STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **7**, June 2019, pp 520–529.
Published online in International Academic Press (www.IAPress.org)

Variable Selection in Count Data Regression Model based on Firefly Algorithm

Zakariya Yahya Algamal

Department of Statistics and Informatics, University of Mosul, Mosul, Iraq

Abstract Variable selection is a very helpful procedure for improving computational speed and prediction accuracy by identifying the most important variables that related to the response variable. Count data regression modeling has received much attention in several science fields in which the Poisson and negative binomial regression models are the most basic models. Firefly algorithm is one of the recently efficient proposed nature-inspired algorithms that can efficiently be employed for variable selection. In this work, firefly algorithm is proposed to perform variable selection for count data regression models. Extensive simulation studies and two real data applications are conducted to evaluate the performance of the proposed method in terms of prediction accuracy and variable selection criteria. Further, its performance is compared with other methods. The results proved the efficiency of our proposed methods and it outperforms other popular methods.

Keywords Variable selection; count data; Poisson regression; negative binomial regression; firefly algorithm.

AMS 2010 subject classifications 62J12, 62J07

DOI: 10.19139/soic.v7i2.566

1. Introduction

In regression modeling, data in the form of counts are usually common. Count data regression modeling has received much attention in medicine, behavioral sciences, psychology, and econometrics [1, 2, 3, 36]. The Poisson and negative binomial regression models are the most basic models under count data regression models [4](Wang et al. 2014). The problem of overdispersion usually occurs in count data. Unlike Poisson regression model, negative binomial regression can handle the overdispersion issue [5, 6, 31].

In many real applications, recent developments in technologies have made the possibility to measure a large number of variables. In the regression modeling, the existence of huge number has a negative effect by overfitting the regression model. Therefore, identification of a small subset of important variables from a large number of variables set for accurate prediction is an important role for building predictive regression models [7, 35].

Recently, the naturally inspired algorithms, such as genetic algorithm, particle swarm optimization algorithm, firefly algorithm, and crow search algorithm, have a great attraction and proved their efficiency as variable selection methods [12]. This is because that the main target in variable selection is to minimize the number of selected variables while maintaining the maximum accuracy of prediction, and, therefore, they can be considered as optimization problems [13].

Several researchers have employed the naturally inspired algorithms for variable selection in regression models. [14] employed the genetic algorithm for variable selection in linear and partial least squares regression models, with application in chemometrics. Drezner et al. [15] proposed to use tabu search algorithm in model selection in

*Correspondence to: Zakariya Yahya Algamal (Email: zakariya.algamal@uomosul.edu.iq). Department of Statistics and Informatics, University of Mosul, Mosul, Iraq

ISSN 2310-5070 (online) ISSN 2311-004X (print)

the linear regression model. On the other hand, a hybrid algorithm of genetic algorithm and simulated annealing was proposed as a subset selection method in linear regression model by [16]. [17] did a comparison of simulated annealing algorithms for variable selection in principal component analysis and discriminant analysis. Besides, the di?erential evolution algorithm was used as a variable selection in linear regression model by [18]. In generalized linear models, the natural inspired algorithms for variable selection are also used, such as, logistic regression model [19, 20], Poisson regression model [21, 22, 32], and gamma regression model [23].

The purpose of this paper is to propose firefly algorithm, which is a swarm intelligence technique, as an alternative variable selection method for use in count data regression model. The proposed algorithm will efficiently help in identifying the most relevant variables in the count data regression model with a high prediction". The superiority of the proposed algorithm is proved though different simulation settings and a real data application.

2. Count data regression model

Consider that "we have a data set $\{(y_i, \mathbf{x}_i)\}_{i=1}^n$ where $y_i \in \mathrm{R}$ is a response variable and $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ip}) \in R^p$ is a $p \times 1$ known explanatory variable vector.Assume that y_i is counts data and it has a Poisson distribution with probability density function

$$f\left(y_i, \theta\right) = \frac{e^{-\theta_i} {\theta_i}^{y_i}}{y_i!}, \quad y_i = 0, 1, \ldots, \tag{1}$$

where $\theta > 0$ is the parameter of the Poisson distribution. In Poisson regression model (PRM), represents the conditional mean as $\theta_i = \exp({\mathbf{x}_i}^T \beta(p+1) \times 1)$, where $\beta = (\beta_0, \beta_1, ..., \beta_p)$ is a vector of unknown regression coefficients. The PRM can be defined as

$$y_i = \exp({\mathbf{x}_i}^T \beta). \tag{2}$$

The log-likelihood function of Eq. (2) is defined as

$$\ell(\beta_{\mathrm{PRM}}) = \sum_{i=1}^{n} \left\{ y_i {\mathbf{x}_i}^T \beta_{\mathrm{PRM}} - \exp({\mathbf{x}_i}^T \beta_{\mathrm{PRM}}) - \ln y_i! \right\}. \tag{3}$$

