Proposing a hybrid approach to predict, schedule and select the most robust project portfolio under uncertainty

 $\mathbf{M}.\mathbf{R}.\mathbf{Ghodoosi}^{1^{*}},\mathbf{Ramin\, \mathbf{Maftahi}^{2}$, $\mathbf{VahidReza\, \mathbf{Yousefi}^{3}}$

¹Information Technology Management, Management Faculty, University of Tehran, Iran; **²**Information Technology Management, PayamNoor University, Iran; **³**Architecture Faculty, University of Tehran, Iran *E-mail: ghodoosi@ut.ac.ir

Received for publication: 22 July 2016. Accepted for publication: 10 December 2016.

Abstract

Suitable project portfolio selection in inconsistent economy that can reduce the portfolio risks and increasing utilities for investors has gained significant research attentions. This article addresses the project portfolio selection in which conventional certain (1) prediction, (2) optimization and (3) clustering approaches cannot be used to face uncertainty. To predict the real value of affecting project risk parameters, neural network has been used; Then to determine the optimized sequences and procedures, the proposed model have been evaluated using the multiobjective shuffle frog leaping algorithm (SFLA) by robust optimization approach; To suggest different risk criteria, K-means algorithm utilized to categorize the candidate projects and differentiating the clusters. As the proposed hybrid methodology is studied on 420 different construction projects in an Iranian construction company in two economic stable years and an instable year in Iran real estate market. The results show 96 percent prediction-optimization capability due to different desired criteria.

Keywords: project portfolio selection; uncertainty; neural network; multi-objective shuffle frog leaping algorithm; K-means; robust optimization.

Introduction

Projects and its concepts as the means of development and goal achievements for organizations have been proposed in industrial era (Salo et al, 2011). In such situations, although project-based enterprises are formed to do their routine processes, projects have been introduced to maintain developing actions; Many success and failure stories have been told to share the experiences while to face uncertainty (Mohaghegi et al, 2015), asking the expert's opinion have been suggested widely (Lourenço et al, 2012- Carazo et al, 2010- Tavana et al, 2015).

Although these approaches can be accepted in sustainable economies, in in-sustainable situations can be an important source of risk for investments (Melina et al, 2016 -Gładysz et al, 2015). Reviewing the literature of prediction and optimization shows many novel and accepted certain and uncertain methods while the literature gap in hybrid methodologies simultaneously risk predication, optimization and clustering is apparent (Liesiö and Salo, 2012- Düzgün, Thiele, 2010- Arasteh et al, 2014).

In section 2, the literature review of the close articles and researchers have been reviewed. The third section describes the hybrid methodology, notations, models and used tools. In section four, the results obtained by the methods have been reported and analyzed; While the fifth section, conclude the results and suggests the forthcoming directions of possible researches in this research extension.

Literature review

To develop a novel hybrid methodology for simultaneous parameter prediction, optimization of project tasks and selecting the project portfolio based on risk is a necessity.

Project outcome and parameter prediction

Cepra et al (2015) believe that while classification models are widely accepted in scientific methods but little tries to project outcomes predictions have been done. They have proposed classification models to predict software outcomes. The results show that the prediction models are appropriate to forecast the outcomes and parameters of such projects. Hu et al (2015) also have used prediction models to estimate the project risks occurred during an outsourced software project. They have used SVM and decision tree to find the rate of failure and estimating the time and cost of such projects. Walters et al (2015) have suggested using artificial intelligence methods to forecast project durations. They have used a combination method that contains Monte-Carlo simulation, earnedvalue system and AI-based algorithms. Son and Kim (2014) proposed a data mining approach to predict the success or failure of green building projects. The Proposed SVM, neural network, decision tree and logistic regression algorithms have been compared and results show significant differences. Chora et al (2013) has tried to improve the prediction accuracy by a two-phase forecasting system combined with fuzzy C-means and SVM optimized by genetic algorithm (GA). The observed and tested results showed efficient credibility to be utilized in such area. Cheng et al (2010) by proposing a support vector machine inference system, estimated the 415 construction projects success rate. The results gained by empirical studies shows sufficient validity to use such method.

