Using the Firefly Algorithm to Select Variables in a Gamma Regression Model

Rehab Hamza Obeid, Asst. Prof. Dr. Nazik J. Sadik
rehab.hamza1201a@coadec.uobaghdad.edu.iq, dr.nazik@coadec.uobaghdad.edu.iq
Statistics Department/ Collage of administration and Economics/ University of Baghdad

Article Info
Page Number: 5060 – 5070
Publication Issue:
Vol 71 No. 4 (2022)

Article History
Article Received*: 25 March 2022*
Revised*: 30 April 2022*
Accepted*: 15 June 2022*
Publication*: 19 August 2022*

Abstract

The selection of variables in regression models plays an essential role in dealing with high-dimensional data. There were several methods for selecting the variables in order to obtain the best model. Regression modeling is of great interest in scientific fields, in addition to a problem in the selection of variables. In this paper, the FA is proposed to select a variable to model the gamma model.

Keywords: variable selection, high dimensional data, gamma regression model, firefly algorithm

1. Introduction

Choosing a feature or identifying variables plays a vital role in machine learning and data extraction, and the significant growth in the volume of data collected and the vast amount of data is often associated with irrelevant information affecting the performance and accuracy of systems. Algorithms are one of the most important modern methods for selecting variables for regression models, including the gamma regression model.

Gamma's regression is used when the dependent variable is distributed non- normally or is continuous with positive value or positively skewed, and therefore gamma decline supposesthat avariable has a latent distribution that often shows(positively skewed) data in social, epidemiological and economics, where the data of this type of study are non-negative values. Gamma distribution is a distribution that fits very well this type of data Gamma Regression Model (GRM) used to modeling the relationship between a positively response variable potentially regresses .When the y_i response variable~(gamma distribution) with the non-negative shape parameter τ, the non-negative measurement parameter θ this means that y_i ~ Gamma (τ, θ) .

Variable selection is an active search field in the regression analysis context. Gamma regression is full scale applied model for studying many real data problems, like automobile insurance claims, healthcare economics, and medical science [11], [12], [14]. Specifically, gamma regression model is used when the response variable under the study is not distributed as normal distribution or the response variable is positively skewed. Consequently, the gamma regression assumes that the response variable has a gamma distribution [2], [20].

In the recent period, nature-inspired algorithms such as theGA, ant algorithm, crow algorithm and others have been characterized by their effectiveness in selecting variables and have proven high efficiency as methods of selection [17]. The main purpose of selection is that we reduce the variables chosen so that we maintain the accuracy of the prediction so that they are studied as problems for improvement. [18].

In this paper, firefly Algorithm (FA) and variable selection methods will be reviewed using chaotic maps, gamma distribution review and all conditions to be met, and gamma distribution will also be linked to the linear regression model using the interconnection functions (reverse binding function) and parameters are estimated using one of the methods of estimation, (maximum likelihood) method .

2. Gamma Regression Model

Gamma distribution is a continuous distribution, known by Stacy in 1962, when variable values (Yi) are within the period $(0,\infty)$ and gamma distribution can be used in the analysis of positive random variables and is adopted in most medical fields[19], [9]. When the random variable (y) takes the form of a two-parameter gamma dist [9]

$$f(y;\theta,\tau)=\frac{\theta^{\tau}}{\Gamma_{(\tau)}}y^{\tau-1}\,exp\,exp(-\theta y)I_{(0,\infty)}y \qquad \tau,\theta>0 \quad (1)$$

In other words, equation 1 can be written as: [9]

$$f(y;\theta,\tau)=\frac{\theta}{\Gamma_{(\tau)}}(\theta y)^{\tau-1}\,exp\,exp(-\theta y)I_{(0,\infty)}y \qquad (2)$$

Where the shape parameter is (τ) and scale parameter is, $\Gamma_{(.)}$ gamma's function. With $E(Y_i)=\frac{\tau}{\theta}$ and $V(Y_i)=\frac{\tau}{\theta^2}=\mu^2\left(\frac{1}{\tau}\right)=\sigma^2(E(Y_i))^2$, and CDF IS given by:

$$F(y)=\frac{1}{\Gamma_{(\tau)}}\int_0^y u^{\tau-1}e^{-u}\,du$$

Let $Y_i\sim G(\mu_i,\tau)$ and $i=1,2,\ldots,n$ dependent random variables and that τ represent the shape parameter and here is a fixed value and therefore the equation of gamma slope it is the computational medium of the Y [1].

$$\eta_i=g(\mu_i)=x_i'\beta$$

$\beta=(\beta_{0,}\beta_{1,\ldots,}\beta_p)'$ are unknown regression parameters

$x_i=(x_{i1},\ldots,x_{ip})'$are p of independent variables, η_i is liner combination.

