Branch and Bound Method in Feature Selection Process for Models of Financial Condition Evaluation

Zofia Wilimowska[1], Marek Wilimowski[1], Tatiana V.Kalashnikowa[2]
[1] University of Applied Sciences in Nysa, Nysa, Poland
zofia.wilimowska@pwsz.nysa.pl; marek.wilimowski@pwsz.nysa.pl
[2]Kharkiv National Agrarian University, Kharkiv, Ukraine
kalashnikova.tv@gmail.com

Abstract. Uncertainty and risk which are associated with company's activity require a special instrument supporting process of the managers' decision making. The objective of the research is to determine the number features characterizing the financial risk – financial condition of companies. A quality of financial condition evaluation depends on the selection of variables (features) and criteria of the assessment. The choice of financial ratios in the study of financial standing of companies is crucial. The article presents the proposal to apply branch and bound method to choose suboptimal subsets of financial ratios that best describe the subject of the research, which is the company, under the assumption that classification algorithms are used for evaluation function creation. The aim of this study is to present a solution that allows the selection of financial ratios with a very high cognitive value, enabling the building of integrated measures assess the financial condition of the company.

Keywords selection of information, financial ratios, optimization, discriminatory models, branch and bound method

1 Introduction

In the rapidly changing market economies continuous assessment of financial phenomena occurring in businesses, in particular continuous evaluation of their financial condition is expected. Proper evaluation of the processes occurring in the enterprise enables prediction of the financial situation of the company and taking pre-emptive action which could protect the company from bankruptcy, it means enables risk reduction of the activity. A primary source of risk in human activities is a feeling of uncertainty connected with future unknown events, due to the fact that decisions are made today and the effects of the decision will be known in the future.

There have been carrying out calculations of financial ratios of public and private companies since the 19th century. As the years passed the development of statistical methods, that were used to predict business failure, were followed. The sixties of the twentieth century were a turning point in the study of the early diagnosis of the symptoms of risk of bankruptcy.

High volatility of the business environment and high risk of management result in a large number of bankruptcies. The tools of economic analysis allow for the rapid assessment of financial condition of companies and their financial risk. For the past twenty five years many models have been constructed to examine the financial condition of companies and to classify them as "ones with good condition" as the ones with "bad condition" which mean high risk of bankruptcy.

Many social groups are interested in the effects of company's operation. And each group is interested in various types of risk of organization activity. There is no profit without risk. Generally speaking, risk is a concept that denotes a potential negative impact that may arise from a future event, especially a total or partial loss of invested resources.

2 Feature Selection for Financial Condition Evaluation

Enterprises can be described by certain characteristics, features that can be financial and non-financial indicators, ratios. The use of synthetic indicators in the assessment process allows the assessment of a company financial standing, this is integrated assessment. Of course, it is clear that not every financial indicator (feature) is equally important in the evaluation of companies, therefore is crucial in this respect to choose (select) financial indicators most valuable, useful and crucial from the point of view of the assessing enterprise.

Multicriterial methods for company condition evaluation – discriminant methods, taxonomic methods, classifications (discrete risk assessments) – require the definition of a size vector (vector of features) that is the basis for assessments.

Why some indicators are more often used than others? Various aspects effect the frequency of their use. One of them is the availability of data, for example not all companies are listed on the stock exchange, what means that mostly the market ratios of companies are not known, and therefore should be removed from the set of financial ratios.

The literature suggests several methods of selection features (indicators) to build discriminatory models. Very often correlation matrix is used for features selection, but keep in mind that a strong correlation dependence between x_1 and x_2 does not exclude a weak relationship between x_1 and x_3, as well as between x_2 and x_3.

The second technique is to set yourself up as an expert in the selection of appropriate indicators. Currently, the authors are inspired by these indicators, which are often used to assess the insolvency of companies, something discussed in a number of publications.

A company has specific characteristics (in the assessment of the financial condition it can be financial ratios) that describe the object. These characteristics are expressed by a sequence *s* of *N* variables $x_1, x_2, \dots, x_N$. The larger the *N*, eg. the number of features, more difficult to choose of financial indicators that can be used to build the synthetic indicator, which is more difficult to make a selection.

