For office use only	Team Control Number	For office use only
T1 ________________	**89439**	F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	**B**	F4 ________________

2018
MCM/ICM
Summary Sheet

Different languages, One earth

Summary

In this paper, we bulid LP (Language Prediction) model and LO (Location Office) model to predict the development trends of different languages in the next 50 years and locate six new international offices worldwide.

The LP model is divided into three submodels. We select 12 (regions) languages and 23 corresponding and main countries based on the geographical distribution of languages, and we divide total language speakers into native speakers and non-native speakers. In submodel 1, we apply Lesley Matrix Population Prediction model and linear fitting to predict the number of native speakers. In submodel 2, we use BP Neural Network to predict the number of non-native speakers acoording to four influencing factors. Then we obtain the number of total speakers of different languages. We find Japanese will be replaced by French. In submodel 3, we bulid Screening model to screen out languages whose geographic distribution will change over the same period of time, and they are Japanese, Italian ,German and French.

In LO model, we adopt discrete programming and Analytic Hierarchy Process to select six new international offices. In the short term, the locations are Mumbai (Hindi) ,Sydney (English) ,Mexico (Spanish) ,Moscow (Russian) ,Paris (French) ,Tokyo (Japanese).In the long term, Tokyo (Japanese) is replaced by Dubai (Arabic). Considering the changeful nature of global communications, we put forward the strategy whether the company open less than six international offices.

In LP model, we analyze the sensitivity of the weight of economy for submodel 2. And it has slight effect on non-native speakers. So the native language propotion P_{Ni} and stability coefficient S_i for submodel 3 are. In LO model, only when $\triangle$C is relative small can our model be stable. Then we discuss strengths and weaknesses.

Keywords: native speakers; Neural Network; discrete programming

Contents

1 Introduction

1.1 Restatement of the Problem

We are required to build models to predict the number of native speakers and total speakers in the next 50 years, and plan six new international offices all over the world.

In our paper, the problem is divided into four parts.

- Build a population prediction model to predict the population of reigons. And find the ralationship between population and the number of native speakers.
- Build the other prediction model considering the factors and influences mentioned in the problem, and predict the number of non-native speakers.Therefore, we can calculate the number of total speakers in the next 50 years.
- Build an evaluation model to screen the languages that will change their geographic distributions.
- Build a programming model to locate six new international offices in short term and in long term, and advise whether the company can open less than six new offices.

2 Assumptions

To simplify the problem and make it convenient for us to model the com-plex conditions, we make following basic assumptions.

- **Ignore extreme events such as asteroid collisions**, which would cause a catastrophic jump in evolutionary trends over time, and possibly render all languages extinct. These disasters are unpredictable or of low possibility.
- **Governments language policy maintain stability and continuity in a long time**, which means official language of one country remain unchanged in the next 50 years.
- **There is a positive correlation between native speakers and population of corresponding region.**Since present language policy is continued , the influence of other factors such as social pressures is rather small compared to the population. So the number of native speakers depends on population of corresponding region.

- **Immigrants will master the local language after a period of time**. There is no doubt that learning native language is beneficial for immigrants to make a living.
- The proportion of all dialects in one country remain relatively stable in a period of time.

3 Notations

Table 1: Notations

Symbols	Definition
i	The i-th region (language) in the list
j	The j-th country in the region
λ	Factors that influence the number of non-native speakers
CN_i	The number of countries of the i-th region
CR	The number of the regions
N_i	The number of the native speakers of the i-th region
M_i	The number of the non-native speakers of the i-th region
$Total_i$	The number of total speakers of the i-th region
P_i	The population of i-th region
P_{Ni}	The native language propotion of i-th region
S_i	The stability coefficient of i-th region
C_i	The priority index of i-th region

4 LP Model:Language Prediction

We divide total speakers into two parts: native speakers and non-native speakers. The factors that influence them are different.

4.1 Submodel 1: Native speakers

Based on the assumptions above, population is the dominant factor in the number of native speakers. In order to study their quantitative relationship, we select 12 languages by the following standards.

- **The top 10 languages currently mentioned in the problem are included.**
- **The added language is the official language of at least one country.**Since the possibility that official language becomes extinct is very small.

- The added language's mainly speaking countries have a great influence internationally. Hence **these added languages may replace the 10 languages in the next 50 years.**

- For lack of specific statistics, we see different dialects spoken in the same country as one language, which means **Chinese includes Mandarin, Wu Chinese and Yue Chinese, Hindi includes Hindustani and Punjabi**.

As to 12 selected languages, we only choose those countries that have leading influences on population, policy and economy.

Table 2: Languages and Regions

Number	Language	Region(Main countries)
1	Chinese	China
2	Spanish	Spain,Argentina,Mexico,Chile,Venezuela
3	English	The United Kingdom, the United States, Australia,Canada
4	Hindi	India,Pakistan
5	Arabic	Saudi Ar0abia
6	Bengali	Bangladesh
7	Portuguese	Portugal,Brazil
8	Russin	Russia
9	Japanese	Japan
10	French	Congo , France
11	German	Germany,Austria
12	Italian	Italy

