For office use only	Team Control Number	For office use only
T1 ________________	**80812**	F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	**B**	F4 ________________

2018
MCM/ICM
Summary Sheet

Determine the Location of International Office Based on Language Distribution

Summary

We build two models to investigate the distribution of different languages.

MODEL 1 Immigration is an important contributing factor of language distribution. We build an immigration gravity model to estimate the attraction from one country to another, the parameters of which include medical condition, employment condition, environment condition, education condition, safety condition, distance and language family. We determine their weights by AHP and design our own algorithm to solve the model. Then we determine the number of emigrants and calculate the immigration population to each destination country.

MODEL 2 According to countries' birth rate, death rate, education condition and the influence of social pressures and globalization, we build a dynamic population system to calculate the number of different language speakers over time.

Through MATLAB, We make two new top-ten lists(native speakers and total speakers) in 50 years. We also make several world maps to compare the current geographical language distribution with that of 50 years later.

MODEL 3 In order to determine the locations of 6 new offices, we build a maximum profits model. We use language distribution, as well as GDP, to measure the comprehensive national strength of each country. Taking into consideration distance between a country and the nearest office, we calculate the profits gained in each country and then sum them up.

For every scheme of 6 office locations, we calculate the total profits of the company. Then we pick the best scheme which leads to the maximum profits. The recommendations are as follows:

In the short term: Toronto, Canada; Paris, France; Sydney, Australia; Tokyo, Japan; Moscow, Russia; Buenos Aires, Argentina. In the long term, Moscow and Buenos Aires are replaced by Mumbai, India and Rio de Janeiro, Brazil.

Finally, we analyze the sensitivity of two important parameters. One is c_{ed}, which describes the education strength of a country. The other is $distance$, which describes the decreasing profits caused by distance.

Keywords: **AHP; differential equation; mathematical optimization**

Contents

1 Introduction

1.1 Problem Background

Research founds that 96 % of languages are used by only 4 % of the total world population, which suggests that many languages are quickly dying out. The total number of speakers of a language may increase or decrease over time because of a variety of reasons, such as the population of the native speaker, immigration and emigration, globalization, the assistance of translation tool and so on. In view of these infulences, the pattern of languages in the future may become quite different.

1.2 Description of Terminology

- Pan-country: Countries that use the same official language belong to the same pan-country. The pan-country's parameters (such as GDP, birth rate) equal to the average parameters of its typical countries.
- Pan-country language: The offical language used in this pan-country.
- Destination country: The pan-country which is the destination of migration.
- Origin country: The pan-country which is the origin of migration.

1.3 Our work

Each language is unique and has its own character. In order to reduce the difficulty of research while take these differences into consideration at the same time, we simplify the target.
First, in view of the popularity and wide use of English, we divide languages into two categories:

- English
- non-English

Second, when we talk about the distribution of a specific language in the future, we take these following influences into our consideration:

- Does the government promote it?
- Do schools use it to teach?
- Do social pressures force people to learn it?
- Does it influence other languages within a cultural circle?
- How do immigration and emigration influence the use of it?
- How does globalization influence the use of it?
- Does the translation technology influence the use of it?

- Which language family does it belong to?

Third, based on the number of native speakers and total speakers, We select 16 different languages as our objects, and we also select 21 different countries as the representative of these languages:

Table 1: Object pan-countries and their typical countries

Pan-country	Typical Country
Chinese-speaking country	China
Spanish-speaking country	Spain, Argentina
English-speaking country	the United States, South Africa, Canada
Hindi-speaking country	India
Arabic-speaking country	Saudi Arabia, Egypt
Bengali-speaking country	Bangladesh
Portuguese-speaking country	Portugal, Brazil
Russian-speaking country	Russia, Kazakhstan
Punjitan-speaking country	Pakistan
Japanese-speaking country	Japan
French-speaking country	France
Malay-speaking country	Malaysia
Hausa-speaking country	Nigeria
German-speaking country	Germany
Persian-speaking country	Iran
Italian-speaking country	Italy

1.4 Model overview

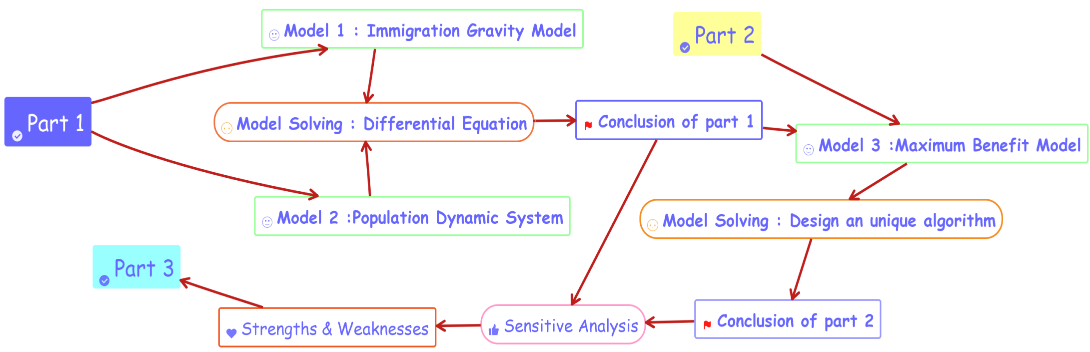


Figure 1: Model overview

2 Assumptions

Our model makes the following assumptions:

- Immigrants's offsprings are native speakers of the country.

- Only one language can become a universal language.
- People in non-English pan-countries will not learn to speak a foreign language other than English unless they emigrate to other pan-countries.
- Every language has dialects.
- We neglect the effect of minority languages.
- We neglect immigrants who become emigrants.

3 Nomenclature

We use the nomenclature in Table 1 to describe our model. Other symbols that are used only once are described later.

Table 2: Table of symbols

Parameter	Meaning	Unit
I_{ij}	the number of immigrants per year from i country to j country	*million*
F_{ij}	j country's attraction to people in i country	1
P	the population of a certain pan-country	*million*
P_o	people who use official language as native language	*million*
P_{od}	people who are native dialect speakers and second(or 3rd,etc) official language speakers	*million*
P_d	people who use only dialect	*million*
SP	the growing number of English second speaker caused by social pressures per year	*million*
G	the growing number of English second speaker caused by globalization per year	*million*

4 Task I

4.1 Model design

In order to describe the distribution of various language speakers over time, we divide the growth of a language into different parts. If we calculate all part's annual growth, we can describe the change of the language over time. As we have stated in 1.3, these two different categories are divided diversely, as is show in **Figure 2** and **Figure 3**. We mainly design two models, one for caculating the influence of immigration and emigration, the other for caculating the influence of population growth. Other factors like migration and assimilation of cultural groups and policy, we only select typical examples to explain their influences perceptually.

