Control and Cybernetics

vol. **40** (2011) No. 1

Hybrid approach to supporting decision making processes in companies*†

by

Wiesław Pietruszkiewicz, Michał Twardochleb and Maciej Roszkowski

West Pomeranian University of Technology
Faculty of Computer Science and Information Technology
Zolnierska 49, 71-711 Szczecin, Poland
e-mail: {wpietruszkiewicz, mtwardochleb, mroszkowski}@wi.zut.edu.pl

Abstract: This article presents the advantages of hybrid approach to the support decision making by analyzing three areas of business decision problems, solved by combination of well-known algorithms into the new hybrid constructions: cascade optimization hybrid, parallel classification hybrid and hybrid multicomponent attribute selection. Each of them solved a different problem: the cascade optimization hybrid allowed for finding an extreme of a composite objective function, the parallel classification hybrid was used to choose a proper class through voting, the multicomponent attribute selection robustly chose significant decision variables. A hybrid approach to the problem of supporting the decision making processes is more effective than using each of the component methods alone, even for the sophisticated ones. A combination of several methods with different characteristics and performance makes it possible to take advantages of their strong sides and simultaneously eliminate the weak ones, resulting in a better computational support of decision making.

Keywords: data mining, cascade optimization hybrid, parallel classification hybrid, hybrid multicomponent attribute selection.

1. Introduction

Business is a continuous decision making process, which, as a multi-stage task, might be formally described through specification of the steps and component elements meant to support the decision-maker. Thanks to such formalism it is possible to present the decision making process in a precise form, which might

*Submitted: July 2010; Accepted: March 2011.

†This is an extended and amended version of the paper, presented at the 5^{th} Congress of Young IT Scientists (Międzyzdroje, 23-25.IX.2010).

be analyzed with regard to its accuracy, be used as guidelines for other persons or be implemented in the form of an IT system.

Thus, from the formal point of view, the decision making process can be divided into five main stages:

1. **Identification of the decision situation** – its aim is to specify the set of all factors influencing the decision making, which, depending on the decision-maker's influence, can be divided into two groups: factors dependent on the decision-maker and factors independent of the decision-maker. In the subsequent stages of the decision making process, they are treated as limiting conditions (the group of factors independent of the decision-maker) or included as an element of the decision evaluation criteria (the group of dependent factors).
2. **Formulation of the decision making problem** – describes the decision problem, which might be used to create the decision model in later stages. The component elements of the description are the following: Specification of the decision-maker(s), their decision making demands on the model, identification of the conditions limiting the decision possibilities, determination of the sets of permissible decisions and criterion (criteria) for decision evaluation. The criterion is a function which allows for ranging the elements of the set of permissible decisions. The order is specified by assigning quantitative or qualitative evaluation of attractiveness to each decision. It is also necessary to specify the way of interpretation of the assigned attractiveness measure.
3. **Construction of a model** – presents a chosen fragment of reality in the form of a mechanism illustrating the decision problem with the necessary precision and allowing for specification of the sets of permissible and optimal decisions. The mechanism is usually constructed by specifying the mathematical relations between the elements, although its functional modeling might also occur, which identifies logical relations between the components.
4. **Determination of the decisions** – specifies the set of decisions: permissible, sufficient and optimal. The type of determined sets depends on the requirements of the decision-maker. For some problems it is enough to determine the decisions meeting specific conditions, while for others there is a need for decisions with attractiveness above/below a certain threshold. The most difficult task is determination of the optimal decisions, i.e. the best ones from the point of view of the decision making criterion.
5. **Decision making** – evaluates the results of previous stages and defines the ultimate decision, which is implemented after its final verification.

To reduce the subjectiveness in the decision making process, causing the results to biased, it is recommended to use multiple opinions from several experts or to examine the observations and deduce the respective knowledge using an artificial intelligence approach.

The subsequent sections of this paper present three examples of using the hybrid approach to supporting the decision making processes, which made it possible to facilitate various decision making processes in companies by showing different decision problems and possibilities for the hybrid use of optimization, classification or selection methods. Hybridization of algorithms is, in this case, the combination of distant problems, which, with such juxtaposition, reveal the benefits of a synergic joint use of various algorithms, while dissimilarity of the decision problems shows great possibilities for the use of computer mechanisms for supporting the decision making processes in companies.

It must me noted that the precise implementation of a decision supporting system depends on the type of problem solved and the process presented above is a general guideline, being also problem independent.

There exist many methods of computational decision support. However, they have some limitations that might be overcome using a hybrid approach that, by combining properly selected elementary methods, allows one to:

- achieve more accurate results,
- construct more robust decision mechanisms,
- deal with problems too complicated for elementary methods,
- reduce the level of required computational resources.

Hence, the hybrid methods, build over correctly selected basic ones (even simple), perform better than single, even sophisticated, methods applied in the decision problems. To prove this proposition, herein three different examples will be analyzed and by presenting the successful applications of hybridization in various kinds of decisional problems, they will confirm the above statement.

2. Hybrid method of function optimization

Hybridization of multiple optimization methods (especially when facing large-dimensional problems) may lead to high synergy between included methods. When several simple method are linked together, a very powerful optimization tool can be created, able to solve a rich variety of problems, independent of its characteristics (continuity, differentiability, multiple local extremes etc.)

