

Estimation of Project Success Using Bayesian Classifier

Seiya Abe, Osamu Mizuno, Tohru Kikuno
Graduate School of Information Science and
Technology
Osaka University
1-5 Yamadaoka, Suita, Osaka, 565-0871, Japan
{s-abe, o-mizuno, kikuno}@ist.osaka-u.ac.jp

Nahomi Kikuchi, Masayuki Hirayama
Software Engineering Center
Information-technology Promotion Agency
2-28-8 Honkomagome, Bunkyo-Ku, Tokyo,
113-6591, Japan
{n-kiku, m-hiraya}@ipa.go.jp

ABSTRACT

The software projects are considered to be successful if the cost and the duration are within the estimated ones and the quality is satisfactory. To attain project success, the project management, in which the final status of project is estimated, must be incorporated.

In this paper, we consider estimation of the final status (that is, successful or unsuccessful) of project by applying Bayesian classifier to metrics data collected from project. In order to attain high estimation accuracy rate, we must select only a set of appropriate metrics to be applied. Here we consider two selection methods: the first method by the experts and the second method by the statistical test.

Then we conducted an experiment using 28 project data and 29 metrics data in an organization of a certain company. The result showed that the method by the test gave higher accuracy rates than the method by the experts, and Bayesian classifier with the test method is effective to estimate project success.

Categories and Subject Descriptors: K.6.3 [Software Management]: Software development

General Terms: Measurement

Keywords: project success, estimation, Bayesian classifier

1. INTRODUCTION

From the standpoint of industries, it is deeply desired to lead all of the software projects into success ones. Generally speaking, project success is considered from three distinct viewpoints: the quality of the product, the cost of development and the duration of the project.

Recently it is said that the projects become more and more difficult to be accomplished successfully. As the main reasons for this, the increases in the functionality and the complexity of the projects are pointed out [2]. Additionally the demands on the short term development and the small amount of budget are clearly other practical reasons. In fact, many studies reported several unsuccessful projects in software development [1, 4].

On the other hands, there are many innovative researches which try to lead projects into success [3, 5–7]. The most fundamental way is an improvement in the test activity or the review activity of the development process. It is easy to incorporate into the project, but unfortunately the effects

are rather limited. Thus the prediction of project success at early stage is one of the important tasks for project management.

In this paper, we try to predict project success at the early stage of the software development. If a project is estimated to be unsuccessful, then appropriate treatment can be applied to change the project into successful one.

Here we apply the Bayesian classifier to metrics data collected from development process in order to predict project success. By the result of the preliminary experiments, the capability of the prediction was verified within limit, but the accuracy rate of prediction was not satisfactory. Thus in this paper, we try to select an appropriate set of metrics in order to increase the accuracy rate of the prediction.

We propose two selection methods and compare the accuracy rates of two methods. The first method is to select metrics by the experts. Thus the key idea is exactly the know-how or the experiences by the experts. On the other hand, the second method applies statistical test to project data in order to select the metrics. Then we evaluate the effectiveness of two methods by applying the Bayesian classifier with two methods to 28 practical project data in an organization of a certain company.

2. PRELIMINARIES

2.1 Project Context

Software development projects must be performed by developers under the constraint of limited resources. They must be executed by balancing the competing demands for quality, time, and cost. In order to terminate the projects successfully, a project plan must be prepared and the projects must be controlled by the managers according to the plan.

In this research, we deal with 28 software development projects in an organization of a certain company. Here we describe project context of these 28 projects as follows: (1) They were the development and integration of software system for specific corporate customers. (2) As the final product, high reliability system is strongly needed. (3) The development was performed in the same development processes. (4) The size of all projects are middle or large. (5) The company is responsible for all these projects.

2.2 Software Metrics

Software measurement and quantification are crucial to enable control of projects or processes. Software metrics provide a quantitative basis for the development. Metrics are indicators which reflect the situation of process execu-

tion. Since the result of process execution influences on the quality of products, the cost, and the period of the development projects, it is thought that metrics are related to these three aspects.

