For office use only	Team Control Number	For office use only
T1 ________________	**87583**	F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	**B**	F4 ________________

2018
MCM/ICM
Summary Sheet

Office Determination Based on Language Developemnt

Summary

In this work, we start with constructing a prediction system for analyzing the future distribution of global languages. Based on the prediction results, we develop evaluation models to suggest potential options of office locations.

First, the preliminary task of the prediction system is to characterize the existing distribution of global languages. We choose 10 factors that have direct impact on people's language choices (We choose 10 factors that people rely on to make language choices). All factors of interest will be integrated to form a dominant factor, based upon/relying on which we can build the model for generating language distribution by using available data.

Second, according to the preceding model, we incorporate PSO algorithm into a BP neural network to predict the time-dependent language distribution. Suggested by our prediction model, Punjabi should be replaced by Malay in the list of top-ten languages.

Third, we propose using the method of "Extrapolation of Mathematic Models" to calculate the nation's natural population growth rate. By combining with global population growth rate provided on UN's official website, we can calculate global population statistics for next 50 years. Meanwhile, we use Markov chain model to predict the population migrations in world's 15 major countries, constructing a pattern of most significant global migrations. The resultant pattern and statistics naturally produce the distributions of various language changes over time.

In addition to aforementioned prediction system, we further propose a comprehensive evaluation criterion to extract different cities'features. By analyzing the metrics extracted by SPSS, we identify that there exist a multiple collinearity among these metrics. Accordingly, we use PCA to select and simplify metrics. Then, we use Topsis algorithm to rank cities and choose the new locations for the company in short term — Los Angeles, Beijing, Washington, Hangzhou, Paris and Zurich. The long-term options of the new locations are Los Angeles, Beijing, Hangzhou, Washington, Mumbai and New Delhi.

Finally, we need to consider whether the number of required offices can be reduced. Depending on the scale of number, we could use regression (for small-scale numbers) or neural network (for large-scale numbers) to predict the future development of individual offices. If an office tends to cause deficit, it should be closed. In this manner, the number of offices could be possibly less than 6.

1 Introduction

1.1 Background

There are about 6900 languages being used at present all over the world. Mandarin, Spanish, English, Hindi, Arabic, Bengali, Portuguese, Russian, Punjabi and Japanese are spoken by half of the world's total population approximately. Many people not only speak native language, but also speak other language. As time goes by,different language's user group will change. The functions of the language will also change to two distinct directions. One situation is that the function of the language will expand. The number of its users will rise. Accordingly, its social influence is increase. The other situation is entirely opposite, which means the language may be in danger of extinction. A variety of factors, such as policies, migration and the like, really have an impact on the use of language. Besides, under the background of globalization, the use of language will also be influenced by tourism, network and other factors. Language is a tool for communication. Thus, the distribution of languages in the future cannot be neglected when a multinational company chooses the locations of new offices.

1.2 Restatement of the Problem

Language is an irreplaceable factor which has a significant impact on the selection of the office location. For the purpose of providing reliable suggestions to the large multinational service company, we are required to predict the trend of global language development. On this basis, we choose the location for the new offices.

In the prediction, we apply Markov chain and PSO with BP neural network to predict the number of population and values of other metrics in the coming years.

In the evaluation, we combine the Principal Component Analysis (PCA) and improved Topsis algorithm to rank the preliminary cities. After that, we select 6 best cities and determine the languages that offices ought to use. According to the different conditions of the long and the short term, we propose many useful suggestions.

With the goal of saving resources and making profits, we also consider the interaction between the multiple cities selected in one country, which may cause waste of resource. Afterwards we combine relevant metrics to predict future profits and deficit of each office. The preceding steps help us develop the evaluation system to judge whether an office is qualified or should be closed.

2 Assumptions

- Assume that in each country, everyone can speak their native language.
- No unexpected events affact on the population of countries during the prediction period.
- People's preference of using language do not change in the prediction time.
- During the research, structure of people's choice of language in various countries remained stable.

- Small countries ignored will have a trival impact on the result.
- The people being counted as an language user are those who can skillfully use the language in their life, excluding those who learn the language but not use it as a daily language.
- Due to changes in the living environment, after the migration, the languages that people use will change because of the influence of the new national languages.
- In the modeling process, we do not consider the various regional dialects of each country.

3 Prediction System

3.1 Language Distribution over Time

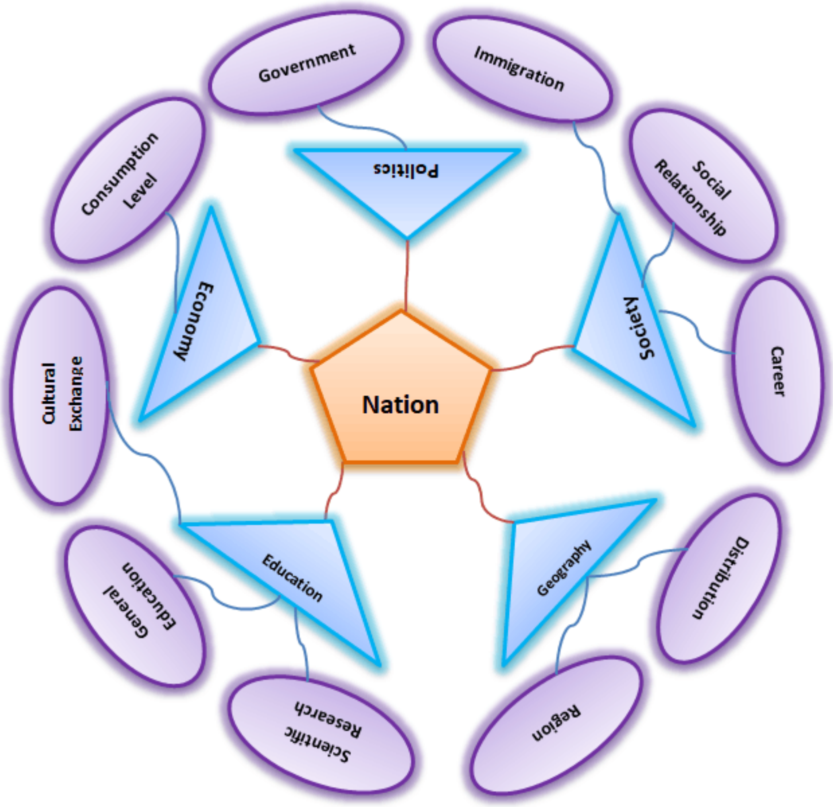

Figure 1: Factors

In this section, our task is to develop a model for various language distribution over time. To deal with this task, our first step is to depict the language distribution. Then we can analyze how the spatial distribution changes over time.

As the problem mentioned, many factors will influence language distribution. Our goal is to find out the most significant factor which has a remarkable impact on people's choice of languages. To achieve this goal, we find some direct factors. By analyzing and generalizing these factors, we conclude the most important factor about language use.

Firstly, we refer to some books and acquire information from the Internet, summarizing ten basic factors — local government, social relationship, career, migration, distribution, region, scientific research, general eduction and consumption level — which directly affect people's choice of languages. Some factors among our summarize may have correlation with each other, such as government is related to region and people's consumption level is also associated with the region.

On this basis, we have to cut down the factors, classifying these factors and generalizing them to more general factors, such as politics and economy, than former step.

Finally, we find all these factors can come down to one major and conspicuous factor — people's nation, which could be the decisive factor that influences the use of languages (Figure 1).

Since people's nation is the only factor we need to consider through our discussion above, we can build the worldwide language distribution model by figureing out major countries' language and search for census in every country (Figure 2).

