Expert Systems With Applications 80 (2017) 75–82

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Genetic algorithm based model for optimizing bank lending decisions

Noura Metawa[a], M. Kabir Hassan[a], Mohamed Elhoseny[b,*]

[a] College of Business Administration, University of New Orleans, New Orleans, USA
[b] Faculty of Computers and Information Sciences, Mansoura University, Mansoura, Egypt

ARTICLE INFO

Article history:
Received 16 January 2017
Revised 8 March 2017
Accepted 9 March 2017
Available online 16 March 2017

Keywords:
Lending decision
Genetic algorithm
Loan portfolio
Bank objectives

ABSTRACT

To avoid the complexity and time consumption of traditional statistical and mathematical programming, intelligent techniques have gained great attention in different financial research areas, especially in banking decisions' optimization. However, choosing optimum bank lending decisions that maximize the bank profit in a credit crunch environment is still a big challenge. For that, this paper proposes an intelligent model based on the Genetic Algorithm (GA) to organize bank lending decisions in a highly competitive environment with a credit crunch constraint (GAMCC). GAMCC provides a framework to optimize bank objectives when constructing the loan portfolio, by maximizing the bank profit and minimizing the probability of bank default in a search for a dynamic lending decision. Compared to the state-of-the art methods, GAMCC is considered a better intelligent tool that enables banks to reduce the loan screening time by a range of 12%–50%. Moreover, it greatly increases the bank profit by a range of 3.9%–8.1%.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

It is clear that the financial crisis was accompanied by a reduction in the credit supply available to all customers (Aguilar, Valenzuela, & Ortiz, 2015). According to recent studies (Judit & Wang, 2012; Michael & Rohwedder, 2010), the nature of the latest financial crisis in US has brought to the fore concerns regarding bank's ability to continue with its traditional bank lending strategies. The main channel through which a banking crisis affect the real economy relates to its ability to provide the credit needed given credit constraints imposed on them. Hence, a key question at the heart of the any financial crisis is whether and how did the banking sector managed to distribute the limited credit available in a way that maximizes their profits in the time of crisis (Manju, Rocholl, & Steffen, 2011). Therefore, there is a need to set an optimal mechanism of bank lending decisions that will maximize the bank profit in a timely manner. The inability of banks to manage loan portfolio efficiently may result in a credit crunch. A credit crunch is often caused by a sustained period of careless and inappropriate lending, resulting in losses for lending institutions and investors in debt when the loans turn sour and the full extent of bad debts becomes known. These challenges have led to a rise in more formal and accurate methods to optimize the lending decision and minimize loan risks. In conjunction, bank lending decision has become a primary tool for financial institutions to increase profit, reduce possible risks, and make managerial decisions.

Unlike the traditional statistical and mathematical programming techniques such as discriminant analysis, linear and logistic regression, linear and quadratic programming, or decision trees; intelligent systems have proven their ability to overcome different challenges in financial research areas, specially in banking loan portfolio optimization (Eletter, Yaseen, & Elrefae, 2010; Ghodselahi & Amirmadhi, 2011; Marque, Garci, & Nchez, 2013; Nazari & Alidadi, 2013). However, most of them are focused on either credit scoring - to determine whether the applied customer is eligible to get the required loan - (Ghodselahi & Amirmadhi, 2011; Marque et al., 2013), or portfolio selection (Berutich, Francisco, Luna, & Quintana, 2016; Lwin, Qu, & MacCarthy, 2017; Saborido, Ruiz, Bermúdez, Vercher, & Luque, 2016) aiming to choose the optimum stocks that maximize the customer profit.

Thus, the problem of bank lending decision in a credit crunch environment- where all applicable customers are eligible to get the desired loan - is an NP-hard optimization problem that can be solved using meta-heuristic algorithms such as evolutionary algorithms (Bhargava, 2013; Ghosh & Tsutsui, 2003; Metawa, Elhoseny, Kabir Hassan, & Hassanien, 2016). The evolutionary computing methods are highly capable of extracting meaning from imprecise data and detecting trends that are too complex to be discovered by either humans or other conventional techniques. For that purpose, this paper proposes an intelligent model based on Genetic Algorithm (GA) to organize bank lending decision in a highly competing environment with credit crunch constraint.

* Corresponding author.
E-mail addresses: nmetawa@uno.edu (N. Metawa), mhassan@uno.edu (M.K. Hassan), mohamed_elhoseny@mans.edu.eg (M. Elhoseny).

http://dx.doi.org/10.1016/j.eswa.2017.03.021
0957-4174/© 2017 Elsevier Ltd. All rights reserved.

The main contribution of this paper is the creation of a GA model that facilitates how banks would make an efficient decision in case of a cut back on lending supply when faced with a liquidity shock, while staying focus on the main objective of bank profit maximization. The main focus of the GA model is two-fold: to stabilize systemically banks while achieving maximum profit, and to establish the capital base so that banks would increase lending efficiently. The GA model takes into account the margins along which banks adjust their loan portfolios in response to the crisis, either through imposing credit supply contractions across all types of borrowers, or disproportionally cut back credit for some specific types of loans. Multiple factors including loan characteristics, creditor ratings and expected loan loss are integrated to GA chromosomes and validation is performed to ensure the optimal decision.

The remaining of this paper is organized as the following: Section 2 discusses the literature review of the bank lending decision's problem and the recent work to solve this problem using GA. Section 3 describes GAMCC and the GA data representation. Then, Section 4 discuss the experimental results. Finally, Section 5 concludes the paper.

2. Literature review

The optimal loan portfolio selection requires a solution of a high-dimensional nonlinear program and is computationally challenging (Elhachlouf, Guennoun, & Hamza, 2012; Sefiane & Benbouziane, 2012). Loan portfolios commonly have a large number of different loans ranging from hundreds to thousands of loans (Melennec, 2000). Every loan is characterized by a highly dimensional vector of loan characteristics including; credit score, collateral, interest rate, loan balance, loan purpose, loan age and size, payment history, and location.

