For office use only	Team Control Number	For office use only
T1 ________	63648	F1 ________
T2 ________		F2 ________
T3 ________	Problem Chosen	F3 ________
T4 ________	E	F4 ________

2017

MCM/ICM

Summary Sheet

How to Scientifically Construct A Smart City

Smart growth covers a range of strategies that build cities economically prosperous, socially equitable, and environmentally sustainable. In this project, we seek to explore how to implement smart growth theories into city design around the world.

First of all, we define a novel metric to measure the success of smart growth of a city. Specifically, we identify seven important indicators which are closely related with the principles of smart growth, and collect data from 16 cities with these indicators. Outlier detection is used to discover anomalies and clean up data. Principal component analysis is used to convert the collected data into a set of linearly uncorrelated principal components. Then, we learn a prediction model to obtain the metric based on support vector regression model.

Secondly, we study the current growth plan of two selected mid-sized cities including Chaozhou and Saint Paul. Based on their existing development levels of each indicator, we find the advantages and disadvantages of their specific plans. The proposed metric can be used as an effective tool to evaluate the successful outcomes.

Thirdly, by exploring the correlationship between the proposed metric and these indicators, we recommend new plans for both cities. Our experiments evidence that the optimal smart growth strategies vary widely in different cities, based on their individual specific indicators such as the population density, percentage of unemployment, and average wage per person. To solve the fourth problem, our model defines the current gradient of our metric function as the potential of smart growth of a city. The ranking results of their initiatives are also provided and discussed.

Based on the empirical validation, our proposed model shows good generalization ability, robustness, and dynamics. Recently, with the development of artificial intelligence, scientists predict the computer-based intelligences are possible to exceed the existing technologies and even the sum total of human brainpower. Hence, our future work is to develop creative strategies and tools based on the new computer-based technologies for smart growth.

Key words: smart growth, outlier detection, principal component analysis, regression

Content

1 Introduction

1.1 Background

In the late 1990s, Americans realized the problems of their “suburbanization” development. Europe’s “compact development” has kept many historic towns in a compact, high-density form. This strategy is widely regarded as an ideal environment for living and working. Therefore, Americans proposed a “smart growth” concept. Smart growth covers a range of development and conservation strategies that make our cities, towns, and neighborhoods more attractive, economically stronger, and more socially diverse [1]. In this project, our mission is to implement smart growth theories into city design around the world.

1.2 Problem Restatement

Above all, smart growth is about helping every town and city become a more economically prosperous, socially equitable, and environmentally sustainable place to live. This approach looks different for every community, but can help neighborhoods of any kind flourish, make towns and cities competitive in a 21st-century economy, and improve lives by improving neighborhoods. Furthermore, smart growth means reinvesting in America’s downtowns and Main Streets, the economic engines of big cities and small towns alike. Smart growth means creating homes for families of all income levels alongside one another. Smart growth means diversifying our transportation system so Americans have a choice in how they get around.

However, to implement a smart growth plan for a city, we must first establish a standard to measure the city’s current success degree of smart growth. This measure needs to cover the 10 principles of smart growth and covers all aspects of urban development. Since 1990s, only a small number of scholars established mathematical models to analyze how to evaluate a city’s success of smart growth. Therefore, in order to meet the needs of The International City Management Group (ICM), we have several understandings about the given tasks shown in below.

- The most important task is to define a metric to measure the success of smart growth of a city. This metric should be based on smart growth principles. Based on this metric, we can accurately estimate or predict the success of smart growth of a city.
- We should collect the current development plans of two medium-sized cities. And these two selected cities should be on different continents. We assume the proposed growth plans remain unchanged in the planning period.
- For the current plans of two cities, we should analyze their impact on the proposed metric. These effects may be enhanced or weakened. Most importantly, we need explore the influence such as geographical location on urban smart growth. Finally, we propose initiatives to the current growth plans or redesign current plans.
- Similarly, for each of our proposed planned items, we need explore which proposed

initiatives contribute to the smart growth metric most. And then we should rank the proposed initiatives as the most potential to the most potential.

- Finally, we should discuss the influence of population growth on our model. We need to discuss the impact of explosive population growth on our proposed plan. On the other hand, we should discuss the impact of population growth on the metric we define. Finally, we can revise plans according to this extreme case for two mid-sized cities.

1.3 Assumptions and Hypothesis

To establish our mathematical model, we have several assumptions below.

- We assume the values of any indicate are from a normal distribution.
- The population density is inhomogeneous in a city.
- The government of two mid-sized cities has sufficient funds to support their current growth plans.
- The city's current growth plans will not change in a time range.
- The missing data can be estimated and filled by interpolation method.

1.4 Detailed definitions

In Table 1, we provide the variable definitions with its specific meanings of our model.