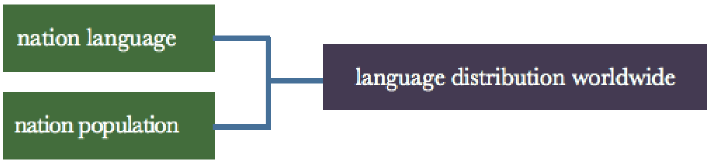

Figure 2: Judgment System

We make a simple graph to illustrate our result (Figure 3):

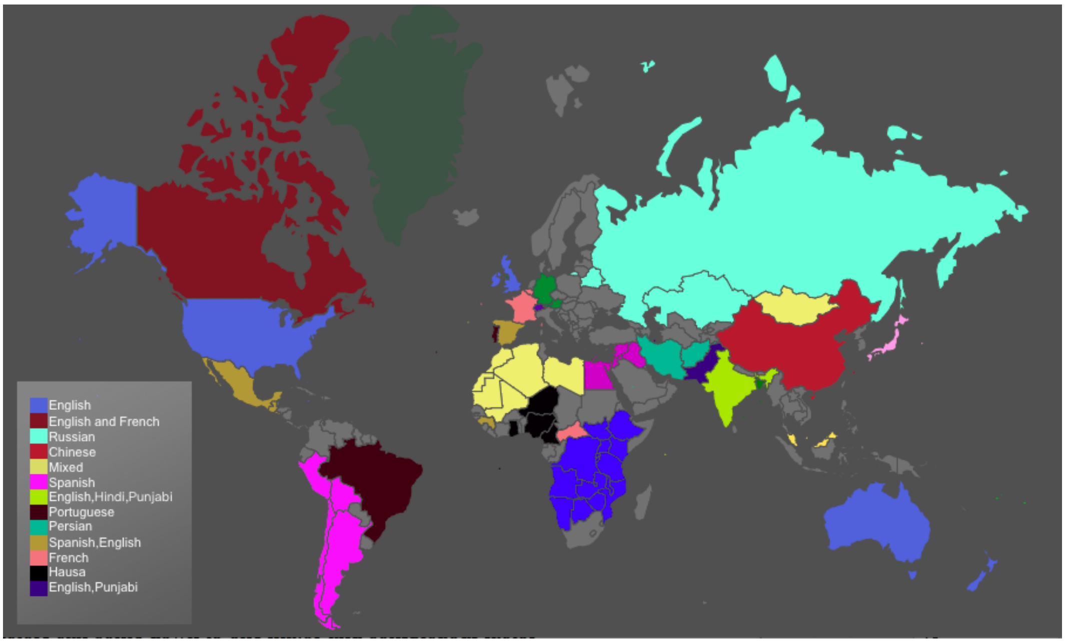

Figure 3: Worldwide Language Distribution

(We only calculate the world's major language and country)

By searching the census data, we can draw the line chart of every major country's population change over time. Accordingly, combining the census data and distribution, we build the language distribution model over time (Figure 4).

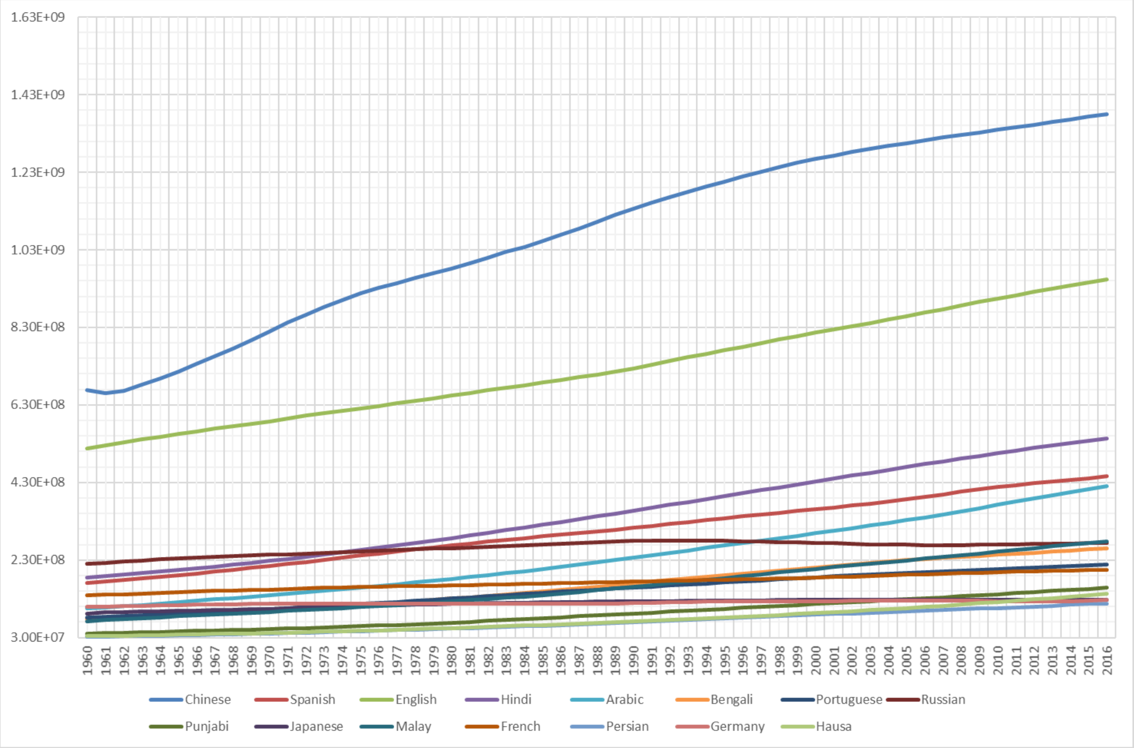

Figure 4: The Changes in Total Number of Language Users in 1960-2016

3.2 Prediction of Language Distribution Change

Based on the model we build in preceding section, in this section, we use Particle Swarm Optimizer and Back Propagation Neural Network(BP Neural Network) Model to predict the numbers of native speakers and total speakers in the next 50 years.

3.2.1 Particle Swarm Optimizer and Back Propagation Neural Network

(1) **BP Neural Network.** Neural Network (Artificial Neural Network) is a common algorithm to fix complicated problems. It mainly solves problem as classification, recognition, prediction, etc. In this section we need to use data from the past to make predications for future.

Back propagation, combining with optimize algorithm, such as descendent decrease, is usually used to train Neural Network. Its most prominent feature is that it could use only sample data to make predictions with highly accurate nonlinear mapping.

(2) **Particle Swarm Optimizer.** In the process of solveing descendent decrease algorithm in BP Neural Network, searching could be limited to partial optimal. So we decide

to use PSO to avoid this problem, since PSO can jump out to local optimal and continue searching for global optimal. Moreover, PSO has a faster convergence speed than normal neural network.

Therefore, we choose Particle Swarm Optimizer and Back Propagation Neural Network to predict the change in language use.

3.2.2 Details of Neural Network

We use language's user distribution in the past 57 year to predict future distribution. Specifically, we use the statistics recorded users' number of every language in past 57 years to predict its future number in 50 years.

The neural network contains 4 layers, including 1 input layer, 2 middle layers and 1 output layer. The middle layers contain 6 nodes.

More details are show in Figure 5.

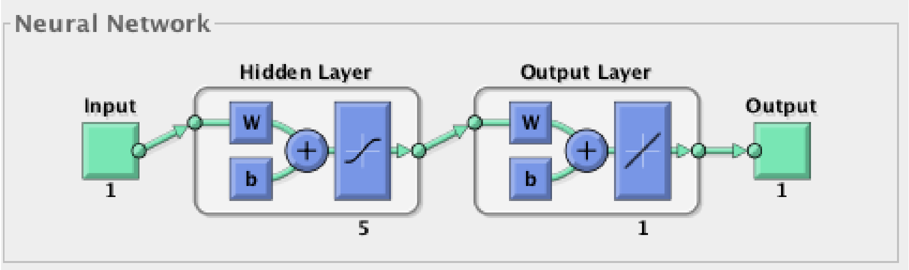

Figure 5: The Details of Neural Network

3.2.3 The Numbers of Speakers in the Next 50 years

From the line chart, we can find that in the future, English will have largest number of users. For total speakers, English, Mandarin, Hindi, Arabic, Spanish, Malay, Bengali, Russian, Persian, Portuguese will be on the top 10 lists(Figure 6 and Figure 7).

