Using a Genetic Algorithm in the Optimal Investment Portfolio Decision-making Process

Timur M. Zainashev

Abstract: This article looks at optimizing an oil company's investment portfolio based on a genetic algorithm (GA) gradient search. The algorithm is then integrated into a group decision-making process.

***Index terms*: genetic algorithm, investment portfolio, vertically integrated oil company (VIOC), decision theory**

I. Introduction

COMPOSING an investment project portfolio is one of the most important stages of investment management in a vertically integrated oil company. One of the key challenges of optimizing a VIOC's portfolio composition is the large number of investment project requests. The process is further complicated by the need for long-term planning. Current computing capacities are not yet able to solve these problems quickly by using enumerative techniques alone[1]. An heuristic method of gradient search, based on a genetic algorithm, is therefore applicable [2], [3].

The GA concept is adapted from a model that occurs in the natural world. It is a method of organizing different stages of the investment process based on the principles of natural selection, resulting in optimal fitness for purpose. A VIOC fitness-function is a multi-criterion investment portfolio. It is comprised of key performance indicators (KPIs) such as net present value (NPV), capital expenditure (CAPEX), and operational expenditure (OPEX). To stick with the Darwinian analogy, a population is a set of individuals or is a set of portfolios, linked to key research area. In the same way that chromosomes identify individual species, switching a matrix identifies an individual portfolio. As evolution in the natural world leads to a stronger species, a financial analyst can create a stronger investment portfolio, selecting projects to be included, or excluded, from the portfolio using a switching matrix.

A lot of research into certain aspects of GA has already been published. This includes studies into the dependency between optimum searching efficiency and such GA parameters as crossover and mutation operators, population size and selection, and new population creation strategies. There is, however, a lack of research into the use of a GA search to optimize the decision-making process.

The decision-making process is dependent upon certain factors, which are particularly important in the context of large scale organizations, such as a VIOC. An investment portfolio must be aligned with many departments and subsidiaries and therefore it becomes a complicated political process. Transparency and objectivity are a top priority. T. K. Kravchenko distinguishes the following classifications within the decision making process in [6].

1. The number of problematic situations and the ability to assess the probability that they will occur.
2. Awareness of the consequences of taking a certain decision. The consequences of a decision are not always obvious, and the probability of potential consequences may be difficult to define.
3. Whether a decision is made by a single person or a group.
4. Whether a decision is taken based on a single criterion or multiple criteria.
5. The comparability of multiple criteria.
6. Criteria alignment (Pareto / majority / criteria sequence).
7. The alignment of experts' estimations (Pareto / majority / criteria sequence).
8. The set type of relevant variants (finite / continuum / enumerable).
9. Scale (numerical / ordinal).
10. Human involvement. Is it possible to fully automate the process or might manual correction be required?

Research areas will be grouped according to the classifications described above. The task is to find the optimal structure for a VIOC investment portfolio, taking into account the following conditions of the decision-making process:

1. There are several potential macroeconomic problems, such as the price per barrel of oil, fluctuating exchange rates and so on. Financial analysts are able to measure the probability of such events occurring and how they could impact on each other.
2. The technical and economic basis of an investment request is outlined in detail, allowing for a comprehensive calculation of the effects and consequences of an investment.
3. A decision is made by a team of technical experts and financial analysts, therefore it is a group decision.
4. The main criteria are the net present value (NPV), operational expenditure (OPEX) and capital expenditure (CAPEX) of the portfolio. Optimization is therefore a multiple-criteria task.
5. Main criteria are not comparable, therefore weight rate defines value of each.

Manuscript received December 07, 2010; revised January 04, 2011.
T. M. Zainashev was with High School of Economics, Moscow, Russia (e-mail: zainashev@gmail.com).

6. The majority principle is used for alignment of multiple criteria.
7. The majority principle is used for alignment of experts' estimations.
8. If a portfolio consists of N projects, then the maximum number of portfolios is 2^N. This number is finite, consequently, the composition of a set of portfolios is also finite.
9. Criteria are measured in real numbers. As a result, a VIOC uses an exact numerical scale.
10. The business is flexible, and the portfolio modeling is complex, hence manual corrections are also deployed.