Table 1: Variable definitions of our model

Name	**Definition**	**Denotation**
Smart growth metric	A metric that measures the success of smart growth of a city	$\boldsymbol{SGM}$
Population density	An indicator that describes the number of people per square kilometer of a year in a city	$\boldsymbol{X_1}$
Ratio of industrial land to housing land	An indicator that describes ratio of industrial land to housing land of a year in a city	$\boldsymbol{X_2}$
Percentage of unemployment	An indicator that describes percentage of unemployed persons of total population of a year in a city	$\boldsymbol{X_3}$
Average wage per person (the unit is USD)	An indicator that describes average income per person of a year in a city	$\boldsymbol{X_4}$

Green area (hectares) per 100,000 persons (core)	An indicator that describes the amount of green area per 100,000 persons of a year	X_5
Annual number of public transport trips per person (core)	An indicator that describes the average number of bus trips per person per year in a city	X_6
Average price for urban housing	An indicator that describes the average prices for urban housing of a year in a city	X_7
SGM contributing variable (principal component)	A variable that affects smart growth metric (SGM) with statistical proof	$F_1, F_2 \ldots F_m$
Cumulative contribution rate	A variable which is made up of contribution rates of multiple principal components ($F_1, F_2 \ldots F_m$)	$G(m)$
Growth rate of ***SGM***	A variable that describes the growth rate of ***SGM*** of a city	K
GDP	A variable that measure the economic level of the city of a year	X_{gdp}
Livability indicator	A variable that describes the degree of livability of the city of a year	X_{liv}

1.5 The Advantages of Our Model

There are several advantages in our model.

- The proposed method is a unified framework with high flexibility and portability. Although the current model focuses on constructing metrics and specific plan for smart growth, different sections of this method can be easily transplanted or incorporated into a new model to tackle other applications.
- Anomaly detection is used to identify items or observations which do not conform to an expected pattern. This step in preprocessing to remove anomalous data often results in a statistically significant increase in accuracy in data mining. In our task, it helps us to obtain a robust model even with the limitation of a small sample size.
- In empirical validations, we provide the model comparisons with several classical algorithms, such as logistic regression, quadratic regression, support vector regression. Our model achieves performance improvement over other methods.
- We study the dynamics of the model on two representative mid-sized cities. From our experiments, we could find our metric and its corresponding prediction is consistent with the existing evaluation results of smart growth of these cities. And our model could be easily extended into a longer time range.

2 Our Work

Smart growth focuses on building cities that embrace the E's of sustainability—Economically prosperous, socially Equitable, and Environmentally Sustainable. In this section, we explain how we implement smart growth principles into city design around the world.

First of all, we define a novel smart growth metric called ***SGM*** to measure the success degree of smart growth of a city. Specifically, we identify seven important indicators which are closely related with the principles of smart growth, and then collect data from 16 cities with these indicators. We use the outlier detection method to clean up our data. Then, the principle component analysis method is used to reduce the dimension of the original data. By comparing the results from multiple regression models, we construct our prediction model to represent and calculate ***SGM*** based on linear regression model.

As ***SGM*** can be used to evaluate the level of smart growth, we use ***SGM*** to analyze the current plans of two selected cities. We also recommend our new smart growth plan by considering the initiatives for each selected city by taking into consideration of their different initiatives.

The flow chart of our model is presented in Figure 1 below.

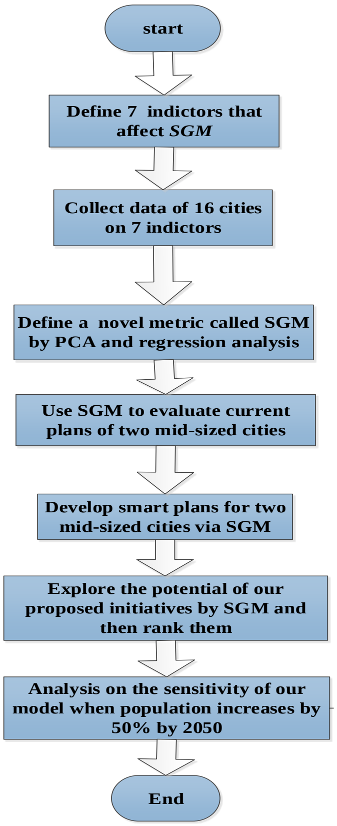

Figure 1: The flow chart of our model is presented in Figure 1 below

3 Task 1: Construction of the Smart Growth Metric

In this section, we derive a composite index, called the ***SGM* (smart growth metric),** to measure the success of smart growth of a city. As we assume above, the smart growth metric considers the three E's of sustainability and the 10 principles of smart growth theory [6].