Portfolio risk optimization

Borthen et al (2010) by supposing interdependencies among IT and software projects, proposed an optimization model to select the projects portfolio. They believe in that their approach would be helpful and meaningful for IT managers to decide effectively. Liesiö and Punkka (2014) have studied multi-attribute analysis based on projects baselines and used computational methods to deal with this approach. Fernandez et al (2015) believes that the performance of algorithms in selecting the project portfolio are questionable. They have proposed a new version of ant colony optimization (ACO) to overcome such problem. The results show good impacts of such version on different sizes success. Tavana et al (2015) also suggested a hybrid method contains different risk criteria. They have used multi-criteria decision making (MCDM) approach to select the portfolio. The efficiency of proposed methods have been verified under a real world case study. Hassnzadeh et al (2014) used a robust optimization model to decide the project portfolio in pharmaceutical research and development outsourcing. A mathematical programming model have been proposed and a robust optimization approach have been conducted. Hassenzadeh et al (2014) also conducted R%D project portfolio optimization research to model and evaluate different objectives under uncertainty. The robust approach is then applied to investigate the most desirable solutions. Rabbani et al (2014) proposed a multi-objective version of PSO algorithm for solving the cost and time risk simultaneously. Rafiee et al (2014) also studied the stochastic programming of project portfolio selection and scheduling. Browning and Yassine (2014) have investigated the priority rules for product portfolio selection with the aim of minimizing the delays on many projects. Shou et al (2014) used multi-agent systems as a means of project portfolio selection and scheduling. Mild et al (2015) also provided a multi-criteria as multi-objective model assuming projects interdependencies and uncertainty to select the most robust solution. Fliedner and Liesiö (2016) have studied uncertain project parameters under multi-attribute project portfolio selection with dependencies among the projects scores. Fernandes et al (2016) also have conducted the research on the robust portfolio

selection with uncertain returns. The following table have summarized the literature in the intended aspects:

Table 1: Summerized Literature Review

As table 1 figures out, the two gaps (boxes A and B) can be seen in the literature. The Reason for such gap would be different viewpoints about the phenomena and their approaches to deal with it. Therefore, the optimization articles have tried to propose models and solving corresponding algorithms while the data-mining or artificial intelligence approaches have aimed at parameter prediction or project clustering. In order to cover these gaps, the following hybrid methodology has been proposed.

Research Methodology

As the literature review implicitly figure out, a complete methodology that conducts the investors from reviewing the project proposals to select the most profitable portfolio, a three-step

hybrid method is a necessity. The following flowchart suggests the hybrid configuration of the methods.

Figure 1: The flowchart of the proposed hybrid project portfolio risk management methodology

As the figure 1 demonstrates, the first need is to investigate and predict the affecting parameter estimation of projects tasks. To predict the most relevant parameters, three groups of data have been used to train, check and test the performance and validate the results of the formed neural network by examining them through different validation criteria. The input of this stage would be gathering the most similar completed projects during recent years, the process will result in finding the weights of neural networks arcs to show the most predictable needed resources amount, time, cost and income.

The first stage outputs will then got used by the proposed model to get optimized by shuffle frog leaping algorithm. The results will be optimized by robust optimization of different scenarios which have been predicted by the extracted neural network arcs. In such optimization method, different criteria will be tested and optimized independently. In other words, to evaluate the risks of projects in different aspects, each criteria will be chosen as the objective function which will be optimized by SFLA robust optimization version (Sarkheyli et al, 2015). As the third stage has the project risks evaluated in different criteria, a clustering algorithm (K-means) would be able to find the most suitable and sustainable project portfolio. The outcomes of portfolio can be advised to the no-taking risk investors in developing economics. The methods, model and tools are described as following:

Prediction Phase:

As a means of prediction and making relations among some inputs to gain outputs throughout a network configuration, neural network and its improved versions has been accepted

widely (Cao, 2015- Beigmoradi et al, 2014). In this paper, gathered parameters of same and most recent completed projects in the corporation or in that working region have been used as the inputs. The outcomes of prediction which has formed architecture would be able to estimate parameters. To use the prediction, adaptive neuro-fuzzy inference system is chosen.

