Applied Soft Computing 40 (2016) 206–220

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A multi-objective genetic optimization for fast, fuzzy rule-based credit classification with balanced accuracy and interpretability

Marian B. Gorzałczany*, Filip Rudziński

Department of Electrical and Computer Engineering, Kielce University of Technology, Al. 1000-lecia P.P. 7, 25-314 Kielce, Poland

ARTICLE INFO

Article history:
Received 8 June 2015
Received in revised form
18 September 2015
Accepted 22 November 2015
Available online 30 November 2015

Keywords:
Accuracy and interpretability of credit classification systems
Financial decision support
Multi-objective evolutionary optimization
Fuzzy rule-based systems
Genetic computations

ABSTRACT

Credit classification is an important component of critical financial decision making tasks such as credit scoring and bankruptcy prediction. Credit classification methods are usually evaluated in terms of their accuracy, interpretability, and computational efficiency. In this paper, we propose an approach for automatic designing of fuzzy rule-based classifiers (FRBCs) from financial data using multi-objective evolutionary optimization algorithms (MOEOAs). Our method generates, in a single experiment, an optimized collection of solutions (financial FRBCs) characterized by various levels of accuracy-interpretability trade-off. In our approach we address the complexity- and semantics-related interpretability issues, we introduce original genetic operators for the classifier's rule base processing, and we implement our ideas in the context of Non-dominated Sorting Genetic Algorithm II (NSGA-II), i.e., one of the presently most advanced MOEOAs. A significant part of the paper is devoted to an extensive comparative analysis of our approach and 24 alternative methods applied to three standard financial benchmark data sets, i.e., *Statlog* (*Australian Credit Approval*), *Statlog* (*German Credit Approval*), and *Credit Approval* (also referred to as *Japanese Credit*) sets available from the UCI repository of machine learning databases (http://archive.ics.uci.edu/ml). Several performance measures including accuracy, sensitivity, specificity, and some number of interpretability measures are employed in order to evaluate the obtained systems. Our approach significantly outperforms the alternative methods in terms of the interpretability of the obtained financial data classifiers while remaining either competitive or superior in terms of their accuracy and the speed of decision making.

1. Introduction

Among the important issues in credit research area, credit classification that supports credit decision making plays a fundamental role [82]. In general, credit classification consists in categorizing credit applicants as either good (with credit accept decisions) or bad (with credit reject decisions). The classification is performed on the basis of credit applicant's characteristics such as annual income, type of bank account and bank balance, type of occupation, marital status, age and education. Naturally the required applicant's information may vary depending on, e.g., the amount of credit, its purpose, the legislation in a given country, etc. Credit classification is directly related to two critical financial decision making problems such as credit scoring (or rating) and bankruptcy prediction. Credit scoring modelling is focused on determining whether credit customers belong to either good or bad customer group. Sometimes, more specifically, the term 'application scoring' is used for predicting the repayment behavior of a credit applicant whereas the term 'performance scoring' is used for monitoring and predicting such a behavior of a customer with already granted credit [36]. Similarly, the objective of bankruptcy prediction modelling is to predict if a new customer (both individuals and companies can be considered) will go bankrupt or not. Hence, credit scoring and bankruptcy prediction can be regarded as corresponding to each other binary classification problems, i.e., the classification of a given applicant by a credit scoring model to, say, bad customer group corresponds, in a way, to forecasting by a bankruptcy prediction model that the applicant will go bankrupt [83].

Credit classification techniques are usually evaluated in terms of three aspects [36,99]: their accuracy, their transparency and interpretability, and their computational efficiency (the speed of classification). The accuracy (i.e., the ability to adequately represent the modelled decision making process and to generate the highest possible number of correct decisions) for obvious reasons is the fundamental requirement; even a small increase in the number of correct decisions may result in significant savings [87]. The transparency and interpretability (i.e., the ability

* Corresponding author. Tel.: +48 41 34 24 217; fax: +48 41 34 24 152.
E-mail addresses: m.b.gorzalczany@tu.kielce.pl (M.B. Gorzałczany), f.rudzinski@tu.kielce.pl (F. Rudziński).

http://dx.doi.org/10.1016/j.asoc.2015.11.037
1568-4946/

to generate a compact and understandable explanation and justification of the proposed decisions including the selection of the most essential input attributes) is important for both a decision maker and applicant. The decision maker can verify and possibly improve his or her knowledge and expertise in a given decision making domain, whereas the applicant obtains clear and understandable explanation of the decision made. It is particularly important when credit has been denied to the applicant since, according to the legislature in some countries, vague and indefinite reasons of credit denial are illegal [63]. As far as the speed of classification is concerned, according to [36], "an instant decision is much more appealing to a potential borrower than is having to wait for several days". Obviously, if the credit decisions are made within fractions of seconds or seconds by various computerized systems, the differences between them in the classification speed are meaningless from the practical point of view.

Huge credit scoring databases are available, e.g., according to [36], databases with well over 100,000 applicants characterized by more than 100 attributes are quite common, whereas performance scoring databases that contain information about past repayment behavior can even be much larger. It is obvious, that a great amount of knowledge on various aspects of financial decision making is buried within such data sets. Therefore, the development of effective financial knowledge discovery tools in such data or financial data mining methods has a strong rationale. The knowledge discovery tools are able to automatically reveal valid and understandable patterns, trends and decision making mechanisms hidden in financial data. As fas as the representation of such a knowledge is concerned, the most useful structures are conditional rules and particularly linguistic fuzzy conditional rules [17]. Their essential advantages are high modularity and readability as well as easy-to-grasp comprehension and interpretation by humans. However, the transparency and interpretability of rule-based systems become limited when excessive numbers of complex rules (with many antecedents) are generated. Therefore, both the accuracy as well as the transparency and interpretability (as discussed earlier in this section) are the main objectives in designing rule-based systems (in particular, fuzzy rule-based classifiers (FRBCs)) from financial data sets. Taking that into account, in this paper we present an original approach to designing FRBCs that is equipped with an effective mechanism of their accuracy-interpretability trade-off optimization. It means that FRBCs characterized by various optimized levels of such a trade-off can be obtained – including highly accurate systems of better interpretability (in regard to the existing approaches) – for financial decision support. It is worth emphasizing that such aspects have not yet been addressed in the literature on financial decision making. In order to achieve the assumed goals: (a) an automatic designing of FRBCs from financial data will be considered as a structure and parameter optimization task or search problem in a large search space and (b) a multi-objective evolutionary optimization algorithm (MOEOA) will be employed to solve that task. Evolutionary methods including genetic algorithms have been successfully applied for solving various complex-space-search and optimization problems, e.g. [14,28–31,40,73]. As far as the FRBCs' accuracy-interpretability trade-off is concerned, both objectives are to some extent complementary/contradictory ones. Hence, it is possible to formulate a single-objective evolutionary optimization task using a fitness function defined as a combination of both objectives. As shown in [30] such a solution gives very good FRBCs' accuracy-interpretability trade-offs in many areas of applications. Nevertheless, for obvious reasons, it explores only a limited part of the whole search space and thus some valuable solutions may never be discovered. MOEOAs that generate in a single run of the algorithm a set of solutions (approximating Pareto-optimal solutions) characterized by various levels of accuracy-interpretability trade-off are much better suited for the considered task.

The proposed hybrid and synergistic combination of FRBCs and MOEOAs that maximizes their advantages while minimizing their limitations belongs to the computational intelligence (CI) [25,49] area that is a modern extension and generalization of the conventional artificial intelligence (AI) field. According to the authors of [12] (having extensive 14-years experience in the financial industry) "...AI techniques have rarely been used in credit rating research to generate comprehensive decision rules, particularly when compared with statistical methods...". Therefore, our approach contributes to filling the knowledge gap related to effective modelling (with optimally balanced accuracy and interpretability) of financial decision making problems by means of modern hybrid computational-intelligence techniques.

In MOEOA-based automatic designing of FRBCs from data still several open problems exist [19]. First, some measures of FRBCs' interpretability must be defined (there are already well-defined and generally accepted measures of FRBCs' accuracy). In the case of system's interpretability, which is much more subjective property than its accuracy, usually two aspects are considered [21]: the complexity of FRBCs' rule bases and the semantics associated with membership functions of fuzzy sets representing linguistic terms that describe particular systems' attributes. We propose an original measure of complexity-dependent interpretability including the average length of rules, the overall number of attributes, and the overall number of fuzzy sets (linguistic terms) in the system. We also present a computationally efficient implementation of the so-called strong fuzzy partition (SFP) condition [74] that perfectly meets the semantics-dependent interpretability constraints [21]. Second, we propose an original, special-coding-free, direct representation of the classifier's fuzzy rule base structure and original crossover and mutation genetic operators for its processing by the MOEOA used. Third, we present an implementation of our ideas in the framework of the well-known and presently one of the most advanced MOEOAs, i.e., Non-dominated Sorting Genetic Algorithm II (NSGA-II) [15].

A broad comparative analysis of our method and as many as 24 alternative approaches applied to three well-known benchmark financial data sets (*Statlog* (*Australian Credit Approval*), *Statlog* (*German Credit Approval*), and *Credit Approval*) is also carried out in this paper. The *Credit Approval* data set is sometimes referred to as *Japanese Credit* set (see, e.g., [83,94]); for this reason, we'll call it *Credit Approval* (*Japanese*) data set. All data sets are available from the UCI repository of machine learning databases (http://archive.ics.uci.edu/ml). The results of the aforementioned 24 alternative techniques are collected in three papers published in recent years in this journal, i.e., [83] of 2014 by C.-F. Tsai et al., [10] of 2012 by S.-Y. Chang and T.-Y. Yeh, and [69] of 2011 by Y. Peng et al. as well as in two other papers, i.e., [99] of 2013 by X. Zhu et al. and [94] of 2008 by L. Yu et al. Therefore, our paper continues and extends the up-to-date research in the important field of financial data mining and decision support.

The remainder of the paper is organized as follows. Section 2 presents a review of the related works on financial data classification methods and the accuracy-interpretability trade-off issues in designing (fuzzy) rule-based systems from data. Section 3 presents our FRBC (its fuzzy knowledge base and fuzzy approximate inference engine) for financial decision making. Main components of the FRBC's genetic learning and MOEOA-based optimization procedure are discussed in Section 4. Section 5 presents the experimental results and the aforementioned broad comparative analysis. Finally, Section 6 formulates some conclusions and outlines our further research directions.

2. Related works

In this paper, we propose a fuzzy rule-based classification approach for financial decision support with genetically optimized trade-off between its accuracy and interpretability. For this reason, in the review presented in this section – after a brief survey of accuracy-oriented techniques – we consider alternative interpretable (usually "if-then"-rule-based) solutions of the considered problem. We also briefly address the important aspect that has not yet been covered in the literature on financial decision support, i.e., the accuracy-interpretability trade-off optimization in designing (fuzzy) rule-based systems from data.

Many credit risk evaluation and credit classification techniques concentrate almost exclusively on the accuracy aspects neglecting the interpretability issues. From among them, statistical and optimization techniques were first employed, including discriminant analysis [5,70], logistic regression (logit) [37,48,89], probit regression [32], *k*-nearest neighbor method [11,38,39], linear programming [22,35], and integer programming [52]. In general, according to [99], the main drawback of statistical methods is their non-sufficiently high accuracy, whereas their main advantage is simplicity. Some statistical models are still being employed in financial decision support, see, e.g., [47,48] for financial distress classification. Moreover, recently, an extensive review [8] of traditional statistical methods for the bankruptcy prediction has been published.

Several accuracy-oriented and black-box-type methods for financial decision support have emerged in the framework of the computational intelligence (CI) area being a modern extension and generalization of traditional artificial intelligence (AI) field. They include: artificial neural networks (ANNs) with various topologies [6,50,62,77,90] (according to [57], the most widely applied models), support vector machines (SVMs) [55,76,98] and least square SVMs [93,95,96] (according to [57], one of the most hotly researched models recently (besides hybrid systems – see below)), genetic algorithms (GAs) [1,53,67], genetic programming [56,66], artificial immune systems [10], and some case-based reasoning (C-BR) approaches [57,59]. In order to enhance the effectiveness of stand-alone black-box-type approaches, their ensembles as well as various hybrid combinations are considered. The ensemble systems are based on combination (usually, majority voting, bagging or boosting combination methods [60,83] are used) of single approaches, yielding, e.g., ANN ensembles [51,65,84,88], SVM ensembles [65,68], *k*-nearest neighbor ensembles [65], C-BR ensembles [58], etc. (see [83] for review). In turn, hybrid systems are constructed by (i) cascading different approaches, (ii) combining clustering and classification methods, or (iii) integrating, in a synergistic way, different techniques into a "single" approach [60]. Selected examples of such systems from particular groups are: (i) ANNs and discriminant analysis hybrid [54], GAs and C-BR hybrid [3], (ii) *k*-means and ANNs hybrid [42], (iii) GAs and SVMs hybrid [64], GAs and ANNs hybrid [43], and fuzzy sets and SVMs hybrid [86] (see [60] for review). In general (cf. [99]), the aforementioned approaches perform satisfactorily in terms of accuracy, but their results are not (or, hardly) interpretable in terms of transparent and clear decision mechanisms based on selected, most essential input attributes.

