Scientific paper/Znanstveni rad

GENETIC ALGORITHM AND OPTIMIZATION OF THE SALES ASSORTMENT STRUCTURE

Professor **BRANO MARKIĆ**, Ph.D.[1]
[1]Faculty of Economics University of Mostaru
e-mail: brano.markic@sve-mo.ba

ABSTRACT

The genetic algorithm belongs to a group of evolutionary algorithms that find inspiration in Darwin's theory of maintaining the best species. It is based on only three operations: selections, crossover and mutations. The application space is found in all areas that require optimization by finding the values of the variables that optimize the target function. Each genetic algorithm therefore uses a fitness function in order to choose the best crossover units from the population in the next generation. The optimal structure of sales assortment is a theoretical and pragmatic challenge to the sales function in every market-oriented organizational system. In this paper, the optimization of the sales assortment structure is viewed as the task of determining the share of individual products in a group of products sold on a particular market. There is a hypothesis that the structure of the sales assortment can be optimized using the genetic algorithm. In the paper is developed software solution in C# to verify the main hypothesis and the solution is open to new extensions and demonstrates a satisfying application power.

KEYWORDS: evolutionary algorithms, selection, crossing, mutation, sales assortment.

INTRODUCTION

Optimization of the sales assortment is a complex economic, mathematical, marketing, accounting and IT task. The market's "correct" choice of sales assortment structure and the quantity share of individual products in the assortment structure results in economic effects that certainly increase the economic and organizational performance of companies. Therefore, sales knowledge should always be complemented by appropriate methods and algorithms to minimize procurement errors, both in content and in quantitative terms. In order to assist with marketing procurement activities, various mathematical approaches and sales forecasting methods (ARMA models, ARIMA models, exponential smoothing, moving averages etc.) have been developed, knowledge management systems seeking to provide expert knowledge of sales in the form of knowledge base expert system. However, it is rarely seen in the literature that the choice of quantity of individual products in a given sales assortment is being monitored in order to minimize procurement risk and meet demand [2]. For solution, knowledge from more scientific disciplines is needed. In particular, marketing knowledge on procurement, demand, market, and mathematical knowledge of modeling and optimization can be distinguished [5]. There are two simultaneous goals: maximizing the margin rate realized by the sales assortment (the sales limit is the market absorption power or demand for the products in that sales assortment) and minimizing the risk of sales. It is therefore necessary to observe and analyze the market and, based on the data on the quantities sold and the margin rate of each product in the sales assortment, select a structure that will ensure satisfactory sales in the margin rate of the entire assortment and acceptable sales risk. Such research for the task of the mathematical usually solves quadratic programming.

The paper analyzes and presents a different approach to optimizing the sales assortment using a genetic algorithm. It was developed and implemented a special software solution proposing the structure of the sales assortment, and then the relative share of each product in the sales assortment. The sales assortment can consist of a group of products such as ice cream (different manufacturers and packaging), beer (different manufacturers and packaging), cosmetics, footwear etc. In theory, it is important to minimize the risk of product procurement. The risk is always directly correlated with the average quadratic deviation of the sales of the individual product in the sales assortment (or margin rate) from their average. This measure is called a variance. Larger value of variance (or it square root - standard deviation) means higher risk, and smaller standard deviation means less risk. Risk and standard deviation are directly proportional. The complexity of the problem of the optimal assortment is the consequence of economic logic that the lesser risk is always associated with the smaller margin rate, and then the lower profit or profitability. There may be more questions to ask. How to maximize the margin rate range for the pre-accepted risk expressed by the standard deviation of the margin rate in the sales assortment? How to minimize the risk for a predetermined margin rate? What is the best and most acceptable relationship between risk and expected margin rate? The logical consequence of such questions are two mathematical, quantitative goals: minimizing the risk of sales and maximizing the realized margin rate of the sales assortment. It is therefore necessary to present and analyze a mathematical model that calculates the expected return and standard deviation of the assortment as a measure of procurement risk.