Usually information about the assessed object is collected in excess. Therefore, there are many unnecessary features. Features may be superfluous because

• they do not introduce any differentiating information risk levels,

• sometimes it is enough to use one feature, if others are very strongly dependent on the former, or

• they have no relation to the purpose of classification – company condition evaluation.

In the "bankruptcy" models, Polish and foreign authors, there is a lot of talk about the quality of their assessment of enterprises, but not much about the selection of indicators in these models. Dozens of attempts to use models are carried out, estimating their diagnostic quality, but not much about the diagnostic quality of selected indicators. Of course, expert knowledge should be appreciated, but attention should also be paid to the possibility of using the methods already used, eg the methods of selecting information and selecting the features for example branch and bound method.

In this paper we propose a branch and bound method for selecting features for the construction of the synthetic index of company financial condition evaluation using classification methods.

2.1 Feature Selection

Before undertaking a discreet risk assessment of an enterprise (classification) [4, 5, 6], competent persons and experts are given a set of characteristics of objects to be classified on the basis of these features. However, the number of elements of such a set is usually very large. Some of these elements do not provide any or very little information about the class of the object. They are, from the point of view of recognition of company financial condition, not very useful and even complicate the algorithm. Therefore, it is necessary to select from a complete set a certain subset of features on the basis of which appropriate algorithms are built.

The set of features describing a given phenomenon should be constructed in such a way that it fully describes them. If this condition is met, then it will be possible to talk about the accuracy of assessments and analyzes, predictions as well as the accuracy of decisions made by the user on their basis. The reduction in the number of features is carried out on the basis of certain selection criteria.

One of the methods of reducing the number of features is the reduction of the set of features. The reduction of the set of features is understood as a procedure of reducing the set of features on the basis of which the classification algorithm will be implemented, most often by non-linear transformation of the *n*-dimensional space into *m*-dimensional, $n > m$.

The particular case of reduction is selection. Selection is understood as a method of selecting a subset of features from a larger set, ensuring minimization or maximization of the appropriate criterion. In other words, the selection task is to select a subset of features that bring as much amount information as possible (the amount of information is understood here as the value of the appropriate criterion). One of the basic selection criteria is the probability of erroneous classification, that is, the classification of an object into a class i and while it belongs to the class j.

Feature selection methods try to find a minimal subset of the features that meet the following criteria:

- the classification quality will not be significantly reduced,
- class distribution obtained using only selected features is as close as possible to the original distribution of these classes.

Depending on the a priori information about the recognized object, there are different selection algorithms. The algorithms proposed in the literature are very similar and differ only in the form of the selection criterion. Typically, the criteria proposed by the authors of the algorithms are different estimates of incorrect classification, so the algorithms are suboptimal to incorrect classification.

2.2 Combinatorial Method of Selecting Features in Classification Based on an Increase in Risk of Incorrect Classification

Analytical methods of selecting features do not provide absolute certainty as to the correctness of the choice of features – they are only the best, in the sense of a given selection criterion, approximations of the optimal selection. If the goal of the selection is to reduce the *n*-element set, to (*n*-*m*)-element set, it is only an overview of all possible combinations of features, i.e. analysis of all subsets $\binom{n}{n-m}$ can give such a guarantee. To reduce the number of reviewed solutions, you can use the branch and bound method.

For this case, two basic tasks have been formulated [3]:

1. rejection m features from n, $m < n$, such that the increase in risk is minimal,
2. rejection of as many features as possible, so that the increase in risk does not exceed the preset number $\varepsilon > 0$.

The proposed selection algorithms are mainly based on the following three methods:

1. Calculation for each feature x_i, $i = 1,2,\dots,n$ appropriate estimates and rejection of those features for which these estimates reach extreme values.
2. Designation for every possible solution of an appropriate estimation and selection of the one for which it assumes the smallest value, in other words a direct review of all solutions.
3. Subsequent rejection of traits, that is, the rejection of one feature, the choice of the best solution, then the rejection of two features, the previously chosen and the next and again the choice of the best solution, etc.

