Sigma J Eng & Nat Sci 7 (2), 2016, 263-270

Papers Produced from Turkish Articles and PhD Theses Presented at
Graduate School of Natural and Applied Sciences, Yıldız Technical University
Yıldız Teknik Üniversitesi, Türkçe Makaleler ve Fen Bilimleri Enstitüsü
Doktora Tezi Kapsamında Hazırlanan Yayın

Research Article / Araştırma Makalesi

DIFERENTIAL EVOLUTION ALGORITHM FOR NONLINEAR REGRESSION MODELS

Didem TETİK KÜÇÜKELÇİ*[1], Atıf EVREN[2]

[1]*Yildiz Technical University, Graduate School of Natural and Applied Sciences, İSTANBUL,*
[2]*Yildiz Technical University, Faculty of Science and Literature, Department of Statistics, İSTANBUL*

Received/Geliş: 21.09.2016 Accepted/Kabul: 21.10.2016

ABSTRACT

Due to the inherent difficulties of nonlinear modelling, the studies for finding more practical methods on parameter estimation become more and more important. As well as numerical methods like Gauss-Newton, The Steepest Descent, Newton-Raphson, Levenberg-Marquardt Compromise algorithms etc., some methods based on artificial intelligence optimization get popularity among scientists increasingly. In this study, differential evolution algorithm (DEA) as one of the main artificial intelligence algorithms is used in nonlinear modelling. Then the parameter estimates by this method have been compared with those obtained by classic Gauss-Newton method. We have used three growth models, namely, Gompertz, Logistic and Weibull, in modeling. In the end, our emphasis is the similarity of parameter estimates realized by both methods. Hence DEA may be advocated for finding the similar results with greater simplicity.
Keywords: Nonlinear regression, differential evolution algorithm, Gauss-Newton method.

DOĞRUSAL OLMAYAN REGRESYON MODELLERİ İÇİN DİFERANSİYEL GELİŞİM ALGORİTMASI

ÖZ

Doğrusal ve doğrusal olmayan regresyon modellerinin çözümü birçok bilim dalında araştırma konusu olmaktadır.
Özellikle gerçek hayata dayalı olan doğrusal olmayan modellerin optimizasyonunda pratik yöntemler araştırmak zorunlu hale gelmiştir. Gauss-Newton, En dik iniş, Newton-Raphson, Levenberg-Marquardt uzlaşımı gibi iteratif çözümlere dayalı nümerik yöntemler paralelinde yapay zeka optimizasyon algoritmaları da önem kazanmaktadır.
Bu çalışmada literatürden alınan doğrusal olmayan regresyon modellerine örnek olan üç büyüme modeli kullanılmıştır: Gompertz, Lojistik, Weibull. Modellere Gauss-Newton metodu ile yapay zeka algoritmalarından biri olan diferansiyel gelişim algoritması uygulanmış ve sonuçlar karşılaştırılmıştır.
Anahtar Sözcükler: Doğrusal olmayan regresyon modelleri, diferansiyel gelişim algoritması, Gauss-Newton yöntemi.

* Corresponding Author/Sorumlu Yazar: e-mail/e-ileti: didemkucukelci@yahoo.com.tr, tel: (286) 218 00 18 / 2837
Doktora Öğrencisi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü

1. INTRODUCTION

A list of problems related to modeling data involves optimization procedures inevitably. Although linear models are wellcome by various scientists because of the easiness they provide , some relations are unfortunately nonlinear in nature.

1.1. Nonlinear Regression Models

A nonlinear regression model can be specified as

$$Y_i = f(X_{ij}, \gamma) + \varepsilon_i \tag{1.1}$$

Here Y_i is the ith observed value of dependent variable . $f(X_{ij}, \gamma)$ is the nonlinear function of parameters $\gamma_0, \gamma_1, \dots, \gamma_{p-1}$. X_{ij} is the ith observed value of the jth independent variable (j=1,2,…,q). The matrix of the observations of independent variables is

$$X_{qn} = \begin{bmatrix} X_{11} X_{12} \dots X_{1n} \\ X_{21} X_{22} \dots X_{2n} \\ \dots \\ X_{q1} X_{q2} \dots X_{qn} \end{bmatrix} \tag{1.2}$$

We suppose that the model has p parameters and the parameter vector is denoted by γ

$$\gamma = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \dots \\ \gamma_{p-1} \end{bmatrix} \tag{1.3}$$

