A New Optimization Strategy for Urbanization Construction Layout Based on Multiple Linear Regression Model

Zhao Fuyang

School of Economics and Management,
Dalian Jiaotong University
Dalian, Liaoning, 116028, China

Abstract — A new strategy for urbanization construction layout is of great importance for healthy development of economy and urbanization. In this paper we aim to design an effective new optimization strategy for urbanization construction layout. Firstly, we introduce the multiple linear regression model, and point out that we utilize the maximum likelihood estimator to compute the model parameters. Secondly, eight variables of the multiple linear regression model are defined and then EM algorithm is adopted to estimate the maximum likelihood for the weight vector. In order to speed up the new optimization process, parallel computing is utilized in our proposed maximum likelihood estimator. Thirdly, to test the effectiveness of our proposed optimization strategy, a case study based on a city in China is provided. After analyzing the current area and optimization area for different construction land types, the conclusions can be drawn that the proposed optimization strategy is able to optimize the new urbanization construction layout design.

Keywords -- Multiple linear regression; New urbanization construction; maximum likelihood estimator; layout optimization strategy.

I. INTRODUCTION

As is well known that land is belonged to the limited and non-renewable resources, and it is the most important factor in the development of economy and modern society. In general, with the rapid development of economy and urbanization, there are more and more land resources which are required [1][2].How to effectively plan and utilize land resources relies on optimizing the structure and layout of land and realize suitable land utilization. Hence, It is important to land management and the development of economy and society only if the land use planning is scientific [3]. That is to say, optimize the internal structure and layout mode of construction land between urban area and rural area is a key problem in modern society development [4].

Currently, the metropolitan city controls the pattern of urbanization, and it tends to the regionalization. Particularly, China's urban development has been in the era of "metropolitan area", meanwhile, the big cities achieve high urbanization level, its urban consolidation may extend to the entire city limitation, moreover, big cities represents the advanced form of current urban development of our country[5][6].

On the other hand, wasting resources and unreasonable exploting land resources can cause great conflicts and contradictions in the society, which has been paid more and more attentions[7].Hence, how to optimize the structure and layout of urban and rural construction land is very crucial in the economic healthy development[8]. Furthermore, it also plays an important role in the urban and rural economic and social development and then is may help to effective integration urban areas and rural areas.

The rest of the paper is organized as follows. Section 2 presents related works of multiple Linear Regression model. Section 3 gives overview of the multiple regression model. In section 4, the proposed method for new urbanization construction layout optimization problem is provided. Section 5 provides experimental results to verify the effectiveness of the proposed approach. In the end, section 6 concludes our works.

II. RELATED WORKS

As is a powerful computing tool, Multiple Linear Regression model has been successfully and widely utilized in many fields. Therefore, in this section, we try to discuss related works about applications of the Multiple Linear Regression model.

Chong et al. exploited Central Composite Design and adaptive neuro-fuzzy inference system are adopted to identify and forecast the output intensity ratio of light which passes through a plastic optical fiber sensor in Remazol Black B dye solution of various concentrations. The predictive performances of these models are compared with the Multiple Linear Regression model[9].

Yano et al. tested the bitterness intensity of a panel of P-gp substrates and non-substrates using different taste sensors, and then exploited multiple linear regression model to estimate the correlation between P-gp-inhibitor/substrate status and different physical properties. Particularly, the authors computed the first principal component analysis score as the representative value of bitterness [10].

Moon et al. analyzed the estimated calibration accuracy with indoor environmental conditions exploiting multiple temperature sensors for a W-band direct detection radiometer system. The proposed approach is designed to compute the radiometer calibration coefficient using the Multiple Linear Regression model[11].

Mahmoud et al. provided two kinds of approaches to tackle the problem of profile monitoring with sample sizes lower than the number of regression parameters. Experimental results demonstrate that both these approaches perform better than the existing approaches which are utilized to detect multiple linear regression profile[12].

Bhosale et al. test the accuracy of artificial neural networks and multiple linear regression model for forecast of first lactation 305-day milk yield adopting monthly test-day milk yield records of 443 Frieswal cows. Particularly, the performance of ANN is discovered to be higher than the multiple linear regression model to forecast milk yield[13].

Li et al. proposed a multivariate sparse group lasso variable selection and estimation approach for data with high-dimensional predictors and high-dimensional response variables. In general, this approach is developed through a penalized multivariate multiple linear regression model through an arbitrary group structure[14].