In this research, we consider 29 metrics shown in Table 1 that are classified into five categories: development process, project management, company organization, human factor and external factor. These metrics are applied to the 28 projects and the data are collected. Each metric is evaluated by ordered scale with 5 degrees. We cannot present the detailed definitions of these metrics according to the contract with the company.

Table 1: All metrics

1	Metrics on Development Process
	M1: Familiarity of technology used
	M2: Necessity of adjustment to the other parties
	M3: Requirements elaboration
	M4: Pending requirements
	M5: Potential requirements change
	M6: Complexity of system features
	M7: Difficulty of integration of COTS
	M8: Performance and HW constraints
	M9: Feasibility of performance requirements
	M10: Experience of platform
	M11: Stability of platform
	M12: Amount of documentation
	M13: Design consideration of reliability, safety, security
	M14: Risks of tools used
2	Metrics on Project Management
	M15: Planning of the whole project
	M16: Planning of test environment
	M17: Size and cost estimation
	M18: Availability of system and support
	M19: Communication among project stakeholders
3	Metrics on Company Organization
	M20: WBS
	M21: New customer
	M22: Customer's participation to requirements definition
	M23: Customer's participation to acceptance test
	M24: Existence of process manuals and adherent execution of process
4	Metrics on Human Factor
	M25: Skill and experience of developers
	M26: Skill and experience of PM
	M27: Skill and experience of designers
	M28: Continuity of key members
5	Metric on External Factor
	M29: Requirement volatility

2.3 Quality, Cost and Duration

It is strongly required that all projects deliver the required products on the scheduled time within the estimated budget. Thus we should evaluate project success from three viewpoints: the quality of product, the cost of development and the duration of project. The followings show the criteria to judge successfulness of a given project.

Quality: The quality is evaluated by the number of defects found during acceptance test by user. If the number is less than #1 (a threshold), then the project is evaluated to be successful. Here we set #1 to be 0.

Cost: Since the cost is not available for us, we adopt the value of development effort instead of the cost itself. If the difference between the actual effort and the estimated one is less than #2(a threshold), the project is evaluated to be successful. Here we set #2 to be 10 percentage.

Duration: The duration is evaluated by the development period. If the difference between the actual period and the estimated one is less than #3(a threshold), the project is evaluated to be successful. Here we set #3 to be 1 day.

For the threshold #1, #2, #3, we determined actual values such as 0, 10 percentage, and 1 day according to discussions with managers in the company.

Table 2 shows the judgement of 28 projects. The judge-

ment was done according to the criteria mentioned above. In the table, "-" indicates that the judgement was not done for the viewpoint since necessary data is not available.

Table 2: Judgement of project success

Project ID	Viewpoints		
	Quality	Cost	Duration
P1	Unsuccessful	Successful	Successful
P2	-	Unsuccessful	Successful
P3	Successful	Successful	Successful
⋮	⋮	⋮	⋮
P28	Successful	Successful	-

2.4 Research Objective

The final goal of this research is to predict the final status(that is successful or unsuccessful) of a project using metrics data. As we mentioned in subsection 2.3, there are three key viewpoints to evaluate project success: quality of product, cost of development and duration of project. Thus we consider the status of the project from these three viewpoints.

In this paper, we apply Bayesian classifier to metrics data to predict the final status of the project. Then, as we showed in subsection 2.2, we can use 29 metrics for each project.

By observing the metrics carefully, we notice that some metrics are strongly related to a certain viewpoint, but are not effective to others. It brings us an important research objective that how to select useful metrics in order to predict the final status with respect to three viewpoints.

3. BAYESIAN CLASSIFIER

The naive Bayesian classifier is one of the most common approach to classify categorical data into several classes. Let Q_1, Q_2, \dots, Q_n be the variables denoting the observed attribute values to predict a discrete class C . Furthermore, let c represent a particular class. Given the values q_1, q_2, \dots, q_n , we can use Bayes' theorem to calculate the probability $P(C = c|Q_1 = q_1 \wedge \dots \wedge Q_n = q_n)$ and then predicts most probable class. By the Bayes' theorem, this probability is expressed as follows:

$$\frac{\prod_{i=1}^n P(Q_i = q_i|C = c)}{P(Q_1 = q_1 \wedge \dots \wedge Q_n = q_n)} \times P(C = c).$$