The vector of initial estimates for p parameters g is as given below :

$$g = \begin{bmatrix} g_0 \\ g_1 \\ \dots \\ g_{p-1} \end{bmatrix} \tag{1.4}$$

Besides the difference between the kth parameter (k=0,1,2,…,p-1) and its initial estimate before iterations is

$$\beta_k^{(0)} = \gamma_k - g_k^{(0)} \tag{1.5}$$

If the matrix whose entries are the first derivatives of the expectation function with respect to kth parameter initially is denoted by

$$D_{ik}^{(0)} = \left[\frac{\partial f(X_i,\gamma)}{\partial \gamma_k}\right]_{\gamma = g^{(0)}} \tag{1.6}$$

by Taylor series expansion, one can get the following linearized form as

$$Y_i \cong f_i^{(0)} + \sum_{k=0}^{p-1} D_{ik}^{(0)} \beta_k^{(0)} + \varepsilon_i \tag{1.7}$$

If the ith residual is defined as

$$Y_i^{(0)} = Y_i - f_i^{(0)} \tag{1.8}$$

another version of (1.1) is obtained as follows:

$$Y_i^{(0)} \cong f_i^{(0)} + \sum_{k=0}^{p-1} D_{ik}^{(0)} \beta_k^{(0)} + \varepsilon_i \tag{1.9}$$

By using matrix notation;

$$Y^{(0)} \cong D^{(0)} \beta^{(0)} + \varepsilon \tag{1.10}$$

Here the derivative matrix **D** plays the same role as **X** matrix does in linear models. Hence parameter estimates and hypothesis tests are realized by this analogy (7). Following this argumentation, estimators of β can be realized initially:

$$b^{(0)} = \left(D^{(0)^T} D^{(0)}\right)^{-1} D^{(0)^T} Y^{(0)} \tag{1.11}$$

For the sake of generality, the superscripts (zero's) at (1.11) is simply replaced by "j" for the jth iteration.

$$b^{(j)} = \left(D^{(j)^T} D^{(j)}\right)^{-1} D^{(j)^T} Y^{(j)} \tag{1.12}$$

2. SOME DIFFERENCES IN GEOMETRIES OF LINEAR AND NONLINEAR MODELS

In linear regression models, the expected response vectors form a p-dimensional expectation surface in the n-dimensional response space. This surface is a linear subspace of the response space. For nonlinear models however, the expectation surface is a p-dimensional curved surface in the n-dimensional response space (1) .

In nonlinear regression models, the residuals play the same role as the observed values of explained variable of linear models. In linear models, observed values of dependent variable are projected onto two orthogonal subspaces of n dimensional Euclidean space to obtain residuals and parameter estimates. In nonlinear models, however, residuals themselves are projected to get parameter estimates iteratively. At the end of each iteration, current parameter estimate vector is updated by the following equation:

$$g_k^{(j)} = g_k^{(j-1)} + b_k^{(j-1)} \qquad \text{for j=1,2,…} \tag{1.13}$$

The iterative search procedure is finished as soon as iterative estimates (or alternatively some summarizing statistics like sum of squares, etc.) are convergent, although the solution reached at the end would rather coincide with a local optimum unluckily.

If the error terms are independently and normally distributed with zero expectation and the common variance σ^2, then the asymptotic distribution of g can be approximated well by a multivariate normal distribution with the expectation vector and variance-covariance matrix respectively as

$$E(g) \cong \gamma \tag{1.14}$$

$$s^2(g) = MSE(D^T D)^{-1} \tag{1.15}$$

Here MSE is the mean squares for the error term as usual:

$$MSE = \frac{\sum_{i=1}^{n} e_i^2}{n-p} \tag{1.16}$$

The validity of parameter estimates realized by(1.10), (1.11) and (1.12) depends on how justifiable the linearization technique is. If the intrinsic nonlinearity is high, then the results from the analysis of residuals may be highly misleading (9). In practice, a lot of competing algorithms are used so that (1.10), (1.11) and (1.12) are useful in a pedagogical way. The formulas in (1.14) and (1.15) are used in hypothesis testing and confidence interval estimation procedures . For bootstrap approach in nonlinear modeling one can refer to (5). One special difficulty to go forward by any estimation method is that one must initially determine starting values, step sizes. etc. Determining step size in each iteration requires another optimization procedure (2). Some other problems that have to be taken into account and some practical ways to remedy can be found in (6) and (8). Some of these issues can be listed as follows:

i) For nonlinear models, the objective function may have more than one optimum.

ii) R-square statistic (and some other versions of it) may be highly misleading for nonlinear models. For nonlinear models, R-square values should be very close to 1 for a good fit. For this reason comparing linear and nonlinear models by only checking R-square statistics is not appropriate (6).

iii) Initial parameter estimates are to be introduced the model exclusively. This may require another estimation procedure initially.

iv) Assumption on the normality of dependent variable does not guarentee normally distributed parameter estimates. Therefore confidence interval estimates require bigger sample sizes compared to linear models.

v) If intrinsic curvature is high, the results obtained by the analysis of residuals may be misleading .