$g(.)$ Link functions

Gamma Regression Model has three links function: [1]

$log\ link\ function: g(\mu) = log\ log(\mu)$and

$identity\ link\ function: g(\mu) = \mu$ and $inverse\ link\ function: g(\mu) = \frac{1}{\mu}$

In the case of the figure parameter τ is not fixed, i.e. contains a linear composition, this will be addressed in this message, i.e. it can be modeled and the gamma slope model allows for the joint modeling of the average and shape parameters of the gamma distribution variable i.e.

$Y_i \sim G(\mu_i, \tau_i)$, where $i = 1,2, \dots, n$, [9]

$$\eta_{1i} = g(\mu_i) = x_i' \beta \quad ,$$

$$\eta_{2i} = h(\alpha_i) = z_i' \gamma$$

$\beta = (\beta_{0,}\beta_{1,\dots,}\beta_p)', \gamma = (\gamma_{0,}\gamma_{1,\dots,}\gamma_k)'$are the regression parameters related to the mean and the figure respectively, where g(μ) the binding function of the average (natural binding function). $h(\alpha)$ The binding function of the shape (usually the logarithmic binding function), η_{1i}, η_{2i} : linear predictions.$x_i = (x_{i1}, \dots, x_{ip})'$:p is vector of independent variables:

$z_i = (z_{i1}, \dots, z_{ik})'$.: k is vector of independent variables .

3. Maximum Likelihood Method

This method is one of the important methods because of its extensive applications for estimating the parameters of statistical models and this method is characterized by several evidentiary characteristics, including consistency, stability and lack of impartiality in most cases if the function of the possibility of distributing gamma can be written in the following form: [9]

$$L = \prod_{i=1}^{n} \quad f(y; \theta, \tau)$$

$$L = \prod_{i=1}^{n} \quad \frac{1}{\Gamma_{(\tau_i)}} \left(\frac{\tau_i}{\mu_i}\right)^{\tau_i} {y_i}^{\tau_i - 1} exp\left(-\frac{\tau_i}{\mu_i} y_i\right)$$

$$log\ log(L) = \sum_{i=1}^{n} \quad \left\{- log\ log(+\tau_i\ log\ log\left(\frac{\tau_i y_i}{\mu_i}\right) - log\ log(y_i) - \left(\frac{\tau_i}{\mu_i}\right) y_i \right\}$$

Where $\mu_i = x\acute{}\beta$ and $\tau_i = exp(z\acute{}\gamma)$

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{n} -\frac{\tau_i}{\mu_i}\left(1-\frac{y_i}{\mu_i}\right)x_{ij} \quad ; j = 1, \ldots, p$$

$$\frac{\partial L}{\partial \gamma_k} = \sum_{i=1}^{n} -\tau_i\left[\frac{d}{d\tau_i} log\ log\Gamma_{(\tau_i)} - log\ log\left(\frac{\tau_i y_i}{\mu_i}\right) - 1 + \frac{y_i}{\mu_i}\right]z_{ik} \quad ; k = 1, \ldots, r \quad , \quad j \geq k \quad , \quad p \geq r$$

And through the Hessian Matrix, which is the second-class partial derivation matrix of a multi-variable numerical function that includes all possible second-degree partial derivatives of the function [9] .

