International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Identification of Least Risk Path using GA-SVM for the Software Project Management

K Amandeep Singh, T.V. Ananthan

Abstract*: Risk management is an important part of the development cycles for high quality applications. Most specific threats are incidents that may adversely affect the plan or organizational climate growth. The major risk factors contains time, budget and resources can affect adversely by events. Important considerations such as plan, time and cost are generally impacted. Essentially, risk assessment includes recognizing, assessing, preparing and monitoring incidents that affect the atmosphere of the project. Risk is the danger of volatility, lack of knowledge regarding events, activities and lack of appropriate technologies for managing measures and activities. Therefore, both exogenous and endogenous influences contribute in the venture risks and uncertainties. The high task failure rates due to poor planning of project which can limit the teams and future wealth creation, while project managers should allow for the plan being to anticipate potential risks when preparing their project achievements based on their own past experiences. This paper addresses the Supervised Learning mechanism with multi-label Support Vector Classifier (SVC) to predict the project risks and apply Genetic algorithm for providing avoidance action as recommendation.

***Keywords* : About four key words or phrases in alphabetical order, separated by commas.**

I. INTRODUCTION

Progressively companies are based on Information Systems (ISs) to enhance business operations, promote decision-making in leadership, and execute business strategies. Dependency has grown in the current business climate and digitally a number of transactions containing trade in services and goods are being conducted. Engineering projects including those centered on buildings, bridges, viaducts, highways, railways, ports and other structures in the building industry are complex product and center on future delivery [1]. In general, signed contracts projects among the controller and financier play a role of significant in-charge for the construction. In the complex projects in future deliveries which are subjected for deviations among specifications over early plan and get some case studies in the executed projects [2].

System risk management includes six main strategies as phases, including risk detection, risk analysis, risk prioritization, risk resolution, risk management scheduling, risk control computer and error forecasting is part of the risk identification process and can aid to detect system failures at an early stage, enabling efficient allocation of test assets and optimizing software architecture[3]. A number of identification approaches are developed and implemented to classify software defects as a major data mining activity. Software defect prediction is usually modeled by categorizing software units using historical data [4] as either Fault-Prone (FP) or Non-Fault-Prone (NFP) are the two-group classification problem. The dynamic context in which an organization exists implies a high importance to the risk management process. The necessity of risk management comes from the organizational medium analysis which is in a continuous transformation. Risk management means to identify, analyze, evaluate and then treat all significant risks in order to create a framework which sustains the achievement of organizational objectives [5]. Therefore, this research has proposed Genetic Algorithm with the combination of SVC in order to identify and rank the least risk path for the better execution of projects.

II. LITERATURE REVIEW

In the model of software defect prediction proposed CBA-SVM has the ability of non-linear computation of bat algorithm (CBA) with centroid strategy based on optimization efficiency. SVM model's output depends largely on parameters based on C and σ which uses RBF as its kernel feature. Yet SVM design typically approves the approach of trial-and-error to evaluate its parameters, which can technically ignore direction. Therefore, CBA to refine SVM parameters to improve the prediction model's accuracy. The goal of optimization algorithm as a particle in CBA which has to treat a pair of SVM parameters, changing particles themselves until the algorithm hits its terminal condition. The calculating particle frequency under this pair of parameters is the precision of the prediction model [6]. This article discusses the structure of Naive Bayesian Classification which is one of the approaches of Supervised Learning. The essence of Naive Bayesian identification is to test the variables and in many complicated real-world situations it frequently achieves better. The Bayesian method is attempting to more reliably and comprehensively classify the computer risk.

Manuscript published on November 30, 2019.
* Correspondence Author
K Amandeep Singh*, Research Scholar Department of Computer Science and Engineering. Dr. M.G.R. Educational and Research Institute,Chennai 600095, Tamil Nadu, India.
T. V. Ananthan, Professor, Department of Computer Science and Engineering. Dr. M.G.R. Educational and Research Institute, Chennai 600095, Tamil Nadu, India.