Wang et al. utilized Multiple linear regression to choose the best subset of descriptors and to construct linear models. Moreover, nonlinear models are designed through an artificial neural network. The given models with 6 descriptors demonstrate good predictive power: the linear model has the average absolute relative deviation for the training and test sets[15].

Apart from the above works, there are some other researches about Multiple Linear Regression model in recent years, such as factors affecting the simulated Basic Life Support (BLS) performance with Automated External Defibrillator (AED) in Flemish lifeguards[16], Quantitative Volumetrics Reveal Abnormalities in the Inferior Lateral Ventricles[17], estimate time-frequency electrophysiological responses in single trials[18],solubility of some fullerenes derivatives[19], Predicting internal bond strength of particleboard under outdoor exposure[20], Pan Evaporation Simulation Based on Daily Meteorological Data[21], Tropical Cyclone Intensity[22].

Overview of the multiple linear regression model

Multiple linear regression model is defined as follows.

$$y_i = x_i^{'}\beta + \varepsilon_i, i \in \{1,2,\cdots,n\} \tag{1}$$

where y_i refers to responses, x_i denotes the $p \times 1$ vectors of covariates, and

$$\beta = (\beta_1, \beta_2, \cdots, \beta_p)^*$$

refers to an unknown $p \times 1$ vector of the parameters of the multiple regression model. Particularly, ε_i means the unobservable random errors, and ε_i is also defined as a first order random coefficient autoregressive process as follows.

$$\varepsilon_i = (\theta + z_i)\varepsilon_{i-1} + e_i, i \in \{1,2,\cdots,n\} \tag{2}$$

where parameter θ denotes auto regression parameter, and z_i, e_i denote zero and mean independent process respectively.

In particular, maximum likelihood estimator is usually exploited to calculate parameters in multiple regression model. In maximum likelihood estimator model, parameter β is defined as follows.

$$\hat{\beta} = \arg\min_{\beta} (Y - X\beta)^T (Y - X\beta) \tag{3}$$

We also can represent $\hat{\beta}$ by the following equation.

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{4}$$

Supposing that $s_1, s_2, \cdots, s_J$ denote weights of maximum likelihood estimator for innermost intervals $H_1, H_2, \cdots, H_J$. Afterwards, the following condition should be satisfied.

$$L^*(s) = \prod_{i=1}^{n} \sum_{j=1}^{J} \alpha_{ij} \cdot s_j \tag{5}$$

where $\alpha_{ij} = I[H_j \subset A_i]$ and $I(\ )$ refers to the usual indicator function. Furthermore, the aim of multiple regression model is to maximize the likelihood satisfying the following constraints:

$$\sum_{j=1}^{J} s_j = 1, s_j \geq 0, j \in \{1,2,\cdots,J\} \tag{6}$$

III. THE PROPOSED METHOD FOR NEW URBANIZATION CONSTRUCTION LAYOUT OPTIMIZATION PROBLEM.

To study on the new urbanization construction layout optimization strategy, variables of the multiple linear regression model are defined in advance (shown in Table. 1).

TABLE I. TABLE 1. VARIABLES OF THE MULTIPLE LINEAR REGRESSION MODEL

Variable	Factor
X1	F1: Second industrial added value
X2	F2: Third industrial added value
X3	F3: Industrial added value
X4	F4: Increase in transportation
X5	F5: Construction land
X6	F6: Urban land
X7	F7: Independent industrial and mining land
X8	F8: Land used in transportation

Multiple linear regression model is belonged to regression analysis, and it is used to analyze the relationships between Independent variables and Dependent variables. Mathematical model of multiple linear regression model is defined as follows.

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n \tag{7}$$

where y refers to a Dependent variable, x_i denotes an Independent variable, and n means the number of variables.

DOI 10.5013/IJSSST.a.17.16.02 ISSN: 1473-804x online, 1473-8031 print

Therefore, we should constrain the search space, and then utilize symbol Ω to represent the parameter space as follows.

$$\Omega = \left\{ s \in R^J, \sum_{j=1}^{J} s_j = 1, s_j \geq 0, j \in \{1, 2, \cdots, J\} \right\} \quad (8)$$

Next, in order to estimate the maximum likelihood for vector s, EM algorithm is exploited.