Optimization of the function of several variables is used not only to calculate the function minimum or maximum. Optimization procedures are also used to find sub-optimal solutions which meet a certain quality criterion. This is particularly useful in situations where an approximate solution is sought to problems of NP class.

Although the use of optimization procedures requires a description of the studied problem in a mathematical form, thanks to the use of modeling, it is possible to bring typical decision problems to such a form. A created mathematical model does not have to describe a function which is unimodal, continuous and differentiable in its domain – a much larger class of functions is those which (today) cannot be optimized with analytical methods and their values for input data can only be determined with the use of many iterations of the algorithm simulating the studied phenomenon.

Function optimization is used both for problems occurring on the financial markets (e.g. the choice of an investment portfolio consisting of a given number of securities) and in decision support systems.

A mathematical definition of optimization comes down to finding (with given limitations) the minimum of the function (an equivalent problem is to find its maximum). Although the definition is simple, in practice, determining the minimum of any function is not a trivial task. Over the years, many optimization algorithms have been created.

The simplest type of optimization tasks is optimization of the function of one variable. At the same time, tasks of this type have an important place in the theory of optimization not only due to their frequent use in the engineering practice, but also due to the fact that the problem quite often appears as a sub-task during optimization of functions in multidimensional spaces.

Among the many types of optimization algorithms, the ones worth attention are genetic algorithms, which are characterized with efficiency exceeding other forms. However, in most cases, the optimization process must be preceded by choosing the best optimization method and its parameters. As a result, they are used especially in dedicated systems, where information is provided on the characteristics of the studied problem.

Combinations of algorithms of various types are called hybrid optimization procedures. Their use is justified when a suitable combination of component algorithms leads to the use of their strong sides and elimination of the weak ones.

The idea of using several different optimization methods for solving complex problems occasionally appears in the available literature (see e.g.: Pinter, 2002; Bagirov and Rubinov, 2004). A combination of the elements of global and local optimization is proposed. Two approaches to the combined use of various types of optimization methods might be observed in international literature:

1. Local optimization is used to **find a stationary point**, which is a local optimum, and, then an attempt is made to "jump" out of neighbourhood of the the local minimum with the use of a global optimization technique.
2. **Starting points are found** for a further local optimization with the use of global optimization.

The here presented cascade hybrid method, de facto, combines both of these approaches. In fact, due to the assumed cyclicity of the optimization process, the global and local optimization techniques are used alternately.

The prerequisite for the construction of this hybrid is the use of various methods in subsequent phases of the optimization process. The structure of the hybrid algorithm should also meet the following conditions:

1. All component methods should represent **different classes of optimization methods** as they will, then, represent different approaches to problem solving. Thanks to that it is possible for methods of different characteristics to "support" each other, i.e. the strong sides of each method will be used and their disadvantages will be eliminated in the process of synergy,
2. Methods used in the initial stage of the optimization process ("rough" search) should be characterized by high resistance to **problems of the local extreme of objective function and the starting point**. A consequence of this is departure from the convergence requirement of the search process in this stage. Methods of the initial stage should allow for equal sampling of the domain area of the optimized function. In this stage high precision of methods is not required as the aim of this stage is the possibly precise determination of the starting point (or starting points) for further search in the subsequent stages of the optimization process.
3. Methods used in the subsequent stage of search should allow for **more precise analysis of the neighbourhood of the starting points** obtained previously in the initial stage search. In this stage, high convergence of search is also not required.
4. Optimization methods used in the final stage (precise search takes place in this stage) should be characterized by **high convergence of search to the extreme point**. These methods should effectively and precisely determine the local extreme point located in the surrounding of the starting point obtained in the previous stages. Due to the assumed even "combing" of the domain in the previous stages, the obtained point will be the global extreme of the optimized function with high probability. Thus, these methods (even in spite of high analytical complexity of the algorithm) should provide a highly precise result in a relatively short time.
5. **A hybrid method should be characterized by its cyclical nature.** This means that, after finding the extreme point (as a result of all optimization stages), the method should make it possible to continue the search. When starting the next optimization cycle by returning to the stage of "rough" search, one must check these areas of the domain, which have not been analyzed previously and omit the areas, which have already been examined. The possibility of cyclical search is especially important in the case of multi-modal tasks of very high dimensionality as it provides a possibility for finding a better solution than the one obtained before in the next search cycle. In the case of a priori optimization (such, where the result of search is unknown), the researcher is not certain as regards obtaining the function extreme.

Based on the assumed criteria, a three-stage cascade hybrid method might be proposed consisting of the following stages:

1. Monte Carlo stochastic method
    - The random points have homogenous probability distribution within the function domain
2. Genetic method – with the use of competition selection
    - An individual's chromosome corresponds to a record of individual coordinates (dimensions) of points
3. Rosenbrock method
    - Rotation of coordinates realized in accordance with the Gram-Schmidt method.

A combination of three methods with different characteristics was aimed at elimination of their weak sides with simultaneous use of their assets, which significantly improves the convergence and reduces the search time.

