

Smart Growth Prediction Based on Support Vector Regression

Feiyang Li, Wenjie Chen, Weijian Chen and Nian Cai*

School of Information Engineering, Guangdong University of Technology, Guangzhou, P. R. China

*Corresponding author

Abstract—Smart growth is a technique to improve the quality of development for a city. To effectively measure the degree of smart growth, an evaluation model is proposed based on principle component analysis (PCA) in this report. We use support vector Regression (SVR) to predict the components of smart growth and measure the degree of smart growth in the future. Our experimental results indicate that the proposed model is feasible to measure the degree of smart growth of a city and predict the trends of smart growth.

Keywords—smart growth; principle component analysis; support vector regression

I. INTRODUCTION

Globally, more people live in urban areas than in rural areas, with 54 percent of the world's population residing in urban areas. The urban population of the world has grown rapidly since 1950, from 746 million to 3.9 billion in 2014. Continuing population growth and urbanization are projected to add 2.5 billion people to the world's urban population by 2050^[1].

Due to the increasing urban population, urban sprawl will inevitably occur, which is the expansion of human populations away from central urban areas into low-density and usually car-dependent communities in a process called suburbanization. However, human daily activities still depend on the center of the city. For example, people living in the suburban area have to drive to work or purchase in the downtown, which leads to air pollution and traffic jam. Furthermore, the overdevelopment for the human residential area will decrease the farmland, forest and open squares rapidly, which invades the habitats of animals and botany. Thus, urban sprawl will significantly influence the environment and economics of the city when the overpopulation becomes serious in the city.

In order to develop economic continuously and sustain the quality of human living condition, city and state governments in US have utilized an urban planning theory called smart growth to slow suburban sprawl since 1990s^[2]. Smart growth is a theory of land development that accepts that growth and development will continue to occur, and so seeks to direct that growth in an intentional, comprehensive way. Its proponents include urban planners, architects, developers, community activists, and historic preservationists^[3]. There are ten principles for smart growth^[4], which are (1)mix land uses, (2)take advantage of compact building design, (3)create a range of housing opportunities and choices, (4)create walkable neighborhoods, (5) foster distinctive, attractive communities with a strong sense of place,(6)preserve open space, farmland, natural beauty, and critical environmental areas, (7)strengthen

and direct development towards existing communities, (8)provide a variety of transportation choices, (9)make development decisions predictable, fair, and cost effective, (10)encourage community and stakeholder collaboration in development decisions. Smart growth concerns on the 3 aspects which are (1) economically prosperous, (2) socially equitable, (3) environmentally sustainable. (In other word, they are three Es)

To effectively measure the degree of smart growth, we develop an evaluation model based on principle component analysis (PCA) and support vector Regression (SVR). The proposed model is also determined by the three Es. Here, we define the three Es as three kinds of factors in our proposed model, which are economic factor, social factor and environment factor. Each kind of factor consists of several vital indicators. We can predict the indicators of our model and analyze the future degree of smart growth.

II. THE METRIC OF SMART GROWTH

In this section, we construct a metric to evaluate the prosperity of a city, named total metric of smart growth (SGTM). SGTM is influenced by three vital factors, which are economic factor (ECF), social factor (SOF) and environmental factor (ENF). Here, we assume that SGTM can be formulated by a linear combination of three factors.

$$SGTM = \frac{ECF + \alpha SOF + \beta ENF}{1 + \alpha + \beta} \quad (1)$$

where α and β are the weights, which imply the contributions of SOF and ENF to SGTM, respectively. When $\alpha = \beta = 1$, it indicates that three factors make the same contribution to SGTM. It distinguishes from different city due to the degree of development.

A. Indicators of Factors

Each factor of the proposed model contains some indicators to evaluate the factor.

ECF:

- (1) Estimated Median House or Condo Value (EMH)
- (2) Unemployment Rate (UR)
- (3) Median Household Income (MHI)

In the process of city modernization, the most accurate and direct reflections of the resident economy are the price of the house and the income of household. And low unemployment

rate of the city guarantees the steady development^[5]. Therefore, we choose EMH, UR and MSI as the indicators of ECF.