3.3 Prediction of Geographical Distribution of Languages

In order to facilitate the analysis, we collect the most populous countries in the world to analyze their development of languages. Based on previous population and its migration patterns, we make a quantitative forecast of their future development.

3.3.1 The World's Major Migration

Note: The color of the dotted line is the same as the color of the destination country.

According to the information collected, we can draw the following conclusions about population migration pattern:

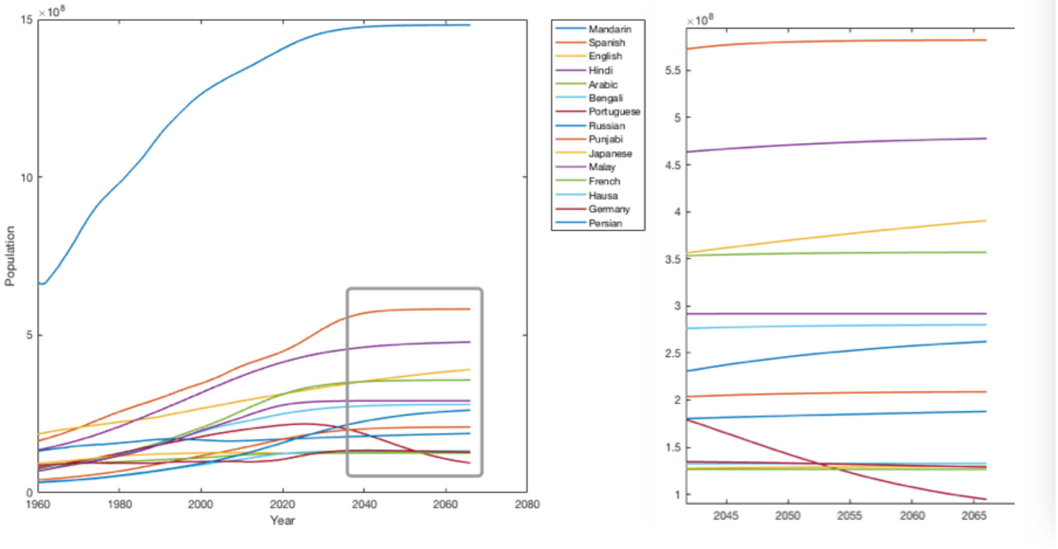

Figure 6: Number of Native Speaker Change Over Time in 50 Years

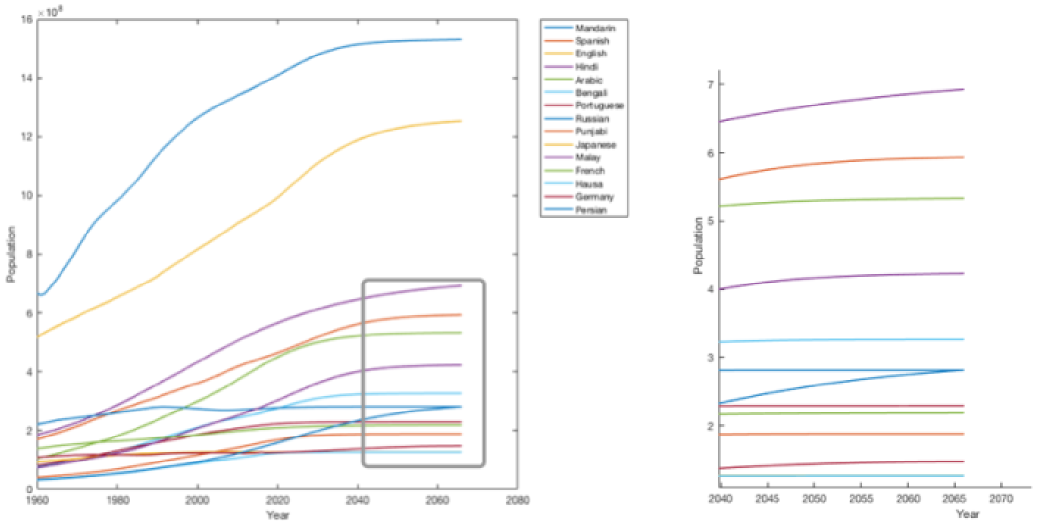

Figure 7: Number of Total Speaker Change Over Time in 50 Years

First, many people migrate from Southeast Asia to the Middle East. This phenomenon is driven by the numerous oil industry of Arabian Peninsula. Second, the United States is the largest destination for immigrants, owing to its developed economy and diverse culture. Third, there is a huge cycle of migration in African countries all year around. Additionally, more immigrants moved to China from Japan, Vietnam, India and Brazil than the number of immigrants from China who moved into these countries than the past. These countries are mainly gathered in Asia.

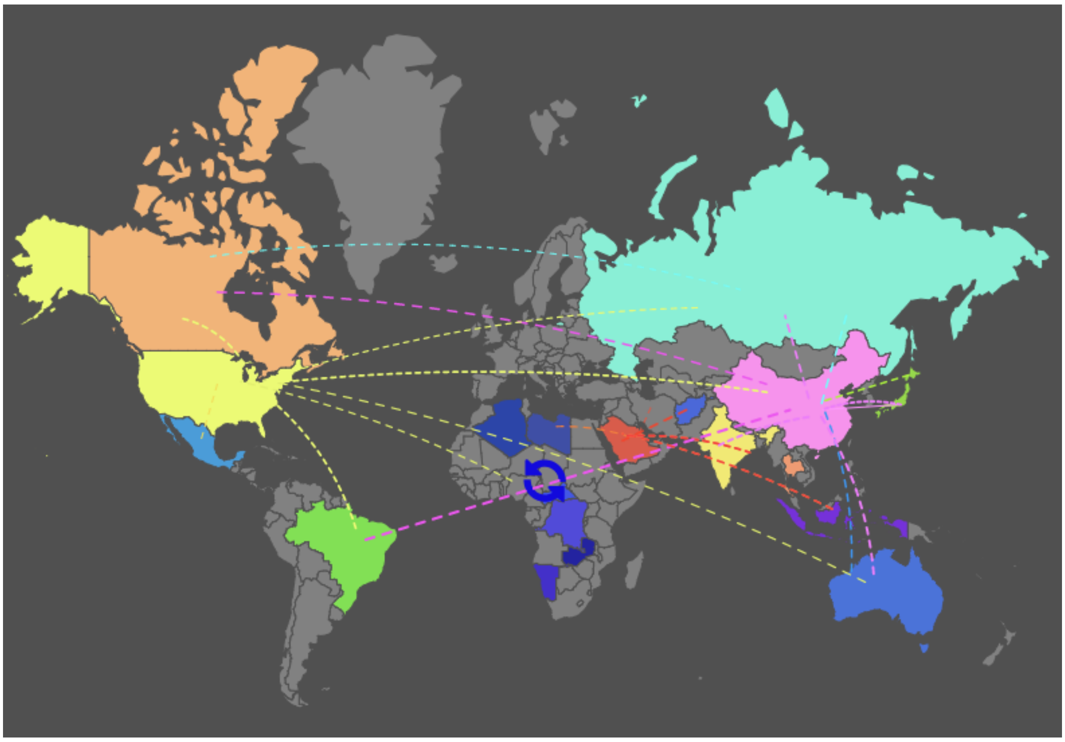

Figure 8: A Sketch of the World's Migration

3.3.2 The Forecast of Global Natural Population Growth

On the basis of preceding steps, the natural population growth is divided into two levels. One is the natural population growth in the world which can use the direct data released by United Nations. Statistics from the United Nations are obtained and used in the population factor forecasting method (the cohort component method). Given that the population is the most influential factor, as for the national level of natural growth, we first select 15 most populous countries. Then, we use the natural growth rates of various countries in 2010-2015 to derive the forecast results through mathematical modeling. However, because of the various factors affecting the population over time will change significantly, the defects of using mathematical modeling to make long-term prediction always exist. As a result, we use the global forecast of population to regulate and optimize the population projections of each country. Under the assumption that we have eliminated the demographic factors in other countries, we have scaled down the global population so that the total of those 15 countries will equal the total global population.