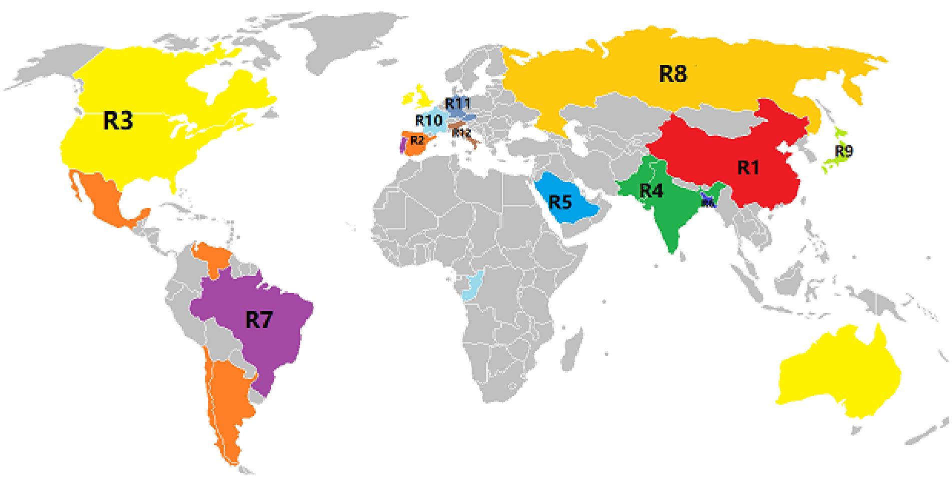

Figure 1: The distribution of the countries and regions

4.1.1 Lesley Matrix Population Prediction model

The basic principle of the selection of population prediction model should meet the natural characteristics of population reproduction, which is determined by three basic elements, namely fertility, death and migration. We find the above countries' immigration rate, birth rate and death rate from World Health Organization. This model has considered **the influence of immigration**.

Now we use the **Lesley Matrix Prediction model** to predict population changes in 21 countries from region R_1 to region R_{12}. The basic expression of the Lesley model is

$$P(t+1) = AP(t) + G(t), \tag{1}$$

where t represents the year.

The lesley matrix is as below.

$$\left\{\begin{array}{c} P_{0(t+1)} \\ P_{1(t+1)} \\ P_{2(t+1)} \\ P_{3(t+1)} \\ \vdots \\ P_{(\omega-1)(t+1)} \end{array}\right\} = \left\{\begin{array}{ccccccc} B_0 & B_1 & B_2 & \cdots & \cdots & B_{\omega-2} & B_{\omega-1} \\ S_0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & S_1 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & S_2 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & S_{\omega-2} & 0 \end{array}\right\} \cdot \left\{\begin{array}{c} P_{0(t+1)} \\ P_{1(t)} \\ P_{2(t)} \\ P_{3(t)} \\ \vdots \\ P_{(\omega-1)(t)} \end{array}\right\} + \left\{\begin{array}{c} g_{00} \\ g_0 \\ g_1 \\ g_2 \\ \vdots \\ g_{\omega-2} \end{array}\right\}$$

where B_x is the number of children who are x years old,x=0, 1, 2, $\cdots$, $\omega - 1$.

S_x is the survival rate of the people who are x years old, $S_x = 1 - m_x$, x =0, 1, 2, $\cdots$, $\Omega - 2$.

P_x is the number of people who are x years old, x=0, 1, 2, $\cdots$, $\Omega - 1$.

G_x is the net immigrants who are x years old, x=00, 0, 1, $\cdots$, Ω-2.

4.1.2 P-N coefficient

Since there is a positive correlation between the change in population and the change in native speakers, firstly, we guess the number of native speakers is proportional to the number of population. However, after **linear fitting**, we find it's more suitable to apply the linear function, namely $N_i(t) = \alpha_i \times P_i(t) + \beta_i$, to describe the relationship.

Here we take Chinese as an example. (Figure 2) The population and the number of native speakers are basically linear.

However, from table3 we can see α_2 and α_8 are bigger than others. In fact, in addition to the selected countries, Russian and Spanish are also the official language of many other countries. Because the distribution of these countries with little international influence is very disperse, we ignore them for simplification.

Table 3: coefficient factors α_i and β_i

Region	α_i	β_i
1	1.9822	-1686.1
2	4.4718	-775.03
3	1.8949	-416.4
4	0.4449	-212.09
5	2.3070	-334.05
6	3.6553	-369.79
7	3.202	488.04
8	5.1324	-565.96
9	1.0103	-21.38
10	1.4529	-114.06
11	0.8794	42.234
12	1.6049	-35.229

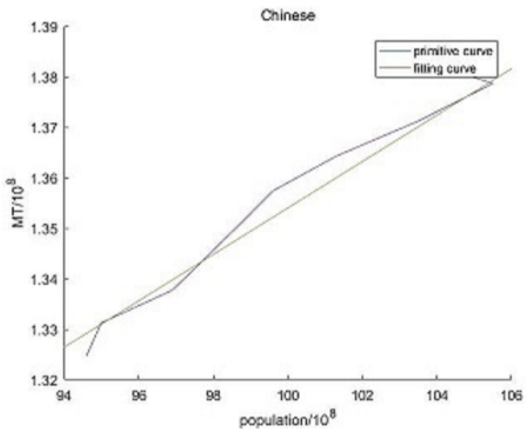

Figure 2: Chinese

4.1.3 Model testing

While calculating P-N coefficient factors, we use collected data from 2008 to 2016. Then we predict the number of native speakers of 12 languages of 2017. We compare the predicted value with real value.

The testing result is as below.

Region	Real value	Predicted value
1	1049	1045
2	436	433
3	371	373
4	477	470
5	290	330
6	242	230
7	218	216
8	153	177
9	128	106
10	76	80
11	76	87
12	63	62

We can see the relative error is rather small, so the sub-model 1 could meet the requirements of prediction.

4.2 Submodel2:Non-native speakers

4.2.1 Weights of factors

By analysis, the factors mentioned that influence the number of non-native speakers could be divided into three categories: economic factors,technical factors, cultural factors. Besides, the difficulty of language is also important.

