For office use only	Team Control Number	For office use only
T1 ________	**87843**	F1 ________
T2 ________		F2 ________
T3 ________	Problem Chosen	F3 ________
T4 ________	**B**	F4 ________

A Prediction Model Based on Multivariate Simulation

2018 MCM/ICM Summary Sheet

In order to accomplish this project, we have totally built five models with several algorithms.

Firstly, we came up with the gray prediction model and advanced logistic model to forecast the change of population over future 50 years because of the strong correlation between the population and the number of native language speakers.

After that, we started to use multivariate time series model to forecast the future quantity of the second language speakers. In this model, we assume that the number is almost completely determined by two variables which are GDP factors and Internet factors and the Multivariate polynomial regression is able to fit a quadratic surface eventually. What should be pay attention to is that these two features are not just the GDP of one country and the popularity of it. We transformed them into the percentage of the whole world as you can see in the text.

Then for the purpose of obtaining the future GDP factors and the Internet factors, our team firstly took the cubed exponential smoothing to get the short-term data.In addition, it is easily for us to combine the real data and the predicted data in short term to forecast the long-term change of these two variables by linear regression. After gaining the future GDP and Internet factors' data, they can be taken into the quadratic surface which was established by the above-mentioned regression and finally answer the question of the future quantity of second language speakers of various languages.

At this point, the total number of native language speakers and second language speakers is on our hand.

Next, the spotlight is focused on the changing geographical distribution of each popular languages. Analogy to the gravitation formula, it is natural for us to devise the derivative gravity model. Our team consider five variables as key factors which are the population of both countries, the ratio of two countries' GDP and the Area of the destination country. According to the convention, we apply the algorithm of stochastic gradient descent to train the parameters and finally obtain a second order asymmetric tensor T_{ij} which means the number of people migrating from i th country to the j th country. This tensor helps us to quantitatively described the situation of migration throughout the whole world.

At last, we found that analytical hierarchy process(AHP model) has the capability to reliably identify the importance rank of different variables. So it is an amazing tool for us to help decide which region is reasonable and valuable for a company to establish international offices.

KEYS: Gray Prediction Model, Advanced Logistic Model,Multivariate Time Series Model, Derivative Gravity Model, AHP

Contents

1 Introduction

1.1 Problem Background

In our current world, more than 6900 kinds of languages exist in the world. However, with the fast progress of the globalization, we have witnessed drastic changes of various languages.

On one hand, with the growth of population, the number of native speaker of some languages such as Mandarin has inclined. While the number of native speaker of some languages has declined because of the low birth rates. On the other hand, the number of world's population speaking a second language has also been influenced by some factors. These factors include: government policy, education, social influences, migration and assimilation of cultural groups, and the use of electronic communication and social media, international business relations, increased global tourism, and the use of technology of translation based on artificial intelligence.

1.2 Restatement of the Problem

The majoring problem we are faced with is how to use mathematical models to get us know about the future development of the languages used in large scale all around the world. Our purpose is to address language changes in the global areas from several interrelated perspectives and discern the impact these factors bring. The key point is that we need to find a way to qualify those factors and visualize the result geographically. Then, in the next part, we give our recommendations for the location of the international offices and the languages used in the office.

Two major problems are discussed in this paper, which are:

- What will be the distribution of various language speakers over time?
- Recommendations for locations of the newly opened international offices based on the model built.

1.3 Our Work

Because it is a problem of studying the world's language, we spent quite a bit of time initially looking for data and information. Data found include native speakers, second language usage rankings, world language distributions, demographic distributions, etc.

After collecting the data, we use the GM model and the logistic model to predict the top 10 world language users over time. For second language usage, we use multivariable regression algorithms.

Then we build a model to predict the immigration between some typical countries. Given the global population and human migration patterns predicted for the next 50 years, we use the AHP Model to evaluate each potential locations.

When the forecast was done, we use the heated picture to show which places are the most suitable or potential places to locate the international offices and which languages are used in the office.

1.4 Logic Framework of Our Paper

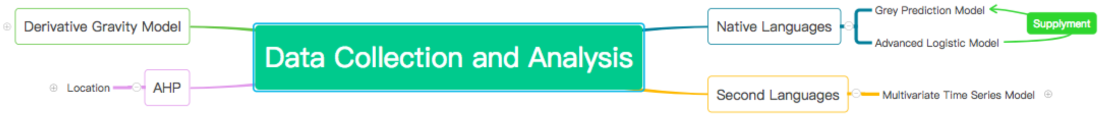

Figure 1: The Logic Framework of Our Paper

2 Preparation of the Models

2.1 Assumptions

Advanced Logistic Model

For the Advanced Logistic Model, due to limited data and to avoid overcomplicating the time series of the native language, we make the following assumptions:

1. The age distribution of the population in each of the countries concerned is the same.
2. Population policies are the same in all countries.
3. People of each country has the same expectation of having children.

Grey Prediction Model

1. Not considering the large-scale international migration
2. Igonring some events with low probability,like asteroid collisions.

Multivariate Time Series Model

1. The overall national strength of China and English-speaking countries as a whole will stay ahead of the next 50 years.
2. In the next 50 years, without any political turmoil or war in the major countries, the overall GDP growth will not be fiercely turbulent.
3. The surface we want to fit is a quadratic surface

2.2 Notations

The primary notations used in this paper are listed in **Table 1**.

Table 1: Notations

Symbol	Definition
P_i	the ith country's population
G_i	the ith country's GDP
A_i	the ith country's land area
$G^{(i)}$	the ith language's GDP
$I(n)$	the nth language's Internet users per 100 people
I_M	$I^{(1)}$, $I^{(1)}$ is the data of Mandarin
$I_{(E)}$	$I^{(2)}$, $I^{(2)}$ is the data of English

2.3 Population Visualizing

The number of native speakers is largely decided by the number of population of the country. To make the problem more clear, we visualize the distribution of population.

Figure 2: World Population(Part 1)

Figure 3: World Population(Part 2)

In the picture above, each red point represent a city or village with a relatively large residents.

2.4 Data Collecting

From[1], we get the number of people using the top 13 kinds of languages in the world. The data is shown as follows(millions):

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Chinese	1120	1036	1120	1120	1120	1120	1120	1115	1030	1090	1090
English	480	618	480	450	480	480	480	480	840	942	983
Russian	285	278	285	285	285	285	285	280	0	275	267
French	265	213	265	265	270	265	265	270	220	274	229
Hindi	250	487	250	350	250	250	250	0	160	208	544
Arabic	221	285	221	221	221	221	221	220	490	385	422
Portuguese	198	0	188	188	248	188	188	190	0	262	229
Bengali	195	207	185	185	185	185	185	0	7210	380	261
German	170	91	109	109	100	109	109	120	86	210	129
Japanese	133	127	133	133	133	130	130	130	130	130	129
Spanish	0	376	320	350	320	320	320	315	490	570	527
Indonesian	0	234	0	0	0	0	0	0	0	250	0
Punjabi	0	0	0	0	130	130	130	0	0	146	148

Then we visualize the data as follows:

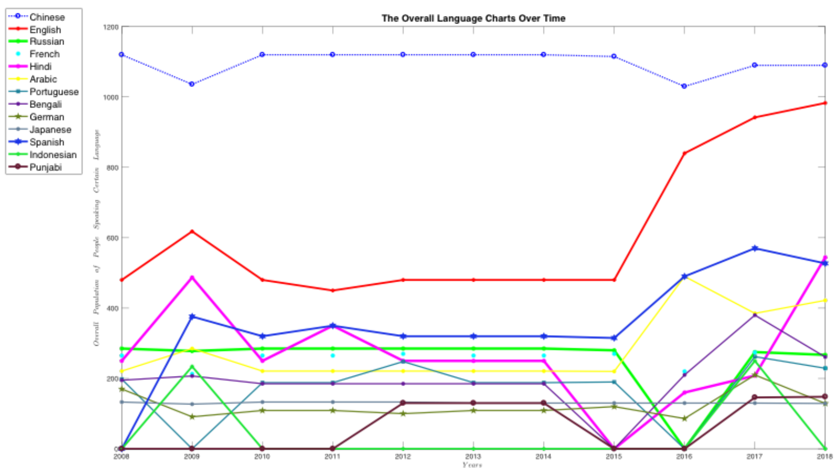

Figure 4: The Overall Language Chart Overtime

Then, we adjust the data according to the statistical viewpoint: we can use the Max Likelihood method to make up these missing data and draw the charts, where we delete two languages: Indonesian and Punjabi for we only have a little data of these two languages.