Let us assume that the Bayes algorithm is used to assess the financial condition of the company.

The measure of incorrect classification for the Bayes algorithm is expressed as follows [3]:

$$R = \bar{P} = \bar{P}(i = \Psi(x)) = 1 - \int_{\mathbf{x}} \max_{j=1,\dots,L} p_j f\left(\frac{x}{j}\right) dx$$

Let

X^i– means $(n-1)$-dimensional space, after elimination the feature x_i, $i = 1, \dots, n$;

$R^{(i)}$ – means the measure of incorrect classification, which will be obtained by carrying out (recognizing) the classification based on $(n-1)$ the characteristics, after rejecting the feature

Let it continue

X_i – means the one-dimensional space for the x_i, $i = 1, \dots, m$ feature, whereas

x^i – indicates the current value of the $(n-1)$ dimensional vector after eliminating from the vector x the features with the number *i*,. $i = 1, \dots, n$.

Then by definition

$$R^{(i)} \stackrel{\Delta}{=} 1 - \int_{\aleph^i} \max_{j=1,\dots,L} \left[\int_{\aleph_i} f(x/j) p_j dx_i \right] dx^i \tag{1}$$

Analogously, let

$$X^{k_1, k_2, \dots, k_s} \tag{2}$$

means the $(n-s)$ dimensional space of objects obtained from space *X* by rejection of index $k_1, k_2, \dots, k_s \in \{1, \dots, n\}$ features.

Next, let $x^{k_1, k_2, \dots, k_s}$ (3)

means the current value of the $(n-s)$ dimensional feature vector (3) after eliminating (rejecting) the characteristics of the indexes $k_1, k_2, \dots, k_s$, then

$$R^{(k_1, k_2, \dots, k_s)} \tag{4 }$$

means the measure of incorrect classification that will be obtained as a result of a feature-based on $(n-s)$ features classification after elimination $x_{k_1}, x_{k_2}, \dots, x_{k_s}$

$$R^{(k_1, k_2, \dots, k_s)} \stackrel{\Delta}{=} 1 - \int_{\aleph^{k_1, k_2, \dots, k_s}} \max_{j=1,\dots,L} \left[p_j \int_{\aleph_{k_1}} \dots \int_{\aleph_{k_s}} f(x/j) dx_{k_1} \dots dx_{k_s} \right] dx^{k_1, \dots, k_s} . \tag{5}$$

Let

V^n means the set of indexes of individual components

$$V^n = \{1, 2, \dots, n\}, \tag{6}$$

While

$$U_t^q = \{r_1, r_2, \dots, r_t\}^q \tag{7}$$

means the means q^{th} *t*-elements combination of *n*-elements index set V^n,

Considering (7), the follow form will also be used

$$R^{U_t^q} \equiv R^{(r_1, \dots, r_t)^q} \equiv R^{(r_1^q, \dots, r_t^q)}, \tag{8}$$

$$r_i^q \in V^n, \quad i = 1,\ldots,t.$$

With these signs, the basic selection tasks can be formulated in the following way.

Task 1. Let there be a given vector of measurable *n*-attributes $x = (x_1, x_2, \ldots, x_n)$ and a number of features that should be rejected, m

The task is to find such a *m*-element combination of the set V^n, so that the rejection of features with indexes included in this combination resulted in the smallest increase in the measure of incorrect classification of R, i.e. it should be found q_0 that

$$R^{U_m^{q0}} = \min_q R^{U_m^q}, \quad q = 1,2,\ldots,\binom{n}{m}. \tag{9}$$

Task 2. Let there be a *n*-element feature vector and the number ε by which the risk may increase

$$\varepsilon = R_1 - R, \tag{10}$$

where

R_1 – means the measure of incorrect classification after selection.

The task is to find the maximum number of features that can be rejected on the assumption that the risk will not increase more than ε, i.e. it should be found q_0, t_0 that

$$\left.\begin{aligned} & t_0 = \max_t, \quad t = 1,\ldots,n-1, \\ & R^{U_{t0}^{q0}} - R \le \varepsilon \end{aligned}\right\} \tag{11}$$

The feature selection tasks defined in this way can be treated as combinatorial tasks.

