



# EVALUATION OF COMPETING PERSONNEL IN SOFTWARE COMPANY USING EXPECTATION MAXIMIZATION APPROACH

Krishankumar R., Ravichandran K. S. and Anil Bala K.  
School of Computing, SASTRA University, Thanjavur, Tamil Nadu, India  
E-Mail: [krishankumar@sastra.ac.in](mailto:krishankumar@sastra.ac.in)

## ABSTRACT

Human effort is an integral part of any organization. The success of the project is directly influenced by the candidates involved in the project. Due to the constraint of time and cost, selecting the right candidate for the task is a key factor to be investigated. Researchers have proposed different optimization model for achieving the task. But, all these models consider the input vector to be a known parameter. With the help of this vector, knowledge is derived. In this paper, we deviate from the traditional setup and make efforts to analyze the input with the help of known pattern. The pattern is generally binary in nature. We propose a novel research model that uses Expectation Maximization (EM) and Maximum Likelihood (MLE) approaches for evaluating the input vector. We test the model over a personnel evaluation setup in Startup Company. Finally, results from confusion matrix infer that, the proposed model is effective in guessing the right personnel for the project.

**Keywords:** EM, MLE, confusion matrix, personnel selection.

## 1. INTRODUCTION

### 1.1 Background

Human Resource Management (HRM) is a process of managing human effort in the organizational setup to meet the emerging demand (Thorndike, 1949). There is a separate department that works for the effective maintenance of human power. Researchers have explored different parameters for effective handling of the resource. Such parameters are claimed as attributes that define the persons' professional and social character. (Safari *et al.* 2014) gave a new dimension to HRM by defining their job as the process of selecting, finding, hiring, evaluating and training a suitable work force for a certain specific task. Following this idea, (Delaney & Huselid, 1996) claimed that the concept of personnel selection can be viewed as a strategy based decision making process where feasible decision regarding human effort is made in systematic

manner. (Kabak *et al.* 2012) also gave a new perspective for personnel selection, which follows, the process of selecting the right work groups from the available set.

Based on these definitions researchers came up with novel ideas to select the right person for the job. They adopted both attribute based selection and methodical based selection. The first strategy selects candidates for the job based on the attributes possessed by the candidate. The preference threshold is set by the decision maker and candidates who satisfy the threshold are elected for the process. On the other hand, in methodical based strategy, a systematic procedure is adopted, which takes input from attribute based strategy for evaluation. The latter idea is considered effective and widely used in present day research. To summarize the different attributes used for personnel selection, we conduct a longitudinal survey process. The results obtained from the survey process are given in Table-1.

**Table-1.** Survey on personnel selection and its attributes.

Ref #	Preferred attributes	Realization approach	Discussion
(Blue et al. 2013)	Attitude	Interdisciplinary Education Perception and Teamwork Score	The role of skill and attitude was analyzed in a teamwork setup. Results show that, these two were key factors in team building
(Little et al. 2007)	Appearance	Thematic Survey	The role of facial appearance was analyzed in terms of voting behavior.
(Garcia et al. 2013)	Behavior	Artificial Intelligence	A tool named Groupware was designed to effectively assign role to teammates based on their behavior.
(Rathore et al. 2016)	Benevolence	Immune system model	Inspired from human psychology, sensor nodes were tested for trust using benevolence factors. Results prove to be effective.
(Castillo et al. 2008)	Beauty	Empirical and Thematic cases	Experiments were conducted to study the role of beauty in teaming. Empirical inferences were made rather than statistical inferences.
(Lee et al. 2013)	Culture	Field survey	The experiment was conducted with software developers to understand the effects of dispersion and the role of culture in balancing such issues.
(Krishankumar & Ravichandran 2016b)	Communication	Ant colony optimization and Hungarian method	The study analyzed the effects of communication in team and project success. It also investigated the role of closed and open communication in teams.
(Kamberg 2001)	Ethics and Ethnicity	Thematic survey	The experiments investigate the role of ethics and ethnicity in team work and team management
	Ego	Empirical study	Ensemble method was proposed for identification of ego centric community. Results show that, proposed approach performs better than other methods.
(Giordano & George 2013)	Experience	Empirical analysis	Experiments revealed that deception was less common among experienced people and thus virtual teams needed people with adequate experience.
(Karsh & Eyal 2015)	Emotion	Thematic study	The role of emotions in team performance was analyzed. Results showed that, positive emotions were boosting team work and performance.
(Assudani 2011)	Familiarity	Empirical analysis	The research analyzes the effects of familiarity in virtual team set up and infers that, this parameter plays a key role in boosting performance.
(Fernández-Sanz & Misra 2012)	Gender	Empirical study	Software teams were empirically gauged for their performance using gender as a metric. Results infer that, teamwork was affected by gender.
(Adair 2004)	Hatred	Empirical study	The research addresses the effects of hatred in team performance and concludes that, hatred is a delimiting attribute to performance.
(Krishankumar & Ravichandran 2016a)	Iron Star Factors	Research model	The factors like cost, time, quality, risk, resource and scope constitute the iron star factors. Ensemble classifiers were adopted to test their effects on team performance.
(Mihalcea 2014)	Job satisfaction	Thematic study	Research on job satisfaction and performance was carried out to understand the proportionality relationship between the two factors.
(Hausknecht et al. 2004)	Mindset/ Mood	Research model	The research deals with understanding the mood of the application during selection process.
(Schmit & Ryan 1993)	Personality	Big Five Trait model	The five factors namely, openness, consciousness, extraversion, agreeableness and neurotics were investigated and results proved that, consciousness and agreeableness are essential for high team performance.
(Chan et al. 2008)	Skill	Research model	The research proposes a model for understanding the effects of task skill on project performance. The model proves the fact that skill is an essential
(Zhang & Zhang 2015)	Trust	Classification model	Support Vector Machine classifier is used to analyze the role of trust in overall performance of cross functional teams
(Fagerholm et al. 2015)	Work Performance	Research framework	A new research framework is developed to understand the effects of different attributes towards performance. The model forms a holistic view of the attributes that affects project performance.