Besides the accuracy and speed of the credit classification methods, their good interpretability has become an increasingly important requirement in recent years [99]. The interpretability and explanatory power of a given approach represents its ability to provide a user with understandable and compact explanation and justification of generated decisions based on the selection of the most essential decision attributes. As briefly discussed in the Introduction of the paper, it is important for both a decision maker and an applicant. Single techniques equipped with some explanatory power that have already been applied to financial decision making problems include decision trees (DTs) and rough sets (RSs).

DT methods use recursive partitioning of a given data set and some pruning measures to generate a top-down tree structure, which in turn results in a set of "if-then" rules operating on selected input attributes (both numerical and categorical attribute can be processed). Single DT approaches were used in financial decision support problems mainly in earlier research [9,20]. In recent years, various DT ensembles (using bagging [85,97] and boosting [4,79] combination methods) and DT-based hybrid systems [80] dominate in this field. In turn, RS approaches use lower and upper approximations to represent any vague concept and the idea of an indiscernibility relation (later, a dominance relation was also used) for inducing "if-then" decision rules based on reduced set of input attributes. Similarly as in the case of DTs, in financial decision making single RS techniques were applied mainly in earlier research [16,27], whereas recently various hybrid combinations of RSs and other methods are developed, including RSs and random forests [92] (a cascade approach), RSs and SVMs [91] (a clustering-and-classification approach), as well as RSs and ANNs [2] (an integrated approach). Review papers, such as [36,60,71,82], present more details on some of these techniques.

An effective way of the representation of a human-oriented financial knowledge in the form of linguistic decision rules is offered by fuzzy rule-based systems (FRBSs). However, they themselves are not able to extract such a knowledge from data and thus they usually operate within hybrid solutions. The first generation of such solutions are neuro-fuzzy systems and the second – genetic-fuzzy ones (in the considered financial research area, see, e.g. [41,61], respectively). In this paper, we propose even more advanced approach, i.e., the hybrid of FRBSs and multi-objective evolutionary optimization algorithms (MOEOAs) for an effective optimization of two essential and contradictory/complementary objectives such as accuracy and interpretability in financial decision support. Such aspects have not yet been addressed in the literature on financial decision systems.

Early approaches to designing FRBSs from data (viz. neuro-fuzzy systems; see, e.g., [23,24,26,33,75]) concentrated, first of all, on systems' accuracy. In turn, in order to improve their interpretability, the removal (usually, in a manual way) of low active rules and, possibly, some antecedents from selected rules (the rule-based pruning) was performed. The next-generation techniques of designing FRBSs from data (viz. genetic-fuzzy systems; see, e.g., [14,29,30,40,78]) are much more flexible in terms of the optimization objectives that are represented by appropriate fitness functions (both the FRBSs' parameter and structure optimization can be carried out). Initially, single-objective optimization tasks were considered with the fitness function defines as a weighted sum of several components representing various aspects of system's accuracy and interpretability, e.g., [30,44,45]. At present, Pareto-based MOEOAs [15,46] are of particular interest; see, e.g., recent review papers [13,19]. Our paper attempts to demonstrate high usefulness of MOEOA-based FRBSs in financial data mining and financial decision support.

3. Fuzzy rule-based classifier (FRBC) for financial decision making

In this section, we present two main components of the proposed FRBC, i.e., its knowledge base and its fuzzy approximate inference engine that generates FRBC's response to input data. Moreover, a special-coding-free and computationally efficient representation of a rule base (a subcomponent of the knowledge base) is also proposed.

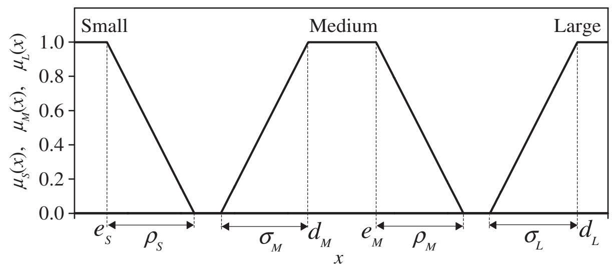

Fig. 1. Trapezoidal membership functions of S-type, M-type, and L-type fuzzy sets and their parameters.

3.1. FRBC's knowledge base

A FRBC with n input attributes (in short, attributes or, alternatively, variables, features, etc.) $x_1, x_2, \ldots, x_n$ and an output, which has the form of a fuzzy set over the set $Y = \{y_1, y_2, \ldots, y_c\}$ of c class labels is considered. Both numerical and categorical attributes x_i ($x_i \in X_i$), $i = 1, 2, \ldots, n$ can be processed.

The FRBC's knowledge base consists of R linguistic fuzzy classification rules (the overall number R of rules in the knowledge base changes during the genetic learning and optimization process). The rth rule ($r = 1, 2, \ldots, R$) has the following form:

$$\begin{aligned}
\textbf{IF}\quad & [x_1\ is[not]_{(sw_1^{(r)}<0)}\ \ A_{1,|sw_1^{(r)}|}]_{(sw_1^{(r)}\neq 0)}\ \textbf{AND}\ldots\textbf{AND}\\
& [x_i\ is[not]_{(sw_i^{(r)}<0)}\ \ A_{i,|sw_i^{(r)}|}]_{(sw_i^{(r)}\neq 0)}\ \textbf{AND}\ldots\textbf{AND}\\
& [x_n\ is[not]_{(sw_n^{(r)}<0)}\ \ A_{n,|sw_n^{(r)}|}]_{(sw_n^{(r)}\neq 0)}\\
& \textbf{THEN}\ y \text{ is } B_{(singl.)j^{(r)}}.
\end{aligned} \tag{1}$$

The components $[expression]_{(condition)}$ in (1) represent conditional inclusion, i.e., *expression* is included into the rule if and only if *condition* is fulfilled. $|\cdot|$ returns the absolute value. $sw_i^{(r)} \in \{0, \pm 1, \pm 2, \ldots, \pm a_i\}$, where a_i denotes the number of fuzzy sets (linguistic terms) defined for the ith attribute, $i = 1, 2, \ldots, n$. $sw_i^{(r)}$ is a switch, which controls the presence or absence of the ith attribute in the rth fuzzy rule. If $sw_i^{(r)} = 0$ then the ith attribute is excluded from the rth rule (or equivalently, is not active in that rule). On the other hand, if $sw_i^{(r)} \neq 0$ then the ith attribute is included (or, active) in the rth rule, however, (i) for $sw_i^{(r)} > 0$, the component $[x_i \text{ is } A_{ik_i}]$ is included in the rule ($k_i = |sw_i^{(r)}|$), and (ii) for $sw_i^{(r)} < 0$, the component $[x_i \text{ isnot } A_{ik_i}]$ is included in the rule; not A_{ik_i} ($\bar{A}_{ik_i}$, for short) is the complement of A_{ik_i}, i.e., $\mu_{\bar{A}_{ik_i}}(x_i) = 1 - \mu_{A_{ik_i}}(x_i)$, where $\mu_{A_{ik_i}}(x_i)$ and $\mu_{\bar{A}_{ik_i}}(x_i)$ are the membership functions of fuzzy sets A_{ik_i} and $\bar{A}_{ik_i}$.

Let $F(X_i)$ denote the family of all fuzzy sets defined in the universe X_i, which represents the ith attribute. For each numerical attribute x_i, $i \in \{1, 2, \ldots, n\}$, a collection of a_i fuzzy sets $A_{ik_i} \in F(X_i)$, $k_i = 1, 2, \ldots, a_i$ is defined. These sets create a fuzzy partition $FP(X_i) = \{A_{i1}, A_{i2}, \ldots, A_{ia_i}\}$ of X_i. The first set A_{i1} from that collection represents the linguistic term "Small" and is referred to as S-type fuzzy set. The last set A_{ia_i} – referred to as L-type one – represents the term "Large", whereas the remaining $a_i - 2$ fuzzy sets (the so-called M-type sets) represent several "Medium"-type linguistic terms, i.e., "Medium1", "Medium2", etc. Trapezoidal membership functions of S-, M-, and L-type fuzzy sets are presented in Fig. 1. Each categorical attribute $x_i \in X_i = \{x_{i1}, x_{i2}, \ldots, x_{ia_i}\}$ is characterized by a_i fuzzy singletons $A_{(singl.)ik_i} \in F(X_i)$, $k_i = 1, 2, \ldots, a_i$ represented by the following membership functions:

$$\mu_{A_{(singl.)ik_i}}(x_i) = \begin{cases} 1, & \text{for } x_i = x_{ik_i}, \\ 0, & \text{elsewhere.} \end{cases} \tag{2}$$

$B_{(singl.)j^{(r)}}$ in (1) is the fuzzy singleton that represents the class label $y_{j^{(r)}}$, $j^{(r)} \in \{1, 2, \ldots, c\}$; its membership function $\mu_{B_{(singl.)j^{(r)}}}$ is defined in an analogous way as in (2), i.e., $\mu_{B_{(singl.)j^{(r)}}}(y) = 1$ for $y = y_{j^{(r)}}$ and 0 elsewhere.

A simple example showing the FRBC's knowledge base (1) in the context of financial applications is now presented. Four fuzzy rules enable us to classify applicants as either "good credit risk" ones or "bad credit risk" ones based on two categorical attributes: "status of existing checking account" ("status", for short) and "credit history" as well as two numerical attributes: "credit amount" and "installment rate in percentage of income" ("installment", for short), in the following way:

$$\begin{aligned}
&\textbf{IF}\quad [\text{“status''}\,\textit{is}\,\text{“no checking account''}]\,\textbf{THEN}\,\text{“good credit risk''},\\
&\textbf{IF}\quad [\text{“credit history''}\,\textit{is}\,\text{“other credit existing (not at this bank)''}]\\
&\qquad \textbf{THEN}\,\text{“good credit risk''},\\
&\textbf{IF}\quad [\text{“credit amount''}\,\textit{is}\,\text{“Large''}]\,\textbf{THEN}\,\text{“bad credit risk''},\\
&\textbf{IF}\quad [\text{“status''}\,\textit{is}\,\text{“less than 0''}]\,\textbf{AND}[\text{“installment''}\,\textit{is}\,\text{“Large''}]\\
&\qquad \textbf{THEN}\,\text{“bad credit risk''}
\end{aligned} \tag{3}$$

Therefore, the system (3) identifies applicants without checking accounts or with credits in other banks as the good-credit-risk ones and those applying for large credits or with negative bank balance and large installment rates as the bad-credit-risk ones (see Section 5.3 for more details).

During the process of genetic learning and optimization of the FRBC, particular attributes in various rules are being disabled and re-enabled by appropriate changes of their switches $sw_i^{(r)}$ (including possible replacement of A_{ik_i} by $\bar{A}_{ik_i}$ and vice versa). In different rules, different (in general) nonempty subsets of attributes are included. The rules with all switched-off attributes (the "empty" rules) are removed from the system. In such a way, the final minimal collection of rules contains only the most compact subset of selected, essential attributes for the financial data classification purposes. In experiments presented later in the paper, the rules (1) will be presented without switches $sw_i^{(r)}$; simply, only the active attributes will be shown.

Two basic components can be distinguished in the FRBC's knowledge base, i.e., its rule base (RB), which contains the information on the structure of particular rules and its data base (DB), which stores the parameters of fuzzy sets occurring in the rules. The RB-DB distinction is also related to the structure and parameter optimization of the FRBC. In our approach, the RB is represented by the following set of parameters:

$$RB = \{sw_1^{(r)}, sw_2^{(r)}, \ldots, sw_n^{(r)}, j^{(r)}\}_{r=1}^{R}. \tag{4}$$

Therefore, we propose simple, direct, special-coding-free and thus computationally efficient RB's representation. Dedicated original genetic operators (crossover, mutation, and the so-called repairing) for its processing are presented in the subsequent part of the paper.

The DB in our approach contains parameters: e_{i1}, ρ_{i1} (for S-type fuzzy set), $d_{i2}, e_{i2}, \sigma_{i2}, \rho_{i2}$ (for the first M-type fuzzy set), $\ldots$, d_{i,a_i-1}, $e_{i,a_i-1}, \sigma_{i,a_i-1}, \rho_{i,a_i-1}$ (for the last M-type fuzzy set), and d_{ia_i}, σ_{ia_i} (for the L-type fuzzy set) of rule-antecedents membership functions, $i = 1, 2, \ldots, n$, for numerical attributes. The DB contains also information on domains of categorical attributes $X_i = \{x_{i1}, x_{i2}, \ldots, x_{ia_i}\}$ and the set of class labels $Y = \{y_1, y_2, \ldots, y_c\}$; these parameters, obviously, are not being tuned during the learning process.