Functioning of the hybrid method for a given number N, being the parameter of the overall method:

- a choice of the best "shots" from among N samples – Monte Carlo,
- use of the best samples as a starting population (filling of the chromosomes) and creation of N generations,
- initiation of the optimization with the Rosenbrock method for the best ultimate individual.

In order to check the efficiency of the proposed method, it was subject to a number of tests with the use of many benchmarking functions. Their aim was to prove the effectiveness of the proposed hybrid method in solving optimization tasks with various characteristics. These studies also aimed at evaluating the effectiveness of the proposed method in comparison with other optimization methods.

The testing functions included are shown in Figs. 1-4.

Shekel's "Foxholes"

$$f(x) = -\sum_{i=1}^{30}\left(\sum_{j=1}^{n}(x_j - a_{ij})^2 + c_i\right)^{-1} \quad (1)$$

n – number of variables
(for specific values a_{ij}, c_i)

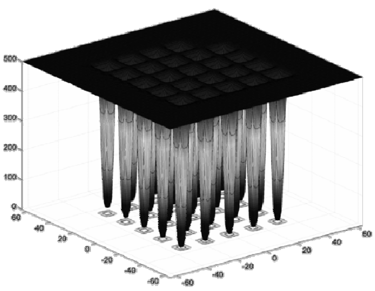


Figure 1. An example of the Sheckel's function graph

Rastrigin's function

$$f(x) = 10n + \sum_{i=1}^{n} (x_i^2 - 10\cos(2\pi x_i)) \quad (2)$$

n – number of variables
-5.12 $\leq x_i \leq$ 5.12

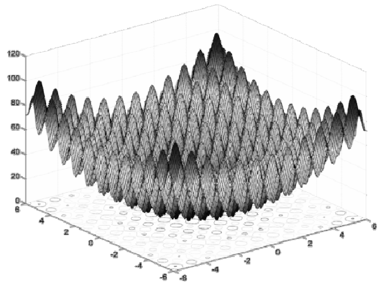


Figure 2. An example of the Rastrigin's function graph

Schwefel's function

$$f(x) = -\sum_{i=1}^{n} x_i \sin\left(\sqrt{|x_i|}\right) \quad (3)$$

n – number of variables
-500 $\leq x_i \leq$ 500

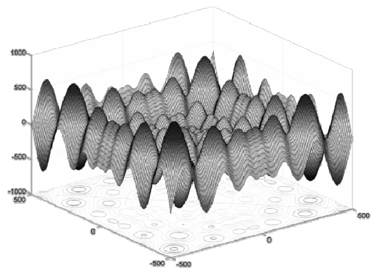


Figure 3. An example of the Schwefel's function graph

Ackley's function

$$f(x) = -a\exp\left(-b\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}\right) - \exp\left(\frac{\sum_{i=1}^{n}\cos(cx_i)}{n}\right) + a + \exp(1) \quad (4)$$

n – number of variables
-1 $\leq x_i \leq$ 1
(for specific values a, b, c)

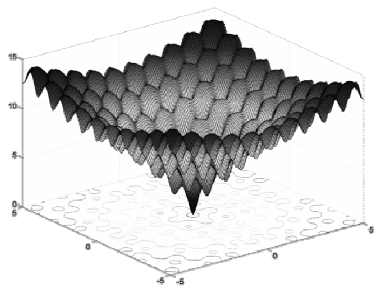


Figure 4. An example of the Ackley's function graph

Table 1. Efficiency of the hybrid compared to the components (percentage of correct solutions, i.e. obtaining the assumed global minimum neighbourhood with 20 independent samples, depending on the required precision). Self-elaboration based on Twardochleb and Rychcicki (2009)

Name	**Monte Carlo**	**Genetic algorithm**	**Rosenbrock method**	**Hybrid metod**
Precision criterion:	***0.01, 0.007, 0.0025, 0.0001, 0.00007***	***0.01, 0.007, 0.0025, 0.0001, 0.00007***	***0.01, 0.007, 0.0025, 0.0001, 0.00007***	***0.01, 0.007, 0.0025, 0.0001, 0.00007***
Foxholes	{0, 0, 0, 0, 0}	{100, 100, 100, 5, 5}	{0, 0, 0, 0, 0}	{100, 100, 100, 5, 5}
Rastrigin 2var	{100, 100, 90, 10, 10}	{100, 100, 100, 100, 100}	{10, 10, 10, 10, 10}	{100,100,100,100,100}
Rastrigin 3var	{45, 30, 5, 0, 0}	{100, 100, 100, 100, 100}	{0, 0, 0, 0, 0}	{100, 100, 100, 100, 100}
Schwefel	{0, 0, 0, 0, 0}	{50, 50, 35, 15, 15}	{60, 60, 60, 60, 60}	{100, 100, 100, 100, 100}
Ackley 2var	{75, 60, 35, 0, 0}	{100, 100, 95, 50, 50}	{45, 45, 45, 45, 45}	{100, 100, 100, 100, 100}
Ackley 4var	{0, 0, 0, 0, 0}	{90, 90, 70, 10, 5}	{0, 0, 0, 0, 0}	{100, 100, 100, 100, 100}
Ackley 6var	{0, 0, 0, 0, 0}	{55, 35, 25, 0, 0}	{0, 0, 0, 0, 0}	{100, 100, 100, 100, 100}

A developed testing application makes it possible to determine the current optimization result after a given number of iterations (and after performing a given number of calculations of the optimized function values) and specification of the minimum number of iterations after which a definite accuracy has been reached (calculations were made both of the difference between the global minimum value and the current result of the method as well as of the Euclidean distance from the global minimum). A detailed description of these studies was included in Twardochleb and Rychcicki (2009). Table 1 presents the efficiency of the hybrid with its components for chosen benchmarking functions depending on the given criterion of precision.