$$\frac{\partial^2{}_L}{\partial \beta_k \beta_j} = \sum_{i=1}^{n} \frac{\tau_i}{{\mu_i}^2}\left(1-\frac{2y_i}{\mu_i}\right)x_{ij}\,x_{ik} \quad ; j, k = 1, \ldots p$$

$$\frac{\partial^2{}_L}{\partial \gamma_k \beta_j} = \sum_{i=1}^{n} -\frac{\tau_i}{\mu_i}\left(1-\frac{y_i}{\mu_i}\right)x_{ij}\,z_{ik} \quad ; k = 1, \ldots r, j = 1, \ldots, p$$

$$\frac{\partial^2{}_L}{\partial \gamma_k \gamma_j} = \sum_{i=1}^{n} -\tau_i\left[\frac{d}{d\tau_i} log\ log\ \Gamma(\tau_i) - log\ log\left(\frac{\tau_i y_i}{\mu_i}\right) - 1 + \frac{y_i}{\mu_i}\right]z_{ij}\,z_{ik}$$

$$-\sum_{i=1}^{n} \tau_i\left[\tau_i \frac{d^2}{d{\tau_i}^2} log\ log\ \Gamma(\tau_i) - 1\right]z_{ij}\,z_{ik} \quad ; j, k = 1, \ldots, r$$

Used (Fisher information matrix) to calculate the contrast matrix associated with the greatest possible estimates of my agencies [9]:

$$I(\beta) = \left[-E\left(\frac{{\partial_L}^2}{\partial \beta_k \beta_j}\right) - E\left(\frac{{\partial_L}^2}{\partial \gamma_k \beta_j}\right) - E\left(\frac{{\partial_L}^2}{\partial \gamma_k \beta_j}\right) - E\left(\frac{{\partial_L}^2}{\partial \gamma_k \gamma_j}\right)\right]$$

$$-E\left(\frac{{\partial_L}^2}{\partial \beta_k \beta_j}\right) = \sum_{i=1}^{n} \frac{\tau_i}{{\mu_i}^2}x_{ij}\,x_{ik}$$

$$-E\left(\frac{{\partial_L}^2}{\partial \gamma_k \beta_j}\right) = 0 \quad , k = 1, \ldots, r \ ; \ j = 1, \ldots, p$$

$$-E\left(\frac{\partial_L{}^2}{\partial\gamma_k\gamma_j}\right)=\sum_{i=1}^{n}\quad \tau_i{}^2\left[\frac{d^2}{d_{\tau_i}{}^2}log\ log\ \Gamma(\tau_i)-\frac{1}{\tau_i}\right]z_{ij}\,z_{ik}\quad ;\ j,k=1,\dots,r$$

$$I(\beta)=\left[\sum_{i=1}^{n}\quad \frac{\tau_i}{\mu_i{}^2}x_{ij}\,x_{ik}00\sum_{i=1}^{n}\quad \tau_i{}^2\left[\frac{d^2}{d_{\tau_i}{}^2}log\ log\ \Gamma(\tau_i)-\frac{1}{\tau_i}\right]z_{ij}\,z_{ik}\right]$$

We note that (Fisher information matrix) is a diagonal matrix where one block matches the parameters of the medium regression and the other with the shape regression parameter and this leads to the maximum likelihood estimator of β,γ independent of each other and through the above we note that the parameters of the gamma regression model (GRM) cannot be estimated in the usual methods and we will use fisher score, a repetitive algorithm to obtain the maximum estimate of the probability of downhill model parametersgamma is similar to Newton-Raphson method or the algorithm of the iterative weighted least square using the expected value of the second matrix derivatives and through the algorithm we get an estimate of the parameters $\hat{\beta}\ \ and\ \hat{\gamma}$ [1] .

$$\hat{\beta}^{(k+1)}=(\acute{X}W_1{}^{(k)}X)^{-1}\acute{X}W_1{}^{(k)}Y$$

A diagonal matrix with elements of diameter containing is$W_1{}^k$

$$W_1{}^k=\frac{(\mu_i{}^2)^{(k)}}{\tau_i{}^{(k)}}$$

$$\hat{\gamma}^{(k+1)}=\left(\acute{Z}W_2{}^{(k)}Z\right)^{-1}\acute{X}W_2{}^{(k)}Y$$

$W_2{}^k$ diagonal matrix with elements of diameter containing $W_2{}^k=\frac{1}{\tau_i{}^{(k)}}$

$$di=\tau_i{}^{-2}\left[\frac{d^2}{d_{\tau_i}{}^2}log\ log\Gamma_{(\tau_i)}-\frac{1}{\tau_i}\right]^{-1}.$$

The Z matrix can contain the same variables as the X matrix [10]

4. Firefly algorithm

In recent years has been proposed many nature-inspired algorithms to solve improvement problems.

