The application of cat swarm optimisation algorithm in classifying small loan performance

Eka N Kencana¹, Nyoman Kiswanti², and Kartika Sari²
¹ Tourism Research Consortium, University of Udayana – Denpasar 80234, Indonesia
² Department of Mathematics, Udayana University – Jimbaran, Badung 80361 - Indonesia
E-mail: i.putu.enk@unud.ac.id, mangkisz@gmail.com, sari_kartika@yahoo.co.id

Abstract. It is common for banking system to analyse the feasibility of credit application before its approval. Although this process has been carefully done, there is no warranty that all credits will be repaid smoothly. This study aimed to know the accuracy of Cat Swarm Optimisation (CSO) algorithm in classifying small loans’ performance that is approved by Bank Rakyat Indonesia (BRI), one of several public banks in Indonesia. Data collected from 200 lenders were used in this work. The data matrix consists of 9 independent variables that represent profile of the credit, and one categorical dependent variable reflects credit’s performance. Prior to the analyses, data was divided into two data subset with equal size. Ordinal logistic regression (OLR) procedure is applied for the first subset and gave 3 out of 9 independent variables i.e. the amount of credit, credit’s period, and income per month of lender proved significantly affect credit performance. By using significantly estimated parameters from OLR procedure as the initial values for observations at the second subset, CSO procedure started. This procedure gave 76 percent of classification accuracy of credit performance, slightly better compared to 64 percent resulted from OLR procedure.

1. Introduction

Extracting valuable information such as recognising pattern from hidden data is one aim of data mining technique. This technique belongs to multidisciplinary field that intersects statistical methods, machine learning, pattern recognition, and data visualisation. One purpose in data mining technique is classifying objects based on their attribute(s). The classification problem can be viewed as a problem to classify objects into a set of clusters without any priori knowledge [1].

To classify an object into its appropriate group can be done by several methods. Among these methods, K-means algorithm is very popular. Despite of its popularity, refers to [2], the major drawback of this algorithm is frequently found it gets stuck at local minima and classification results is heavily depended on the choice of initial cluster center. To overcome this limitation, some alternative techniques in classifying objects were developed. Metaheuristic algorithms such as genetic algorithm as well as particle swarm optimisation had been operationalised to classify objects regarding their attribute(s) [3][4].

This work implements Cat Swarm Optimisation (CSO) algorithm in classifying small loans performance were approved by Bank Rakyat Indonesia, one of several public bank in Indonesia. CSO is belonged to swarm intelligence method that models cats behavior in solving optimisation problem [5]. Researchers had been applied this technique to solve various problem dimensions,
but to our best knowledge, CSO is rarely used to classify financial data. In addition, to elaborate its performance, we used Ordinal Logistic Regression (OLR) to predicts loan’s classification from the loan’s attributes.

Chu and Tsai [5] propose CSO as a new metaheuristic technique in optimisation problems by observing the behavior of cats. Basically, CSO algorithm consists of two states of cats, namely (a) seeking mode, represents cats during a resting period but being alert, and (b) tracing mode, represents mode of cats while tracing their target. According to [5], cats which are awake spend most of their time resting and spend very little time to chase target. The proportion on cats in tracing mode is defined as mixture ratio (MR).

2. Research Method
The small loan data from 200 lenders for this work were acquired from Bank Rakyat Indonesia, Melati branch office at Denpasar. The data consists of 9 independent variables represent credit’s profile and one categorical dependent variable represents the performance of credit. The operational definition of the variables is listed on Table 1.

Prior to analysis stage, the data is randomly divided into two subset with equal size. In order to make fair comparison, before OLR and CSO algorithm were applied, we check the proportion of loans are classified as smoothly paid, less smoothly, and doubtful in each of data subset. The maximum proportion differences is allowed for same classification between subset is five percent. The first subset is analysed with OLR technique to identify influential variables in determining performance of credit, and the second subset is analysed using CSO algorithm with number of dimensions to move is determined by the number of influential variables that is identified with the OLR. The last stage is to compare the classification accuracy is resulted from both techniques. Classification accuracy of both methods were counted through the Apparent Correct Classification Rate (ACCR) measure [6], which is stated by equation

$$ACCR = \frac{\sum_{i=1}^{c} n_i}{N}$$

In equation (1), N represents total observation numbers, \(n_i\) is the number of observations were correctly classified in class \(i\), and \(c\) represents number of class.

