Evaluation of Project Quality: A DEA-Based Approach

Shen Zhang[1,2], Yongji Wang[1,3], Jie Tong[1,3], Jinhui Zhou[1,3], and Li Ruan[1,3]

[1] Laboratory for Internet Software Technologies, Institute of Software,
The Chinese Academy of Sciences, Beijing 100080, China
{zhangshen, ywang, tongjie, jinhui,
ruanli} @itechs.iscas.ac.cn
[2] Key Laboratory for Computer Science,
The Chinese Academy of Sciences, Beijing 100080, China
[3] Graduate University, The Chinese Academy of Sciences, Beijing 100039, China

Abstract. The evaluation of project quality exhibits multivariable, VRS (variable return to scale) and decision maker's preference properties. In this paper, we present a Data Envelopment Analysis (DEA) based evaluation approach. The DEA VRS model, which handles multivariable and VRS effectively, is used to measure project quality. And the DEA cone ratio model, which utilizes Analytical Hierarchy Process (AHP) to constrain quality metrics with respect to decision maker's preference, is also adopted to analyze the return to scale of the projects. A case study, which assesses 10 projects from ITECHS and 20 "Top active" projects on sourceforge.net with the novel method, is demonstrated. The results indicate that our approach is effective for quality evaluation and can get accurate estimates of future possible improvements.

1 Introduction

Evaluation of project quality can lead to a better control of the schedule, cost and resources allocation, furthermore smooth the way for process improvement efforts. However, there are three characteristics embedded in the evaluation problem.

Firstly, defect, which is a key measure of software quality, consists of multiple attributes, such as defect severity, defect priority, etc. Thus, the quality evaluation has to deal with multi-attribute problem. Secondly, to evaluate project quality, we usually take software scale and defect attributes as input and output. However, as is stated in [5][6], the relationship between system size and the number of defects or defect-density is nonlinear. Thus, the problem of evaluation exhibits VRS (variable return to scale, i.e. the relationship between the input and the output is non-linear). Thirdly, generally speaking, the evaluation should be consistent with managerial goal of the organization. Thus, incorporating subjective managerial preference into quality assessment must be taken into account [2]. In a word, an efficient evaluation method is needed to fulfill these requirements of multivariate, VRS and decision maker's preference properties.

Data Envelopment Analysis (DEA) developed by A. Charnes and W. W. Cooper [12] in 1978 is a non-parametric mathematical programming approach. It can be used to evaluate the relative performance of a number of decision making units (DMU), which may have multivariate input and output. Henceforth, dozens of DEA extension

Q. Wang et al. (Eds.): SPW/ProSim 2006, LNCS 3966, pp. 88 – 96, 2006.

models have been brought into the world, Banker, Charnes and Cooper improved the basic theory and established the first DEA VRS model (BCC) [9] in 1984. Five years later, the C^2WH cone ratio model [11] with respect to "preference of decision maker" was brought forward by Charnes in 1989. At present, DEA has been widely accepted in the computing industry.

In this paper, we present a DEA-based approach to evaluate the project quality. The approach utilizes DEA CCR model and its extension models to calculate the quality score, which is the basis of the evaluation result. Since the datasets used for studies and analysis are collected from defect report and tracking systems, where cost and schedule information is insufficient, we only extract defect-related attributes from defect reports as input/output metrics in our approach. And then the quantitative results to measure the further possible improvements of low quality projects are discussed. Furthermore, the return to scale of each project with respect to decision maker's preference is also investigated.

2 Relate Work

[1] proposes to use DEA VRS model to measure the performance of ERP projects. Their method can handle multivariate data and VRS well, but doesn't take into account subjective managerial goal. Since they only evaluate the productivity as performance score, quality measurement is recommended to improve their work. Our work can be thought an extension of their study.

[10] presents a case study on an OSS(Open Source Software) development project, the FreeBSD project, and then compares the quality of OSS projects with that of commercial projects. But the evaluation only focuses on defect-density, which is the key quality metric, and ignores the impact brought about by other defect attributes. Also, their measurement can't deal properly with VRS.