We obtain:

Table 4: influences and factors

symobl	factor	component
λ_1	Economy	social pressures,international business relations and increased global tourism
λ_2	Communication	the use of electronic communication and social media and technology to assist in quick and easy language translation
λ_3	Culture	migration and assimilation of cultural groups
λ_4	Difficulty	The rank of language's difficulty

Before calculating the weights of regions of any factor, we need to quantify these four factors.

- **Economy**

GDP is often recognized as the best indicator of the state of the economy. Social pressures and international business relations reflect the economic situation at home and abroad respectively. So we use **GDP** to generalize the effects of economic factors.

The weight of i-th region's GDP is

$$\lambda_{1it} = \frac{\sum_{j=1}^{CNi} GDP_{ijt}}{\sum_{i=1}^{CR} GDP_{it}}, \tag{2}$$

Where GDP_{ijt} is GDP of j-th country in i-th region, GDP_{it} is GDP of i-th region, CN_i is the number of countries in i-th region, CR_i is the number of regions.

- **Communication**

  Advanced communication technology promotes the development of second or third languages. According to a new research by Zenith, a media market research firm, the proliferation of smartphones and other mobile devices means making it easier for consumers to access media information. Therefore, **the penetration rate of smartphones** can be a sign of technological factors.

  The weight of i-th region's penetration rate of smartphones is

$$\lambda_{2it} = \frac{\sum_{j=1}^{CNi} Rate_{ijt}}{\sum_{i=1}^{CR} Rate_{it}}, \tag{3}$$

  Where $Rate_{ijt}$ is the penetration rate of smartphones of j-th country in i-th region, $Rate_{it}$ is the rate of i-th region.

- **Culture**

  Language is a good carrier of culture,the countries with high cultural competitiveness are more attractive to the second or third language learners. Hence COV(Cultural output value) is the indicator of cultural factor.

  The weight of i-th region's COV is

$$\lambda_{3it} = \frac{\sum_{j=1}^{CNi} COV_{ijt}}{\sum_{i=1}^{CR} COV_{it}}, \tag{4}$$

  Where COV_{ijt} is COV of j-th country in i-th region, COV_{it} is COV of i-th region.

- **Difficulty**

  There is no doubt that the characteristics of the language itself also affects the number of non-native speakers. That is to say, the difficulty of language impede foreigners' learning.

The weight of i-th region's difficulty rank is

$$\lambda_{4i} = \frac{\sum_{j=1}^{CNi} Rank_{ij}}{\sum_{i=1}^{CR} Rank_i} \tag{5}$$

Where $Rank_{ij}$ is the difficulty rank of j-th country in i-th region, $Rank_i$ is the rank of i-th region.

The data used in the calculation process of λ_1, λ_2 and λ_3 is divided into two parts: statistical data from CEIC's global database, and predicted data from Goldman Sachs Group. What's more, the language difficulty ranking in the world is offered by UNESCO (United Nations Educational, Scientific and Cultural Organization) in 2012.

4.3 Calculating Results

4.3.1 submodel results

We predict 12 regions' population in the next 50 years by Lesley Matrix Prediction model.

In order to cut down the amount of computation and make up for the lack of data, we replace missing data with regional averages, and apply appropriate human intervention to ensure that the final results will be commonsensible.

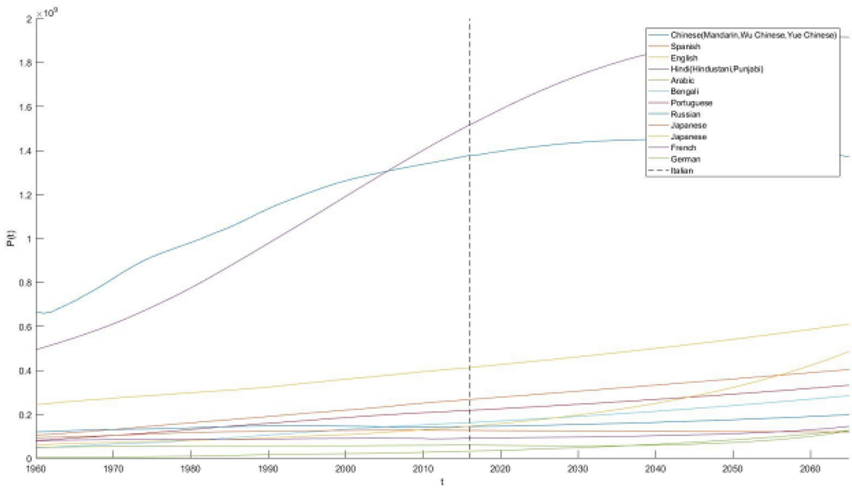

Figure 3: Population of 12 regions

By the relationship equation $N_i(t) = \alpha_i \times P_i(t) + \beta_i$, and P-N coefficients α_i and β_i that we have already calculated, we can obtain the number of native speakers of 12 regions (languages) in the next 50 years.

The specific data of results is in appendix A.