Based on the survey, we identified these 20 attributes that are closely related to personnel selection and team formation. The fields in the table are straightforward and are easy to understand. From the analysis, we infer that, prediction of attribute nature is an

interesting and challenging research area. Empirical cases demonstrate the nature of each attribute, but deciding whether an attribute is of benefit zone or cost zone still remains questionable. From the literature, we found that, quality, satisfaction, trust and personality are the dominant



attributes used by researchers for gauging personnel at work.

Before proceeding further, it is worth to note some key attributes used by researchers for electing the better work force for the project.

### 1.2 Problem Statement for MLE

We take an interesting scenario from the software company where groups are being gauged based on their previous success rates. Let us consider an illustrative example where 3 sets are considered each set involving 5 project results. Two groups are chosen and they are given these 15 projects for completion. Projects involve logic development and coding. The success or failure is the class that is determined for each trial. Based on this record a data set is to be developed. The objective of this illustration is to find out whether group A or B is viable/feasible for the next project.

### 1.3 Problem Statement for EM

In this section, we modify the scenario and get the data set for the 15 projects (3 set each having 5 projects). In this case, the data is known but the group that

performed the data is kept anonymous. Now, the interesting challenge here is to find out the possible group that adheres to the sample space.

The rest of the article is constructed as, Proposed research model in Section 2, Procedure for MLE in Section 3, Procedure for EM in Section 4, Results and Discussion in Section 5 and Conclusion in Section 6.

## 2. PROPOSED RESEARCH MODEL

Let us now propose the generic research model for the investigation of teams in the organization. The architecture is given in Figure-1. It consists of the data feed phase where the initial data is fed for processing. The teams' previous project success rates are fed as input based on the interviews. The process phase formulates a work cube for understanding the efficacy of each team. The procedure for two methods in the work cube is given in section 3 and 4. The results from the work cube reach the response phase where an optimal team is chosen for the next task. The research model is simple and straightforward. Henceforth, any organization can easily incorporate the model and make rational decisions during team selection process.

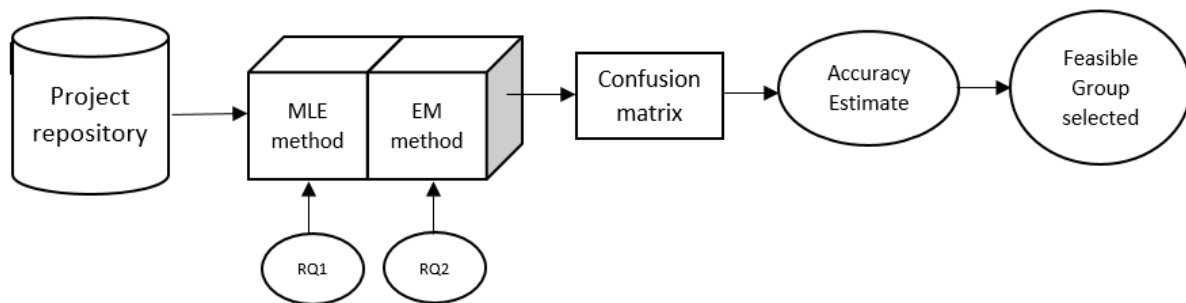


Figure-1. Proposed research model.