J.C. Paradi et al. [2] introduce a DEA-based model to measure the performance of a group of software development projects and investigate the effect of quality on software maintenance projects. Decision maker's preference is incorporated into their model as well. However, the definition of quality used in their paper is quite narrow and omits other quality indicators, which can be easily extracted from defect reports.

In a word, compared with the existing models and methods for performance evaluation, our approach has the advantage of dealing with multivariate, VRS and decision maker's preference issues properly at the same time.

3 The DEA-Based Project Quality Evaluation Approach

In this section, we present our DEA-based project quality evaluation approach, which can be divided into four steps: constructing project dataset; establishing the input/output of DMUs; assessing project quality; analyzing Return to Scale. Figure 1 illustrates the flow chart of our approach.

3.1 Constructing Project Dataset

Constructing project dataset is to determine reference DMU sets. For the purpose of project quality measurement, we select each project under evaluation as a DMU. Moreover, because our DEA-based approach evaluates the relative quality among the similar DMUs, the basic requirement of the DMU selection is that the DMUs must be *homogenous*. The homogenous DMUs mean that they are project sets satisfying the same conditions, such as they are both object oriented projects and developed by the same language, so that the DMUs are comparable in quality.

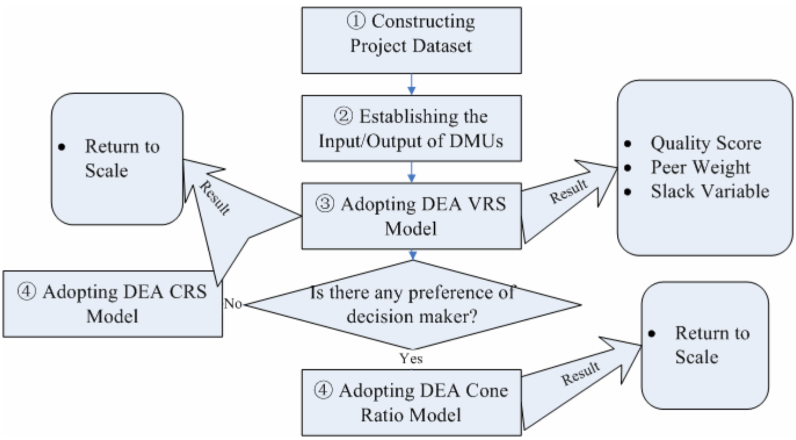

Fig. 1. DEA-based project quality evaluation approach

3.2 Establishing the Input/Output of DMUs

After constructing project dataset, we will establish input/output of DMUs [8] which largely depends on the selection of quality metrics.

Firstly, the defect reports specification of the projects under evaluation should be taken into account. It is because the selection of quality metrics is based mainly on the indicators provided by these defect reports. For example, when we choose quality metrics for the projects on sourceforge.net, we can’t gather the information of defect priority and defect life-cycle, since defect reports on sourceforge.net don’t provide any indicators of defect priority and defect life cycle at all.

Secondly, we must consider the relationship of the quality metrics. Because these metrics are not isolated, they may influence the cognizance of other variables. For example, we should discard a variable if its information has been covered by other several variables or has strong relationship with some other input/output variables.

Thirdly, we filter out the metrics that can’t be quantified easily, for example, the customer satisfaction (corresponding to the comments submitted by customers in defect reports) and so on. Then we can generate the remaining metrics value for all the DMUs. Note that they are all positive values.

Fourthly, according to the efficiency ratio principle of DEA model, we prefer the smaller input values and bigger output values.

3.3 Assessing Project Quality

In order to evaluate the project quality, we adopt DEA VRS Model (BCC) [9] to deal with the nonlinear relationship inherent in the evaluation issue. The BCC model is written as:

$$\left(D^{O}{}_{BC^2}\right)=\begin{cases}\max\left[\theta+\varepsilon\left(\sum_{i=1}^{m}s_i^-+\sum_{k=1}^{s}s_k^+\right)\right]\\ \sum_{j=1}^{n}X_j\lambda_j+\sum_{i=1}^{m}s_i^-=X_0\\ \sum_{j=1}^{n}Y_j\lambda_j-\sum_{k=1}^{s}s_k^+=\theta Y_0\\ \sum_{j=1}^{n}\lambda_j=1\\ \lambda_j\geq 0, j=1,\dots,n\\ s_i^+\geq 0, i=1,\dots,m\\ s_k^+\geq 0, k=1,\dots,s\end{cases} \tag{1}$$