Results show that the hybrid method finds very precisely the correct solution for almost every function (except only for the very high precision required from Foxholes function optimization), which is better or equal to that of any of the single component method.

Combining the optimization methods with different characteristics into a hybrid made it possible to obtain a highly universal optimization tool. Unlike the methods used individually, the hybrid method allows for effective optimization of both unimodal and multimodal functions. This feature of the hybrid is of particular importance in the situation when characteristics of a new optimization task are unknown and, as a result, it is impossible to determine in advance the type of the optimization method which would be suitable for its solution. It was noticed that, in certain cases, the number of iterations for the hybrid method is bigger than for the composite methods although higher efficiency in reaching the global extreme point makes it possible to accept this cost, especially as an increased advantage of the hybrid method was observed together with an increase in the number of variables. A dramatic decrease in the efficiency of stochastic methods with a growing number of variables or increased precision proves that the hybrid method is less sensitive to this factor. The use of an initial phase of random search, in turn, reduces the risk of "getting stuck" in the neighbourhood of a local extreme, which is a significant drawback of the organized methods (e.g. Rosenbrock method).

The proposed hybrid approach might be extremely useful in the case of problems with unrecognized characteristics with many equality and inequality constraints (also nonlinear). These definitely include a range of decision problems regarding company management.

3. Parallel classification hybrid

The parallel classification hybrid made use of methods that represent different approaches to the problems of supporting decision making processes in companies that can be categorized as consisting in classification. The examples are the Naive Bayes and K-Nearest Neighbors methods. Naive Bayes method is based on probability distribution in particular classes. The K-Nearest Neighbors method assumes memorizing of the whole training set in the process of training. An other example, of diversity of these methods but with some level of similarity, is the idea of functioning of Neural Networks MLP and SVM Method. Neural Network MLP makes possible splitting the space of input data with the use of hyper planes. The working of the SVM method is similar, but firstly the data space is mapped into the properties space with the use of a hypersphere. These two methods are often presented together because of high quality of classification comparing to other classification methods. All methods used in the parallel classification hybrid have many complimentary characteristics and properties. A combination of particular methods in the form of a hybrid model gives hope for achieving better solutions than achieved by each method separately.

Calculations made use of the commonly available data on testing the quality of wine by experts (Cortez et al. 2009). The input variables were 11 quantitative physicochemical properties of wine. The output variable was evaluation of wine quality performed by experts. For the needs of the experiment, a change was made in the values of the output variable. The "quality" quantitative output variable (values from 0 to 10) was changed into a qualitative output variable with two values: "bad quality" (equivalent to values from 0 to 5) and "good quality" (equivalent to values from 6 to 10). Two sets of data were used in the test: red wines (1599 samples) and white wines (4898 samples). In the set of red wines, the cardinality of wine classes was similar (53.5% "good quality", 46.5% "bad quality"), while in the set of white wines, the cardinality of classes was more differentiated (66.5% "good quality", 33.5% "bad quality"). The set of white wines was used while training the hybrid (random division of the set: 50% training data, 50% testing data), while the set of red wines was used to test the hybrid. Input data was subject to standardization so that the values of individual input properties do not influence classification.

3.1. Construction of the parallel classification hybrid

Six methods were used for constructing the parallel classification hybrid (Fig. 5): C & RT (Classification and Regression Trees), NB (Naive Bayes), NN (Neural Networks), K-NN (K-Nearest Neighbors), MARS (Multivariate Adaptive Regression Splines), SVM (Support Vector Machine). Implementation of the parallel classification hybrid was performed in StatSoft Statistica 9.0. program.

Models of classification and regression trees use a classical C&RT algorithm (Breiman et al. 1984) for constructing classification and regression trees. Models of classification and regression trees use different approach from the linear and non-linear methods. They make it possible to present the classification results in the form of logical conditions of if-then type, so that it is easy to find a set of logical conditions for the division of the classified object. Most frequently, the number of logical conditions is not too high and, as a result, the model is not too complicated, which allows for fast and easy classification of new cases.

Naïve Bayes method (NB) makes use of Bayes theorem for classification of objects into the most probable class. Probability of belonging to one of the classes might be estimated as a ratio of the number of cases belonging to this class to the number of all cases. The Naïve Bayes method is based on a "naďve" assumption that values of the individual input properties are conditionally independent with a specified decision making class. This assumption is usually not met, but works well as approximation.