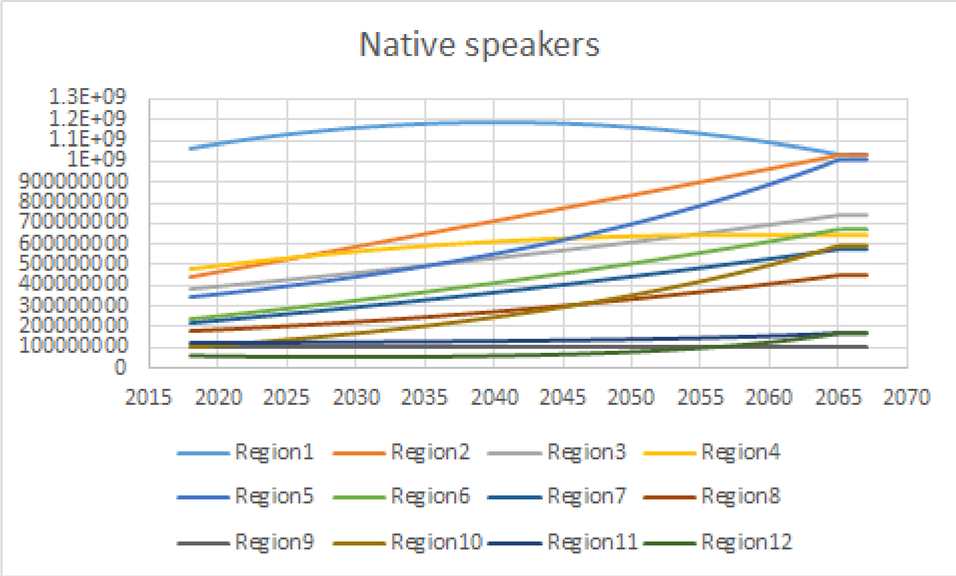

Figure 4: the number of native speakers

4.3.2 Conclusion

We apply **BP Neural Network** to predict the number of non-native speakers. Because $Total_i = N_i + M_i$, we can obtain the number of total language speakers in the next 50 years.

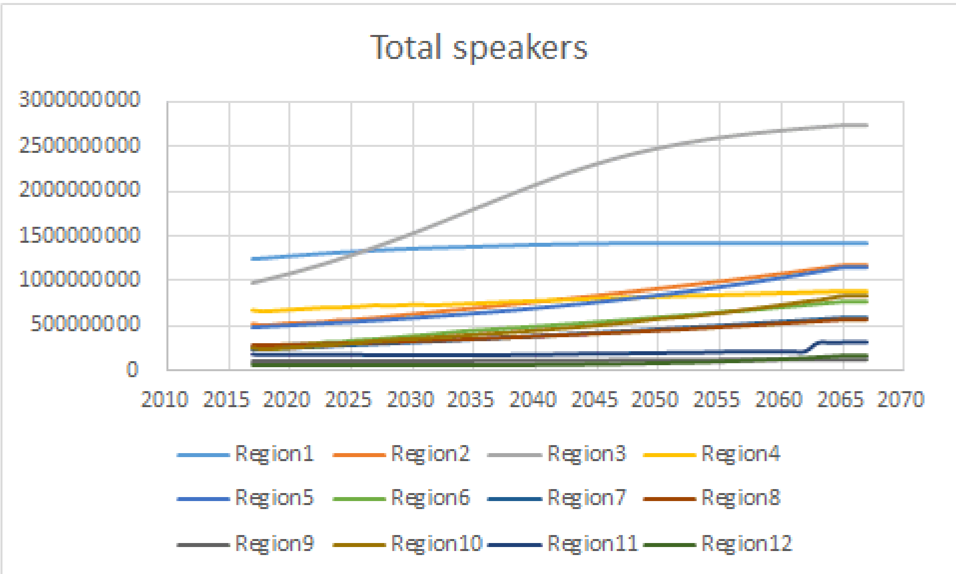

Figure 5: the number of total language speakers

We rank the number of native speakers and total language speakers for 12 regions(languages) in 2067.

From the table 5, for native speakers, no languages in the current top-ten lists

Table 5: Number and Rank

Language	Native speakers	Rank	Total speakers	Rank
English	1005454959	3	2736085829	1
Chinese	1031868606	1	1423868606	2
Spanish	1030742139	2	1168742139	3
Arabic	672179146	5	1141454959	4
Hindi	741085829	4	8912328956	5
French	169989357	10	8249270287	6
Bengali	640232896	6	7631791457	7
Portuguese	591927029	7	6020784495	8
Russian	577078450	8	5734739519	9
German	166483419	11	3059893571	10
Italian	993769863	12	1734834193	11
Japanese	451473952	9	1203769863	12

will be replaced, for total speakers, Japanese will be replaced by French.

4.4 Screening model

The current global human migration patterns is shown in figure 6.

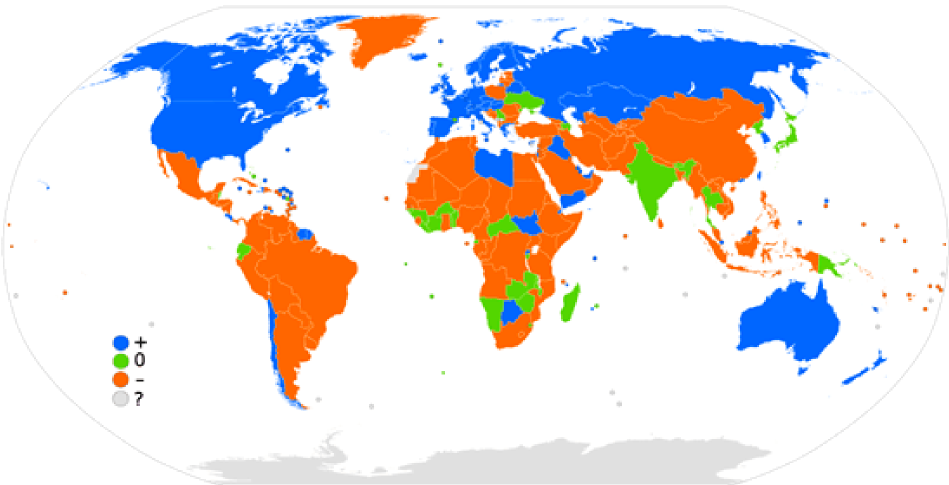

Figure 6: Human migration patterns

The color represents a country's migration type.

$$Color = \begin{cases} Blue & immigration > emigration \\ Green & immigration \approx emigration, \\ Orange & immigration < emigration \end{cases} \tag{6}$$

There is a common trend that people migrate from developing countries to developed countries in a long period. So the migration types will not change in the next 50 years.