The disadvantage of GA methodology at its current level of development is the weakness of the multiple criteria assessment of fitness for purpose. There is a description of the use of Pareto principle alignment in [4],[5],[7], however the Pareto principle is too strict for portfolio selection apart from on a majority principle. Sometimes it is difficult to find a single optimal portfolio, thereby the situation becomes a stalemate. Such strict selection leads to early convergence and finding of local extremes [8], [9]. The purpose of current research, therefore, is to develop both a model and a procedural method of decision-making regarding VIOC investment the optimal structure selection of portfolios using the majority principle.

II. Investment portfolio optimization model

This chapter describes endogenous, exogenous variable and target variables of the model.

Input variables definition is below. Basically it is endogenous, if it is not marked as exogenous.

$BasicKPIs = \{CF_B, CAPEX_B, OPEX_B\}$ – basic matrixes of revenues CF_B and expenditures $CAPEX_B$, $OPEX_B$ of the whole projects set.

$CF_B = ||CF_{jht}||$ - revenue hypermatrix of each project j in year t in problematic situation h.

$CAPEX_B = ||CAPEX_{jht}||$ - investment expenditures hypermatrix of each project j in year t in problematic situation h.

$OPEX_B = ||OPEX_{jht}||$ - operational expenditures hypermatrix of each project j in year t in problematic situation h.

$j \in \{1,\ldots,J\}$ – project counter in portfolio. J – number of project candidates to portfolio. It is not fixed. It depends on number of raised investment requests in company. J is exogenous variable.

$t \in \{1,\ldots,T\}$ – year counter in planning horizon timeline of 30 years. T is exogenous constant by the company investment regulation.

$h \in \{1,\ldots,H\}$ – problematic situation counter. H is exogenous constant and depends on number of macroeconomics exogenous factors, which are taken in consideration (exchange rate of local currency and USD, BRENT oil price, tax interest, etc).

P_h – probability of problematic situation occurrence. Condition of standardization is $\sum_{h=1}^{H} P_h = 1$.

$BasicPop = (\Gamma_1, \ldots, \Gamma_i, \ldots, \Gamma_I)$ – basic population with more, than two creatures. It is the base for initial population.

$\Gamma_i = ||\gamma_{ijt}||$ - basic switching investment matrix of portfolio i.

$\gamma_{ijt} \in \{0,1\}$ – indicator of including / excluding project j in portfolio i in year t.

$i \in \{1,\ldots,I\}$ – portfolio counter in population.

$I \in \{2,..,fixed\_size\}$ number of creatures (portfolios) in population. This number is limited by range from 2 to *fixed_size*. 2 and more parents in input basic population and *fixed_size* creatures are always in population of each GA iteration.

$Consts = \{r, int, NPV_{min}, OPEX_{max}, CAPEX_{max}, Z, W, fixed\_size\}$ – different exogenous constants.

fixed_size – constant value of population size. Only fixed number of selected creatures are able to live in population.

$r = 15\%$ - corporate standard discount rate.

$int = 13\%$ - bank interest rate.

$Z = \{Z_{npv}, Z_{opex}, Z_{capex}\}$ – criteria weight (value rate). Condition of standardization $Z_{npv} + Z_{opex} + Z_{capex} = 1$.

NPV_{min} – minimal limit of optimal portfolio discount profit.

$OPEX_{min}$ – minimal limit of optimal portfolio expenditures on operational stage.

$CAPEX_{min}$ – minimal limit of optimal portfolio expenditures on investment stage.

$W = \{w_k\}$ – set of competency constants.

$w_k \in (0,1)$ – competency of expert k (weight).

GA will select the best portfolio set. Afterwards, the investment committee will decide which one is to beput forward as an investment program. The best portfolio of population is the best from the point of view of GA formal criteria, but another one can be selected, based on the experience and knowledge of experts. Portfolios of population can be different in quality features apart from formal quantitative KPIs. Therefore we will define matrix *Eval* for experts' evaluations, and matrix will be filled after completion of GA processing.

$Eval = ||\beta_{ik}||$ – experts' estimations matrix.

β_{ik} - estimation of portfolio i by expert k.

Target output model variable is $\Gamma^* = ||\gamma^*_{jt}||$. It is switching matrix with optimal projects structure. γ^*_{jt} is indicator of including / excluding project j in year t of optimal portfolio Γ^*.