The probability $P(C = c)$ can be easily estimated from training data. In this research, we deal with numeric attributes. Each numeric attribute is modeled by some continuous probability distribution over the range of its values. We assume that within each class, the values of numeric attributes are normally distributed. Generally $P(Q_1 = q_1 \wedge \dots \wedge Q_n = q_n)$ is not estimated as it is just a normalizing factor. Instead, we can ignore it and then normalize so that the sum of $P(C = c|Q_1 = q_1 \wedge \dots \wedge Q_n = q_n)$ over all classes is 1. Thus we can compute $P(C = c|Q_1 = q_1 \wedge \dots \wedge Q_n = q_n)$ and predict the value of C if the values of q_i are given.

In this paper, we let Q_i denote a risk factor or a metric and C denote the status of a project with two values: s (successful) and u (unsuccessful). From the calculated probability, we can classify the test data into two classes. Here, we determine a project is in class s if $P(C = s|Q_1 = q_1 \wedge \dots \wedge Q_n = q_n) \geq P(C = u|Q_1 = q_1 \wedge \dots \wedge Q_n = q_n)$.

Table 3: Evaluation by expert method

Quality			Cost			Duration		
Actual	Predicted (# of Projects)		Actual	Predicted (# of Projects)		Actual	Predicted (# of Projects)	
	Successful	Unsuccessful		Successful	Unsuccessful		Successful	Unsuccessful
Successful	16	4	Successful	14	5	Successful	7	7
Unsuccessful	3	1	Unsuccessful	8	0	Unsuccessful	6	6
Accuracy rate: 0.708			Accuracy rate: 0.519			Accuracy rate: 0.500		

Table 4: Evaluation by test method

Quality			Cost			Duration		
Actual	Predicted (# of Projects)		Actual	Predicted (# of Projects)		Actual	Predicted (# of Projects)	
	Successful	Unsuccessful		Successful	Unsuccessful		Successful	Unsuccessful
Successful	20	0	Successful	14	5	Successful	11	3
Unsuccessful	3	1	Unsuccessful	3	5	Unsuccessful	3	9
Accuracy rate: 0.875			Accuracy rate: 0.704			Accuracy rate: 0.769		

4. EXPERT METHOD

4.1 Selection by the Experts

In order to find a set of metrics from 29 metrics which is used in Bayesian classifier, three experts discussed the effectiveness of each metric with respect to three viewpoints in brainstorming. All of the experts have a wealth of business experiences in software development and understand metrics quite well.

For each metric, they determine the viewpoints for which the metric seems to be effective. Please note that some metrics are determined to be effective for several viewpoints.

As the result, the selection of metrics are done. For the quality of product, 12 metrics, M7, M8, M9, M10, M11, M12, M13, M19, M24, M25, M27, and M28 are selected to be effective. For the cost, only 3 metrics, M12, M14, and M29 are selected. Finally for the duration, 18 metrics, M1, M2, M3, M4, M5, M6, M7, M15, M16, M17, M18, M19, M20, M21, M22, M23, M25, and M26 are selected. It is observed that each metric is considered to be effective to at least one viewpoint.

4.2 Application of Bayesian Classifier

In order to apply Bayesian classifier to predict final status of projects with respect to viewpoints, we must prepare two kinds of data: the data for learning and the data for prediction. In this paper, we adopt the leave-one-out cross validation method. Thus we choose one project as the data for prediction, and predict its status using the data of the rest 27 projects. By executing the prediction interchangeably for each project, we get all predicted results.

In case of prediction with respect to the quality of product, we consider M7, M8, M9, M10, M11, M12, M13, M19, M24, M25, M27, M28 as $Q_1, Q_2, Q_3, Q_4, Q_5, Q_6, Q_7, Q_8, Q_9, Q_{10}, Q_{11}, Q_{12}$, respectively.

The result of quality prediction is shown in Table 3. The table shows that 19 projects are predicted to be successful. Among them, 16 projects are actually successful and 3 projects are unsuccessful. On the other hand, 5 projects are predicted to be unsuccessful, and 4 projects are actually successful and 1 project is unsuccessful among them. The accuracy rate is calculated to be $(16 + 1)/24 = 0.708$.