The maximum likelihood estimation of the PRM is obtained by taking the first derivative of Eq. (3) and solving it as

$$\frac{\partial \ell(\beta_{\mathrm{PRM}})}{\partial \beta_{\mathrm{PRM}}} = \sum_{i=1}^{n} \left[y_i - \exp(\mathbf{x}_i^T \beta_{\mathrm{PRM}}) \right] \mathbf{x}_i = 0. \tag{4}$$

The iteratively weighted least squares algorithm can be used to obtain the maximum likelihood estimators (MLE) of the PRM as

$$\hat{\beta}_{\mathrm{PRM}} = (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{W}} \hat{\mathbf{u}}, \tag{5}$$

where $\hat{\mathbf{W}} = \mathrm{diag}(\hat{\theta}_i)$ and $\hat{\mathbf{u}}$ is a vector where ith element equals to $\hat{u}_i = \log(\hat{\theta}_i) + ((y_i - \hat{\theta}_i)/\hat{\theta}_i)$.

One of the most important assumptions in PRM is that the mean and variance of the response variable are equivalent. When this assumption is violated, then the PRM suffers from the overdispersion issue. In real applications, the conditional variance can exceed the conditional mean, and, therefore, the negative binomial regression model (NBRM) be more appropriate than a PRM for modeling count data [24, 25, 33, 34]. In NBRM, there is a random variation, α_i, in the Poisson conditional mean as $\alpha_i = z_i \theta_i$, where z_i is a random variable having gamma distribution such that $z_i \sim \Gamma(\lambda, \lambda)$.

Assume that y_i is counts data and it has a negative binomial distribution with probability density function

$$f\left(y_i\right) = \frac{\Gamma(y_i + \tau^{-1})}{\Gamma(y_i + 1)\Gamma(\tau^{-1})} \left(\frac{\tau^{-1}}{\tau^{-1} + \theta_i} \right)^{\tau^{-1}} \left(\frac{\theta_i}{\tau^{-1} + \theta_i} \right)^{y_i}, \tag{6}$$

where $\tau \geq 0$ is the overdispersion parameter which is defined as $\tau = \lambda^{-1}$. The estimation of NBRM coefficients is usually estimated by the ML estimator which is obtained by maximizing the log-likelihood function

$$\ell(\beta_{\mathrm{NBRM}}) = \sum_{i=1}^{n} \left\{ \begin{array}{l} \left(\sum_{j=0}^{y_i-1} \log(j+\tau^{-1}) \right) - \ln y_i! - (y_i+\tau^{-1}) \ln(1+\tau \exp({\mathbf{x}_i}^T \beta_{\mathrm{NBRM}})) \\ +y_i \ln(\tau) + y_i {\mathbf{x}_i}^T \beta_{\mathrm{NBRM}} \end{array} \right\}. \tag{7}$$

Then "the ML estimator can be obtained by solving Eq. (7) as

$$\frac{\partial \ell(\beta_{\mathrm{NBRM}})}{\partial \beta_{\mathrm{NBRM}}} = \sum_{i=1}^{n} \left[\frac{y_i - \exp(\mathbf{x}_i^T \beta_{\mathrm{NBRM}})}{1+\tau \exp(\mathbf{x}_i^T \beta_{\mathrm{NBRM}})} \right] \mathbf{x}_i = 0, \tag{8}$$

and $\hat{\beta}_{\mathrm{NBRM}}$ is

$$\hat{\beta}_{\mathrm{NBRM}} = (\mathbf{X}^T \hat{\mathbf{Q}} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{Q}} \hat{\mathbf{v}}, \tag{9}$$

where $\hat{\mathbf{Q}} = \mathrm{diag}(\hat{\theta}_i)$ and $\hat{\mathbf{v}}$ is a vector where ith element equals to $\hat{v}_i = \ln(\hat{\theta}_i) + ((y_i - \hat{\theta}_i)/\hat{\theta}_i)$.