From (1) we calculate the quality score θ , the peer weight λ and slack variable s. The quality score is between 1 and $+\infty$, A project with quality score of 1 is of *relative high quality*, otherwise the project is of *relative low quality*. Each project can be presented by a linear combination of the DMU sets, such as: $DMU_{j0}=\lambda_i DMU_i+\lambda_k DMU_k+\cdots+\lambda_j DMU_j$. The peer weight λ_i provides the degree that high-quality project i for the relatively low-quality project j_0 to emulate. The slack variable s can be divided into two parts: input slack variable s^- and output slack variable s^+. The former represents the over use of work effort scale, while the latter represents the insufficient quality metrics. Since we focus on defect elimination, we present the formula (2) to calculate the quantitative improvement of every quality metric for low-quality projects:

$$\Delta=\theta y_{j0}+s_j^+-y_{j0} \tag{2}$$

3.4 Analyzing Return to Scale

After computing the results using DEA VRS model, we analyze return to scale between software scale and the quality metrics represented by defect attributes. For this purpose, we should take into account whether some specific managerial preference exists. When there is no impact of managerial preference, we can combine the results of DEA CRS model and VRS model to judge return to scale. First, calculate the quality score δ with DEA CRS model, then compare δ with θ , there are three conditions: 1) $\delta<\theta$, the project exhibits IRS ; 2) $\delta=\theta$, the project exhibits CRS; 3) $\delta>\theta$, the project exhibits DRS. —IRS (DRS) indicates that an increase in one unit's inputs will yield a greater (or less) proportionate increase of its outputs.

Otherwise, when it is necessary to incorporate subjective managerial preference in return to scale analysis, we should utilize the DEA cone ratio model [11] to fulfill managerial goals. In order to constrain the weights of quality metrics according to managerial preference, we adopt AHP (Analytical Hierarchy Process) [7]. Firstly, we gather opinions of several project managers on "the importance of each quality metrics", then establish the AHP Decision Matrix A_m and calculate the max latent root $\lambda_{\max}$ of A_m. Secondly, we construct weight constraint

$$\Gamma = \{\mu \mid (A_m - \lambda_{\max} E_m)\mu \geq 0\} \tag{3}$$

where μ in Γ means the weights of quality metrics. Thirdly, incorporating Γ into DEA cone ratio model (4),(5) and calculate the parameter μ_0 which is the indicator of return to scale. There are also three conditions: 1) $\mu_0<0$, the project exhibits DRS; 2) $\mu_0=0$, the project exhibits CRS; 3) $\mu_0>0$, the project exhibits IRS;

$$\left(\hat{P}_{C^2R}\right) = \begin{cases} V_1 = \min(\omega^T X_0) \\ \omega^T X_j - \mu^T Y_j \geq 0, j = 1,\ldots,n \\ \mu^T Y_0 = 1 \\ (A_m - \lambda_{\max} E_m)\mu \geq 0 \end{cases} \tag{4}$$

$$\left(\hat{P}_{BC^2}\right) = \begin{cases} V_2 = \min(\omega^T X_0 - \mu_0) \\ \omega^T X_j - \mu^T Y_j + \mu_0 e^T \geq 0, j = 1,\ldots,n \\ \mu^T Y_0 = 1 \\ (A_m - \lambda_{\max} E_m)\mu \geq 0 \end{cases} \tag{5}$$

4 Case Study

In this section, an empirical study is presented based on the sequence in Section 3.

Firstly, we construct the evaluation data sets. The first dataset consists of 10 projects from one single organization —ITECHS [3]. On the contrary, our second dataset consists of 20 "Top active" projects on sourceforge.net [4], which are developed by different organizations. These projects of the two datasets are all developed in Java. Especially the projects in the first dataset are all J2EE Web Applications, so the DMUs can be regarded as *homogenous.*

Secondly, according to the specification of defect reports of selected projects (13 metrics in total), we have chosen the following metrics for the first dataset. While only defect severity, system size and work effort are used in the second dataset as its Input/Output metrics.