Neural Networks (NN) are a very simplified implementation of human brain connections. The system of neuron connections has a specific structure. Input signals are given to the network. The signals reaching the neurons have some weight. The signal activating the neuron is calculated as a sum of the products of the input signal value and its weight. The signal activating the neuron is an argument of an activation function with a specified activation threshold. In the training phase, the input data is processed and solutions are written into internal mapping of the neural network. The hybrid made use of an MLP (Multi Layer Perception) unidirectional neural network trained with the Back Propagation algorithm.

K-Nearest Neighbors method (K-NN) classifies a given object into the most numerous class corresponding. Cardinality of classes is calculated based on the nearest neighborhood. K of the nearest objects are taken into consideration, assuming that the objects belonging to the same class must neighbor upon each other. The method involves memorizing of the whole training set, implying its high memory complexity. The analysis of the memorized set starts once the required classification appears. The subsequent classifications do not make the model adjust to them. All intermediate calculations are done again for the new classifications. The classification schema requires calculation of the distance of the classified object from all objects. Next, K-nearest cases are chosen after sorting. The object is classified into the appropriate class with the highest number of cases.

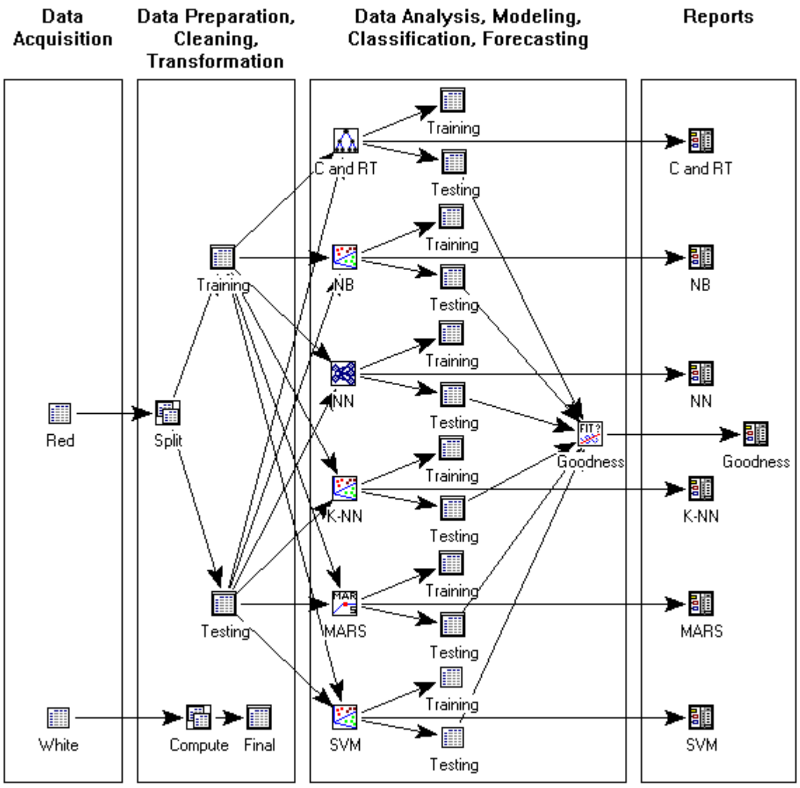


Figure 5. Parallel classification hybrid. Implementation for data on wine quality classification. Self-elaboration.

Red / White – input data regarding red/white wines.
Split – division into training and testing data.
C and RT, NB, NN, K-NN, MARS, SVM – classification methods.
Goodness – goodness-of-fit statistics.
Compute – compute best prediction from all models
Final – final result of classification of the whole hybrid.

Multivariate Adaptive Regression Splines method (MARS) makes use of multiple segment linear regression (Friedman, 1991) for classification of input data. This method uses base functions in the form of two-sided cut linear functions for approximation of relations between input and output variables.

Support Vector Machine method (SVM) classifies the objects with the use of hyperplanes which separate the cases belonging to different classes in multidimensional space. Location of the hyperplane in the space is optimal if it allows for obtaining a maximum margin during separation of example data. Such hyper plane "rests" on the points located on the edge of separable areas. The use of the supporting vectors method is possible in the case of separating classes with the use of a hypersphere when data space is mapped into properties space. For this mapping to be successful Cover's theorem about data reparability must be satisfied. According to Cover‘s theorem, a data space might be transformed into a properties space (where their linear separation will take place) when space dimension is high and transformation is non-linear.

Identical sets of training and testing data were given at the input of each method. Each method processed input data as a single dichotomous classifier. In this way, each method created its own model fitting the data from the training sample.

In order to evaluate the models generated by each method, goodness-of-fit statistics was used expressed in percentages of accuracy, i.e. the quotient of the number of properly classified cases and the total number of cases. Goodness-of-fit statistics was calculated based on the classification of the testing sample.

Based on the goodness-of-fit statistics values, implementation was initiated on the set of white wines. Having taken into consideration the rates of wrong classifications from the testing sample, it was possible to use model voting. With the use of such voting it is possible to obtain aggregation of predictions for many models of different types for the same data. The prognosis was set in four variants by: vote of all models, vote of best three models, vote of best two models and choice of the best model.

3.2. Quality evaluation of the classifications performed by the parallel hybrid

In order to compare the classification results obtained by the hybrid four variants of voting, these results were correlated with the results obtained by the six classifier models used in the hybrid (Table 2).