The Objective and constraint functions of bank profit maximization are nonlinear and sometimes non convex which can be costly to evaluate using traditional methods. In addition, the securities that are backed by loan pools-collateralized loan and mortgage obligations - are mostly considered complex derivatives that make the underlying loan portfolio more complicated to estimate (Sirignano & Giesecke, 2015).

Despite the importance of loan portfolio selection decisions, the tradeoff between risky loans with higher returns and safe loans with lower returns has not received nearly as much attention as the equity portfolio selection decisions. According to Altman (1996), loan portfolio selection is based on the standard mean-variance formulation of the portfolio, as well as the unexpected loss from defaults as a risk measure. On the other side, Paris (2005) focuses on the consumer loan portfolio selection problem using one-period, discrete-state and the objective of expected utility. In addition, the mean-variance and utility-based formulations when loans are placed into groups of loans with similar characteristics are studied at Mencia (2012). A multi-objective evolutionary algorithm (MODE-GL) is proposed at Lwin et al. (2017) imposing real-world constraints; cardinality, quantity, pre-assignment, round-lot and class constraints. MODE-GL model uses value-at-risk (VaR) as a market risk measure to assess the financial asset price fluctuations. However, the processing complexity represents the main challenge of this method.

An improvised risk evaluation of Multidimensional Risk prediction clustering Algorithm is proposed at Kavitha (2016) to determine the good and bad loan applicants whether they are applicable or not. The proposed clustering method aims to find the risk percentage to determine whether loan can be sanctioned to a customer or not. During recession and financial crisis era, the illiquidity problem arises leading to credit becoming less available for borrowers, as banks tend to adapt strategies of credit rationing focusing on optimizing their loan portfolio. The insolvency of some banks can lead to a reduction in bank loans, not only because of the bad bank failures but also because the healthy banks tend to be more cautious in their lending practices through cutting back on lending (Teimouri et al., 2016).On this matter, at Deyoung, Gron, Torna, and Wintion (2015) the researchers find that typical US commercial banks reduced their business lending during the financial crisis, this decline was due to increased risk overhang effect (related to higher levels of illiquidity), and increased levels of credit rationing, which is consistent with the higher lender risk aversion. Also, according to Cingano, Manaresi, and Sette (2016) on quantifying the real effect of credit crunch on banks after the financial crisis, they find that the negative shock to bank liquidity significantly deteriorated credit conditions: high-exposure banks tightened credit supply more than low-exposure banks did, even when looking at the same borrower.

Targeting loan portfolio optimization focusing on the goal of loan default and bad loans elimination, the researchers at Agarana, Bishop, and Odetumibi (2014) used goal programming technique to determine the boundaries of loan portfolio decisions that would reduce loan default rate in Nigeria banks. Focusing on Italian banks, an importance sampling (IS) technique for loan portfolio optimization problem is used (Clemente, 2014). The IS technique objective is to minimize the loan portfolio volatility in 23 Italian banks during risky and uncertain economic conditions. In addition, VAR is used at Ning, Wang, and Pan (2014) to measure the uncertain environments of loan portfolio when all the loan rates are subject to uncertain changes.

A loan price differentiation impact on the behaviour of banks potential borrowers in the market for consumer lending is analyzed at Romanyuk (2015). Accordingly, a concept of a decision support system for loan granting based on the use of a continuous loan price function of the borrower's creditworthiness rating is proposed. However, the required time to get the optimum decision using that system represents its main challenge.

At Jat and Xoagub (2016), a fuzzy logic-based automated artificial system is proposed to help in decision-making for disbursement of the bank loan and give the reason for the inferred decision. However, this system is not design to work with credit crunch environments. In addition, uncertain simulations to test VAR technique solutions for loan optimization are designed. Moreover, the authors explained the cases where VAR can be used to maximize loan portfolio returns in dynamic environment. On the other hand, the application of a multi-objective optimization GA model addressing the problem of the risk-return tradeoff in bank loan portfolio is presented at Mukerjee and Deb (2002). The authors of Mukerjee and Deb (2002) obtained an approximation to the set of Pareto-optimal solutions which increases the decision flexibility available to the bank manager and provides a visualization tool for one of the tradeoffs involved. However, that work (Mukerjee & Deb, 2002) did not consider interest rate risks as a factor in the trade off of loan portfolio optimization.

On the same sequence, a genetic algorithm model of loan portfolio multiple constrains optimization of banks' risk-return is suggested at Misra and Sebastian (2013). This model built portfolio with mean-variance dominating for both AAA rating and AA rating and calculating for the loan portfolio risk and return. Their GA technique is applied to a leading bank of India and designed as per Indian Banking Regulations has been outperformed the current portfolio of the bank.

Depending on the above discussion, it is noticed that most of the related work focused on either credit scoring to determine whether the applied customer is eligible to get the required loan, or portfolio optimization which aims to choose the optimum stocks that maximize the customer profit. For that, this paper aims to organize bank lending decision using real world credit crunch

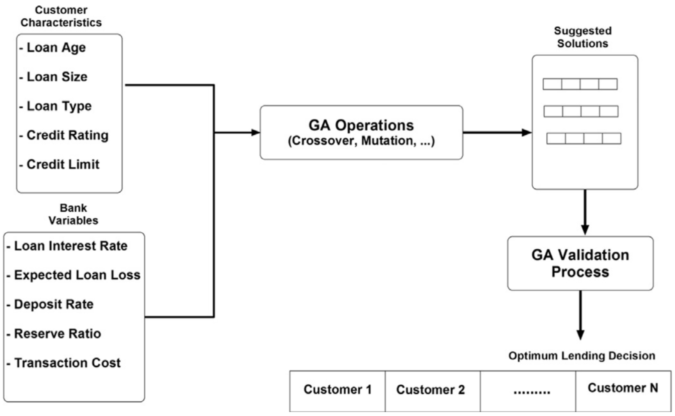

Fig. 1. The working steps of GAMCC. The arrow depicts the data flow.