Cross-validation research has shown that the reliability of the equation has very strong and that the risk factors can be correctly represented by input variables. Due to the complex spatial and temporal aspects of risk management, this design implementation utilizes a variety of approaches and information, which are still some work to be done in future research. This measures computer threat to combat imitation by using information's spatial and temporal characteristics [7]. This paper argues that there are significant consequences for several system factors in the selection of risk quantification approach. We are reviewing existing risk management complexity models and proposing an extension to meet criteria centered on the choice of quantification method based on other system parameters and to advise the choice of quantification method tested by 3 case firms. This accomplished the objective by linking RM maturity concepts to previous research on product development, project management, and RM methods, deriving five categories to guide practitioners in the choice of the appropriate method. Our framework advances the state of the art by taking into account the quality of the available data, the corporate culture and awareness of risk, and the way responses are planned. We preliminarily tested the validity of our approach in three different companies, showing its value in tailoring RM to the specific needs and challenges of the companies [8]. In this paper used for the Software Cost Estimation (SCE) some meta-heuristic Bat Algorithm (BA) and GAs that researchers have already used in combination with other SCE algorithms in [9, 10] GAs. BA in SCE was proposed and a model based on enhancing GA with BA was suggested [11]. A highly recommended tool for cost and schedule control is EVM of planned research, actual completed research, an estimate of costs at completion and the estimate of costs to complete the research [12][13][14]. Borges [15] proposed to demonstrate that deliberate schedules to guide the implementation of engineering research can be calculated and controlled through EVM technique.

III. RESEARCH METHOLODGY

In the proposed work, the project is split into stages and specific procedures of performing an empirical analysis of the venture whereas the proposed solution aims to use multi-label SVC to predict possible project risks, and Genetic Programming to provide recommendations on actions to prevent the slippage. The research team coordinates with project teams from few well-known software service provider companies to gather project data like identified risks, tracked smilestones, technology complexities, environmental factors etc. every day. This will enable the research team gather metrics data such as schedule deviation and variance, effort deviation and variance, cost deviation and variance, and defect density. The proposed block diagram has been illustrated given below in fig.1.

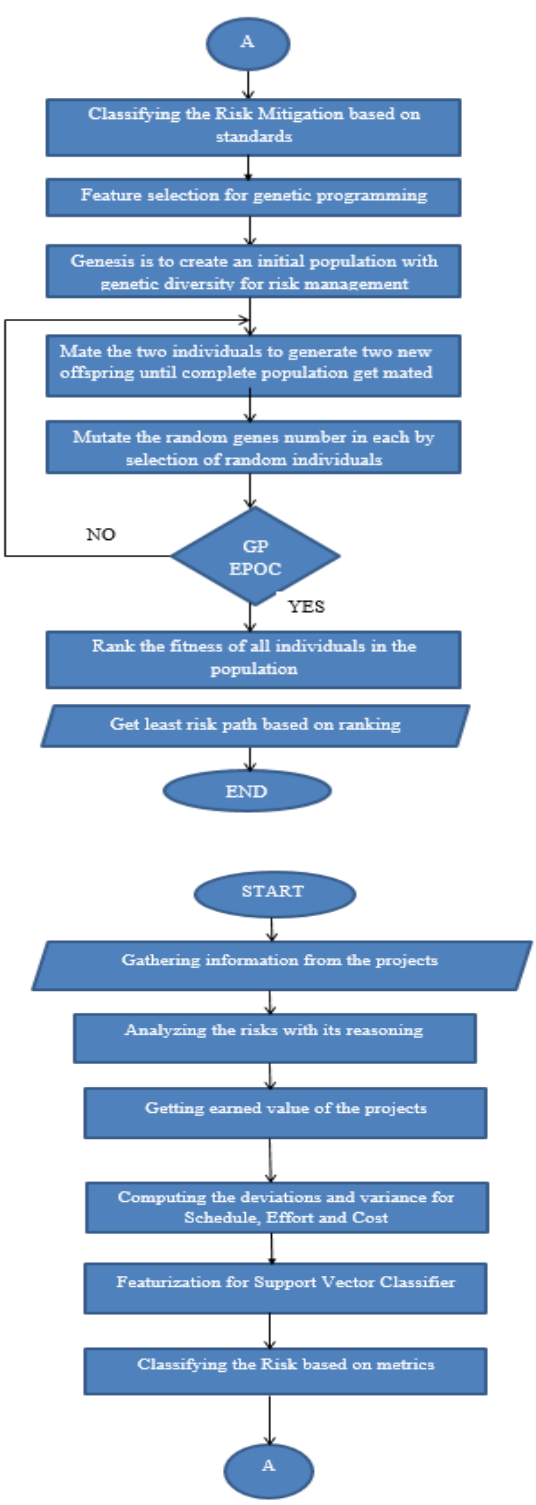

Figure.1 Flow diagram for proposed model

An SVM is a mathematical classification method that initially maps through several kernels in a high-dimensional space by non-linear information and then tries to find the hyperplane dividing data with the highest distance from this hyperplane. Once the project information has been collected from the organization, the analyzing the risk present in the project is explained with reasoning and the earned value of project which get computed using various formulas namely