The components of the language in the current world is shown in the pie chart:

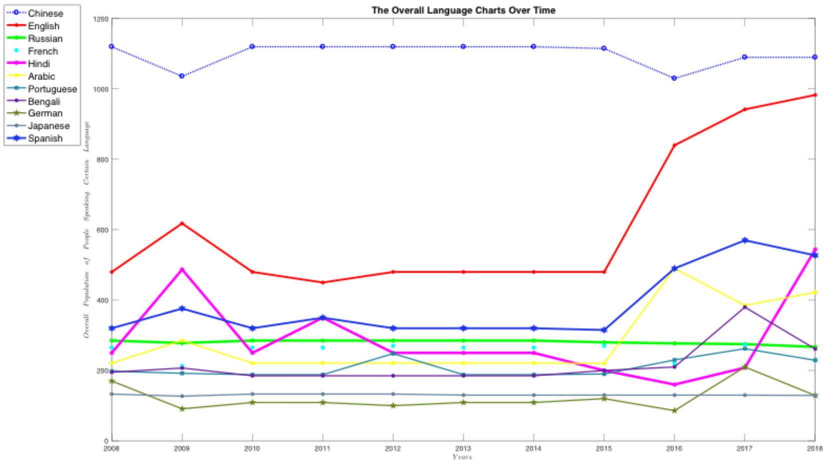

Figure 5: The Overall Language Chart Overtime(modified)

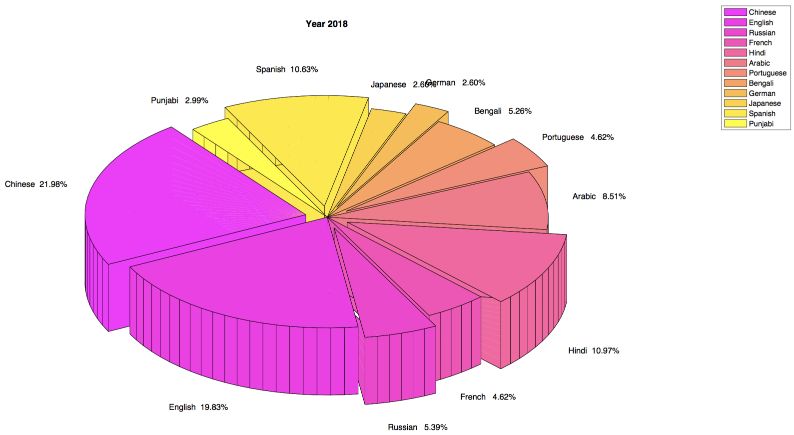

Figure 6: The Component of Language in the Current World

3 The Models

3.1 Gray Prediction Model

The GM(1,1) model is the this model's core. It is a first-order differential equation model predicted by a single variable, whose discrete time influence function approximates an exponential law.

The basic method of establishing a GM (1,1) model is: $X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), ...$ $X^{(0)}(N)\}$, which are the basic non-negative time series.

$X^{(1)}(t)$ is the accumulated sequence, which means

$$X^{(1)}(t) = \sum_{m=1}^{i} X^{(0)}(m), t = 1, 2, ..., n$$

The differential equation is $\frac{dX^{(1)}}{dt}+aX^{(1)} = u$, where a and u are two parameters to be identified. Let $\hat{a} = \begin{bmatrix} a \\ u \end{bmatrix}$, we can use least squares method to calculate the $\hat{a} = (A^TA)^{-1}A^TB$, where $A = \begin{bmatrix} -\frac{1}{2}(X^1(1)) + X^1(2) & 1 \\ -\frac{1}{2}(X^1(2)) + X^1(3) & 1 \\ ...... & \\ -\frac{1}{2}(X^1(n-1)) + X^1(n) & 1 \end{bmatrix}$,$B = \begin{bmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ ... \\ X^{(0)}(n) \end{bmatrix}$.

Then, the discrete function of Grey model is $X^{(1)}(t+1) = (X^{(1)} - \frac{u}{a})e^{-at} + \frac{u}{a}$.

Then we can get the prediction data: $\hat{X}^{(0)}(t+1) = \hat{X}^{(1)}(t+1) - \hat{X}^{(1)}(t), t = 1, 2, 3...n$

We use these time series to deal with the following analysis:$X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), ...X^{(0)}(N)\}$

We can calculate the level ratio: $\lambda(k) = \frac{x^{(0)(k-1)}}{x^{(0)}(k)}$, $\lambda = \lambda(2), \lambda(3), ..., \lambda(N)$

All the $\lambda(k) \in$ accordingly certain range, k=2,3...N, therefore, we can use these data to form our model.

$$\hat{a} = \begin{bmatrix} a \\ u \end{bmatrix} = (A^TA)^{-1}A^TB. \quad (1)$$

Then we can get the formulation of prediction, which we can take advantage of to continue our analysis.

3.2 Advanced Logistic Model

We know that the number of native speakers is closely related to the number of nationals using the language as their official language. Therefore, our group decided to use demographic factors to roughly compare the number of native language users over time.

In the demographic model, the growth rate will start to decline after the population reaches a certain number. It is noted that natural resources, environmental conditions and national medical conditions all have a retarding effect on the growth of the population. Although the role of population growth is getting bigger and bigger, the blockade is greater. The blocking effect is mainly reflected in the population growth rate, so we have equations:

$$\frac{dx}{dt} = r(x)x, \quad x(0) = x_0. \quad (2)$$

Make a simple assumption about $r(x)$:

$$r(x) = r - sx(r, s > 0). \quad (3)$$

Here r is the inherent growth rate (constant), then we define a country's maximum population X_m (the maximum of a country), and finally we take $r(x)$ into (1).

Then, we get

$$\frac{dx}{dt} = rx(1 - \frac{x}{x_m}), x(0) = x_0. \tag{4}$$

Then we use the method of separation of variables, it is easy to get the solution:

$$x(t) = \frac{x_m}{1 + (\frac{x_m}{x_0} - 1)e^{-rt}}. \tag{5}$$

Now we use this equation to fit in the countries which use top 15 languages as the official language or primary language so that we can get the number of people who use these language as their first language in several years.

Please note that for a language, some are native speakers and others use it as second languages. We have only discussed the changes in the number of native speakers over time, not involving the second foreign language. In the following sections, we use other models and algorithms to discuss the changes in the number of second language users.

3.3 Multivariate Time Series Model

In this model, we combine multivariate polynomial regression and linear regression to analyze the changing number of second language speakers over time. We initially consider only GDP(Gross Domestic Product), the rate of Internet penetration, and the number of population. We think these three factors can comprehensively reflect the potential of a language being chosen as a second language for other countries: GDP can reflect the economical influences of the regions using the language, the rate of Internet penetration reflects the level of cultural exchanges one language has, the number of population reflects the potential of the language to be chosen to learn by others. One thing deserves to be noted is that we get this data by adding the data of main countries that using this language. The specific list of countries is as follows:

Chinese: China

English: Australia, Canada, U.K., Ghana, India, Nigeria, U.S., South Africa

Russian: Russia

Urdu: Pakistan

Hindi: Pakistan

Arabic: Algeria, Egypt, Iraq, Morocco, Saudi Arabia, Sudan, Syrian, Tunisia, Yemen

Portuguese: Angola, Brazil, Mozambique, Portugal

Bengali: Bangladesh

Japanese: Japan

Spanish: Argentina, Chile, Colombia, Cuba, Ecuador, Spain, Guatemala, Mexico, Peru, Venezuela

The following is the note of the factors:

(1)GDP: The data can be found in the website of United Nations. To avoid the inaccuracy brought by the boost of GDP, we use the ratio of one language and the sum of all languages.

$$G_f^{(i)} = \frac{G^{(i)}}{\sum_1^8 G^{(i)}}. \tag{6}$$

Constrained by the data, we only consider 8 languages.