3.1 Data Collection

First, we need to determine which indictors have a direct impact on ***SGM*** (smart growth metric). Take 10 principles of smart growth theory [2] or the three E's of sustainability into consideration, we select 7 smart growth indicators in this model to construct the smart growth metric. These indicators are defined as $X_1, X_2, \ldots X_7$. In **Table 2**, we provide the data sources of these seven indicators. Empirical studies have shown that these seven indicators are in line with the 10 principles of smart growth [2].

Table 2: The Data Sources of 7 Smart Growth Indicators

Indicators	**Data Sources**
Population density	United States Census Bureau
Ratio of industrial land to housing land	World Council on City Data
Unemployment rate	United States Census Bureau & World Council on City Data
Average wage per person(the unit is USD)	UN Data Center
Green area (hectares) per 100 000 population(core)	United States Census Bureau & UN Data Center
Annual number of public transport trips per capita(core)	World Council on City Data
Average prices for Urban housing	United States Census Bureau & World Council on City Data

As we know, there is no empirical study provides the logical relationship between the seven indicators and ***SGM***. To determine the relationship between the seven indicators and the smart growth metric ***SGM*** and construct the computational model, we collect data of 16 cities from three continents on 7 indicators. To make sure the data are in the same scale, a normalization procedure is used transform the value ranges from 0 to 100. In **Table 3**, we provide the original data of 16 cities on 7 indicators.

To construct the smart growth metric, we construct another index based on the three E's of sustainability and the 10 principles of smart growth. This index tries to combine the measure of the social equity, the environmental sustainability, and the economic prosperity. It is used as the ground truth of the smart growth metric. This index is defined as follows:

$$T_{output} = \frac{X_{gdp} + X_{liv}}{2} \tag{1}$$

where $\boldsymbol{X}_{\text{liv}}$ is the livability indicator of the city and $\boldsymbol{X}_{\text{gdp}}$ is the GDP of a city in a year. We simply use their linear combination to express the development level of smart growth.

Table 3: Original data of 16 cities on 7 selected indicators

City	Country	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Toronto	Canada	4430	1.37	0.0952	$54,234	445.67	201.9	$226
Amsterdam	Netherlands	4994	1.18	0.08	$50,352	375.59	255.18	$5,798
Barcelona	Spain	15777.	1.18	0.1701	$43,757	180.59	441.86	$4,236
Boston	USA	5100.13	2.62	0.053	$90,167	240.4	214.87	$9,225
Los Angeles	USA	2983.46	1.22	0.077	$81,214	376.23	53.07	$5,172
London	UK	5341.74	1.65	0.0599	$46,425	871.89	563.03	$18,777
Buenos Aires	Argentina	14450.8	1.47	0.0436	$4,541	62.46	725.75	$2,528
Amman	Jordan	3800.88	0.83	0.1005	$34,470	10.47	2.3	$1,309
Beijing	China	802.11	2.94	0.009	$13,817	519.12	391.992	$4,798
Tianjin	China	842.47	2.70	0.074	$10,119	283.88	167.041	$1,745
Suzhou	China	770.31	2.51	0.016	$9,197	643.96	192.085	$1,744
Shanghai	China	2259.21	0.88	0.040	$13,361	913.06	199.110	$3,635
Shenzhen	China	1554.68	3.00	0.008	$11,305	3232.93	880.669	$3,126
Dalian	China	470.38	1.99	0.061	$9,165	609.63	349.437	$1,745
Guangzhou	China	1119.6	2.46	0.091	$10,989	1926.48	386.096	$2,472
Qingdao	China	685.76	2.76	0.053	$8,813	766.89	276.857	$1,963

3.2 The Flowchart of Smart Growth Metric Construction

In this section, we will introduce the details in each step of our methods to construct the smart growth metric ***SGM***.

- In the first step, we use the outlier detection method to detect the anomalous data. Outlier refers to the items that are significantly different from other values.
- The second step is to apply PCA to reduce the dimension of our data and determine the principal components from our indicators.
- In the third step, we apply various regression methods to learn and construct the smart metric.

3.3 Outlier Detection

In statistics, an outlier [8] is an observation point that is distant from other observations. An outlier may be due to variability in the measurement or it may indicate experimental error. Anomaly detection (also outlier detection) is the identification of items, events or observations which do not conform to an expected pattern or other items in a dataset [8]. This step in preprocessing to remove anomalous data often results in a statistically significant increase in accuracy in data mining. In our task, it helps us to obtain a robust model even with the limitation of a small sample size. In this section, we use outlier detection method to discover anomalies and clean up data.