(1) E-step: Supposing that the expected value of α_{ij} is defined as the parameter $\beta_{ij}(s)$, which is calculated as follows.

$$\beta_{ij}(s) = \frac{\alpha_{ij} s_j}{\sum_{m=1}^{J} \alpha_{im} \cdot s_k} \quad (9)$$

(2) M-step: In this maximiztion process, the overall proortion of failures for H_j is defined as follows.

$$\gamma_j(s) = \frac{1}{n} \sum_{i=1}^{n} \beta_{ij}(s) \quad (10)$$

To speed up the new urbanization construction layout optimization process, we introduce the parallel computing technology in maximum likelihood estimator, and the computing process of the parallel maximum likelihood estimator is illustrated in Fig. 1 as follows.

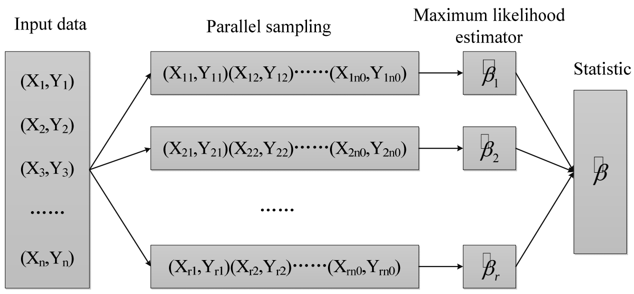

Figure 1. Figure 1. Structure of the parallel maximum likelihood estimator

In this section, we provide a case study to test performance of our proposed new urbanization construction layout optimization strategy. Particularly, we choose a city in east of China to construct this case study, and the economy and society index and each type construction land are listed as follows.

TABLE II. TABLE 2. VALUES OF EACH ECONOMY AND SOCIETY INDEX OF A GIVEN CITY FOR DIFFERENT YEARS.

Year	X1	X2	X3	X4	X5	X6	X7	X8
2000	90902.88	14901.27	82323.29	451.92	421.11	115.50	460.27	338.33
2001	143256.27	22534.67	104508.84	542.63	2008.06	373.34	464.60	341.20
2002	152258.35	24762.11	135724.24	3295.43	2721.41	1019.91	479.82	356.38
2003	158672.33	27453.04	153662.79	3646.44	2950.08	1120.52	522.12	359.30
2004	175898.10	65133.40	174443.09	9144.20	4766.12	1201.05	547.25	370.55
2005	195108.52	81900.09	213673.18	10708.17	5057.01	1236.07	564.97	378.46
2006	206798.43	96768.82	232522.79	12834.46	5851.87	1447.40	579.49	395.41
2007	229792.92	113694.16	312190.54	14861.85	6139.71	1616.03	581.33	398.06
2008	278351.55	118436.88	317309.72	16914.91	6222.37	1644.37	591.47	408.57
2009	360545.46	122284.97	341481.92	19070.60	7592.02	1701.52	617.10	426.02
2010	459192.12	136596.54	344232.96	19490.10	9160.23	1834.60	628.48	427.65
2011	462504.67	154263.49	376270.79	21053.35	13794.35	1864.00	635.61	438.30
2012	475751.03	203383.26	412183.71	22280.31	14019.73	1916.50	645.80	445.89
2013	513269.57	228736.20	414640.76	22785.97	20492.51	1918.46	648.96	480.65
2014	534011.38	234339.46	449435.30	25703.74	22522.14	2112.24	653.57	488.90

Using the correlation analysis of Independent variables and Dependent variables

TABLE III. TABLE 3. FACTOR CORRELATION COEFFICIENT TABLE FOR DEPENDENT VARIABLE

Factor	Variable
X1	0.8745
X2	0.9208
X3	0.9517
X4	0.9166
X5	0.8270
X6	0.9439
X7	0.9589
X8	0.8976

Then, based on the above settings, we propose the new urbanization construction layout optimization results for the given city in the year 2015. Particularly, we utilize six land types in the new urbanization construction layout optimization problem.

TABLE IV. TABLE 4. CURRENT AREA AND OPTIMIZATION AREA FOR DIFFERENT CONSTRUCTION LAND TYPE IN 2015 OF THE GIVEN CITY

Construction land type	Current area	Optimization area
Urban land	2209.12	2550.51
Rural residential land	3148.42	3183.09
Independent industrial and mining land	784.22	1367.68
Special ground	73.58	77.00
Transportation land	512.05	985.63
Water conservancy facilities land	16197.27	17042.05
Total	22924.66	25205.95

Afterward, based on the experimental resutls in Table. 3, we provide the ratio of different land types before optimization and after optimization.