(2)Internet penetration rates: The data can be found in the website of United Nations as well, but after a simple estimation, we find that the penetration of the Internet in all countries in the world has exceeded 80% and then remain stable. Therefore, we can see that this change does not work very well in a long time. So we define a new kind of variable to reflect the level of Internet penetration. Details are as follows:

$$I_f^{(n)} = \frac{G^{(n)}}{I_M + I_E}. \tag{7}$$

The explanation for the definition: We write the denominator into the sum of English and Chinese instead of writing only the English or Chinese because we want to avoid making a variable of Internet penetration remains 1. Under the assumption, the new variable we define will change between 0 and 1.

After we have done the definition changes, we begin the work of multivariate polynomial regression.

In three-dimensional space, we firstly assume that what we want to fit is a quadratic surface, but due to the scarcity of data, we will eventually find out the existing data is the projection of the surface in a specific direction, that is, a curve.

The definition of the surface is as follows:

$$Z = \theta_0 + \theta_1 G_f^{(i)} + \theta_2 I_f^{(i)} + \theta_3 [G_f^{(i)}]^2 + \theta_4 G_f^{(i)} I_f^{(i)} + \theta_5 [I_f^{(i)}]^2 \tag{8}$$

$$\theta_j \quad are \quad all \quad constants, j = 0, 1, ..., 5.$$

Next, all the data can be expanded into the high dimension space$(x_1, x_2 to 1, x_1, x_2, x_1^2, x_1 * x_2, x_2^2)$, so that we turn the problem into a multiple linear regression problem.According to the convention, we use the gradient descent method to optimize the parameters. The specific formula is as follows:

$$h_\theta(x) = \begin{bmatrix}\theta_0, ..., \theta_5\end{bmatrix} \begin{bmatrix} x_0 \\ ... \\ x_5 \end{bmatrix} = \theta^T x \tag{9}$$

$$x^{(i)} = \begin{bmatrix} 1 & G_f^{(i)} & I_f^{(i)} & [G_f^{(i)}]^2 & G_f^{(i)} I_f^{(i)} & [I_f^{(i)}]^2 \end{bmatrix}^T \tag{10}$$

$$J(\theta) = J(\begin{bmatrix}\theta_0, ..., \theta_5\end{bmatrix}^T) = \frac{1}{2 \times 8} \sum_{i=1}^{8} (h_\theta(x^{(i)}) - y^{(i)})^2 \tag{11}$$

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta), j = 0, ..., 5. \tag{12}$$

After we get this surface, we project it to the existing GDP factor and Internet factor so that we can get several curves of different languages. The result is shown on the picture*

It should be aware that the curves for each language is on different flats. That is, they can't be shown on 2 dimension canvas.

After that, we want to predict the number of second language speakers in the future 50 years, according to our multivariate polynomial regression. But now we need for changes in GDP factors and Internet factors over the next 50 years.

Before exhibiting the algorithms, we have to explain why the predictions of the second language speakers in number over future 50 years are not directly based on time series analysis. In our model, the change of the number of native language speakers is predicted by time-series analysis, because the change in the number of native language speakers is mainly caused by the development of population, which can be easily understood. But there are many influencing factors affecting the number of second language speakers. In order to directly predict the change of second language speakers, we ought to find out the data of many years, but we don't. Therefore, we choose two main influencing factors to indirectly predict the potential of second language speakers. In addition, GDP data are completely recorded on UNdata website or other statistical stations, as well as the Internet factors, so the time-series analysis of these two changes will be accurate and simple.

For the purpose of simplifying the model, we still adopt regression and exponential smoothing algorithm to predict the change of these two variables with time.

Take China as an example. At this stage, the growth rate of China's GDP is 6% and 7%. The popularity of the Internet is also steadily rising with the rate of increase slightly declining. Then on the function graph, the previous phase should be linear and the last phase should be the same as the shape to the right of the sigmoid curve.Therefore, we first take cubed exponential smoothing to predict the trend from 2017 to 2020. Combined with this prediction, we can deduce the slope of the linear function, and also avoid the limitation of exponential smoothing method that can only predict the short-term limitation accurately.

The formula for cubed exponential smoothing is as follows:

$$S_t^{(1)} = \alpha x_t + (1-\alpha)S_{t-1}^{(1)} \tag{13}$$

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)} \tag{14}$$

$$S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha)S_{t-1}^{(3)} \tag{15}$$

$$x_{t+T} = A_T + B_T T + C_T T^2 \tag{16}$$

$$A_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \tag{17}$$

$$B_t = (\frac{\alpha}{2(1-\alpha)^2})[(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)}] \tag{18}$$

$$C_t = (\frac{\alpha^2}{2(1-\alpha)^2})[S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}] \tag{19}$$

$$S_0^1 = \frac{x_1 + x_2}{2} \tag{20}$$

$$variable \quad x_i \quad can \quad either \quad denote \quad G_f^{(i)} \quad or \quad I_f^{(i)}$$

After forecasting the change of GDP factors and Internet factors over time[2], we put the forecast result of the future 50 into the fitted surface by the polynomial regression algorithm, and get the potential diagram of the result of each number of second language speakers in the next 50 years. The result is shown on picture10.

In the graph, we can find that the number of second language speakers such as Mandarin and Arabic is on the rise. In contrast, the second language speaker's number of English and Russian decline, which seems a bit unreasonable. However, the existing data can be found some clues. In English-speaking countries, the popularity of the Internet is stable and confronting the rising popularity of other countries, its number of second language speakers will be extruded by other languages.

Looking at the decline of the Russian language, at this stage,Russia's economy is sluggish and the growth rate is slowing down. In other countries, the economy is rising strongly (especially China), it is also reasonable that the second language speaker's number of Russian decrease.

3.4 Derivative Gravity Model

Few people have use this kind of model to predict the flow or migration of people in the world. The Derivative Gravity Model is in analogy with Newton's law of gravity. In our model, T_{ij} denotes the number of people born in the ith country and later migrate into the jth country. Like Newton's law of gravity, we give the equation of T_{ij}:

$$T_{ij} = \frac{\theta_1 P_i^{\theta_2} P_j^{\theta_3}}{\frac{1}{A_j}(\frac{G_i}{G_j})^{\theta_4}} \tag{21}$$

$$lnT_{ij} = ln\theta_1 + \theta_2 lnP_i + \theta_3 lnP_j + lnA_j - \theta_4[G_i - G_j] \tag{22}$$

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (lnT_{ij} - lny_{ij})^2 \tag{23}$$

In this model, we first use Gradient Descent Method to fit the best $\theta_1, \theta_2, \theta_3, \theta_4$. Then we consider the population and GDP trends of the next 50 years in the United States, Russia, China, the United Kingdom, Saudi Arabia, Australia, Spain, Brazil and India. From the trend we can come to the result of the T_{ij} between these countries in next 50 years.

3.5 Analytical Hierarchy Process(AHP)

As we all know, the Analytical Hierarchy Process(AHP) is a quite useful and rigorous framework for dealing with multi-variable problems.

This method partly depends on the preference of the decision makers, which, to some extent, is a little subjective because the degree of importance of the factors is determined by them. Therefore, we add the consistency test to make the process more rigorous, largely decreasing the error.

The overall 1-9 scale for each variable is presented in the following table.