The global population forecast from the UN website is as follows:

Table 1: The Forecast of UN of Global Population Change from 2015 to 2050(Million)

	2020	2025	2030	2035	2040
Global Population	7795.48	8185.61	8551.20	8892.70	9210.34
	2045	2050	2055	2060	2065
Global Population	9504.21	9771.82	10011.17	10222.60	10409.81

The basic process of population forecasting for all countries is:

Step 1: Calculating the population growth rate between 2010 and 2015, we assume that the natural population growth rate will remain unchanged for the next five years:

$$Q_{n,2020} = Q_{n,2015} \times (1 + U_{n,201s5})$$

Step 2: In order to ensure the quantitative consistency between the global population projected by the UN and the population projected by countries, we use the national average as a guideline and distribute the difference equally among the national population projections obtained from the first step:

$$Q_{sum,t} = \sum_{n=1}^{N} Q_{n,t}$$

$$\delta_{sum,t} = Q_{UN,t} - Q_{sum,t}$$

$$\delta_{n,t} = \frac{\delta_{sum,t}}{Q_{sum,t}} \times Q_{n,t}$$

$$Q^{'}_{n,t} = Q_{n,t} + \delta_{n,t}$$

Among them, n and t respectively represent a country and a specific time node (take 5 years as a span). $Q_{n,t}$ is the population at time t of country n. $Q_{sum,t}$ is the total population of all countries at time t. $Q_{UN,t}$ is the UN forecast at time t. $\delta_{sum,t}$ is the error between the two at time t. $\delta_{n,t}$ is the adjusted population of country n at time t. $Q^{'}_{n,t}$ is country n's total population after adjustment at time t. Step 3: Use the adjusted results to correct the natural population growth rate, and then go through the second step many times to get the predicted population of these countries in 2065:

$$U_{n,t} = \sqrt[5]{\frac{Q^{'}_{n,t}}{Q_{n,t-1}}} - 1$$

$$Q_{n,t+5} = Q^{'}_{n,t} \times (1 + U_{n,t})^5$$

Among them,$U_{n,t}$ is the revised natural population growth rate. $Q_{n,t+5}$ is the prediction of population 5 years later. Compare our outcome with the data released by UN (Table 2).

Most country's forecast is similar to the statistics of UN. It suggests that our outcome meets the requirements. Nevertheless, some of these countries, such as China, Russia and German, still have relatively large error compared to the others. The population is still showing a growing trend between 2010 and 2015, and the global population is also growing during the period. Therefore, in general, the population projections of these countries also show an increasing trend. Nonetheless, if we take the other factors such as population policy into consideration, these countries will experience negative growth in the future. This is the source of the error. However, these errors are still within the acceptable range.

Table 2: The Estimation and Verification of World's Population in 2065(Million)

Nation	Model Prediction (2065)	UN Projections (2065)	Error
China	1420.95	1248.12	0.138
India	1710.15	1675.74	0.021
The United States	442.03	412.05	0.073
Brazil	241.25	223.04	0.082
Pakistan	352.66	337.01	0.046
Nigeria	555.40	534.36	0.039
Bangladesh	216.84	201.53	0.076
Russia	143.91	127.96	0.124
Japan	111.10	99.54	0.116
Mexico	173.77	167.25	0.039
Germany	85.07	75.95	0.120
Egypt	184.94	172.96	0.069
Iran	100.48	89.72	0.119
France	77.36	71.56	0.081
The United Kingdom	79.48	77.59	0.024

3.3.3 International Population Movement Forecast

Considering that population migration of the past is dependent from the future, we use the Markov chain model, which has a character that is no after-effectiveness, to predict the population movements in 15 countries. The idea is to emphasize that population migration in different years is a random process. Also, if the Markov chain transition probability matrix is stable in time series, it can be used for long-term prediction. In this question, we need to make a long-term forecast of the number of people in each country. First, we obtain the number of migrants of each country and the whole world from 2000 to 2015 from the UN website. After that, By dividing the number of migrants in each country by the total number of migrants in the world, we acquire the transfer probability matrix taking every five years as a time span:

$$TPM = \frac{M}{n}$$

Among them, TPM is the transition probability matrix. M is the number of population migration matrix. n is the total number of migrants in the worldwide. Having obtained the transition probability matrices of three time periods respectively, we find out that they show a steady state. Afterwards, the average of the three matrices as the transition probability matrix is used to predict the number of migration in each country in 2065:

$$T\bar{P}M = \frac{TPM_{2000-2005} + TPM_{2005-2010} + TPM_{2010-2015}}{3}$$

In the end, the number of migrants from all countries in 2065 is got. Having finished the model tests, it is found that the errors with the Population Division's forecast data are within acceptable limits. Meanwhile, the analysis of the impacts that the migrants in various countries have on the total population of the each country is made.

Table 3: The Total Population and Population Migration in Each Country in 2065

Nation	Model Forecast (Thousand)	UN Forecast (Thousand)	Error	Total Population (Million)	The proportion of Migrants (%)
China	-258.51	-271.20	0.05	1420.69	0.02
India	-347.21	-340.00	0.02	442.86	0.19
The United States	833.22	850.00	0.02	442.86	0.19
Brazil	1.77	1.80	0.02	241.25	0.00
Pakistan	-126.11	-136.00	0.07	352.53	0.04
Nigeria	-47.67	-51.00	0.07	555.35	0.01
Bangladesh	-239.51	-255.00	0.06	216.60	0.11
Russia	91.67	85.00	0.08	144.00	0.06
Japan	40.21	42.40	0.05	111.14	0.04
Mexico	-39.51	-42.40	0.07	173.73	0.02
Germany	164.91	170.00	0.03	85.24	0.19
Egypt	-35.34	-38.20	0.07	184.90	0.02
Iran	-36.87	-34.00	0.08	100.45	0.04
France	64.91	68.00	0.05	77.42	0.08
The United Kingdom	151.57	144.40	0.05	79.64	0.19

Regarding the changes in the geographical distribution of languages, we conclude that it will not change in the forecast time.

3.3.4 The Geographical Distribution of Languages Basically Remains Stable

According to the Question A in Part I, there comes to the conclusion that countries are the biggest factor affecting the geographical distribution of languages. The population of countries and the proportion of people using different languages affect the global geographical distribution of languages. In the assumptions, we ignore the situation which will lead to a huge change in the number of people using a given language. Therefore, only the consideration of the impact of changes in total population is needed.

From the above table, we are able to see that the migration will not have much influence on the total population. The key point, which affects the geographical distribution of languages, is the natural population growth rate. The natural rate of population growth in each country does not affect the geographical distribution of languages, but affects the number of speakers of each language based on their original location.

All in all, the languages' geographic location will not change. The number of people using a certain language at a given location varies with the growing of total population.

4 Evaluation System

4.1 Index System and Model for Evaluating Urban Fitness Based On PCA and Improved Topsis

In this section, we combine the Principal Component Analysis (PCA) and improved Topsis to evaluate the overall situation of 20 cities we selected all over the world from a macro perspective.