Firefly FA is one of the most recent proposed nature-inspired effective algorithms introduced for the first time by [21]. Applying the FA algorithm is easy to solve improvement problems compared to other nature-inspired algorithms and is inspired by the collective conduct of the lighters across bright lights.

FA has enabled agroup of (low-light) fireflies to move on the way to the brightest neighbors, fireflies with highersearch capabilities to solve improvement problems [7], [15], [13], [6], [3], [4], [8] .

There are three rules in the FA[22] ,the first rule is: that one of the fireflies is attracted to the other fireflies regardless of their gender, while the next rule is : that the attractiveness of the firefly is proportional to its brightness, that is, for any 2 flashing shepherds, the lower-brighter firefly moves towards the brighter fireflies, or in the absence of a single firefly brighter than a particular shepherd, the firefly will move randomly. The last rule states : that the brightness of the firefly depends on the analytical form of the trapping function, with regard to the problem of maximization, the brightness of each firefly is related to the value of the fitness function.

Let's say that d after the object function that will be improved, representsn_f (the number of fireflies) and (δ) denotes to the (light absorption factor) and represents (I_i) the intensity of light andrepresents r the space between the 2 sites $i(s_i)$& $j(s_j)$ This Cartesian distance can be described by:

$$r(s_i, s_j) = \sqrt{\sum_{c=1}^{d} (s_{i,c} - s_{j,c})^2} \qquad (3)$$

Since $I_i$decrease as soon as the space from the source rises, differences in I_i $must$be a monotonously decreasing function ,that means , most applications, $I_i$be able to rounded up as follows :

$$I_{(r)} = I_O e^{-\delta r^2} \qquad (4)$$

Since I_O is the original intensity of light, since the gravity of the firefly is proportional toI_O, the gravity of the firefly is description as follows:

$$\varphi_{(r)} = \varphi_o e^{-\delta r^2}$$

Where the $\varphi_o$represents gravity when $r = 0$.

The movement of any firefly to the best location, will be attracted to other people who are more attractive through*:*

$$s_i^{(t+1)} = s_i^{(t)} + \varphi_o e^{-\delta r_{i,j}^2}\left(s_j^{(t)} - s_i^{(t)}\right) + \alpha(k_1 - 0.5) \qquad (5)$$

Where the α andk_1 respectively, it is the random distribution parameter and the random number resulting from regular distribution [1, 0].

FA was originally proposed to solve continuous improvement problems, however, in variable selection, the improvement problem is intermittent. zhan, Gao suggested a BFA [23], algorithm to deal with the variable selection problem when the location is binary because the problem selecting the variable is in determining a particular variable or not, the result is therefore expressed like a binary vector, where value 1 refers to avariable that is Selected and represents 0 otherwise, in (BFA).

$$\varphi_o e^{-\delta r_{i,j}^2}\left(s_j^{(t)} - s_i^{(t)}\right) + \alpha(k_1 - 0.5) \quad (6)$$

It will move to the probability vector by the sigmoid function accordingly:

$$\text{Sigmoid} = \frac{1}{1+exp\,[\varphi_o e^{-\delta r_{i,j}^2}\left(s_j^{(t)} - s_i^{(t)}\right)+\alpha(k_1 - 0.5)]} \quad (7)$$

According to the fireflies position in the previous equation they will be replaced as follows:

$$s_i^{(t+1)} = \{1 \ if\ sigm\ \geq\ k_2 0\ otherwise$$

k_2 :which is a r.n. resulting from uniform dist. [1,0]

5. The suggested chaotic FA

Chaos theory explains irregular conductin non-linear systems for this reason, chaotic maps are used .Chaotic maps are conceived and can be transmitted as particles in the limited range of non-linear& Dynamic&non-linear systems ,with no specific way for regular travel of these units [16].

Chaos strategy is applied to avoid falling into the trap of local optimization and improve the quality of global optimization search. Therefore, chaos has been used in many optimization applications. Given that the feature the selection problem is an improvement problem with the search range [0, 1]; chaos can be used to improve this problem [16].