Table 1
Loan age categories.

Category (α)	Value
1	$1 \le \alpha \le 3$
2	$3 < \alpha \le 5$
3	$5 < \alpha \le 10$
4	$10 < \alpha \le 20$

constraint in order to maximize bank profit and minimize bank default.

3. Methodology

3.1. Problem representation

This paper proposes an efficient, GA-based model to maximize bank profit in lending decision. In our proposed method, the lending decision is dynamically decided based on customer's loan characteristics. With the assumption that all customers are applicable (good) to get the required loan, GA is employed to search for the most suitable customers depending on a set of factors such as loan age, loan size, loan interest rate, loan type, and borrower credit rating. A set of other depending variables, i.e., expected loss on the loan and loan interest rates, are considered as well. Using the chosen customers, the optimum bank lending decision and the loans details are determined. Fig. 1 clarifies the flow of our proposed model.

This paper assumes that each customer has a set of predefined loan characteristics. These characteristics represent the variables of GAMCC. Mathematically, we can represent our problem using these variables to maximize the bank profit. For each customer C the decision variables are:

3.1.1. Loan age α

With an assumption that the maximum number of years for any loan is 20 years, the loan age is divided into four categories depending on its expected period time. Table 1 clarifies these different categories and their corresponding values. Mathematically, let $\vec{A}$ is the set of the good customers who are applicable for loans. So, the loan age can be represented as Eq. (1).

$$\forall\ C \in \vec{A} \longleftrightarrow \alpha : 1 \le \alpha \le 20 \tag{1}$$

3.1.2. ***Credit limit***

The credit limit represents the maximum loan amount that can be given to the customer based on his income and occupation. The credit limit is used to determine the loan size in each case which is also categorized in micro, small, medium and large credit limits as shown in Table 2.

Table 2
Loan size categories.

Category (L)	Value
Micro	$ $0 \le L \le$ $ 13,000
Small	$ 13,001 $< \alpha \le$ $ 50,000
Medium	$ 50,001 $< \alpha \le$ $ 100,000
Large	$ 100,001 $< \alpha \le$ $ 250,000

Table 3
Loan interest rate assignment.

φ	α	r_L Value
M	1	NA
	2	NA
	3	NA
	4	$0.021 \le r_L \le 0.028$
P	1	$0.0599 \le r_L \le 0.0601$
	2	$0.0601 < r_L \le 0.0604$
	3	$0.0604 < r_L \le 0.0609$
A	1	$0.0339 \le r_L \le 0.03349$
	2	$0.0349 < r_L \le 0.0379$
	3	$0.0379 < r_L \le 0.0399$

Table 4
The credit rating and the corresponding expected loan loss.

Credit Rating	λ Value
AAA	$0.0002 \le \lambda \le 0.0003$
AA	$0.0003 < \lambda \le 0.001$
A	$0.001 < \lambda \le 0.0024$
BBB	$0.0024 < \lambda \le 0.0058$
BB	$0.0058 < \lambda \le 0.0119$

3.1.3. ***Loan size*** L

Depending on the credit limit, the loan size is determined. The loan size (L) determines the amount of loan requested by a specific customer (C). This amount is considered an effective variable in lending decision due to its effect of the total allowable loan amount of the bank which is determined by a required reserved deposit ratio called (r_D). (L) is predefined based on the purpose for which (C) is requested the loan. For that, we classifies (L) in four different categories as shown at Table 2. The maximum allowable loan for each customer is one million dollar.

3.1.4. ***Loan type*** φ

In our model, there are three types of loans: Mortgage (M), Personal (P), and Auto (A). In future, we are planning to add corporate loans as the fourth type. It is not allowed for customer to request a personal or an auto loan with more than 10 years of loan age which can be represented as a mathematical constraint in our model (See Eq. (2)).

$$\forall \varphi \in (P, A) \longleftrightarrow L \le 10 \tag{2}$$

3.1.5. ***Loan interest rate*** $\mathbf{r_L}$

Based on the values of φ and α, the loan interest rate r_L is assigned using interest rates data from world banks (Chase, and capital One), the rate is averaged for the years 2015–2016. In our model r_L is classified into three different categories and its value is determined according to the average loan rates of the largest US banks, i.e., Chase, and Capital One. The details of value assignment are listed at Table 3.

3.1.6. ***Expected loan loss (λ) and borrow credit rating***

Borrow Credit Rating is used to measure the range of the expected loan loss (λ). Good creditor customers are divided into five credit ratings ranges as shown at Table 4.

Table 5
Set of parameters to control the GA evolution process.

Objective	- Determine a bank lending decision that maximizes the bank profit. - Determine a bank lending decision that minimizes the crediting cost.
Raw Fitness	$F_x = \vartheta + \varpi - \beta - \sum_{i=0}^{n} \lambda$
Parameters	- Population Size $(n) = 60$ - GA Generations $(\Re) = 60$ - Crossover ratio $= 0.8$ - Mutation ratio $= 0.006$ - Reproduction ratio$(n) = 0.194$ - Basic selection method is spanning the weighted roulette wheel selection six times.
Stopping Criteria	-The evolution continues until: $F_x(t) = \Re$ where t is the time interval

Table 6
Set of variables to generate simulated data. The values of D is represented in millions. The improvement ratio of F_x is shown at the last two columns.