The best accuracy was obtained for the vote of two best classification models (70.68%). The second best result was obtained with the vote of all models (69.54%). Let us add that "Best prediction" and NN are actually the same methods. Performance of K-NN is highly influenced by the choice of K. Good result of the method based on neural networks in comparison with other methods is only a support for similar studies regarding dichotomic classification with the standard (Roszkowski, 2006; Witkowska, 2002).

The results of the parallel hybrid look promising. An exception is only the variant of the vote of three best models, the result of which was probably influenced by weak classification accuracy of the component models of the hybrid.

It is worth mentioning that a promising result of the classification accuracy obtained by the parallel hybrid does not mean that it is a universal classifier for all data. Good results were obtained for data on the wine quality test performed

Table 2. Accuracy of classification of the parallel classification hybrid for four vote variants and for six individual classification models (NN, SVM, NB, MARS, C and RT, K-NN). Self-evaluation

	Classification Methods	Number of classification	Number of samples	Accuracy
1	Vote of 2 best predictions	3462	4898	70.68%
2	Vote of all preditions	3406		69.54%
3	Best prediction	3369		68.78%
	NN	3369		68.78%
4	Vote of 3 best predictions	3365		68.70%
5	SVM	3329		67.97%
6	NB	3285		67.07%
7	MARS	3190		65.13%
8	C and RT	2868		58.55%
9	K-NN	2829		57.76%

by experts (Cortez et al. 2009). For each classifier it is possible to find input data for which the results of classification accuracy will be very good as well as such input data for which the results of classification accuracy will be very weak (Duch, 2000). However, the use of the hybrid might significantly minimize the risk of weak classification accuracy.

4. Hybrid selection of decision making attributes

The example presented in this section will analyze the advantages of hybridization applied to feature selection, where hybrid selection was expected to give more robust results and have lower complexity. The problems mentioned are the most significant ones relating to the standard methods of feature selection that either require time-consuming search in the feature space (impossible for high-dimensional problems) or due to selection based on a single criterion do not guarantee optimal selection or the results of selection are strongly biased by this criterion, thus being inferior when used by other processing methods e.g. at the stage of modeling.

One type of decision making processes are those which require creation of a model combining the input and output attributes. A typical step in this process is choosing significant input (decision making) attributes. The choice means a usual selection of attributes out of all potential ones, i.e. leads to reduction in dimensionality of the problem by choosing an optimal sub-set of attributes.

There are several reasons for dimensionality reduction, e.g.:

- reducing the curse of dimensionality – convergence of estimators used for machine training is lower for problems with high dimensionality than for problems with lower number of dimensions,
- reducing the memory requirements,
- increasing the training speed – calculation complexity of the machine training method is over-linear for most cases and, thus, reduction of dimensions is manifested with shorter training time,
- simplifying the model – models with simpler structure, created with the use of a smaller number of attributes are easier to understand and use,
- removing the information noise – unimportant attributes are the information noise, which worsens the quality of the model functioning,
- increasing the generalizing abilities – unnecessary attributes reduce the ability of the model to work with new samples.

Available methods of attribute selection can be divided into three types (Fig. 6), according to Guyon and Elissee (2003). The first one is filtration, whose functioning is based on calculation of the indexes measuring the strength of relation between pairs of attributes. Measures such as correlation, entropy or statistical tests are used in order to do this. Another type of selection algorithms are *wrappers*, which are multi-step algorithms testing the usefulness of attributes in the modeled process by examining their various combinations. The last type of selection algorithms are the inbuilt methods, which are components used inside other algorithms, e.g. selection of attributes used in decision trees for choosing attributes placed in the nodes.

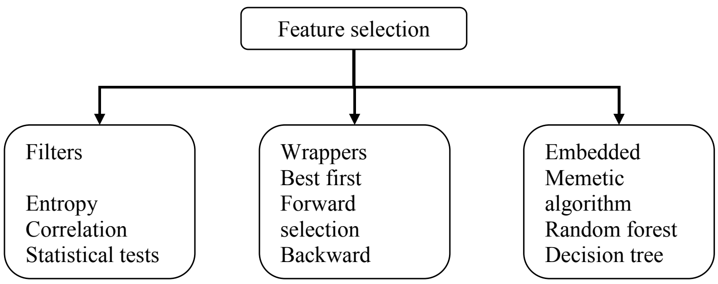


Figure 6. Division of attribute selection algorithms. Source: Pietruszkiewicz (2010).

All of these methods were subject to many studies, e.g. comparison of filters based on correlation with *wrappers* (Hall and Smith, 1999), compilation of selection results based on correlation, InfoGain and Chi2 (Zheng, 2004) or comparison of different entropy-based algorithms (Duch, 2004).

The use of filters and *wrapper* algorithms involves important problems. For filters, it is being based on a chosen criterion of attribute evaluation, which does not have to provide an optimal set. On the other hand, for *wrappers*, it is expensive search in the attribute space. For these reasons, this article examines the hybrid attribute selection algorithm. The hybrid contains three popular algorithms: InfoGain, GainRatio and Chi2 as well as weighing block, which, based on rankings provided by all methods, selects attributes by voting, taking into consideration multi-criteria evaluation of their importance.