We assume the values of any indicate are from a normal distribution, and we identify observations which are deemed “unlikely” based on mean and standard deviation. Specifically, we remove the data whose indicator value is out of 3 standard deviations from the mean. According to our calculation, we remove the data from London and Shenzhen. The value of X_5 for Shenzhen is much larger than other values. The same situation happens in the value of X_7 for Shenzhen.

3.4 Dimensionality Reduction on Smart Growth Indicators via PCA

Principal component analysis (PCA) [7] is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Although the dimension of our data is not very high, the small number of sample size makes us difficult to analyze the relationship between indicators and furtherly construct an effective metric.

Therefore, in this section, we use PCA to reduce the dimension of our data and determine the principal components from our indicators. Specifically, the PCA algorithms used in our model consists of the following steps:

Step 1: Calculate the covariance matrix of data of 14 cities. The covariance matrix Σ has 14 rows (14 cities) and 7 columns (seven indicators).

$$\sum = \left(S_{ij}\right) p \times p \tag{2}$$

where

$$S_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_{ki}-\overline{x}_i\right)\left(x_{kj}-\overline{x}_j\right) \tag{3}$$

And p is the dimension of covariance matrix.

Step 2: Calculate the Eigenvalues λ of Σ and the orthogonalization of unit eigenvectors a_{ij}.

Step 3: Choose the principal components:

At this step, we select several principal components based on the equation below. Let the principal components be $F_1, F_2, \ldots F_m$.

$$G(m) = \sum_{i=1}^{m}\lambda_i \Big/ \sum_{k=1}^{p}\lambda_k \tag{4}$$

According to PCA method, the top M (where $M < 7$) principal components are mutually independent. When the cumulative contribution rate $G(m)$ is greater than **95%**, we can assume that the top M principal components are enough to represent all the characteristics of the original data. In our experiments, we use PCA algorithm to reduce the dimensions from 7 to 2.

3.5 Determine SGM Function by Regression Methods

In this section, we try to construct different regression methods to determine the SGM function based on the indicators. Here, we build different regression models including linear regression model [9], quadratic regression model, and support vector machine regression model (SVR) [10] to explore the relationship between 7 indicators and existing values of SGM for different cities. Our purpose is to find an optimal regression model which achieves lowest the average error.

To learn the regression model, we apply the leave-one-out cross-validation to give an almost unbiased estimator of the generalisation properties of our model. The difference between the predicted value and the actual value is calculated to evaluate different models. All the statistical experiments are repeated for five times, and the average results are reported. The average predicted error is given in Table 4. In this table, we demonstrate the results of various regression models with or without dimensionality reduction.

Table 4: The average predicted error comparisons of different regression models

	Regression models	**The average error**
No dimensionality reduction	Linear regression	29.0576
	Support vector regression (***Linear kernel***)	29.0384
	Support vector regression (***Gaussian kernel***)	15.8562
Using ***PCA*** for dimensionality reduction	Linear regression	10.0807
	Quadratic regression	14.4157
	Logistic regression	10.3070
	Support vector regression (***Linear kernel***)	9.8799
	Support vector regression (***Gaussian kernel***)	13.2811

From **Table 4**, it is obviously that using PCA for dimensionality reduction could significantly improve our performance. We can also find the linear regression and the SVR model with linear kernel can achieve comparable performance. Since SVR has no explicit equation, we select the linear regression method to construct our final evaluation criterion ***SGM***. The fitting results of the equation are shown as follows:

$$SGM = 20.46 + 0.62 \times F_1 + 0.11 \times F_2 \tag{5}$$

In this equation, the parameter $\boldsymbol{F_1}$ is the first principal component and $\boldsymbol{F_2}$ is the second principal component.