Let denote $\overline{w}$ a discrete finite set. Each element of this set can accept only two values 0 or 1

$$w_i = \{0,1\}, \quad i = 1,\ldots,n. \tag{12}$$

Let the set $\overline{w}$ further correspond to the set of indices V^n. The values of the elements of the set $\overline{w}$ are defined as follows:

$$w_j = \begin{cases} 0, \text{ when the feature } j \text{ was rejected}, \\ 1, \text{ when the feature } j \text{ was not rejected}, \end{cases} \tag{13}$$

$$j = 1,\ldots,n.$$

The problem of the optimal selection of features in relation to the measure of incorrect classification consists in finding such a set $\overline{w}$, for which the objective function, in this case the measure of incorrect classification, achieves the smallest value. So it is a combinatorial task. The goal function, which should be minimized when selection tasks is defined as combinatorial tasks, can not be presented in a general way. It should be defined for each set $\overline{w}$ separately.

When considering **task 1**, i.e. the task of rejecting from among n features, m features, such that the increase in risk is minimal, the function of the goal can be formulated as follows.

It is assumed that only m elements of the set $\overline{w} = \{w_1, w_2, \ldots, w_m\}$ can simultaneously accept zero values. If $w_{k_1}, w_{k_2}, \ldots, w_{k_m} = 0$, $k_1, \ldots, k_m = 1, \ldots, n$, $k_1 \neq k_2 \ldots \neq k_m$ the other elements of this set are equal to 1, then the objective function for such a set has the form

$$g(\overline{w}) = R(\overline{w}) = R^{k_1, \ldots, k_m}. \tag{14}$$

The limitation in this case is in the form

$$\sum_{j=1}^{n} w_j = n - m \cdot \tag{15}$$

The task is to find such a set $\overline{w}$ that satisfies the constraint (15) so that objective function (14) will be the smallest.

Task 2 can be formulated in the following way. It is assumed that t elements of the set $\overline{w}$ can take zero values, $t = 1, \ldots, n-1$.

The goal function to be minimized is in the form

$$g(\overline{w}) = \sum_{i=1}^{n} w_i \tag{16}$$

with limitation

$$R(\overline{w}) - R \leq \varepsilon, \quad \varepsilon > 0, \tag{17}$$

where $R(\overline{w})$, if $w_{k_1}, w_{k_2}, \ldots, w_{k_t} = 0$, $k = 1, \ldots, n$, $t < n$, $i = 1, \ldots, t$, and remaining elements of the set $\overline{w}$ equal to one are expressed as follows

$$R(\overline{w}) = R^{(k_1, k_2, \ldots, k_t)}. \tag{18}$$

From the formulation of the optimal selection task in the above-mentioned way, it follows that only some of the discrete programming methods can be used in these tasks. Both tasks have nonlinearities. In **task 1** the objective function is non-linear, while in **task 2** there are restrictions. It is possible to reject all methods of cutting planes, because it was shown that these tasks should be treated as a combinatorial type task and look for a suitable method among the methods that solve this type of task.

One of the most suitable methods for solving selection tasks is the branch and bound method, and this method has continued to be adapted to the selection tasks mentioned. The branch and bound method is aimed at finding in a large set of solutions characterized by a specific feature, an element with an extreme value of this trait. This is done by appropriately sequentially dividing the set of all solutions into smaller and smaller subsets to receive a one-piece collection, which is a sought-after solution to the task. Finding the shortest way to get a solution requires the adoption of an appropriate selection strategy. For this purpose, depending on the formulation of the objective function, subsets are subdivided according to the minimum or maximum values limiting the values of the features occurring in individual subsets.

The ideas of this method can be presented in the following way.

Minimize the function

$g(z)$, under the condition $Z = (z_1, z_2, \ldots, z_n) \in G$,

where *G* – it is some set of finite sets.

At the basis of this method are the following activities, which in many cases allow to reduce the area of the review:

1. Calculation of the lower bound (estimation) $\xi(G)$,
   $g(z) \geq \xi(G)$.
2. Division into subsets of the set of solutions *G*.
3. Conversion of bounds (estimates).