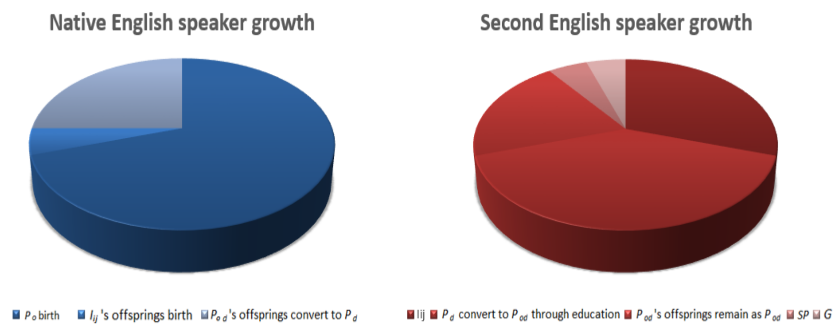


Figure 2: English part

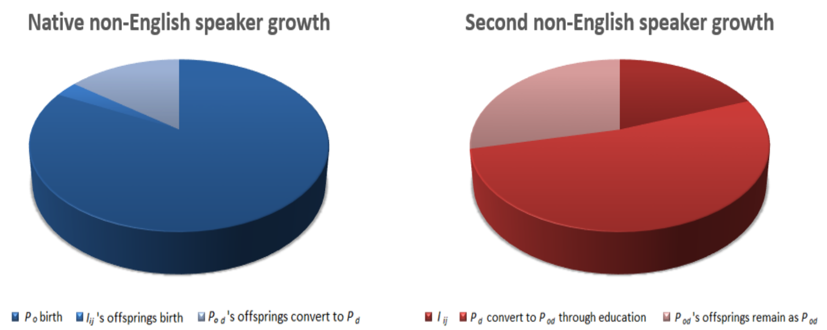


Figure 3: Non-English part

4.2 Model I: Immigration Gravity Model

Immigrants are the second language speakers of the country and their offsprings are the native speakers of the country(according to our assuption in chapter 2). So both the number of immigrants and their population growth rate are related with the total number of speakers of a language. If we know the total number of immigrants in a year, we will know the growing number of the second language speakers; If we know the birth rate of immigrants' offsprings, we will know the growing number of the native language speakers due to immigration; If we know the number and distributions of immigrants, we will know the geographic distributions of various language , as immigrants bring their native language(aslo their native country's language) to other parts of the world.

4.2.1 Model assumption

According to *International Migration Outlook 2012 - Non-Profit Data*, most immigrants are in the United States. To simpilfy our model, we select 6 pan-countries as immigration destination pan-countries, including English-speaking countries, Itlian-speaking countries, Spanish-speaking countries, Russian-speaking countries, German-speaking countries and French-speaking countries. And all the pan-countries in **Table 1** are origin countries. We assume that immigration does not happen within the same pan-country, because this kind of immigration will not affect the distribution of languages.

4.2.2 Factors affecting F_{ij}

Consider the influence of immigration and emigration, we select 7 different factors to describe F_{ij}: difference of medical condition, difference of education condition, difference of environment condition, difference of employment environment, difference of security condition, whether their languages belong to the same Family and the distance between them. The first four factors have positive relationship with F_{ij}, while the other effects have negative relationship with F_{ij}.
We use gravity model to describe this relation:

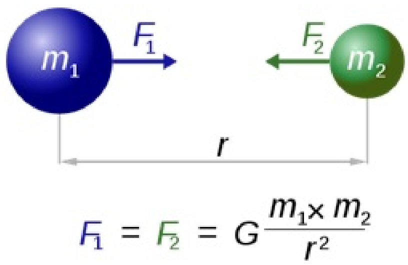


Figure 4: Law of universal gravitation

We use the ratio of i pan-country's parameters and j pan-country's parameters to describe the different condition of the two pan-countries. Besides, every parameter has its own weight. So we have:

$$F_{ij} = G \cdot \frac{M^{\alpha_1} \cdot J^{\alpha_2} \cdot En^{\alpha_3} \cdot Ed^{\alpha_4}}{f \cdot (S^{\alpha_5} \cdot D^{\alpha_6})^2}. \tag{1}$$

NOTE:

- All the following socres come from https://www.usnews.com/news/best-countries/data-explorer#:
  - M equals to the rate of j pan-country's medical condition score and i pan-country's medical condition score.
  - J equals to the rate of j pan-country's employment condition score and i pan-country's employment condition score.
  - En equals to the rate of j pan-country's environment condition score and i pan-country's environment condition score.
  - Ed equals to the rate of j pan-country's education condition score and i pan-country's education condition score.
  - S equals to the rate of i pan-country's safety condition score and j pan-country's safety condition score.
  - D equals to the distance between the two pan-countries.
- α_1, α_2, α_3, α_4, α_5, α_6 are the weights of M,J,En,Ed,S,D.
- f measures the difference between the two pan-countries in the language family. If they belong to the same language family, $f = 1$; otherwise, $f = 1.02$.
- G does not influence our conclusions, so we suppose $G = 1$.

4.2.3 Determine α_1, α_2, α_3, α_4, α_5, α_6

Datas in *International Migration Outlook 2012 - Non-Profit Data* show that work migration inflows and family migration inflows account for 35.5% and 39.6% respectively. So we chose work migration (W_m) and family migration (F_m) as our objects and determine α_1, α_2, α_3, α_4, α_5, α_6 for each one.
We use Analytic Hierarchy Process (AHP) and the results are as follows:

	Medical	Jobs	Environment	Education	Safety	Distance
Medical	1	1/3	7	9	1	6
Jobs	3	1	5	8	2	6
Environment	1/7	1/5	1	4	1/6	3
Education	1/9	1/8	1/4	1	1/9	1/3
Safety	1	1/2	6	9	1	8
Distance	1/6	1/6	1/3	3	1/8	1

Table 3: W_m judgment matrix

	Medical	Jobs	Environment	Education	Safety	Distance
Medical	1	1/4	5	1/4	1	7
Jobs	4	1	5	1/2	3	8
Environment	1/5	1/5	1	1/6	1/5	5
Education	4	2	6	1	4	9
Safety	1	1/3	5	1/4	1	7
Distance	1/7	1/8	1/5	1/9	1/7	1

Table 4: F_m judgment matrix

Factors	Medical	Jobs	Environment	Education	Safety	Distance
Weight	α_1	α_2	α_3	α_4	α_5	α_6
W_m	0.2425	0.3715	0.0680	0.0244	0.2539	0.0397
F_m	0.1294	0.2797	0.0516	0.3837	0.1329	0.0227

Table 5: The weight of α_1, α_2, α_3, α_4, α_5, α_6

As we have showed in **4.2.1**, we have 16 pan-countries as our objects. We select 6 pan-countries as migration destination and all pan-countries as migration origin. We don't consider immigration within the same country, which means $F_{ii} = 0, 1 \leq i \leq 6$. So we get a 16×6 matrix with 96 different elements as follows:

$$F = \begin{bmatrix} 0 & F_{1,2} & \cdots & F_{1,6} \\ \vdots & \ddots & \ddots & \vdots \\ F_{6,1} & \cdots & \cdots & 0 \\ F_{7,1} & \cdots & \cdots & F_{7,6} \\ \vdots & \ddots & \ddots & \vdots \\ F_{16,1} & \cdots & \cdots & F_{16,6} \end{bmatrix} \tag{2}$$

4.2.4 Model updating

F_{ij} changes over years The world is changing all the time. So F_{ij} also changes since the difference between i pan-country and j pan-country changes over years. We assume that the change of F_{ij} has positive relationship with time and negative relationship with the rate of annual GDP growth in i pan-country and j pan-country.So we have:

$$F_{ij}^{(t+1)} = \begin{cases} F_{ij}^{(t)} \cdot \left(\frac{\overline{GDP_i}}{\overline{GDP_j}}\right)^{-\gamma} & \frac{\overline{GDP_i}}{\overline{GDP_j}} \geq 1 \\ F_{ij}^{(t)} \cdot \left(\frac{\overline{GDP_i}}{\overline{GDP_j}}\right)^{\gamma} & \frac{\overline{GDP_i}}{\overline{GDP_j}} < 1 \end{cases} \tag{3}$$

NOTE:

- T means one year, also the iteration times.
- γ is the coefficient of F_{ij} changes.

We perform this operation on all the elements in the matrix F to get the matrix F'.

Determine the percentage of immigrants in 6 destinations The rate of i pan-country's emigrants in each destination is determined by the rate of its F_{ij} with other destinations' F_{ij}. So we have:

$$F_{ij}^{(t+1)} = \frac{F_{ij}^{(t)}}{\sum_{j=1}^{6} F_{ij}} \tag{4}$$

We perform this operation on all the elements in the matrix F' to get the matrix F''.