Scheduling Phase

As the literature review has advised, the most usable approach to deal with the project portfolio scheduling problem would be mathematical programming and solving the obtained model by the metaheuristics.

Table 3: Parameters

Mathematical Model Description

To consider the beneficiary relations among different projects and optimizing the best project portfolio while considering different objectives, the following assumptions have been made:

Each project has costs having uniform distribution in occurrence.

 Each project has incomes and benefits that gained during time and each occurrence have uniform distribution.

The project implementation duration is related to allocated resources for such project.

 The competition in project finish date will result in more willingness among customers and also higher price possibility. In other words, the upper limit for benefits is related to competitiveness in project duration.

 Each project has its own type results in different demand and competition among constructors.

- The budget constraint exists.
- The number of in-process projects must not exceed a pre-defined number.

Based on these assumptions, the model tries to select projects to (I) maximize the total benefits gained by projects incomes and other advantages and (II) minimizing the total cash flow that withdraw from bank account. The first step is to present the indices, parameters and variables.

By defining the indexes, parameters and variables, the proposed model would be as follows:

C₁: Max |
$$
\frac{incommax_i-incommini_i}{(1+interest)^{(RDA_i-tz_{i,t})}}
$$
 |
\n(C₂: Min | $\frac{incommax_i-costmin_i}{RDA_i-tz_{i,t}}$ | + | $\frac{incommini-costmax_i}{RDA_i-tz_{i,t}}$ |
\n(C₃: Min VaR |
\nRDA_i = Durationmax_i - $\sum_k \gamma_k$ used_{i,k}needs_{i,k}
\n(2) $RDI_i = \text{Durationmin}_i - \sum_k \gamma_k$ used_{i,k}needs_{i,k}
\n(3) \sum_i used_{i,k}needs_{i,k} \leq available_k
\n(4) used_{i,k} \geq needs_{i,k}
\n(5) incomemin_i = PA_i - θ $\sum_i \sum_t$ company_i is_{i,j}demand_{j,t}
\n(6) incomemax_i = PA_i - θ $\sum_i \sum_t$ company_i is_{i,j}demand_{j,t}
\n(7) costmin_i = \sum_k used_{i,k} costIr_{k,t}
\n(8) costmax_i = \sum_k used_{i,k} costIr_{k,t}

௧݀ݑܤ ≥ ݔ݉ܽݐݏܿ (9)

(10)
$$
\in \{0,1\}, RDA_i, RDI_i, costmin_i, costmax_i, incomemax_i, incomemin_i \in R^+, used_{i,k} \in Z
$$

As the model describes, three different criteria have been made to evaluate the portfolio. The first and second constraints calculate the maximum and minimum duration of the ith project respectively. The third one ensures that the used amount of each resource are less than the total available amounts. The fourth constraint checks that if the usage of resources happened, the needs are suitable. The fifth and sixth constraints calculate the maximum and minimum incomes of the ith project respectively, while the seventh and eighth, calculate the maximum and minimum costs of the ith project. The constraint 9 ensures that the maximum calculated cost is less than the total assigned budget at each period of time.

The presented model is non-linear mixed integer programming model. As many portfolio optimization models in the literature suggest, the model could be categorized in the NP-Hard class. Therefore, the metaheuristics approach have been proposed to solve this problem with three different criteria in a timely manner attempts.

Solving algorithm:

 $Z_{i,t}$

To find the best solutions obtained from the assumed model, the Shuffled Frog Leaping Algorithm (SFLA) as a metaheuristics approach have been used. This algorithm combines the concepts of local search in Memetic algorithm and global search of Particle Swarm Optimization (PSO) to achieve a better answer. While the literature has shown many applications of such algorithm, less have been told about the usage in solving robust optimization problems. To use it in such a way, the scenarios resulting from the neural network estimation phase must be applied as described in Pishvaee (2011). To do so, each population of frogs have been studied under different estimated parameters to achieve the most robust solution (having the least deviation from other solutions). By the computation of the objective function, robust optimization approach have been applied and different risk criteria have been calculated. The Following Pseudo-code has been proposed relatively:

	abie entiam objective binamed 110g Ecuping Algoriumi 1 beduo code						
	Generate random population of Frogs {assigning resources to the projects as a frog}						
	FOR each individual feasible frogs do:						
	Calculate the fitness of frog by different criteria $\{C_1, C_2, \text{ and } C_3\}$.						
Initial Phase	END						
	Sort the population P in descending order of their different fitness criteria						
	Divide frogs into m memeplexes due to three different categories $\{C_1, C_2, \text{ and } C_3\}$.						
Local Search and improvement	WHILE the stopping criteria has not been met:						
	FOR each memeplex (different categories)						
	Find the best and worst frogs of that category.						
	Modify the worst frog position due to the best frog of that category.						
	IF this modification also suggests the improvement in other memeplex:						
	Accept the modification and add the resulted frog to the next generation.						
	END						
	END						
	END						
	Sort the solutions due to their categories and pass it to the K-Means algorithm.						

Table 5:Multi-objective Shuffled Frog Leaping Algorithm Pseudo-code

To determine the efficiency of such algorithm, the performance of SFLA has been compared with the results obtained from a commercial solver in the empirical result.

Selection Phase:

By achieving the different risk criteria of the sample projects, K-Means algorithm have been applied to find the segments of projects risk level and clustering them due to their similarities. At last, the cluster with the least risk level chosen as the suitable cluster has been compared with the real values of projects and their associated risks. The following section describes the empirical study and results of the proposed hybrid methodology in practice.

Empirical study and results

To evaluate the proposed hybrid method to optimize the project portfolio risk, data of some real-world projects have been gathered during a 9-year study of a construction company that has constructed 420 different projects and also evaluated many other construction candidates that have been archived. To use the data, 60 percent of the gathered data have been subjected to train, 30 percent have been used to check and 10 percent have been used to test the gained neural network. The gained network have been used to predict the 3-year (2012 and 2013 as the stable years and 2014 as the instable year) real estate data in Iran.

The following table demonstrates each step's error in different sizes of suggested projects while comparing the risks obtained by different and combined risks criteria.

Size (number) of candidate projects)	Phase One		Phase two		Phase three		criteria (rank)			Cumulative Error
	error	CPU ime	error	G ime.	error	CPU ime	C ₁	C ₂	C ₃	
30	0.3	5s	0.8	$\overline{3}$	0.7	9	1	3	$\overline{2}$	2.11250575
90	1.2	10	1.5	8	1.4	15	1	$\overline{2}$	3	4.30039884
150	1.5	14	1.8	15	1.9	27	$\overline{2}$	3		5.39651439
240	1.6	18	2.4	19	2.5	41	1	$\overline{2}$	3	7.37583747
300	1.9	22	3.2	26	4.1	49	$\overline{2}$	1	3	8.6643861
360	2.3	24	5.1	31	5.2	58	3	$\overline{2}$	1	9.6570071
420	3.9	30	6.2	37	6.1	69	1	3	$\overline{2}$	10.4623913

Table 6: The trends among different phases and different criteria

As the results show, by increasing the number of suggested candidate projects, the error and differences between criteria increases proportionally. A 4-percent error in the biggest size (420 number of construction candidate projects) shows a reasonable error level to use such hybrid methodology.

The following figure shows the trends of SFLA error in contrast with the global optimum obtained from ILOG CPLEX commercial solver.

Figure 2: Comparing the performance of SFLA with ILOG CPLEX

The calculated clusters have been figured out as follows:

Figure 3: Clusters due to first and second criteria

Figure 4: Clusters due to first and third criteria

0 20 40 60 80 100 120 140 160 180 200 Clusters due to criteria two and three

Figure 5: Clusters due to second and third criteria

Figure 6: Clusters due to all criteria

As the above figures suggest, the clusters have been determined due to different criteria. By assuming all criteria, the clusters get closer to each other. The best cluster (as shown in Figure 6) would be the blue cluster.

Conclusion and future research

This article addressed the problem of selecting the most robust portfolio in instable industries and economics. To overcome the uncertainty in parameters, unknown projects schedules and facing

different risk criteria, the hybrid methodology has been proposed. Neural network as a data mining approach to estimate parameters in the first phase, scheduling project and finding the best portfolio based on the proposed model

and evaluating the model performance by SFLA in the second phase and clustering the most suitable project portfolio based on K-Means algorithm in the third step have made up the methodology.