3. Firefly algorithm

In recent years, "numerous nature-inspired algorithms have been proposed as powerful approaches to solve the continuous optimization problems. Minimizing the number of variables with maximizing the accuracy of prediction is an optimization problem [27]. Firefly optimization algorithm (FA) is one of the recently efficient proposed nature-inspired algorithms, which is firstly introduced by [26]. The application of FA is an easy algorithm for solving the optimization problems compared with other algorithms. FA is inspired by the social behavior of fire?ies through flashing lights. FA enables a swarm of fireflies with low light intensities to move towards the neighbor brighter fireflies possessing superior search abilities in solving optimization problems. Three rules are held in FA [27]. The first rule is that all fireflies are unisex meaning that one firefly will be attracted to other fireflies regardless of their sex. The second rule is that the degree of the attractiveness of a firefly is proportion to its brightness, therefore for any two flashing fireflies, the less bright one will move towards the brighter one and the more brightness. If there is no brighter one than a particular firefly, it will move randomly. The third rule is that the brightness of a firefly is somehow related to the analytical form of the fitness function. For a maximization problem, the brightness of each firefly is proportional to the value of the cost function. Let d represents the dimension of the object function that will optimized, $n_f i$ represents the number of fireflies, δ refers the light absorption coefficient, I_i is the light intensity, and r is the distance between any two firefly locations $(s_i j)$ and $(s_j r(s_i, s_j) = \sqrt{\sum_{c=1} {I_i}^d (s_{i,c} - s_{j,c})^2}.)$. This Cartesian distance can be defined as

$$r(s_i, s_j) = \sqrt{\sum_{c=1}^{d} (s_{i,c} - s_{j,c})^2}. \tag{10}$$

Because I_i decreases when the distance from the source increases, the variations of should be monotonically decreasing function. As a result, in most applications, the I_i can be approximated as

$$I(r) = I_0 \, e^{-\delta r^2}, \tag{11}$$

where I_0 is the original light intensity. Because the attractiveness of a firefly is proportional to the I_i, the attractiveness φ of a firefly is defined as

$$\varphi(r) = \varphi_0 \, e^{-\delta r^2}, \tag{12}$$

where φ_0 represents the attractiveness at $r = 0$.

FA originally is proposed to solve continuous optimization problems. However, in variable selection, the optimization problem is discrete. A binary firefly algorithm (BFA) is proposed by [28] to deal with the problem of variable selection where the position is binary. Because variable selection problem is to select a specific variable or not, thus the solution is expressed as a binary vector, where the value 1 indicates a variable to be selected and 0 otherwise.

Accordingly, the position of a firefly will be replaced as follow:

$$s_i^{(t+1k_2)} = \begin{cases} 1 & \text{if sigm} \geq k_2 \\ 0 & \text{otherwise,} \end{cases} \tag{13}$$

where k_2 represents a random number generated from uniform distribution with [0, 1]. The pseudo code of the BFA is given in Figure 1.

Consequently, our proposed algorithm setting is as follows:

Step 1: The number of fireflies is $n_f = 40$, $\varphi_0 = 1$, $\delta = 0.2$, $\alpha = 0.1$, and the maximum number of iterations is $t_{\max} = 500$.

Step 2: The positions of each firefly are randomly generated from uniform distribution with 0 and 1. The representation of the positions of a firefly is depicted in Figure 2.

Step 3: The fitness function is defined as

$$\text{fitness} = \min \left[\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \right]. \tag{14}$$

Step 4: The positions of the fireflies are updated using Eq. (13).

Step 5: Steps 3 and 4 are repeated until a $t_{\max}$ is reached".

Input: n_f (number of fireflies);
$t_{\max}$ (number of iterations);
δ (light absorption coefficient);
α (randomization parameter)
Output: Optimal selected variables (firefly position).
Initialize randomly a population of n_f positions.
Find the best solution based on fitness
while Stopping criteria not met do
  for each firefly i do
  for each firefly j do
  if firefly j is brighter than firefly i then
    Move firefly i towards firefly j using Eq. (15).
  else
  end
  Evaluate the positions of each firefly.
end

Figure 1. Pseudo code of the Firefly algorithm.

x_1	x_2	x_3	.....	x_{p-1}	x_p
1	0	0	.....	1	0

Figure 2. The representation of the firefly position.