3.2. FRBC's fuzzy approximate inference engine

During the genetic learning, an evaluation of particular individuals (fuzzy knowledge bases) in the framework of our Pittsburgh-type approach, cf. [14,40], must be performed in each generation. For this purpose: (i) a fuzzy-set-theory representation of the set of linguistic rules (1) must be formulated and (ii) a fuzzy approximate inference scheme must be employed. In general, there are two different interpretations of linguistic "if-then" rules [18,25]. The first one, referred to as the conjunction-based approach, treats particular rules as independent local statements and thus they are aggregated disjunctively. In the second one, referred to as the implication-based or logical approach, the rules are treated as fuzzy constraints and, therefore, based on the minimum-specificity principle, they are aggregated conjunctively; cf. [17]. Fuzzy models are learned from collections of input–output data pairs, which are independent local examples of the system's behavior. Therefore, the conjunction-based model is more consistent than the logical one with the nature of the learning data and, for this reason, it is almost exclusively used (especially the Mamdani's model with min-type t-norms for combining many antecedents and antecedents with consequents of the rules as well as with max-type t-conorm for rule aggregation). As far as fuzzy approximate inference schemes are considered, two of them prevail in the literature (see [7]), i.e., compositional rule of inference and similarity-based reasoning; both can be implemented in our approach. Concluding, employing the aforementioned Mamdani's model, we obtain – for the input numerical data $\mathbf{x}'=(x'_1, x'_2, \ldots, x'_n)$ – a FRBC's fuzzy-set response B' characterized by its membership function $\mu_{B'}(y)$, $y \in Y=\{y_1, y_2, \ldots, y_c\}$:

$$\mu_{B'}(y) = \max_{r=1,2,\ldots,R} \mu_{B'^{(r)}}(y) = \max_{r=1,2,\ldots,R} \min[\alpha^{(r)}, \mu_{B_{(singl.)j^{(r)}}}(y)], \tag{5}$$

where

$$\alpha^{(r)} = \min_{\substack{i=1,2,\ldots,n,\\ sw_i^{(r)}\neq 0}} \alpha_i^{(r)}, \tag{6}$$

and

$$\alpha_i^{(r)} = \begin{cases} \mu_{A_{i,sw_i^{(r)}}}(x'_i), & \text{for } sw_i^{(r)} > 0, \\ \mu_{\bar{A}_{i,|sw_i^{(r)}|}}(x'_i), & \text{for } sw_i^{(r)} < 0. \end{cases} \tag{7}$$

$\alpha^{(r)}$ is the activation degree of the rth fuzzy rule by the input numerical data $\mathbf{x}'$, whereas $\alpha_i^{(r)}$ for i such that $sw_i^{(r)} \neq 0$ are the activation degrees of particular attributes in that rule.

Usually, a non-fuzzy response y' of the FRBC is required; it is calculated in the following way:

$$y' = \underset{y \in Y}{\mathrm{argmax}}\, \mu_{B'}(y). \tag{8}$$

4. Main components of genetic learning and optimization of FRBC designed from financial data

This section presents basic components of the FRBC's genetic learning and optimization procedure such as learning data set, genetic optimization objectives and genetic operators. Original measure of FRBC's complexity-related interpretability, efficient implementation of some condition on semantics-related interpretability, and original genetic operators for the processing of RBs (rule bases) (4) are also introduced.

4.1. Learning data set

The FRBC is designed from K input–output learning samples:

$$\mathbf{L} = \{\mathbf{x}_k^{(lrn)}, y_k^{(lrn)}\}_{k=1}^{K}, \tag{9}$$

where $\mathbf{x}_k^{(lrn)} = (x_{1k}^{(lrn)}, x_{2k}^{(lrn)}, \ldots, x_{nk}^{(lrn)}) \in \mathbf{X} = X_1 \times X_2 \times \ldots \times X_n$ ($\times$ stands for the Cartesian product of ordinary sets) is the set of input attributes and $y_k^{(lrn)}$ is the corresponding class label ($y_k^{(lrn)} \in Y$) for the kth data sample.

4.2. Objectives of FRBC's genetic optimization

4.2.1. FRBC's accuracy

The following measure (an objective function subject to maximization) of the accuracy of the FRBC is used:

$$Q_{ACC}^{(lrn)} = \frac{1}{K}\sum_{k=1}^{K} \delta(y'_k, y_k^{(lrn)}), \tag{10}$$

and

$$\delta(y'_k, y_k^{(lrn)}) = \begin{cases} 1, & \text{for } y'_k = y_k^{(lrn)}, \\ 0, & \text{otherwise}, \end{cases} \tag{11}$$

where y'_k is the class label that is the system's non-fuzzy response (8) for the learning data sample $\mathbf{x}_k^{(lrn)}$ and $y_k^{(lrn)}$ is the desired class label from that sample taken from $\mathbf{L}$ (9). Therefore, $\sum_{k=1}^{K} \delta(y'_k, y_k^{(lrn)})$ is the number of correctly classified learning data samples and $Q_{ACC}^{(lrn)} \in [0, 1]$. An analogous measure $Q_{ACC}^{(tst)}$ for the test data can also be calculated.

4.2.2. FRBC's interpretability

4.2.2.1. Complexity-related interpretability. We propose a new measure (an objective function subject to maximization) of the FRBC's complexity-related interpretability in the following form:

$$Q_{INT} = 1 - Q_{CPLX}, \tag{12}$$

where $Q_{CPLX} \in [0, 1]$ describes the fuzzy rule base complexity ($Q_{CPLX}=0$ and 1 represent minimal and maximal complexity, respectively). Q_{CPLX} is determined on the basis of three indices that measure the average complexity of particular rules (Q_{RATR}) and the complexity of the whole system in terms of active attributes (Q_{ATR}) and active fuzzy sets (Q_{FS}):

$$Q_{CPLX} = \frac{Q_{RATR} + Q_{ATR} + Q_{FS}}{3}, \tag{13}$$

where

$$Q_{RATR} = \frac{1}{R}\sum_{r=1}^{R} \frac{n_{ATR}^{(r)} - 1}{n-1}, \quad n > 1, \tag{14}$$

$$Q_{ATR} = \frac{n_{ATR} - 1}{n-1}, \quad n > 1, \tag{15}$$

$$Q_{FS} = \frac{n_{FS} - 1}{\sum_{i=1}^{n} a_i - 1}, \quad n > 1. \tag{16}$$

$n_{ATR}^{(r)}$ is the number of active attributes (rule antecedents) in the rth rule. $Q_{RATR} \in [0, 1]$; $Q_{RATR}=0$ means that the rule base contains only the rules with one attribute, whereas for $Q_{RATR}=1$ – all rules have all attributes in their antecedent parts. n_{ATR} and n_{FS} are the numbers of active attributes and fuzzy sets (linguistic terms), respectively, in the whole system. $Q_{ATR} \in [0, 1]$ and $Q_{ATR}=0$ means that only one attribute is active in the whole system, while $Q_{ATR}=1$ – that all attributes are active. Q_{FS} has the same interpretation for fuzzy sets in the system.

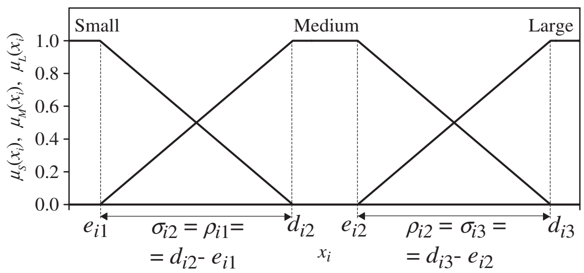

Fig. 2. Implementation of three-fuzzy-set SFP.

4.2.2.2. Semantics-related interpretability. In our approach, the semantics-related interpretability of the FRBCs is addressed by implementing the SFP (strong fuzzy partition) requirement (see remarks in the Introduction). Fuzzy partitions in which the sum of the values of all membership functions for any domain value is equal to 1 (referred to as SFPs), satisfy the desired semantics-related interpretability demands at the highest level. We propose simple and thus computationally efficient implementation of SFP condition for trapezoidal membership functions as follows (see Fig. 2 for three-set SFP of x_i):

$$\sigma_{ik_i} = \rho_{i,k_i-1} = d_{ik_i} - e_{i,k_i-1}, \quad k_i = 2, 3, \ldots, a_i \tag{17}$$

and, obviously,

$$e_{i1} \le d_{i2} \le e_{i2} \le \ldots \le d_{i,a_i-1} \le e_{i,a_i-1} \le d_{ia_i}, \quad i = 1, 2, \ldots, n. \tag{18}$$

4.3. Genetic operators for FRBC processing

A population of single individuals (each of them represents one FRBC) is processed during the genetic learning process. Each individual consists of two parts that represent the RB (rule base) (4) and DB (data base), respectively, of a single FRBC as shown in Section 3 of the paper. Non-binary genetic operators – defined separately for the RB (4) and DB processing – will now be presented.

4.3.1. RB processing

4.3.1.1. Original RB-crossover operator (RB-Cro). The proposed RB-Cro operator processes two individuals (two RBs) of R_1 and R_2 rules, respectively, by performing one of five randomly selected sub-operations RB-Cro1, RB-Cro2, . . ., or RB-Cro5:

1) RB-Cro1 (exchange of many rules): In the first stage, for the rth rule in both RBs, $r = 1, 2, \ldots, min(R_1, R_2)$, the so-called *random-switch condition* (equivalent to the random selection of 1 from the set {0, 1}) is checked. If this condition is fulfilled, the rth rules from both RBs are exchanged. In the second stage, each of the remaining rules of the larger RB, assuming that the *random-switch condition* is fulfilled, is moved to the smaller RB.
2) RB-Cro2 (exchange of a single rule): Analogous as RB-Cro1 but the activities of RB-Cro1 are performed unconditionally only once for randomly selected rth rule in the larger RB.
3) RB-Cro3 (exchange of many fuzzy sets in many fuzzy rules): If the same condition as in the first stage of RB-Cro1 is fulfilled then – for the ith attribute, $i = 1, 2, \ldots, n$ and for the output class label – the *random-switch condition* is run independently again. If it is fulfilled, the fuzzy sets describing a given attribute or class label in both RBs are exchanged.
4) RB-Cro4 (exchange of many fuzzy sets in a single rule): Analogous as RB-Cro3 but the activities of RB-Cro3 are performed unconditionally only once for randomly selected rth rule in both RBs.
5) RB-Cro5 (exchange of a single fuzzy set): Analogous as RB-Cro4 but the activities of RB-Cro4 are performed unconditionally only once for randomly selected ith attribute or output class label.

4.3.1.2. Original RB-mutation operator (RB-Mut). The proposed RB-Mut operator processes a single RB by performing one of four randomly selected sub-operations: RB-Mut1, RB-Mut2, RB-Mut3, or RB-Mut4:

1) RB-Mut1 (rule insertion): It inserts into RB a new rule (1) with randomly selected values of switches sw_i and class label j ($sw_i \in \{0, \pm 1, \ldots, \pm a_i\}$, $i = 1, 2, \ldots, n$, $j \in \{1, 2, \ldots, c\}$).
2) RB-Mut2 (rule deletion): It removes a randomly selected rule from the RB.
3) RB-Mut3 (change of a single fuzzy set): It randomly selects one rule from the RB and its one ith attribute or output class label j. Next, it randomly selects a new value of switch sw_i or class label j.
4) RB-Mut4 (change of an attribute in a fuzzy rule): It randomly selects: one rule in the RB, its one active (i.e., with $sw_{i_1} \neq 0$) and one non-active (i.e., with $sw_{i_2} = 0$) attributes. Then, the first attribute is switched off ($sw_{i_1} = 0$) and the second is switched on ($sw_{i_2} \neq 0$) in that rule.

4.3.1.3. Original RB-repairing operator (RB-Rep). The proposed RB-Rep operator processes all RBs of descendant population by performing the following three operations:

1) RB-Rep1 (removal of empty fuzzy rules): The rules with all switched-off attributes ("empty" rules) are removed from RB.
2) RB-Rep2 (removal of rule duplicates): All identical rules, except for one, are removed from RB.
3) RB-Rep3 (replenishment of RB): The principle of "at least one fuzzy rule per class" must be preserved. For each class label that is not represented in RB's consequents, one rule is added to RB using the RB-Mut1 sub-operation.

4.3.2. DB processing

4.3.2.1. DB-crossover operator (DB-Cro). DB-Cro operator randomly selects two fuzzy sets, each from one DB. Let $A^{(1)}_{ik_i}$ be the fuzzy set from the first DB and $A^{(2)}_{ik_i}$ – from the second DB, $i \in \{1, 2, \ldots, n\}$, $k_i \in \{1, 2, \ldots, a_i\}$. $d^{(1)}_{ik_i}$, $e^{(1)}_{ik_i}$, $\sigma^{(1)}_{ik_i}$, $\rho^{(1)}_{ik_i}$ are the membership function parameters of $A^{(1)}_{ik_i}$ and $d^{(2)}_{ik_i}$, $e^{(2)}_{ik_i}$, $\sigma^{(2)}_{ik_i}$, $\rho^{(2)}_{ik_i}$ are the analogous parameters for $A^{(2)}_{ik_i}$ (assuming that both fuzzy sets are the M-type sets; analogously for S-type and L-type fuzzy sets). DB-Cro, first, calculates linear combinations of d- and e-parameters to obtain their new values:

$$\underset{new}{d^{(1)}_{ik_i}} = \gamma d^{(1)}_{ik_i} + (1-\gamma) d^{(2)}_{ik_i}, \tag{19}$$

and, analogously, for $\underset{new}{e^{(1)}_{ik_i}}$, $\underset{new}{d^{(2)}_{ik_i}}$, and $\underset{new}{e^{(2)}_{ik_i}}$. $\gamma \in [0, 1]$ is a randomly selected value independently for each calculation. The new values of those parameters must also fulfil condition (18). Then, DB-Cro calculates the new values of σ- and ρ-parameters using formula (17).