To evaluate the proposed methodology in practice, datasets of 420 different construction projects have been gathered. Based on project characteristics, the projects tasks have been estimated based on the trained neural network architecture. The outputs, then have been utilized to find the solutions in the mentioned model and SFLA has contributed to solve such models as a robust optimization model. The outputs then have been inserted in K-Means algorithm to find clusters based on three different risk criteria and evaluate their risks.

To further this study, it can be advised to (1) assume the decision making role for contractor as a Stackelberg game, (2) assume the value of different resources of the suppliers and contractors while (3) making a goal programming approach for finding the balance between different risk criteria. Using other data mining approaches or metaheuristics algorithms in addition to substituting other models in the second phased of the proposed methodology can make extensions and improve the results.

References

- Arasteh, A., Aliahmadi, A., Mohammadpour-Omran, A. (2014). Application of Gray Systems and Fuzzy Sets in Combination with Real Options Theory in Project Portfolio Management. Arabian Journal for Science and Engineering, 39 (8), 6489-6506.
- Bardhan, I.R., Kauffman, R.J., Naranpanawe, S. (2010). IT project portfolio optimization: a risk management approach to software development governance. IBM Journal of Research and Development. 54 (2).
- Beigmoradi, S., Hajabdollahi, H., Ramezani, A. (2014). Multi-objective aero acoustic optimization of rear end in a simplified car model by using hybrid Robust Parameter Design, Artificial Neural Networks and Genetic Algorithm methods. Computers & Fluids. 90, 123–132.
- Browning, T. R. and Yassine, A. A. (2016), Managing a Portfolio of Product Development Projects under Resource Constraints. Decision Sciences, 47: 333–372. doi: 10.1111/deci.12172.
- Carazo, A.F., Gómez, T., Molina, J., Hernández-Díaz, A.G., Guerrero, F.M., Caballero, R. (2010). Solving a comprehensive model for multi-objective project portfolio selection. Computers & Operations Research, 37(4), 630–639.
- Cao, F., Ye, H., Wang, D. (2015). A probabilistic learning algorithm for robust modeling using neural networks with random weights. Information Sciences. 313, 62–78.
- Cerpa, N., Bardeen, M., Astudillo, C.A., Verner, J. (2016). Evaluating different families of prediction methods for estimating software project outcomes. Journal of Systems and Software. 112, 48–64.
- Chang, P.T., Hung, L.T., Pai, P.F., Lin, K.P. (2013). Improving project-profit prediction using a two-stage forecasting system. Computers & Industrial Engineering. 66 (4), 800–807.
- Cheng, M.Y., Wu, Y.W, Wu, C.F. (2010). Project success prediction using an evolutionary support vector machine inference model. Automation in Construction. 19 (3), 302–307.
- Düzgün, R., Thiele, A. (2010). Robust optimization with multiple ranges: theory and application to R&D project selection. Technical report, Lehigh University, Bethlehem, PA, USA.
- Liesiö, J., (2014). Measurable Multi-attribute Value Functions for Portfolio Decision Analysis. Decision Analysis. 11, 1–20.

