Bankruptcy prediction in Colombian case, using multilayer perceptron trained with memetic algorithm

Student:Iván Andrés Trujillo Abella,

Proffesors: Eliana María González Neira & Gabriel Mauricio Zambrano Rey

Abstract—Literature about Bankruptcy prediction is still incipient, therefore this work try to fill this gap by using machine learning and metaheuristics techniques to find an optimal set of weights in a MLP model.

Index Terms—Machine learning, Bankruptcy, Metaheuristics, Evolutionary Algorithms, Local Search, Memetic algorithms, Neural Networks, Multilayer Perceptron.

I. INTRODUCTION

BANKRUPTCY is an important economic problem, given that means the inability of a firm to fill its financial obligations either permanently or temporarily. Therefore the ability to predict it, allows lenders to avoid important monetary losses, but this also has a social cost if we classify a company as risky when it is not, denying it the access to financing **C1**.

Since Altman's publication **I2** on forecasting bankruptcy using Multiple Discriminant Analysis (MDA), more flexible and accurate methodologies have been considered in bankruptcy prediction, such as logistic regression **I3** and hazard models **I4**. Moreover, the problem also has attracted the use of Machine Learning (ML) techniques. According to the findings described in **I5**, **I6**, **I7**, **I8** the ML algorithms outperform traditional statistical techniques in bankruptcy prediction, given that do not imply assumptions such as normality, linearity, and independence among predictors.

II. NEURAL NETWORKS

Among the most popular ML techniques to predict bankruptcy appear the Neural Networks (NN) that have been shown outperform traditional statistical techniques **I8**. However, its most common learning algorithms are based on gradient methods that could get trapped in local minima **BPO**, this problem attracted the uses of metaheuristics as mechanism to reach global solutions.

A. Multilayer Perceptron(MLP)

MLP is a model of a Fed Forward Neural Network (FFNN) that given its universal approximator proof **universal** has been used to implement metaheuristcs as learning algorithms. Genetic Algorithms (GAs), Particle Swarm Optimization, Cuckoo Search and Grey Wolf Optimizer are some examples of such metaheuristics used as learning algorithms to train MLPs **REVM**.

According to the Figure: 1 an MLP is composed of *input, hidden* and *output layers* that aim to mimic nervous system. Its objective is minimize a loss function iteratively by updating the weights $w_{j,i}$ based on the error.
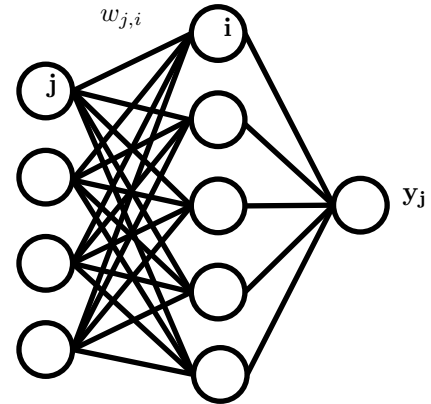

Fig. 1: Multilayer Perceptron illustration

An example of the possible learning algorithm is Gradient Descent(GD):

$$\vec{w}_t = \vec{w}_{t-1} - \alpha \nabla f(\vec{w}_{t-1}) \qquad (1)$$

In summary, training an MLP is an optimization problem. However, as mentioned above, gradient-based methods are not guaranteed to find the global solution. Therefore, some works use metaheuristics such as genetic algorithms to train MLPs.

III. GENETIC ALGORITHMS

GAs are utilized for solving optimization problems and were designed based on the principles of Darwin's theory of evolution, for instance;*"individuals with better features have major probability of have offspring"*.

According to **ES1** an elementary genetic algorithm can be described as follows: *population initialization, evaluation, selection, crossover, mutation and replacement*.

To implement a GA, two elements are required; a solution representation known as *chromosome* and a measure of its capacity to solve the problem or degree of accuracy to fit the environment named *fitness value*.

In MLP training, a potential chromosome representation can be illustrated as depicted in Figure 2. In this representation, each $w_{i,j}$ represents a gene, and its specific value corresponds

to an allele. In classification tasks, some studies utilize the Root Mean Squared Error (RMSE) as the fitness measure. Thus, the fitness function for $i-th$ MLP in $t-th$ iteration will be written as $f_i(t)$.