Fitness-function consist of three investment KPIs $Crit^* = \{NPV^*, OPEX^*, CAPEX^*\}$.

$NPV^* \rightarrow max$ – Net Present Value of optimal portfolio Γ^*.

$CAPEX^* \rightarrow min$ – discounted Capital Expenditures of optimal portfolio Γ^*.

$OPEX^* \rightarrow min$ – discounted Operational Expenditures of optimal portfolio Γ^*.

Each matrix (portfolio) must be calculated, taking in account the following limitations

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$Lim^* = \{ NPV^* \geq NPV_{min},\ OPEX^* \leq OPEX_{max},\ CAPEX^* \leq CAPEX_{max}\}$.

Here, we should state the main assumption of the model and the method. Costs incurred in halting and resuming a project are not taken into account. In other words, it is considered within the overall scope of financial investment. The GA algorithm can switch off financial supplementation of a project for one year, following which it can be resumed. As a result, no cash flow and expenditure will occur over the single year of such exclusion. The project is not deferred indefinitely pending financial resources, it is just switched off for one definite year. Nevertheless, the model is still valuable for supporting the decision–making process which leads to optimal portfolio structure.

III. Investment Portfolio Optimization Method

We will formulate method of optimal investment portfolio search, basing on model above. Method integrates GA and decision making process with majority principle. We will describe auxiliary model variables first. They are endogenous variables.

$\Gamma_{NG}= (\Gamma_1, ..., \Gamma_i, ... , \Gamma_I)$ - set of switching matrixes of new generation (candidates to optimal portfolio).

$\Gamma_{CG}= (\Gamma_1, ..., \Gamma_i, ... , \Gamma_I)$ - set of switching matrixes of current generation (candidates to optimal portfolio).

$Crit = (Crit_1, …, Crit_i, …, Crit_I)$ – values of criteria of each portfolio i.

$Lim = (Lim_1, …, Lim_i, …, Lim_I)$ – values of actual limits of each portfolio i.

CurGeneration = (Γ_{CG}, *Crit*, *Lim*) – the set of portfolios of current parents population with criteria and limits values.

NewGeneration = (Γ_{CG}, *Crit*, *Lim*) – the set of portfolios of new child population with criteria and limits values.

Method is defined by following function *Hybrid*.

Γ^* = *Hybrid*(*BasicPop, Lim, Crit, BasicKPIs, Consts).*

CurGeneration = Init(BasicPop, Lim, Crit, BasicKPIs, Consts) – initializes the first population of portfolios using basic set.

Do – processing loop of optimal portfolio search.

NewGeneration = *Crossover*(*CurGeneration*) – generates new creatures (portfolios) using crossover operator.

NewGeneration = *Mutation*(*CurGeneration*, *NewGeneration*) – generates additional creatures using the mutation operator.

- *NewGeneration* = *CritLimComp*(*NewGeneration*, *Crit*, *Lim*, *BasicKPIs*, *Consts*) – computes criteria and limits values (*Crit* and *Lim* accordingly) for new generation creatures.
- *NewGeneration* = *FilterOverLimitPortfoilios* (*NewGeneration*, *Lim*, *Consts*) – filters overlimited portfolios from the new population.
- *NewGeneration* = *CommitNewGen* (*NewGeneration*, *CurGeneration*, *Consts*) – defines rank of portfolios on numerical scale using the elitism principle. Afterwards function filters weak portfolios, which are not meet fixed population size. See details of function below.

CurGeneration = *NewGeneration* – new population replaces old one.

Until *StopConditions* (*NewGeneration*) = True – number of fixed iterations is reached.

Financial analytics fill matrix *Eval* for all found strongest portfolios, i.e. they estimate each found portfolio, basing on their experience and knowledge.

Γ^*= *Major* (*NewGeneration*.Γ_{NG}, *Eval*, *Consts*) – selects optimal portfolio using majority principle for expert's estimations alignment.

Thus optimal investment portfolio is calculated, using modified GA. GA modification and research novelty consist in two points. New generation selection procedure (function *CommitNewGen*) and experts' estimations alignment addition (function *Major*). The rest of the functions are described in [19], [20]. We will consider modification functions in details.

We will define auxiliary endogenous variables of *CommitNewGen* function first.