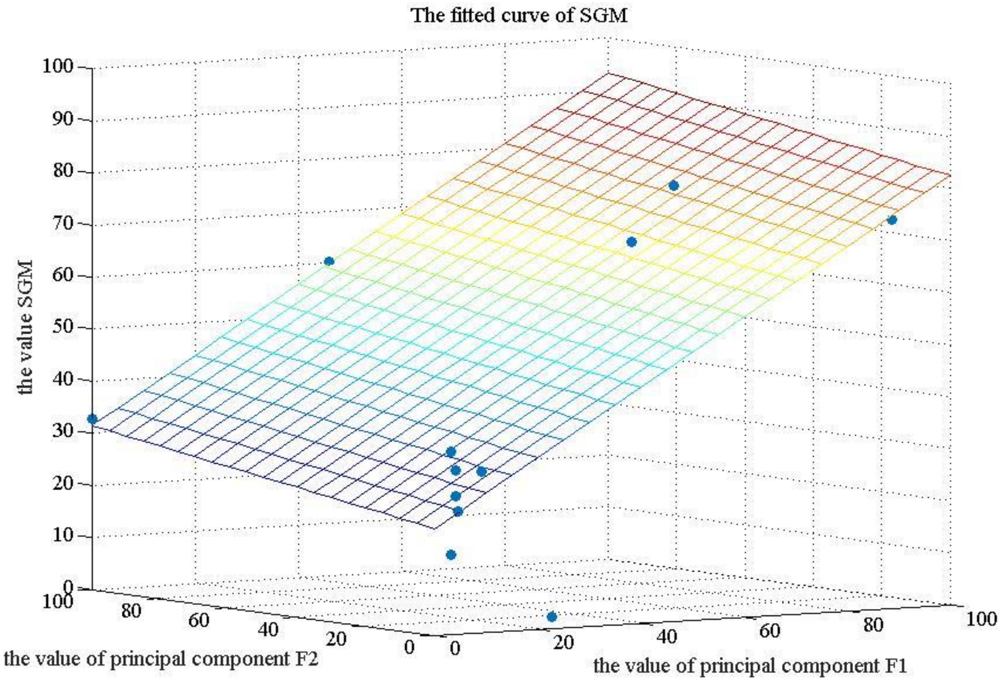

Figure 2: The best-fitted plane of all data for ***SGM*** prediction.

We show the fitted curve of all data in three dimensions in **Figure 2**. In this figure, we can find that the first principal component and the second principal component could effectively construct a two-dimensional plane to represent the value of **SGM**.

As a result, we can easily use the function of ***SGM*** to evaluate the successful degree of smart growth of a city. The greater the value of SGM is, the better the smart growth of the city is. To verify the generalization ability of our proposed evaluation metric ***SGM***, we also compare the ranking results with the World City Livable Index [3]. Our results show that the predicted result is highly consistent with this widely acknowledged ranking index.

4 Task 2: Evaluation the Current Plans of Two Mid-Sized Cities via SGM

In the previous section, we have constructed a effective evaluation metric and its function called ***SGM*** to measure the degree of urban sophistication or smart growth. Then, in this section, we use ***SGM*** to measure the current plans of the two selected cities. As a result, we can determine the level of smart growth of the current plans. Finally, we discuss how the current plans of the two cities meet the ten principles of smart growth theory.

4.1 Select Two Mid-Sized Cities and Analyze Current Growth Plans

In this section, we will analyze the growth plans of two mid-sized cities. To meet the International City Management Group (ICM)'s requirement, the population of two mid-sized cities is between 100,000 and 500,000 persons and they should be located in different continents.

In the next step, we obtain the current development plans of the selected city from the National Bureau of Statistics of China and the official website of the Municipal Government of Sao Paulo. The development plans of these cities cover various fields such as economic growth planning, environmental renovation, urban land use reform and improving social employment. The specific information of two selected cities is shown in **Table 5.**

Table 5: Specific information of two selected mid-sized cities

City	State	Country	Continent	Population (in 2016)
Saint Paul	Minnesotain	USA	North America	273535
Chaozhou	Guangdong	China	Asia	426158

We obtain growth plans of the two selected cities—Chao Zhou & Saint Paul from their official websites [4][5]. Their growth plans are all ranged from 2010 to 2020. They have carried out a detailed planning on multiple aspects like economic development, social equity, environmental protection. After we obtain the specific growth plan, we digitize these two cities' growth indicators and map them into our seven indicators. The specific plans of two cities are shown in Table 6.

Table 6: Detailed growth plans of the two mid-sized cities

City	Year	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Chaozhou	2015	865	0.968	0.024	2741.52	38.85	231.2	872.4
	2020	---	0.822	---	13813	45	293.7	---
Saint Paul	2015	3157.30	0.251	0.07	67,407	62.42	201.3	3,448.50
	2020	3661.01	---	0.152	---	70.97	201.3	3,448.50

Note: the symbol of --- means the corresponding data is missing for the corresponding indicator.

Actually, the current growth plans of these two cities cannot cover all data in these seven indicators. However, in our model, we need to get the values of all seven indicators of a city for that year to calculate ***SGM***. For example, the indicators of X_2 (ratio of industrial land to housing land), X_4 (average wage per person ($)) are missing. In our experiments, we try to use the interpolation method to estimate and fill those missing parts.

4.2 Measure and Analyze the Current Plans

We have obtained and compiled the current plans of two mid-sized cities. Therefore, we need to measure and discuss how their plans meet the 10 principles of smart growth theory. In **Table 6**, we can see that in the 2020 growth plan of Chaozhou, the indicators of X_2 (ratio of industrial land to housing land), X_4 (average wage per person), X_5 (green area(hectares) per 10000 persons) and X_6 (average number of bus trips per year) are all increased. Therefore, for Chaozhou City, we will discuss and measure how it meets the 10 principles of intelligence.