Scale	Meaning
1	importance of two factors: former equals latter
3	importance of two factors: former slightly larger than latter
5	importance of two factors: former larger than latter
7	importance of two factors: former quite larger than latter
9	importance of two factors: former significantly larger than latter
2,4,6,8	the median of adjacent judgments above
Reciprocal	$a_{ij} = \frac{the \quad importance \quad of \quad factor \quad i}{the \quad importance \quad of \quad factor \quad j}, then \quad a_{ij} = \frac{1}{a_{ji}}$

$$A = \begin{bmatrix} 1 & 1/2 & 1/7 & 1/5 & 1/5 \\ 2 & 1 & 1/4 & 1/3 & 1/3 \\ 7 & 4 & 1 & 2 & 3 \\ 5 & 3 & 1/2 & 1 & 1 \\ 5 & 3 & 1/3 & 1 & 1 \end{bmatrix} = (a_{ij})_{55}$$

A_{ij} stands for the ratio of influence a_i to a_j towards the location problem.

We can calculate $\lambda_{max}(A) = 5.073$

The consistency index $CI = \frac{max(\lambda)-n}{n-1}$, here $n = 5$, then $CI = \frac{\lambda_{max}-5}{5-1} = 0.018$, and we can find $RI = 1.12$

Therefore, $CR = \frac{CI}{RI} = \frac{0.018}{1.12} = 0.016 < 0.1$, which means our matrix can be accepted.

Then we standardize the eigenvector $U = (-0.8409, -0.4658, -0.0951, -0.1733, -0.192)$, and we can easily get the final weight $(0.126, 0.103, 0.051, 0.263, 0.475)$.

4 Analysis and Results

4.1 Number of Native Speakers

4.1.1 Grey Prediction Model

First, we try Grey Prediction model to get the prediction. The result is shown as follows (only results of English and Spanish are given):

4.1.2 Advanced Logistic Model

However, we think that the number of population cannot increase without limitation.

So, in this part, we use the logistic model which has been illustrated in Part3.

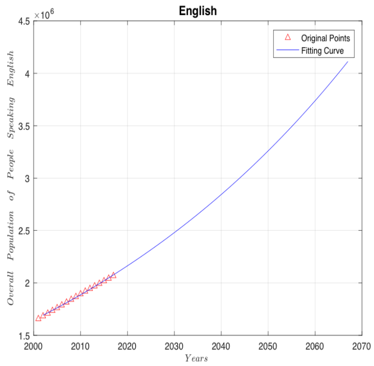

Figure 7: English

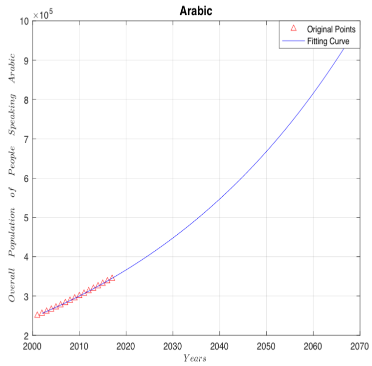

Figure 8: Arabic

Our object is to predict the number of population in a long time. We use the data we have to solve the equation (4) which is an algebraic equation. The prediction we get is as follows:

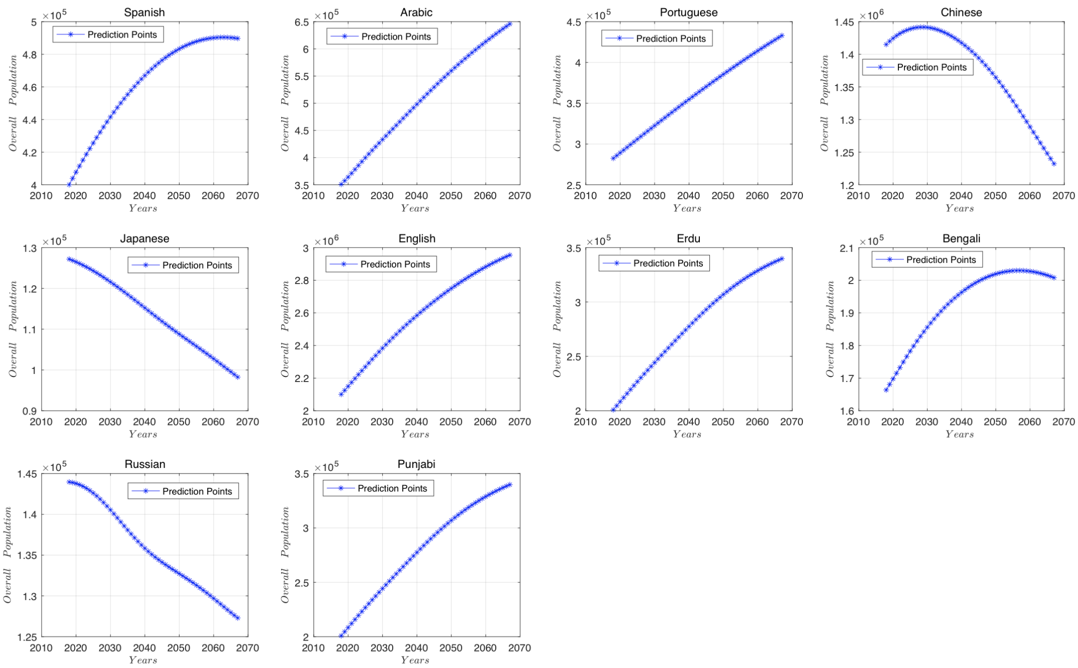

Figure 9: The Future Trend of Native Speakers Predicted By Logistic Model

4.2 Number of Second Language Speakers

We use the Multivariate Time Series Model and get the following result: Then we get the number of second language speaker along with time. The result is as follows:

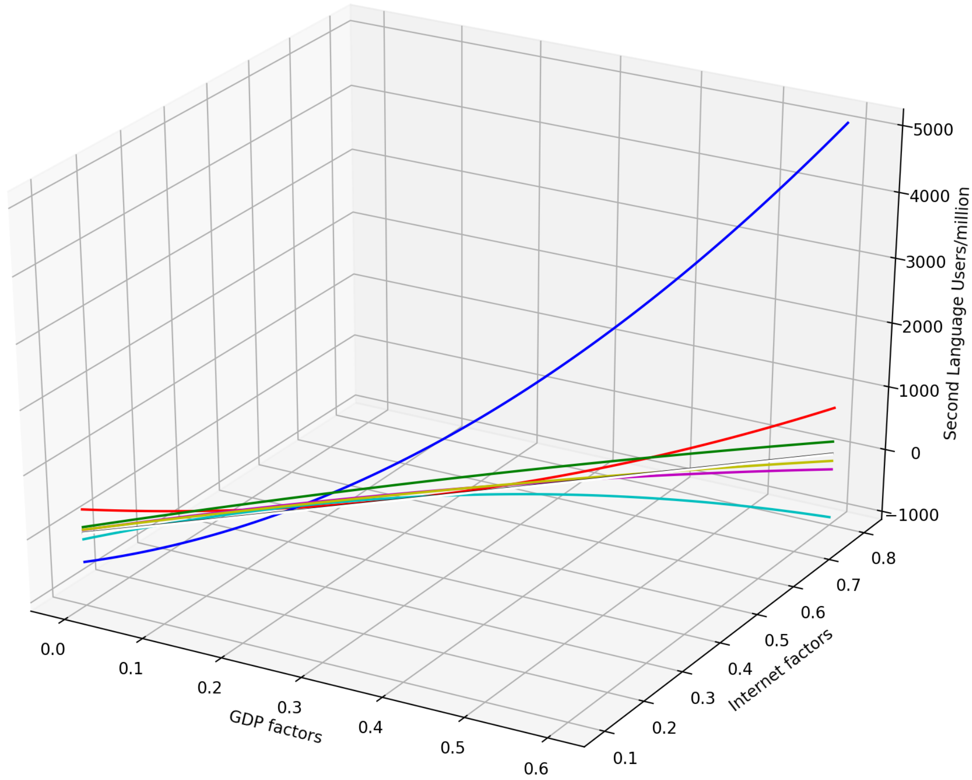

Figure 10: The Further Development Along With the Two Factors.