MixedGeneration = { Γ_{CNG}, $Crit_{CNG}$, $IntCrit_{CNG}$ } – set of portfolios, each single and integral criteria values of current and new generation.

$\Gamma_{CNG} = \{\Gamma_1, …, \Gamma_i, …, \Gamma_{2I}\}$ – set of switching matrixes Γ_i of current and new generation.

$Crit_{CNG} = \{ Crit_1, …, Crit_i, …, Crit_{2I} \}$ – set of single criteria values for each portfolio of current and new generation.

$IntCrit_{CNG} = \{ IntCrit_1, …, IntCrit_i, …, IntCrit_{2I} \}$ – set of integral criteria values for each portfolio of current and new generation.

Function details are the following.

NewGeneration = *CommitNewGen* (*NewGeneration*, *CurGeneration*, *Consts*) – creates new generation based on elitism rule i.e. all parents will be included to selection. Function uses majority principle for alignment of criteria taking in account problematic situations.

Loop i=1, 2*I – processing loop at projects of current and new generation.

- *MixedGeneration.IntCrit*$_i$ = Z_{npv} · *MixedGeneration.Crit*$_i$*.NPV* + Z_{capex} · *MixedGeneration.Crit*$_i$*.CAPEX* + Z_{opex} · *MixedGeneration.Crit*$_i$*.OPEX* – compute value of integral criteria for portfolio i using majority principle. Problematic situation probability is taken in account in formula of three main criteria NPV, OPEX and CAPEX below.

Where:

Net Present Value of portfolio i in mixed generation

$$NPV_i = NPV(\Gamma_i) = \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{h=1}^{H}\frac{CF_{jht}\cdot\gamma_{ijt}\cdot P_h}{(1+r)^t} - CAPEX_i - OPEX_i$$

Investment (capital) discounted expenditures of portfolio i in mixed generation

$$CAPEX_i = CAPEX(\Gamma_i) = \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{h=1}^{H}\frac{CAPEX_{jht}\cdot\gamma_{ijt}\cdot P_h}{(1+\text{int})^t}$$

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Operational discounted expenditures of portfolio *i* in mixed generation

$$OPEX_i = OPEX(\Gamma_i) = \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{h=1}^{H}\frac{OPEX_{jht}\cdot\gamma_{ijt}\cdot P_h}{(1+\text{int})^t}$$

End loop.

MixedGeneration.Γ_{CNG} = *Sort* (*MixedGeneration.*Γ_{CNG} , *MixedGeneration.IntCrit*$_{CNG}$) – sorts portfolio according to integral criteria.

NewGeneration.Γ_{NG} = *TopN* (*MixedGeneration.*Γ_{CNG} , *fixed_size*) – filters mixed generation of parents and children with fixed number *fixed_size* of strongest portfolios.

As a result we got new generation of portfolios, which will be use GA further processing. We will describe mathematical method of second GA extension below.

Γ^* = *Major*(*NewGeneration.*Γ_{NG}, *Eval*, *Consts*) - selects optimal portfolio using majority principle for alignment of experts' estimations.

Loop i=1,*I* – processing loop at last population portfolios, which was calculated as a result of GA.

$MrjEval_i = \sum_{k=1}^{K}\beta_{ik}\cdot w_k$ - calculates integral estimation of portfolio *i* among *K* experts.

End loop

$\Gamma^* = \max_{i=1,I} MrjEval_i$ - selects portfolio with maximum of integral estimation.

Where: MrjEval = {*MrjEval*$_1$, …, *MrjEval*$_i$,…, *MrjEval*$_I$} – set of each portfolio integral estimation.

The target portfolio Γ^* is calculated. This is optimal in terms of projects-structure for the multicriteria case. The time-consuming process of portfolio-composition enumeration will be computed by GA and the final decision will be made by a group of experts. The novelty of the proposed model and method consists of integration of GA techniques within the decision-making process. The functions *CommitNewGen* and *Major* demonstrate this extension.

IV. Model and Method Verification

The method consists of two heterogeneous parts: generation of optimal population using GA and optimal single portfolio selection. Therefore we will describe verification in two stages, taking optimal portfolio selection verification first and GA verification afterwards.