SOF:

- (1) Education Gini Index (EGI)
- (2) Healthy Diet Rate (HDR)
- (3) Crime Index (CI)

In order to determine SOF, we take education, health and safety into consideration to ensure the equity of the society. EGI is an equitable index to measure the fairness of education. The lower EGI implies the more fair education of the city^[6]. Health life style is the fundament of resident social life and represents the level of resident living standard. Therefore, we choose EGI, HDR and CI as the indicators of SOF.

ENF:

- (1) Air Quality Index (AQI)
- (2) Utility Gas Rate (UGR)

In the modern city, what the residents mostly concern about is air pollution. The quality of the air directly affects our daily life. Therefore, we use AQI and UGR to evaluate ENF. AQI is calculated for four major air pollutants regulated by the Clean Air Act: ground-level ozone, particle pollution, carbon monoxide, and sulfur dioxide^[7]. UGR is a reflection of the energy consume, which can also evaluate ENF.

B. Principle Component Analysis

In this section, we evaluate the smart growth of the cities according to the indicators of the cities. We synthesize the 3 vital factors to determine the weight of each indicator via principle component analysis (PCA).

Assume that there are m indicator variables and n evaluation objects.

Min-Max Normalization. Because the attribute of each index is different from each other, we pre-process the collected data via the following strategy. Generally, attributes of indexes are benefit-oriented and cost-oriented. Benefit-oriented indexes are commonly the higher, the better. On the contrary, cost-oriented indexes are the lower, the better. In our model, MSI, AQI and HDR are benefit-oriented indicators and the rest are cost-oriented. For fair analysis and easy linear combination, the data should be normalized in the range from 0 to 1.

For any benefit-oriented index,

$$\tilde{a} = \frac{a - a_{min}}{a_{max} - a_{min}} \quad (2)$$

where \tilde{a} and a are the variables after and before normalization, respectively. a_{max} / a_{min} is the maximum/minimum of the variable a .

For any cost-oriented indexes,

$$\tilde{b} = \frac{b_{max} - b}{b_{max} - b_{min}} \quad (3)$$

where \tilde{b} and b are the variables after and before normalization, respectively. b_{max} / b_{min} is the maximum/minimum of the variable b .

After normalization, the higher the preprocessed index is, the more contribution to SGTM it has.

- 1) Calculate the correlation coefficient matrix \mathbf{R} .

$$\mathbf{R} = (r_{ij})_{m \times m} \quad (4)$$

$$r_{ij} = \frac{\sum_{k=1}^n \tilde{a}_{ki} \tilde{a}_{kj}}{n-1}, i, j = 1, 2, \dots, m \quad (5)$$

where r_{ij} is the correlation coefficient between the i th indicator and the j th indicator. $r_{ii} = 1, r_{ij} = r_{ji}$.

- 2) Calculate eigenvalues and eigenvectors. Calculate the eigenvalues of the correlation coefficient matrix \mathbf{R} $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ and the corresponding eigenvectors u_1, u_2, \dots, u_m . Here, $u_j = [u_{1j}, u_{2j}, \dots, u_{mj}]^T$ contains m eigenvectors of the new indicator variable.

$$y_1 = u_{11}\tilde{x}_1 + u_{21}\tilde{x}_2 + \dots + u_{m1}\tilde{x}_m,$$

$$y_2 = u_{12}\tilde{x}_1 + u_{22}\tilde{x}_2 + \dots + u_{m2}\tilde{x}_m,$$

⋮

$$y_m = u_{1m}\tilde{x}_1 + u_{2m}\tilde{x}_2 + \dots + u_{mm}\tilde{x}_m$$

A method of eigenvalue decomposition is used to calculate eigenvalues and eigenvectors. Suppose a matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$, it can be decomposed as

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \quad (6)$$

where $\mathbf{U} = [u_1, u_2, \dots, u_m]$. The column vectors of \mathbf{U} are the eigenvectors of \mathbf{A} . $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$

- 3) Choose $p (p \leq m)$ principle components, and calculate the integrated assessment value.