Determine the number of emigrants How to determine the number of emigrants of each origin? Apparently, it is closely related with the population of the origin pan-country and the living quality of its people. When GDPPC is too high, people are less likely to migrate; When GDPPC is too low, people can not afford the price for migration. So we decide to use the following function to describe the relationship between the number of immigrants,GDPPC and population.
The number of immigrants has positive correlation with the population of the pan-country, and has inverted U relation with GDPPC, the function is as follows:

$$I_i = f(GDPPC_i, P_i) = \sigma \cdot P_i \cdot \left(GDPPC_i + \frac{k}{GDPPC_i}\right)^{-1} = \frac{\sigma \cdot P_i \cdot GDPPC_i}{GDPPC_i{}^2 + k} \tag{5}$$

NOTE:

- σ is the coefficient of migration.
- P_i is the population of i pan-country.
- $GDPPC_i$ is GDPPC of i pan-country.
- k is the coefficient of immigration and GDPPC.

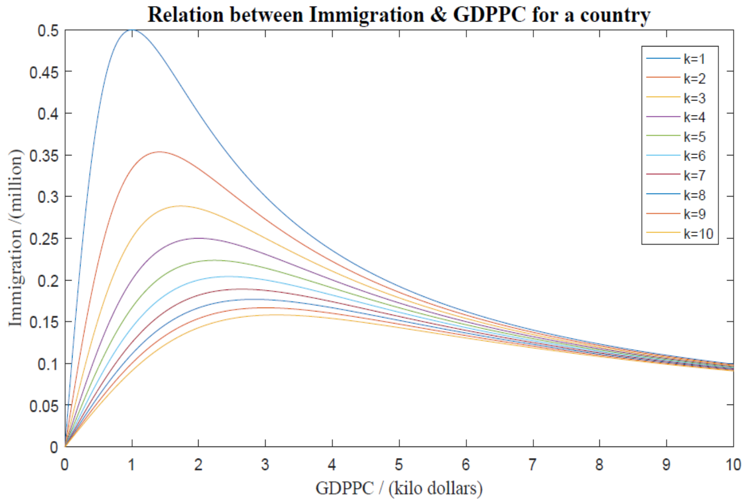


Figure 5: $f(GDPPC_i, P_i)$

When $\sigma = 1, P_i = 1$, $f(GDDPC_i, P_i)$ is as follows:

We can see that the number of immigrants is both small when GDPPC is low and high, which fits our expectations.

Determine the matrix I By scalar multiplication of equation(4)'s results and matrix F'', we get the final matrix I as follows:

$$I = \begin{bmatrix} I_1 \\ \vdots \\ I_{16} \end{bmatrix} \cdot \times F'' = \begin{bmatrix} 0 & I_{1,2} & \cdots & I_{1,6} \\ \vdots & \ddots & \ddots & \vdots \\ I_{6,1} & \cdots & \cdots & 0 \\ I_{7,1} & \cdots & \cdots & I_{7,6} \\ \vdots & \ddots & \ddots & \vdots \\ I_{16,1} & \cdots & \cdots & I_{16,6} \end{bmatrix} \tag{6}$$

So we get the total number of immigrants in 6 destination pan-countries every year:

$$I_j = \sum_{i=1}^{16} I_{ij} \tag{7}$$

4.3 Model II: Dynamic Population System

4.3.1 Determine P_o, P_{od}, P_d

We divide a pan-country's people into 3 categories: P_o, P_{od} and P_d (their meanings have already been explained in **1.2**). They can change over time:

- P_d can change into P_{od} through education.
- P_{od}'s offsprings have two options:

  - P_{od} can change into P_o if their parents teach them official language only.
  - P_{od} can also remain as P_{od} if their parents teach them to use dialect first.

- P_o's offsprings remain as P_o.

Based on these analysis, we have the following dynamic model:

$$(P_d)_{i+1} = (1 + r_b - r_d - s_{ed} \cdot c_{ed}) \cdot (P_d)_i \tag{8}$$

$$(P_{od})_{i+1} = [1 + r_b(1 - c_l) - r_d] \cdot (P_{od})_i + s_{ed} \cdot c_{ed} \cdot (P_d)_i \tag{9}$$

$$(P_o)_{i+1} = (1 + r_b - r_d) \cdot (P_o)_i + r_b \cdot c_l \cdot (P_{od})_i \tag{10}$$

NOTE:

- $(P_d)_i$, $(P_{od})_i$ and $(P_o)_i$ means P_d, P_{od} and P_o in i year.
- r_b and r_d means the birth rate and death rate of the pan-country.
- s_{ed} means education score of the pan-country.
- c_{ed} means the education coefficient, which reflect the variation trend of illiterates.
- c_l means the dialect loss coefficient of the pan-country, which reflect the loss of dialect in the pan-country.

4.3.2 Model updating

The model II describe the change of certain language speakers caused by the growth of the pan country's population. In order to calculate more precisely, we consider the following 3 things and add them to our model.

Amend the number of immigrants According to model I, we have got I_{ij} and these immigrants become second pan-country language speakers. We also know that immigrants' offsprings become native pan-country speakers (according to **Chapter 2**). So we have:

$$\begin{aligned} P_{od}^{(t+1)} &= P_{od}^{(t)} + I_{ij}^{(t)} \\ P_o^{(t+1)} &= P_o^{(t)} + (r_b - r_d) \cdot I_{ij}^{(t)} \end{aligned} \tag{11}$$

Difference caused by social pressures and globalization Consider the influence of social pressures and global communication, we assume that people in non-English pan-countries will not learn to speak a foreign language other than English unless they emigrate to other pan-countries (according to **Chapter 2**). So we have to amend the number of Second English speakers. We define:

$$\begin{aligned} SP_{neo} &= \alpha \cdot (P - SP - G) \cdot \frac{\rho^{\frac{1}{2}}}{GDPPC^{\frac{1}{4}}} \\ G_{neo} &= \beta \cdot (P - SP - G) \cdot s_c \\ SP^{(t+1)} &= SP^{(t)} + SP_{neo} \\ G^{(t+1)} &= G^{(t)} + G_{neo} \end{aligned} \tag{12}$$

NOTE:

- SP_{neo} equals to the growth of second English speakers caused by social pressures.
- G_{neo} equals to the growth of second English speakers caused by global communication.
- s_c is the connection socre of a certain pan-country.
- ρ is the population density of a certain pan-country.
- α and β are coefficients.

So when calculating second English speakers, we should add these two inluences into consideration:

$$P_{od}^{(t+1)} = P_{od}^{(t)} + SP^{(t)} + G^{(t)} \tag{13}$$

The influence of policy, cultrual circle and technology Even though it is universally acknowledged that our world change quickly over time, we do not consider the influence of policy due to its unpredictablity.
As for the influence within the same cultrual circle, it is also hard to describe. Since the difficulties of quantification the cultrual difference between pan-countries, we have hard time determining each coefficent of the influence of cultrual. So we do not take this effect into consideration.
The same is true with technology. Nowadays, we have witness rapid development of our translation technology. But it is also hard to predict the trend in next 10 years not to mention 50 years. So we do not take this factor into consideration, either.