1 OPTIMIZATION OF SALES ASSORTMENT

The sales assortment is made up of products whose average margin rate is denoted as $RuC_3, \ldots, RuC_n$, the individual product share in the total sales assortment $\mu_1, \mu_2, \mu_3, \ldots, \mu_n$ respectively. The expected margin rate of the whole sales assortment is (Rp):

$$R_p = \sum_{i=1}^{n} \mu_i RuC_{ii} \quad (1)$$

If the sales assortment consists of only two products, then the variance of the sales assortment is:

$$\sigma^2_p = E(RuC_i - E(RuC_i))^2 = E[\mu_1(RuC_1 - \rho_1) + \mu_2(RuC_2 - \rho_2)]^2 =$$
$$= \mu_1^2 E(RuC_1 - \rho_1)^2 + \mu_2^2 E(RuC_2 - \rho_2)^2 + 2\,\mu_1\,\mu_2 (RuC_1 - \rho_1)\,(RuC_2 - \rho_2) =$$
$$= \mu_1^2\,\sigma^2_1 + \mu_2^2\,\sigma^2_2 + 2\,\mu_1\,\mu_2\,\sigma_{12} \quad (2)$$

gdje je:
σ^2_p= the variance of the margin rate of the sales assortment,
σ^2_1= the variance of the margin rate of the first product,
σ^2_2= the variance of the margin rate of the second product,

Rp=margin rate of sales assortment,

μ_i= the share of the i-th product in the sales assortment ($\sum_{i=1}^{n}\mu_i = 1$),

ρ_1= the arithmetic mean of the margin rate of the first product in the sales assortment,
ρ_2= the arithmetic mean of the margin rate of the second product in the sales assortment,
$\mu_1 + \mu_2 = 1 \Rightarrow \mu_2 = 1 - \mu_1$,
RuC_i= the expected margin rate of the i-th product in the sales assortment.
The calculated profit margin of the sales assortment is a relatively simple task (1) while the standard deviation of the sales assortment (2) is more complex. Namely, the variance of the profit margin of the sales assortment is not the arithmetic mean of the variance of profit margins of individual products in the sales assortment. A sophisticated analysis is needed in order not to derive the wrong conclusion that the arithmetic mean variables of the profit margins of individual products of the sales assortment is the variance of the sales assortment [3]. How then is the standard deviation of the profit margin of the sales assortment calculated? Therefore, the simplest sales assortment will be analyzed first, with only the two products with the average profit margins RuC_1 and RuC_2, relative shares in sales assortment ρ_1 i ρ_2, standard deviations of weekly profit margins σ_1, σ_2 and covariances of profit margins of the first and second products σ_{12} .
The goal is to minimize the function (2). Therefore, the first function is derived and equated with zero:

$$\frac{\partial((\sigma_p^2)}{\mu_1} = 2\mu_1\sigma_1^2 + 2(1- \mu_1)\sigma_2^2 + 2(1-\mu_1)\sigma_{12} = 0 \quad (3)$$

Solving equation (3) by μ_1 we obtain:

$$\mu_1 = \frac{\sigma^2{}_2 - \sigma_{12}}{\sigma^2{}_1 + \sigma^2{}_2 - 2\sigma_{12}} \quad (4)$$

If we use the correlation coefficient as a ratio between the covariance profit margins in the sales assortment and the standard deviations, then the equation (4) can be written:

$$\mu_1 = \frac{\sigma^2{}_2 - r\sigma_1\sigma_2}{\sigma^2{}_1 + \sigma^2{}_2 - 2r\sigma_1\sigma_2} \quad (5)$$

where r is the coefficient of correlation between the profit margins of the first and second products in the sales assortment.