## 3. PROCEDURE FOR MLE

The MLE is the process of finding a suitable value for the likelihood function that maximizes the distribution. It is represented mathematically as  $\mu^{\wedge}$  and its given in (1).

$$\mu^{\wedge} = \theta^{\sum x} (1 - \theta)^{n - \sum x} \quad (1)$$

where  $\theta$  is the estimate value.

The steps for MLE are given below:

- Step 1:** Consider the set of projects and its completion rates. Form a matrix using the given data. The instances are binary in nature with 1 for success and 0 for failure.
- Step 2:** Find the probability of success and the probability of failure separately. This estimate determines the chance for success and failure by a group. Mathematically it is given by (2-3).

$$\lambda_A = \frac{\# \text{ of events occurred}}{\# \text{ of total events}} \quad (2)$$

$$\lambda_B = \frac{\# \text{ of events occurred}}{\# \text{ of total events}} \quad (3)$$

**Step 3:** Using this estimate, the managers can make rational decisions about future. This value indicates the previous success rates. If the success value is higher, then the future scope for the group is high. This estimate is however not fully right as the success in current task is an independent parameter of time. But still, this estimate gives some confidence to the manager regarding the working nature of the group and hence can be used as an opinion metric for gauging groups.

## 4. PROCEDURE FOR EM

The data associated with time factor are well defined using Markov models. The EM is an algorithm that can be deployed with this model to obtain computationally effective result. The main purpose of EM is to handle missing and incomplete data. Researchers have proved that, optimization in EM is difficult when compared to optimization in MLE. The EM is the



generalization of MLE. In general, EM is the log probability of the given set of data.

The complete set of data forms an objective function which is of log term. This function is solved using MLE and we obtain global optimum. Whereas, the incomplete type of data forms an objective function that cannot be solved using such mechanisms. The result for such data reach local optima and choosing global optimum from these local optimum points is often tough and misleading. To overcome this issue, EM concept is used. According to this approach, the problem of optimization is divided into modules with each module capable of reaching global optimum. These sub modules are further integrated to obtain the net global outcome. The EM has 2 stages namely E-stage and M-stage. In the E-stage, a feasible point  $\lambda$  is determined which sets the lower bound for the log term. On the other hand, the M-stage fine tunes the feasible point to find global solution. This mechanism is iterative and the steps for implementation are given below (Do & Batzoglu 2008):

**Step 1:** Choose arbitrary probabilities for the events as initial points. We name them as  $\lambda_0^A$  and  $\lambda_0^B$ .

**Step 2:** Now, we begin the process for E-stage: With the random points from Step 1, we calculate the new probability for the events. We adopt binomial distribution for this. We then normalize the results of binomial distribution to get the new probabilities for each of the event. Mathematically, it is given by (3-4).

$$B^d = P^x Q^{n-x} \quad (3)$$

$$B_d^* = \frac{B^d}{\sum B^d} \quad (4)$$

where, P is the success, Q is failure probability and  $B_d^*$  is the normalized estimate.

**Step 3:** Calculate the values of success and failure in each set using (5). Use this and apply (2-3) to estimate another set of probabilities for each event. This is again fed to Step 1 of EM.

**Step 4:** We follow, Step 1 to Step 3 repeatedly until we obtain an optimal point. The stopping condition for this iterative procedure is determined when for two successive iterations the value does not change.

## 5. RESULTS AND DISCUSSIONS

Let us now deploy the two methods to solve the research question under study. The RQ1 is solved using MLE and RQ2 is solved using EM. The two questions are challenging and interesting. These questions arise in managers' mind every time a new project is to be initiated and completed successfully.

The RQ1 deals with finding a suitable group for the project provided their history of previous success rate. Now, we apply the MLE procedure for solving this.

Consider 3 sets with 5 projects in each set for two groups A and B. Table 2 clearly shows the scenario.

**Table-2.** Success rates for each set of project.

Scenario	Set of projects		
	I	II	III
Group A	10011	01001	11010
Group B	00110	10010	01110

**Table-3.** Probability estimates.

Probability estimate	Set of projects		
	I	II	III
Group A	0.6(s), 0.4(f)	0.4(s), 0.6(f)	0.6(s), 0.4(f)
Group B	0.4(s), 0.6(f)	0.4(s), 0.6(f)	0.6(s), 0.4(f)

The Table-3 above shows the probability of success and failure for each set of projects. We obtain this by using (2) and (3). Now, we apply MLE to estimate the final probability for the two groups.