4.3.2.2. DB-mutation operator (DB-Mut). DB-Mut operator randomly selects one fuzzy set from the DB and one of its two parameters d and e. Assume that d is selected. Its new value $d_{new} = d + rand(-0.2, 0.2)[x_{i,max} - x_{i,min}]$, where $rand(\cdot)$ returns a random number from the assumed interval and $[x_{i,min}, x_{i,max}]$ is a range of the domain of the selected set. d_{new} must fulfil condition (18). New values of σ and ρ are then calculated using formula (17).

5. Application to financial benchmark data sets and comparative analysis

There are three publicly available standard financial benchmark data sets (see, e.g., discussion in recently published review paper [60]), i.e., *Statlog* (*Australian Credit Approval*), *Statlog* (*German Credit Approval*), and *Credit Approval* sets (the last one will be called *Credit Approval* (*Japanese*) set – see remarks in the Introduction). These sets – available from the UCI repository of machine learning databases (http://archive.ics.uci.edu/ml) – have often been processed by various classification techniques. The application results of 24 such techniques are collected (as already mentioned in the Introduction) in five papers published in recent years. They include three papers published in this journal ([83] of 2014 by C.-F. Tsai et al., [10] of 2012 by S.-Y. Chang and T.-Y. Yeh, and [69] of 2011 by Y. Peng et al.) and two other papers ([99] of 2013 by X. Zhu et al. and [94] of 2008 by L. Yu et al.). The aforementioned collection of results creates an excellent platform for comparison with new classification methods such as our approach and thus three earlier-mentioned standard financial benchmark data sets will also be employed in our experiments.

5.1. Experimental design

In order to fully understand the value and performance of our approach and to increase the reliability of the conclusions drawn, our main evaluation experiments will consist in using the k-fold cross-validation method with various values of k. Such a method minimizes bias associated with a random split of a given data set into the learning and test subsets. The whole data set is randomly divided into k mutually exclusive subsets with approximately equal number of records and with similar class distribution. The learning process is repeated k times; each time one of the k subset is used as the test set and the remaining as the learning set. The learning process implements our ideas in the framework of one of the presently most advanced MOEOAs, i.e., NSGA-II [15]. In each learning experiment, the best Pareto-front approximation is obtained. Pareto front is the set of Pareto-optimal (i.e., non-dominated by any other) solutions; they are characterized by various levels of accuracy-interpretability trade-off. Having obtained a collection of solutions in the form of the best Pareto-front approximation, a single solution (FRBC) characterized, first, by the highest accuracy $Q_{ACC}^{(tst)}$ in the test data set (defined analogously as $Q_{ACC}^{(lrn)}$ (10) for the learning data set) and, second, by the highest interpretability Q_{INT} (12) is selected from the front. Then, the average results from k partial experiments are computed. We refer to such an experiment as a single k-fold cross-validation (single k-fcv). The single k-fcv experiment is repeated 10 times for different divisions of the original data set into k subsets. Moreover, in order to minimize bias associated with an initialization of our approach, each of the aforementioned k-fcv experiments is repeated 10 times for randomly selected initial values of parameters in our approach. Therefore, altogether 100 single k-fcv experiments are carried out for a given k. Additionally, in order to increase the performance evaluation reliability and for the purpose of comparative analysis with alternative approaches reported in the literature, four values of k are considered: $k=10$ (i.e., learn-to-test ratio 9:1), $k=5$ (ratio 4:1), $k=3$ (ratio 2:1), and $k=2$ (ratio 1:1). For the *Credit Approval* (*Japanese*) data set also a non-typical 2.6:1 ratio is considered – the same as in several alternative approaches presented in [94].

Before the presentation of the afore outlined broad cross-validation experiment, in order to demonstrate the operation of our approach in some detail, the experiment for a single learning-test data split (with 9:1 ratio) of a given data set is presented. It includes the presentation of the best Pareto-front approximation and the details of the selected, best (first, in terms of the test-data accuracy and then in terms of the interpretability) solution (FRBC), i.e., its fuzzy rule base and fuzzy sets (linguistic terms) that describe particular attributes occurring in fuzzy rules.

In all experiments, the genetic learning process is carried out through 1000 generations, the initial population contains 10,000 individuals, the tournament selection with selection pressure equal to 2 is used, and the crossover and mutation probabilities are equal to 0.7 and 0.5, respectively.

The FRBC's performance is evaluated in terms of its accuracy and interpretability. The basic measure of the accuracy is $Q_{ACC}^{(lrn)}$ (10) for the learning data and analogously defined $Q_{ACC}^{(tst)}$ for the test data. The FRBC's accuracy can also be expressed as percentage of correct decisions: $ACC^{(lrn)} = Q_{ACC}^{(lrn)} \cdot p100\%$ and $ACC^{(tst)} = Q_{ACC}^{(tst)} \cdot p100\%$. Some additional accuracy measures regarding both the learning and test data can be used in two-class classification tasks [34], including true positive rate (TPR) or sensitivity and true negative rate (TNR) or specificity that are defined as follows:

$$TPR = \frac{TP}{TP+FN}, \quad TNR = \frac{TN}{TN+FP}, \tag{20}$$

where TP (true positive) is the number of correctly classified positive (or abnormal) cases, TN (true negative) is the number of correctly classified negative (or normal) cases, and FP (false positive) and FN (false negative) are the numbers of misclassified positive and negative cases, respectively. Obviously, the aforementioned overall accuracy $Q_{ACC}=(TP+TN)/(TP+FN+TN+FP)$.

The basic measure of the FRBC's interpretability is Q_{INT} (12). We also use several other interpretability indices such as the number of rules (R) in the rule base, the number of attributes (n_{ATR}) used by the classifier, the number of fuzzy sets (linguistic terms) describing the classifier's attributes (n_{FS}), and the number of attributes per rule ($n_{ATR/R}$).

Our comparative analysis is based on the result of the following alternative approaches applied to the classification of the benchmark financial data sets (the results are published in the enclosed references): linear discriminant analysis (LDA) [99], quadratic discriminant analysis (QDA) [99], logistic regression [69], (LogR [99,94]), Bayesian network [69], naive Bayes classifier [10,69], k-nearest neighbor (k-NN) method [99,69], artificial neural network (ANN) [94], backpropagation neural network (BPN) [10], radial basis function (RBF) network [69], multilayer perceptron (MLP) [83], support vector machine (SVM) [83,99,94,10,69], least square SVM (LSSVM) [99], artificial immune network of Timmis et al. (model of Timmis et al.) [10], artificial-immune-network-based (AINE-based) classifier [10], simple artificial immune system (SAIS) [10], the technique for order preference by similarity to ideal solution for classification tasks (C-TOPSIS) [99,83], decision tree (DT) methods [99] including C4.5 algorithm [10,69], the rule induction by the repeated incremental pruning to produce error reduction (RIPPER rule induction) [69], the ensemble approaches: SVM-, MLP-, and DT-based ensembles [83], voting-based and reliability-based neural network ensembles [94,69] and, finally, the hybrid approaches: SVM-genetic algorithm (SVM-GA), SVM-Grid, and SVM-Grid-F-score hybrids [10] as well as neuro-fuzzy and fuzzy SVM hybrids [94]. Concluding, 24 alternative techniques arranged in 28 experimental set-ups related to various learn-to-test ratios are considered.

Table 1
Interpretability and accuracy measures of solutions from Fig. 3 (*Statlog* (*Australian Credit Approval*) data set).

No.	Objective functions complements		Interpretability measures				Accuracy measures	
	$1 - Q_{INT} = Q_{CPLX}$	$1 - Q_{ACC}^{(lrn)}$	R	n_{ATR}	n_{FS}	$n_{ATR/R}$	$ACC^{(lrn)}$	$ACC^{(tst)}$
1.	0	0.1449	2	1	1	1	85.5%	85.5%
2.	0.0307	0.1417	2	2	2	1	85.8%	85.5%
3.	0.0666	0.1384	4	3	4	1	86.1%	86.9%
4.	0.0730	0.1368	2	2	2	1	85.8%	85.5%
5.	0.0769	0.1336	5	3	5	1.2	86.6%	88.4%
6.	0.0820	0.1320	5	3	6	1.2	86.7%	89.8%
7.	0.1025	0.1304	5	4	4	1.4	86.9%	86.9%
8.	0.1247	0.1288	6	4	7	1.6	87.1%	89.8%
9.	0.1333	0.1239	5	5	5	1.4	87.6%	86.9%
10.	0.1435	0.1207	5	5	6	1.6	87.9%	86.9%
11.	0.1516	0.1191	7	5	7	1.7	88.0%	89.8%
12.	0.1604	0.1175	7	5	8	1.8	88.2%	89.8%
13.	0.1717	0.1159	6	6	7	1.5	88.4%	89.8%
14.	0.1760	0.1143	6	6	7	1.6	88.5%	89.8%
15.	0.1812	0.1127	6	6	8	1.6	88.7%	89.8%
16.	0.1824	0.1111	7	6	8	1.7	88.8%	89.8%
17.	0.1912	0.1078	7	6	9	1.8	89.2%	89.8%
18.	0.2000	0.1062	8	6	10	2	89.3%	89.8%
19.	0.2079	0.1046	9	6	11	2.1	89.5%	89.8%
20.	0.2692	0.1030	10	8	13	2.1	89.6%	89.8%
21.	0.2769	0.1014	10	8	14	2.2	89.8%	89.8%
22.	0.3025	0.0998	10	9	14	2.2	90.0%	89.8%
23.	0.3102	0.0982	10	9	15	2.3	90.1%	89.8%

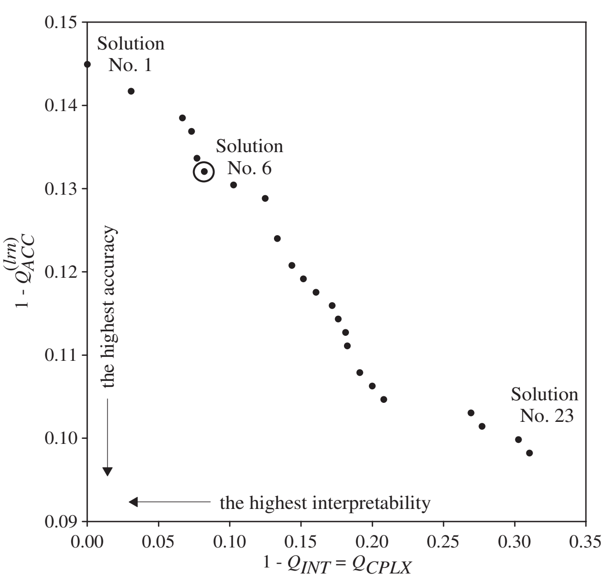

Fig. 3. The best Pareto-front approximation (*Statlog* (*Australian Credit Approval*) data set).

Table 2
Fuzzy rule base for the solution No. 6 from Table 1 (*Statlog* (*Australian Credit Approval*) data set).

No.	Fuzzy classification rules	
1.	**IF**	x_8 *is* 0 **THEN** Class "−"
2.	**IF**	x_{12} *is not* 2 **AND** x_{13} *is not* Small **THEN** Class "−"
3.	**IF**	x_{13} *is* Medium2 **THEN** Class "−"
4.	**IF**	x_{12} *is* 3 **THEN** Class "+"
5.	**IF**	x_{13} *is* Medium3 **THEN** Class "+"

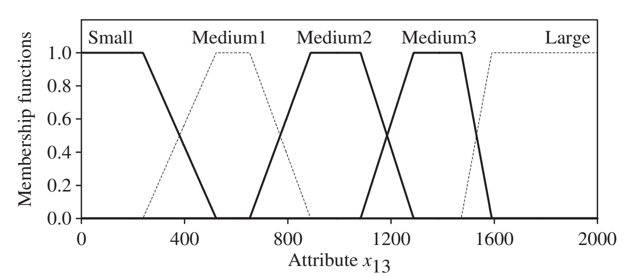

Fig. 4. Final shapes of membership functions for strong fuzzy partition (SFP) of numerical attribute x_{13} used in the rule base of Table 2 (*Statlog* (*Australian Credit Approval*) data set).

5.2. Statlog (Australian Credit Approval) data set

This data set concerns credit card applications and contains 690 instances (applicant data records). Each data record is characterized by 6 numerical and 8 categorical attributes; their names and values have been changed to meaningless symbols to protect the data confidentiality (they are referred to as $x_1, x_2, \ldots, x_{14}$). Particular data records belong to one of two classes: application rejected (class label "−") or application approved (class label "+"). In general, the "−" cases occupy 55.5% of the whole set, whereas the remaining 45.5% represent the "+" cases.