We set native language propotion P_{Ni} and stability coefficient S_i to analyze i_{th} language'change over geographic distributions.

P_{Ni} is defined by $P_{Ni} = \frac{N_i}{N_i+M_i}$.

This index weighs the ability that resist the influence brought by immigrantions.

S_i is defined by $S_i = \frac{\sigma_i}{N_i}$.

where σ_i is prosperity coefficient, which is related to economy and culture. Therefore, $\sigma_i = \lambda_{1i} + \lambda_{3i}$.

This index weighs the ability that maintain stable.

However, while solving migration problem, maximum environmental population capacity is also needed to be considered as threshold value. On one hand, current environmental population capacity can meet people's need. On the other hand, the technology is developing. So we ignore it.

4.5 Calculating results and conclusion

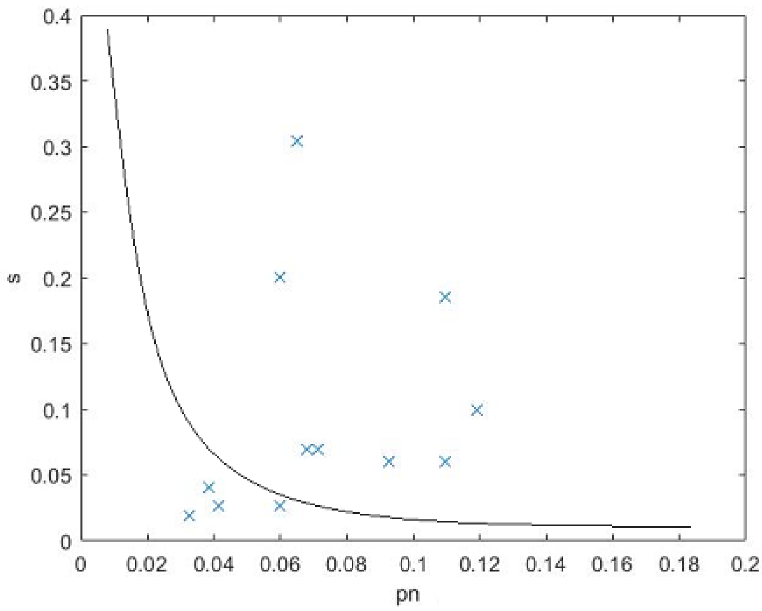

Figure 7: Screening curve

If the value of P_{Ni} is big, the i-th region's proportion of native speakers is large. If the value of S_i is big, the i-th region is stable. Therefore, it's hard for migration to change the geographic distribution of i-th language.

Only when the value of P_{Ni} and S_i are both small, will the i-th language's geographic distribution change.

From the figure 7,Japanese, Italian, German and French will change.

5 LO Model:Location Office

5.1 Discrete programming

We apply AHP and discrete progarmming to locate six new international offices.

5.1.1 The weights of factors

The locations of six new international offices have to consider many aspects. We select four factors to analyze by using data of 2017. However,these factors will change over time. So we predict data 25 years later and 50 years later.

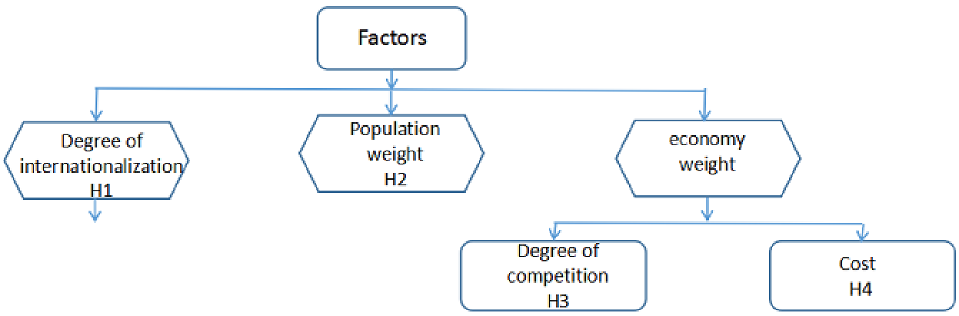

Figure 8: Hierarchy figure

- **Degree of internationalization**

  Client company is desired to open new international offices in the country or region with high degree of internationalization. Since people are more likely to accept foreign capital. What's more, this company desires to have the employees of each office speak both in English and one or more additional languages.So it's necessary to take the degree of internationalization into consideration.

  Based on model 1, the factor can be represented by the number of non-native speakers, namely, $1 - P_i = \frac{M_i}{N_i + M_i}$.

- **Population weight**

  Larger population means more opportunities for client company to expand its market.

- **Degree of competition and cost**

  Both degree of competition and cost are economic factors. Economically prosperous areas are usually highly competitive and high-cost, which makes it harder for companies to make profits.

We construct comparison matrix by comparison method, then we get the weight of each factor.

Factor	Internationalization	Population	Economy
Internationalization	1	2	1
Population	$\frac{1}{2}$	1	$\frac{1}{2}$
Economy	1	2	1

Factor	Degree of competition	Cost
Degree of competition	1	2
Cost	$\frac{1}{2}$	1

Consistency ratio:0.0956<0.1

5.1.2 Priority Index and coefficient of resource waste

We set the priority index $C_{it} = W_1H_{1t} + W_2H_{2t} + W_3H_{3t} + W_4H_{4t}$.Obviously, larger C_i means client company is likely to make more profits in i-th region.

We use smooth curve to connect C_{it} to obtain the function $C_i(t)$.

Considering the fact that closer two offices are, more additional resources will be wasted, we use d_k to represent the distance between the k-th office and the closest one.