Total Success of Group A = 1.6

Total Failure of Group A = 1.4

MLE of Group A  $\theta_A = \frac{1.6}{3} = 0.5333$

Total Success of Group B = 1.4

Total Failure of Group B = 1.6

MLE of Group B  $\theta_B = \frac{1.4}{3} = 0.4666$

Thus, the net MLE is given by Table-4. From this, we observe that, Group A is a better choice for the next project under consideration.

**Table-4.** Estimation using MLE method.

Groups	Calculation of MLE	
	Success (%)	Failure (%)
Group A	53	47
Group B	47	53

Now consider RQ2 which is another interesting challenge in the organization. According to RQ2, we have a set of events indicating project success and failure rates. But, we are unaware of the group that yields the result. Now, the objective is to find the group that yields this result. For this, we apply the EM procedure. The process is an iterative mechanism and the implementation of EM is given below.

**Table-5.** Formulation of RQ2.

Set	Value
Set 1	01011
Set 2	10011
Set 3	11000

Consider the initial probability,  $\lambda_0^A = 0.5$  and  $\lambda_0^B = 0.6$ . Now, apply the procedure for E-stage. From (3) and (4) we get,

$$A = (0.5)^3(1 - 0.5)^2 = 0.03125$$

$$B = (0.6)^3(1 - 0.6)^2 = 0.03456$$

$A^* = 0.475$  and  $B^* = 0.525$  are the two success probabilities that are normalized. Similarly, set 2 and 3 are also constructed. For set 2  $A^* = 0.475$  and  $B^* = 0.525$ . For set 3  $A^* = 0.576$  and  $B^* = 0.424$ .

Now, we start the M-stage by multiplying the normalized terms with the respective probabilities. Table-5 shows the estimation of the adjusted probability.

**Table-6.** Estimation using EM approach.

Set	Estimated values using EM
Set 1	2.85 (A) and 3.15 (B)
Set 2	2.85 (A) and 3.15 (B)
Set 3	2.30 (A) and 1.70 (B)

The new probability estimate for the two groups are given by,  $\lambda_1^A = 0.5$  and  $\lambda_1^B = 0.5$ . The process is done iteratively to optimize the feasible points. The feasible point thus obtained is  $\lambda^A = 0.55$  and  $\lambda^B = 0.4$ . This infers that Group A is a better candidate for the new project compared to Group B and also we can observe that, Group A is the optimal candidate for the set of projects.

In order to test the power of the proposed model, we deploy the model to different test scenarios. These test scenarios are projects that are previously done by the groups. Now, our objective is to test whether the proposed model predicts the group correctly or not. To achieve this, we form confusion matrix. We take 15 sets for our test case with each set containing 5 projects and perform the analysis. Table-6 shows the performance analysis for each of these trails.

**Table-7.** Performance analysis between methods.

Set (s)	MLE (%)	EM (%)
Set 1	60	80
Set 2	60	80
Set 3	60	100
Set 4	40	80
Set 5	40	80
Set 6	60	60
Set 7	40	60
Set 8	60	100
Set 9	80	100
Set 10	80	80
Set 11	60	60
Set 12	40	100
Set 13	40	80
Set 14	60	80
Set 15	80	100
Net Accuracy Estimate	57.33	82.67

From Table-6, we observe that, EM approach performs better as it is an iterative procedure for finding the right candidate for the task. The net accuracy of the EM approach yields a value of 82.67%.

## 6. CONCLUSIONS

In this paper, we propose a research model which employs EM and MLE approaches for solving common and challenging issues in the organization. These are framed as RQ1 and RQ2 which are solved using the proposed research model. To further test the efficacy between methods, we take up a common question and test for different trials. We form confusion matrices to determine the accuracy for each method. The results infer that EM approach performs better compared to MLE as the method follows iterative mechanism. As a part of future work, different research models deploying different methods can be proposed and other challenging questions like strategy selection, policy making and performance optimization can be tackled. Also, fuzzy based approaches are planned for effective handling of vagueness and uncertainty in decision process.

## ACKNOWLEDGEMENT

We would like to thank the funding agency, University Grants Commission for their financial support for the research from Rajiv Gandhi National Fellowship of Grant no: F./2015-17/RGNF-2015-17-TAM-83. We also thank SASTRA University for providing us with an excellent infrastructure to carry our research work. We also thank the participants for spending their valuable time for our research work. We thank the editor and the



anonymous reviewers for their constructive comments which helped us modify our work into an effective research.