Number of Customers	D	K	Accepted Customers	M(%)	P(%)	A(%)	F_x	
							TLP	MODE-GL
200	5	0.00	77	31	38	31	1.9	1.4
400	5	0.00	82	41	37	22	2.4	1.7
600	10	0.00	153	32	35	33	4.1	2.1
800	10	0.00	299	41	39	20	0.2	0.01
1000	15	0.15	588	32	36	32	1.7	0.9
2000	100	0.20	1115	27	41	32	5.3	1.8

Table 7
Impact of D on the lending decision. The values of D is represented in millions.

Number of Customers	D	K	Accepted Customers	M(%)	P(%)	A(%)
	5	0.00	77	31	38	31
	10	0.00	96	38	30	32
200	15	0.15	115	39	31	30
	20	0.20	141	42	30	28
	5	0.00	82	41	37	22
	10	0.00	101	28	35	37
400	15	0.15	200	41	29	30
	20	0.20	314	33	25	42
	5	0.00	141	35	33	32
	10	0.00	153	32	35	33
600	15	0.15	374	45	21	34
	20	0.20	427	30	28	42
	5	0.00	157	33	38	29
	10	0.00	299	24	41	35
800	15	0.15	438	32	30	38
	20	0.20	615	29	37	34
	5	0.00	289	40	29	31
	10	0.00	314	29	32	39
1000	15	0.15	588	32	36	32
	20	0.20	703	38	37	25

Table 8
Impact of credit rating on the lending decision.

Number of Customers	Accepted Customers	AAA(%)	AA(%)	A(%)	BBB(%)	BB(%)
	77	39	21	19	12	9
	96	41	20	10	12	17
200	115	35	24	32	7	2
	141	52	15	12	10	11
	82	67	10	4	12	7
	101	28	12	24	21	15
400	200	52	22	17	6	3
	314	47	41	11	1	0
	141	62	24	5	5	4
	153	35	41	15	7	2
600	374	54	20	20	0	6
	427	51	24	15	7	3
	157	35	14	33	6	12
	299	39	32	21	4	4
800	438	24	29	23	14	10
	615	40	27	17	12	7
	289	61	21	15	3	0
	314	48	43	9	0	0
1000	588	56	28	12	2	2
	703	40	36	19	4	1

3.2. Chromosome encoding

The chromosome of the GA contains all the building blocks to a solution of the problem at hand in a form that is suitable for the genetic operators and the fitness function. In genetic algorithms, a chromosome is a set of parameters or factors that determine a proposed solution to the intended problem. In our proposed model, each gene in the chromosome indicates to a candidate customer. The value of a gene can be either 1 or 0, where 1 indicates that the corresponding customer is elected to get the loan and 0 indicates a non-selected customer. Fig. 2 depicts a chromosome for a field with 25 customers.

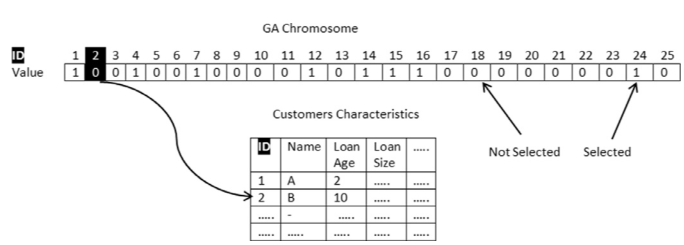

Fig. 2. Chromosome representation for 25 customers.

As shown at Fig. 2, each customer is uniquely identified by ID number. The ID number is used for mapping between GA chromosome and the customer dataset.

Table 9
Bank lending decision using the real data.

Applied Customers	Accepted Customers	Selected M (%)	Selected P (%)	Selected A (%)	F_x	
					TLP	MODE-GL
100	72	41	19	40	3.9	2.1
300	163	36	29	35	4.5	2.8
500	205	40	31	29	7.2	3.9
700	351	29	35	36	8.1	5.4
1000	395	33	28	39	2.9	2.8
1300	503	38	31	31	5.0	2.4
1500	710	45	16	39	4.2	3.5
1700	798	22	41	37	6.8	3.9
2000	871	50	22	28	4.1	3.2

3.3. Genetic algorithm fitness function

The GA's fitness function (F_x) simply consists loan revenue (ϑ), loans cost (μ), total transaction cost (ϖ), and cost of demand deposit (β). The main objective is to maximize $F(x)$. As shown at Eq. (3), the value of (ϑ) is calculated depending on the values of r_L, L, and λ.

$$\vartheta = \sum_{i=0}^{n}(r_L L - \lambda) \tag{3}$$

where N is the number of customers. In addition, the loan cost μ is determined by L and the predetermined institutional cost (δ) as shown at Eq. (4).

$$\mu = \sum_{i=0}^{n} L\delta \tag{4}$$

The total transaction cost of the expected lending decision ϖ is determined by the institutional transaction cost (T) and the customer transaction rate (r_T) as shown at Eq. (5).

$$\varpi = \sum_{i=0}^{n} r_L T \tag{5}$$

The value of the institutional transaction cost can be obtained as defined at Eq. (6) where K is the required reserved ratio of the financial institution's deposit D. In addition, r_T is determined based on the bank's deposit interest rate r_D. However, in this paper we assume that $r_T = 0.01$ in all our experiments.

$$T = (1 - K)D - L \tag{6}$$

Finally, the cost of demand deposit β is simply defined at Eq. (7).

$$\beta = r_D D \tag{7}$$

Where r_D is given by a weighted average of all different deposits rates based on deposit type (checking, saving) and based on deposit age ranging from (3 months till 10 years). In the GA model we left this rate to vary randomly for each bank. And for the fitness function calculation we used average rate of 0.9% for the return on deposit.

Dependently, the GA's fitness function is hence defined as Eq. (8) with the aim of profit maximization.

$$F_x = \vartheta + \varpi - \beta - \sum_{i=0}^{n} \lambda \tag{8}$$

3.4. Genetic algorithm operations and lending decision validation

To be self-content, we briefly review the main operations of GAs (Elhoseny et al., 2015; Kucukkoc, Karaoglan, & Yaman, 2013).