3. Results and Discussion
3.1. OLR Analysis
OLR is an extension of binary logistic regression (BLR). OLR as well as BLR utilised maximum likelihood method to estimate the parameters of regression model [5]. Using STATA to analyse the first subset data, we obtained three out of nine predictor variables significantly affect small loan performance. The OLR’s result is listed on Table 2. The Nagelkerke pseudo \(R^2\) showed the variance of loan’s performance can be explained by its predictors as much as 61.5 per cent while the rest is unexplained.

The number of small loan performance were predicted correctly by using this OLR model as smoothly, less smoothly, and doubtful credits are 34, 24, and 6 observation respectively. By applying eq. (1), the classification accuracy is resulted from OLR method as much as 64 per cent.

3.2. CSO Analysis
Arbitrary, we create 30 cats in our work. In initialisation stage, these cats were randomly assigned between tracing and seeking mode according to MR value. As aforementioned explain, in seeking mode where cats are in resting period, four factors have to be considered and defined as follows [6]:

[2]
Table 1. List of Small Loan Variables

| Code | Description          | Measurement | Values                          |
|------|----------------------|-------------|---------------------------------|
| PER  | Loan performance     | Categorical | 1: Smoothly paid                |
|      |                      |             | 2: Less smoothly                |
|      |                      |             | 3: Doubtful                     |
| X1   | Amount of loan       | Continuous  | -                               |
| X2   | Business type        | Categorical | 1: Services                     |
|      |                      |             | 2: Trading                      |
| X3   | Loan’s period        | Continuous  | -                               |
| X4   | Lender’s education   | Categorical | 1: Elementary                   |
|      |                      |             | 2: Junior school                |
|      |                      |             | 3: High school                  |
|      |                      |             | 4: Undergraduate                |
| X5   | Income per month     | Continuous  | -                               |
| X6   | Business’ duration   | Continuous  | -                               |
| X7   | Number of children   | Continuous  | -                               |
| X8   | Home ownership       | Categorical | 1: Owned                        |
|      |                      |             | 2: Family                       |
|      |                      |             | 3: Rent                         |
| X9   | Lender’s age         | Continuous  | -                               |

Table 2. The OLR’s Estimators and Their Significances

| Code | Estimate | Wald’s Statistic | p-Values |
|------|----------|------------------|----------|
|      | Constant 1 | -1.901        | 0.505    | 0.478    |
|      | Constant 2 | 0.572          | 0.046    | 0.830    |
| X1   | 0.972     | 6.874           | 0.009    |
| X2   | -0.406    | 0.441           | 0.507    |
| X3   | 0.025     | 0.872           | 0.351    |
| X4   | 0.035     | 0.020           | 0.887    |
| X5   | -5.743    | 10.955          | 0.001    |
| X6   | -0.589    | 16.171          | 0.000    |
| X7   | 0.196     | 1.299           | 0.254    |
| X8   | 0.044     | 0.038           | 0.845    |
| X9   | 0.001     | 0.000           | 0.985    |

(i) Seeking Memory Pool (SMP), represents the seek’s memory size of each cat, and denotes the number of neighboring positions are considered by a cat;
(ii) Count of Dimension to Change (CDC), represents number of dimension will be mutated;
(iii) Seeking Range (SR), represents mutative ratios for the selected dimension; and
(iv) Self-position Consideration (SPC), is a Boolean variable which its value indicates the current position of cat is considered as a new neighbor position. If SPC = 1 then current position is viewed as a new neighbor position, otherwise cat should consider a different position.
The computational steps of cat in seeking mode can be listed as follows [5][6]:

(i) Create \( n \) copies of current position of cat\(_i\), \( n = \text{SMP} \). If SPC = 1 then \( n = n - 1 \);
(ii) For each of dimension according to CDC’s value, replace the SR’s old value for each cat with randomly plus or minus \( x \) percent of the old value;
(iii) Randomly select point to move to from the candidates points and replace the position of cat\(_i\).