In this work, chaotic maps are studied to develop the working of the BFA in terms of avoiding falling into the trap of local optimization and make better the convergence speed of VS in the GRM .Two messy maps are used in this work. These maps are described in Table (1)

The table below represents two chaotic maps of the firefly algorithm to select variables for training data:

Table(1) The description of the two used maps

Method	Definition	Rang
Chebyshev	$x_{k+1} = \cos\ (k\cos^{-1}(xk))$	(-1, 1)
Sinusoidal	$x_{k+1} = 2.3x_k\sin(\pi x_k)$	(0, 1)

Our suggested Algorithm setting is defined by:

- The number of (fireflies) isn_f =5, φ_o= 1, δ = 0.2, α = 0.2

.

- The fitness function is express as:

$$fitness = min[\frac{1}{n}\sum_{i=1}^{n} \quad (yi-\widehat{yi})^2] \qquad (8)$$

- The locations of the fireflies are updated using Eq. (5).

6. Experimental side

The simulation aspect was used to find out the quality of the method where a number of variables were generated that distributed a natural multiple distribution, 20 independent variables were selected, 5 of them important and 15 unimportant where both of:

yi$\sim gamma(50,1)$

fn=5 fireflies population size

iter=5 iterations

gamma=0.2: light absorption coefficient

beta0=1 : attraction coefficient base value

alpha=0.2 : randomization parameter

pMin=0, Lower Bound of the position

pMax=1 Upper Bound of the position

Table)1(represents the simulation results, where the first column represents the independent variables, which represents the number of sites looking for a solution, and the second column represents 1, the chosen variable and 0 symbolizes the variable that has not been chosen.

The variables are selected according to two criteria: the first criterion is the number of correct variables in the original model that were selected in the estimation model. In this study, the number of real variables in the original variable is 5; hence we note that the quality of the method is better compared to other methods whenever this criterion is approximately 5. The second criterion for the selection of variables is the number of unreal variables (zero) that have been excluded and not entered into the estimated model.

Table (1)

Variables	Beta. True	Solution map 1	
X1	1	1	T
X2	1	1	T
X3	1	1	T
X4	1	1	T
X5	1	0	F
X6	0	1	F
X7	0	1	F
X8	0	1	F
X9	0	1	F
X10	0	1	F
X11	0	0	T
X12	0	1	F
X13	0	1	F
X14	0	1	F
X15	0	0	T
X16	0	0	T
X17	0	0	T
X18	0	1	F
X19	0	1	F
X20	0	1	F
MSE=0.6392139			

Table (2)

Variables	Beta. True	Solution map 2	
X1	1	1	T
X2	1	1	T
X3	1	1	T
X4	1	1	T
X5	1	1	T
X6	0	0	T
X7	0	1	F
X8	0	1	F
X9	0	1	F
X10	0	1	F
X11	0	0	F
X12	0	0	T
X13	0	1	F
X14	0	0	T
X15	0	1	F
X16	0	0	T
X17	0	1	F
X18	0	1	F
X19	0	1	F
X20	0	0	T
MSE=0.6329162			

Table (3)

Method	MSE
Chebyshev	0.6392139
Sinusoidal	0.6329162

Table (3(represents the third criterion for comparing the two methods, showing that the method of selecting variables according to the second method is better than the first method because the error rate is lower.

7. Conclusion

In this paper, the problem of VS in GRM is considered,achaotic FA with two maps was suggested as (VS) method.

The results found through the simulation show that the second method of selecting variables is the best because the number of correct variables in the original model was chosen by the estimation model according to the first criterion for selecting variables.