- a. Calculate the information contribution rates and accumulate the contribution rates of eigenvalues $\lambda_j (j = 1, 2, \dots, m)$.

$$b_j = \frac{\lambda_j}{\sum_{k=1}^m \lambda_k} \quad (7)$$

$$\alpha_p = \frac{\sum_{k=1}^p \lambda_k}{\sum_{k=1}^m \lambda_k} \quad (8)$$

where b_j is the information contribution rate of the principle component y_j and α_p is the accumulating contribution rate of principle components y_1, y_2, \dots, y_p . When α_p approaches to 1, p indicator variables y_1, y_2, \dots, y_p are selected as the principle components to replace the first m indicator variables. In general, when the accumulating contribution rate of p th is more than 80%, we

can neglect its subsequent indicator variables, and use p indicator variables to represent all the indicator variables.

b. Calculate the overall score as

$$Z = \sum_{j=1}^p b_j y_j \quad (9)$$

where b_j is the information contribution rate of the j th principle component. Thus, we can evaluate the SGTm via the overall score.

Finally, we can construct the metric of three factors as follows:

$$ECF = 0.6290y_{ECF1} + 0.2785y_{ECF2}$$

where $y_{ECF1} = 0.6695\tilde{x}_1 + 0.0459\tilde{x}_2 + 0.7414\tilde{x}_3$,

$$y_{ECF2} = 0.4880\tilde{x}_1 + 0.7797\tilde{x}_2 - 0.3924\tilde{x}_3$$

$\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$ respectively represent Estimated Median House or Condo Value (EMH), Unemployment Rate (UR), Median Household Income (MHI).

$$SOF = 0.4096y_{SOF1} + 0.3310y_{SOF2} + 0.2621y_{SOF3}$$

where $y_{SOF1} = 0.7013\tilde{x}_4 - 0.0459\tilde{x}_5 + 0.7414\tilde{x}_6$,

$$y_{SOF2} = 0.5638\tilde{x}_4 - 0.6048\tilde{x}_5 + 0.5624\tilde{x}_6,$$

$$y_{SOF3} = 0.4362\tilde{x}_4 + 0.7964\tilde{x}_5 + 0.4190\tilde{x}_6.$$

$\tilde{x}_4, \tilde{x}_5, \tilde{x}_6$ respectively represent Education Gini Index (EGI), Healthy Diet Rate (HDR), Crime Index (CI)

$$ENF = 0.7182y_{ENF1} + 0.2818y_{ENF2}$$

where $y_{ENF1} = 0.7071\tilde{x}_7 - 0.7071\tilde{x}_8$,

$$y_{ENF2} = 0.7071\tilde{x}_7 + 0.7071\tilde{x}_8$$

\tilde{x}_7, \tilde{x}_8 respectively represent Air Quality Index (AQI), Utility Gas Rate (UGR).

4) After the PCA procedure, we are still unable to calculate the SGTm since the measurement units of three factors are different. Therefore, we need to normalize the values of three factors calculated above. We can normalize each factor by

$$\tilde{c} = \frac{c - c_{min}}{c_{max} - c_{min}} \quad (10)$$

where \tilde{c} and c are the values of each factor after and before normalization, respectively. c_{max} / c_{min} is the maximum/minimum of c .

After data post-processing, we can calculate SGTm, whose values are in the range from 0 to 1. When the value of SGTm approaches to 1, the corresponding city can be considered as a smart-growth city.

III. PREDICTION OF SGTm

In this section, we apply Support Vector Regression (SVR) to predict the indicators mentioned above in a few decades.

Assume the training data as $\{(x_1, y_1), \dots, (x_l, y_l)\} \subset \mathcal{X} \times \mathbb{R}$, where \mathcal{X} denotes the space of the input patterns for instance \mathbb{R}^d . For SVR, our goal is to find a decision function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all training data^[8]. According to the principle of SVR, we can solve a convex optimization problem to build the decision function. In this instance, we use the final form of dual problem of SVR as follows:

$$\begin{aligned} & \text{maximize} -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) \\ & \quad - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i(\alpha_i - \alpha_i^*) \\ & \text{subject to} \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned}$$

where $K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right)$ is radial basis function (RBF) kernel function, $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ are dual variables. The constant $C > 0$ determines the trade off between the flatness of f and the amount up to which deviations larger than ε are tolerated. To solve the problem quickly, we use the *libsvm* toolbox in MATLAB.