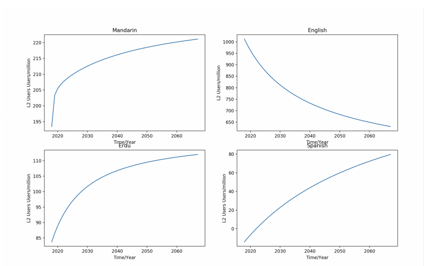

Figure 11: The trend of users along with time(part1)

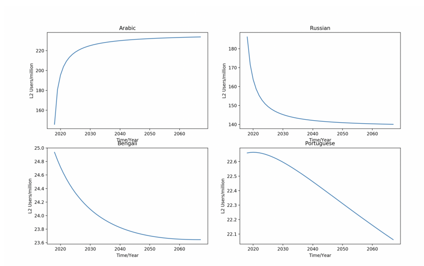

Figure 12: The trend of users along with time(part2)

5 Testing the Model

The Advanced Logistic Model

According to $x(t) = \frac{x_m}{1+(\frac{x_m}{x_0}-1)e^{-rt}}$, we can check our model from the following aspects.

We can first calculate the following second derivative $\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = r^2(1-\frac{x}{x_m})(1-\frac{2x}{x_m})x$

We can find three general circumstances.

(i) $\lim\limits_{t\to+\infty} x(t) = x_m$, and the x_m is the final limit, which is reasonable because the factors of environment, resources and some other limited ones.

(ii)When $0 < x_0 < x_m$, $\frac{dx}{dt} = r(1 - \frac{x}{x_m})x > 0$, this function is increasing in this scope. And the concavity and convexity can be easily known by second derivative, and the demarcation point is $\frac{x_m}{2}$

(iii)When $\frac{dx}{dt}$ get the point of $x = \frac{x_m}{2}$, the speed of growth will generally be smaller and till zero, which accords with the reality.

6 Sensitivity Analysis

In this part, we discuss more about the location of office in short term and long term:

From the discussion above, we know that the influence of Internet is stable after a period of time. So, we use the different factors for the model, and we get different result. The result is shown below:

Short term

According to the AHP analysis, the five factors are Internet penetration, tourism, education, quantum of international trade, migration in sequence. And their weights are the U vector's value accordingly. We have the factors' data processed by time series analysis, which are also normalized. And the normalization method is all the data minus mean and are then divided by standard deviation.

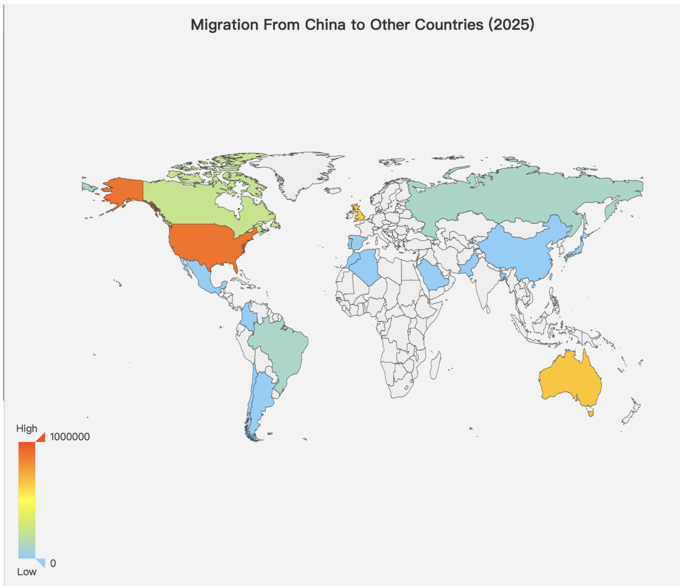

Figure 13: Migration From China to Other Countries(2025)

Country	AHP
United States	3.012
Russia	1.021
China	1.979
United Kingdom	2.898
Saudi Arabia	0.979
Australia	2.776
Spain	2.224
Brazil	0.625
India	0.787

Long term

In the long term, we find the influence of the Internet penetration can be replaced by culture. Therefore, we replace the five elements: culture, tourism, education, culture, quantum of International trade, migration. And then we can use the data we processed to calculate the final AHP results.

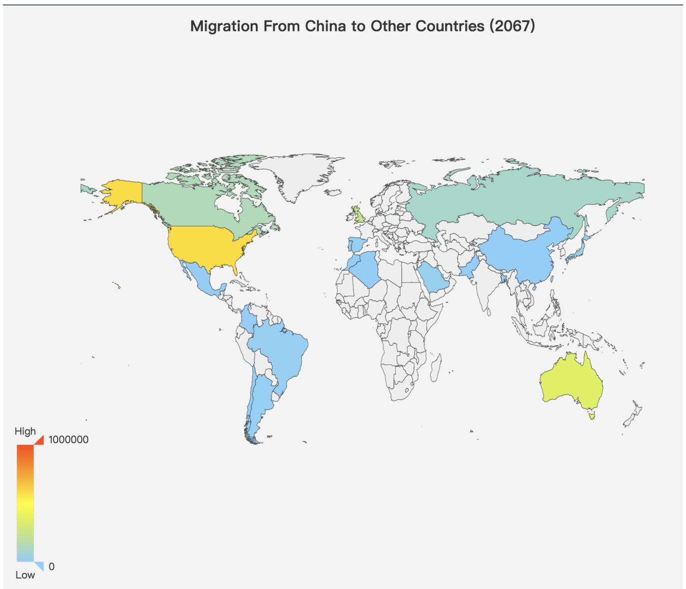

Figure 14: Migration From China to Other Countries(2067)

We choose the office according to the AHP results. According to the AHP results, we can easily find the first six countries with certain language, and the tables above are just the top several locations. Consequently, in the short term, we can easily find that the data in the table are quite close, therefore, we can just choose the top six countries as our candidates. And the languages we choose are the certain countries' native language accordingly because the cultural factors, communication ingredient as well as other reasons. However, in the long term, we can find a large decline on AHP score between Russia and India. Also, associated with our processed and predicted data, we suggest that it is appropriate and economic for the company to open only five offices around the world. What's more, from two heated map above, we can observe that the number of migration from China to other countries largely declines, which means Chinese influence is

Country	AHP
United States	3.041
Russia	2.853
China	3.141
United Kingdom	2.947
Saudi Arabia	1.854
Australia	2.887
Spain	1.765
Brazil	1.767
India	2.027

gradually increasing, so it is reasonable for the company to open a new office in China besides Shanghai. Also, the United States have the same reason to open the second office besides New York.

7 Strengths and Weaknesses

7.1 Strengths

Wide application

The AHP Model and Advanced Logistic Model can be easily modified to accommodate to various situation.

Creative

The Derivative Gravity Model gets inspiration from Newton's law of gravity. It is not only reasonable but also reflects the unity of substance.

Technical supporting

We use many theory and methods to support our work. And each one of them is used reasonably and properly.

7.2 Weaknesses

Laking data

Because we just find 5 groups of data, it is quite difficult for us to get an accurate prediction for the next 5o years.

Inaccuracy

When considering the value of parameters, we obtain them through a lot of different research papers which may influence the result of our model.

We lack a rigorous derivation of T_{ij} in Derivative Gravity Model. Apart from that, lacking theoretical guidance, we just use the data we get to fit the function.

Simplifying assumptions

To simplify the model, we make a few assumptions which may affect the result of our model.

Letter

Dear Chief Operating Officer:

We know you are investigating the development of languages in the current world to get more information about the location of the six new international offices. We are here to help you. This letter explains our optimum solution model, and provides suggestions both in short term and in long term.

As we all known, the main purpose of researching the problem is to determine where to locate the new offices and which language to be chosen as the official language except English and the native language. To answer the question, we need to know which country or region is the most suitable place in the short term and which country or region is the most potential place to locate in the long term.

Thus we build AHP Model to quantify the potential of countries whose languages are the top 10 languages in the world. They are the United States, Russia, China, the United Kingdom, Saudi Arabia, Australia, Spain, Brazil and India. We know that our company has already had offices in New York City in the United States and Shanghai in China. We still think it is reasonable to open new offices in some other cities in these two countries.

According to the AHP analysis, we consider five factors influencing the location. In short term, they are Internet penetration, tourism, education, quantum of international trade, migration. In long term, they are culture, tourism, education, culture, quantum of International trade, migration. That's because during the experiments, we find that the rate of Internet penetration remains stable after just around 10 years. That means this element will not work after a short period of time.