The required data for the calculation of the optimal shares should be shaped as a data set in which the columns represent the profit margins of the first (RuC1) and the second (RuC2) product and the corresponding equal time periods (T):

T	RuC_1	RuC_2
t_1	a_{11}	a_{12}
t_2	a_{21}	a_{22}
…..		
t_n	a_{n1}	a_{n2}

Table 1. Margin rate of sales assortment for two products

It is obvious that the share of sales volume and profit margins depend on the variations in the sales volume (ie profit margin) of the product in the sales assortment and their coefficients of correlation. Therefore, appropriate variances and coefficients of correlation are calculated. Then the result is included in equation (5) and the share of the first product is obtained. If it is assumed that the profit margin variance of the first and second products is $\sigma^2_1= (0,6)^2$ a $\sigma^2_2= (0,8)^2$ respectively, and the correlation coefficient of profit margins of the first and second product in the sales assortment r = 0.5 then the relative share of the quantity of the first and the second product can be calculated on the basis of:

$$\mu_1 = \frac{\sigma^2{}_2 - r\sigma_1\sigma_2}{\sigma^2{}_1 + \sigma^2{}_2 - 2r\sigma_1\sigma_2} = \frac{0,8^2 - 0,5*0,6*0,8}{0,6^2 + 0,8^2 - 2*0,5*0,8*0,6} = \frac{0,64 - 0,24}{1 - 0,48} = \frac{0,4}{0,52} = 0,769$$

The profit margin variance (sales volumes) of the assortment is:

$\sigma^2{}_p=\mu_1{}^2\ \sigma^2{}_1+\ \mu_2{}^2\ \sigma^2{}_2+2\ \mu_1\ \mu_2\ \sigma_{12}$ =0,769*0,6^2 +0,231*0,8^2+2*0,769*0,231*0,5*0,6*0,8=

=0,27684+ 0,14784+ 0,0853 = 0,432

The variance of the sales assortment is 43.2% while the profit margins of the product range 60% and the other 80%. If the sales assortment optimization is generalized to n products, then based on the data behind the dataset for n products:

T	RuC_1	RuC_2	…..	RuC_{n-1}	RuC_n
t_1	a_{11}	a_{12}	…..	$a_{1.n-1}$	$a_{1,n}$
t_2	a_{21}	a_{22}	…..	$a_{2,n-1}$	$a_{2,n}$
…..	…..	…..	…..	…..	…..
t_m	a_{m1}	a_{m2}	…..	$a_{m,n-1}$	$a_{m,n}$

Table 2. Margin rate of assortment for n products

can calculate the matrix covariance σ_{ij}, and the variances s of margins rates σi of each i-th product in the sales assortment.

The goal is to minimize the variance of the sales assortment:

$$\min \sum_{i=1}^{n}\sum_{j=1}^{n}\mu_i\mu_j\sigma_{ij} \quad (6)$$

with constrains

$$R_p = \sum_{i=1}^{n}\mu_i RuC_i \geq A \quad (7)$$

$$\sum_{i=1}^{n}\mu_i = 1 \quad (8)$$

The variance of the sales assortment is not just the arithmetic mean of the variance of the profit margins of the product in the sales assortment but rather of the more complex function (6).

In practical situations, the sales assortment may be very deep and there may be a large number of products in it. Therefore, it is reasonable to assume that the sales assortment is formed on the basis of the profit margin data obtained by the n product ie it is selected a certain number of products (eg 15). Therefore, in the paper, the 15 products that earn the highest profit margin in the last 60 periods (the period may be a day, a week, a month) are selected.

2 OPTIMIZATION OF THE SALES ASSORTMENT USING A GENETIC ALGORITHM

The genetic algorithm belongs to a group of evolutionary algorithms. Its key idea is to imitate the survival of the most sophisticated and most powerful individuals in the population. It implements only three operations: selection, crossover, mutation and fitness function. It "checks" every single population of the generation and chooses among them the ones that maximize the fitness function. The best individuals with their genetic material have the highest probability of crossover choices and thus reproduce the new generation (individuals of the population) with the best genetic material (genes).
The genetic algorithm has four key features [2]: parameter coding, parallelism, fitness function, the choice of individuals and their evaluation. The block diagram of the genetic algorithm steps shows the following picture:

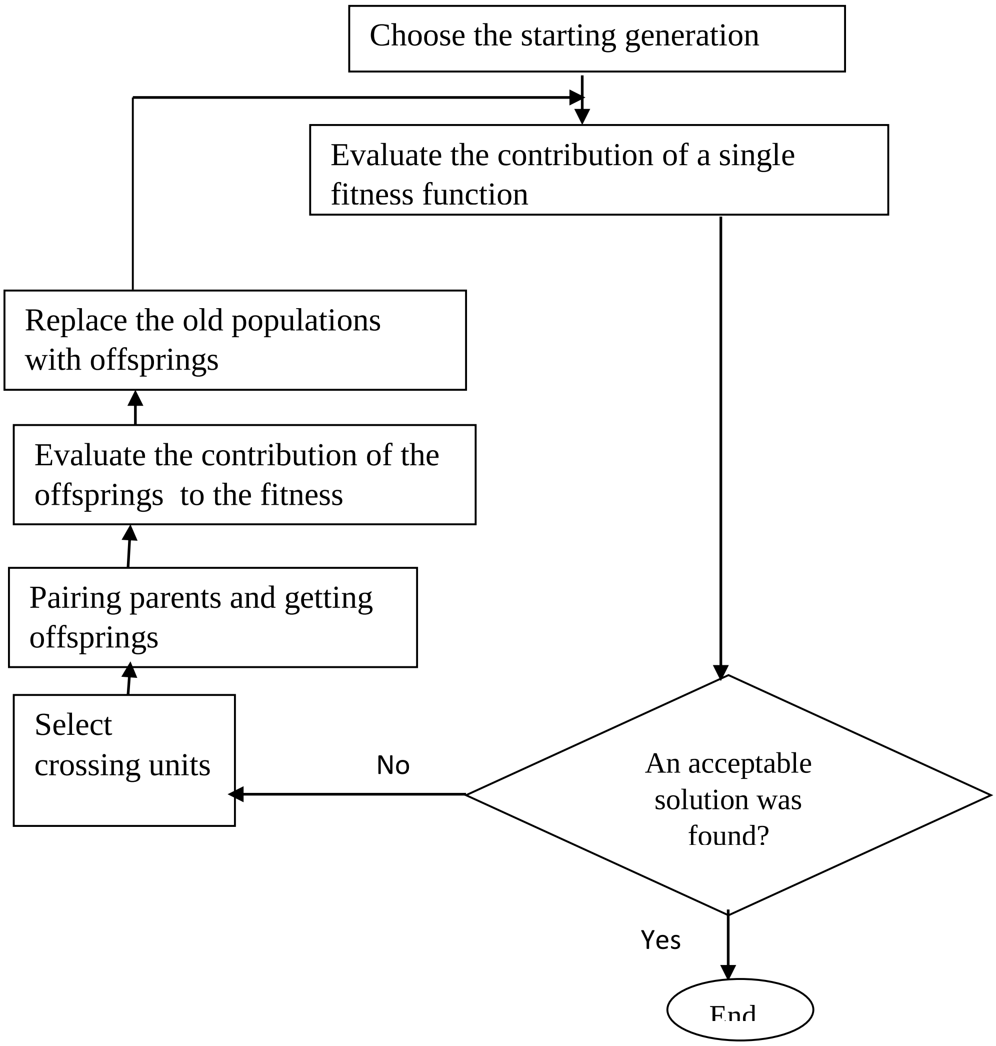

Figure 1. The steps of the genetic algorithm

Below is a brief description of the genetic algorithm's properties and a description of its steps [1].

1. Encoding parameters

Parameters are the objects of optimization of genetic algorithms. The main goal of optimization is to find the best combination of parameters for the set problem. Parameters are most often displayed (coded) with binary digits (as binary digits 0 and 1). Coded parameters can also be in a decade number system.

2. Parallelism

Genetic algorithms simultaneously analyze the points distributed in different parts of the space of possible solutions. By contrast, the standard optimization algorithms view (analyze) at one point only one point in the space of possible solutions [4]. This is an example of a linear programming algorithm that improves the function of a goal by moving from one point to another in a local area.