A. Schedule variance

Schedule variance is calculated by taking the Planned Value of your Project from the Earned Value whereas in other word any variation among the completion of schedule for an activity and its exact completion.

$$\text{Schedule variance} = \frac{((\text{Actual days} - \text{Planned days}) + \text{Start variance})}{\text{Planned days}} * 100$$

B. Effort Variance

Effort variance provide the variation among actual effort and the estimated effort which can be manipulated to the stages of project whereas the variation among actual effort t and planned effort to all kind of phases in the project.

$$\text{Effort variance} = \frac{(\text{Actual effort from the phase} - \text{Planned effort for a phase})}{\text{Planned effort from the phase}} * 100$$

C. Cost Variance

This metric is figured by manipulating the variation among the actual cost and the earned value.

Cost

$$\text{Cost variance} = \textit{Earned value } (EV) - \textit{Actual Cost } (AC)$$

D. Defect Density

This metrics is said to be detected number of defects present over software while development divided by the software project size.

$$\text{Defect Density} = \frac{\text{Total defects Number}}{\text{Project Size}}$$

In general, SVM has utilized for two kinds of classes which are adequate to this functions whereas the performance of data classification is considered as fault-finding feature of SVC to classify the risk observed in the metrics. Therefore, the responsibility of SVM is not to be fastened over local minima. In order to determine the classification accuracy through experimental the proposal model and even permit input data of SVM has been directed on a particular field which act as an approximate implementation of the "Structural risk Minimization". The training algorithm of SVM structure is depend upon the inner multiplication care among the support vector namely X_i and the vector "X" that is derived from input space. Hence, the little amount of training data subgroup gets extracted using an algorithm called SVM. The research team provides weights to each parameter (including the metrics) and develops a feature set. They should apply the multinomial SVM on this feature set. They should apply various kernel tricks and evaluate which kernel works better. The mitigation of risk identified for each metrics based on the standards and also from earlier case studies of projects used in the organization. Once this process is completed, the major ranking of least risk identifier of Genetic Algorithm on the same feature set and knowledge base of known solutions at every stage in the development life cycle and train the model. The GP technique will recommend optimal solution from the known knowledge base of solutions. The knowledge base of known recommendations / solutions should be kept very large and the general idea behind genetic algorithms steps as follows.

Step 1: Initializing

According to the dataset for the respective project with probable solution as initial population has been created and the possible solution may be produced randomly. However, the initial population used for this project is created as 1 chromosome consists of genes strings and it represents one possible solution to the problem are modified with random value as 1 gene with respect to population(X_0).

Step 2: Crossover

Before get in to this process the individuals are selected by its individual fitness value. Once the crossover process has created novel individual from the selected population member whereas the crossover occurred has been represented with its crossover rate that can be choose by analysis of sensitivity. However, the utilization of crossover with one point is positioned randomly over chromosome has been chosen as crossover point whereas reproduction of two offspring is formed from the two parents which get selected. When both parents get spliced at an equivalent point, the generation of one offspring from the first parent head and other offspring from the second parent tail. Hence, the population reproduction for subsequent generation require a privilege of chromosome with high and effective fitness but it act as an essential for preserving the diversity of population. Thus, the chromosomes get iterated while the individuals with less fitness assists for analysing novel solution and also dodge the convergence of population over local choice.

Step 3: Mutation

In order to minimize the population diversity performance selection process is utilized whereas the utilization of mutation rate provided for maintains heterogeneity degree over solution has assisted to dodge the convergence of premature population. Once the mutation gets selected, the randomly selected value for gene present on chromosome has been altered slightly. However, the mutation probability for a chromosome in this research is 0.1%. During the occurrence of mutation, novel state for gene gets choose randomly from the feasible states.