4.4 Calculation

4.4.1 Determining Parameters

We determine the parameters through deep thought, the results are as follows:

Parameters	**Meaning**	**Value**
α	the English learning coefficient caused by social pressures	10^{-5}
β	the English learning coefficient caused by globalization	10^{-7}
γ	the coefficient of F_{ij} changes	0.044
θ	the coefficient of migration	0.03
c_l	the dialect loss coefficient of a certain pan-country	0.2
c_{ed}	the education coefficient	8×10^{-4}
k	the coefficient of immigration and GDPPC	100

Table 6: The Value of α, β, γ, θ, c_l, c_{ed} and k

4.4.2 Model Algorithm

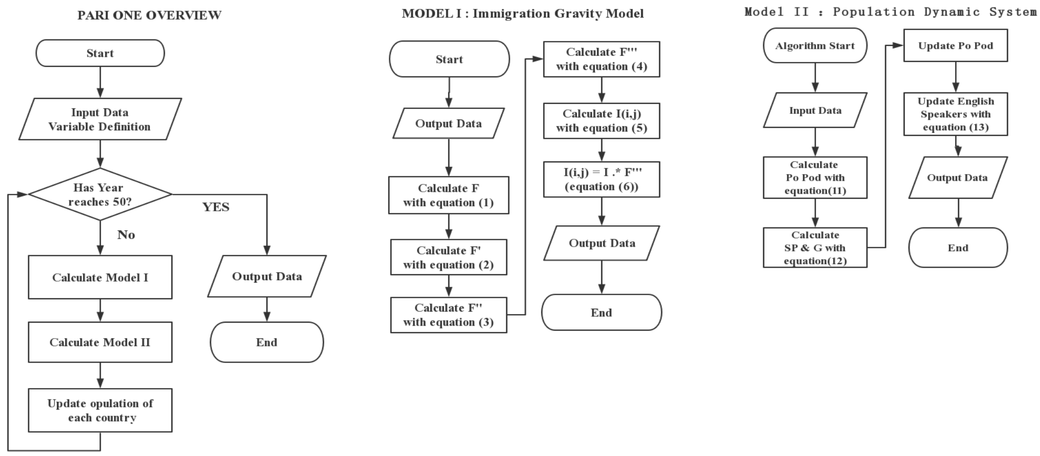
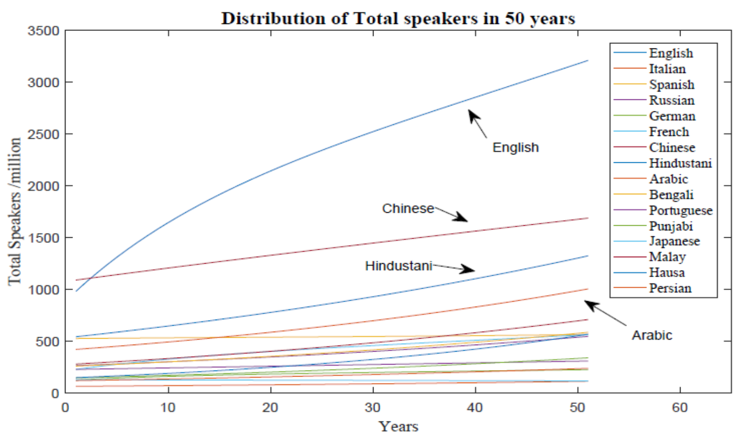


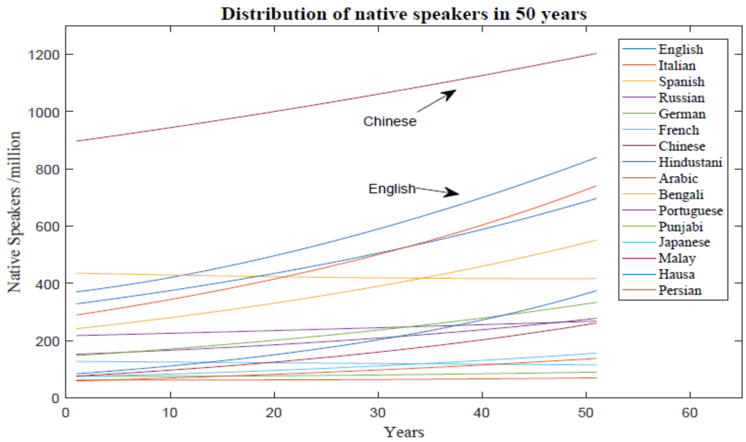
Figure 6: Model Algorithm for Task I

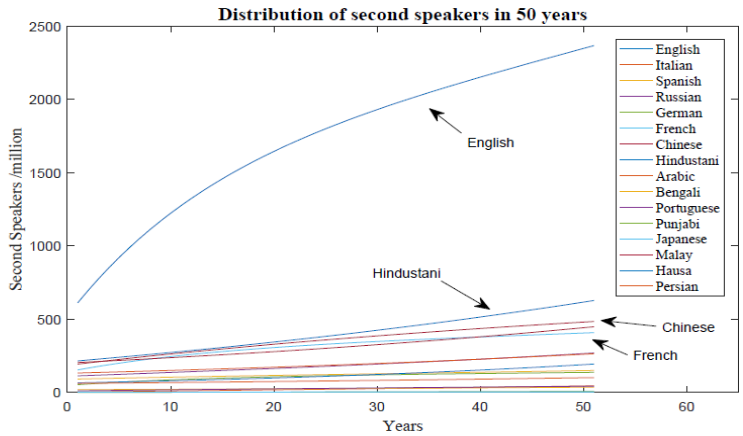
4.5 Conclutions

4.5.1 Part I problem A

Based on our Model I and Model II, we get our conclusions through MATLAB. Belows are the graghs which describe the distribution of total speakers, native speakers and second speakers of various languages in 50 years:







4.5.2 Part I problem B

We also calculate the rank of total speakers and native speakers in 50 years:

Table 7: Total Speakers Rank

Languages	**Population**	**Rank**
English	3207	1
Chinese	1687	2
Hindi	1324	3
Arabic	1005	4
Malay	709.8	5
Bengali	586.7	6
Hausa	568.3	7
Spanish	566.3	8
French	564.7	9
Russian	547.4	10

Table 8: Native Speakers Rank

Languages	**Population**	**Rank**
Chinese	1203	1
English	839.8	2
Arabic	741.1	3
Hindi	696.7	4
Bengali	551.5	5
Spanish	417.0	6
Hausa	374.9	7
Punjabi	334.2	8
Russian	278.6	9
Portuguese	268.4	10

Based on the graphs and tables, we find the results are quite convincing:

- Population has a great influence on the number of both native speakers and second speakers:

  - It is reasonable that Chinese has the most native speakers and the second most total speakers.
  - It is reasonable that Malay has a large number of second speakers due to its high birth rate and hight c_{ed}.
  - It is reasonable that Japanese and Portuguese have not changed greatly due to the low birth rate.

- Based on the number of immigrants, we find the results are quite convincing:
  - It is reasonable that English has the most second speakers.
  - It is reasonable that French are on the list of top ten total speakers but are not in the list of top ten native speakers since French pan-country is the second largest destination country.

4.5.3 Part I problem C

Based on matrix I (**equation 6**), we can calculate the geographic distribution of various languages in 50 years. We only show the results of English and Chinese due to shortage of time:

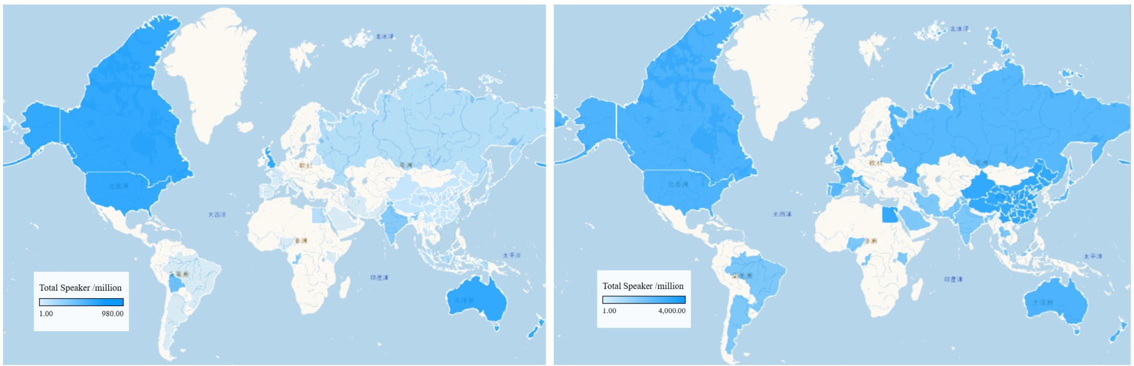


Figure 7: Current (left) and Future (right) Geographical Distribution of English

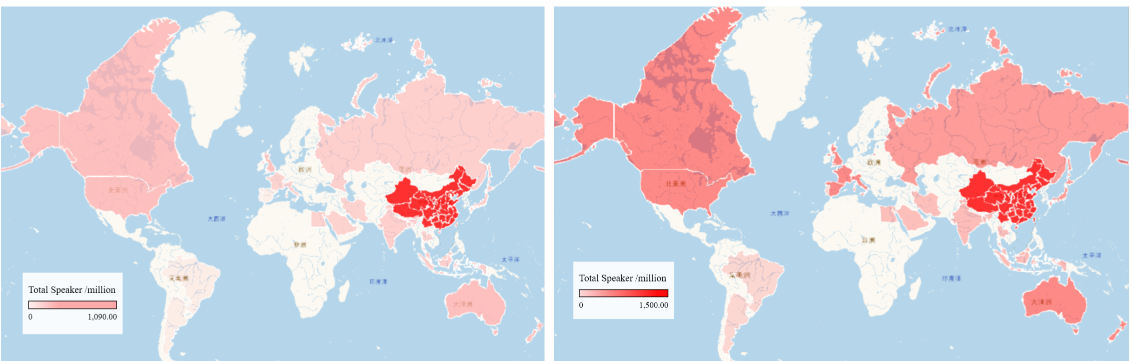


Figure 8: Current (left) and Future (right) Geographical Distribution of Chinese

5 Task II

5.1 Model III: Maximum Benefit Model

5.1.1 Model III Design

Company always wants to maxmize the total profits. So the criterion of our model is to evaluate the score of each plan. It is obvious that 6 offices should be placed in developed areas. However, if offices are too close to each others, the effiency of our campany will decrease since lots of areas are far from our offices.

5.1.2 Model III Updating

Determine the comprehensive power of a country We use geographic distribution results in Part I to describe the comprehensive strength of each country:

- If a country have a large rate of P_o with P and a low rate of P_d with P, this country has a developed education system.
- If a country has large number of people who can speak English and another languages other than dialects and official language, this country has high opening degree.
- Both the population and the economic power has positive relationship with its influence, so we add GDP to our model.

We define:

$$s_i = k_1 \cdot \frac{p_1 \cdot p_3}{p_2} \cdot GDP \tag{14}$$

NOTE:

- s_i is the comprehensive power of i country.
- p_1 is the rate of P_o with P.
- p_2 is the rate of P_d with P.
- p_3 is the rate of the number of people who can speak English and another languages other than dialects and official language with P.
- k_1 is ratio coefficient. We make $k_1 = 1$.

The influence of distance Apparently, if our offices are in countries with high s_i, we can get good profits. We assume that people will go to the nearest office. So if the route between one office and i country is too long, then i country's contribution is obviously low. We define:

$$B_i = k_2 \cdot \frac{s_i}{d_i^{distance}} \tag{15}$$

NOTE:

- B_i is the contribution to our campany, which shows how much profits we can get from people in i country.
- d_i will be explained in **Chapter 5.1.2**.
- $distance$ is the weight of distance, which controls the weight of s_i and d_i.
- k_2 is ratio coefficient. We make $k_2 = 1$.

Determine d_i We use Google Map to determine the distance between 23 countries, we define d_{ij} as the distance between i country and j country. So we have distance matrix D_1:

$$D_1 = \begin{bmatrix} d_{1,1} & \cdots & d_{1,23} \\ \vdots & \ddots & \vdots \\ d_{23,1} & \cdots & d_{23,23} \end{bmatrix} \tag{16}$$

We exclude the United States and China and choose 6 countries from 21 countries, where our 6 offices are located in. Then we make this 6 countries and the United States and China as our 8 offices, so we have a new distance matrix D_2, and $d_{(ij)}$ is the distance between i country and j office:

$$D_2 = \begin{bmatrix} d_{1,1} & \cdots & d_{1,8} \\ \vdots & \ddots & \vdots \\ d_{23,1} & \cdots & d_{23,8} \end{bmatrix} \tag{17}$$

We assume that people will only choose to go to the nearest office, so we select the minimum distance and get the final distance matrix D_3 :

$$D_3 = \begin{bmatrix} d_1{}' \\ \vdots \\ d_{23}{}' \end{bmatrix} = \begin{bmatrix} min\{d_{1,1}, d_{1,2}, \cdots d_{1,8}\} \\ \vdots \\ min\{d_{23,1}, d_{23,2}, \cdots d_{23,8}\} \end{bmatrix} \tag{18}$$

We still have to alter $d_i{}'$ before we get d_i for calculation:

- if i country has a office, we make $d_i = 1$, so we have:

$$B_{i1} = k_1 \times s_i$$

- If $d_i{}'$ is bigger than $5,000km$, which is the average width of continent, people in i country nearly have no office to go to, so we think the profits from i country equals to zero. So we have:

$$B_{i2} = k_i \times \frac{s_i}{d_i{}^2} = 0.01B_{i1},$$

So we get $d_i = 10$, so we have:

$$d_i = \frac{d_i{}'}{500}.$$

Based on analysis stated above, consider the range of d_i, we have:

$$d_i = \begin{cases} 1 & d_i' < 500 \\ \frac{d_i'}{500} & d_i' \geq 500 \end{cases} \tag{19}$$

5.2 Calculation

5.2.1 A Brief Description

We exclude the United States and China from the object 23 countries, and select 6 countries to locate our offices. So we have C_{21}^{6} schemes. In p scheme, our office use the country's official language to talk. We calculate score for each scheme and select the highest as the best scheme. So we first calculate i country's S_i, then we calculate its contribution to our company, and sum all the countries up. So we have:

$$B_p = \sum_{i=1}^{23} B_{ip}, \tag{20}$$

and this is total profits of p scheme. So the maximum B_p is the best scheme.

5.2.2 Model Algorithm

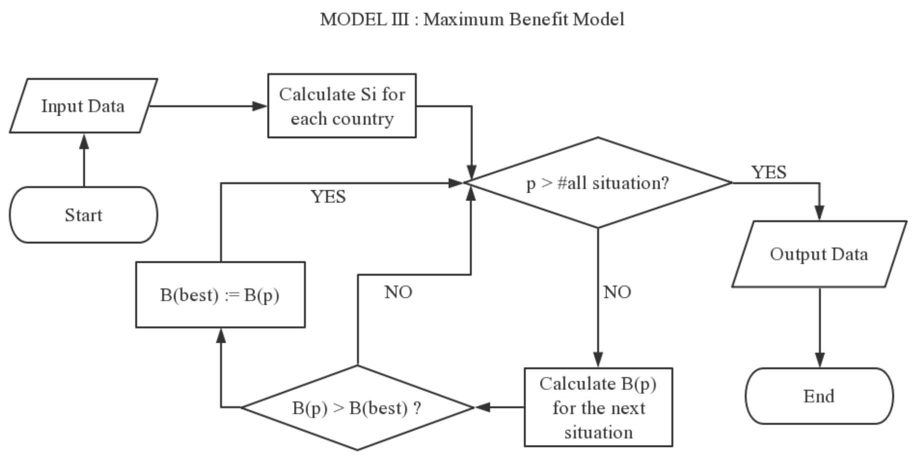


Figure 9: Model Algorithm for Part II

5.3 Conclutions

5.3.1 Part III Problem A

We discuss results in two situation: one in 10 years (short term), the other in 50 years (long term), the results are as follows:

- In short term, we locate these offices in:

Countries	America	China	Canada	France	Australia	Japan	Russia	Argentina
Cities	New York	Shanghai	Toronto	Paris	Sydney	Tokyo	Moscow	Buenos Aires

Table 9: Short term scheme

- In long term, we locate these offices in:

Countries	America	China	Canada	France	Australia	Japan	India	Brazil
Cities	New York	Shanghai	Toronto	Paris	Sydney	Tokyo	Bombay	Rio de Janeiro

Table 10: Long term scheme

It is clear that our 8 offices are worldly spread. Besides, due to the poor comprehensive power of countries in Africa, it is not surprising that no office is located in Africa. In view of the rapid development of India, it is reasonable that it replace the state of Russia. All in all, the results are reasonable.