2.3 Selection Algorithms

It is assumed that the purpose of the relevant algorithms is to determine the features that should be rejected. The following points will be adopted as the basis for the adaptation of the division method and constraints:

1. The lower bound (estimation) of the objective function on the set G is R
   $$\xi(G) = R \cdot \quad (19)$$
2. The division of the set G into subsets will be made according to the following rule
   $$G \equiv G^0 = G_1 \cup G_2 \cup \ldots \cup G_n, \quad (20)$$
   where:
   G_i – a set of these solutions contained in the "superior" solution, in which the *i*th, $i = 1, \ldots, n$ trait is additionally rejected (by the superior solution is meant a solution in which the features were rejected $j \in U_s^q$).

3. The lower bound of the objective function (estimation) for each set G_i will be $R^{(U_s^q, i)}$ – the risk after rejection of the features contained in U_s^q and the *i*th traits.

Optimization of algorithms for both tasks results from the optimality of the branch and bound method [3].

In order to clearly present the idea of applying the branch and bound method to the selection of features, the selection task marked with the number 1 was analyzed. The algorithm is based on the branch and bound method according to the strategy, in which two stages can be distinguished. In the first stage, the solutions are reviewed along one branch and the optimal solution is found in it, in the second stage, on the basis of the solution obtained in the first stage, the remaining branches are reviewed [3].

The algorithm is showed in [3].

Figures 1 and 2 show an illustration of stage 1 and stage 2 of the branch and bound method on the example of the selection of three features of five; $n = 5, m = 3$. Of course, the example is demonstrative and its task is to show the mechanism of the algorithm – the way to choose the next solutions.

It is easy to see that in this case there are features to "leave" than to "reject" if only the algorithm is reversed. It was assumed here that the next "best" sets of solutions are collections:

G_3^1 – in the first step G_2^2 – in the second step and G_2^3 – in the third step, i.e. the solution $\bar{Z}$ obtained in the first stage is:

$z = \{3,2,4\} = G_2^3 = U_3^2$

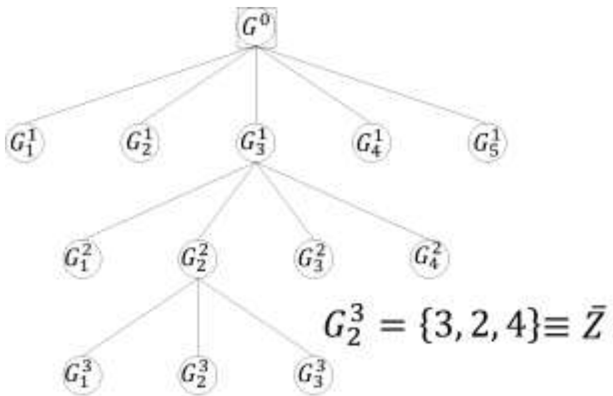

Fig. 1. Graphical representation of stage 1

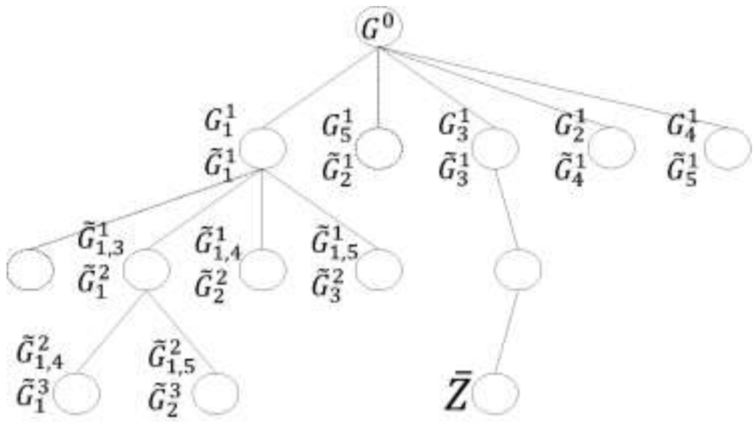

Fig. 2. Graphical representation of stage 2

Fig. 2 shows the "renumbering" of the index set;