4. Computational results

In this section, "the performance of our proposed variable selection method, FA-BN is tested. Further, the performance of FA-BN is compared with the Akaike information criteria (AIC), corrected Akaike information criteria (CAIC), and Bayesian information criteria (BIC) that are defined as, respectively,

$$\text{AIC} = -2\ell(\hat{\beta}) + 2 \times q, \tag{15}$$

$$\text{CAIC} = -2\ell(\hat{\beta}) + \frac{2q\,(q+1)}{n-q-1}, \tag{16}$$

$$\text{BIC} = 2\ell(\hat{\beta}) + \log(n) \times q, \tag{17}$$

where $\ell(\hat{\beta})$ is the log-likelihood for either PRM or NBRM, and q is the number of selected variables".

4.1. simulation results

In this section, "the same simulation settings of [29] and [4] are used. Each simulation setting is considered for PRM and NBRM. The sample size is considered with $n \in \{50, 100, 200\}$.

Simulation 1: In this simulation, 20 explanatory variables are generated from multivariate normal distributions with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$ which elements $\rho(x_i, x_j) = \rho^{|i-j|}$ with $\rho = 0.5$. The true vector of parameters is given by $\beta = (0.5, -0.5, 0.5, -0.6, 0.5, \underbrace{0, ..., 0}_{5}, 0.5, -0.5, 0.5, -0.6, 0.5, \underbrace{0, ..., 0}_{5})^T$ with 10 true explanatory variables and the rest in non-true variables.

Simulation 2: Here, The true vector of parameters is given by $\beta = (1.25, -0.95, 0.90, -1.10, 0.60, \underbrace{0, ..., 0}_{15})^T$ with 5 true explanatory variables and 15 non-true variables. The explanatory variables are generated as same as simulation 1 with $\rho(x_i, x_j) = 0.5$.

Simulation 3: In this simulation, 8 explanatory variables are generated as same as simulation 1 with $\rho(x_i, x_j) = 0.5^{|i-j|}$. The true parameter vector is given by $\beta = (\underbrace{0.17, ..., 0.17}_{8})^T$.

For all the simulation examples 1-3, the response variable is generated according to PRM as $y_i \sim Po(\exp(\mathbf{x}_i{}^T\beta))$. Simulations 4-6 are the same as the setting of simulations 1 C 3 where the response variable is generated according to NBRM with conditional mean $\exp(\mathbf{x}_i{}^T\beta)$ and $\tau = 2$. For performance evaluation of the FA-BN, the mean squared error (MSE) is used as a prediction accuracy criteria, which is defined as $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2/n$.

In terms of variable selection performance, the number of the truly nonzero coefficients which are incorrectly set to zero (I), and the number of the true zero coefficients which are correctly set to zero (C). The higher the values of C, and the lower the values of I, the better the variable selection performance is. All computations of this paper were conducted using R. Based on 500 times of repeating simulation, the averaged MSE, I, and C with their associated standard deviations (the number in parentheses) are listed in Tables 1-6, respectively, for PRM and NBRM.

It shows from these tables that the FA-BN method there has a significant improvement where it has a much better average of MSE than those AIC, CAIC, and BIC methods. For instance, in Table 1 when $n = 50$, the MSE

reduction by FA-BN was about 56.91%, 49.07%, and 44.43% comparing with AIC, CAIC, and BIC, respectively. Further, regardless of the value of n, the FA-BN often shows the smallest MSE among the competitor methods.

In terms of variable selection performance, our proposed method obviously selects a very few irrelevant variables comparing with AIC, CAIC, and BIC, where the number of the true zero coefficients which are correctly set to zero is high comparing with others. For example, in Table 4 when $n = 200$, FA-BN does not select, on average, about 8 irrelevant variables out of 10 irrelevant variables. While AIC, CAIC, and BIC select more than 4 irrelevant variables. On the other hand, FA-BN performs very well with the smallest I (the number of the truly nonzero coefficients which are incorrectly set to zero) among all the used methods. This indicates that FA-BN misses a very few important variables. In Table 4 when $n = 50$, FA-BN does not select on average one important variable out of 10 important variables. In the same case, AIC, CAIC, and BIC select on average more than 6 important variables.

From the results of simulation 3 (Table 3) and simulation 6 (Table 6), the model is dense, and, therefore, all the methods have zero values for the criterion C. On the other hand, FA-BN is the best because the number of nonzero variables that have been identified as irrelevant variables is smaller compared with AIC, CAIC, and BIC. It is worth noting that AIC has inferior performance in all simulation examples comparing with CAIC, BIC, and FA-BN methods.