Integrating all the above experimental results, we can find that using the proposed strategy, more reasonable new urbanization construction layout results can be achieved.

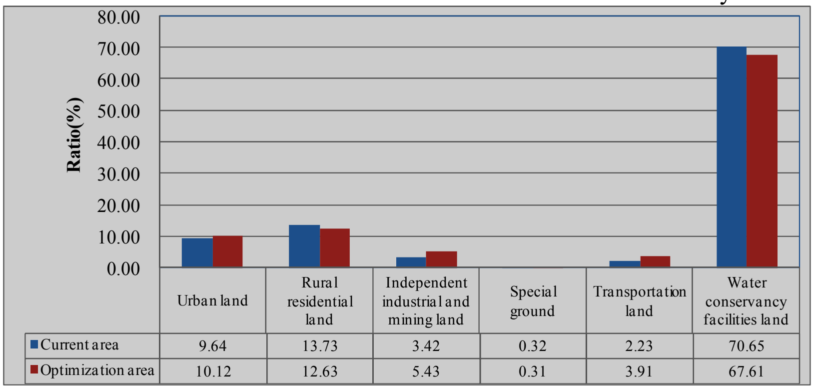

Figure 2. Ratio of different type of land in 2015 of the given city

Integrating all the above experimental results, we can find that using the proposed strategy, more reasonable new urbanization construction layout results can be achieved.

IV. CONCLUSION

This paper aims to design an effective new urbanization construction layout optimization strategy. The multiple linear regression model is given in this paper, and we use maximum likelihood estimator to compute parameters in multiple linear regression model. Next, eight variables of the multiple linear regression model are defined and then EM algorithm is used as well. Next, to speed up the new urbanization construction layout optimization process, parallel computing is integrated with the maximum likelihood estimator. Finally, a case study demonstrates that our method can provide an optimal new urbanization construction layout strategy.

ACKNOWLEDGEMENT

This work is supported by Social Science Fund of Liaoning Province (No.L14CJL031）.

REFERENCES

[1] Buckley Michelle, “On the Work of Urbanization: Migration, Construction Labor, and the Commodity Moment”, Annals of the Association of American Geographers, vol. 104 , No. 2, pp. 338-347, 2014.

[2] Fry Matthew, “From Crops to Concrete: Urbanization, Deagriculturalization, and Construction Material Mining in Central Mexico”, Annals of the Association of American Geographers, vol. 101, No. 6, pp. 1285-1306,2011.

[3] Zhao Jingzhu, Dai Dongbao, Lin Tao, Tang Lina, “Rapid urbanisation, ecological effects and sustainable city construction in Xiamen”, International Journal of Sustainable Development and World Ecology, vol. 17, No. 4, pp. 271-272,2010.

[4] Qiu Bingkui, Li Huilei, Zhou Min, Zhang Lu, “Vulnerability of ecosystem services provisioning to urbanization: A case of China, Ecological Indicators”, vol. 57: , No.22, pp. 505-513,2015.

[5] Ren Lijun, Wang Wenjun, Wang Jianchun, Liu Rutao,” Analysis of energy consumption and carbon emission during the urbanization of Shandong Province, China”, Journal of Cleaner Production, vol. 103 , No.4, pp. 534-541,2015,:.

[6] Pan Jiahua, Liu Jie, Zhu Chun, “Budgeting carbon for urbanization in China”, Wiley Interdisciplinary Reviews-energy and Environment, vol. 4, No. 5, pp. 406-409,2015.

[7] Enloe Heather A., Lockaby B. “Graeme, Zipperer Wayne C., Somers Greg L., Urbanization effects on soil nitrogen transformations and micro bialbiomass in the subtropics”, Urban Ecosystems, vol. 18 , No. 3, pp. 963-976,2015.

[8] Li Zhen, Yan Zhongwei, Tu Kai, Wu Hongyi, “Changes of Precipitation and Extremes and the Possible Effect of Urbanization in the Beijing Metropolitan Region during 1960-2012 Basedon Homogenized Observations”, Advances in Atmospheric Sciences, vol. 32, No.9, pp. 1173-1185, 2015.