Using AHP, we build a comprehensive evaluation model to evaluate various countries, top of which are listed above. From the result of evaluation, we get the optimum choice for location.

In short term, we think that the top six countries deserve to be located. They are United States, Russia, China, United Kingdom, Spain and Australia. The AHP value is at least 1 for these six areas. The value has a drastic drop when it comes to other areas. However, we still think there is not so much differences in those countries. So we recommend your company to locate in those six places to broaden the market.

However, in long term, we can find a large decline on AHP score between Russia and India. Also, associated with our processed and predicted data, we suggest that it is appropriate and economic for the company to open only five offices around the world. What's more, from the two heated map reflecting the number of people born in China and migrating to other countries, we can observe that the number of migration from China to other countries largely declines, which means Chinese influence is gradually increasing, so it is reasonable for the company to open a new office in China besides Shanghai. Also, the United States have the same reason to open the second office besides New York.

From the perspective of choosing a office language, we think that apart from English, the native language spoken in the located country should be chosen as the second office language. That is because no matter how many people migrate to a country, the number is still too tiny compared with the number of native speakers.

Wish our optimal investment strategy can inspire you at the key point of solving the probable solution of offices locating. We are very eager to hear your opinion on our performance and to have more communication about it. We look forward to hearing from you.

Yours sincerely,

A group of modelers who are enthusiastic about mathematical modeling.

References

[1] https://en.wikipedia.org/wiki/List_of_languages_by_total_number_of_speakers

[2] http://www.8pu.com/gdp/per_capita_gdp_1996.html

Appendix

The Matlab codes of Grey Prediction Model:

```
%y=input('Please input data ')
clear
clc
A=xlsread('Urdu.xlsx');
B=A;
%B=reshape(A,17,4);
C=sum(B,2);
D=flipud(C);
y=D.';

n=length(y);
yy=ones(n,1);
yy(1)=y(1);
for i=2:n
yy(i)=yy(i-1)+y(i);
end
B=ones(n-1,2);
for i=1:(n-1)
    B(i,1)=-(yy(i)+yy(i+1))/2;
    B(i,2)=1;
end
BT=B';
for j=1:n-1
    YN(j)=y(j+1);
end
YN=YN';
A=inv(BT*B)*BT*YN;
a=A(1);
u=A(2);
t=u/a;
t_test=input('How many years do you want to predictïijŽ');
i=1:t_test+n;
yys(i+1)=(y(1)-t).*exp(-a.*i)+t;
yys(1)=y(1);
for j=n+t_test:-1:2
    ys(j)=yys(j)-yys(j-1);
end
x=2001:(n+2000);
xs=2002:n+t_test+2000;
yn=ys(2:n+t_test);
plot(x,y,'^r',xs,yn,'-b');
title('\bf{Urdu}','FontSize', 14)
xlabel('$Years$ ', 'FontSize', 11,'Interpreter','Latex');
ylabel('$Overall \quad Population\quad of\quad People\quadSpeaking\quad
Urdu$', 'FontSize',11,'Interpreter','Latex');
legend('Original Points','Fitting Curve')
grid on
det=0;
for i=2:n
    det=det+abs(yn(i)-y(i));
end

det=det/(n-1);
```

```
disp(['Predictionïijž ',num2str(ys(n+1:n+t_test))]);
predict=ys(n+1:n+t_test);
p=predict.'
```

The codes for picturing the Logistic Model are as follows:

```
clear
clc
A=xlsread('logistic_result.xlsx');
label={'Spanish','Arabic','Portuguese','Chinese','Japanese',
'English','Erdu','Bengali','Russian','Punjabi'};

x = A(1:end,1);

figure
title('\bf{Prediction Based on Logistic Model}');

for k=1:10
    y = A(1:end,k+1);

     subplot(3,4,k);
     plot(x,y,'-*b');
     title(label(k));
     xlabel('$Years$ ', 'FontSize', 11,'Interpreter','Latex');
     ylabel('$Overall \quad Population$', 'FontSize', 11,'Interpreter','Latex');
     grid on;
     legend('Prediction Points')

end
```