IV. RESULTS

In this section, we use the data collected online (data source: <http://www.city-data.com>) to evaluate the degree of smart growth several cities in the world via SGTm and predict the components of SGTm to calculate SGTm.

TABLE I. THE VALUES OF DIFFERENT CITIES IN THE WORLD IN 2013

City	SGTM	City	SGTM
Honolulu	0.893764393	Brooklyn	0.382307056
Ewa	0.774364947	Fairfield	0.376292427
Florence	0.644552132	Pittsburgh	0.371259124
Orlando	0.528524676	New York	0.309428859
Wichita	0.432944166	Cincinnati	0.249981286
Lincoln	0.409095299	Chicago	0.091735735
Spring Valley	0.392449088	Newark	0.06764242

In order to calculate the SGTms for the references, we need to predict the future trend of indicators. Due to lack of the data, we predict the future trend of indicators based on the data from 2010 to 2014. If the original trend of indicators decreases, we use a logarithmic function to match the future trend. If the original trend of indicators increases, we use a linear function to match the future trend. In our prediction, indicators will improve every decade. Therefore, SGTms for the references in both cities are ideal and improve every decade.

We take Florence for example. We predict HDR and EMH separately via SVR in next few decades and assume other components grow or decade as reference.

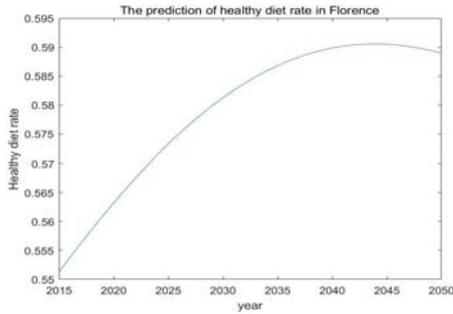


FIGURE I. THE PREDICTION OF HDR

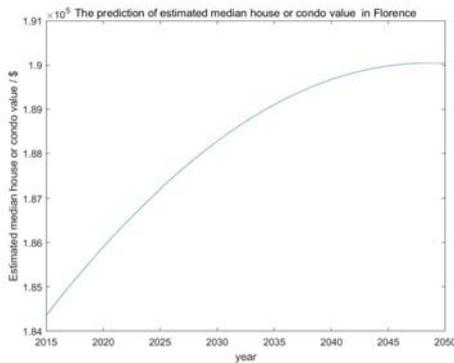


FIGURE II. THE PREDICTION OF EMH

TABLE II. SGTM OF FLORENCE

Year	Flo_Reference	HDR	EMH
2020	0.529254	0.530219	0.528510
2030	0.539457	0.540622	0.560708
2040	0.549022	0.550326	0.588377
2050	0.556897	0.558426	0.604379

V. CONCLUSION

In this report, we develop a model named SGTM to evaluate the degree of smart growth. SGTM contains three key factors that are the reflections of economically prosperous, socially equitable, and environmentally sustainable. Each factor consists of several components. We determine the relationship of each factor and its components via PCA. After separate the SGTM into components, we can conveniently to predict the SGTM of a city in the future.

REFERENCES

[1] United Nations. Dept. of Economic and Social Information and Policy Analysis. World urbanization prospects.[M]. United Nations, Dept. of International Economic and Social Affairs, 2012.

[2] Geoff Anderson. WHY SMART GROWTH: A PRIMER. International City/County Management Association.

[3] Boeing G, Church D, Hubbard H, et al. LEED-ND and Livability Revisited[J]. Social Science Electronic Publishing, 2014, 27(1):31-55.

[4] Smart Growth: A Guide to Developing and Implementing Greenhouse Gas Reduction Programs[J]. Local Government Climate & Energy Strategy Guides, 2010.

[5] Lan Cheng. A study on the dynamics of American house prices based on natural expectation[D]. Fudan University, 2012.

[6] Thomas V, Wang Y, Fan X. Measuring Education Inequality: Gini Coefficients of Education[J]. Social Science Electronic Publishing, 2001, 1(100):43-50.

[7] NJEDL. Air quality index: a guide to air quality and your health[J]. 2003.

[8] Smola A J, Lkoph B. A tutorial on support vector regression[J]. Statistics and Computing, 2004, 14(3):199-222.