First, the genetic learning experiment for a single learning-test data split (with 9:1 ratio) of the considered data set is presented in order to reveal some details of our approach. Fig. 3 presents a set of 23 solutions (FRBCs) obtained in the final generation of a single run of our method. They form the best approximation of Pareto-optimal solutions characterized by various levels of accuracy-interpretability trade-off. The values of the interpretability and accuracy measures of solutions from Fig. 3 are presented in Table 1. Analysing particular solutions from Table 1, one can find that the solution No. 6 is, first, the most accurate for the test data (89.8%) and, second, the most transparent (i.e., the simplest in terms of interpretability). Table 2 presents its rule base; it contains 5 fuzzy rules with only 3 attributes, 3 fuzzy sets (linguistic terms) for 1 numerical attribute (x_{13}) and 3 fuzzy singletons (see (2)) for the remaining 2 categorical attributes (x_8, x_{12}), and, on average, 1.2 attribute per rule. The membership functions of fuzzy sets for numerical attribute x_{13} that occurs in the rules are additionally shown in Fig. 4. Some details on the labels of categorical attributes, i.e., 0 for x_8, 3 and 2 for x_{12} can be found in the UCI repository (http://archive.ics.uci.edu/ml).

Table 3
Results of our approach and comparison with alternative approaches for *Statlog* (*Australian Credit Approval*) data set.

Source	Method	Learn-to-test ratio	Number of single *k*-fcv experiments	Average accuracy measures for test data			Average interpretability measures			
				$\overline{ACC}$ [%]	$\overline{TPR}$ [%]	$\overline{TNR}$ [%]	$\overline{R}$	$\overline{n}_{ATR}$	$\overline{n}_{FS}$	$\overline{n}_{ATR/R}$
[69] (Y. Peng et al., 2011)	Bayesian network	9:1	1	85.22	79.80	89.56	–	–	–	–
	Naïve Bayes	9:1	1	77.25	58.63	92.17	–	–	–	–
	SVM	9:1	1	85.51	92.51	79.90	–	–	–	–
	Logistic regression	9:1	1	86.23	86.64	85.90	–	–	–	–
	k-NN	9:1	1	79.42	77.52	80.94	–	–	–	–
	C4.5	9:1	1	83.48	79.48	86.68	n/a	n/a	n/a	n/a
	RBF network	9:1	1	83.04	75.24	89.30	–	–	–	–
	RIPPER rule induction	9:1	1	85.22	85.34	85.12	n/a	n/a	n/a	n/a
	Ensemble	9:1	1	85.51	82.74	87.73	–	–	–	–
[10] (S.-Y. Chang, T.-Y. Yeh, 2012)	AINE-based	9:1	1	85.36	n/a	n/a	–	–	–	–
	BPN	9:1	1	86.83	n/a	n/a	–	–	–	–
	C4.5	9:1	1	82.50	n/a	n/a	n/a	n/a	n/a	n/a
	Naïve Bayes	9:1	1	84.90	n/a	n/a	–	–	–	–
	SAIS	9:1	1	85.20	n/a	n/a	–	–	–	–
	SVM	9:1	1	84.70	n/a	n/a	–	–	–	–
	Model of Timmis et al.	9:1	1	85.20	n/a	n/a	–	–	–	–
	SVM-GA hybrid	9:1	1	86.90	n/a	n/a	–	–	–	–
	SVM-Grid hybrid	9:1	1	85.51	n/a	n/a	–	–	–	–
	SVM-Grid-F-score hybrid	9:1	1	84.20	n/a	n/a	–	–	–	–
[99] (X. Zhu et al., 2013)	LDA	4:1	100	85.79	80.64	92.28	–	–	–	–
	QDA	4:1	100	80.02	91.37	65.70	–	–	–	–
	DT	4:1	100	83.18	85.39	80.39	n/a	n/a	n/a	n/a
	LogR	4:1	100	86.25	85.73	86.91	–	–	–	–
	k-NN	4:1	100	69.31	79.84	56.01	–	–	–	–
	SVM	4:1	100	86.00	83.99	88.54	–	–	–	–
	LSSVM	4:1	100	86.78	85.01	89.03	–	–	–	–
	C-TOPSIS	4:1	100	86.52	88.00	84.68	n/a	n/a	n/a	n/a
[83] (C.-F. Tsai et al.,	SVM	9:1	1	85.63	n/a	n/a	–	–	–	–
	SVM ensembles	9:1	1	85.05	n/a	n/a	–	–	–	–
	MLP	9:1	1	82.44	n/a	n/a	–	–	–	–
	MLP ensembles	9:1	1	84.62	n/a	n/a	–	–	–	–
	DT	9:1	1	84.91	n/a	n/a	n/a	n/a	n/a	n/a
	DT ensembles	9:1	1	87.23	n/a	n/a	n/a	n/a	n/a	n/a
This paper	Our approach	1:1	100	86.8	85.1	88.9	5.2	4.8	6.2	1.4
		2:1	100	87.3	84.6	90.6	5.5	4.9	6.7	1.4
		4:1	100	88.0	86.4	90.0	5.7	5.3	7.1	1.5
		9:1	100	89.1	87.1	91.5	5.2	4.7	5.9	1.4

n/a stands for not available; – stands for not applicable, e.g., the number of rules for non-rule-based systems.

The results of our main evaluation experiment based on multiple cross-validation and the results of 24 alternative approaches are collected in Table 3 for the purpose of comparative analysis. The reference papers [10,69,83,94,99] not always give a clear information on how many times the basic cross-validation experiment (we call it the single *k*-fcv – see Section 5.1 for comments) was repeated. For this reason, in non-clear cases we assume that the number of single *k*-fcv experiments is equal to 1. We'd like to emphasize that in our investigations each single *k*-fcv experiment is repeated 100 times (see Section 5.1 for comments) and the averaged results are presented. The results of Table 3 show that our approach outperforms all the alternative 24 methods in terms of both the accuracy and interpretability and for any learn-to-test ratio. For the 9:1 ratio, our approach gives 89.1% average accuracy for the test data using fuzzy rule base that contains on average 5.2 rules with 4.7 attributes, 5.9 fuzzy sets (linguistic terms), and 1.4 attribute per rule. The best alternative approach for the 9:1 learn-to-test ratio is DT ensemble [83], which achieves only 87.23% accuracy for the test data but no information on its interpretability (e.g., the number of nodes, leaves, selected attributes, etc. of DTs) is available in [83] and thus its interpretability evaluation and comparison is not possible. In general, the overwhelming majority of alternative approaches are the black-box-type techniques. As far as some interpretability indices are concerned, only [99] introduces a qualitative and subjective binary "interpretability score". According to that score, DT and C-TOPSIS [99] are graded 1, whereas LDA, QDA, LogR, *k*-NN, SVM, and LSSVM (i.e., the black-box-methods) are graded 0. Therefore, the "interpretability score" is not an interpretability measure but a distinction index between non-black-box and black-box methods. On the contrary, our approach provides easy-to-comprehend measures of the level of the classifier's interpretability. Moreover, it is worth emphasizing that our approach is also an effective tool for the attribute (feature) selection and, even more, for finding softly defined (by means of fuzzy sets) areas of attribute values that are essential in financial decision making.

For the 4:1 learn-to-test ratio, our approach with 88.0% average accuracy for the test data and the interpretability measures as shown in Table 3 also outperforms the best alternative approach, i.e., the black-box LSSVM method with 86.78% accuracy. For possible further comparisons, we also present the results of our approach for 2:1 and 1:1 learn-to-test ratios.

Another aspect of accuracy analysis is related to the fact that misclassification of a bad applicant usually represents significantly higher costs for the credit granting institution than the opposite case [81]. The additional accuracy measures, i.e., *TPR* and *TNR* (20) address this problem. Our approach outperforms the majority of the remaining methods also in this aspect. For 9 : 1 learn-to-test ratio only SVM gives higher average *TPR* ($\overline{TPR} = 92.51\%$) but much worse average *TNR* ($\overline{TNR} = 79.90\%$); in the case of our approach,

Table 4
Interpretability and accuracy measures of selected solutions from Fig. 5 (*Statlog* (*German Credit Approval*) data set).

No.	Objective functions complements		Interpretability measures				Accuracy measures	
	$1-Q_{INT}=Q_{CPLX}$	$1-Q_{ACC}^{(lrn)}$	R	n_{ATR}	n_{FS}	$n_{ATR/R}$	$ACC^{(lrn)}$	$ACC^{(tst)}$
1.	0	0.31	2	1	1	1	69.0%	79%
2.	0.0037	0.3088	2	1	2	1	69.1%	74%
3.	0.0213	0.2766	2	2	2	1	72.3%	76%
4.	0.0426	0.2666	3	3	3	1	73.3%	77%
5.	0.0464	0.2633	4	3	4	1	73.6%	77%
6.	0.0508	0.2622	4	3	4	1.2	73.7%	77%
7.	0.0639	0.2577	4	4	4	1	74.2%	78%
8.	0.0677	0.2544	5	4	5	1	74.5%	78%
9.	0.0712	0.2533	5	4	5	1.2	74.6%	78%
10.	0.0721	0.2522	4	4	5	1.2	74.7%	80%
11.	0.0774	0.25	6	4	6	1.3	75.0%	69%
12.	0.0812	0.2477	6	4	7	1.3	75.2%	68%
13.	0.0841	0.2466	6	4	7	1.5	75.3%	69%
14.	0.0879	0.2433	6	4	8	1.5	75.6%	72%
15.	0.0987	0.24	6	5	7	1.3	76.0%	67%
16.	0.1016	0.2333	6	5	7	1.5	76.6%	70%
17.	0.1042	0.2322	7	5	8	1.4	76.7%	69%
...	...	...	...	...	...	...	...	...
24.	0.1267	0.2211	6	6	9	1.5	77.8%	73%
...	...	...	...	...	...	...	...	...
31.	0.1531	0.2088	7	7	11	1.5	79.1%	70%
...	...	...	...	...	...	...	...	...
38.	0.2677	0.1955	8	11	19	2.3	80.4%	67%

Table 5
Fuzzy rule base for the solution No. 10 from Table 4 (*Statlog* (*German Credit Approval*) data set).

No.	Fuzzy classification rules	
1.	**IF**	status of existing account *is* A14 **THEN** Class 1
2.	**IF**	credit history *is* A34 **THEN** Class 1
3.	**IF**	credit amount *is* Large **THEN** Class 2
4.	**IF**	status of existing account *is* A11 **AND** installment rate (% of income) *is* Large **THEN** Class 2

$\overline{TPR}=87.1\%$ and $\overline{TPR}=91.5\%$. Similarly, for 4 : 1 learn-to-test ratio only QDA and C-TOPSIS give better $\overline{TPR}$s but much worse $\overline{TNR}$s comparing to our approach.

5.3. Statlog (German Credit Approval) *data set*

The problem of the classification of credit applications into one of two groups labelled as "good credit risk" (Class 1) and "bad credit risk" (Class 2) is addressed in the data set considered in this subsection. The data set contains 1000 instances (700 representing Class 1 and 300 – Class 2) characterized by 20 attributes: (1) status of existing account, (2) credit duration, (3) credit history), (4) credit purpose, (5) credit amount, (6) savings, (7) present employment, (8) installment rate (% of income), (9) personal status and sex, (10) other debtors/guarantors, (11) present residence, (12) property, (13) age, (14) other installment plans, (15) housing, (16) number of credits, (17) job, (18) number of people liable to provide maintenance for, (19) telephone, and (20) foreign worker.

Fig. 5 shows the best Pareto-front approximation with 38 solutions (FRBCs) representing various levels of accuracy-interpretability trade-off and obtained from a single run of our approach for a data split with the 9:1 learn-to-test ratio. Table 4 presents the details of selected solutions. The solution No. 10 is, first, the most accurate for the test data (with $ACC^{(tst)}=80.0\%$) and, second, the most interpretable one. Table 5 presents its fuzzy rule base with only 4 rules using 4 attributes, 2 fuzzy sets (linguistic terms) for 2 numerical attributes and 3 fuzzy singletons (see (2)) for

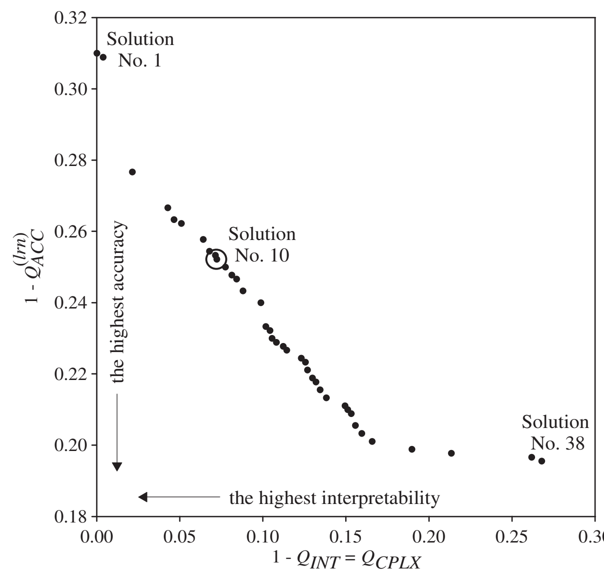

Fig. 5. The best Pareto-front approximation (*Statlog* (*German Credit Approval*) data set).

the remaining 2 categorical attributes, and on average 1.2 attribute per rule. Fig. 6 shows the membership functions of the fuzzy sets for numerical attributes used in the rules. Information on labels A14, A34, and A11 of categorical attributes in the rules can be found in the UCI repository (http://archive.ics.uci.edu/ml).