The python codes for Multivariate Time Series Model:

```
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import csv

gdp_data = []
internet_data = []
y_data = []
with open('GDP-language.csv', "r") as f:
    csv_reader = csv.reader(f)
    for row in csv_reader:
        row = [float(i) for i in row[1:]]
        gdp_data.append(row)

with open('Internet-language.csv', "r") as f2:
    csv_reader2 = csv.reader(f2)
    for row2 in csv_reader2:
        row2 = [float(i) for i in row2[1:]]
        internet_data.append(row2)

with open('L2_data.csv', "r") as f3:
    csv_reader3 = csv.reader(f3)
    for row3 in csv_reader3:
```

```
        row3 = [float(i) for i in row3]
        y_data.append(row3)
print(gdp_data)
# print(internet_data)
# print(y_data)
X = []
Y = []

'''
def max_all_1dim(dim_1_list):
    max_all = max(dim)
    return max_all

def min_all_2dim(dim_2_list):
    min_all = min([min(t) for t in dim_2_list])
    return min_all
'''

def scailing(dim_2_list):
    new_list = []
    new_list.append(dim_2_list[0])
    # max_all = max_all_2dim(dim_2_list)
    # min_all = min_all_2dim(dim_2_list)
    for q in range(1, len(dim_2_list)):
        max_all = max(dim_2_list[q])
        min_all = min(dim_2_list[q])
        new_list.append([(k-0.99*min_all)/(max_all-min_all) for k in
                                                    dim_2_list[q]])

    return new_list


gdp_data1 = scailing(gdp_data)
internet_data1 = scailing(internet_data)
y_data1 = scailing(y_data)


def get_X(n):
    for w in range(len(gdp_data1[0])):
        sum_w = sum([gdp_data1[i][w] for i in range(1, 9)])
        sum_eng_chi = internet_data1[1][w] + internet_data1[2][w]
        X.append([gdp_data1[n][w]/sum_w, internet_data1[n][w]/
                                                    sum_eng_chi])
    return np.array(X)


def get_Y(n):
    for j in range(len(gdp_data1[0])):
        Y.append([y_data1[n][j]])
    return np.array(Y)


'''
regressor = LinearRegression()
regressor.fit(X_train, y_train)
xx = np.linspace(0, 26, 100)
yy = regressor.predict(xx.reshape(xx.shape[0], 1))
```

```
plt.plot(X_train, y_train, 'k.')
plt.plot(xx, yy)
'''
xx1 = np.linspace(0, 0.6, 100)
xx2 = np.linspace(0.1, 0.8, 100)
xx = np.array([[xx1[i], xx2[i]] for i in range(100)])


# print(xx2)


def get_Z(m):
    quadratic_featurizer = PolynomialFeatures(degree=2)
    X_train_quadratic = quadratic_featurizer.fit_transform(get_X(m))
    # X_test_quadratic = quadratic_featurizer.transform(X_test)
    regressor_quadratic = LinearRegression()
    regressor_quadratic.fit(X_train_quadratic, get_Y(m))
    xx_quadratic = np.array(quadratic_featurizer.transform(xx))
    # plt.plot(X, regressor_quadratic.predict(xx_quadratic), 'r-')
    # plt.show()
    # print(regressor_quadratic.predict(xx_quadratic))
    # Z = np.array([i for i in regressor_quadratic.predict(xx_quadratic
                                        )])
    Z1 = np.array([i[0] for i in regressor_quadratic.predict(
                                        xx_quadratic)])

    return Z1


def get_z(m, x1, x2=100):
    quadratic_featurizer = PolynomialFeatures(degree=2)
    X_train_quadratic = quadratic_featurizer.fit_transform(get_X(m))
    regressor_quadratic = LinearRegression()
    regressor_quadratic.fit(X_train_quadratic, get_Y(m))
    # xx_quadratic = np.array(quadratic_featurizer.transform(xx))
    # plt.plot(X, regressor_quadratic.predict(xx_quadratic), 'r-')
    # plt.show()
    # print(regressor_quadratic.predict(xx_quadratic))
    # Z = np.array([i for i in regressor_quadratic.predict(xx_quadratic
                                        )])
    # Z1 = np.array([i[0] for i in regressor_quadratic.predict(
                                        xx_quadratic)])
    z = regressor_quadratic.predict(quadratic_featurizer.transform([[x1
                                        , x2]]))
    return z


fig = plt.figure(1)
'''
axsub = []
for i in range(1, 9):
    axsub.append(plt.subplot(110+i))
'''
# ax = fig.add_subplot(111, projection='3d')
ax = Axes3D(fig)
# ax.plot_surface(xx1, xx2, Z)
ax.plot(xx1, xx2, get_Z(1)*(max(y_data[1])-min(y_data[1]))+0.99*min(
                                    y_data[1]), color='r')
```

```
ax.plot(xx1, xx2, get_Z(2)*(max(y_data[2])-min(y_data[2]))+0.99*min(
                                    y_data[2]), color='b')
ax.plot(xx1, xx2, get_Z(3)*(max(y_data[3])-min(y_data[3]))+0.99*min(
                                    y_data[3]), color='g')
ax.plot(xx1, xx2, get_Z(4)*(max(y_data[4])-min(y_data[4]))+0.99*min(
                                    y_data[4]), color='c')
ax.plot(xx1, xx2, get_Z(5)*(max(y_data[5])-min(y_data[5]))+0.99*min(
                                    y_data[5]), color='m')
ax.plot(xx1, xx2, get_Z(6)*(max(y_data[6])-min(y_data[6]))+0.99*min(
                                    y_data[6]), color='y')
ax.plot(xx1, xx2, get_Z(7)*(max(y_data[7])-min(y_data[7]))+0.99*min(
                                    y_data[7]), color='k')
ax.plot(xx1, xx2, get_Z(8)*(max(y_data[8])-min(y_data[8]))+0.99*min(
                                    y_data[8]), color='w')
ax.set_zlabel('Second Language Users/million')
ax.set_ylabel('Internet factors')
ax.set_xlabel('GDP factors')

plt.show()

gdp_data2 = []
gdp_data2.append([2018+i for i in range(50)])

for j in range(1, 9):
    regressor2 = LinearRegression()
    regressor2.fit([[gdp_data[0][o]] for o in range(len(gdp_data[0]))],
                   [[gdp_data[j][p]] for p in range(len(gdp_data[j]))])
    gdp_data2.append([regressor2.predict([[2018+i]]) for i in range(50)
                                        ])

gdp_data_tmp = []
gdp_data_tmp.append(gdp_data2[0])
sum_j1 = []
for j in range(len(gdp_data2[0])):
    sum_j1.append(sum([gdp_data2[i][j] for i in range(1, 9)]))
for n in range(1, 9):
    gdp_data_tmp.append([gdp_data2[n][p]/sum_j1[p] for p in range(len(
                                        gdp_data2[0]))])

gdp_data2 = gdp_data_tmp


# print(gdp_data2)


def make_zlist(n, gdp_datalist, internet_datalist = None):
    max_tmp = max(y_data[n])
    min_tmp = min(y_data[n])

    # max_int = max(internet_datalist[n])
    # min_int = min(internet_datalist[n])

    z_pre = [get_z(n, gdp_datalist[n][i], internet_datalist[n][0])*(
                                        max_tmp-min_tmp)+0.99*min_tmp
                                        for i in range(50)]
    return z_pre
```

```
# print(make_zlist(1, gdp_data2))

'''
def process_z_pre(z):
    return z*(max_all_2dim(y_data)-min_all_2dim(y_data))+min_all_2dim(
                                                    y_data)


z_pre = [process_z_pre(get_z(1, gdp_data2[1][i], 90)) for i in range(50
                                            )]
print(z_pre)
'''
'''
plt.figure(2)

plt.subplot(221)
plt.plot([2018+i for i in range(50)], [make_zlist(1, gdp_data2,
                                    internet_data1)[h][0][0] for h in
                                    range(50)])
plt.xlabel("Time/Year")
plt.ylabel("L2 Users/million")
plt.title('Mandarin')

plt.subplot(222)
plt.plot([2018+i for i in range(50)], [make_zlist(2, gdp_data2,
                                    internet_data1)[h][0][0] for h in
                                    range(50)])
plt.xlabel("Time/Year")
plt.ylabel("L2 Users/million")
plt.title('English')

plt.subplot(223)
plt.plot([2018+i for i in range(50)], [make_zlist(3, gdp_data2,
                                    internet_data1)[h][0][0] for h in
                                    range(50)])
plt.xlabel("Time/Year")
plt.ylabel("L2 Users/million")
plt.title('Erdu')

plt.subplot(224)
plt.plot([2018+i for i in range(50)], [make_zlist(4, gdp_data2,
                                    internet_data1)[h][0][0] for h in
                                    range(50)])
plt.xlabel("Time/Year")
plt.ylabel("L2 Users/million")
plt.title('Spanish')

plt.show()
'''

plt.figure(3)

plt.subplot(221)
plt.plot([2018+i for i in range(50)], [make_zlist(5, gdp_data2,
                                    internet_data1)[h][0][0] for h in
                                    range(50)])
plt.xlabel("Time/Year")
plt.ylabel("L2 Users/million")
```

```
plt.title('Arabic')

plt.subplot(222)
plt.plot([2018+i for i in range(50)], [make_zlist(6, gdp_data2,
                                    internet_data1)[h][0][0] for h in
                                    range(50)])
plt.xlabel("Time/Year")
plt.ylabel("L2 Users/million")
plt.title('Russian')

plt.subplot(223)
plt.plot([2018+i for i in range(50)], [make_zlist(7, gdp_data2,
                                    internet_data1)[h][0][0] for h in
                                    range(50)])
plt.xlabel("Time/Year")
plt.ylabel("L2 Users/million")
plt.title('Bengali')

plt.subplot(224)
plt.plot([2018+i for i in range(50)], [make_zlist(8, gdp_data2,
                                    internet_data1)[h][0][0] for h in
                                    range(50)])
plt.xlabel("Time/Year")
plt.ylabel("L2 Users/million")
plt.title('Portuguese')

plt.show()
```