In the same way, the cost and the duration are predicted by applying Bayesian classifier, and the evaluated results are shown in Table 3. The accuracy rates for the cost and the duration are 0.519 and 0.500, respectively.

5. IMPROVEMENT IN ESTIMATION

5.1 Selection by Statistical Test

In order to improve the accuracy rate of prediction, we propose to perform the Wilcoxon rank sum test on project data to select more effective metrics out of 29 metrics.

Generally, the Wilcoxon rank sum test can determine if one population is statistically different from another by comparing the locations of two populations. In this paper for each metric, we apply the test to check if there is a statistical difference of the distributions between the group of successful projects and the group of unsuccessful projects. If such a difference is observed for a certain metric, then the metric seems to have strong influence on the prediction.

We adopt the probability value(that is, p-value) to check if there is a statistical difference. The smaller the p-value is, the more statistical difference is observed between two groups. However since we need several metrics to be applied to Bayesian classifier, we set p-value to be 0.3(relatively large compared with ordinary values 0.05 or 0.1).

At first the Wilcoxon rank sum test is applied to project data shown in Table 2 with respect to the quality and data for metrics M1 through M29. As the result, 10 metrics, M1, M9, M18, M19, M20, M25, M26, M27, M28 and M29 are selected to be effective with respect to the quality. Then, for the cost the Wilcoxon test is applied similarly and 4 metrics, M6, M14, M21 and M28 are obtained. Finally for the duration the Wilcoxon test is applied and 8 metrics, M7, M9, M10, M19, M21, M22, M25 and M27 are obtained. It is clearly observed that 13 metrics are not selected for any viewpoint by the Wilcoxon test, although all metrics are selected by the experts in subsection 4.1.

5.2 Application of Bayesian Classifier

In order to predict final status of projects, we applied Bayesian classifier. In the application, we again adopt the leave-one-out cross validation method to project data shown in Table 2. Here we use metrics data which are newly selected by the Wilcoxon rank sum test.

The result of application of Bayesian classifier is shown in Table 4. For the quality of product, 21 projects are predicted correctly and only 3 projects are failed to predict. Thus the accuracy rate is 0.875 and thus it is improved quite well compared with 0.708 by the experts. It should be noted that for 20 projects, which are actually successful, the Bayesian classifier can predict correctly.

Next, for the cost of development, the accuracy rate be-

Table 5: Comparison of accuracy rate

Method	Quality	Cost	Duration
Expert method	0.708	0.519	0.500
Test method	0.875	0.704	0.769

comes 0.704. In this case, compared with the accuracy rate 0.519 by the experts, it is also improved very much. By just looking at the resultant tables, we notice that out of 8 projects, which are actually unsuccessful, the Bayesian classifier can predict 5 projects correctly. However for the rest 3 projects the classifier failed to predict.

Finally for the duration of project, the accuracy rate becomes 0.769. In this case, compared with the accuracy rate 0.500 by the experts, it is highly improved.

6. DISCUSSIONS

Table 5 summarizes the accuracy rates of two methods and Table 6 shows comparison of metrics selected by two methods. At first, we can see that significant improvements in accuracy rates are attained by the test compared with the one by the experts. The reason for this improvement is the difference between the selected metrics, since both methods apply the same Bayesian classifier to the project data.

We start to analyze the characteristics of the comparison in Table 6. Out of 29 metrics which are selected by the experts, only 16 metrics are chosen by the test. Thus the rest 13 metrics are not selected in the method using the test. For the viewpoint of the quality, 12 metrics are selected by the experts and 10 metrics are selected by the test. Comparing 12 metrics and 10 metrics, 5 metrics are commonly contained. For the viewpoint of the cost, 3 metrics are selected by the experts and 4 metrics are selected by the test. These two sets of metrics are different so much, since only one metric M14 is contained commonly. For the viewpoint of the duration, 18 metrics are selected by the experts and only 8 metrics are selected by the test. Out of 18 metrics, 5 metrics are commonly selected by the test, and 8 metrics are completely removed.