- **Meet the principle of creating a range of housing opportunities and choices.**

Smart growth policies and practices that advocate more compact and mixed-use communities, more transportation options, and the preservation of green space have the impact on the energy consumption in multiple ways. Another important strategy is to provide a variety of transportation choices. According to Chaozhou's current plan, the value of ratio of industrial land to housing land will decrease from 0.968 to 0.822. That means there will be more residential areas in the city.

- **Meet the principle of preserving open space, farmland, natural beauty, and critical environmental areas.**

Smart growth means reimagining the places we have already built, and protecting our open green spaces for generations to come. According to Chaozhou's current growth plan, the value of green area (hectares) per 10000 persons will be increased from 38.85 to 45 by 2020. This means that Chaozhou will provide more open space and green space without expanding the size of the city. This is in line with the principle of smart growth.

Similarly, we also measure Sao Paulo about how its current plan meets 10 principles of smart growth.

- **Meet the principle of creating a range of housing opportunities and choices.**
  According to the current plan of Saint Paul, the value of ratio of industrial land to housing land will decrease from 0.07 to 0.152. That means there will be more residential areas in the city.

Next, based on our metric ***SGM,*** we can measure the level of smart growth of a city. As we have concluded in task 1, ***SGM*** is used to measure the level of smart growth of a city. On the other hand, we define a variable *K*. This variable *K* can be used to represent the rate of the improvement of the level of smart growth.

$$K = \frac{SGM_{y_1} - SGM_{y_2}}{SGM_{y_2}} \times 100\% \tag{6}$$

where $\boldsymbol{SGM}_{y1}$ is the value of start year of current plan and $\boldsymbol{SGM}_{y2}$ is the value of ending year of current year. In Table 7, we will show the value of ***SGM*** and the growth rate of ***SGM***. We can draw a conclusion that the smart degree of Saint Paul is higher than that of Chaozhou both in 2015 and 2020. However, the growth rate of ***SGM*** of Chaozhou is higher than that of Saint Paul.

Table 7: The growth rate of SGM

City	SGM_{2015}	SGM_{2020}	*K(growth rate of SGM)*
Chaozhou	21.3559	28.7794	34.76%
Saint Paul	66.5847	74.7259	12.23%

5 Task 3: Develop Smart Plans for Two Mid-Sized Cities Based on SGM

In this section, to meet the need of the International City Management Group (ICM), we should propose several plans for both cities. Based on the metric ***SGM*** we have defined, we propose initiatives to the current growth plans or redesign current plans.

5.1 Proposed Plans for Chaozhou City via *SGM*

First of all, we collect the data of seven indicators of the Chaozhou City from 2005 to 2012. Then, in order to analyze which indicators have great impact on the smart growth of Chaozhou City, we increase the growth rate of each indicator by 50%. Through the control variable method, we can get the smart growth success rate of Chaozhou City in 2020. The growth rate of smart growth in 2020 is shown below.

Table 8: The growth rate of smart growth of indicators

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	*APVB*
Growth rate	-2.1%	19.8%	-7%	8.4%	6.3%	5%	10.5%	3.2%

Note: APVB means Annual Passenger Volume of the Bus.

From Table 8 we can find that the increases in X_2 (personal wages) can achieve the most promotion of the smart growth of Chaozhou. Taking the geographical location of Chaozhou City into account, it has a superior advantage of the port trade. In addition, this geographical advantage meets the 5th principle of smart growth. On the other hand, the city is close to high-speed rail stations and airports, which provides people of Chaozhou more choices of transportation. Based on the above analysis, we develop our growth plan for Chaozhou.

- Strengthen cooperation with the neighboring coastal cities and create a more open maritime market and port trade.
- Rebuild the ancient area of the city and promote traditional culture vigorously. In

addition, the government should expand the green area to attract more tourists.

- The government should formulate plans to rebuild the old factory. In addition, some of the old buildings in the city center can be converted into housing, which will increase the number of affordable housing.
- The government should invest some funds to support the development of the Internet industry.

After the implements of these specific plans, the growth rate of several indicators is expected to be changed as follows.

Table 9: The expected growth rate of each indictor after the implements of the proposed plans

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	***APVB***
Gaining rate	0.00014	1258	75	1.6	0.7619	0.1179	27	118.8
New gaining rate	0.00004	2500	75	2.0	1.0	0.1179	35	130
Rate of change	-71.4%	98.7%	0.0%	25%	31.2%	0.0%	29.6%	9.4%

Note: ***APVB*** means Annual Passenger Volume of the Bus.

After the implements of our proposed initiatives, the value of ***SGM*** of Chaozhou City in 2020 is 67.23, which achieves 29.8% improvement compared with previous plans.