The model and function *Major* verification principle is the following. $\exists\Gamma^*, \forall\Gamma_i\ F^* \geq F_i$ Γ^* is the best portfolio among all candidates according to criterion *F*. The criterion *F* aligns expert estimations of each portfolio Γ_i and selects the best one using the majority principle.

$F_i = \sum_{k=1}^{K}\beta_{ik}\cdot w_k$ - value of portfolio *i* criteria. *K* experts estimated it.

$\Gamma_i \in \Gamma_{NG}$= (Γ_1,…, Γ_i,… , Γ_I) – portfolio *i* of the last created generation. It is candidate to investment program of VIOC.

Γ_{NG} = *Hybrid* (*BasicPop, Lim, Crit, BasicKPIs, Consts)* – set portfolio-candidates, which generated as result of GA search without using last step of function *Major*.

Verification of GA is impossible due to probable basis of mutation and crossover operators. This method does not guarantee a globally optimal result in a fixed number of iterations. However, if last generation Γ_{NG} is found to be a bit worse or much better, comparing to initial generation based on Schema theorem in [10]. As a result it is deemed that last generation Γ_{NG} is optimal at the moment when the search concludes.

V. Conclusion

The primary result of research is the developed and verified mathematical model and method, which integrate GA within the decision-making process of VIOC investment portfolio optimal structure selection in the following respects:

- multicriteria fitness-function case with criteria alignment using the majority principle;
- problematic scenarios are taken into account by probable macroeconomical parameters;
- group decision making process is put into place;
- alignment of experts estimations is based on the majority principle.

Thus, we contribute to the evolution of computation and decision-making theories. Further model and method improvement is yet possible in the following ways:

- to avoid assumption regarding switching-off financing of a project in any given year;
- to take into account relationships between projects in a portfolio, e.g. realization sequence and necessity, or priority, of projects included;
- to take into account multicriteria case for expert estimations;
- to realize other principles of criteria alignment and expert estimations;
- to realize model and method in software products SIS Merak Capital Planning and existing Investment Management information system based on SAP BI.

Acknowledgment

I would like to thank and express my respect for my friend and advisor K.V. Kravtsov. He has supported and inspired me and had a great deal of input in my research. I would also like to thank my supervisors A.S. Akopov and T.K. Kravchenko, whose comments were very valuable.

References

[1] A. S. Akopov, "System-dynamic approach for oil company investment management," Audit and Financial analysis. Vol.2, 2006, pp.153-188. Акопов А.С. Системно-динамический подход в управлении инвестиционной дея тельностью нефтяной компании. // Аудит и Финансовый анализ, № 2, 2006, - с. 153- 188

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[2] A. V. Eremeev, “Development and analysis of genetic and Hybrid Algorithms for discrete optimization problems solving,” Ph.D. dissertation, OFIM, Omsk, 2000. Еремеев А.В. Разработка и анализ генетических и гибридных алгоритмов для решения задач дискретной оптимизации. Дисс. канд.физ.-мат.наук. Омск, 2000.

[3] Goldberg D. E. Genetic algorithms in search, optimization, machine learning. Reading, MA: Addison-Wesley, 1989.

[4] A. V. Eremeev, “Relationship between dynamic programming and multicriteria evolution algorithms”, OFIM, Omsk, 2008. Ереемев А.В. О связи динамического программирования и многокритериальных эволюционных алгоритмов, Омск, 2008.

[5] M. Garey and D. Johnson, “Computers and Intractability: A theory of NP Completeness,” W.H. Freeman and Company Publishers, 1979.

[6] T. K. Kravchenko, “Expert systems and Decision Support Systems” in Lecture Notes, HSE, Moscow, 2005. Кравченко Т.К., Курс лекций по «Экспертные системы и системы поддержки принятия решений» (раздел 1,4) для магистерской программы ВШЭ, М.2005.

[7] Dimo Brockhoff, Theoretical Aspects of Evolutionary Multiobjective Optimization—A Review N° 7030, INRIA, Sep-2009.

[8] Darrel Whitley "A Genetic Algorithm Tutorial", 1993.

[9] F.Herrera, M.Losano, A.M.Sanches. "Hybrid Crossover Operators for Real-Coded Genetic Algorithms: An Experimental Study".

[10] Robin Biesbroek "Genetic Algorithm Tutorial. 4.1 Mathematical foundations", 1999.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)