Data used in presented experiments regard a decision making problem of risk evaluation of household bankruptcies (Rozenberg and Pietruszkiewicz, 2008), which is a necessary element of credit risk evaluation. All 17 input attributes might be divided in 3 groups: behavioral, demographic and geographical attributes (Table 3).

The output attribute in the modeled process was information on due payments of liabilities by households where three situations were considered: due payments, slight delays and default in payment of liabilities.

Selection of attributes was done with the hybrid method (labeled as Voting) and comparison with sets of attributes of the same cardinality, but determined with three component methods. Fitting of attributes was checked by constructing classifiers on them in the form of C4.5 decision trees and neural networks. The quality of the created models (accuracy in %) was tested with the use of 3-fold cross validation (Fig. 7) and training set (Fig. 8). On both figures the numbers on the attributes axis represent the best subsets of selected inputs (as in Table 3) for each method separately. It means that the accuracy was tested for various prediction models with the dimensionality ranging from 1 to 17.

It may be noticed in both figures that the hybrid selection of attributes made it possible to obtain good results without the selection risk typical to a single attribute evaluation criterion. This method provided stable results observed through in the high scores obtained for different numbers of attributes that verified the initial assumption about universality of the hybrid feature selection. Hence, it confirmed that the hybrid selection is more robust than features filtering and has lower complexity than wrappers. So, this approach might be used as a simple and safe method of dimensionality reduction.

5. Conclusions

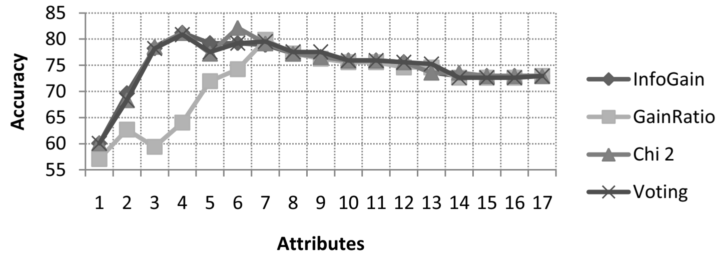
This article presented a hybrid approach used to support decision making process. The three experiments described herein, relating to different problems, confirmed the usefulness of hybrid algorithms supporting decisions and proved the synergy resulting from a proper combination of basic methods. A wide range of decision problems analyzed and solved by hybridization proves in general its effectiveness and promotes it as a simpler and more efficient solution to the computational decision support.

Table 3. Input attributes used in the feature selection experiment

Wi	Name	Variable type	Description	Criterion
W1	Household size	numerical	Number of people in the household	demographic
W2	Children	numerical	Number of children in the household	demographic
W3	Number of working people	numerical	How many people work in the family	demographic
W4	Income	qualitative	Level of total net income in the household	demographic
W5	Sex	qualitative	Respondent's sex	demographic
W6	Number of women	numerical	Number of women in the household	demographic
W7	Number of men	numerical	Number of men in the household	demographic
W8	Family age	numerical	Average family age	demographic
W9	Age	numerical	Respondent's age	demographic
W10	Education	numerical	Education, courses and training (sum of points for education of the whole family)	demographic
W11	Residence	qualitative	Place of residence	geographic
W12	Marital status	qualitative	Marital status	demographic
W13	Denomination	qualitative	Denomination	demographic
W14	Disability	qualitative	Is there a disabled person in the household?	demographic
W15	Illness	qualitative	Is there a chronically ill person in the household?	demographic
W16	Savings	qualitative	Who makes the decisions on family savings?	behavioral
W17	Loans	qualitative	Who makes decisions on loans?	behavioral

Source: Rozenberg and Pietruszkiewicz (2008)

a) C4.5



b) Neutral networks

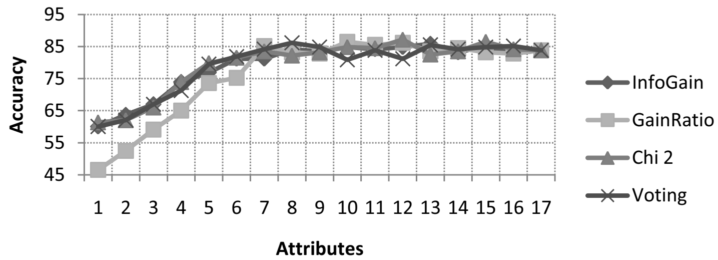
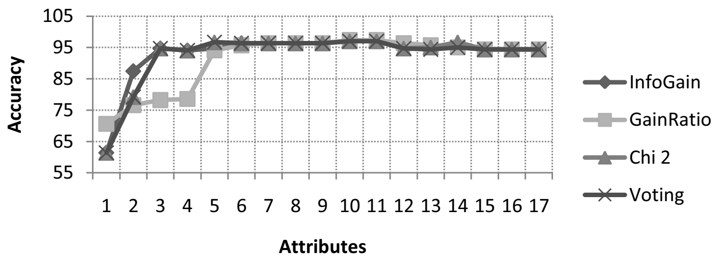