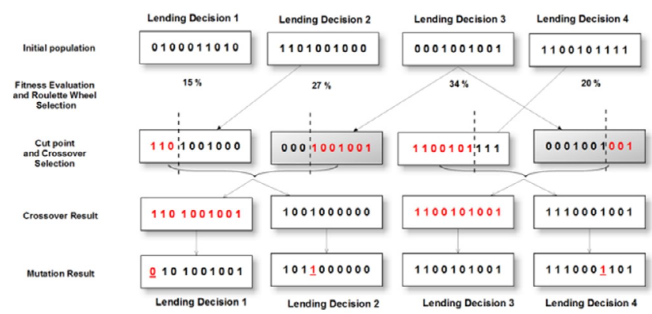

Fig. 3. An example for the GA operations that are used in GAMCC.

GA is adaptive search method that simulates some of the natural processes: selection, information, inheritance, random mutation and population dynamics (Yuan, Elhoseny, Minir, & Riad, 2017). A GA starts with a population of strings and thereafter generates successive populations of strings. A simple GA consists of three operators:

- **Reproduction:** is the process of keeping the same chromosome without changes and transfer it to the next generation. The input and the output of this process is the same chromosome.
- **Crossover:** is the process of concatenating two chromosomes to generate a new two chromosomes by switching genes. The input of this process is two chromosomes while its output is two different chromosomes. A simple one-point crossover operation for binary coded populations have been used. For example, let $I = \{s_1, \ldots, s_j, \ldots, s_n\}$ and $I' = \{s'_1, \ldots, s'_j, \ldots, s'_n\}$ be two different indexes in the current population P. The crossover point was defined by randomly generating an integer j in the range [1, n]. Then the resulting crossed indexes are $I = \{s_1, \ldots, s_{j-1}, s'_j, \ldots, s'_n\}$ and $I' = \{s'_1, \ldots, s'_{j-1}, s_j, \ldots, s_n\}$ (See Fig. 3).
- **Mutation:** is the process of randomly revers the value of one gene in a chromosome. So, the input is a single chromosome and the output is different one. The integer parameter to undergo a mutation, let us say s_j, is selected randomly. Then it mutates into $s'_j = 0$ if $s_j = 1$ and into $s'_j = 1$ otherwise (See Fig. 3).

When crossover is determined not to be conducted, the parent chromosomes are duplicated to the offspring without change. Varying the crossover probability alters the evolution speed of the search process. In practice, the value of crossover is close to 1. Different from the crossover probability, the mutation probability is usually fairly small. Essentially mutation operation could create completely new species, i.e., an arbitrary locus in the fitness landscape. Hence, it is a means to get out of a local optimum.

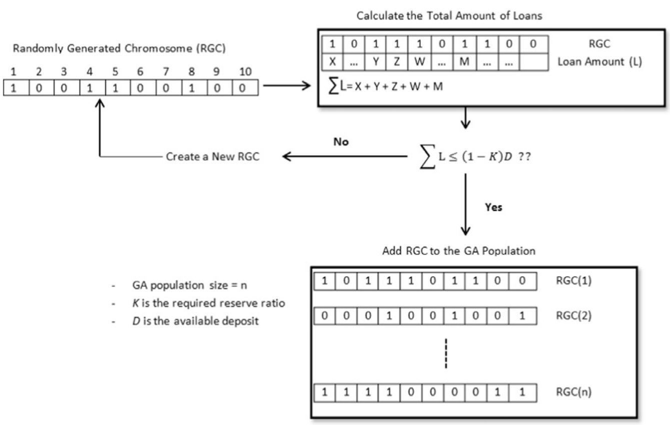

Fig. 4. The GA validation process to generate a GA population with n chromosomes.

Algorithm 1 Bank lending decision algorithm.

Randommly generate a pool of S Lending Decisions $X = \{x_1, x_2, \dots, x_S\}$.
Represent each decision in S by one chromosome to get a pool P of chromosomes $U = \{u_1, u_2, \dots, u_P\}$.
$\forall u_i \in U$, Forming the Lending Decision According to Parameters)
Evaluate the fitness of each $u_i \in U$ using F_x.
for $q = 1, 2, \dots, Q$ **do**
  $\tilde{U} \Leftarrow \emptyset$
  for $p = 1, 2, \dots, P$ **do**
    Randomly select $u_a, u_b \in U$ $(a \neq b)$ based on the normalized fitness $\tilde{f}(u)$:

$$\tilde{f}(u) = \frac{f(u)}{\sum_c f(u)}.$$

    Cross over u_a and u_b according to α

$$\mathcal{C}(u_a, u_b|\alpha) \Rightarrow u'_a, u'_b.$$

    Perform mutation on u'_a and u'_b according to $\vec{m}$

$$\mathcal{M}(u'_a|\beta) \Rightarrow \tilde{u}_a, \quad \mathcal{M}(u'_b|\beta) \Rightarrow \tilde{u}_b.$$

    Evaluate $f(\tilde{u}_a)$ and $f(\tilde{u}_b)$.
    $\tilde{U} \Leftarrow \tilde{U} \cup \{\tilde{u}_a, \tilde{u}_b\}$
  end for
  $U \Leftarrow \{u_i; u_i \in \tilde{U} and f(u_i)\}$
end for
Find the chromosome u^* that satisfies

$$u^* = \arg\max_u f(u), u \in U$$

Return the lending decision x^* by mapping u^* back

Table 10
The average time used to determine the optimum lending decision.