## REFERENCES

- Adair J. 2004. The John Adair Handbook of management and Leadership. Leadership. p. 242.
- Assudani R.H. 2011. Role of familiarity in affecting knowledge gaps in geographically dispersed work. IEEE Transactions on Professional Communication. 54(3): 314-332.
- Blue A. et al. 2013. Interprofessional Teamwork Skills and Attitudes as Predictors of Clinical Outcomes in a Simulated Learning Setting. Journal of Interprofessional Care. 27: 161.
- Castillo M., Petrie R. & Torero M. a. 2008. Beautiful or White? Discrimination in Group Formation. SSRN eLibrary.
- Chan C.L., Jiang J.J. & Klein G. 2008. Team task skills as a facilitator for application and development skills. IEEE Transactions on Engineering Management. 55(3): 434-441.
- Delaney J.T. & Huselid M. a. 1996. The impact of human resource management practices on perceptions of organizational performance. Academy of Management Journal. 39(4): 949-969.
- Do C.B. & Batzoglou S. 2008. What is the expectation maximization algorithm? Nature biotechnology. 26(8): 897-899.
- Fagerholm F. et al. 2015. Performance Alignment Work: How software developers experience the continuous adaptation of team performance in Lean and Agile environments. Information and Software Technology. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0950584915000269>.
- Fernández-Sanz L. & Misra S. 2012. Analysis of cultural and gender influences on teamwork performance for software requirements analysis in multinational environments. IET Software. 6(3): 167.
- Garcia P. et al. 2013. Automatic detection of team roles in computer supported collaborative work. IEEE Latin America Transactions. 11(4): 1060-1065.
- Giordano G. & George J.F. 2013. The effects of task complexity and group member experience on computer-mediated groups facing deception. IEEE Transactions on Professional Communication. 56(3): 210-225.
- Hausknecht J.P., Day D. V. & Thomas S.C. 2004. Applicant reactions to selection procedures: An updated model and meta-analysis. Personnel Psychology. 57(3): 639-683.
- Kabak M., Burmaoğlu S. & Kazançoğlu Y. 2012. A fuzzy hybrid MCDM approach for professional selection. Expert Systems with Applications. 39(3): 3516-3525.
- Kamberg M.-L. 2001. The dynamics of team interaction. Women in Business. pp. 42-45.
- Karsh N. & Eyal T. 2015. How the Consideration of Positive Emotions Influences Persuasion: The Differential Effect of Pride Versus Joy. Journal of Behavioral Decision Making. 28(1): 27-35. Available at: <http://doi.wiley.com/10.1002/bdm.1826>.
- Krishankumar R. & Ravichandran K.S. 2016a. A Novel Trio Combo Strategy for Efficient Team Formation using Hybrid Triangulation Mechanism. ARPJ Journal of Engineering and Applied Science. 11(5): 3041-3048.
- Krishankumar R. & Ravichandran K.S. 2016b. Optimal Pairing of Teammates for Enhancing Communication Rates in Software Projects using Ant Colony Optimization Approach. ARPJ Journal of Engineering and Applied Science. 11(5): 2939-2944.
- Lee G., Espinosa J.A. & Delone W.H. 2013. Task environment complexity, global team dispersion, process capabilities, and coordination in software development. IEEE Transactions on Software Engineering. 39(12): 1753-1771.
- Little A.C. et al. 2007. Facial appearance affects voting decisions. Evolution and Human Behavior. 28(1): 18-27.
- Mihalcea A. 2014. Leadership, Personality, Job Satisfaction and Job Performance. Procedia - Social and Behavioral Sciences. 127: 443-447. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1877042814023787>.
- Rathore H., Badarla V. & J, G.K. 2016. Sociopsychological trust model for Wireless Sensor Networks. Journal of Network and Computer Applications. 62: 75-87. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1084804515002714>.
- Safari S., Karimian M.V. & Khosravi A. 2014. Identifying and ranking the human resources management criteria influencing on organizational performance using MADM Fuzzy techniques. Management Science Letters. 4(7): 1577-1590. Available at: [http://www.growing-science.com/msl/Vol4/msl\\_2014\\_154.pdf](http://www.growing-science.com/msl/Vol4/msl_2014_154.pdf).
- Schmit M.J. & Ryan A.M. 1993. The Big Five in personnel selection: Factor structure in applicant and



nonapplicant populations. *Journal of Applied Psychology*. 78(6): 966-974.

Thorndike R.L. 1949. *Personnel Selection*.

Zhang L. & Zhang X. 2015. SVM-Based Techniques for Predicting Cross-Functional Team Performance: Using Team Trust as a Predictor. *IEEE Transactions on Engineering Management*. 62(1): 114-121. Available at: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84920687130&partnerID=tZOtx3y1>.