3. DIFFERENTIAL EVOLUTION ALGORITHM (DEA)

Although the issue of DEA methodology on optimization problems is relatively older, a considerable amount of effort on the applicability of DEA methods on various fields is a fact. In this type of problem formulation, especially for nonlinear models, there exiss a lot of constraints that should be met and a lot of difficulties in calculations to be considered etc. Besides, not all the methods offered by DEA methodology is suitable in every situation. Recently scientists have focused on nature, natural systems, natural progress and natural ways. As a result, artificial intelligence algorithms, in other words intuitive methods, gain importance. DEA has been developed by Price and Storn in 1995. It provides scientists a simple but powerful tool (10). Especially, it is a direct search algortihm used to optimize functions of real variables on a global basis.

3.1. Parameters of DEA and Implementation

Selection is the step which determines the conditions, under which currently produced vectors will be included in population. In DEA, a new-born vector is kept in population, at least for one generation more, if it is less developed than its vector parents. Crossbreeding is a complementary operation whose main goal is to make the inquiry successful to constitute new vectors by using the existing goal vector parameters. In the first versions of DEA, special nonuniform, seperate crossbreeding operation was adopted. In this operation ,while a new solution (which has the chance of taking the place of its parents) has been generated, some elements of the solution are taken from parents whereas some other elements from mutated vectors. In selecting the new candidate for solution, exponential and binomial operations are done to control the frequency which determines the elements that will be taken from the vectors. These two operations are performed by a control parameter which is kept constant during optimization operation and also called as crossbreeding ratio (CR). At least one of the parameters of the new-born vector is aimed to be different from that of parent vector.

The basic steps of DEA algorithm are as below:

Step 1. Specify the values of control parameters:

D : the number of parameters to be optimized $j = 1,2, \dots, D$

NP : the number of variable vectors ,$NP \geq 4$, $i = 1,2, \dots, NP$

F : scale factor , $F \in (0,2)$

Assign parameter boundaries:

x^{-min} : lower boundaries of parameters.

x^{-max}: upper boundaries of parameters.

Step 2. Create initial population:

G : generation number.

$x_{j,i,G=0}$: jth parameter of ith chromozome of generation G.

$$x_{j,i,G=0} = x_j^{-min} + rand[0,1](x_j^{-max} - x_j^{-min}) \quad (3.1)$$

Repeat the following steps until stopping criterion is met.

Step3. Mutation and crossbreeding:

$u_{j,i,G+1}$:generated chromosome (solution) after generation $x_{j,i,G+1}$

$r_1, r_2, r_3 \in NP \; and \; r_1 \neq r_2 \neq r_3 \neq i$

$$u_{j,i,G+1} = \begin{cases} x_{j,r_3,G} + F(x_{j,r_1,G} - x_{j,r_2,G}), & rand_j[0,1] < CR \\ x_{j,i,G+1} & otherwise \end{cases} \quad (3.2)$$

Step 4. Selection:

$$\bar{x}_{i,G+1} = \begin{cases} \bar{u}_{i,G+1}, \; if \;\; f(\bar{u}_{i,G+1}) \leq f(\bar{x}_{i,G}) \\ \bar{x}_{i,G} \qquad\qquad otherwise \end{cases} \quad (3.3)$$

4. APPLICATION

4.1. Data

There are two data sets for three different growth models used in this study. In each data set there is one dependent, and one independent variable.

4.2. Models

Three widely used growth models are studied here. In model selection, our two criteria are the number of parameters and the difficulty levels of each model in parameter estimation respectively. Models can be classified as models having three or four parameters. Similarly, they can be identified as "models easy to estimate" and "models difficult to estimate".

4.3. Applying DEA

To estimate parameters of each model, DEA methodology has been adopted. The least squares function has been used to determine the degree of goodness of fit.

In other words, the parameter estimates that minimize the goal (least squares) function are found. In achieving this objective, the lower and upper parameter values are introduced by specifying some intervals and then by the help of "rand" function initial values are fixed.

After we fix these values, the value of the goal function is determined. Then by the help of rand function parameter estimates are subjected to mutation and crossbreeding. During this application, a program written on Matlab 7.10 has been run for 100, 200 and 300 iterations.