3. Fitness function

An essential component of the genetic algorithm is a certain function or fitness function. It is the key to the selection of individuals in solving problem because each population is a potential solution.

4. Election and evaluation rules

Space research is based on stochasticity. Thus a genetic algorithm avoids the danger of "capturing" in the local extremes. Units of the population in one generation determine the candidates for the next generation. The basic steps of the genetic algorithm make the following sequence:

1.Generate the starting generation
2. Evaluate each and every single population and her contribution to fitness function
3. Repeat:
  - Select the crossing units (parents) from population
  - Save parents and get offspring
  - Evaluate the contribution of the offspring to the fitness function
  - Replace the descendants' old population until a satisfactory solution is found [6]

3 OPTIMIZE THE SALES ASSORTMENT STRUCTURE WITH A GENETIC ALGORITHM

The optimization of the sales assortment structure with the genetic algorithm is only possible with the development of software solutions that in some programming language encode all of its steps. This paper presents a software solution in language C # ..
The first step is to choose the starting population with 15 products with the highest profit margin in the last 60 time intervals. These 15 products will be a sales assortment and the task of the genetic algorithm is to define their volumes in total procurement, whereby the ratio between the total sales volume of the assortment and the risk is maximized.
The following table shows the shares of individual products (15) in the sales assortment in the last 60 intervals (periods):

T	RuC_1	RuC_2	…..	RuC_{14}	RuC_{15}
t_1	u_{11}	u_{12}	…..	u_{14}	u_{15}
t_2	u_{21}	u_{22}	…..	u_{24}	u_{25}
…..	…..	…..	…..	…..	…..
t_{60}	u_{m1}	u_{m2}	…..	$u_{60,14}$	$u_{60,n15}$

Table 3. Shares of profit margins u_{ij} of products j in time interval i

Based on the share u_{ij} of the product j in the time interval t is calculated the covariance matrix he of the sales assortment for 15 products:

$$cov(i, j) = \frac{\sum_{i=1}^{15}\sum_{j=1}^{15}\sum_{k=1}^{60}(u_{kj} - pj) * (u_{ki} - p_i)}{15}$$

where is:
u_{kj} - array and make it a share of the profit margins of each product included in the sales assortment
p_i – the average profit margin of the i-th product for all 60 time intervals. In C #, enough to use two loops for:

```
double[,] u;
double[v];
for (j=1;j<16;j++)
        {   v=0;
                for(k=1;k<=60;k++)
                {v=v+u(k,j);}
                u(j)= Math.Round(v/60,4);
         }
```

Again, it is necessary to write down the function that calculates the covariance of the profit margins of the sales assortment based on the data from Table 1. (shown as the array data structure):

```
public void covariance()
{
double [,] cov; double [,] u; double [] p;

int i,j,k;
        for (i=1;i<16;i++)
                {
                        for(j=1;j<=60;j++)
                        {cov(i,j)=0;
                                for (k=0;k<61; k++)
                                  {cov(i,j)=cov(i,j)+ (u(k,j)-p(j))*(u(k,i)-p(i));}
                                cov(i,j)=Math.Round(cov(i,j)/15,4):
```

}
}

}

A covariance is the measure of agreeing a profit margin change in the sales assortment over the past 60 time intervals.

3.1. SELECTION AND CROSSING OPERATIONS

After calculating the share of profit margins of each product in the sales assortment (Table 1), the covariance matrix of the profit margins of the sales assortment it is neccassary o select the individuals for crossing from the population.
The selection of chromosome for crossover is performed by randomly generated numbers so that individuals contributing more to fitness function are more likely to choose [2]. The result of the selection operation is the interpopulation where two individuals generates two offsprings. They carry the parent genetic material. Therefore, it is necessary to generate randomly. The crossing point of the chromosome must be a randomly generated number. Replacing the right parts of the chromosome constitutes offsprings for the next population. The chromosome crossover illustrates the following figure:

			Crossover point													
	u_1	u_2	u_3	u_4	u_5	u_6								u_{14}	u_{15}	u_{16}
Parent 1	**0,0808**	**0,0798**	**0,0449**	*0,0724*	*0,092*	*0,0639*								*0,094*	*0,0874*	*0,0279*
Parent2	***0,1134***	***0,0477***	***0,0358***	**0,0763**	**0,0702**	**0,0448**								**0,0377**	**0,0492**	**0,0524**
Child 1	**0,0808**	**0,0798**	**0,0449**	**0,0763**	**0,0702**	**0,0448**								**0,0377**	**0,0492**	**0,0524**
Child 2	***0,1134***	***0,0477***	***0,0358***	***0,0724***	***0,092***	***0,0639***								***0,094***	***0,0874***	***0,0279***

Figure 2. Crossover operation

After crossing the third parent gene, two children are born. They need to be evaluated by the fitness function The fitness function is determined by the profit margin of the sales assortment and it should be maximized by the function of the variation of the sales assortment (the risk of sales of the assortment) that needs to be minimized. In other words, it is necessary to maximize the function of goodness:

$$\max(\text{fitnessF}) = \frac{\sum_{i=1}^{n} \mu_i RuC_i}{\sum_{i=1}^{n}\sum_{j=1}^{n} \mu_i \mu_j \sigma_{ij}} \qquad (9)$$

After 500 generations of sales assortments, a sales assortment with the best results for the fitness function (0) was chosen. Each generation finds chromosomes (unit of sale assortment) with the best ratio of expected margin rate and risk.
The result of applying the genetic algorithm for the given data set and the fitness function is shown in the following table:

Expected profit margin of sales assortment	Margin rate risk (standard deviation) of the sales assortment	Product	Product share in sales assortment (quantities)
O,2871	0,042	P_1	0,09
		P_2	0,06
		P_3	0,15
		P_4	0,04
		….	…….
		P_{15}	0,08

Tablica 4.: Optimal shares of product quantities in sales assortment

After 500 generations, the product shares in the sales assortment that give the best ratio between the expected margin of the sales assortment and the risks are:P_1=9%, P_2=6%,…P_{15}=8% with the expected margin rate 28.71% and the risk of 4.2%. The genetic algorithm and its selection, crossing, and mutation operations have generated a satisfactory solution, information needed in the procurement process. Of course, the number of generations can be over 500, which simply allows the software solution.

4 CONCLUSION

Information on the optimal sales assortment is key to the procurement process and it allows to improve the financial performance of the organizational system. Optimization simultaneously meets two goals: maximizing the expected profit margin and minimizing risk. The genetic algorithm demonstrates simplicity in the application and selection, crossing, and mutation operations to find the best chromosomes representing the product shares in the sales assortment. This paper presents some of the code functions implemented in the C # language. Experimental results with sixty chromosomes after five hundred generations show an acceptable solution. The genetic algorithm uses fitness function and optimizes product shares in sales assortments. The solution is open to extensions of the assortment depth and C # shows an acceptable developmental potential and applicability.

LITERATURE

1. Goldberg, D.E. (1999). *Genetic Algorithms in Search, Optimization & Machine Learning*, Addison-Wesley, Redwood City, CA4, 266–275
2. Kovač,I., , Palić,M.,Lovrić,M., *Primjena modela optimizacije upravljanja asortimanom (OPTAS model) na maloprodajnim mjestima u Republici Hrvatskoj*, International Journal of Sales, Retailing and Marketing, Vol. 5, No. 1, str. 3-13.
3. Markić B., *Optimizacija portfolija i evolucijski algoritmi*, Informatologia 41, 2008.,
4. Markić, B. (2003), *Genetic algorithm for economic modeling,* Chapter 33, DAAM International Scientific Book, Vienna, Austria, 2003. ISBN: 3-901509-36-4, ISSN 1726-9687
5. Marshall Fisher, Ramnath Vaidyanathan (2014), *A Demand Estimation Procedure for Retail Assortment Optimization with Results from Implementations, Management Science 60(10):2401-2415*