Firstly for the quality, Table 5 shows that relatively high accuracy rates are attained by both methods and that a further improvement is accomplished by the method with test. As mentioned earlier in subsection 5.2, the improvement is realized by the correct prediction of successful projects. On the other hand, for unsuccessful projects the accuracy rate is still not satisfactory. The main reason is that the number of projects, which are actually unsuccessful, is only 4 and thus very small compared with 20. Thus it seems to be very difficult to predict unsuccessful projects correctly.

Next for the cost, Table 5 shows that the accuracy rate is improved very much. In this case the improvement is realized by the increase of the accuracy rate of prediction with respect to unsuccessful projects. That is, all unsuccessful projects are predicted incorrectly in Table 3, but 5 out of 8 are predicted correctly in Table 4. This essential change in the prediction may be incurred by the changes in selected metrics as shown in Table 6.

Finally for the duration, Table 5 shows that the accuracy rate is highly improved. It is realized by the increase of the number of the correct predictions for both successful and unsuccessful projects. As pointed out before, the number of metrics in subsection 4.1 is too large compared with the

Table 6: Comparison of selected metrics

Metrics	Expert method	Test method	Metrics	Expert method	Test method
M1	Duration	Quality	M16	Duration	
M2	Duration		M17	Duration	
M3	Duration		M18	Duration	Quality
M4	Duration		M19	Quality, Duration	Quality, Duration
M5	Duration		M20	Duration	Quality
M6	Duration	Cost	M21	Duration	Cost, Duration
M7	Quality, Duration	Duration	M22	Duration	Duration
M8	Quality		M23	Duration	
M9	Quality	Quality, Duration	M24	Quality	
M10	Quality	Duration	M25	Quality, Duration	Quality, Duration
M11	Quality		M26	Duration	Quality
M12	Quality, Cost		M27	Quality	Quality, Duration
M13	Quality		M28	Quality	Quality, Cost
M14	Cost	Cost	M29	Cost	Quality
M15	Duration				

number of project data. Thus some metrics are generally related to the duration, but judging from the project data they are not so associated with it. Therefore the number of metrics is reduced by the test.

7. CONCLUSION

In this paper, we have predicted the final status(that is, successful or not) of a given project by applying the Bayesian classifier to metrics data collected from the project. The final status is evaluated from the three distinct viewpoints: the quality of the product, the cost of development and the duration of the project.

In order to attain high accuracy rate of the prediction, selecting an appropriate set of metrics is a key step. Here we present two methods: the selection by experts and the selection by the statistical test. The experts select metrics using so-called know-how, and the test can select metrics by applying statistical knowledge. The experiments using an actual data in an organization of a certain company showed the superiority of the test over the experts.

The future research works include further analysis of the result in this paper, and evaluation of Bayesian classifier under another project context.

8. ACKNOWLEDGMENTS

Authors would like to thank Oki Electric Industry Co., Ltd. for their cooperation.

9. REFERENCES

- [1] R. L. Glass. How not to prepare for a consulting assignment and other ugly consultancy truths. *Communication of the ACM*, 41(12):11-13, 1998.
- [2] R. L. Glass. Software runaway - some surprising findings. *Journal of Systems and Software*, 41(2):75-77, 1998.
- [3] J. Jiang and G. Klein. Software development risks to project effectiveness. *Journal of Systems and Software*, 52:3-10, 2000.
- [4] S. McConnell. *Rapid Development*. Microsoft Press, 1996.
- [5] J. D. Procaccino, J. M. Verner, S. P. Overmyer, and M. E. Darter. Case study: factors for early prediction of software development success. *Information and Software Technology*, 44:53-62, 2002.
- [6] Y. Takagi, O. Mizuno, and T. Kikuno. An empirical approach to characterizing risky software projects based on logistic regression analysis. *Empirical Software Engineering*, 10(4):495-515, 2005.
- [7] C. Wohlin and A. A. Andrews. Prioritizing and assessing software project success factors and project characteristics using subjective data. *Empirical Software Engineering*, 8:285-303, 2003.