In summary, it is obvious that the simulation results for both PRM and NBRM demonstrated the use of FA-BN in variable selection. Another important point that is concluded from the simulation results is that the variable selection performance of the FA-BN is not changed by changing the sample size".

Table 1. Simulation 1 results, on average, for PRM

Methods	MSE	C	I
	$n = 50$		
FA-BN	3.301 (0.012)	8.421 (0.011)	0.664 (0.009)
AIC	7.661 (0.035)	5.332 (0.017)	4.141 (0.018)
CAIC	6.482 (0.022)	5.071 (0.018)	3.872 (0.018)
BIC	5.941 (0.019)	6.874 (0.017)	3.096 (0.012)
	$n = 100$		
FA-BN	3.114 (0.011)	8.545 (0.013)	0.720 (0.010)
AIC	7.374 (0.031)	5.456 (0.022)	4.197 (0.025)
CAIC	6.285 (0.019)	5.195 (0.019)	3.928 (0.022)
BIC	5.757 (0.019)	6.998 (0.019)	3.152 (0.012)
	$n = 200$		
FA-BN	3.065 (0.013)	8.582 (0.013)	1.295 (0.012)
AIC	7.425 (0.033)	5.493 (0.019)	4.172 (0.023)
CAIC	6.246 (0.021)	5.232 (0.019)	4.503 (0.021)
BIC	5.705 (0.023)	7.035 (0.016)	2.912 (0.017)

4.2. real application results

In this section, "two real applications are considered. The first real application related to the PRM, while the second real application related to the NBRM.

For the first real application, the number of publications produced by Ph.D. biochemists of [30] is considered where the response variable is the number of articles in last three years of Ph.D. Five explanatory variables were used. They are: the gender (x_1), the marital status (x_2), the number of children under age six (x_3), prestige of Ph.D. program (x_4), and the number of articles by the mentor in last three years (x_5). In this application, the response variable is following Poisson distribution. Depending on the PRM analysis, four explanatory variables, x_1, x_2, x_3, and x_5, are significantly related to the response variables with a level of significant 0.05.

In the second real application, we considered the nuts dataset [24]. The dataset contains 52 observation and 7 explanatory variables. The nuts dataset concerning with the squirrel behavior and several features of the forest across different plots in Scotland's Abernathy Forest. The response variables, which is the number of cones stripped

Table 2. Simulation 2 results, on average, for PRM

Methods	MSE	C	I
	$n = 50$		
FA-BN	4.815 (0.021)	13.217 (0.018)	1.201 (0.019)
AIC	9.175 (0.028)	7.907 (0.024)	3.507 (0.024)
CAIC	7.996 (0.024)	7.122 (0.023)	3.118 (0.023)
BIC	7.455 (0.022)	9.493 (0.022)	2.749 (0.023)
	$n = 100$		
FA-BN	4.541 (0.019)	13.281 (0.017)	1.233 (0.018)
AIC	8.801 (0.025)	7.971 (0.024)	3.539 (0.025)
CAIC	7.712 (0.023)	7.186 (0.022)	3.150 (0.023)
BIC	7.184 (0.020)	9.557 (0.021)	2.781 (0.019)
	$n = 200$		
FA-BN	4.466 (0.019)	13.292 (0.018)	1.241 (0.017)
AIC	8.826 (0.022)	7.982 (0.024)	3.547 (0.023)
CAIC	7.647 (0.020)	7.197 (0.023)	3.158 (0.022)
BIC	7.106 (0.019)	9.568 (0.023)	2.789 (0.017)