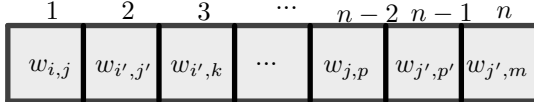| 1	2	3	$\cdots$		$n-2$	$n-1$	n
$w_{i,j}$	$w_{i',j'}$	$w_{i',k}$	$\cdots$		$w_{j,p}$	$w_{j',p'}$	$w_{j',m}$

Fig. 2: Chromosome representation of a MLP

There are different ways to implement each step in GAs. One possible combination of these steps is described as follows:

1) Population initialization: Generate T random samples of chromosomes of the same longitude n.

2) Evaluation: Each solution candidate have associated a fitness value. RMSE is defined as:

$$f_i(t) = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(y_j - \hat{y}_j)^2} \qquad (2)$$

Here, m represents the total number of observations or patterns in the dataset. y_j refers to the target value (for instance if the firm is bankrupt or not), and $\hat{y}_j$ represents the predicted classification using the weights of the $i-th$ MLP.

3) Parents selection: Select possible chromosomes as parents according to its fitness;

$$P_{x_i} = \frac{f_i(t)}{\sum_j^n f_j(t)} \qquad (3)$$

thus P_{x_i} represent the probability that the a chromosome will be selected as parent (to be selected to interchange genetic material).

4) Chromosomal crossover: Interchange genetic material between two parents to create offspring, choosing a random locus (position in chromosome), now α_a and α_b are parents coded as indexing arrays, if the locus is equal to π then the offspring will be $\alpha_a[:\pi]$ concatenate with $\alpha_b[\pi:]$.

5) Mutation: Adding random mutations produce a search of solutions, this process could be summarized as change the allele in a random locus.

$$w_{i,j} = w_{i,j} + N(\bar{x}, \sigma) \qquad (4)$$

6) new generation: Which of the offspring and parent will be selected to be the population in the following generation, an easy strategy is replace all current population by the offspring.

All steps are repeated until a stop criteria is reached, for instance loss of diversity or a number of maximum of iterations.

The *implicit parallelism*, the lack of constraints on differentiability of functions to optimize, and the fact that a full understanding of the phenomena is not necessary make GAs an appealing learning algorithm for training MLPs.

IV. RELATED WORKS

Pendharkar & Rodger **GR1** used a GA to train an MLP with the aim of predicting bankruptcy in a sample of 100 firms, where half were considered bankrupt and the other half non-bankrupt. They also compared the performance of different crossover operators: GA(A), GA(O) and GA(U) that uses arithmetic, one-point and uniform crossover operators respectively. They also used a MLP with backpropagation (MLPBP) as a reference.

The variables used as input to train the models were: *the relation of earnings before interest to taxes, the relation of earnings before interest and taxes to assets, the relation of current assets to current liabilities, the relation of retained earnings to assets, and the relation of market value of equity to book value debt.* The results of correct classifications on the test data were the same for MLPBP, GA(A), and GA(O). However, GA(U) outperformed the aforementioned models by two percentage points (60%), which could be considered significant in a financial context.

Since the late 1980s, GAs have played an important role in training NNs due to their ability to diversify the search and spend less time reaching the region where the global optimum is located. However, GAs are not as fast as gradient-based methods in finding the exact point **NI***. To leverage the exploration capabilities of metaheuristics and the exploitation capabilities of local search algorithms (such as gradient-based techniques), some studies have used hybridizations of these approaches to enhance the quality of the results. For instance:

García et al. **MMN** uses the Gravitational Search Algorithm and the Chaotic Gravitational Search Algorithm each one integrated with Quasi-Newthon (gradient-based algorithm) as learning algorithms to train a set of MLPs, with the objective of taking advantage of the global and local search ability that offer each methodology, namely using the fast rate of convergence of the gradient method avoiding local optima with the metaheuristics. The results demostrated the superiority of proposed *memetic chaotic gravitational search algorithm* regarding another metaheruistics such as PSO and GAs in the XoR problem, in a classification problem and in the approximation of a continuous function as well as in speed as convergence and quality of solutions.

In the field of bankruptcy Naveen & Rao **memetic** uses an hybridization of Cukoo Search algorithm and Particle Swarm Optimization to predict bankruptcy assessing its performance using three different databases and comparing the results of the proposed methodology with a decision tree. However, the results were inconclusive. While the model showed significantly better performance in one dataset, it demonstrated equal performance in another dataset and lower performance in a third dataset.