Once cat\(_i\) goes into tracing mode, its steps can be listed as follows:

(i) Update its velocities for every dimension to consider (CDC) according to equation (2)

\[
v_{i,d} = v_{i,d} + r_1 c_1 (x_{\text{best},d} - x_{i,d}) \tag{2}
\]

where \( x_{\text{best},d} \) and \( x_{i,d} \) are the position of cat with best fitness value and position of cat\(_i\) on dimension \( d \), respectively; \( c_1 \) is an arbitrary constant and \( r_1 \) is a random value in the range [0, 1].

(ii) Update the position of cat\(_i\) at dimension \( d \) \( (x_{i,d}) \) according to equation (3)

\[
x_{i,d} = x_{i,d} + v_{i,d} \tag{3}
\]

3.3. CSO Result

To make fair comparison with OLR technique, we position each cat in five dimensional space \( (R^5) \). These dimensions represents two constants and three significant variables according to OLR analysis. The values from OLS analysis were applied as initial positions in each dimensions of cat. In applying CSO algorithm to classify the loan performance, we set the minimum, increment, and maximum values for CSO’s parameters as follows:

| Parameters | Value(Min; Increment; Max) |
|------------|-----------------------------|
| MR         | (0; 0.03; 1)                |
| SR         | [0; 0.10; 1]                |
| CDC        | [0; 0.20; 1]                |
| SMP        | [1; 1; 10]                  |

After 2000 iterations, for 30 cats and using \( c_1 = 2 \) as suggested by [6] we found the best parameter for MR, SR, SMP, and CDC values in our work are 0.20, 0.50, 3, and 0.20 respectively. Refers to Table 3, the classification accuracy result from CSO method as much as 76 percent, slightly greater than 64 percent was obtained from OLR technique.

4. Conclusion

The application of CSO algorithm in classifying small loan performance by utilizing the significantly estimated parameters from OLR technique proved can increase the classification accuracy. This finding is inline with similar experiment result conducted by [7] in classifying gene expression data. Kumar and Mishra utilised CSO method to update the gene’s weight values resulted from Principal Component Analysis (PCA) and Factor Analysis (FA) for high dimensionality of the datasets.

As a metaheuristic algorithm, CSO and other swarm optimisation techniques were known as an efficient and effective global search technique. However, CSO has never been investigated its
Table 3. Classification Accuracy

| Actual Group | Predicted Group | Total |
|--------------|-----------------|-------|
|              | Smoothly | Less Smoothly | Doubtful |
| Smoothly     | 36      | 1             | 0        | 37 |
| Less Smoothly| 11      | 24            | 3        | 38 |
| Doubtful     | 1       | 8             | 16       | 25 |
| **Total**    | 48      | 33            | 19       | 100 |

performance according to the characteristics of object under studied. We showed by eliminating
unsignificant variables in small loan classification process, the classification accuracy increase.

Despite of its performance to classify small loan data is better than OLR method, CSO as
well as other particle swarm optimisation techniques, its theoretical background is rather weak.
Refers to [8], PSO algorithm has a limited condition for the particle(s) move stable from the
point of math’s.

References
[1] Liu Y and Shen Y 2010 Data clustering with cat swarm optimization J. Convergence Information Technology 5 (8) pp 21-28.
[2] Selim S Z and Ismail M A 1984 K-means-type algorithm: generalized convergence theorem and characterization
of local optimality IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (1) pp.81-87.
[3] Laszlo M and Mukherjee S 2006 A genetic algorithm using hyper-quadtrees for low-dimensional k-means
clustering IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (4) pp. 533-543.
[4] Kao Y T, Zahara E and Kao I W 2008 A hybridized approach to data clustering Expert Systems with
Applications 34 (3) pp. 1754-62.
[5] Chu S C and Tsai P W 2007 Computational intelligence based on the behavior of cats International Journal
of Innovative Computing, Information and Control 3 (1) pp. 163-173.
[6] Hosmer D and Lemeshow S 2000 Applied Logistic Regression 2nd ed (New York, USA: John Wiley & Sons Inc.).
[7] Kumar A and Mishra D 2013 Cat swarm based optimization of gene expression data classification international
Journal of Computer Trends and Technology 4 (5) pp. 1185-90.
[8] Clerc M and Kennedy J 2002 The particle swarm - explosion, stability, and convergence in a multidimensional
complex space IEEE Transactions on Evolutionary Computation 6 (1) pp. 58-73.