Table 6 collects the results of our main multiple-cross-validation-based experiment and, for comparison, also the results of 24 alternative classification techniques. In terms of the average accuracy for the test data: (1) for the 9:1 learn-to-test ratio, our approach (78.5%) gives better results than the best alternative method, i.e., SVM-GA hybrid (77.9%), and (2) for the 4 : 1 ratio, our approach (76.5%) gives comparable (worse by 0.12%) results in regard to the best alternative LogR method (76.62%). However, due to the black-box-nature of both SVM-GA hybrid and LogR methods, our approach significantly outperforms them in terms of the interpretability. As far as the average additional accuracy measures $\overline{TPR}$

Table 6
Results of our approach and comparison with alternative approaches for *Statlog* (*German Credit Approval*) data set.

Source	Method	Learn-to-test ratio	Number of single k-fcv experiments	Average accuracy measures for test data			Average interpretability measures			
				$\overline{ACC}$ [%]	$\overline{TPR}$ [%]	$\overline{TNR}$ [%]	$\overline{R}$	$\overline{n}_{ATR}$	$\overline{n}_{FS}$	$\overline{n}_{ATR/R}$
[69] (Y. Peng et al., 2011)	Bayesian network	9:1	1	72.50	36.00	88.14	–	–	–	–
	Naïve Bayes	9:1	1	75.50	50.67	86.14	–	–	–	–
	SVM	9:1	1	77.40	49.33	89.43	–	–	–	–
	Logistic regression	9:1	1	77.10	49.33	89.00	–	–	–	–
	k-NN	9:1	1	66.90	45.00	76.29	–	–	–	–
	C4.5	9:1	1	71.90	44.00	83.86	n/a	n/a	n/a	n/a
	RBF network	9:1	1	74.00	46.33	85.86	–	–	–	–
	RIPPER rule induction	9:1	1	73.40	45.00	85.57	n/a	n/a	n/a	n/a
	Ensemble	9:1	1	76.20	45.33	89.43	–	–	–	–
[10] (S.-Y. Chang, T.-Y. Yeh, 2012)	AINE-based	9:1	1	77.10	n/a	n/a	–	–	–	–
	BPN	9:1	1	77.80	n/a	n/a	–	–	–	–
	C4.5	9:1	1	72.40	n/a	n/a	n/a	n/a	n/a	n/a
	Naïve Bayes	9:1	1	74.70	n/a	n/a	–	–	–	–
	SAIS	9:1	1	75.40	n/a	n/a	–	–	–	–
	SVM	9:1	1	76.00	n/a	n/a	–	–	–	–
	Model of Timmis et al.	9:1	1	72.40	n/a	n/a	–	–	–	–
	SVM-GA hybrid	9:1	1	77.90	n/a	n/a	–	–	–	–
	SVM-Grid hybrid	9:1	1	76.00	n/a	n/a	–	–	–	–
	SVM-Grid-F-score hybrid	9:1	1	77.50	n/a	n/a	–	–	–	–
[99] (X. Zhu et al., 2013)	LDA	4:1	100	72.08	71.38	72.37	–	–	–	–
	QDA	4:1	100	67.64	69.57	66.82	–	–	–	–
	DT	4:1	100	69.85	49.56	78.54	n/a	n/a	n/a	n/a
	LogR	4:1	100	76.62	49.34	88.31	–	–	–	–
	k-NN	4:1	100	70.70	19.17	92.78	–	–	–	–
	SVM	4:1	100	71.78	64.26	75.00	–	–	–	–
	LSSVM	4:1	100	71.99	61.78	76.36	–	–	–	–
	C-TOPSIS	4:1	100	75.47	58.94	82.55	n/a	n/a	n/a	n/a
[83] (C.-F. Tsai et al., 2014)	SVM	9:1	1	75.68	n/a	n/a	–	–	–	–
	SVM ensembles	9:1	1	75.78	n/a	n/a	–	–	–	–
	MLP	9:1	1	70.57	n/a	n/a	–	–	–	–
	MLP ensembles	9:1	1	76.48	n/a	n/a	–	–	–	–
	DT	9:1	1	73.77	n/a	n/a	n/a	n/a	n/a	n/a
	DT ensembles	9:1	1	75.98	n/a	n/a	n/a	n/a	n/a	n/a
This paper	Our approach	1:1	100	74.4	38.0	90.1	5.7	5.2	8.2	1.7
		2:1	100	75.4	38.6	91.2	6.3	6.0	10.2	2.0
		4:1	100	76.5	42.3	91.2	5.9	5.3	8.7	1.8
		9:1	100	78.5	45.8	92.7	5.9	5.4	8.7	1.8

n/a stands for not available; – stands for not applicable, e.g., the number of rules for non-rule-based systems.

and $\overline{TNR}$ are concerned, our approach – in comparison with alternative techniques – gives similar results in terms of $\overline{TPR}$ but much better – in terms of $\overline{TNR}$. Similarly as in Section 5.2, the results of our approach for 2:1 and 1:1 ratios are included in Table 6 for possible further comparisons.

5.4. Credit Approval *(Japanese) data set*

The third data set – concerning the classification of credit card applications into either group labelled as “+” or “–” – contains 690 instances (307 from class “+” and 383 from class “–”) characterized by 6 numerical and 9 categorical attributes. Similarly as in the case of *Statlog* (*Australian Credit Approval*) data set, their names and values have been changed to meaningless symbols to protect the data confidentiality (we call them $x_1, x_2, \ldots, x_{15}$).

First, we present the results of a single run of our approach for the exemplary data split with the 9:1 learn-to-test ratio. Fig. 7 and Table 7 show the best Pareto-front approximation with 19 solutions (FRBCs) representing various levels of accuracy-interpretability trade-off. The most accurate for the test data and then the most interpretable is the solution No. 4 from Table 7. Its fuzzy rule base – presented in Table 8 – contains 5 rules using only 3 attributes (1 – numerical and 2 – categorical), only 1 fuzzy set (shown in Fig. 8) for the selected numerical attribute, 3 fuzzy singletons (see (2)) for the remaining 2 categorical attributes, and, on average, 1.2 attribute per rule. Some comments on the “t”, “s”, “p” labels of categorical attributes occurring in the rules can be found in the UCI repository (http://archive.ics.uci.edu/ml).

Table 9 presents the results of our main experiment based on multiple cross-validation and also the results of 13 alternative techniques. In terms of the accuracy for the test data, our approach gives slightly better results for the 9:1 learn-to-test ratio (89.0%) than the best alternative technique, i.e., DT ensemble (88.36%). The best alternative method for the 2.6:1 learn-to-test ratio considered in [94], i.e., reliability-based neural network ensemble gives 88.08% of the accuracy for the test data. Our approach, for the same ratio, achieves comparable results. However, in terms of the interpretability, our approach significantly outperforms the black-box-type reliability-based neural network ensemble, whereas DT ensemble's evaluation and comparison is not possible since no information on DTs' complexity is given in [83] (see comments in Section 5.2). Moreover, our approach gives the best results in terms of the average accuracy measure $\overline{TPR}$ and, except for one alternative method, also in terms of $\overline{TNR}$. The results of our approach for 4:1, 2:1, and 1:1 learn-to-test ratios are also included in Table 9 for possible further comparisons.

Table 7
Interpretability and accuracy measures of solutions from Fig. 7 (*Credit Approval* (*Japanese*) data set).

No.	Objective functions complements		Interpretability measures				Accuracy measures	
	$1-Q_{INT}=Q_{CPLX}$	$1-Q_{ACC}^{(lrn)}$	R	n_{ATR}	n_{FS}	$n_{ATR/R}$	$ACC^{(lrn)}$	$ACC^{(tst)}$
1.	0	0.1465	2	1	1	1	85.3%	86.9%
2.	0.0283	0.1401	2	2	2	1	85.9%	86.9%
3.	0.0566	0.1320	3	3	3	1	86.7%	88.4%
4.	0.0658	0.1304	5	3	4	1.2	86.9%	89.8%
5.	0.0849	0.1223	4	4	4	1	87.7%	82.6%
6.	0.0934	0.1191	6	4	5	1.1	88.0%	81.1%
7.	0.1177	0.1175	6	5	6	1	88.2%	88.4%
8.	0.1217	0.1143	6	5	6	1.1	88.5%	86.9%
9.	0.1256	0.1111	7	5	7	1.1	88.8%	86.9%
10.	0.1301	0.1095	7	5	8	1.1	89.0%	86.9%
11.	0.1500	0.1078	6	6	7	1.1	89.2%	89.8%
12.	0.1539	0.1062	6	4	5	1.1	88.0%	81.1%
13.	0.1573	0.1046	7	6	8	1.2	89.5%	88.4%
14.	0.1618	0.1030	7	6	9	1.2	89.6%	86.9%
15.	0.1701	0.1014	9	6	10	1.4	89.8%	88.4%
16.	0.1936	0.0998	7	7	10	1.4	90.0%	85.5%
17.	0.2011	0.0982	9	7	11	1.5	90.1%	85.5%
18.	0.2082	0.0966	9	7	12	1.6	90.3%	84.0%
19.	0.2392	0.0950	9	8	13	1.7	90.4%	84.0%

Table 8
Fuzzy rule base for the solution No. 4 from Table 7 (*Credit Approval* (*Japanese*) data set).

No.	Fuzzy classification rules	
1.	**IF**	x_9 *is not* t **THEN** Class "−"
2.	**IF**	x_{13} *is* s **THEN** Class "−"
3.	**IF**	x_9 *is not* t **AND** x_{13} *is not* p **THEN** Class "−"
4.	**IF**	x_{13} *is* p **THEN** Class "+"
5.	**IF**	x_{14} *is* Small **THEN** Class "+"

5.5. Some remarks on computational efficiency of our approach

Apart from the accuracy and interpretability of computerized financial decision making systems also their computational efficiency, i.e., the speed of automatic decision making is an important aspect of their operation. From that point of view, we must clearly distinguish the stage of the construction of our approach from the stage of its use. The construction, i.e., the genetic learning and optimization process must be performed only once for a given financial data set. As a result of that, a collection of FRBCs characterized by various levels of accuracy-interpretability trade-off (the best approximation of the Pareto-front solutions) is obtained. Each of those FRBCs can be used as the decision support system. In our investigations, FRBC characterized, first, by the highest accuracy for the test data and, second, by the highest interpretability was selected in each application area. Once a specific FRBC is chosen, a new decision is obtained by calculating formulas (5)–(8). Table 10 presents the average, standard deviation, minimum, and maximum of computing time in seconds of FRBCs presented in Tables 2, 5, and 8 for the processing of a single data record from *Statlog* (*Australian Credit Approval*), *Statlog* (*German Credit Approval*), and *Credit Approval* (*Japanese*) data sets, respectively. 100 runs have been performed for each data record from a given data set using

Table 9
Results of our approach and comparison with alternative approaches for *Credit Approval* (*Japanese*) data set.

Source	Method	Learn-to-test ratio	Number of single k-fcv experiments	Average accuracy measures for test data			Average interpretability measures			
				$\overline{ACC}$ [%]	$\overline{TPR}$ [%]	$\overline{TNR}$ [%]	$\overline{R}$	$\overline{n}_{ATR}$	$\overline{n}_{FS}$	$\overline{n}_{ATR/R}$
[94] (L. Yu et al., 2008)	LogR	2.6:1	10	75.82	76.36	74.58	–	–	–	–
	ANN	2.6:1	10	80.77	82.26	80.08	–	–	–	–
	SVM	2.6:1	10	79.91	81.43	78.41	–	–	–	–
	Neuro-fuzzy hybrid	3:1	10	77.91	n/a	n/a	n/a	n/a	n/a	n/a
	Fuzzy SVM hybrid	1.6:1	10	83.94	85.43	82.70	–	–	–	–
	Voting-based ens.	2.6:1	10	85.22	86.58	84.37	–	–	–	–
	Reliability-based ens.	2.6:1	10	88.08	87.44	88.86	–	–	–	–
[83] (C.-F. Tsai et al., 2014)	SVM	9:1	1	86.37	n/a	n/a	–	–	–	–
	SVM ensembles	9:1	1	86.37	n/a	n/a	–	–	–	–
	MLP	9:1	1	84.38	n/a	n/a	–	–	–	–
	MLP ensembles	9:1	1	87.60	n/a	n/a	–	–	–	–
	DT	9:1	1	86.37	n/a	n/a	n/a	n/a	n/a	n/a
	DT ensembles	9:1	1	88.36	n/a	n/a	n/a	n/a	n/a	n/a
This paper	Our approach	1:1	100	86.7	90.0	84.0	5.8	4.6	6.2	1.5
		2:1	100	87.2	89.9	85.1	6.2	5.2	7.2	1.5
		2.6:1	100	87.6	90.1	85.5	6.0	5.1	7.1	1.6
		4:1	100	88.2	90.5	86.4	6.2	5.5	7.4	1.5
		9:1	100	89.0	91.4	87.0	5.3	4.5	5.7	1.3

n/a stands for not available; – stands for not applicable, e.g., the number of rules for non-rule-based systems.