We set r_k as the coefficient of resource waste.

$$r_k = \begin{cases} \frac{d_k}{d_0}, & d_k < d_0 \\ 1, & d_k \geq d_0 \end{cases} \tag{7}$$

5.1.3 Objective function and constrains

Objective

$$max \int_{t_0}^{t_1} \sum_{t_0}^{t_1} C_k(t) r_k X_k df$$

In order to reduce the amount of computation, we simplify the objective function.

$$\max r_k X_k(t_1 - t_0) \sum_k \left(\frac{1}{2}(C_k(t_0 + C_k(t_1))\right)$$

Constrains

$$\sum_{i=1}^{NR} X_i = 6$$
$$X_i = 0, 1$$

$t_0 = 2017$. In the short term, we only consider the condition in the next 3 years, $t_1 = 2020$; in the long term, we only consider the condition in the next 50 years, $t_1 = 2067$.

5.2 Conclusion

In the initial calculation,we found that the six selected regions include China. Since there is already an international office in Shanghai in China,we modified the constration $\sum_{i=1}^{NR} X_i = 7$ to select seven regions included China.

Table 6: C_i and Rank

Language	Short term C_i	Long term C_i
Chinese	0.510300787	0.45357119
Spanish	0.208629658	0.216234645
English	0.252978181	0.198330495
Hindi	0.742305825	0.685316579
Arabic	0.124543963	0.135993461
Bengali	0.175510104	0.182203787
Portuguese	0.204704541	0.212156034
Russian	0.147708112	0.149576781
Japanese	0.088905298	0.134488939
French	0.138413294	0.164152775
German	0.095511941	0.135968236
Italian	0.110488296	0.132007078

Our recommendations are different in the short term versus the long term.

Comprehensively considerating the traffic, population and other factors, we selected a country's developed cities as the result of site selection.

In the short term, we advise that the company should locate the new offices in Mumbai (Hindi) , Sydney (English) , Mexico (Spanish) , Moscow (Russian) , Paris (French) , Tokyo (Japanese).

In the long term,we advise that the company should locate the new offices in Mumbai (Hindi) ,Sydney (English) ,Mexico (Spanish) ,Moscow (Russian) , Paris (French) , Dubai (Arabic).

5.3 Strategy considering global communications

There is no doubt that the nature of global communications will also be of great change in the next 50 years . Like how mobile phones in the past ten years changing the way of human life , advanced communication technology, such as Quantum communication, XG networks, Internet of things and Instant translation, will greatly change the pattern of work , trade and relation. Specially, with the development of artificial intelligence technology, the instant translation will be more accurate and more intelligent. This means that communication technology can automatically and accurately transform one kind of language into another when people make a long distance video call or holographic projection. Language will no longer be the barriers among people from different regions.

Therefore, in the long run the company should open less than six international offices, but we still need following information to make a final judgement:

- **The nature of our client's service or the scope of the service.** If my clients mainly provide consulting service for the customers, then six international offices is not necessary . If the business is entity, namely containing specific products or after-sales service, the number of international offices should not decrease.
- **The average profitability of our employers and the expected return level of each region.** If the region's expected earnings are much lower than the average profitability of my client, then decreasing the number of international offices is necessary so that my clients can save resource and invest elsewhere.
- **The climate, policies, customs and social environment of location of international offices.** If local policy does not support foreign investment, or the social environment is very turbulent, or the local customs, religious beliefs conflict with my clients service, for example, selling the refrigerator to the eskimo who live in the North Pole.

6 Sensitivity Analysis

We analyze three kinds of sensitivity in our paper,namely Non-native speakers prediction model,Screening mode and Location Office model.

6.1 Sensitivity analysis for Non-native speakers model

We take Russian non-native speakers as an example. The number of non-native speakers has four influencing factors and we use λ_1 namely economy weight. Shown in figure 9, the greater the economy weight, the more Russian

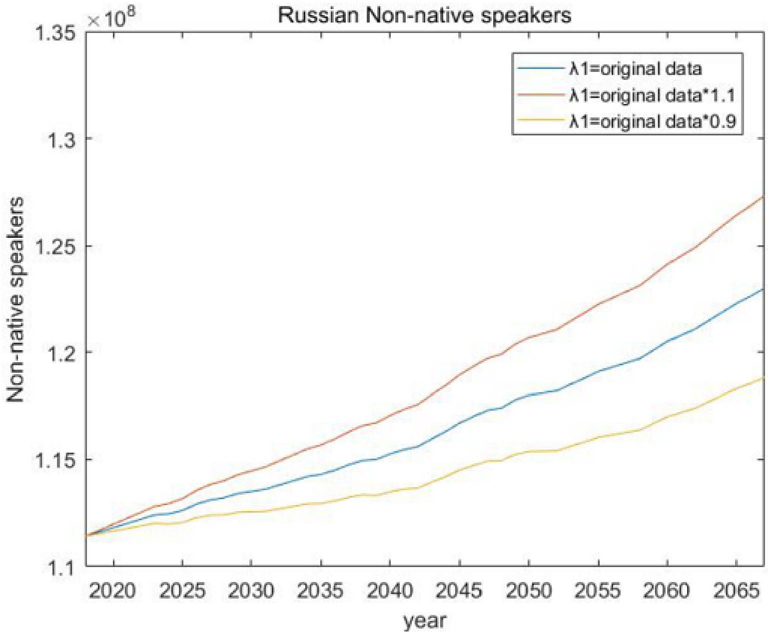

Figure 9: Non-native speakers:Russian

non-native speakers. With the increase of λ_1, the curve has a bigger slope,which is in line with our expectation. When the direction of change is reversed, the slope decreses. This model is stable with the economy weight.