5.3.2 Part III Problem B

What else will influence our model?

- Consider global communication:

With the development of transportation, the price for long time travel is becoming lower and lower. In model III, the influence of d_i decrease through years. So we change the value of $distance$:

$$B = \sum_{i=1}^{n} k_1 \times \frac{s_i}{{d_i}^{distance'}}$$

So the additional information we need is the degree of traffic development, which determine the change of $distance$.

- Consider resources or costs:

Running an additional office needs lots of budget. If the income of one office is lower than its costs, it should be closed down. So it is likely that the best scheme includes less than six international offices. So the additional information we need is the cost of running all the offices c_p.

In conclusion, with the help of additional informations, we amend our model as follows:

$$B_p{}' = k_1 \times \sum_{i=1}^{n} \frac{s_i}{d_i^{distance'}} - c_p \tag{21}$$

6 Sensitivity Analysis

6.1 c_{ed} in Part I Model II

c_{ed} is in **equation (8)**. We assume the number of illiterates equals to P_d, so c_{ed} describe the speed of the change of P_d into P_{od} . We make c_{ed} equals to 0.0008 (in **Table 6**) according to some statistic datas. The number of illiterates in different countries in 50 years are as follows:

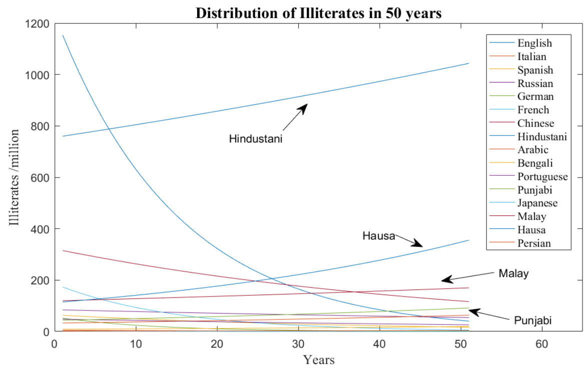


Figure 10: The number of illiterates in different countries in 50 years

It is showed in the graph that only the illiterates in Hindustani pan-country,Hausa pan-country, Malay pan-country and Punjabi pan-country increase. We think it is caused by the low score of s_{ed} of these four pan-countries. We describe the variation of illiterates in Hindustani pan-country and Hausa pan-country when we change c_{ed} :

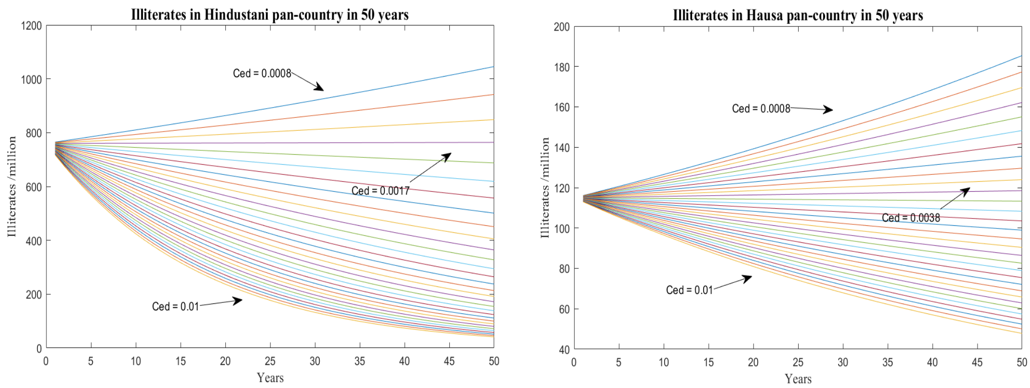


Figure 11: The number of illiterates variation due to c_{ed}

We can see that with the increase of c_{ed}, the number of illiterates in both pan-countries decrease. We also find that only when the education strength of Hindi pan-country increase to $0.0017/0.0008 = 2.125$ times, will the number of illiterates not increase; Only when the education strength of Hausa pan-country increase to $0.0038/0.0008 = 4.750$ times, will the number of illiterates not increase.

6.2 $distance$ in Part II Model III

As we have stated in **Chapter 5.1.2**, the value of $distance$ is closely related with the result. We change the value of $distance$ and find results change:

- When $distance = 1.72$, results of short term scheme change:

Countries	America	China	Canada	**Italy**	Australia	Japan	Russia	Argentina
Cities	New York	Shanghai	Toronto	**Rome**	Sydney	Tokyo	Moscow	Buenos Aires

Table 11: The change of short term scheme caused by $distance$

- When $distance = 1$, results of long term scheme change:

Countries	America	China	Canada	France	Australia	**German**	India	Brazil
Cities	New York	Shanghai	Toronto	Paris	Sydney	**Berlin**	Bombay	Rio de Janeiro

Table 12: The change of long term scheme caused by $distance$

7 Strengths and Weaknesses

7.1 Strengths

- The prediction about language distribution is basically consistent with our expectation.
- The recommended 6 office locations are generally distributed all over the world, which shows that our model comprehensively considers the respective national strength and overall distribution.
- When calculating F_{ij}, We take into consideration many factors, the weights of which are determined according to their importance.
- F_{ij} changes over time with the development of two countries.
- We test every plan of 6 office locations to make sure we find the best scheme.

7.2 Weaknesses

- It is hard to find out every member of a pan-country, especially one whose language is widely uesd in the world. We have to choose several typical countries to investigate. As a result, the average parameters may not precisely estimate the reality of a pan-country.

- The population of a certain pan-country(P) is hard to obtain. So the data we use is not accurate enough.
- People who learn other languages because of personal interest are neglected.

7.3 Future Works

- To make our models work better, we need to find out as many members of a pan-country as possible. It is clear that with more members joining in our models, the models will perform more accurately.
- We can study deeper into different culture circles on Earth. Considering the influence of these culture circles deeper can strengthen our models.

8 References

References

[1] CountriesŠ GDP, GDP per capita, medical condition score, employment condition score, environment condition score, education condition score, safety condition score, connection score:
https://www.usnews.com/news/best-countries/data-explorer#

[2] GDP growth rate:
http://www.kuaiyilicai.com/stats/global/yearly_overview/g_gdp_growth.html

[3] Density of population:
https://goo.gl/VJRVEC

[4] Birth rate and death rate:
http://www.chyxx.com/industry/201801/607184.html

[5] Distance between two countries:
http://www.google.cn/maps

[6] OECD, International Migration Outlook 2012 - Non-Profit Data, 2012

To: Chief Operating Officer

From: 80812

Date: February 13, 2018

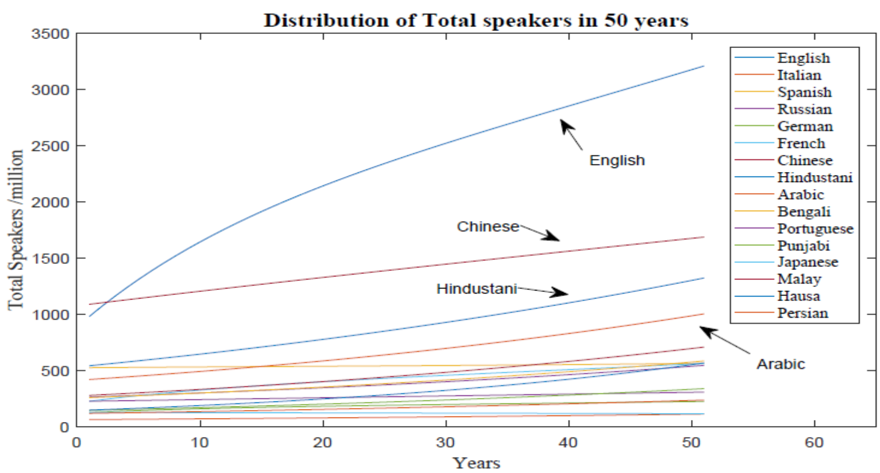
Subject: Determine the location of international office based on language distribution

