0	0	0

Task ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
9	1	0	1	1	0	0	0	1	-	0	0	0	0	0
10	0	0	0	1	0	1	1	1	0	-	0	0	0	0
11	0	0	0	0	0	1	1	1	0	1	-	0	0	0
12	1	0	0	0	0	1	1	0	0	1	1	-	0	0
13	1	0	0	0	1	0	0	0	0	0	0	1	-	0
14	1	1	1	1	1	1	1	1	1	1	1	1	1	-

Table 4. Resource information in past data in case study 2.

Task ID	Resource information (person)	Resource capacity (work amount / day)
1	Intergrate product development team and additional experts	15
2	Configuration designer (2)	2
3	Configuration designer (1), Structure designer (1)	2
4	Aerodynamic engineer (1)	1
5	Structure engineer (6), stress engineer (6)	12
6	Structure engineer (1), stress engineer (1)	2
7	Structure analyst (4)	4
8	Weight engineer (2)	2
9	Control engineer (1), Configuration engineer (1)	2
10	Aerodynamic engineer (1), FEM engineer (1), General load engineer (1)	3
11	FEM engineer (2)	2
12	Strength enginner (4)	4
13	Manufacuture engineer (4), Procurement people (1), Manufacture expert (8)	13
14	Intergrate product development team	10

The 14 tasks have certain interdependencies based on the information exchange among them. The task dependencies are shown in Table 3 as a Dependency Structure Matrix. Using the simulation model proposed in the same literature, past data from 10 projects are generated and recorded as gantt charts in Figure 7. It is assumed that the past projects have adopted the same human resource strategy for each tasks, listed in Table 4.
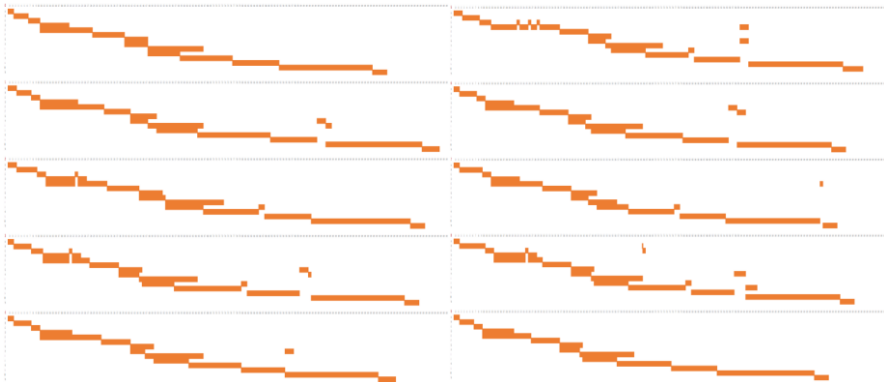
Figure 7. Generated past data in case study 2.

2.2.2 Resource strategy

Such situation is presumed before deciding the concrete resource strategies: In future projects, external organizations will execute the tasks by outsourcing contract. There are external organizations that can perform multiple tasks, and there are also external organizations that can only perform specialized single tasks. Given such situation, the resource strategies to consider are:

1. Strategy 1: Organizie many resource teams with a single skill.
2. Stretegy 2: Organize few resource teams with multiple skills.

To compare the strategies, certain restrictions are imposed on the number of the human resources. The resource information for simulation input Table 5 and Table 6.

Table 5. Resource information for resource strategy 1.

Task ID \ Team ID	1	2	3	4	5	6	7	8	9	10	11	12
1	15	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0	0
3	0	0	2	0	0	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0	0	0
5	0	0	0	0	12	0	0	0	0	0	0	0
6	0	0	0	0	12	0	0	0	0	0	0	0
7	0	0	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	2	0	0	0	0
10	0	0	0	0	0	0	0	0	3	0	0	0
11	0	0	0	0	0	0	0	0	0	1	0	0
12	0	0	0	0	0	0	0	0	0	0	1	0
13	0	0	0	0	0	0	0	0	0	0	0	13
14	15	0	0	0	0	0	0	0	0	0	0	0

Table 6. Resource information for resource strategy 2.

Task ID \ Team ID	1	5	12	13	14	15
1	15	0	0	0	0	0
2	0	0	0	1	0	0
3	0	0	0	2	0	0
4	0	0	0	1	0	0
5	0	12	0	0	0	0
6	0	12	0	0	0	0
7	0	0	0	0	1	0
8	0	0	0	0	1	0
9	0	0	0	0	0	2
10	0	0	0	0	0	4
11	0	0	0	0	0	2
12	0	0	0	0	1	0
13	0	0	13	0	0	0
14	15	0	0	0	0	0

2.2.3 Result

For each strategy, 1000 simulations were conducted and the result is shown in Figure 8. The predicted durations for strategy 1 and strategy 2 is 255 days (SD = 15 days) and 266 days (SD = 11 days), and the resource cost is 11.88 million yen(SD = 0.80 million JPY) and 12.52 million JPY (SD = 0.78 million JPY). Overall, compared to strategy 2, strategy 1 has resource teams distributed and there is no resource competition, so that a few tasks can be performed concurrently. As a result, a relatively short estimation period and low cost was output on average. From the simulation result obtained from the proposed method, it can be judged that strategy 1 should be adopted in order to achieve shorter project duration and lower cost.
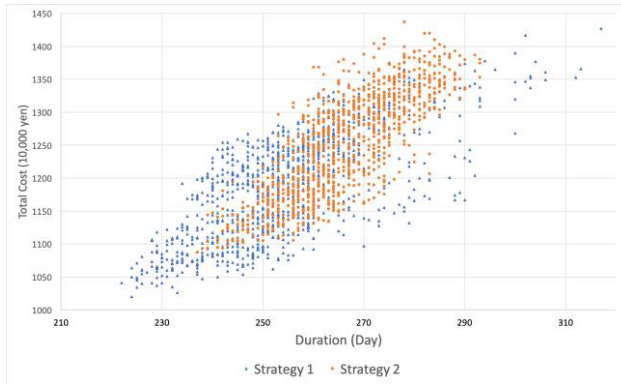

Figure 8. Simulation result in case study 2.

3. Discussion and Conclusion

For the discussion of the proposed method, we defined the information of past data applied to the parameter extraction method, and indicated that what kind of data is required when estimating the project duration when taking the uncertainty into account. However, the acquisition of such data will be inevitably the premise of the proposed method, which is considered to be difficult in real world sometimes.

As the conclusion of this paper, A method that predicts project duration based on parameters extracted from past data is proposed in this paper. The project simulation model considering uncertainty as delay and rework with parameter configuration method by past data is proposed. Multiple duration prediction results can support decision-making process in resource arrangement.

References

[1]   C. Earl, J. Johnson, C. Eckert., 'Complexity' in J.P. Carlson, C. Eckert (eds.) *Design Process Improvement – A review of current practice*, 2005, pp. 174-197.
[2]   T.R. Browning, *Modelling and analyzing cost schedule and performance in complex system product development*. Ph.D. Thesis, Massachusetts Institute of Technology, 1998.
[3]   T. Mitsuyuki, K. Hiekata, T. Goto, Evaluation of project architecture in software development mixing waterfall and agile by using process simulation, *Journal of Industrial Integration and Management*, 2017, Vol. 2, No. 2, 1750007.