6.2 Sensitivity analysis for Screening model

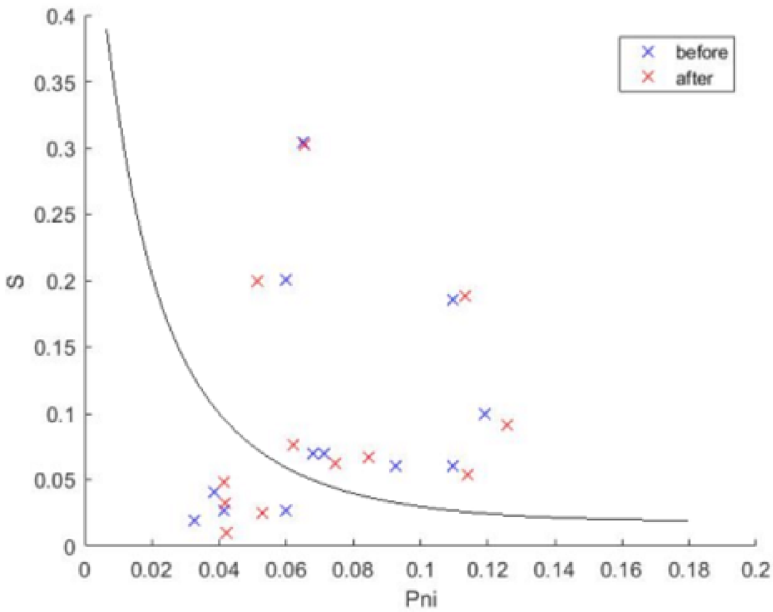

Figure 10: Screening model

In analyzing the sensitivity of the screening model, we use MATLAB to generate random numbers so that P_{Ni} and S_i can make a slight change. In figure 10, the red X represents that there is a small disturbance, and the blue X represents the original results. The languages whose geographic distributions change over time are unchanging. Our model is not sensitive to P_{Ni} and S_i.

6.3 Sensitivity analysis for Location Office model

	short term				long term				
	△C=0	△C=0.002	△C=0.004	△C=0.006	△C=0	△C=0.002	△C=0.004	△C=0.006	△C=0.008
Chinese	√	√	√	√	√	√	√	√	√
Spanish	√	√	√	√	√	√	√	√	√
English	√	√	√	√	√	√	√	√	√
Hindi	√	√	√	√	√	√	√	√	√
Arabic				√	√	√	√	√	
Bengali									
Portuguese									√
Russian	√	√	√	√	√	√	√	√	√
Japanese	√	√	√						
French	√	√	√	√	√	√	√	√	√
German									
Italian									

Figure 11: Location programming model

The sensitivity analysis of this model is mainly to analyze the priority index C_i, and △C represents the degree of change of C_i. As we can see from the figure 11, our results do not change in the short term when △C is less than 0.004, and when △C is greater than 0.006, the calculation results change.The Arabic language replaces Japanese. Howerver,in fact, we use the normalization method to do data processing and the value of C_i is small. Since △C= 0.006 is large, our results are not sensitive to C_i.

7 Strengths and weaknesses

7.1 Strengths

- **Two models for total speakers**

  Considering the influences of the number of native speakers and non-native speakers are different,we bulid two models to predict.

- **Significantly fitting well**

  According to the results of model testing,there is a good linear correlation between population and the number of native speakers.So the predicting results are reliable.

- **The solution to immigration problem**

Immigration problem is complex and hard to be quantified while calculating the number total speakers.Hence we add this factor into population prediction model.

7.2 Weakness

- **Lack of accurate data**

  While processing data,we use difference method to solve the problem of missing data.However,it still affects the accuracy of the results.

- **Limited to time**

  We do not have time to study Mandarin,Wu Chinese,Yue Chinese, Hindustani and Punjabi seperately,which needs further discussion in the future.

8 Memo to the Chief Operating Officer

To:Chief Operating Officer of the service company

Date: February 12, 2018

Subject: Our research results and recommendations

In response to your questions regarding to investigate trends of global languages and location options for new offices, we are writing to inform you of our findings.

Our research conclusions contain following two parts.

Firstly, we research the trends of global languages. We generally divide total speakers into native speakers and non-native speakers, and we select 12 (regions) languages and 23 corresponding and main countries based on the geographical distributions of languages. Two sub-models are built to predict the number of native speakers and non-native speakers respectively.

The first submodel is Native speakers model(We consider the impact of immigration), which helps us to predict the number of native speakers. As to the second submodel, we use our algorithm to predict the number of non-native speakers acoording to four influencing factors. Screening model evaluating the stability of every language is based on human migration patterns. Only when the value of native language propotion and stability coefficent are both small, will the languages geographic distribution change. On the whole, we come to the following three conclusions:

- In the next 50 years, the total number of English speakers will reach 2.7 billion and English is dominant. The followings are Hindi, Chinese and Spanish.
- Among top 10 languages, Japanese will be replaced by French regarding to total speakers.
- The geographic distributions of Japanese, Italian ,German and French will change over time.

Secondly ,we research the location options for new offices. Based on the model above, we bulid another model to find the optimum regions for the six new offices in the short term and in the long term. Our goal of the programming is to make maximum profits when the locations are enough dispersive. And we locate the new office in the main city of the main country of the selected region.

In the short term(3 years), we advise that you should locate the new offices in Mumbai (Hindi) , Sydney (English) , Mexico (Spanish) , Moscow (Russian) , Paris (French) , Tokyo (Japanese).

In the long term(50years), we advise that you should locate the new offices in Mumbai (Hindi) , Sydney (English) , Mexico (Spanish) ,Moscow (Russian) , Paris

(French) , Dubai (Arabic). Considering the changeful nature of global communications, we suggest you to open less than six international offices. However, we still need additional information to make the final judgment, such as the nature of our clients service or the scope of the service.