V. MEMETIC ALGORITHMS

In general terms, a Memetic Algorithm (MA) can be considered as a combination of two algorithms: an Evolutionary Algorithm (EA) and a Local Search (LS) that examines a neighborhood to determine if there is a better solution **memes1**.

$$EA + LS \subset MA \quad (5)$$

An example of a simple MA implementation considers all the steps mentioned earlier for a GA, but applying a LS in the operations of *Chromosomal crossover* and *mutation* to improve the quality of solutions in each step.

VI. COLOMBIAN BANKRUPTCY STUDIES

In Colombia bankruptcy prediction has been study mainly with traditional statistical techniques, such as MDA, logistic and probit regression; however, there are few studies comparing the performance between traditional statistical and Ml techniques. Nevertheless, is important remark that none of the ML techniques used by the works presented here are Neural Networks.

Most of the works are elaborated with financial statements specifically using financial ratios retrieved from Super Intendencia De Sociedades (in spanish).

Reference	Year	Methodology
C8	2002	MDA
C7	2003	Probit heterocedastic
C11	2009	Cox regression
C14	2013	Logistic regression
C13	2013	• Logistic regression • Probit regression
C1	2017	• Logistic regression • Adtive generalized model • Genralized model of extreme value • Binary generalized model of extreme value
C2	2017	Rough sets
C15	2019	• Logistic regression • Boosting algorithm
C16	2019	Logistic regression and another variants.
C3	2020	• Logistic regression • Boosting algorithm • Support vector machine

TABLE I: Source: Own elaboration

In a short overview of some found articles: the first works is **C8** that tries to discriminate between weak and strong companies using the Altman model, defining in an ambiguous way the classification of "strong-weak" firms and using an MDA model find a specificity and sensitivity of 82% and 100% respectively.

García et al. **C1** uses the legal declaration of bankruptcy or the declaration of started a restructuring process, and using financial ratios from one year before to bankruptcy and estimating a heterocedastic probit model, shows the variables of *profitability ratio and financial indebtedness of the asset* as significant with a specificity and sensitivity of 82.4% and 81.5%.

Peréz et al. **C13** make a benchmark among logit and probit in a final sample of 631 bankrupt and 631 no bankrupt. The results in test validation, indicated that the best performance is obtained with the logit model, with a sensitivity of 77% and a specificity of 88%. In contrast **C16** make a benchmark among different logit models finded that the *mixted logit model* outperform the traditional logistic model, given that also can capture outlier values.

According to Table I the only works that make a benchmark among traditional and not traditional technique are **C3**, **C15**.

Mejía et al. **C3** using information retrieved from Super Intendencia de Sociedades for 58 firms that entered a process of restruncting and 11,754 that did not. Performs a benchmark between logistic regression, boosting algorithm and support vector machine.

Using 10 financial ratios, including *relation operational net working capital to sales, free cash flow, asset turnover, return on assets, return on equity, debt to assets ratio, debt concentration, gross profit margin, operating profit margin, net profit margin* and using accuracy as assessment metric, it was found that boosting(92.1%) outperform logistic(54.1%) and support vector machine (50.0%) to predict bankrupt with a year of anticipation.

Mejía & Castaño **C15** using financial statements retrieved from Super Super Intendencia de Sociedades, to carry out a benchmark among Boosting and Logistic regression, with the financial ratios in 2017 for companies that began a restructuring process in 2018, in this work were used 15 financial ratios for 127 bankrupt companies and 2988 no-bankrupt companies. Using sensitivity it was found that boosting (80%) outperform logist regression (0%), the type two error was 19% for boosting and 100% for logistic regression, the authors claim that, this is given logistic can not handle well imbalance datasets.

The results confirm the superiority of ML techniques to predict bankruptcy; however, there is not yet sufficient comparasion of different ML techniques in the country.

Given the gap between international and Colombian research on bankruptcy prediction, and the attractiveness of metaheuristics as algorithms to train networks (given they do not imposed differentiability on loss function and could optimized the network architecture and weights simultaneously). We proposed to test an evolutionary strategy with a no gradient search local technique, to train a MLP model in Colombian bankruptcy prediction.

Main goal: To compare the prediction performance of a memetic MLP with state-of-the-art methodologies in Colombian bankruptcy.

Specific Objectives
Retrieve and consolidate information to understand data in bankruptcy context.
Design and implementation of the memetic algorithm.
Benchmark among models and validation with another datasets of bankruptcy.