Figure 7. The accuracy of prognosis precision for C4.5 decision trees (a) and neural networks (b) for testing with 3-fold cross validation. Self-elaboration.

a) C.4.5



b) Neural networks

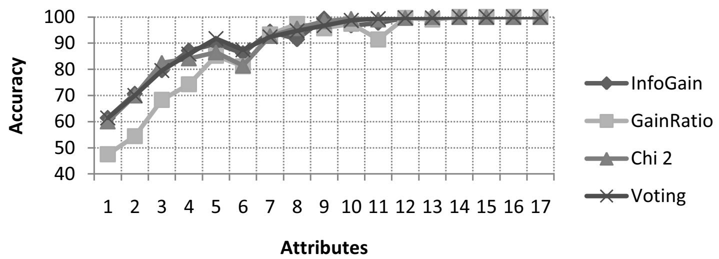


Figure 8. Results of prognosis precision for C4.5 decision trees (a) and neural networks (b) while testing with the training set. Self-elaboration.

Hybrid approach to the optimization of the function of several variables made it possible to obtain tools with very high effectiveness. The proposed cascade hybrid finds satisfactory solutions for different initial stages (specified only by the state of a pseudo-random generator) independent from the characteristics of the analyzed function. Owing to this it might be used successfully for a wide spectrum of optimization problems, including decision problems with properties which either make it impossible to use traditional methods or require an initial method "calibration" depending on specific conditions of the task.

The presented parallel classification hybrid made it possible to obtain good classification accuracy thanks to the use of results aggregation of six models. Among the considered voting variants of classification models, the best choice was voting of all models simultaneously.

The last example presented in the article was the hybrid selection of attributes. Hybridization of the selection involves using several filters evaluating the importance of attributes whose results co-decide about the final selection, thus, creating a multicriteria hybrid filter. The hybrid algorithm was characterized by the stability of results obtained for attributes selected with its use, reducing the risk of non-optimal choice of characteristic attributes subset for the selection based on one criterion.

The choice of various decision problems, presented in this paper was dictated by the will to obtain both effects i.e. wide possibilities of applications and the variety of hybrid algorithms. The research plans of the authors also include the use of hybridization for other types of decision problems and study on the possibilities of synergic combination of other types of algorithms.

References

Bagirov, A.M. and Rubinov, A.M. (2004) Local Optimization with Global Multidimensional Search for Descent. School of Information Technology and Mathematical Science, University of Ballarat.

Breiman, L., Friedman J., Stone Ch.J. and Olshen R.A. (1984) *Classification and Regression Trees.* Chapman and Hall/CRC.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T. and Reis, J. (2009) Modelling wine preferences by data mining from physicochemical properties. *Decision Support Systems* **47** (4), 547-553.

Duch, W., Wieczorek, T., Biesiada, J. and Blachnik, M. (2004) Comparison of feature ranking methods based on information entropy. *Proceeding of International Joint Conference on Neural Networks. Proc. of International Joint Conference on Neural Networks (IJCNN)*, Budapest 2004, IEEE Press, 1415-1420.

Duch, W., Korbicz, J., Rutkowski, L. and Tadeusiewicz, R. (eds.) (2000) Biocybernetics and biomedical engineering. Neural networks. Warsaw, Exit, vol.6 (in Polish).

Friedman, J. (1991) Multivariate adaptive regression splines (with discussion). *Annals of Statistics*, **19**, 1-141.

Guyon, I. and Elissee, A. (2003) An introduction to variable and feature selection. *Journal of Machine Learning Research* **3**, 1157-1182.

Hall, M.A. and Smith, L.A. (1999) Feature Selection for Machine Learning: Comparing a Correlation-based Filter Approach to the Wrapper. *Proceedings of the Twelfth International FLAIRS Conference*. AAAI Press, 235-239.

Pietruszkiewicz, W. (2010) The Hybrid Feature Selection for the Prediction of Household Bankruptcy. *Proceeding of 5th International Conference on Neural Networks and Artificial Intelligence*. Brest' State Technical University, Brest.

Pinter, J.D. (2002) Global and Convex Optimization in Modelling Environments: Compiler-Based, Excel, and Mathematica Implementations. *Research Report, PCS Inc.*
www.optimization-online.org/DB_FILE/2002/08/523.pdf.

Roszkowski, M. (2006) A method for identification of objects state with the use of neural networks based on the example of classifying the e-market clients. *Doctoral dissertation, Faculty of Information Technology,* Technical University of Szczecin, Szczecin.

Rozenberg, L. and Pietruszkiewicz, W. (2008) *The methodology of diagnosis and prognosis of household bankruptcy* (in Polish). Difin, Szczecin.

Twardochleb, M.R. and Rychcicki, R. (2009) Efficiency of hybrid optimization method in solving tasks of verified characteristics (in Polish). *Metody Informatyki Stosowanej*, 4, 135-145.

Witkowska, D. (2002) *Artificial neural networks and statistical methods. Selected financial problems.* C.H. Beck, Warsaw (in Polish).

Zheng, Z., Wu, X. and Srihari, R. (2004) Feature selection for text categorization on imbalanced data. *ACM SIGKDD Explorations Newsletter*, **6** (1). ACM, New York, 80-89.