Number of Customers	Method	Mean	STD
	TLP	0.66	0.21
200	MODE-GL	0.48	0.18
	GAMCC	0.42	0.10
	TLP	0.74	0.31
400	MODE-GL	0.54	0.26
	GAMCC	0.45	0.22
	TLP	1.03	0.21
600	MODE-GL	0.71	0.12
	GAMCC	0.51	0.14
	TLP	1.22	0.25
800	MODE-GL	0.79	0.31
	GAMCC	0.65	0.28
	TLP	1.29	0.33
1000	MODE-GL	0.88	0.38
	GAMCC	0.74	0.47

An example that clarifies the GA operations in our proposed model is shown at Fig. 3. First, the GA creates the initial population by randomly generates a set of lending decisions. Then, each decision is evaluated using the proposed fitness function. The fitness function's value of a lending decision determines its probability to be selected as a chromosome at the new population for the next GA generation as shown at Eqs. (9) and (10). Eq. (9) is used to calculate the probability of select (P_s) for each chromosome (ℓ). While Eq. (10) calculates the expected count of select (π), and the actual count of select for each chromosome as the following:

$$P_s(i) = \frac{f_x(\ell)}{\sum_{i=0}^{n} f_x(i)} \tag{9}$$

$$\pi = \frac{f_x(\ell)}{[\sum_{i=1}^{n} f_x(i)/n]} \tag{10}$$

Where n is the number of the chromosomes in the population. A new generation of the GA begins with reproduction. We select the mating pool of the next generation by spinning the weighted roulette wheel six times. So, the best chromosome representation get more copies, the average stay even, and the worst die off and will be excluded for further processing. GAMCC develops a GA validation algorithm which used to validate each new generated chromosome before adding it to the GA population. As shown at Fig. 4, the validation algorithm calculates the total loan amount of the expected lending decision to make sure that its value is not exceed the required reserved ratio of the bank deposit. The validated chromosome is then added to the population. If the chromosome violates the validation condition, a new randomly generated chromosome (RGC) is created and he process will be repeated until the number of validated chromosome is equal to the population size n.

3.5. Pseudo code of GAMCC

As we mentioned before, our proposed model uses a binary chromosome to specify the selected customer in the lending decision, in which a digit one represents an elected customer and a zero represents a customer that is not eligible. Algorithm 1 presents the pseudo code of GAMCC. In this algorithm, $q \in [1, Q]$ denotes the number of generations, and the population size is P. The pool of chromosome, denoted by U, is initialized with randomly generated individuals. An intermediate pool of chromosomes, denoted by $\tilde{U}$, is used to hold the individuals created in a generation, and depending on the needs user can specify any intermediate population size that is greater than the initial population size P. In crossover operation, two chromosomes are randomly selected from U and, according to the crossover probability; two new chromosomes are created by switching consecutive genes. In mutation operations, the value of a randomly picked gene is altered between 0 and 1 according to the mutation probability $\vec{m}$.

4. Results and discussion

First, all data are obtained from World Bank public database for year 2016. In our evaluation, the GA parameters used in our experi-

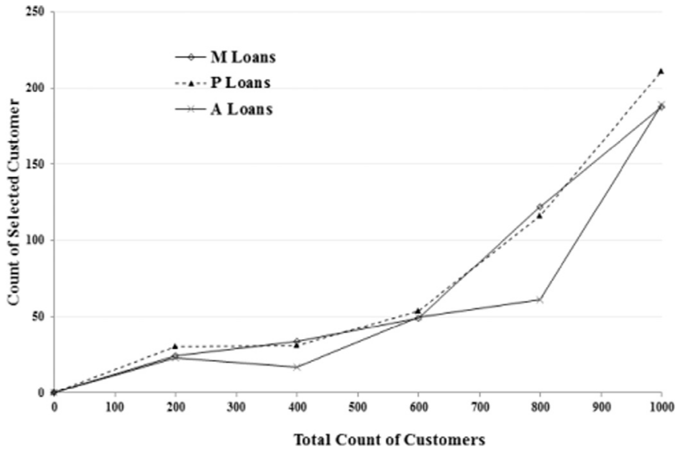

Fig. 5. The number of selected customer for each type of loan.

Fig. 6. The improvement ratio of GAMCC compared to TLP.

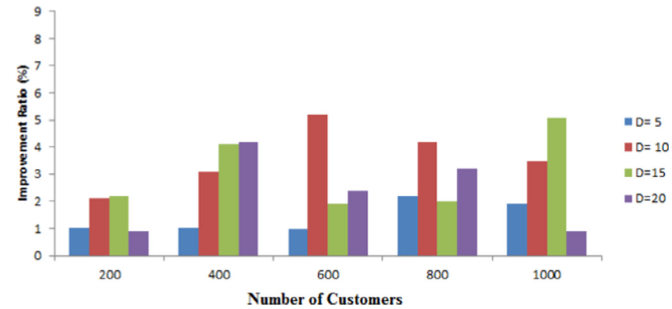

Fig. 7. The improvement ratio of GAMCC compared to MODE-GL.

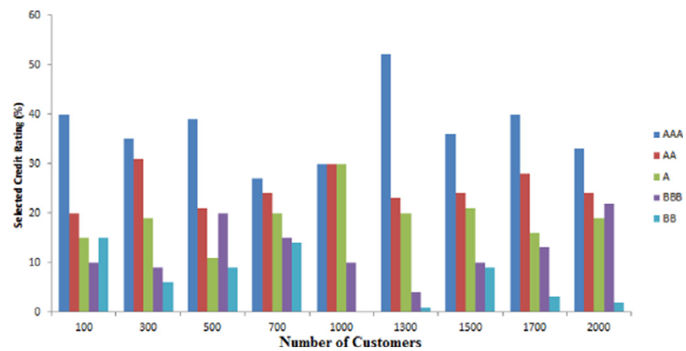

Fig. 8. The ratio of selected customers by GAMCC based on their credit rating compared to TLP.

ments are listed in Table 5. The analysis of the results is conducted based on 10 experiments.