Step 4: Ranking of Fitness function

The evaluation of fitness function based on chromosome as an input with individual and providing value which is proportional for an output as optimum which ranked the fitness of all individuals in the population.

This process is a cyclic process once the genetic programming epoch get accomplish by meeting the fitness evolved from the chromosomes. If the desired fitness is not achieved then it go to **step 2** process of crossover to all individuals are in order to accomplish the fitness. Thus the risk of software projects can be manipulated using SVC as Risk value and Riskless value that can be evaluated using population datasets.

IV. RESULTS AND DISSCUSION

In this paper, the datasets of an organization based on projects with multi-label classification is done through SVM whereas the datasets from the entire department depending upon the projects get segregated as data preparation and are used for the identifying the least risk path from the risk factor using proposed method.

The checklist is designed in accordance with a risk matrix that is entirely varies from usual questionnaires and build using matrix mode. These matrix modes assist through Genetic Algorithm as the concept of GP EPOCH of individual risk factors and their population impacted on a project is considered as parallel. The scaling the risk factor with rating of 10 for each risk factor whereas the checklist consist of 30 questionnaire data of the 20 risk factors as the input data is shown in Table 1 source for SVC and output uses the risk fitness value in project management for each project as classification recognition.

Table 1 Summation of Fitness value based on risk factors

Population	Risk Factor					Fitness Value
	Risk 1	Risk2	Risk 3	…	Risk 20	
Population 1	3	6	8	…	9	181
Population 2	9	4	5	…	2	168
Population 3	8	7	5	…	4	149
Population 4	2	4	3	…	2	74
…	…	…	…	…	…	…
Population 30	6	4	7	…	8	130

When the maximum risk factor is recognized from each risk factor with recognition accuracy rate is obtained using import of the risk factor based on combination screened by the iterative calculation of GA with SVC for ranking the least fitness value as the combination of high importance. During the recognition of SVC, the optimum parameter combination of the RBF kernel function (for SVM) in the training set is searched using a Genetic algorithm, and parameter combination (C=6, γ =0.0078) is used to set the GA with SVC model to identify the least risk fitness value for the project are shown in the figure2. The figure 2 has illustrated that population 17 consists of least fitness value is 68 which act as the optimum risk path for the project.

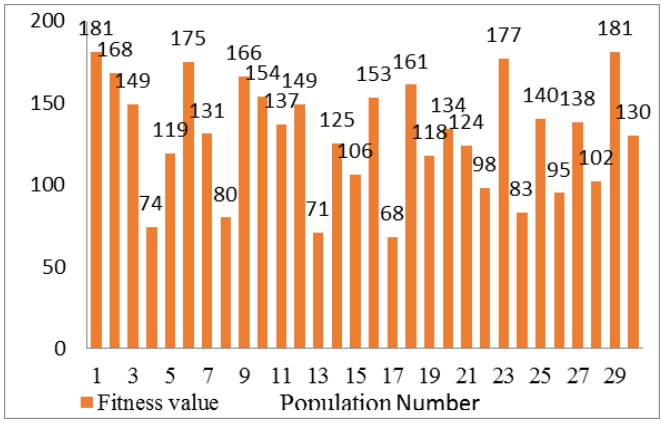

Fig.2. Risk Fitness value for the 30 populations

V. CONCLUSION

The risk management process of the project with group decision making as a decisions of milestone have not defined by personal or few expert whereas project management is a team work which needs labor force along with experiences and professional knowledge for each stakeholder with dissimilar attitudes, personalities, perceptions, motivations and feed back in order to fuse the team efficiently. This paper proposed GA with SVC to integrate the individual working standard using the checklist questionnaires to identify the quantitative and qualitative information in a risk matrix mode and imported in to SVM model for calculation, optimal and least risk path can be accomplished. Hence the least risk path get screened out whereas every risk factor can be discovered for reducing the probability of overall project decision-making mistakes to improve the project success rate. The comparison of evaluating the proposed model with existing will be focused as the future work.