Table 10
Computing time results of our approach.

Financial data sets	Time (s)			
	Average	Standard deviation	Minimum	Maximum
Statlog (*Australian Credit Approval*)	2.11×10^{-6}	7.02×10^{-7}	1.54×10^{-6}	7.69×10^{-6}
Statlog (*German Credit Approval*)	5.97×10^{-6}	8.03×10^{-7}	5.13×10^{-6}	1.69×10^{-5}
Credit Approval (*Japanese*)	2.64×10^{-6}	8.78×10^{-7}	2.05×10^{-6}	9.75×10^{-6}

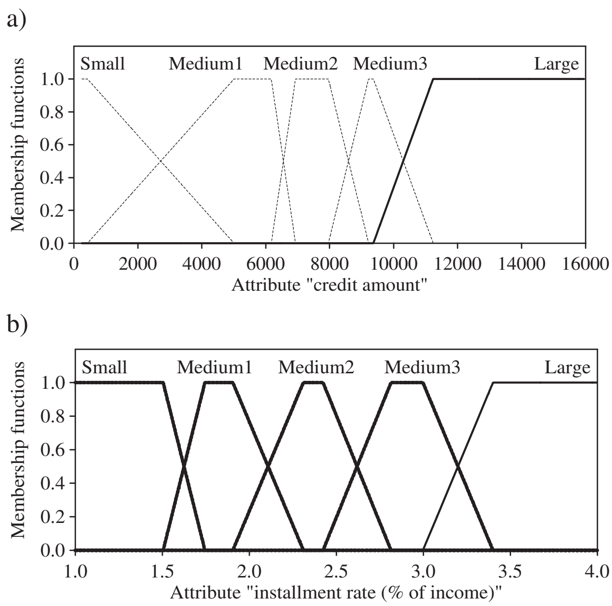

Fig. 6. Final shapes of membership functions for strong fuzzy partitions (SFPs) of numerical attributes: (a) "credit amount" and (b) "installment rate (% of income)" used in the rule base of Table 5 (*Statlog* (*German Credit Approval*) data set).

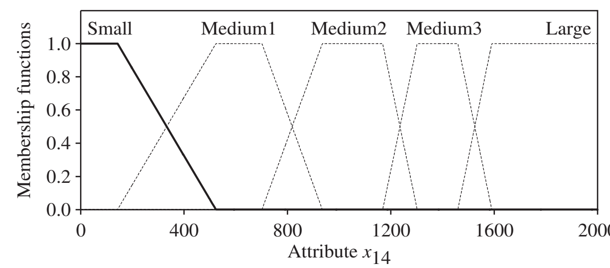

Fig. 8. Final shapes of membership functions for strong fuzzy partition (SFP) of numerical attribute x_{14} used in the rule base of Table 8 (*Credit Approval* (*Japanese*) data set).

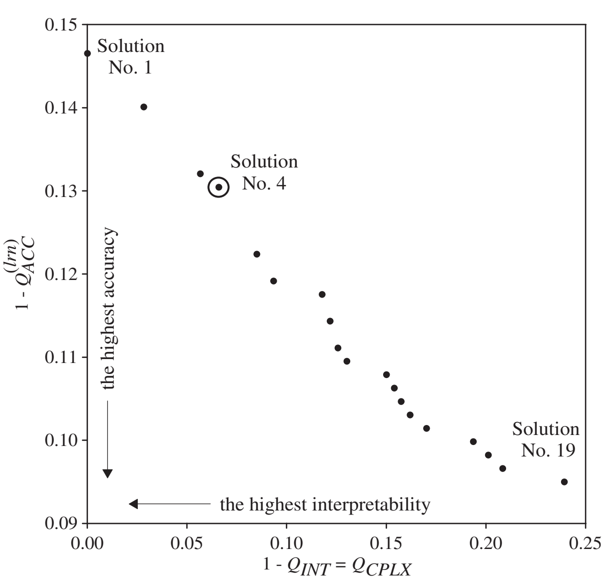

Fig. 7. The best Pareto-front approximation (*Credit Approval* (*Japanese*) data set).

a computer with Intel i7 X920 2GHz CPU and the results have been averaged. The results of Table 10 show that our approach is also highly competitive in terms of the speed of generating decision comparing with alternative methods (see, e.g., [99] for comparison).

6. Conclusions

Credit classification is an important component of critical financial decision making tasks such as credit scoring and bankruptcy prediction. The performance of various credit classification methods is usually evaluated in terms of their accuracy, interpretability, and the speed of generating decisions. The classifier's accuracy represents its ability to adequately represent the decision making process, whereas the interpretability – its ability to provide compact and understandable explanations and justifications of the proposed decision. Linguistic fuzzy classification rules – due to their easy-to-grasp interpretation by humans and high modularity – are particularly well-suited for the transparent representation of financial knowledge.

In this paper, we propose an approach for designing, in an automatic way, fuzzy rule-based classifiers from financial data. In order to address the aforementioned evaluation criteria and taking into account that the classifier's accuracy and interpretability are to some extent contradictory requirements, we employ a multi-objective evolutionary optimization algorithm in the process of the classifier's construction. Within a single experiment, our approach generates a collection of non-dominated solutions that belong to the best approximation of the front of Pareto-optimal solutions of the considered decision-making problem. Our solutions (financial fuzzy rule-based classifiers) are characterized by various levels of accuracy-interpretability trade-off including classifiers of significantly superior interpretability while still remaining either competitive or also superior in terms of the accuracy in comparison with the existing classification methods. Moreover, our approach is also highly competitive in terms of the decision-making speed. After the completion of the genetic learning and the selection of a single classifier – due to computational simplicity of the fuzzy approximate inference engine – new decisions can be obtained almost instantaneously.

The contribution of this paper includes the introduction of the approach that involves and optimizes both the accuracy and interpretability requirements already at the stage of designing the classifier from data. The overwhelming majority of the existing financial-data-classification methods concentrate almost exclusively on the accuracy issues. The important problem of formulating compact and understandable explanations and justifications of the proposed decisions is either not involved in the process of the classifiers' design or it is ignored at all.

Another aspect, as far as the contribution of this paper is concerned, is the introduction of several original components that are essential in designing fuzzy rule-based classifiers for various (including financial) applications by means of multi-objective evolutionary optimization. In particular, we introduce: (i) the new complexity-related interpretability measure, (ii) the efficient implementation of the conditions for the strong fuzzy partitions of the attribute domains in order to address the semantics-related interpretability, (iii) simple, special-coding-free, and thus computationally efficient representation of the classifier's rule base, (iv) original genetic operators for the processing of rule bases, and (v) the implementation of our ideas in the context of one of the presently most advanced multi-objective evolutionary optimization algorithms, i.e., the NSGA-II method.

This paper also contributes to the important field of the attribute (feature) selection in decision making. Our approach is an effective tool for the attribute selection and, even more, for discovering in the domains of numerical attributes softly defined (by means of fuzzy sets) areas of attribute values that are essential in decision making process. Similarly, in the domains of selected categorical attributes (represented by sets of labels), also the essential subsets of labels are discovered.

The multi-objective evolutionary optimization method is an important part of our approach. It directly affects the quality of the classifier's accuracy-interpretability balance. Therefore, our further research will concentrate on improving the optimization algorithm itself. Our generalization of the well-known Strength Pareto Evolutionary Algorithm 2 (SPEA2) [100] proposed in [72] is the first attempt to achieve that goal. The generalization consists in exchanging the environmental selection procedure in SPEA2 by a new original algorithm which aims to determine the final non-dominated solutions with a high spread and well-balanced distribution in the objective space. Such an improvement supports our generalized multi-objective evolutionary optimization method in more effective addressing the accuracy-interpretability balance problems.

References

[1] R. Aguilar-Rivera, M. Valenzuela-Rendón, J.J. Rodríguez-Ortiz, Genetic algorithms and Darwinian approaches in financial applications: a survey, Expert Syst. Appl. 42 (21) (2015) 7684–7697.

[2] B.S. Ahn, S.S. Cho, C.Y. Kim, The integrated methodology of rough set theory and artificial neural network for business failure prediction, Expert Syst. Appl. 18 (2) (2000) 65–74.

[3] H. Ahn, K.-J. Kim, Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach, Appl. Soft Comput. 9 (2) (2009) 599–607.

[4] E. Alfaro, N. Garcia, M. Gámez, D. Elizondo, Bankruptcy forecasting: an empirical comparison of AdaBoost and neural networks, Decis. Support Syst. 45 (1) (2008) 110–122.

[5] E.I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, J. Financ. 23 (4) (1968) 589–609.

[6] A.F. Atiya, Bankruptcy prediction for credit risk using neural networks: a survey and new results, IEEE Trans. Neural Netw. 12 (4) (2001) 929–935.

[7] M. Baczyński, B. Jayaram, Fuzzy Implications, Studies in Fuzziness and Soft Computing, Springer, 2008.

[8] S. Balcaen, H. Ooghe, 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems, Brit. Account. Rev. 38 (1) (2006) 63–93.

[9] C. Carter, J. Catlett, Assessing credit card applications using machine learning, IEEE Expert 2 (3) (1987) 71–79.

[10] S.-Y. Chang, T.-Y. Yeh, An artificial immune classifier for credit scoring analysis, Appl. Soft Comput. 12 (2) (2012) 611–618.

[11] S. Chatterjee, S. Barcun, A nonparametric approach to credit screening, J. Am. Stat. Assoc. 65 (329) (1970) 150–154.

[12] Y.-S. Chen, C.-H. Cheng, Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry, Knowl.-Based Syst. 39 (2013) 224–239.

[13] O. Cordón, A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: designing interpretable genetic fuzzy systems, Int. J. Approx. Reason. 52 (6) (2011) 894–913.

[14] O. Cordón, F. Herrera, F. Hoffman, L. Magdalena, Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases, vol. 19 of Advances in Fuzzy Systems – Applications and Theory, World Scientific, Singapore, New Jersey, London, Hong Kong, 2001.

[15] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197.

[16] A.I. Dimitras, R. Slowinski, R. Susmaga, C. Zopounidis, Business failure prediction using rough sets, Eur. J. Oper. Res. 114 (2) (1999) 263–280.

[17] D. Dubois, H. Prade, What are fuzzy rules and how to use them, Fuzzy Sets Syst. 84 (2) (1996) 169–185.

[18] D. Dubois, H. Prade, L. Ughetto, A new perspective on reasoning with fuzzy rules, Int. J. Intell. Syst. 18 (5) (2003) 541–567.

[19] M. Fazzolari, R. Alcala, Y. Nojima, H. Ishibuchi, F. Herrera, A review of the application of multiobjective evolutionary fuzzy systems: current status and further directions, IEEE Trans. Fuzzy Syst. 21 (1) (2013) 45–65.

[20] H. Frydman, E.I. Altman, D.-L. Kao, Introducing recursive partitioning for financial classification: the case of financial distress, J. Financ. 40 (1) (1985) 269–291.

[21] M.J. Gacto, R. Alcala, F. Herrera, Interpretability of linguistic fuzzy rule-based systems: an overview of interpretability measures, Inf. Sci. 181 (20) (2011) 4340–4360.

[22] F. Glover, Improved linear programming models for discriminant analysis, Decis. Sci. 21 (4) (1990) 771–785.

[23] M.B. Gorzałczany, On some idea of a neuro-fuzzy controller, Inf. Sci. 120 (1–4) (1999) 69–87.

[24] M.B. Gorzałczany, A computational-intelligence-based approach to decision support, in: H. Bunke, A. Kandel (Eds.), Neuro-Fuzzy Pattern Recognition, World Scientific Publishing, Singapore, London, 2000, pp. 51–73.

[25] M.B. Gorzałczany, Computational Intelligence Systems and Applications: Neuro-Fuzzy and Fuzzy Neural Synergisms, Studies in Fuzziness and Soft Computing, Physica-Verlag, Heidelberg, New York, 2002.

[26] M.B. Gorzałczany, A. Głuszek, Neuro-fuzzy systems for rule-based modelling of dynamic processes, in: H.-J. Zimmermann, G. Tselentis, M. Someren, G. Dounias (Eds.), Advances in Computational Intelligence and Learning, vol. 18 of International Series in Intelligent Technologies, Springer, Netherlands, 2002, pp. 135–146.

[27] M.B. Gorzałczany, Z. Piasta, Neuro-fuzzy approach versus rough-set inspired methodology for intelligent decision support, Inf. Sci. 120 (1-4) (1999) 45–68.