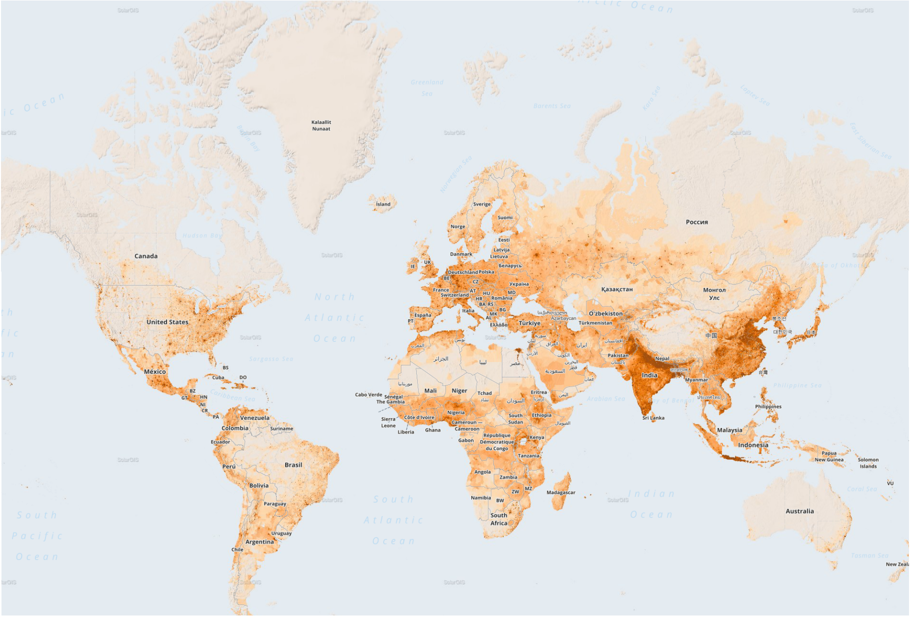

Figure 9: The Map Global Population Distribution

According to the data we obtained in Part I, the world population distribution density map and the related information from the Internet, we select 20 representative cities of 16 countries as candidates. They are Beijing and Hangzhou, China; Mumbai and New Delhi, India; Washington and Los Angeles, the United States; Rio, Brazil; Islamabad, Pakistan; Abuja, Nigeria; Moscow, Russia; Tokyo and Osaka, Japan, Mexico City, Mexico; Munich, Germany; Cairo, Egypt; Tehran, Iran; Paris, France; London, England; Zurich, Switzerland; and Toronto, Canada. Trough checking the statistics websites of selected cities and the UN, we acquire the data of 7 indicators from 2006 to 2015 (Missing data was complemented with SPSS software). Then, use the data to construct the evaluation system.

4.1.1 The Construction of Index System

The choice of the geographical location of the new offices needs to be considered in light of economic and social factors and the like to make a comprehensive assessment of candidates. It is important to consider the comprehensiveness, independence, and their comparability of indicators. After consulting the scholars' research on the company's location on the Internet and combining with the service-oriented premise of the office, we set up an evaluation index system consisting of three guideline levels and seven indicator levels.

Table 4: Index System

Target Layer	Guidelines Layer	Indicator Layer
$NewOffices'$ $Location$	Economic Factors	Residents Consumption Expenditure
		GDP
	Social Factors	Population Density
		Types of Languages
		Number of Second Language Speakerss
	Geographical Factors	Road Network Density
		Land Price

4.1.2 The Determination of the Weight

Have used SPSS to calculate the correlation matrix, we learn that there are serious multiple linear problems between the evaluation indicators. Based on the phenomenon, we use PCA to assign the weights objectively in order to get objective evaluation results.

This approach takes advantage of the idea of dimensionality reduction, extracting independent and not multi-collinear comprehensive indicators form all indicators. The main components contain most of the information in the original data and have no relationships with each other. Therefore, the variance contribution of each index can be used as the index weight.

The procedure for calculating weights based on PCA is as follows:

(1) The Decomposition of Total Variance

Table 5: The Main Components' Contribution Rate

Main Component	Variance Contribution Rate (%)	Cumulative Variance Contribution Rate (%)
1	67.42	67.42
2	18.55	85.97
3	9.36	95.33
4	3.71	99.04
5	0.96	100.00

From the table, the first three main components' contributions are 67.42%, 18.55% and 9.36%, including 95.33% of the original information totally. The bigger variance contribution rate is, the more information it contains. For the aim of achieving dimensionality reduction within the condition of less information loss, we select the first three main components standing for the original data.

(2) The First Three Main Components' Coefficients and Variance Contributions After further calculation and analysis, three main components' coefficients and variance contributions are shown in Table 6.

The first main component can be expressed as:

$$Z_1 = 0.2760x_1 + 0.2221x_2 + 0.2273x_3 - 0.0974x_4 - 0.0176x_5 + 0.1971x_6 - 0.1774x_7$$

Table 6: Coefficients and Variance Contributions of the First Three Main Components

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Z_1	0.2760	0.2221	0.2273	-0.0974	-0.0176	0.1971	-0.1774
Z_2	0.2776	-0.0067	0.1446	0.0487	-0.0426	-0.1072	-0.2272
Z_3	-0.0561	-0.0730	0.0846	0.1416	-0.1290	0.0730	-0.0716

The second main component can be expressed as:

$$Z_2 = 0.2776x_1 - 0.0067x_2 + 0.1446x_3 + 0.0487x_4 - 0.0426x_5 - 0.1072x_6 - 0.2272x_7$$

The third main component can be expressed as:

$$Z_3 = -0.0561x_1 - 0.0730x_2 + 0.0846x_3 + 0.1416x_4 - 0.1290x_5 + 0.0730x_6 - 0.0716x_7$$

(3) Weight Calculation

B_{kj} is the value of the j^{th} indicator of the k^{th} principal component. U_k corresponds the contribution rate of the k^{th} main component.

$$C_j = \frac{\sum_{k=1}^{3} B_{kj}U_k}{\sum_{k=1}^{3} U_k}, j = 1, \cdots, 7$$

Accordingly, we set up a comprehensive score model:

$$F_s = \sum_{j=1}^{7} C_j x_j$$

C_j are the final coefficients after dong the above-mentioned treatment which takes the contribution rates into consideration. Normalize the coefficients for each indicator. After that, the normalized coefficient can be seen as the weight of each indicator. The determinate weight of city evaluation model is listed as follows:

$$W_1 = 0.5302, W_2 = 0.3233, W_3 = 0.4291, W_4 = -0.0990$$

$$W_5 = -0.0727, W_6 = 0.2735, W_7 = -0.3845$$

For long term, we have predicted the 2006-2025 data in the short-term evaluation, so we use the PSO neural network to predict the index data from 2026 to 2040. Finished re-performing the above series of principal component analysis, the weight determination is established:

$$W_l = 0.3945W_2 = 0.2225, W_3 = 0.3800, W_4 = -0.1157$$

$$W_5 = -0.1817, W_6 = 0.3375, W_7 = -0.2015$$

4.1.3 Rank the Cities Based on Improved Topsis Method

After determining the weight , we use the Topsis method to rank cities.