5.2 The proposed Plans for Saint Paul City via *SGM*

Similarly, we will propose a redesigned plan for Saint Paul via ***SGM***. First of all, we collect the data of seven indicators of Saint Paul from 2005 to 2012. Then, in order to analyze which indicators have greater impacts on the smart growth of Saint Paul, we increase the growth rate of each indicator by 50%. Through the method of control variable, we can obtain the smart growth success rate of Saint Paul City in 2020. The growth rate of smart growth in 2020 is shown in below.

Table 10: The growth rate of smart growth of seven indicators

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Growth rate	3.20%	-2.10%	-1.91%	18.27%	10.3%	4.31%	-10.9%

From the content shown in Table 10, it is clear that the average wage of per person will significantly improve the evaluation value of ***SGM***. Saint Paul is an important shipping port in USA. It is surrounded by airport and high-speed ways, which meets the 5th and 8th principles of smart growth. Besides, Saint Paul is the main livestock processing center in the US. If the government makes use of land resources and improves the land utilization, we expect the economy of Saint Paul will grow quickly.

This strategy meets the 1st principle of smart growth. Based on the above analysis, we can develop a grow plan for Saint Paul City.

- The Government should actively trade with neighboring cities, which will increase the city's income and provide more employment opportunities for the residents.
- Urban planning department should apply rational plans on land using and begin to recycle of those abandoned land. This measure is expected to increase the area of residential land to 20%.
- Due to the fact that high housing price has a negative impact on the ***SGM***, the government should reduce the housing prices. Our experiments show that it is better to reduce housing prices by 50%.

After the implements of these strategies above, the growth rate of several indicators is expected to be changed as shown in Table 11.

Table 11: The expected growth rate of each indictor after implementation of proposed plans

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Gaining rate	-2.29	0.000301	0.004653	5318	0.4271	9.131	217.4
New gaining rate	-2.29	0.003612	0.0041	6807.04	0.4271	9.131	108.7
Rate of change	0.0%	20%	-12%	28%	0.0%	0.0%	-50%

Note: ***APVB*** means Annual Passenger Volume of the Bus.

After the implements of our proposed initiatives, the value of ***SGM*** of Saint Paul City in 2020 is 32.4, which is 12.1% higher than previous plans.

6 Task 4: Ranking the Proposed Initiatives via *SGM*

We have already developed a series of individual initiatives for both cities in the above section. In this section, we will rank these individual initiatives of the two cities separately. The principle of ranking these individual initiatives is based on the value of ***SGM***. The greater the value of ***SGM*** is, the greater potential the initiative is. Next, we calculate the value of ***SGM*** of different initiatives we propose by the control variable method. Finally, by sorting the values of the ***SGM***, the final ranking results are obtained. The ranking result of Chaozhou City is presented in Figure 3. And the ranking result of Saint Paul City is presented in Figure 4.