- Liesiö, J., & Punkka, A. (2014). Baseline value specification and sensitivity analysis in multiattribute project portfolio selection. European Journal of Operational Research, 237 (3), 946– 956.
- Liesiö, J., Salo, A., (2012). Scenario-Based Portfolio Selection of Investment Projects with Incomplete Probability and Utility Information. European Journal of Operational Research. 217, 162–172.
- Lourenço, J.C., Morton, A., Bana e Costa, C.A., (2012). PROBE A Multi-criteria Decision Support System for Portfolio Robustness Evaluation. Decision Support Systems. 54, 534– 550.
- Lowman, M., Trott, P., Hoecht, A., Sellam, Z. (2012). Innovation risks of outsourcing in pharmaceutical new product development. Technovation. 32(2), 99–109.
- Fernandes, B., Street, A., Valladão, D., Fernandes, C. (2016). An adaptive robust portfolio optimization model with loss constraints based on data-driven polyhedral uncertainty sets. European Journal of Operational Research. 255 (3), 961-970.
- Fernandez, E., Gomez, C., Rivera, G., Cruz-Reyes, L. (2015). Hybrid metaheuristics approach for handling many objectives and decisions on partial support in project portfolio optimization. Information Sciences. 315 (10), 102–122.
- Fliedner, T., Liesiö, J. (2016). Adjustable robustness for multi-attribute project portfolio selection. European Journal of Operational Research. 252 (3), 931–946
- Gładysz, B., Skorupka, D., Kuchta, D., Duchaczek, A. (2015). Project risk time management a proposed model and a case study in the construction industry. Procedia Computer Science. $64, 24 - 31.$
- Hassanzadeh, F., Modarres, M. (2009). A robust optimization approach to R&D project selection. World Applied Sciences Journal. 7(5), 582–592.
- Hassanzadeh, F., Modarres, M., Nemati H.R., Amoako-Gyampah, K. (2014). A robust R&D project portfolio optimization model for pharmaceutical contract research organizations. International Journal of Production Economics, 158, 18–27.
- Hassanzadeh, F., Nemati H.R., Sun, M. (2014). Robust optimization for interactive multiobjective programming with imprecise information applied to R&D project portfolio selection. European Journal of Operational Research. 238 (1), 41–53.
- Hu, Y., Feng, B., Mo, X., Zhang, X., Ngai, E.W.T., Fan, M, Liu, M. Cost-sensitive and ensemblebased prediction model for outsourced software project risk prediction. Decision Support Systems. 72, 11–23.
- Mild, P., Liesiö, J., Salo, A. (2015). Selecting infrastructure maintenance projects with Robust Portfolio Modeling. Decision Support Systems. 77, 21–30.
- Melina, G., Yang, S.H.S., Zanna, L.P. (2016). Debt sustainability, public investment, and natural resources in developing countries: The DIGNAR model. Economic Modelling, 52, 630–649.
- Mohagheghi, V., Mousavi, S.M, Vahdani, B. (2015) A New Optimization Model for Project Portfolio Selection Under Interval-Valued Fuzzy Environment. Arabian Journal for Science and Engineering, 40 (11), 3351-3361.
- Pishvaee, M.S., Rabbani, M., Torabi, S.A. (2011). A robust OPTIMIZATION approach to closedloop supply chain network design under uncertainty, Applied Mathematical Modelling, 35, 637-649.
- Rabbani, M., Aramoon Bajestani,M., Baharian Khoshkhou,G., (2010). A multi-objective particle swarm optimization for project selection problem. Expert Systems with Applications, 37 (1), 315–321.

- Rafiee, M., Kianfar, F., Farhadkhani, M. (2014). A multistage stochastic programming approach in project selection and scheduling. The International Journal of Advanced Manufacturing technology, 70 (9), 2125-2137.
- Salo, A., Keisler, J., Morton, A. (2011). Portfolio Decision Analysis: Improved Methods for Resource Allocation, Springer, New-York.
- Sarkheyli, A., Mohd Zain, A., Sharif, S. (2015). The role of basic, modified and hybrid shuffled frog leaping algorithm on optimization problems: a review. Soft Computing, 19 (7). 2011- 2038.
- Shou, Y., Xiang, W., Li, Y., Yao, W. (2014). A multi agent Evoluntiary algorithm for the resource Constrained project portfolio selection and scheduling problem. Mathematical Problems in Engineering, 2014, 1-9.
- Son, H., Kim, C. (2015). Early prediction of the performance of green building projects using preproject planning variables: data mining approaches. Journal of Cleaner Production, 109 (16), 144–151.
- Tavana, M., Keramatpour, M., Santos-Arteaga, F.J, Ghorbaniane, E. (2015). A fuzzy hybrid project portfolio selection method using Data Envelopment Analysis, TOPSIS and Integer Programming. Expert Systems with Applications, 42 (22), 8432–8444.
- Wauters, M., Vanhoucke, M. (2016). A comparative study of Artificial Intelligence methods for project duration forecasting. Expert Systems with Applications, 46, 249–261.