When using the traditional Topsis method to solve multi-objective decision-making problems, there are unavoidable limitations, such as the determination of weight of indicators is to subjective. So we apply the improved Topsis method to evaluate. The final idea of the method is as the follows:

(1) Build a matrix based on the raw data obtained. The spatial matrix X is composed of 20 countries and 7 indicators. Then, In order to make the data comparable, we do communalities of the original matrix to get the normalized matrix $X^{'}$.

$$X = |X_{ij}|_{n\times m}\,, i = 1, 2, \cdots, 20; j = 1, 2, \cdots, 7$$

$$X^{'} = \left|X^{'}_{ij}\right|_{n\times m}\,, i = 1, 2, \cdots, 20; j = 1, 2, \cdots, 7$$

(2) Make the indicators non-dimensional and normalized, then there comes the standard matrix A:

$$A = |a_{ij}|_{20\times 7}\,, a_{ij} = \frac{X^{'}_{ij}}{\sqrt{\sum_{i=1}^{20}(X^{'}_{ij})^2}}$$

(3) After the weight of each index is determined, a matrix A^* of weighted specifications is constructed:

$$A^* = (a^*_{ij})_{n\times m}, a^*_{ij} = w_j \times a_{ij}$$

w_j is the weight of the jth indicator.

(4) Calculate the positive ideal solution and the negative ideal solution in the finite solutions:

$$A^{*+} = (a^{*+}_{i1}, a^{*+}_{i1}, \cdots, a^{*+}_{i7}), a^{*+}_{ij} = max(a^{*+}_{i1}), 1 \leqslant i \leqslant 20, j = 1, 2, \cdots, 7$$

$$A^{*-} = (a^{*-}_{i1}, a^{*-}_{i1}, \cdots, a^{*-}_{i7}), a^{*-}_{ij} = min(a^{*-}_{i1}), 1 \leqslant i \leqslant 20, j = 1, 2, \cdots, 7$$

(5) Calculate the Euclidean distances D^{*+}_i to the positive ideal solution and the Euclidean distances D^{*-}_i to the negative ideal solution of various index values of each candidates respectively:

$$D^{*+}_i = \sqrt{\sum_{j=1}^{m}(a^{*+}_{ij} - a^*_{ij})^2}$$

$$D^{*-}_i = \sqrt{\sum_{j=1}^{m}(a^{*-}_{ij} - a^*_{ij})^2}$$

(6) Calculate the proximity C^*_i of each evaluation object to the positive ideal and sort them. And the bigger C^*_i is, the better the city is.

$$C^*_i = \frac{D^{*-}_i}{D^{*+}_i + D^{*-}_i}, 0 \leqslant C^*_i \leqslant 1$$

Next, Sort cities by size of C_i.

For short-term forecast, the order of the cities ranked in the top six is Los Angeles, Beijing, Washington, Hangzhou, Paris and Zurich (Ordered by grade).

There is another set of cities to choose if we put the predicted data of 2040: Los Angeles, Beijing, Hangzhou, Washington, Mumbai and New Delhi (Ordered by grade).

From the results we can see that the first four cities do not change. This is because these four cities have unsurpassed advantages over other cities. Most obviously, Paris and Zurich are replaced by Mumbai and New Delhi.

It is known that India's population, gross domestic product and the basic transport construction have shown a rapid growth trend in recent years. In contrast, the overall pace of development in Europe is obviously not as good as that in India. This phenomenon is also ultimately reflected in our forecast data, many indicators'data in India exceed the other cities. Owing to it, the result of long-term prediction makes sense.

4.1.4 Proposals for the Choice of Office Collection

Based on the evaluation results, there are two options for the company to choose from: O_1: Open offices in the cities of the US, Europe and China, such as Los Angeles, Beijing, Washington, Hangzhou, Paris and Zurich mentioned above if more attention is paid to short-term benefit. In light of the types and proportions of languages that each city uses, the languages that corresponding offices of the above cities opt are English, Mandarin Chinese, English, Mandarin Chinese, French and German; (Appendix A)

O_2: If we combine the speed of development in various cities and the long-term benefit, it's better to replace those places of countries with slow growth rate with the cities of emerging countries such as India. The chosen cities are Los Angeles, Beijing, Hangzhou, Washington, Mumbai and New Delhi. The languages that corresponding offices of the first four cities opt are English, Mandarin Chinese, Mandarin Chinese, English. Meanwhile, both Mumbai office and New Delhi office need to utilize not only Hindi language but also English. (Appendix B)

4.2 Whether the Number of Offices can be Reduced

In question A above, we decide six cities to set up 6 new offices in the short and long term respectively. Take the company's service-oriented nature and the original intention of establishing the company into consideration, we set up relevant models with the goal of saving resources and making profits in this problem. We not only think about the respective factors of the six cities, but also the interaction between the multiple cities selected in one country, which may cause waste of resource. Additionally, in order to increase the flexibility of the model, we can judge whether we can establish fewer than 6 offices by making predictions on the future profit and loss of each office.

4.2.1 Criterion of Site Selection

Take short term as an example, we obtain six new office locations from the short-term forecast of the previous question, plus the two existing offices, shown as follows:

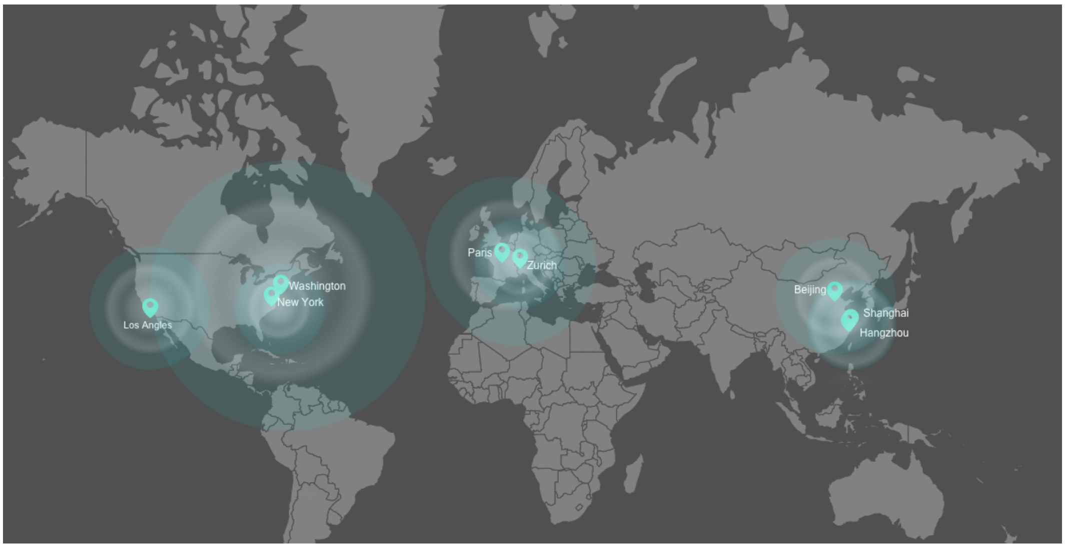

Figure 10: Location of New Offices All Over the World

It's obvious that the service range of each office has a radiation radius. If the distance of two offices is too close, it is likely to cause the overlapping of radiation range. The overlapping part will lead to the waste of resources allocation. When considering about a single city, the best benefit we may gain from setting the office can be quantify. But when the resource waste cost by the new office is greater than the new revenue from it, we need to compare the profits between the two offices to select a city as the location.

From the figure, the locations of the newly established offices are all located in the United States, European and China. Nevertheless, from a geographical point of view, there is a phenomenon of centralization. It is possible to result in waste of resources.