VII. METHODOLOGY

A. *Retrieve and consolidate information to understand data in bankruptcy context*

1) Retrieve Data: The data is public and is available in the official web site of Superintendencia de sociedades, for this work will be retrieved the years 2018 and 2019.

2) Business understanding: Bankruptcy have social and economical cost, both costs could be measure among error type 1 and error type 2, the first measure is the cost derived from classify a bankrupt as no-bankrupt, the other is the cost of classify no-bankrupt as bankrupt, the aim is construct

a model that allow us predict with anticipation bankruptcy to reduce those errors using financial ratios, therefore the financial ratios mus be constructed based on related works and financial theory.

3) Comparation financial ratios among bankrupt and no-bankrupt: The table one is constructed to indicate if there are significative differences among the main financial ratios used in the literature about bankruptcy prediction between bankrupt and no-bankrupt, first is evaluated the normality using shapiro-wilk test if the varaible is determined normal the t-test will be used, otherwise Kruskall-Wallis will be used. significance will be established with $p < 0.05$.

4) Bias selection: to assess if exist bias selection we constructed a comparative table but among groups that fill and not fill inclsusion-exclusion criteria evaluated the normality using shapiro-wilk test if the varible is determined normal the t-test will be used, otherwise Kruskall-Wallis will be used. signficance will be established with $p < 0.05$.

B. Design and implementation of the memetic algorithm

1) Study theorical and empirical properties of local searchs: Understand theorical and empirical properties of some local search techniques to integrate with genetic algorithms to train MLP.

2) Work sessions with professors: Sessions to understand and discuss how integrate the methodologies.

3) Implementation: The implementation will be using Python to solve and test the performance on XOR problem.

C. Benchmark among models and validation with another datasets of bankruptcy

1) Missing values: Using *KKN imputation* in minority class and remove observations majority class.

2) Standardization and outlier: The variables are transformed in terms of standard deviations namely using the following equation:

$$z_i = \frac{x_i - \bar{x}}{\sigma_x} \tag{6}$$

After using *Shevyshev theorem* will be studied possible outliers in variables given this theorem state that *The probability of a observation is to standard deviaton k of mean is:*

$$1 - \frac{1}{k^2} \tag{7}$$

3) Balancing data: The balancing strategy will be *SMOTE+Tomek link* given that consider *under and over sampling*

4) Benchmark models: The models selected to perform the benchmark are; *logistic regression, support vector machine, boosting and Backpropagation MLP*.

5) feature selection: The selection of variables will be performed with *Random Forest-Recursive Feature Elimination algorithm*.

6) Hyperparameter tunning: To perform hyperparameter tunning a $k = 10$ fold cross validation will be implemented.

7) Evaluation: The total data is split in 80% to training and 20% to test, an important point is the test data will be selected in equal proportion to get a more accurate evaluation, the model with the lowest type 1 and 2 error will be selected.

The measures of performance are constructed with the following confusion matrix;

	Bankrupt	No-Bankrupt
Positive prediction	True Positive	False Positive
Negative prediction	False Negative	True Negative

TABLE II: Confusion matrix

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1 = 2\left(\frac{Precision * Recall}{Precision + Recall}\right) \tag{10}$$

$$Type1 = \frac{FP}{FP + TN} \tag{11}$$

$$Type2 = \frac{FN}{TP + FN} \tag{12}$$

VIII. SCHEDULE

The proposed timeline to perform each activity is presented in Table III and is planned to be completed in two semesters.

Week(s)	Activities	Description
[1-5]	Retrieve, preprocessing, merging and understanding data	Retrieve the information from the official site of Super Intendencia de Sociedades, Consolidate data to construct financial ratios, Compare financial ratios among bankrupt and no-bankrupt firms, Determine if there is selection bias in data, Remove missing values, Sandardize variables and determine outliers, Balance data.
[6-16]	Design and implementation of the proposed algorithm	Design of the GA algorithm with local search to train MLP
[17-24]	Experiments and benchmark	Feature selection, Hyperparameter tuning , Benchmark with: logistic regression, boosting, supportvector machine and backpropagation multilayer perceptron.
[25-28]	Validation	Validation of the algorithm with another bankruptcy datasets.
[28-30]	Write results and discussion	Write results and discussion in the official document.
[30-36]	Submit the article to a journal ranked in Scimago	Search and submit the manuscript to a journal

TABLE III: Schedule