Table 1. Input/Output metrics for evaluation

Metrics	Type	Meaning
Defect Severity	Output	Defects can be divided into four levels by severity: C,S,N,M.
Defect Life Cycle	Output	Defects can be divided into five class by the length of its life cycle: I,S,M,L,E
Defect Priority	Output	Defects can be divided in to three level by priority: H,M,L
System Size	Output	
Work Effort	Input	

Thirdly, the results of the quality measures on the ITECHS dataset using DEA VRS model are presented in Table 2. We observe that only two DMUs 6,7 are of relative low quality, while other eight DMUs are all of relative high quality. Moreover, the relative low quality projects can be improved under relative high quality projects' guidance in the future. For example, DMU6 can be shown in the following form: 0.06*DMU1+0.48*DMU2+0.25*DMU9+0.20*DMU10, so the DMU2 is of more benefit to help quality improvement since its peer weight is larger than others'.

Table 2. Quality scores and peer weights obtained from DEA VRS model (Dataset 1)

DMU	Quality score	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
1	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
5	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
6	1.36	0.06	0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.20
7	1.22	0.15	0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.55
8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
9	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
10	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

In Table 3, the output slack variables s^+ can be used to calculate the margin of quality improvement for each quality metric. For example, in order to reduce the defects whose life-cycle is 6-10 days ("M"—Medium in defect life-cycle defined in table 1) in project 6, we combine the slack variable s_8^+ =217 with formula(2) in section 3.3, then calculate the $\triangle$= 9. The result means that the defects, whose life-cycle is "M" in project 6, can get an optimal reduction by 9 under the relative high quality projects' guidance in the future development.

Based on the Sourceforge dataset, we get the similar aggregate result. In Table 4, we only show the quality scores of the 20 projects.

Table 3. Slack variables obtained from DEA VRS model (Dataset 1)

DMU	Work Effort	System Size	Defect Severity				Defect Life-Cycle					Defect Priority		
			C	S	N	M	I	S	M	L	E	H	M	L
	s_1^-	$s_1^+, s_2^+, \ldots\ldots, s_{13}^+$												
6	248	0	3	10	43	49	0	62	217	36	25	0	98	0
7	0	0	2	34	72	106	4	77	83	75	59	0	131	174

Table 4. Quality scores obtained from DEA VRS model (Dataset 2)

DMU	1	2	3	4	5	6	7	8	9	10
θ	3.79	5.59	5.38	3.02	3.65	1.00	1.98	2.78	7.39	6.53
DMU	11	12	13	14	15	16	17	18	19	20
θ	5.02	10.18	1.00	3.77	7.64	3.79	10.49	6.74	5.49	3.58

Fig.2 illustrates a comparison of two methods for quality evaluation of the two datasets. The first method is our DEA-based approach, while the second is to assess quality by defect-density (abbreviated as DD). In the chart, x-axis denotes project number, y-axis denotes quality score. Fig.2 reveals that DEA-based approach can make a more fair evaluation than DD, which can't handle VRS. For example, the third project in the second dataset is regarded as a project of the lowest quality by DD, since its defect-density is nearly 20 times greater than that of project 13, whose defect-density is the lowest. But using DEA-based approach, the quality score is only 5 times greater than that of the highest quality project. The reason for this is that project 3 is the biggest project with 409829 lines of code and 2113 defects, while project 13 has only 115144 lines of code and 43 defects. It is obvious that the comparison between a large project like 3 and a small project like 13 in defect-density is inappropriate, since the evaluation problem exhibits VRS. In general, it seems more reasonable to compare a project with other projects of similar size. So applying our VRS approach is more appropriate to solve the problem. Besides, as can be seen in Fig.2, the curve of dataset 1 is much smoother and closer to 1 than that of dataset 2 in our approach. It means that the process performance of ITECHS is significantly higher than that of the projects in dataset 2.