Last but not least, we use Matlab to check the sensitivity and the results of the check are acceptable. For respectively considering influencing factor of native speakers and non-native speakers, the calculation is greatly reduced. Meanwhile, however, the accuracy of our model is influenced.

We hope our findings will be helpful to you.

Sincerely,

Your employees

References

[1] Chen Xiuzhen, Zheng Qinghua. Hierarchical network security threat situation quantification assessment method. *Journal of software*. 2006(04):885-897.

[2] Gao Ziping. Theoretical construction of international migration of specialized human capital and case analysis in India. *East China normal university*.2006.

[3] `https://www.ceicdata.com/zh-hans/products/global-economic-database`

[4] `https://en.wikipedia.org/w/index.php?title=World_language&action=history`

[5] `https://en.wikipedia.org/wiki/Population`

[6] `http://stats.oecd.org/`

[7] `https://www.cnblogs.com/huangfeihome/archive/2012/12/13/2817190.html`

[8] Li Yongsheng. Model selection and parameter identification in population prediction. *Science of finance and economics*.2004(02):68-72.

[9] Ruan Zhengyu. Transnational migration and national security - a study of non-traditional security factors. *Jinan university*.2003

[10] Shi yun. Matlab of BP Neural Network. *Journal of xiangnan college*2010,31(05):86-88+111.

Appendices

Appendix A Prediction Results

	2017	2022	2027	2032	2037	2042	2047	2052	2057	2062	2067
Chinese(M	1049406794	1102797004	1143346027	1171053863	1185920512	1187945973	1177130248	1153473335	1116975235	1067635948	1031487505
Spanish	433063695	493428000	554214739	615430521	677082240	739177091	801722579	864726538	928197139	992142909	1030107463
English	373237794	404954250	438027278	472521169	508503710	546046397	585224673	626118179	668811016	713392039	741075833
Hindi(Hinc	470167720	509229692	543936990	573668454	598015715	616807701	630124120	638297236	641902134	641736568	640168081
Arabic	341100867	376606706	417386925	464225034	518020928	579808185	650773920	732281592	825897203	933419383	1005469048
Bengali	230749016	266228475	303804038	343599540	385746136	430382725	477656416	527723006	580747497	636904642	672193235
Portugues	216378219	246866900	278819546	312303156	347387864	384147083	422657659	463000024	505258370	549520819	577092539
Russian	177268921	193259398	212056051	233687353	258195684	285634680	316066572	349559432	386184260	426011854	451488041
Japanese	106740635	105951975	105168429	104389960	103616537	102848122	102084684	101326188	100572600	99823887.4	99391075
French	101907692	125664962	153344138	185645037	223393746	267565445	319311341	379990488	451207345	534856127	591941118
German	123098641	124630104	126384983	128442833	130920331	133994270	137940732	143203821	150520111	161151371	170003446
Italian	61877922.4	57774776.5	55619822.5	55570031.5	57916352.6	63157179.2	72114264	86126136.5	107383982	139535155	166497508

Figure 12: The number of native speakers

	2017	2022	2027	2032	2037	2042	2047	2052	2057	2062	2067
1	1242406794	1297797004	1342346027	1372053863	1394920512	1413945973	1424130248	1428473335	1426975235	1424635948	1424415515
2	524063695	538428000	592214739	658430521	726082240	795177091	866722579	942726538	1023197139	1111142909	1169289048
3	984237794	1163954250	1383027278	1639521169	1914503710	2174046397	2385224673	2534118179	2633811016	2702392039	2736632738
4	685167720	706229692	740936990	740668454	771015715	797807701	821124120	841297236	859902134	878736568	891779805
5	473100867	511606706	552386925	599225034	653020928	714808185	785773920	867281592	960897203	1069419383	1142001868
6	249749016	292228475	341804038	403599540	457746136	505382725	555656416	608723006	664747497	724904642	763726055
7	227378219	257866900	290819546	325303156	361387864	400147083	439657659	482000024	526258370	572520819	602625359
8	290268920. 8	305259398. 1	325056050. 7	347687353	373195683. 8	401634679. 8	433066572. 1	467559432. 2	506184259. 8	547011853. 8	574020861
9	107740635. 1	107951975. 1	109168428. 5	110389960. 5	111616536. 5	112848122. 4	115084684	116326187. 8	118572600	119823887. 4	120923895
10	254907691. 5	275664961. 8	333344137. 8	374645036. 6	418393746. 3	469565444. 6	528311340. 7	593990487. 7	670207344. 6	763856127. 5	825473938
11	175098641. 2	172630103. 6	166384982. 7	164442833. 3	167920331	175994270. 4	182940732. 2	192203821. 4	202520111	207151371. 1	306536266
12	64877922. 42	60774776. 46	58619822. 55	59570031. 51	61916352. 62	67157179. 25	77114264. 01	91126136. 51	113383982. 4	145535155	174030328

Figure 13: The number of total speakers

Appendix B The code of BP Neural Network

```
clc;
A=xlsread('sjwl.xlsx','A1:DP4');
goal=xlsread('sjwl.xlsx','A5:DP5');
m=max(goal);
goal=goal/m;
pr=minmax(A);
net=newff(pr,[4,1],'logsig','logsig');
net.trainParam.show = 10;
net.trainParam.lr = 0.05;
net.trainParam.goal = 1e-100;
```

```
net.trainParam.epochs = 500000;
net = train(net,A,goal);
y0=sim(net,A) ;
B=xlsread('xls2.xlsx','A1:WB4')
y1=sim(net,B);
y1=y1';
y1=ceil(y1*m);
y1=reshape(y1,12,50);
```