[9] Chong Su Sin, Aziz A. R. “Abdul, Harun Sulaiman W., ArofHamzah, ShamshirbandShahaboddin, Application of multiple

linear regression, central composite design, and ANFIS models in dye concentration measurement and prediction using plastic optical fiber sensor", Measurement, vol. 74, No.6, pp. 78-86,2015.

[10] Yano Kentaro, MitaSuzune, Morimoto Kaori, "Haraguchi Tamami, Arakawa Hiroshi, Yoshida Miyako, Yamashita Fumiyoshi, Uchida Takahiro, OgiharaTakuo, Multiple Linear Regression Analysis Indicates Association of P-Glycoprotein Substrate or Inhibitor Character with Bitterness Intensity, Measured with a Sensor", Journal of Pharmaceutical Sciences, vol. 104 , No.10, pp. 2789-2794,2015.

[11] Moon Nam-Won, KimYong-Hoon, "Temperature Drift Compensation Using Multiple Linear Regression for aW-Band Total Power Radiometer", IEEE Sensors Journal, vol. 15, No.8, pp. 4612-4620,2015.

[12] Mahmoud Mahmoud A., "SaadAbd El Naser, El ShaerReham, Phase II Multiple Linear Regression Profile with Small Sample Sizes", Quality and Reliability Engineering International, vol.31 , No.5, pp. 851-861,2015.

[13] Bhosale Manisha Dinesh, Singh T. P., "Comparative study of feed-forward neuro-computing with multiple linear regression model for milk yield prediction in dairy cattle", Current Science, vol. 108, No. 12, pp. 2257-2261,2015.

[14] Li Yanming, Nan Bin, Zhu Ji, "Multivariate sparse group lasso for the multivariate multiple linear regression with an arbitrary group structure", BIOMETRICS, vol. 71, No. 2, pp. 354-363,2015.

[15] Wang Daxiang, Yuan Ye, DuanSiyao, Liu Ruina, GuShaojin, Zhao Sanping, Liu Li, Xu Jie, "QSPR study on melting point of carbocyclic nitroaromatic compounds by multiple linear regression and artificial neural network", Chemometrics and Intelligent Laboratory Systems, vol.15 , No. 143, pp. 7-15.,2015.

[16] Iserbyt Peter, Schouppe Gilles, Charlier Nathalie, "A multiple linear regression analysis of factors affecting the simulated Basic Life Support (BLS) performance with Automated External Defibrillator (AED) in Flemish lifeguards", RESUSCITATION, vol. 70, No. 89, pp. 70-74,2015.

[17] Tsay Trevor, Wei Angela, Walsh Ayla, "Chang Eric, MulyantoNadya, Wu Joseph, Multiple Linear Regression Model of Quantitative Volumetrics Reveal Abnormalities in the Inferior Lateral Ventricles", Amygdala, and Pallidum of Brain Injury Patients Compared to Normal Controls, Journal of Head Trauma Rehabilitation, vol. 30, No. 3, pp. E106- E107,2015.

[18] Hu L., Zhang Z. G., Mouraux A., Iannetti G. D., 'Multiple linear regression to estimate time-frequency electro physiological responses in single trials", NEUROIMAGE, vol.32 , No. 111, pp. 442-453,2015.

[19] PourbasheerEslam, Aalizadeh Reza, 'Ardabili Jamal Saffar, Ganjali Mohammad Reza, QSPR study on solubility of some fullerenes derivatives using the genetic algorithms - Multiple linear regression", Journal of Molecular Liquids, vol. 204, No.22, pp. 162-169,2015.

[20] Watanabe Ken, Korai Hideaki, Matsushita Yasuhiro, "Hayashi Tomoyuki, Predicting internal bond strength of particleboard under outdoor exposure based on climate data: comparison of multiple linear regression and artificial neural network", Journal of Wood Science, vol. 61, No. 2, pp. 151-158,2015.

[21] Malik Anurag, Kumar Anil, "Pan Evaporation Simulation Based on Daily Meteorological Data Using Soft Computing Techniques and Multiple Linear Regression", Water Resources Management, vol. 29 , No. 6, pp. 1859-1872,2015.

[22] Lee Chia-Ying, Tippett Michael K., "Camargo Suzana J., Sobel Adam H., Probabilistic Multiple Linear Regression Modeling for Tropical Cyclone Intensity", Monthly Weather Review, vol. 143 , No.3, pp. 933-954, 2015.