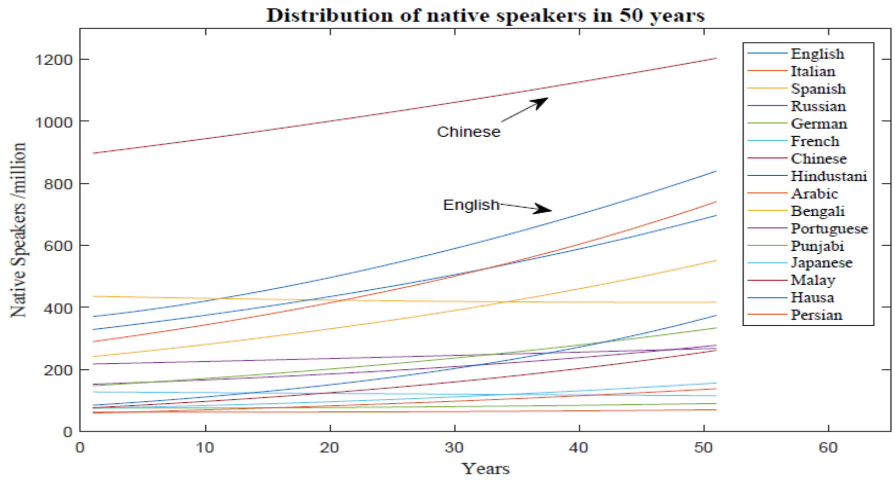
It is our pleasure to help you analyze the trends of language distribution in the future and decide the location of new international offices. We are writing to report the results of our investigation.

As to the trends of global language, an important contributing factor is migration. We build an immigration gravity model to estimate the attraction from one country to another, thus calculating the immigration population.

We take into consideration some other factors, including education, social pressure, globalization, etc. Finally, we make a prediction about the distribution of language(both total speakers and native speakers) in the next 50 years. Results are shown in the figures below.

Apparently, English will be the predominant language in 50 years. People who speaks Chinese or Hindi still take up a significant proportion because of huge population of China and India. The number of people speaking French increases sharply because of immigration.





The total profits of the company consists of the profits gained in each country, which is determined by the national strength of it and the distance between it and the nearest office. We measure the comprehensive national strength of each country according to the distribution of language.

For every scheme of 6 new office locations, we estimate the total profits of the company. Then we pick the best plan which leads to the maximum profits. Here are the recommodations:

In the short term:

Countries	America	China	Canada	France	Australia	Japan	Russia	Argentina
Cities	New York	Shanghai	Toronto	Paris	Sydney	Tokyo	Moscow	Buenos Aires

In the long term:

Countries	America	China	Canada	France	Australia	Japan	India	Brazil
Cities	New York	Shanghai	Toronto	Paris	Sydney	Tokyo	Bombay	Rio de Janeiro

We believe that our model is reasonable because the results are consistent with reality. You are welcome to contact us at any time for further cooperation.

Yours sincerely

Appendices

The programmes we used in our models are as follows:
calculation of Part I

```
%  Solving the Model One & Model Two
%  10/2/2018 8:17
%  Method: Model Difference Equation
clc;
clear all;
%% Parameters Choosing
Alpha = 0.00001;  % Parameters for social pressure
Beta = 0.0000001;  % Parameters for international communication
Gama = 0.044;  % Parameters gravity changes
Sigma = 0.03;  % Parameters for immigration
Loss = 0.2;  % Parameters for dialect lost
Edu = 0.0008;  % Parameters for education
k = 100;  % Parameters for the function Immigrationv.s.GPD


%% Data Input & Variable Definition
load S2
load DistancePanCountry
load Family
Aata = S2;
Aata(:,4:5) = Aata(:,4:5) ./ 1000;
Data = Aata;
[a b] = size(Data);
NUMnative = Data(:,12);  % Native Speakers, Second Speakers, Total Speakers
NUMsecond = Data(:,13);
NUMtotal = NUMnative+NUMsecond;
NUMnone = Data(:,1) - NUMtotal;
for i = 1 : a
    if NUMnone(i) < 0
        NUMnone(i) = - NUMnone(i);
    end
end
Distance = DistancePanCountry;
native = NUMnative;
second = NUMsecond;
none = NUMnone;
PreVal = [];  % #Study English because of social pressure
ConVal = [];  % #Study English because of international communication
ImmigrationWork = [];  % #Immigration because of work
ImmigrationFamily = [];  % #Immigration because of family
ImmigrationTotal = [];  % Table for total immigration
ImmigrationOr = zeros(6,1);  % #Emigration for 6 destination pan-country


%% Problem Solving

%  Calculate the GDP matrix
GDP = [];
for i = 1:a
    for j = 1:6
        GDP(i,j) = Data(j,15) / Data(i,15);
    end
end
learntPre = 0;
```

```
learntConnect = 0;
for T = 1:50 % Study for 50 years step by 1 years

    % Model ONE: consider the social pressure
    NUM = Data(2:end,1);  % Population in a pan-country
    rho = Data(2:end,2);  % Density population of a pan-country
    GPD1 = Data(2:end,3);  % per capita GDP
    Pressure = Alpha * (NUM-learntPre-learntConnect) .* ( rho ./ GPD1 );  % calculate the popul
    learntPre = learntPre + Pressure;
    PreVal(:,T) = Pressure;
    PressureYear = sum(Pressure);
    % Model TWO: consider the international connection
    connect = Data(2:end,6);  % Connection score of each pan-country
    NUM = Data(2:end,1);  % Population of a pan-country
    Connect= Beta .* (NUM - learntConnect-learntPre) .* connect;
    learntConnect = learntConnect + Connect;
    ConVal(:,T) = Connect;
    ConnectYear = sum(Connect);

    % Model THREE: Immigration Gravity Model
    %-----Calculate the gravity matrix-----%
    Fworks = zeros(a,6);  % Gravity matrix for those immigrate for work
    Ffamily = zeros(a,6);  % Gravity matrix for those immigrate for family
    for i = 1:a
        for j = 1:6
            MED = Data(j,7) / Data(i,7);
            JOB = Data(j,8) / Data(i,8);
            ENV = Data(j,9) / Data(i,9);
            EDU = Data(j,10) / Data(i,10);
            SAF = Data(j,11) / Data(i,11);
            DIS = Distance(i,j);
            Family0 = Family(i,j);
            Fworks(i,j) = k * (MED^0.2425)*(JOB^0.3715)*(ENV^0.0680)*(EDU...
                ^0.0244)/(SAF^(2*0.2539))/(DIS^(2*0.0397))/Family0;
            Ffamily(i,j) = k * (MED^0.1294)*(JOB^0.2797)*(ENV^0.0516)*(EDU...
                ^0.3837)/(SAF^(2*0.1329))/(DIS^(2*0.0227))/Family0;
        end
    end
    Fworks = Fworks .* (GDP.^(Gama*-T));  % Gravity change
    Ffamily = Ffamily .* (GDP.^(Gama*-T));
    %------------------------------%

    %-----Calculate the ratio of each destination pan-country-----%
    SUMworks = sum(Fworks,2);
    SUMfamily = sum(Ffamily,2);
    for i = 1:6  % Do not consider a person immigrate to his own country
        Fworks(i,i) = 0;
        Ffamily(i,i) = 0;
    end
    for i = 1:a
        Fworks(i,:) = Fworks(i,:) / SUMworks(i);
        Ffamily(i,:) = Ffamily(i,:) / SUMfamily(i);
    end
    %--------------------------------%

    %-----Calculate the immigration of each country-----%
    %ImmNUM = Sigma .* Data(:,1) .* Data(:,3);  % population for immigration
    ImmNUM = Sigma .* Data(:,1) .* ( Data(:,3) ./ ((Data(:,3).^2+k )));
    for i = 1:a   % Population for i immigrate to j
        NUMworks(i,:) = 0.5 * ImmNUM(i) .* Fworks(i,:);
        NUMfamily(i,:) = 0.5 * ImmNUM(i) .* Ffamily(i,:);
```

```matlab
    end
    NUMtotal = NUMworks + NUMfamily;
    ImmigrationWork = [ImmigrationWork,NUMworks];  % record the immigration data
    ImmigrationFamily = [ImmigrationFamily,NUMfamily];
    ImmigrationTotal = ImmigrationWork + ImmigrationFamily;
    ImmigrationSum = sum(NUMtotal)';
    ImmigrationOr = [ImmigrationOr,ImmigrationSum + ImmigrationOr(:,end)];
    %---------------------------%

    % Model FOUR: Population Dynamic System
    %-----to all pan-country-----%
    NUMnone = NUMnone .* ( 1 + Data(:,4) - Data(:,5) - Data(:,10) * Edu );
    NUMsecond = NUMsecond .* (1 + Data(:,4) - Data(:,5)) - NUMsecond .* ...
        Data(:,4) .* Loss + NUMnone .* Data(:,10) * Edu;
    NUMnative = NUMnative.* (1 + Data(:,4) - Data(:,5)) + NUMsecond .* ...
        Data(:,4) .* Loss;
    %---------------------%
    %-----destination pan-country-----%
    for i = 1:6
        NUMsecond(i) = NUMsecond(i) + ImmigrationSum(i);
        NUMnative(i) = NUMnative(i) + ImmigrationOr(i,end) * Data(i,4);
    end
    %---------------------%
    %-----to English-----%
    NUMsecond(1) = NUMsecond(1) + ConnectYear + PressureYear;
    native = [native,NUMnative];
    second = [second,NUMsecond];
    none = [none,NUMnone];
    %-----update the population for all pan-country-----%
    Data(:,1) = Data(:,1) .* ( 1 + Data(:,4) - Data(:,5) );


end
total = native + second;
x = 1:T+1;
%-----plot the total speakers-----%
plot(x,total);
axis([0 65 0 3500])
legend({'English','Italian','Spanish','Russian','German','French',...
    'Chinese','Hindustani','Arabic','Bengali','Portuguese','Punjabi','Japanese',...
    'Malay','Hausa','Persian'},'FontSize',10,'FontName','Times New Roman');
xlabel('Years','FontName','Times New Roman','FontSize',12);
ylabel('Total Speakers /million','FontName','Times New Roman','FontSize',12);
title('Distribution of Total speakers in 50 years','FontName','Times New Roman','FontSize',14,'
width=800;
height=450;
left=200;
bottem=50;
set(gcf,'position',[left,bottem,width,height])
%-------------------------


% %-----plot the native speakers-----%
% plot(x,native);
% axis([0 65 0 1300])
% legend({'English','Italian','Spanish','Russian','German','French',...
%     'Chinese','Hindustani','Arabic','Bengali','Portuguese','Punjabi','Japanese',...
%     'Malay','Hausa','Persian'},'FontSize',10,'FontName','Times New Roman');
% xlabel('Years','FontName','Times New Roman','FontSize',12);
% ylabel('Native Speakers /million','FontName','Times New Roman','FontSize',12);
% title('Distribution of native speakers in 50 years','FontName','Times New Roman','FontSize',1
```

```
% width=800;
% height=450;
% left=200;
% bottem=50;
% set(gcf,'position',[left,bottem,width,height])
% %--------------------------


% %-----plot the second speaker-----%
% plot(x,second);
% axis([0 65 0 2500])
% legend({'English','Italian','Spanish','Russian','German','French',...
%     'Chinese','Hindustani','Arabic','Bengali','Portuguese','Punjabi','Japanese',...
%     'Malay','Hausa','Persian'},'FontSize',10,'FontName','Times New Roman');
% xlabel('Years','FontName','Times New Roman','FontSize',12);
% ylabel('Second Speakers /million','FontName','Times New Roman','FontSize',12);
% title('Distribution of second speakers in 50 years','FontName','Times New Roman','FontSize',1
% width=800;
% height=450;
% left=200;
% bottem=50;
% set(gcf,'position',[left,bottem,width,height])
% %--------------------------
```