Table 3. Simulation 3 results, on average, for PRM

Methods	MSE	C	I
	$n = 50$		
FA-BN	2.505 (0.011)	0	0.357 (0.013)
AIC	6.865 (0.025)	0	2.622 (0.022)
CAIC	5.686 (0.021)	0	2.483 (0.021)
BIC	5.145 (0.015)	0	1.071 (0.021)
	$n = 100$		
FA-BN	2.231 (0.012)	0	0.343 (0.015)
AIC	6.491 (0.021)	0	2.566 (0.025)
CAIC	5.402 (0.017)	0	2.358 (0.024)
BIC	4.874 (0.015)	0	0.977 (0.018)
	$n = 200$		
FA-BN	2.156 (0.011)	0	0.317 (0.016)
AIC	6.516 (0.024)	0	2.538 (0.021)
CAIC	5.337 (0.020)	0	2.209 (0.021)
BIC	4.796 (0.014)	0	0.914 (0.019)

by squirrels, is following the negative binomial distribution, and, thus the NBRM is been more suitable regression model. The explanatory variables are: the number of trees per plot (x_1), the number of DBH per plot (x_2), mean tree height per plot (x_3), canopy closure (as a percentage) (x_4), standardized number of trees per plot (x_5), standardized mean tree height per plot (x_6), standardized canopy closure (as a percentage) (x_7). Depending on the NBRM analysis, five explanatory variables, x_1, x_2, x_3, x_5, and x_6, are significantly related to the response variables with a level of significant 0.05. Tables 7 and 8 summarize the MSE and the selected variables for each used method for both real data applications, respectively.

As seen from the result of Tables 7 and 8, FA-BN can remarkably reduce the MSE comparing with AIC, CAIC, and BIC. In terms of selected variables, on the other hand, it clearly seen from Table 7 that FA-BN only select 4 variables out of 5 variables when PRM is assumed. FA-BN selected the explanatory variables x_1, x_2, x_3, and x_5. These selected variables are identified as relevant variables to the study. Comparing with AIC and BIC, FA-BN includes extra variable but the MSE is less than them. Further, AIC, CAIC, and BIC selected one irrelevant

Table 4. Simulation 4 results, on average, for NBRM

Methods	MSE	C	I
	$n = 50$		
FA-BN	3.868 (0.011)	8.042 (0.015)	0.981 (0.016)
AIC	8.228 (0.021)	4.953 (0.022)	4.458 (0.023)
CAIC	7.049 (0.019)	4.692 (0.021)	4.189 (0.021)
BIC	6.508 (0.015)	6.495 (0.021)	3.413 (0.019)
	$n = 100$		
FA-BN	3.681 (0.013)	8.166 (0.015)	1.037 (0.014)
AIC	7.941 (0.020)	5.077 (0.023)	4.514 (0.021)
CAIC	6.852 (0.021)	4.816 (0.022)	4.245 (0.019)
BIC	6.324 (0.017)	6.619 (0.022)	3.469 (0.019)
	$n = 200$		
FA-BN	3.632 (0.014)	8.203 (0.014)	1.612 (0.013)
AIC	7.992 (0.022)	5.114 (0.021)	4.489 (0.019)
CAIC	6.813 (0.019)	4.853 (0.019)	4.821 (0.019)
BIC	6.272 (0.017)	6.656 (0.017)	3.229 (0.016)

Table 5. Simulation 5 results, on average, for NBRM

Methods	MSE	C	I
	$n = 50$		
FA-BN	5.378 (0.014)	12.836 (0.015)	1.516 (0.017)
AIC	9.738 (0.022)	7.526 (0.023)	3.822 (0.025)
CAIC	8.559 (0.021)	6.741 (0.021)	3.433 (0.022)
BIC	8.018 (0.019)	9.112 (0.019)	3.064 (0.017)
	$n = 100$		
FA-BN	5.104 (0.013)	12.901 (0.015)	1.548 (0.016)
AIC	9.364 (0.021)	7.592 (0.024)	3.854 (0.023)
CAIC	8.275 (0.021)	6.805 (0.022)	3.465 (0.021)
BIC	7.747 (0.017)	9.176 (0.019)	3.096 (0.018)
	$n = 200$		
FA-BN	5.029 (0.014)	12.911 (0.014)	1.556 (0.016)
AIC	9.389 (0.022)	7.601 (0.024)	3.862 (0.022)
CAIC	8.212 (0.019)	6.816 (0.022)	3.473 (0.022)
BIC	7.669 (0.019)	9.187 (0.021)	3.104 (0.019)

variables (x_4), indicating the possibility of these methods to select unimportant variables. Regarding Table 8, FA-BN has similar results in terms of selected variables. FA-BN only select 4 variables out of 7 variables when NBRM is assumed. FA-BN selected the explanatory variables x_1, , x_5, and x_6. These selected variables are identified as relevant variables to the study. Each of AIC, CAIC, and BIC, on the other hand, shows the ability in selecting irrelevant variables".