$\tilde{x}_1 = x_1, \quad \tilde{x}_2 = x_5, \quad \tilde{x}_3 = x_3, \quad \tilde{x}_4 = x_2, \quad \tilde{x}_5 = x_4.$

The dashed line indicates a solution for which the limit value is greater than the limit value for the solution $\bar{Z}$. In this case, it is a set $\tilde{G}_2^1$ in step one and a set $\tilde{G}_{1,2}^1$ the second. It was further assumed that the limit value for $\tilde{G}_1^2$ is smaller than for $(\bar{z})$. The optimal solution should be sought among the sets $\bar{z}$, $\tilde{G}_1^3$, $\tilde{G}_2^3$.

3 Conclusion

Presenting the task of selecting traits as combinatorial tasks allowed for the construction of an algorithm for optimal risk selection of incorrect classification. The branch and bound method is the optimal method, i.e. it allows finding the best, in the sense of a given criterion, solution. This is particularly important for the solution of **task 2**, where the number of all solutions in the extreme case can be equal $2^n - 2$, if the best solution is to be looked at by reviewing all possible acceptable solutions.

For **task 1**, the reduction in the number of solutions reviewed is less visible, but has the advantage that in some cases the solution is faster, and besides, realizing the algorithm at computer you can interrupt calculations after a while to obtain a suboptimal solution.

The reduction in the number of viewed files in the proposed algorithm can be achieved by using, in place of appropriate estimates, incorrect measures that would satisfy the property of multiplicative or additivity to features. However, finding such an estimate is very difficult. Only for some special cases of recognition it is possible to meet these requirements.

References

1. Sobczak W., Malina W. (1985), Metody selekcji i redukcji informacji, NT, Warszawa.
2. Walesiak M., „Uogólniona miara odległości w statystycznej analizie wielowymiarowej”, Wyd. AE we Wrocławiu, Wrocław, 2002r.
3. Wilimowska Z. (1980), Kombinatoryczna metoda selekcji cech w rozpoznawaniu obrazów na podstawie wzrostu ryzyka. Archiwum Automatyki i Telemechaniki. t. 25, z. 3, s. 405-415.
4. Wilimowska Z. (2002), Względna dyskretna ocena ryzyka w szacowaniu wartości firmy, in: Information Systems Applications and Technology ISAT 2002 Seminar. Modele zarządzania, koncepcje, narzędzia i zastosowania. Materiały międzynarodowego seminarium, Karpacz, 11-13 grudnia, Wrocław.
5. Wilimowska Z**.** (2005), Dyskretny pomiar ryzyka w badaniu kondycji finansowej przedsiębiorstwa. W: Naukovi Praci Kirovograds'kogo Nacional'nogo Technicnogo Universitetu. Ekonomicni Nauki. (Nauk. Pr. Kirovograds'kogo Nac. Tech. Univ., Ekon. Nauki) vyp. 7.
6. Wilimowska Z., Wilimowski M.(2009), Bayesian model of the firm's risk of bankruptcy diagnose, in: Information Systems Architecture And technology : system analysis in decision aided problems, Oficyna Wydawnicza Politechniki Wrocławskiej, pp. 69-82.
7. Wilimowska Z. (2003), Models of the firm's financial diagnose, in: Information Systems Applications and Technology ISAT 2003 Seminar. Proceedings of the 24th international scientific school, Szklarska Poręba, 25-26 September 2003, Wroclaw.

8. Wilimowska Z,. Wilimowski M., Koszałka J. (2009), Integrated models of the firm's financial diagnose, in: Information systems architecture and technology : advances in Web-Age Information Systems, PWr, pp. 303-314.
9. Winakor A., Smith R.F. (1935), Changes in Financial Structure of Unsuccessful Industrial Companies, Bureau of Business Research, Bulletin No. 51.
10. Zhang G., Hu M., Patuwo B., Indro d. (1999), Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis, "European Journal of Operational Research", Vol. 116, No.1, pp. 16-32.
11. Zmijewski M. (1984), Methodological issues related to the estimation of financial distress prediction models, Journal of Accounting Research, Vol. 22(Suppl.), pp. 59–86.