4.2.2 Needful Data and Information

The data and information we need in this model are:

- Cost of setting up the offices (including land price, the number of employees, wage level, etc.)
- Service range of each office
- Relationship between the income and the seven indicators
- Specific values of seven indicators within the radiation range

4.2.3 The Establishment of the Specific Model

The model is divided into two parts–revenue calculation and forecasting:

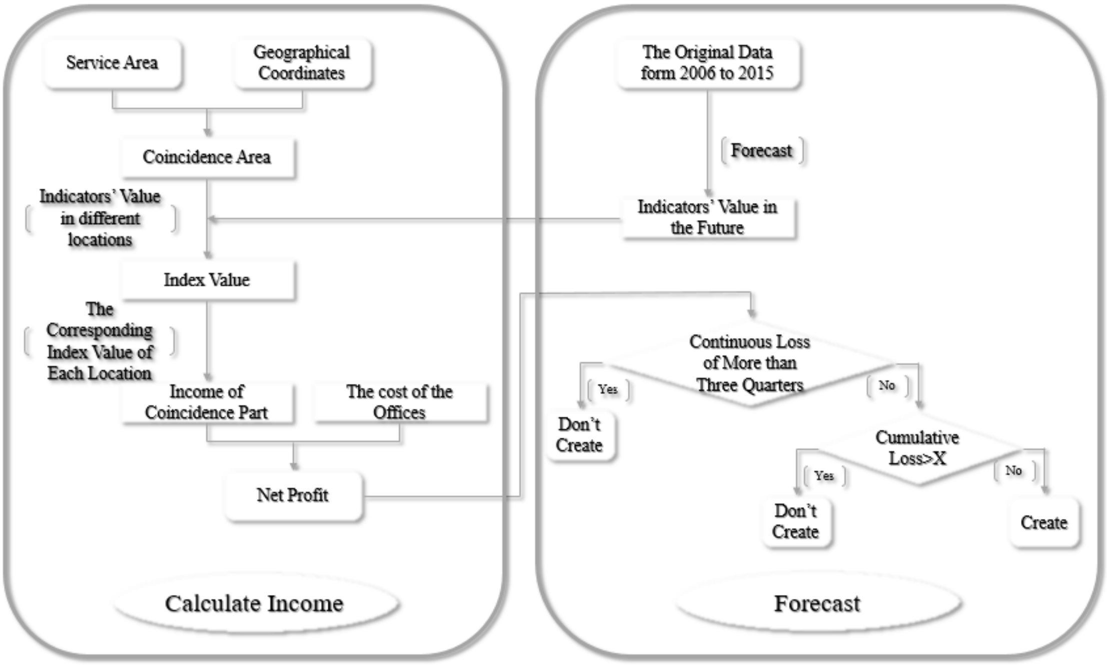
Service Area
Geographical Coordinates
Coincidence Area
Indicators' Value in different locations
Index Value
The Corresponding Index Value of Each Location
Income of Coincidence Part
The cost of the Offices
Net Profit
Calculate Income
The Original Data form 2006 to 2015
Forecast
Indicators' Value in the Future
Continuous Loss of More than Three Quarters
Yes
No
Don't Create
Cumulative Loss>X
Yes
No
Don't Create
Create
Forecast

Figure 11: Decision-Making Model

(1) Revenue Calculation

First, after calculating the overlap area of the service range among the offices, it is possible to calculate the index value of coincidence part using the data of 7 indicators.

Secondly, use the relationship between indicators and income to get the profit of overlap part.

Thirdly, using the total revenue minus the total cost, you can get the net revenue of the new office for a certain period of time.

(2) Forecasting

In the previous question, we predict the future index data by time series and neural network. By substituting these indicator data into the revenue calculation, we can predict the operation of the offices in the future.

And then, we set judgment conditions to decide whether to open the office. Conditions are:

- Three quarters of continual losses in the next ten years
- The cumulative loss reached a certain value and the value needs to be determined through specific data

(3) Suggestions and Conclusions In the event of any of the two cases mentioned above appears, the office will not be opened. At this point, the number of offices is less than 6.

5 Sensitivity Analysis of Predictive Model

We do sensitivity analysis of the prediction model in the first part.

In question B, we use neural network based on particle swarm optimization (PSO) to predict the use of various languages. We put the relevant data into the model and then do the predications for each class of data for 10 times. The average values of the final errors are at the scale of 10_{-3} .

Therefore, it is possible to draw a conclusion that this model has a lower sensitivity.

In Question C, we predict the growth of the population and the number of population migration of each country.

Replace the data and get the results of error(Figure 11 and Figure 12).

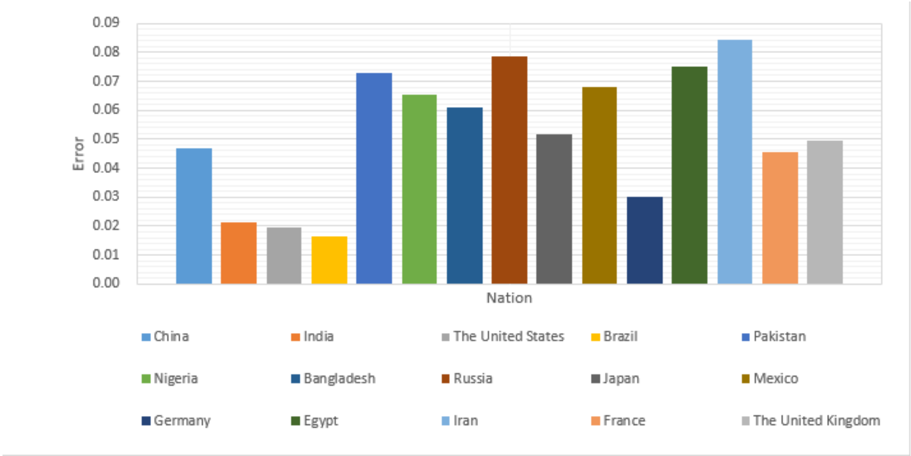

Figure 12: The Prediction Error of Natural Population Growth in Various Countries

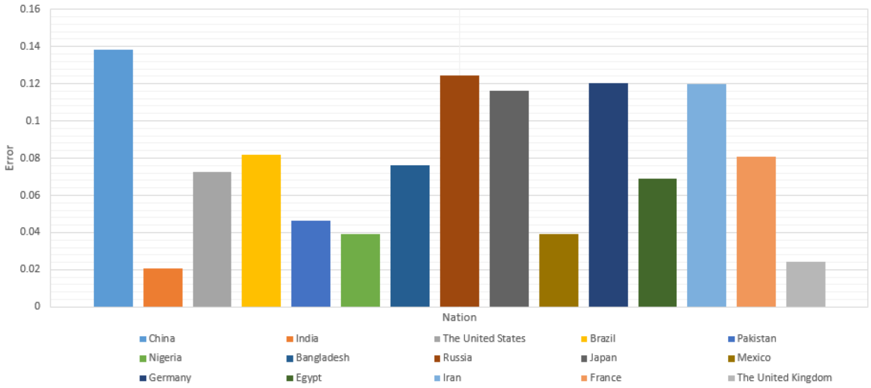

Figure 13: The Prediction Error of the Number of Migrants in Various Countries

Observing the graph, it is obvious that the errors produced by the two models are very small when the data changes.

To sum up, our prediction model is less sensitive and is of universal use, which can be used to predict the data of other companies.

6 Strengths and Weaknesses

6.1 Strengths

- Our models are flexible under different conditions. That is to say, a slight change of parameter will not cause a significant change of the results.
- The application of Neural Network Based on PSO can improve the rate of convergence and avoid local optimum, making the prediction of future data more accurate.
- Our data comes from official websites such as UN Database which is believable. And the combination of PCA and improved Topsis method raise the objectivity of evaluation.
- Based on different conditions for long and short term, we developed the corresponding modulating strategies and come up with suggestions accordingly.

6.2 Weaknesses

- Our models ignore the probability of emergencies and the regional dialects of each country.
- Because of the incomplete migration data, the analysis may be not exhaustive. However, it will not have a dramatic effect on the results.
- Part of the data we use is predicted, which may be a little different from reality.

7 A Memo for the Chief Operating Officer

With the trend of globalization, communication plays a significant role in the development of every company. When a company needs to choose new site for office, it is thorough and wise to think about the language distribution in the world. We are honorable to analyze data for the company with very wisdom.