5. Conclusion

In this paper, the problem of selecting variables in count data regression models is considered. A firefly algorithm was proposed as a variable selection method. The results obtained from simulation examples and real data

Table 6. Simulation 6 results, on average, for NBRM

Methods	MSE	C	I
	$n = 50$		
FA-BN	3.062 (0.011)	0	0.677 (0.010)
AIC	7.422 (0.021)	0	2.942 (0.017)
CAIC	6.243 (0.019)	0	2.803 (0.017)
BIC	5.702 (0.017)	0	1.391 (0.015)
	$n = 100$		
FA-BN	2.788 (0.013)	0	0.655 (0.012)
AIC	7.048 (0.019)	0	2.878 (0.021)
CAIC	5.959 (0.019)	0	2.671 (0.019)
BIC	5.431 (0.016)	0	1.289 (0.019)
	$n = 200$	0	
FA-BN	2.713 (0.011)	0	0.637 (0.013)
AIC	7.073 (0.021)	0	2.858 (0.019)
CAIC	5.894 (0.019)	0	2.529 (0.019)
BIC	5.353 (0.019)	0	1.234 (0.015)

Table 7. MSE and the selected variables for the first real application

Methods	Selected variables	MSE
FA-BN	x_1, x_2, x_3, x_5	1338.21
AIC	x_1, x_2, x_4	1608.86
CAIC	x_1, x_2, x_3, x_4	1577.81
BIC	x_1, x_2, x_4	1571.05

Table 8. MSE and the selected variables for the second real application

Methods	Selected variables	MSE
FA-BN	x_1, x_2, x_5, x_6	79.07
AIC	x_1, x_2, x_3, x_7	110.84
CAIC	x_1, x_2, x_5, x_7	98.16
BIC	x_1, x_2, x_3, x_4	90.53

applications demonstrated the superiority of the FA-BN in terms of MSE, I, and C comparing with AIC, CAIC, and BIC methods.

REFERENCES

1. Z. Y. Algamal, *Diagnostic in Poisson regression models*, Electronic Journal of Applied Statistical Analysis, vol. 5, pp. 178–186, 2012.
2. Y. Asar, and A. Gens, *A New Two-Parameter Estimator for the Poisson Regression Model*, Iranian Journal of Science and Technology, Transactions A: Science, vol. 42, pp. 793–803, 2017.
3. S. Coxe, S. G. West, and L. S. Aiken, *The analysis of count data: a gentle introduction to poisson regression and its alternatives*, J Pers Assess, vol. 91, pp. 121–36, 2009.
4. Z. Wang, S. Ma, M. Zappitelli, C. Parikh, C. Y. Wang, and P. Devarajan, *Penalized count data regression with application to hospital stay after pediatric cardiac surgery*, Stat. Meth. Med. Res., In press., 2014.
5. A. C. Cameron, and P. K. Trivedi, *Regression analysis of count data* Cambridge university press, 2013.
6. J. M. Hilbe, *Modeling count data* Cambridge University Press, 2014.
7. Z. Y. Algamal, and M. H. Lee, *Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification*, Expert Systems with Applications, vol. 42, pp. 9326–9332, 2015.