Table 4.1 Models and data sets

Gompertz y =B1*exp[−exp(B2 –B3*x)] Number of parameters: 3 Level of difficulty:high		Logistic y=B1/(1+exp(B2-B3*x)) Number of parameters:3 Level of difficulty : high		Weibull y=B1-B2*exp(-B3*x^B4) Number of parameters:4 Level of difficulty : medium	
y	x	y	x	y	x
8.93	9	8.93	9	16.08	1
10.8	14	10.8	14	10.8	2
18.59	21	18.59	21	18.59	3
22.33	28	22.33	28	22.33	4
39.35	42	39.35	42	39.35	5
56.11	57	56.11	57	56.11	6
61.73	63	61.73	63	61.73	7
64.62	70	64.62	70	64.62	8
67.08	79	67.08	79	67.08	9
				651.92	10
				724.93	11
				699.56	12
				689.96	13
				637.56	14
				717.41	15

Table 4.1 presents models, data sets, number of parameters to be estimated and the difficulty levels of each model in estimation.

Table 4.2 Parameter estimates for Gompertz growth model.

Parameters	Estimates provided by Gauss-Newton	DEA estimates (after 100 iterations)	DEA estimates (after 200 iterations	DEA estimates (after 300 iterations
B1	82.830	83.532	80.23	83.74
B2	1.224	1.297	1.293	1.264
B3	0.037	0.023	0.018	0.017

Table 4.3. Parameter estimates for logistic growth model.

Parameters	Estimates provided by Gauss-Newton	DEA estimates (after 100 iterations)	DEA estimates (after 200 iterations	DEA estimates (after 300 iterations
B1	72.46	73.89	70.02	74.50
B2	2.618	2.057	1.861	2.611
B3	0.067	0.067	0.035	0.073

Table 4.4 Parameter estimates for Weibull growth model.

Parameters	Estimates provided by Gauss-Newton	DEA estimates (after 100 iterations)	DEA estimates (after 200 iterations	DEA estimates (after 300 iterations
B1	69.96	68.952	67.491	66.691
B2	61.68	62.572	63.304	60.619
B3	0.0001	0.0005	0.0003	0.0002
B4	2.378	2.850	2.489	2.6378

For all of the models, the conditions F=0.8 and CR=1 are assumed. Although initially F=0.5, F=2 and F=1 weighing factors are assumed, it is observed that F=0.8 gives the best results. Secondly, we emphasize that even 100 iterations give satisfactory results for all of the models.

5. RESULTS

In literature, it is pointed out that DEA algorithm provides good solutions for some optimization problems. In this study, DEA algorithm is applied in modelling some nonlinear data. Then the results obtained have been compared to those of Gauss-Newton algorithm. In nonlinear regression models, the performance of the various algorithms depend mostly on the convenient determination of initial estimates of the parameters. The situation is nearly the same for DEA algorithm. Yet as soon as the initial estimates are appropriate, DEA procedure gives consistent final estimates with Gauss-Newton algorithm. In addition, Due to its simple structure in formulation, the great flexibility DEA provides has to be pointed out.

REFERENCES / KAYNAKLAR

[1] Bates D. M., Watts D.G. (1988), *Nonlinear Regression Analysis and Its Applications*, New York, John Wiley&Sons., pp 1-39.

[2] Bevington, P.R., Robinson, D.K. (2003), *Data Reduction and Error Analysis for the Physical Sciences,* McGraw Hill, Third edition, pp 148-151

[3] Billo, E.J. (2007), EXCEL for Scientists and Engineers Numerical Methods, Wiley-Interscience, John Wiley&Sons.

[4] De levie, R. (2004), Advanced Excel for Scientific Data Analysis, Oxford University Press.

[5] Huet, S., Bouvier, A., Gruet, M., Jolivet, E., (1996), *Statistical Tools for Nonlinear Regression: A Practical Guide with S-Plus Examples*, Springer-Verlag, New York, Springer Series in Statistics.

[6] Motulsky, H., Christopoulos, A. (2004), *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting, USA, Oxford University Press.*

[7] Neter J., Wasserman W., Kutner M. H. (1985), *Applied Linear Statistical Models*, Second edition, Illinois, Richard D. Irwin., pp 466-490.

[8] Ross, G.J.S.(1990), *Nonlinear Estimation,* Springer Series in Statistics , Springer-Verlag.

[9] Seber G.A.F., Wild C.J.(1989), *Nonlinear Regression*, USA, John Wiley&Sons., pp 91-102.

[10] Karaboğa, D. (2014), *Yapay Zeka Optimizasyon Algoritmaları*, Nobel Akademik Yayıncılık, 3.Baskı.

INTERNET RESOURCES

http//itl.nist.gov/d.