References

1. Adekanmbi, D. B. (2017), "Generalized Gamma Regression Models with Application to
2. Al-Abood, A. and Young, D. (1986). Improved deviance goodness of fit statistics for a gamma regression model. Communications in Statistics-Theory and Methods, 15(6):1865–1874.
3. Algamal, Z. (2019a). Variable selection in count data regression model based on firefly algorithm. Statistics, Optimization & Information Computing, 7(2):520–529.
4. Algamal, Z. Y. (2019b). A particle swarm optimization method for variable selection in beta regression model. Electronic Journal of Applied Statistical Analysis, 12(2):508– 519.
5. Algamal, Z. Y. and Lee, M. H. (2015). Penalized logistic regression with the adaptive lasso for gene selection in high-dimensional cancer classification. Expert Systems with Applications, 42(23):9326–9332.
6. Al-Thanoon, N. A., Qasim, O. S., and Algamal, Z. Y. (2018). Tuning parameter estimation in scad-support vector machine using firefly algorithm with application in gene selection and cancer classification. Computers in biology and medicine, 103:262–268.
7. Al-Thanoon, N. A., Qasim, O. S., and Algamal, Z. Y. (2019). A new hybrid firefly algorithm and particle swarm optimization for tuning parameter estimation in penal- Electronic Journal of Applied Statistical Analysis 275 ized support vector machine with application in chemometrics. Chemometrics and Intelligent Laboratory Systems, 184:142–152.
8. Al-Thanoon, N. A., Qasim, O. S., and Algamal, Z. Y. (2020). Variable selection in gamma regression model using binary gray wolf optimization algorithm. In Journal of Physics: Conference Series, volume 1591, page 012036. IOP Publishing.
9. Bossio, M. C., & Cuervo, E. C. (2015), "Gamma regression models with the Gammareg R package", Comunicaciones en EstadÃstica, 8(2), 211- 223.

CD4 Cell Counts Data of Aids Patients" , International Journal of Applied Mathematics & Statistical Sciences (IJAMSS), 6(4), 19-36.

10. Cuervo, E. C. (2001), "Modelagem da variabilidade em modelos lineares generalizados" (Doctoral dissertation, Tese de D. Sc., IM– UFRJ, Rio de Janeiro, RJ, Brasil).
11. De Jong, P. and Heller, G. Z. (2008). Generalized linear models for insurance data. Cambridge University Press.
12. Dunder, E., Gumustekin, S., and Cengiz, M. A. (2018). Variable selection in gamma regression models via artificial bee colony algorithm. Journal of Applied Statistics, 45(1):8–16.

13. Kahya, M. A., Altamir, S. A., and Algamal, Z. Y. (2019). Improving firefly algorithmbased logistic regression for feature selection. Journal of Interdisciplinary Mathematics, 22(8):1577–1581.
14. Malehi, A. S., Pourmotahari, F., and Angali, K. A. (2015). Statistical models for the analysis of skewed healthcare cost data: a simulation study. Health economics review, 5(1):11.
15. Qasim, O. S., Al-Thanoon, N. A., and Algamal, Z. Y. (2020). Feature selection based on chaotic binary black hole algorithm for data classification. Chemometrics and Intelligent Laboratory Systems, 204:104104.
16. Sayed, G. I., Darwish, A., and Hassanien, A. E. (2018). A new chaotic whale optimization algorithm for features selection. Journal of classification, 35(2):300–344.
17. Sayed, G. I., Hassanien, A. E., and Azar, A. T. (2019). Feature selection via a novel chaotic crow search algorithm. Neural Computing and Applications, 31(1):171–188.
18. Sindhu, R., Ngadiran, R., Yacob, Y. M., Zahri, N. A. H., and Hariharan, M. (2017). Sine–cosine algorithm for feature selection with elitism strategy and new updating mechanism. Neural Computing and Applications, 28(10):2947–2958.
19. stacy,E. (1962) , " A Generalized of the Gamma distribution " , The Annals of Mathemetical Statistical" , vol.33,No.1 pp.1187-1992.
20. Wasef Hattab, M. (2016). A derivation of prediction intervals for gamma regression. Journal of Statistical Computation and Simulation, 86(17):3512–3526.
21. Yang, X.-S. (2013). Multiobjective firefly algorithm for continuous optimization. Engineering with Computers, 29(2):175–184.
22. Yu, S., Zhu, S., Ma, Y., and Mao, D. (2015). Enhancing firefly algorithm using general ized opposi
23. Zhang, J., Gao, B., Chai, H., Ma, Z., and Yang, G. (2016). Identification of dna-binding proteins using multi-features fusion and binary firefly optimization algorithm. BMC bioinformatics, 17(1):323.