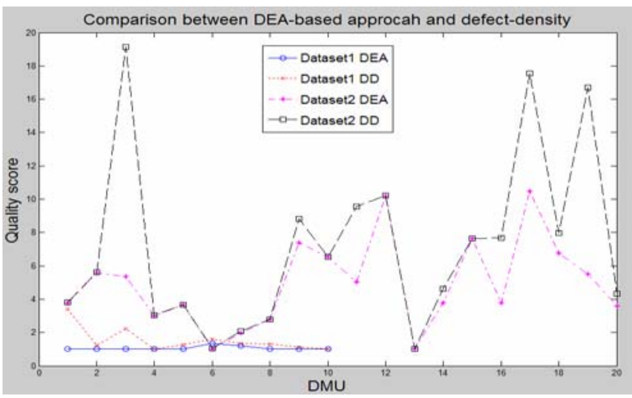

Fig. 2. Comparison between DEA-based approach and defect-density

Table 5. Return to scale obtained from DEA cone ratio model (Dataset 1)

DMU	1	2	3	4	5	6	7	8	9	10
V_1	2.04	1.00	1.38	1.17	1.46	1.86	1.61	1.44	1.35	1.35
V_2	1.00	1.00	1.00	1.00	1.00	1.42	1.31	1.00	1.00	1.00
μ_0	minus	0	minus	minus	minus	minus	minus	minus	minus	minus
result	DRS	CRS	DRS	DRS	DRS	DRS	DRS	DRS	DRS	DRS

In the last step, we present how to use our approach to analyze the return to scale of each DMU. As we have consulted several project managers from ITECHS for their preference on the quality metrics listed in table 1, we are convinced that the cone ratio

DEA model should be adopted to investigate the return to scale for the first dataset. After building the AHP Decision Matrix by incorporating the managerial goals, we use the modified model (4),(5) to calculate the results which is shown in Table 5. As we can see, all the projects except the second have DRS, which means the rate of various defects attributes in these projects increases quicker than the rate of the expending work effort. So the managers should consider of slowing down the scale expansion of these projects, then turn to make improvements in process efficiency.

5 Conclusion

The paper focuses on three intrinsic characteristics of project quality evaluation: multivariable, Variable Return to Scale (VRS) and preference of decision maker. To overcome the difficulties caused by these characteristics, we advocate a DEA-based approach which can fulfill these requirements. A case study illustrates the principle of our approach well. The results of the DEA-based approach is helpful to assess the project quality and estimate the margin of future possible improvement. The return to scale analysis can also help managers to make a decision on an expansion or a reduction in software scale.

Acknowledgements

This paper was partially supported by the National Natural Science Foundation of China (Grant Number: 60373053), 863 Program (Grant Number: 2005AA113140), the Plan of Hundreds Scientists in the Chinese Academy of Sciences, and the key program of the National High-Tech Research and Development Program of China (Grant Number: 2003AA1Z2220).

References

1. Stensrud, E., Myrtveit, I.: Identifying High Performance ERP Projects. IEEE Transaction on Software Engineering, 29(5) (2003) 387-416
2. Paradi, J.C., Reese, D.N., Rosen, D.: Applications of DEA to measure the efficiency of software production at two large Canadian banks. Annals of Operations Research 73(1997)91 – 115
3. http://itechs.iscas.ac.cn/
4. http://sourceforge.net/index.php
5. Malaiya, Y.K. Denton, J.: Module Size Distribution and Defect Density Software Reliability Engineering. Software Reliability Engineering, (2000)62-71
6. Rosenberg, J.: Some Misconceptions About Lines of Code. Software Metrics Symposium, (1997) 137-142
7. Golden, B.L., Wasil, E.A., Harker, P.T. (Eds.): The Analytic Hierarchy Process - Applications and Studies. Springer-Verlag (1989)
8. Liang L, CUI J.C.: Selection of Input-output Items and Data Disposal in DEA. Journal of Systems Engineering, 18(6) (2003) 487-490

9. Banker R.D., Charnes A., Cooper W.W.: Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Management Science, 30(9) (1984) 1078-1092.
10. Dinh-Trong T. Bieman J.M.: Open source software development: a case study of FreeBSD Software Metrics, Software Metrics, (2004) 96 - 105
11. Charnes A., Cooper W.W., Wei Q.L., etc.: Cone Ratio Data Envelopment Analysis and Multi-Objective Programming. International Journal of System Science, Vol. 20. 7 (1989) 1099-1118
12. Charnes A., Cooper W.W., Rhodes E.: Measuring the Efficiency of Decision Making Units. European J. Operational Research, Vol. 2. (1978) 429-444