calculation of Part II

```
% Solving the Model Three
% 11/2/2018 16:37
clc;
clear all;
load Situation1
load Distance
load year50
load year10
Distance = S3;
Distance = Distance'+Distance;
[a b] = size(Memo2);
MAX50 = [];
MAX10 = [];
Final10 = 0;
Final50 = 0;
dist = 0.00001;
for i = 1:a
    Position = find(Memo2(i,:)==1);
    DisNew = [];
    for j = 1:6
        if Position(j) == 1
             DisNew = [DisNew,Distance(:,22)];
        elseif Position(j) == 12
             DisNew = [DisNew,Distance(:,23)];
        else
            DisNew = [DisNew,Distance(:,Position(j))];
        end
    end
    MIN = min(DisNew')'/500;
    for i = 1:23
        if MIN(i)<1
            MIN(i) = 1;
        end
    end
    B10 = year10 ./ (MIN.^dist);
```

```
    B50 = year50 ./ (MIN.^dist);
    SumB10 = sum(B10);
    SumB50 = sum(B50);
    if SumB10 > Final10
        Final10 = SumB10;
        MAX10 = Position;
    end
    if SumB50 > Final50
        Final50 = SumB50;
        MAX50 = Position;
    end
end
MAX10
MAX50
```

Sensitivity Analysis I for c_{ed}

```
% Sensitive Analysis of pan-India country
% 12/2/2018 13:07
clear all;
clc;
Born = 0.0193;
Death = 0.0073;
education = 7;
Edu = 0.0008;
for EDU = 0.0008:0.0003:0.01
    NUMnone = 760;
    none = [];
    for T = 1:50
        NUMnone = NUMnone .* ( 1.012 - education * EDU );
        none = [none,NUMnone];
        T
    end
    x = 1:50;
    plot(x,none);
    hold on;
    xlabel('Years','FontName','Times New Roman','FontSize',12);
    ylabel('Illiteracy /million','FontName','Times New Roman','FontSize',12);
    title('Illiteracy in pan-India among 50 years','FontName','Times New Roman','FontSize',14,'
    width=800;
    height=450;
    left=200;
    bottem=50;
    set(gcf,'position',[left,bottem,width,height])

end
```

Sensitivity Analysis II for c_{ed}

```
% Sensitive Analysis of pan-Pakistan country
% 12/2/2018 13:15
clear all;
clc;
Born = 0.0380;
Death = 0.0128;
education = 3;
Edu = 0.0008;
for EDU = 0.0008:0.0003:0.01
    NUMnone = 115;
    none = [];
```

```
    for T = 1:50
        NUMnone = NUMnone .* ( 1.012 - education * EDU );
        none = [none,NUMnone];
        T
    end
    x = 1:50;
    plot(x,none);
    hold on;
    xlabel('Years','FontName','Times New Roman','FontSize',12);
    ylabel('Illiteracy /million','FontName','Times New Roman','FontSize',12);
    title('Illiteracy in pan-Pakistan among 50 years','FontName','Times New Roman','FontSize',1
    width=800;%£íűÈčňÏñËØÊý
    height=450;%ÿSSűÈ
    left=200;%¿àÆÁÄżŒóÏÂ¡ÇËőÆ¡¿àÀë
    bottem=50;%¿àÆÁÄżŒóÏÂ¡Çt'źÖś¿àÀë
    set(gcf,'position',[left,bottem,width,height])

end
```