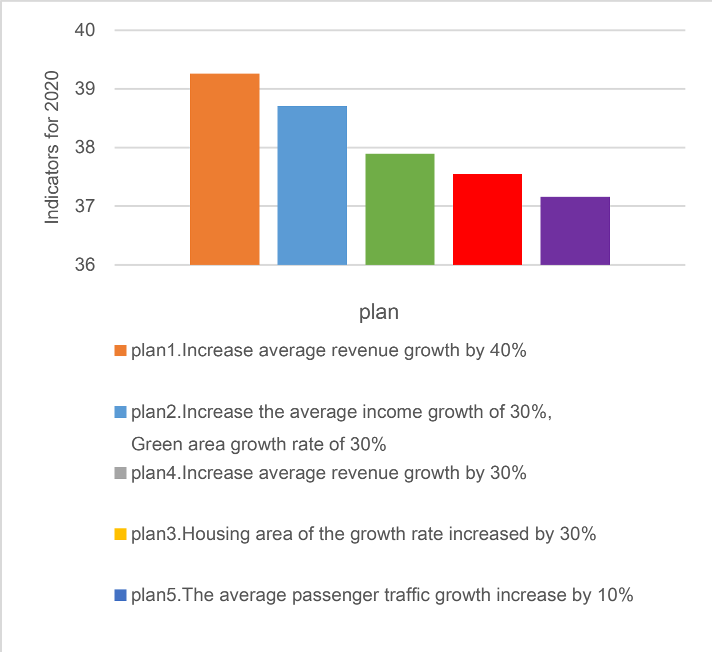

Figure 3: the ranking result of 5 proposed initiatives of Chaozhou City

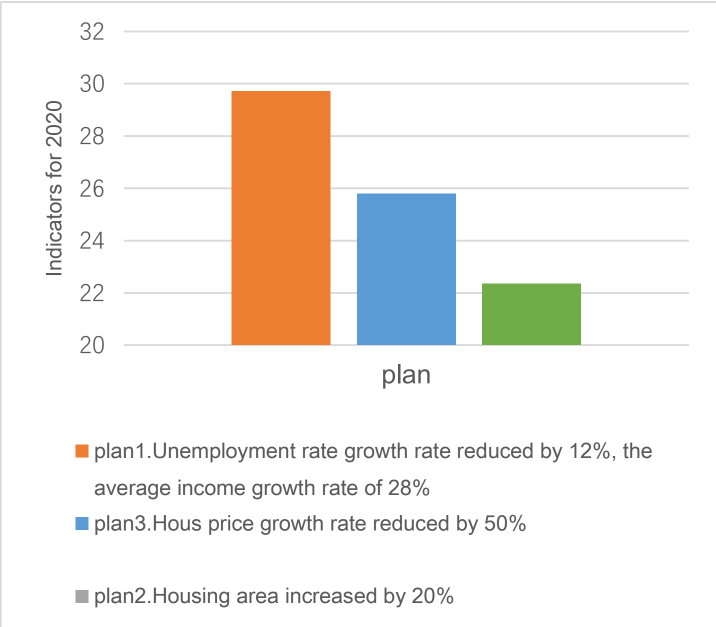

Figure 4: the ranking result of 3 proposed initiatives of Saint Paul City

7 Strength and Weakness

7.1 Strengths of our model

Our proposed model shows good generalization ability, robustness, and dynamics. First, the proposed method is a unified framework with high flexibility and portability. Although the current model focuses on constructing metrics and specific plan for smart growth, different sections of this method can be easily transplanted or incorporated into a new model to tackle other applications.

Second, anomaly detection is used to identify items or observations which do not conform to an expected pattern. This step in preprocessing to remove anomalous data often results in a statistically significant increase in accuracy in data mining. In our task, it helps us to obtain a robust model even with the limitation of a small sample size. Moreover, in empirical validations, we provide the model comparisons with several classical algorithms, such as multiple linear regression, support vector regression and so on. Our model achieves performance improvement over other methods.

Third, the proposed model is not a statical model. We study the dynamics of the model on two representative mid-sized cities. From our experiments, we could find our metric and its corresponding prediction is consistent with the existing evaluation results of smart growth of these cities. And our model could be easily extended into a longer time range.

7.2 Weaknesses of our model

First, we meet some difficulties in collecting data of mid-sized cities, especially the green area ratio in Saint Paul from 2004 to 2010, which is an indicator that describes the amount of green area per 100,000 persons per year. In our experiments, we try to use the interpolation method to fill those missing parts. Although in our experiments, we have showed that the interpolation is useful to calculate the proposed metric, the lack of these data makes it difficult to determine the accurate value of the metric and the corresponding best-fit plan.

Second, the improvements of the new techniques are not considered in our model. In our model, we identify and define seven important factors to construct the proposed metric. Although these factors are selected according to three E's of sustainability and 10 principles of smart growth, and the empirical validation demonstrate they are useful to predict the level of smart growth, all of these factors are extracted based on the past statistical data. The past is not the future! In future, the new technologies, such as AI, will definitely change our cities, towns, and neighborhoods, and even the existing definitions, strategies and principles of smart growth. Therefore, our meaningful future work is to consider the influences of the new technologies on smart growth.

References

[1] EPA, “This is Smart Growth.” 2016 https://www.epa.gov/smartgrowth/smart-growth-publication

[2] Duany, Andres, Jeff Speck and Mike Lydon. The Smart Growth Manual. McGraw Hill. 2010.

[3] The definition of livability. http://media.heraldsun.com.au/files/liveability.pdf

[4] The plan of Saint Paul by 2020, USA. https://www.stpaul.gov/government

[5] The plan of Chaozhou by 2020, China. http://www.chaozhou.gov.cn/index.jhtml

[6] Shaw R P. The impact of population growth on environment: the debate heats up. [J]. Environmental Impact Assessment Review, 1969, 12(1-2):11-36.

[7] Abdi H, Williams L J. Principal component analysis[J]. Wiley Interdisciplinary Reviews Computational Statistics, 2010, 2(4):433-459.

[8] R.H. Outlier Treatment in Data Merging[J]. Journal of Applied Crystallography, 1997, 30(4):421-426.

[9] Jaccard J, Turrisi R, Wan C K. Interaction effects in multiple regression. [J]. Journal of the Royal Statistical Society, 1991, 40(4):461.

[10] Clarke S M, Griebsch J H, Simpson T W. Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses[C]// ASME 2003 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers, 2015:1077-1087.