REFERENCES

1. J. Batselier and M. Vanhoucke, “Construction and evaluation framework for a real-life project database,” Int. J. Proj. Manag.vol. 33, no. 3, pp. 697–710, 2015.
2. Natrayan L, Senthil kumar MS, “Optimization of tribological behaviour on squeeze cast Al6061/Al2O3/SiC/Gr HMMCs based on taguchi method and artificial neural network,” Journal of Advanced Research in Dynamical and Control Systems,11(7), pp. 493-500.
3. Y. Peng, G. Kou, Y. Shi, and Z. Chen, “A descriptive framework for the field of data mining and knowledge discovery,” Int. J.Inf. Technol. Decis. Mak., vol. 7, no. 4, pp. 639–682, 2008.
4. L.Natrayan, P. Sakthi shunmuga sundaram, J. Elumalai. Analyzing the Uterine physiological With MMG Signals using SVM, International journal of Pharmaceutical research, 2019, 11(2); 165-170.
5. P.Sakthi Shunmuga Sundaram , N.Hari Basker , L.Natrayan. Smart Clothes with Bio-sensors for ECG Monitoring, International Journal of Innovative Technology and Exploring Engineering, Volume 8, Issue 4, 2019, Pages 298-301.
6. Natrayan, L., M. Senthil Kumar, and Mukesh Chaudhari. "Optimization of Squeeze Casting Process Parameters to Investigate the Mechanical Properties of AA6061/Al_2O_3/SiC Hybrid Metal Matrix Composites by Taguchi and Anova Approach." Advanced Engineering Optimization Through Intelligent Techniques. Springer, Singapore, 2020. 393-406.
7. Kumar, M. Senthil, et al. "Processing and characterization of AA2024/Al_2O_3/SiC reinforces hybrid composites using squeeze casting technique." Iran. J. Mater. Sci. Eng 16.2 (2019): 55-67.
8. S.Yogeshwaran, R.Prabhu, Natrayan.L, Mechanical Properties Of Leaf Ashes Reinforced Aluminum Alloy Metal Matrix Composites, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 13, 2015.
9. Kaur, J.; Singh, S.; Kahlon, K. S.; Bassi, P., "Neural Network – A Novel Technique for Software Effort Estimation",
10. International Journal of Computer Theory and Engineering (IJCTE), vol. 2, no. 1, pp. 17-19 (2010).
11. Maleki, I.; Ebrahimi, L.; Gharehchopogh, F.S., "A Hybrid Approach of Firefly and Genetic Algorithms in Software Cost Estimation", MAGNT Research Report, vol. 2, no. 6, pp. 372-388 (2014).
12. L.Natrayan, V Sivaprakash, MS Santhosh, Mechanical, Microstructure and wear behavior of the material AA6061 reinforced SiC with different leaf ashes using advanced stir casting method, International Journal of Engineering and Advanced Technology, 2018,8(2S);366-371.
13. M. Ahmedabad, “Analysis of ‘Estimate at Completion’ of a Project’s duration to improve ‘Earned Value Management
14. System’ 1,” vol. 2, no. 10, pp. 1056–1061, 2013.
15. G. M. . Farok and J. A. Garcia, “Developing Group Leadership and Communication Skills for Monitoring Evm in Project Management,” J. Mech. Eng., vol. 45, no. 1, pp. 53–60, 2015.

16. F. Acebes, M. Pereda, D. Poza, J. Pajares, and J. M. Galán, "Stochastic earned value analysis using Monte Carlo simulation and statistical learning techniques," Int. J. Proj. Manag., vol. 33, no. 7, pp. 1597–1609, 2015.
17. Waldevique Franco Borges Jr, "Schedule Variance: A Monetary Value to Determine Time Variance in Construction Project", International Business Management, vol.12,no.6, pp.244-252,2018.

AUTHORS PROFILE

Mr. Amandeep Singh K has completed B.Tech at Saveetha School of Engineering and M.Tech Computer science and Engineering in SRM Kattankulathur. Currently he is pursing research in Dr. M.G.R. Educational and Research Institute. He also published two papers in International Journals and five papers in National conferences.

Dr. T.V. Ananthan is currently working as Professor in Dept. of Comp. Sci. & Eng. and Inf. tech., Dr. M.G.R. Educational and Research Institute, Chennai. He has 26 years of teaching experience. He completed his Doctorate of Philosophy in the year 2012. He has published more than 40 papers in various International and National Journals and Conferences. His area of research is Networking, Wireless Communication, Mobile Communication and Network Security. He is a member in IEEE and IEI. He is also playing the role of reviewer in International Journals.