According to our research, language distribution mainly affect by peopleâĂŹs nation. Therefore the nation population has a direct impact on the global language distribution. We visualize the language distribution, and predict the trend of global language distribution.

In our prediction, most of the language proportion will not change obviously in the future. For the short time, with the explosive speed of population increasing, the usersâĂŹ number of every language will increase. Meanwhile, for long term prediction, language distribution has a tendency to steady, because every countryâĂŹs population tends to become steady in the forecast.

As for the language use top-ten lists, statistics clearly show that most of the current top-ten languages will be replaced in the short term. In other word, language ranking change little inside the list. However, observing data for 50 years, Malays will replace

Punjabi in the top-ten lists, which is a reasonable and acceptable answer for the prediction.

In addition, we further propose the prediction result which combines the migration pattern with the population trend. The result is the distribution of the global population distribution in geography will not change at all. Although the population and migration in the whole world has different develop trend, the distribution is not affected by those changes.

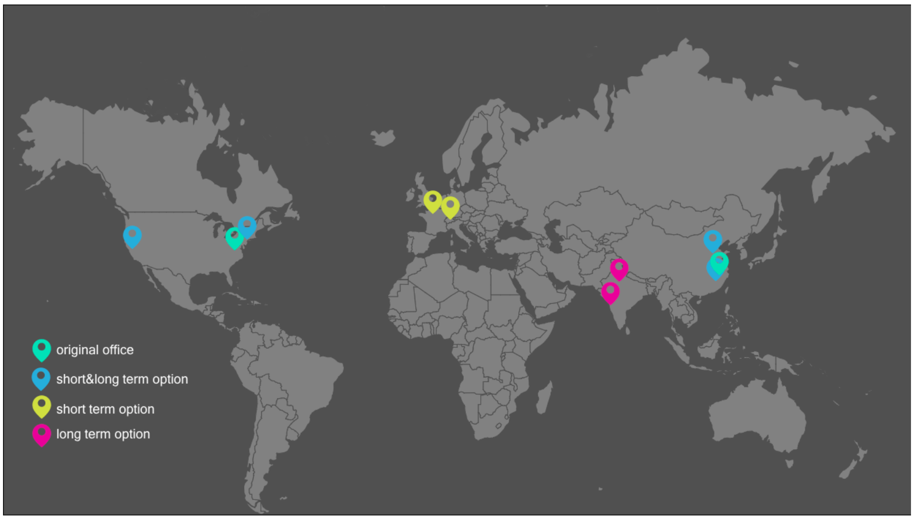

Figure 14: Office Site Option

All results above are actually to serve for the site choice for the company. Since the company already has offices in New York and Shanghai, we propose the suggestion to the company that the office could be locate in these citiesâĂŤLos Angeles, Beijing, Washington, Hangzhou, Paris and Zurich, for the short term benefit. However, if company need to consider set this offices for long term, the decision of cities will change. Los Angeles, Beijing, Hangzhou, Washington, Mumbai and New Delhi will be the better choice.

If the company wants to reduce the number of office, we assume that company would close the office which tends to cause deficit. In this manner, influence between cities should be considered, which means more data, which we are hard to acquire, should be provided to build this model. In short, we already have a comprehensive system to solve this problem, but statistics shortage leads to the difficulty of providing the concise answer. After obtaining the data that needed in this model, the company can apply the model to get result.

References

[1] Dieu Tien Bui, Quang-Thanh Bui, et al. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area[J]. Agricultural and Forest Meteorology, 2017: 32-44.

[2] Esra Akdeniz, Erol Egrioglu, et al. An ARMA Type Pi-Sigma Artificial Neural Network for Nonlinear Time Series Forecasting[J]. Journal of Artificial Intelligence and Soft Computing Research, 2018: 121-132.

[3] Holger R. Maier, Graeme C. Dandy. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications[J]. Environmental Modelling and Software, 2000: 101-124.

[4] K.A. Hoad, A.H. van't Hoog, et al. Modelling local and global effects on the risk of contracting Tuberculosis using stochastic Markov-chain models[J]. Mathematical Biosciences, 2009: 98-104.

[5] Li Li, Lijun Sun, et al. Deterioration Prediction of Urban Bridges on Network Level Using Markov-Chain Model[J]. Mathematical Problems in Engineering, 2014.

[6] Mostafa Mahi, Omer Kaan Baykan, Halife Kodaz. A new approach based on particle swarm optimization algorithm for solving data allocation problem[J]. Applied Soft Computing, 2018: 571-578.

[7] Qingping He, Yibing Lv. Particle Swarm Optimization Based on Smoothing Approach for Solving a Class of Bi-Level Multiobjective Programming Problem[J]. Cybernetics and Information Technologies, 2017: 59-74.

[8] Thamprajamchit S, Ongphiphadhanakul B, et al. A simple prediction rule and a neural network model to predict pancreatic beta-cell reserve in young adults with diabetes mellitus[J]. Medical Association of Thailand. Journal, 2001: 332-338.

[9] Yingjun Wang. Research on Evaluation of Land Intensive Use Based on Improved TOPSIS Method[D]. Wuhan: Huazhong Agricultural University, 2013: 1-87.

[10] Ziyun Wu, Huiming Huang, Xiaoming Yan. Applying TOPSIS to Evaluate the Total Quality of Vocational Students[J]. China Education Innovation Herald, 2009, 16: 238-240.

[11] https://esa.un.org/unpd/wpp/DataQuery/

Appendix

A The Short-term Forecast City Rankings

City	D_i^{*+}	D_i^{*-}	C_i	Result
Washington	0.0044	0.0384	0.8974	3
Los Angeles	0.0005	0.0398	0.9875	1
Beijing	0.0028	0.0395	0.9344	2
HangZhou	0.0057	0.0380	0.8691	4
Toronto	0.0105	0.0264	0.7150	13
Zurich	0.0092	0.0351	0.7915	6
London	0.0102	0.0334	0.7652	8
Pairs	0.0082	0.0346	0.8077	5
Munich	0.0096	0.0303	0.7594	9
Tehran	0.0657	0.0185	0.2197	18
Cairo	0.0797	0.0194	0.1958	19
Mexico City	0.0238	0.0234	0.4959	15
Tokyo	0.0107	0.0349	0.7661	7
Osaka	0.0109	0.0274	0.7154	12
Moscow	0.0230	0.0271	0.5411	14
Abuja	0.0554	0.0134	0.1949	20
Islamabad	0.0324	0.0202	0.3841	17
Rio	0.0252	0.0197	0.4385	16
Mumbai	0.0103	0.0295	0.7421	10
New Delhi	0.0104	0.0296	0.7396	11

B The Long-term Forecast City Rankings

City	D_i^{*+}	D_i^{*-}	C_i	Result
Washington	0.0015	0.0379	0.9625	4
Los Angeles	0.0004	0.0348	0.9884	1
Beijing	0.0011	0.0391	0.9715	2
HangZhou	0.0012	0.0387	0.9703	3
Toronto	0.0080	0.0310	0.7949	12
Zurich	0.0084	0.0327	0.7958	11
London	0.0073	0.0341	0.8228	10
Pairs	0.0059	0.0350	0.8561	8
Munich	0.0061	0.0345	0.8497	9
Tehran	0.0293	0.0214	0.4219	18
Cairo	0.0423	0.0201	0.3219	19
Mexico City	0.0129	0.0297	0.6975	13
Tokyo	0.0059	0.0351	0.8567	7
Osaka	0.0130	0.0279	0.6818	14
Moscow	0.0123	0.0261	0.6801	15
Abuja	0.0443	0.0186	0.2956	20
Islamabad	0.0152	0.0245	0.6174	16
Rio	0.0218	0.0234	0.5175	17
Mumbai	0.0043	0.0375	0.8967	5
New Delhi	0.0059	0.0371	0.8619	6