To evaluate GAMCC, two scenarios are created to study the routine method performance in real and simulated environments. In the simulation scenario, GAMCC randomly generates customer loan's characteristics for a defined number of customers with different values and exports them to a database file. Then, GA starts its work using the generated data.

4.1. First scenario: simulated data

In this scenario, we generate the simulated data for different cases as shown at Table 6. This simulation assumes that $r_D = 0.009$ a weighted average of the deposit interest rates over short and long run. In addition, the institutional cost is 0.0025 in all cases based on the previous literature. The last two columns illustrate the improvement ratio of F_x compared to the traditional linear programming method (TLP) and MODE-GL, respectively.

As shown at Table 6, the loan lending decisions are distributed to the three types of loans. Although, there is no constraints for the ratio of each type, it seems that the GA balanced between them. Compared to MODE-GL, GAMCC achieves an improvement ratio in the bank profit in a range of 0.01%–1.8% using the simulated data. While, the improvement ratio is between 0.0.2% and 5.3% compared to TLP.

Based on Table 6, Fig. 5 shows the number of selected customer for each type of loan. We can say, GAMCC provides a great way to balance between all types of loans.

Depending on the available deposit D, the loan decision is mostly determined. For that, we measured the impact of D on the number of accepted customers. Table 7 shows the effect of changing D on the same number of applied customers.

Using the results shown in Table 7, Figs. 6 and 7 illustrate the improvement ratio of GAMCC using different D values, compared to TLP and MODE-GL, respectively.

On the other hand, the credit rating plays an important role in selecting the corresponding customer. Table 8 shows the ratio of each credit rating category depending on the suggested lending decision.

4.2. Second scenario: real data

The second scenario aims to evaluate the system performance using a pre-known solution for using real data collected from Southern Louisiana Credit Union. Similar to the first scenario, GAMCC greatly increases the bank profit compared to TLP and MODE-GL, using the real data as shown at Table 9. The last two columns of Table 9 show the increasing ratio in F_x that was achieved using GAMCC compared to TLP and MODE-GL, respectively.

In addition, Fig. 8 shows the ratio of selected customer by GAMCC based on their credit rating, compared to TLP.

Efficiency is an important factor in real-world applications. Our experiments are conducted in a computer with Intel core i5 2.6 GHz CPU, 4GB memory, and Windows 8 operating system. The algorithms are implemented in C#. Table 10 lists the average time used to determine the lending decision in each case of run. The time reported is at the last GA generation. The number at the last column is the standard deviation. As shown at Table 10, the average time used by GAMCC is less than the required time to get the optimum solution using either TLP or MODE-GL in all cases. Note that the most time-consuming process in GAMCC is evaluating fitness, which can be implemented with parallel programming to improve efficiency in case of complicated multi-objective problems. Compared to TLP, GAMCC reduces the loan screening time in a range of 36%–50%. While the improvement ratio compared to MODE-GL is in a range of 12%–28%.

Table 11
GAMCC consistency evaluation.

Loan Type	M	P	A
Mean	0.43	0.32	0.25
STD	0.011	0.021	0.064

On the other hand, effective GA model must guarantee the results consistency. For that, we run GAMCC 10 times using the same real data and the same GA parameters. Table 11 lists the average of the selected loan type and its standard deviation at various run times. Small STDs indicate balanced ratio of a specific loan type.

5. Conclusion and future work

The bank lending decision problem in a credit crunch environment is an NP-hard optimization problem that can be solved using meta-heuristic algorithms such as evolutionary algorithms. To avoid the complexity and time consumption of the traditional statistical and mathematical programming techniques such as discriminant analysis, linear and logistic regression, linear and quadratic programming, or decision trees; intelligent techniques have proven their ability to overcome the different challenges in financial research areas, specially in banking loan optimization. This paper proposes an intelligent model based on GA to organize bank lending decision in highly competing environment with credit crunch constraint. GAMCC facilitates how banks would make an efficient decision in case of a cut back on lending supply when faced with a negative liquidity shock, while staying focus on the main objective of bank profit maximization. GAMCC is tested using simulated and real data as well. The results shows that GAMCC greatly increases the bank profit using the suggested lending decision in case of using the real data. Based on different experiments, GAMCC increases the bank profit ratio in both simulated and real environments. Compared to TLP, GAMCC reduces the loan screening time in a range of 36%–50%. While the improvement ratio compared to MODE-GL is in a range of 12%–28%.

For future research, GAMCC can be used for optimizing the corporate lending decisions especially for small and medium companies. Second, it can accommodate the interest rate risk as a dynamic factor in the bank lending decision. Finally, it can be applied to optimize the hedge fund predefined strategies to ensure a profitable return for the fund partners.

References

Agarana, C., Bishop, A., & Odetumibi, A. (2014). Optimization of bank loan portfolio management using goal programing technique. *International Journal of Research in Applied Natural and Social Sciences, 8*(2), 43–52.

Aguilar, R., Valenzuela, M., & Ortiz, J. (2015). Genetic algorithms and Darwinian approaches in financial applications. *Expert Systems with Applications, 42*(21), 7684–7697.

Altman, E. (1996). *Corporate bond and commercial loan portfolio analysis*. Wharton Financial Institutions Center, Wharton School, University of Pennsylvania, USA.

Berutich, J., Francisco, L., Luna, F., & Quintana, D. (2016). Robust technical trading strategies using GP for algorithmic portfolio selection. *Expert Systems with Applications, 46*, 307–315. (http://dx.doi.org/10.1016/j.eswa.2015.10.040).

Bhargava, S. (2013). A note on evolutionary algorithms and its applications. *Adults Learning Mathematics: An International Journal, 8*(1), 31–45.

Cingano, F., Manaresi, F., & Sette, E. (2016). Does credit crunch investment down? New evidence on the real effects of the bank-lending channel. *Review of Financial Studies, 29*(10), 2737–2773. doi:10.1093/rfs/hhw040.