[28] M.B. Gorzałczany, F. Rudziński, Measurement data in genetic fuzzy modelling of dynamic systems, Pomiary Automatyka Kontrola 12 (2010) 1420–1423.

[29] M.B. Gorzałczany, F. Rudziński, A modified Pittsburg approach to design a genetic fuzzy rule-based classifier from data, Lect. Notes Comput. Sci. 6113 (2010) 88–96.

[30] M.B. Gorzałczany, F. Rudziński, Accuracy vs. interpretability of fuzzy rule-based classifiers: an evolutionary approach, Lect. Notes Comput. Sci. 7269 (2012) 222–230.

[31] M.B. Gorzałczany, F. Rudziński, Genetic fuzzy rule-based modelling of dynamic systems using time series, Lect. Notes Comput. Sci. 7269 (2012) 231–239.

[32] B.J. Grablowsky, W.K. Talley, Probit and discriminant functions for classifying credit applicants: a comparison, J. Econ. Bus. 33 (3) (1981) 254–261.

[33] M.M. Gupta, M.B. Gorzałczany, Fuzzy neuro-computational technique and its application to modelling and control, in: Proceedings of IEEE International Conference on Fuzzy Systems, 1992, pp. 1271–1274.

[34] J. Han, M. Kamber, Data Mining: Concept and Techniques, 2nd ed., Morgan Kaufmann, 2006.

[35] D.J. Hand, Discrimination and Classification, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley, Chichester, Brisbane, New York, Toronto, 1981.

[36] D.J. Hand, W.E. Henley, Statistical classification methods in consumer credit scoring: a review, J. Roy. Stat. Soc.: Ser. A (Stat. Soc.) 160 (3) (1997) 523–541.

[37] W.E. Henley, Statistical Aspects of Credit Scoring, The Open University, Milton Keynes, UK, 1995 (Ph.D. thesis).

[38] W.E. Henley, D.J. Hand, A k-nearest-neighbour classifier for assessing consumer-credit risk, Statistician 45 (1) (1996) 77–95.

[39] W.E. Henley, D.J. Hand, Construction of a k-nearest-neighbour credit-scoring system, IMA J. Manage. Math. 8 (4) (1997) 305–321.

[40] F. Herrera, Genetic fuzzy systems: taxonomy, current research trends and prospects, Evol. Intell. 1 (1) (2008) 27–46.

[41] F. Hoffmann, B. Baesens, C. Mues, T. Van Gestel, J. Vanthienen, Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms, Eur. J. Oper. Res. 177 (1) (2007) 540–555.

[42] N.-C. Hsieh, Hybrid mining approach in the design of credit scoring models, Expert Syst. Appl. 28 (4) (2005) 655–665.

[43] Y.-C. Hu, Bankruptcy prediction using ELECTRE-based single-layer perceptron, Neurocomputing 72 (13-15) (2009) 3150–3157.

[44] H. Ishibuchi, T. Murata, I.B. Türkşen, Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems, Fuzzy Sets Syst. 89 (2) (1997) 135–150.

[45] H. Ishibuchi, T. Nakashima, T. Murata, Three-objective genetics-based machine learning for linguistic rule extraction, Inf. Sci. 136 (1-4) (2001) 109–133.

[46] Y. Jin (Ed.), Multi-Objective Machine Learning, vol. 16 of Studies in Computational Intelligence, Springer-Verlag, Berlin, Heidelberg, 2006.

[47] S. Jones, D.A. Hensher, Predicting firm financial distress: a mixed logit model, Account. Rev. 79 (4) (2004) 1011–1038.

[48] S. Jones, D.A. Hensher, Modelling corporate failure: A multinomial nested logit analysis for unordered outcomes, Brit. Account. Rev. 39 (1) (2007) 89–107.

[49] J. Kacprzyk, W. Pedrycz (Eds.), Springer Handbook of Computational Intelligence, Springer-Verlag, Berlin, Heidelberg, 2015.
[50] A. Khashman, A neural network model for credit risk evaluation, Int. J. Neural Syst. 19 (4) (2009) 285–294.
[51] M.-J. Kim, D.-K. Kang, Ensemble with neural networks for bankruptcy prediction, Expert Syst. Appl. 37 (4) (2010) 3373–3379.
[52] P. Kolesar, J.L. Showers, A robust credit screening model using categorical data, Manage. Sci. 31 (2) (1985) 123–133.
[53] V. Kozeny, Genetic algorithms for credit scoring: alternative fitness function performance comparison, Expert Syst. Appl. 42 (6) (2015) 2998–3004.
[54] T.-S. Lee, C.-C. Chiu, C.-J. Lu, I.-F. Chen, Credit scoring using the hybrid neural discriminant technique, Expert Syst. Appl. 23 (3) (2002) 245–254.
[55] Y.-C. Lee, Application of support vector machines to corporate credit rating prediction, Expert Syst. Appl. 33 (1) (2007) 67–74.
[56] T. Lensberg, A. Eilifsen, T.E. McKee, Bankruptcy theory development and classification via genetic programming, Eur. J. Oper. Res. 169 (2) (2006) 677–697.
[57] H. Li, J. Sun, Ranking-order case-based reasoning for financial distress prediction, Knowl.-Based Syst. 21 (8) (2008) 868–878.
[58] H. Li, J. Sun, Majority voting combination of multiple case-based reasoning for financial distress prediction, Expert Syst. Appl. 36 (3, Part 1) (2009) 4363–4373.
[59] H. Li, J. Sun, B.-L. Sun, Financial distress prediction based on OR-CBR in the principle of k-nearest neighbors, Expert Syst. Appl. 36 (1) (2009) 643–659.
[60] W.-Y. Lin, Y.-H. Hu, C.-F. Tsai, Machine learning in financial crisis prediction: a survey, IEEE Trans. Syst. Man Cybern. C 42 (4) (2012) 421–436.
[61] R. Malhotra, D.K. Malhotra, Differentiating between good credits and bad credits using neuro-fuzzy systems, Eur. J. Oper. Res. 136 (1) (2002) 190–211.
[62] R. Malhotra, D.K. Malhotra, Evaluating consumer loans using neural networks, Omega 31 (2) (2003) 83–96.
[63] D. Martens, B. Baesens, T. Van Gestel, J. Vanthienen, Comprehensible credit scoring models using rule extraction from support vector machines, Eur. J. Oper. Res. 183 (3) (2007) 1466–1476.
[64] S.-H. Min, J. Lee, I. Han, Hybrid genetic algorithms and support vector machines for bankruptcy prediction, Expert Syst. Appl. 31 (3) (2006) 652–660.
[65] L. Nanni, A. Lumini, An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring, Expert Syst. Appl. 36 (2, Part 2) (2009) 3028–3033.
[66] C.-S. Ong, J.-J. Huang, G.-H. Tzeng, Building credit scoring models using genetic programming, Expert Syst. Appl. 29 (1) (2005) 41–47.
[67] S. Oreski, G. Oreski, Genetic algorithm-based heuristic for feature selection in credit risk assessment, Expert Syst. Appl. 41 (4, Part 2) (2014) 2052–2064.
[68] G. Paleologo, A. Elisseeff, G. Antonini, Subagging for credit scoring models, Eur. J. Oper. Res. 201 (2) (2010) 490–499.
[69] Y. Peng, G. Wang, G. Kou, Y. Shi, An empirical study of classification algorithm evaluation for financial risk prediction, Appl. Soft Comput. 11 (2) (2011) 2906–2915.
[70] G.E. Pinches, K.A. Mingo, A multivariate analysis of industrial bond ratings, J. Financ. 28 (1) (1973) 1–18.
[71] P. Ravi Kumar, V. Ravi, Bankruptcy prediction in banks and firms via statistical and intelligent techniques – a review, Eur. J. Oper. Res. 180 (1) (2007) 1–28.
[72] F. Rudziński, Finding sets of non-dominated solutions with high spread and well-balanced distribution using generalized strength Pareto evolutionary algorithm, in: Proceedings of the IFSA-EUSFLAT 2015, Gijón, Asturias, Spain, 2015.
[73] F. Rudziński, J. Piekoszewski, The maintenance costs estimation of electrical lines with the use of interpretability-oriented genetic fuzzy rule-based systems, Przegląd Elektrotechniczny 8 (2013) 43–47.
[74] E.H. Ruspini, A new approach to clustering, Inf. Control 15 (1) (1969) 22–32.
[75] L. Rutkowski, Flexible Neuro-Fuzzy Systems: Structures, Learning and Performance Evaluation, Kluwer Academic Publisher, Boston, Dordrecht, 2004.
[76] K.B. Schebesch, R. Stecking, Support vector machines for classifying and describing credit applicants: detecting typical and critical regions, J. Oper. Res. Soc. 56 (9) (2005) 1082–1088.
[77] C. Serrano-Cinca, Self organizing neural networks for financial diagnosis, Decis. Support Syst. 17 (3) (1996) 227–238.
[78] V. Stanovov, E. Semenkin, Hybrid self-configuring evolutionary algorithm for automated design of fuzzy logic rule base, in: in: Proceedings of 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2014, pp. 317–321.
[79] J. Sun, M.-Y. Jia, H. Li, AdaBoost ensemble for financial distress prediction: an empirical comparison with data from Chinese listed companies, Expert Syst. Appl. 38 (8) (2011) 9305–9312.
[80] J. Sun, H. Li, Data mining method for listed companies' financial distress prediction, Knowl.-Based Syst. 21 (1) (2008) 1–5.
[81] L.C. Thomas, D.B. Edelman, J.N. Crook, Credit Scoring and Its Applications, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2002.
[82] L.C. Thomas, R.W. Oliver, D.J. Hand, A survey of the issues in consumer credit modelling research, J. Oper. Res. Soc. 56 (2005) 1006–1015.
[83] C.-F. Tsai, Y.-F. Hsu, D.C. Yen, A comparative study of classifier ensembles for bankruptcy prediction, Appl. Soft Comput. 24 (2014) 977–984.
[84] C.-F. Tsai, J.-W. Wu, Using neural network ensembles for bankruptcy prediction and credit scoring, Expert Syst. Appl. 34 (4) (2008) 2639–2649.
[85] G. Wang, J. Ma, L. Huang, K. Xu, Two credit scoring models based on dual strategy ensemble trees, Knowl.-Based Syst. 26 (2012) 61–68.
[86] Y. Wang, S. Wang, K.K. Lai, A new fuzzy support vector machine to evaluate credit risk, IEEE Trans. Fuzzy Syst. 13 (6) (2005) 820–831.
[87] D. West, Neural network credit scoring models, Comput. Oper. Res. 27 (11-12) (2000) 1131–1152.
[88] D. West, S. Dellana, J. Qian, Neural network ensemble strategies for financial decision applications, Comput. Oper. Res. 32 (10) (2005) 2543–2559.
[89] J.C. Wiginton, A note on the comparison of logit and discriminant models of consumer credit behavior, J. Financ. Quant. Anal. 15 (3) (1980) 757–770.
[90] Z.R. Yang, M.B. Platt, H.D. Platt, Probabilistic neural networks in bankruptcy prediction, J. Bus. Res. 44 (2) (1999) 67–74.
[91] P. Yao, Y.-H. Lu, Neighborhood rough set and SVM based hybrid credit scoring classifier, Expert Syst. Appl. 38 (9) (2011) 11300–11304.
[92] C.-C. Yeh, F. Lin, C.-Y. Hsu, A hybrid KMV model, random forests and rough set theory approach for credit rating, Knowl.-Based Syst. 33 (2012) 166–172.
[93] L. Yu, S. Wang, J. Cao, A modified least squares support vector machine classifier with application to credit risk analysis, Int. J. Inf. Technol. Decis. Mak. 8 (4) (2009) 697–710.
[94] L. Yu, S. Wang, K.K. Lai, Credit risk assessment with a multistage neural network ensemble learning approach, Expert Syst. Appl. 34 (2) (2008) 1434–1444.
[95] L. Yu, X. Yao, A total least squares proximal support vector classifier for credit risk evaluation, Soft Comput. 17 (4) (2013) 643–650.
[96] L. Yu, X. Yao, S. Wang, K.K. Lai, Credit risk evaluation using a weighted least squares {SVM} classifier with design of experiment for parameter selection, Expert Syst. Appl. 38 (12) (2011) 15392–15399.
[97] D. Zhang, X. Zhou, S.C.H. Leung, J. Zheng, Vertical bagging decision trees model for credit scoring, Expert Syst. Appl. 37 (12) (2010) 7838–7843.
[98] L. Zhou, K.K. Lai, L. Yu, Credit scoring using support vector machines with direct search for parameters selection, Soft Comput. 13 (2) (2008) 149–155.
[99] X. Zhu, J. Li, D. Wu, H. Wang, C. Liang, Balancing accuracy, complexity and interpretability in consumer credit decision making: a C-TOPSIS classification approach, Knowl.-Based Syst. 52 (2013) 258–267.
[100] E. Zitzler, M. Laumanns, L. Thiele, SPEA2: Improving the strength Pareto evolutionary algorithm for multi-objective optimization, in: Proceedings of Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems, Barcelona, Spain, 2001, pp. 95–100.