The python codes Derivative Gravity Model:

```
import csv
import numpy as np
import pandas as pd
import pyflux as pf
from datetime import datetime
import matplotlib.pyplot as plt
from math import log

train_country_list = ['Spain', 'China', 'India', 'Japan', 'Argentina',
                                    'Brazil', 'Chile', 'Colombia', '
                                    Morocco']

T_data = {}
A_data = {}
P_data = {}
G_data = {}
with open('Imigrants to us_America_v2.csv', 'r') as f:
    csv_reader = csv.reader(f)
    T_data_tmp = []
    for row in csv_reader:
        T_data_tmp.append(row)

    for i in range(1, len(T_data_tmp)):
        for j in range(1, 4):
            T_data[(T_data_tmp[i][0], T_data_tmp[0][j])] = float(
                                                T_data_tmp[i][j])/10000

with open('Area.csv', 'r') as f1:
```

```
    csv_reader1 = csv.reader(f1)
    for row in csv_reader1:
        A_data[row[0]] = float(row[1])/1000000

with open('Population.csv', 'r') as f2:
    csv_reader2 = csv.reader(f2)
    for row in csv_reader2:
        P_data[(row[0], row[1])] = float(row[3])/10000

with open('GDP_per.csv', 'r') as f3:
    csv_reader3 = csv.reader(f3)
    G_data_tmp = []
    for row in csv_reader3:
        G_data_tmp.append(row)

    for i in range(1, len(G_data_tmp)):
        for j in range(1, 4):
            G_data[(G_data_tmp[i][0], G_data_tmp[0][j])] = float(
                                                                    G_data_tmp[i][j])/
                                                                    100000



alpha = 0.01
theta = [4, 4, 4, 4]
m = len(T_data)
iterations = 200


def T_fun(Pi, Pj, Gi, Gj, Aj):
    T_ij = log(((theta[0]*(Pi**theta[1])*(Pj**theta[2]))*Aj)/(Gi/Gj)**
                                                        theta[3])
    return T_ij


def Tpian_0(Pi, Pj, Gi, Gj, Aj):
    # pian0 = ((Pi**theta[1])*(Pj**theta[2]))*Aj/(Gi/Gj)**theta[3]
    pian0 = 1/theta[0]
    return pian0


def Tpian_1(Pi, Pj, Gi, Gj, Aj):
    # pian1 = (theta[0]*theta[1]*(Pi**(theta[1]-1))*(Pj**theta[2]))*Aj
                                                        /(Gi/Gj)**theta[3]
    pian1 = Pi
    return pian1


def Tpian_2(Pi, Pj, Gi, Gj, Aj):
    # pian2 = (theta[0]*theta[2]*(Pi**(theta[1]))*(Pj**(theta[2]-1)))*
                                                        Aj/(Gi/Gj)**theta[3]
    pian2 = Pj
    return pian2


def Tpian_3(Pi, Pj, Gi, Gj, Aj):
    # pian3 = (theta[0]*theta[3]*(Pi**(theta[1]))*(Pj**(theta[2])))*Aj
                                                        /(Gi/Gj)**(theta[3]+1)
```

```
    pian3 = log(Gi/Gj)
    return pian3

def cost_fun():
    J1 = 0
    for i in T_data:
        J1 += (T_fun(P_data[i], P_data[('United States of America', i[1
                                                    ])],
                     G_data[i], G_data[('United States of America', i[1])
                                                              ], A_data['
                                                              United States of
                                                               America'])\
              -log(T_data[i]))**2

    J1 = J1/(len(T_data)*2)
    return J1


def update_theta(theta_para):
    for j in range(4):
        sum_list = [0, 0, 0, 0]

        for i in T_data:
            var_x = [P_data[i], P_data[('United States of America', i[1
                                                    ])], G_data[i], G_data[
                                                    ('United States of
                                                    America',i[1])], A_data
                                                    ['United States of
                                                    America']]
            sum_list[0] += (T_fun(var_x[0], var_x[1], var_x[2], var_x[3
                                                    ], var_x[4])-log(T_data
                                                    [i])) * Tpian_0(var_x[0
                                                    ],var_x[1], var_x[2],
                                                    var_x[3], var_x[4])
            sum_list[1] += (T_fun(var_x[0], var_x[1], var_x[2], var_x[3
                                                    ], var_x[4]) - log(
                                                    T_data[i])) * Tpian_1(
                                                    var_x[0],var_x[1],var_x
                                                    [2],var_x[3],var_x[4])
            sum_list[2] += (T_fun(var_x[0], var_x[1], var_x[2], var_x[3
                                                    ], var_x[4]) - log(
                                                    T_data[i])) * Tpian_2(
                                                    var_x[0],var_x[1],var_x
                                                    [2],var_x[3],var_x[4])
            sum_list[3] += (T_fun(var_x[0], var_x[1], var_x[2], var_x[3
                                                    ], var_x[4]) - log(
                                                    T_data[i])) * Tpian_3(
                                                    var_x[0],
var_x[1],var_x[2],var_x[3],var_x[4])
        theta_para[j] = theta_para[j] - (alpha/m)*sum_list[j]
    return theta_para

plt.figure(1)
for i in range(iterations):
    J = cost_fun()
    update_theta(theta)
    if i < 15:
        plt.scatter(i, J, s=100, marker='x')
```

```
plt.show()

print(theta)
```

The data used for the model is as follows:

Russia	16377742
China	9326410
United States of America	9147593
Canada	9093507
Brazil	8358140
India	2973193
Argentina	2736690
Australia	7682300
Algeria	2381741
Saudi Arabia	2149690
Mexico	1943945
Colombia	1038700
Morocco	446300
Japan	364485
United Kingdom	241930
Bangladesh	130170
Portugal	91470
Chile	743812
Spain	498980

Figure 15: National Land Area

Region	2013	2014	2015
Spain	2970	3341	3707
China	68410	72492	70977
India	65506	74451	61380
Brazil	10772	10246	11247

Figure 16: The Number of Immigrants to U.S.

language	2015	2014	2013	2012	2011		2010	2009	2008	2007	2006
Russia	9521	14388	16022.65	15425.31	14326.07		11445.13	9178.45	12468.38	9753.26	7420.48
China	8166.76	7701.69	7080.83	6329.46	5582.89		4524.06	3837.9	3467.03	2703	2110.57
United States of America	56436.71	54668.08	52741.73	51403.39	49733.88		48310.34	46909.42	48302.28	47954.53	46351.67
Brazil	8810.5	12111.74	12294.6	12366.95	13242.69		11298.41	8625.13	8854.99	7374.78	5912.96
India	1629	1607.41	1485.6	1481.56	1497.75		1422.93	1153.19	1048.74	1076.84	837.46
Australia	51220.04	61231.97	64733.89	68048.29	66773.11		56360.35	45603.96	49223.87	45162.75	37907.92
Saudi Arabia	21094.59	24580.47	24892.99	25208.16	23654.87		19112.69	16094.68	20157.3	16666.63	15603.99
United Kingdom	43976.4	46478.84	42452.95	41683.67	41259.5		38737.56	38180.57	46890.22	49973.92	44095.77
Spain	25717.56	29663.94	29236.04	28582.17	31867.78		30802.85	32412.23	35724.77	32748.09	28531.25
language	2005	2004	2003	2002	2001		2000	1999	1998	1997	1996
Russia	5713.36	4408.37	3197.69	2556.64	2259.02		1905.95	1432.66	1975.97	2944.47	2842.86
China	1765.72	1512.62	1293.13	1150.21	1053.15		958.564	872.222	827.643	780.839	708.58
United States of America	44218.31	41838.46	39591.87	38113.89	37241.35		36432.51	34601.72	32929.04	31553.62	30047.31
Brazil	4815.89	3659.06	3090.53	2859.41	3183.61		3778.97	3507.74	5132.05	5327.21	5207.94
India	747.335	656.517	571.555	491.989	471.701		462.978	462.218	432.225	434.709	418.63
Australia	36178.74	32796.12	27256.62	21653.29	19442.85		20859.66	21767.42	20373.79	23050.34	23218.03
Saudi Arabia	14068.22	11467.1	9800.53	8822.51	8778.39		9256.54	8092.7	7525.3	8706.35	8527.67
United Kingdom	41566.63	39871.03	34049.76	29654.24	27296.67		27828.38	28159.95	27769.02	26366.88	23974.94
Spain	26550.34	24988.98	21532.45	17083.1	15368.51		14724.65	15774.82	15420.7	14747.59	16057.38

Figure 17: GDP Per Capital

Argentina	2015	43417.765		India	2015	1309053.98
Argentina	2014	42981.515		India	2014	1293859.29
Argentina	2013	42539.925		India	2013	1278562.21
Australia	2015	23799.556		Japan	2015	127974.958
Australia	2014	23474.668		Japan	2014	128162.873
Australia	2013	23150.729		Japan	2013	128312.92
Brazil	2015	205962.108		Morocco	2015	34803.322
Brazil	2014	204213.133		Morocco	2014	34318.082
Brazil	2013	202408.632		Morocco	2013	33824.769
Chile	2015	17762.681		Pakistan	2015	189380.513
Chile	2014	17613.798		Pakistan	2014	185546.257
Chile	2013	17462.982		Pakistan	2013	181712.595
China	2015	1397028.55		Portugal	2015	10418.473
China	2014	1390110.39		Portugal	2014	10471.168
China	2013	1382793.21		Portugal	2013	10527.674
Colombia	2015	48228.697		Spain	2015	46397.664
Colombia	2014	47791.911		Spain	2014	46521.827
Colombia	2013	47342.981		Spain	2013	46697.553
United Sta	2015	319929.162		United Kingdom	2015	65397.08
United Sta	2014	317718.779		United Kingdom	2014	65015.686
United Sta	2013	315536.676		United Kingdom	2013	64641.11

Figure 18: Population