8. R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 58, pp. 267–288, 1996.
9. J. Fan, and R. Li, *Variable selection via nonconcave penalized likelihood and its oracle properties*, Journal of the American Statistical Association, vol. 96, pp. 1348–1360, 2001.
10. H. Zou, and T. Hastie, *Regularization and variable selection via the elastic net*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 67, pp. 301–320, 2005.
11. H. Zou, *The adaptive lasso and its oracle properties*, Journal of the American Statistical Association, vol. 101, pp. 1418–1429, 2006.
12. G. I. Sayed, A. E. Hassanien, and A. T. Azar, *Feature selection via a novel chaotic crow search algorithm*, Neural Computing and Applications, 2017.
13. R. Sindhu, R. Ngadiran, Y. M. Yacob, N. A. H. Zahri, and M. Hariharan, *SineCcosine algorithm for feature selection with elitism strategy and new updating mechanism*, Neural Computing and Applications, vol. 28, pp. 2947–2958, 2017.
14. D. Broadhurst, R. Goodacre, A. Jones, J. J. Rowland, and D. B. Kell, *Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry*, Analytica Chimica Acta, vol. 348, pp. 71–86, 1997.
15. Z. Drezner, G. A. Marcoulides, and S. Salhi, *Tabu search model selection in multiple regression analysis*, Communications in Statistics - Simulation and Computation, vol. 28, pp. 349–367, 1999.
16. H. srkc, *Subset selection in multiple linear regression models: A hybrid of genetic and simulated annealing algorithms*, Applied Mathematics and Computation, vol. 219, pp. 11018–11028, 2013.
17. M. J. Brusco, *A comparison of simulated annealing algorithms for variable selection in principal component analysis and discriminant analysis*, Computational Statistics, vol. 77, pp. 38–53, 2014.
18. E. Dnder, S. Gmstekin, N. Murat, and M. A. Cengiz, *Variable selection in linear regression analysis with alternative Bayesian information criteria using differential evaluation algorithm*, Communications in Statistics - Simulation and Computation, vol. 47, pp. 605–614, 2017.
19. J. Pacheco, S. Casado, and L. Nsez, *A variable selection method based on Tabu search for logistic regression models*, European Journal of Operational Research, vol. 199, pp. 506–511, 2009.
20. A. Unler, and A. Murat, *A discrete particle swarm optimization method for feature selection in binary classification problems*, European Journal of Operational Research, vol. 206, pp. 528–539, 2010.
21. H. Kos, E. Dnder, S. Gmstekin, T. Kos, and M. A. Cengiz, *Particle swarm optimization-based variable selection in Poisson regression analysis via information complexity-type criteria*, Communications in Statistics - Theory and Methods, pp. 1–9, 2017.
22. T. J. Massaro, and H. Bozdogan, *Variable subset selection via GA and information complexity in mixtures of Poisson and negative binomial regression models*, arXiv preprint arXiv:1505.05229, 2015.
23. E. Dunder, S. Gumustekin, and M. A. Cengiz, *Variable selection in gamma regression models via artificial bee colony algorithm*, Journal of Applied Statistics, vol. 45, pp. 8–16, 2016.
24. J. M. Hilbe, *Negative binomial regression* Cambridge University Press, 2011.
25. Y. Asar, *Liu-type negative binomial regression: A comparison of recent estimators and applications*, In Trends and Perspectives in Linear Statistical Inference, Cham, 2018, pp. 23–39.
26. X.-S. Yang, *Multiobjective firefly algorithm for continuous optimization*, Engineering with Computers, vol. 29, pp. 175–184, 2013.
27. S. Yu, S. Zhu, Y. Ma, and D. Mao, *Enhancing firefly algorithm using generalized opposition-based learning*, Computing, vol. 97, pp. 741–754, 2015.
28. J. Zhang, B. Gao, H. Chai, Z. Ma, and G. Yang, *Identification of DNA-binding proteins using multi-features fusion and binary firefly optimization algorithm*, BMC Bioinformatics, vol. 17, pp. 323–337, 2016.
29. Z. Y. Algamal, and M. H. Lee, *Adjusted adaptive lasso in high-dimensional Poisson regression model*, Modern Applied Science, vol. 9, pp. 170–176, 2015.
30. J. S. Long, *The origins of sex differences in science*, Social forces, vol. 68, pp. 1297–1316, 1990.
31. Z. Y. Algamal, and M. H. Lee, *A two-stage sparse logistic regression for optimal gene selection in high-dimensional microarray data classification*, Advances in Data Analysis and Classification, Accepted, 2018.
32. Z. Y. Algamal, *Developing a ridge estimator for the gamma regression model*, Journal of Chemometrics, Vol. 32, pp. 1–12, 2018.
33. O. S. Qasim, and Z. Y. Algamal, *Feature selection using particle swarm optimization-based logistic regression model*, Chemometrics and Intelligent Laboratory Systems, Vol. 182, pp. 41–46, 2018.
34. M. M. Alanaz, and Z. Y. Algamal, *Proposed methods in estimating the ridge regression parameter in Poisson regression model*, Electronic Journal of Applied Statistical Analysis, Vol. 11, pp. 506–515, 2018.
35. M. Kazemi, D. Shahsavani, and M. Arashi, *Variable selection and structure identification for ultrahigh-dimensional partially linear additive models with application to cardiomyopathy microarray data*, Statistics, Optimization & Information Computing, Vol. 6, pp. 373C-382, 2018.
36. E. AVCI, *Flexiblity of Using Com-Poisson Regression Model for Count Data*, Statistics, Optimization & Information Computing, Vol. 6, pp. 278C-285, 2018.