Clemente, A. (2014). Improving loan portfolio optimization by importance sampling techniques: Evidence on Italian banking books. *Economic Notes by Banca Monte dei Paschi di Siena SpA, 2*(43), 167–191.

Deyoung, R., Gron, A., Torna, G., & Wintion, A. (2015). Risk overhang and loan portfolio decisions: Small business loan supply before and during the financial crisis. *The Journal Of Finance, 6*(LXX).

Eletter, S., Yaseen, S., & Elrefae, G. (2010). Neuro-based artificial intelligence model for loan decisions. *American Journal of Economics and Business Administration, 2*(1), 27–34.

Elhachlouf, M., Guennoun, Z., & Hamza, F. (2012). Stocks portfolio optimization using classification and genetic algorithms. *Applied Mathematical Sciences, 6*(94), 4673–4684.

Elhoseny, M., Yuan, X., Yu, Z., Mao, C., El-Minir, H., & Riad, A. (2015). Balancing energy consumption in heterogeneous wireless sensor networks using genetic algorithm. *IEEE Communications Letters, 19*(2), 2194–2197. doi:10.1109/LCOMM.2014.2381226.

Ghodselahi, A., & Amirmadhi, A. (2011). Application of artificial intelligence techniques for credit risk evaluation. *International Journal of Modeling and Optimization, 1*(3), 243–249.

Ghosh, A., & Tsutsui, S. (2003). *Advances in evolutionary computing: Theory and applications*. Springer-Verlag New York, Inc., New York, USA (pp. 175–192).

Jat, D., & Xoagub, A. (2016). Fuzzy logic-based expert system for assessment of bank loan applications in Namibia. In *Proceedings of the international congress on information and communication technology: vol. 2* (pp. 645–652). Springer.

Judit, M., & Wang, C. (2012). *The great recession and bank lending to small businesses*. Federal Reserve Bank of Boston. Working paper. (pp. 11–16)

Kavitha, K. (2016). Clustering loan applicants based on risk percentage using K-means clustering techniques. *International Journal of Advanced Research in Computer Science and Software Engineering, 6*(2), 162–166.

Kucukkoc, I., Karaoglan, A., & Yaman, R. (2013). Using response surface design to determine the optimal parameters of genetic algorithm and a case study. *International Journal of Production Research, 51*(17), 5039–5054.

Lwin, K., Qu, R., & MacCarthy, B. (2017). Mean-VaR portfolio optimization: A nonparametric approach. *European Journal of Operational Research, 260*(2), 751–766. (http://dx.doi.org/10.1016/j.ejor.2017.01.005).

Manju, P., Rocholl, J., & Steffen, S. (2011). Global retail lending in the aftermath of the U.S. financial crisis: Distinguishing between supply and demand factors. *Journal of Financial Economics, 100*, 556–578.

Marque, A., Garci, V., & Nchez, J. (2013). A literature review on the application of evolutionary computing to credit scoring. *Journal of the Operational Research Society, 2013, 64*, 1384–1399.

Melennec, O. (2000). *CBO : CLO: CDO: A practical guide for investors*. Societe Generale ABS Research (pp. 1–14)..

Mencia, J. (2012). Assessing the risk-return trade-off in loan portfolios. *Journal of Banking and Finance, 36*(6), 1665–1677.

Metawa, N., Elhoseny, M., Kabir Hassan, M., & Hassanien, A. (2016). Loan portfolio optimization using genetic algorithm: A case of credit constraints. *12th International Computer Engineering Conference (ICENCO), IEEE*, 59–64. doi:10.1109/ICENCO.2016.7856446.

Michael, D., & Rohwedder, H. (2010). Effects of the financial crisis and great recession on American households. *The national bureau of economic research, working paper no. 16407*.

Misra, A., & Sebastian, J. (2013). Portfolio optimization of commercial banks-an application of genetic algorithm. *European Journal of Business and Management, 6*(5), 120–129.

Mukerjee, A., & Deb, K. (2002). Multi-objective evolutionary algorithms for the risk-return trade-off in bank loan management. *International Transactions in Operational Research, 4*(5).

Nazari, M., & Alidadi, M. (2013). Measuring credit risk of bank customers using artificial neural network. *Journal of Management Research, 5*(2), 17–27.

Ning, Y., Wang, X., & Pan, D. (2014). VaR for loan portfolio in uncertain environment. *Seventh international joint conference on computational sciences and optimization, IEEE*. doi:10.1109/CSO.2014.158.

Paris, F. (2005). Selecting an optimal portfolio of consumer loans by applying the state preference approach. *European Journal of Operational Research, 163*(1), 230–241.

Romanyuk, K. (2015). Concept of a decision support system for a loan granting based on continuous price function. *SAI intelligent systems conference (IntelliSys), IEEE*.

Saborido, R., Ruiz, A., Bermúdez, J., Vercher, E., & Luque, M. (2016). Evolutionary multi-objective optimization algorithms for fuzzy portfolio selection. *Applied Soft Computing, 39*, 48–63. (http://dx.doi.org/10.1016/j.asoc.2015.11.005).

Sefiane, S., & Benbouziane, B. (2012). Portfolio selection using genetic algorithm. *Journal of Applied Finance and Banking, 2*(4), 143–154.

Sirignano, J., & Giesecke, K. (2015). *Risk analysis for large pools of loans*. Stanford University, USA. Working Paper

Teimouri, S., & Dutta, N. (2016). Investment and bank credit recovery after banking crises. *Journal of Financial Stability, 26*, 306–327. (http://dx.doi.org/10.1016/j.jfs.2016.07.013).

Yuan, X., Elhoseny, M., Minir, H., & Riad, A. (2017). A genetic algorithm-based, dynamic clustering method towards improved WSN longevity. *Journal of Network and Systems Management, Springer US, 25*(1), 21–46. doi:10.1007/s10922-016-9379-7.