For office use only	Team Control Number	For office use only
T1 _____	**84802**	F1 _____
T2 _____		F2 _____
T3 _____	Problem Chosen	F3 _____
T4 _____	**B**	F4 _____

2018
MCM/ICM
Summary Sheet

Prediction of languange distribution

Summary

This paper is about to propose a predictable model which is forecasting the number of each native speaker and total speark in the next 50 years. Also, the model would be used to commercial field, it could find a better place to open a office which can make huge benefit towards a service company. And at last, due to the changing nature of global communcation, in order to save the resources of the company, we calculate the profits that the company earn and losses. The results is to suggest the company whether open six office or not.

We first model the trend of number of native speaker based Fuzzy Synthetic Evaluation Model, forecast the change of distribution.

Secondly, we simulate the trend of migration by using Markov Chain Model.

Thirdly, we test the model by using his torical data, we give the criterion that percentage difference should be under 5% As well as we note that the model's strength and weaknesses, which can only refer limited years and the model's creditability would decrease as the time through.

Finally, the model would be detected using the real results.

Keywords: Population; Native speaker ; Fuzzy Synthetic Evaluation Model : Markov Chain Model

Prediction of languange distribution

February 12, 2018

Contents

1 Introduction

1.1 Background

Nowadays,There are 7099 languages around the world. Each of the language has its unique charm. But they are spread unequally throughout the world. That trend is clear whether were looking at whole regions,or individual countries. Under the influence of globalization,the distribution and number of each language speaker are now very different from the past. It is changing all the time. [3]

Moreover,an increasing number of people who learn another languages as second language even third languange or above. Some may know English, Chinese, Spanish and some may know Japanese, Portuguese. These kind of people who require the language job in service company. The head hunter had noticed that the phenomenon.

So we established this model to predict the distribution of the languages in the next 50 years. A further data can improve the business by decreasing the probability of mistakes to open a office. The place would be considered and selected depand on economic index from the model. It could be easy to refer which language would become popular in the corresponding place in the future. Besides, it would be offering the job opportunity directly to someone in need who satisfied the language requirement.Turning job finding to be more convenient. On the other hand, considering the people in these places who can speak more then on language, and they are the main targets to employ. The distribution of number of languages used can be review.

According to the study [2],*"For a company entering a new market, language can be a major barrier that firms may underestimate(Freeman and Sandwell 2008),and understanding language influence across different markets is important for international companies.)"* The sentence above shows the relation between economic development and languange distribution. As we know, the quality of the service demands variety of the languages in the world. The country which attract more tourist visiting , the more understanding and acceptable of foreign language. That's how the top service company can earn more profits from others.

The total number of speaker is mainly affected by the population growth. The following graph shows the relationship between number of speaker and the population

We focus exclusively on the second definition.

1.2 Assumptions

The model is going to ignore unpredictable and high-impact occured, we have to make following assumptions to guarantee the correctness of the model.

- ensure the information is absolutely right,
- the governments won't change the official language in their country,
- ignore the large-scale war, assume it won't break,
- assume the corporate taxes won't change,

Theorem 1.1.

$$\int_{\infty}^{x} x \mathrm{d}F_t(x) \tag{1}$$

Lemma 1.2. *TEX.*

2 Analysis of the Problem

2.1 Overall analysis

The trend of the emigration reviews the trend of language distribution in the next 50 years.

2.2 Key point analysis

2.2.1 Analysis of prediction of native speaker growth

2.2.2 Analysis of prediction of the emigration percentage between countries

The second prdiction is based on the data of Emigration, and we set the model that country to country, and use the percentage of the emigration to form a probability model.According to a study[4] *"It is a quantitative description of the application of Markov mode on the migration"* We note that the trend of emigration is avalible to apply for Markov Chain Model. Trend only depand on the next data but not earlier. The emmigration should satisfy the Markov Chain

Proof. The proof of theorem. □

3 The models

3.1 Notations

We will use the symbols that given in the following table.

Variable	Description
L_i	Number of first(second,third or above) language (i=1 for first,etc)
Eg_i	Number of Emigration of country i
Ig	Number Immigration
δ	Death rate of country j in a year
ℓ_{ij}	The percentage of emigration from Country i to Country j
P	Population
P_{GDP}	Per Capita GDP
Im	Import (dollar)
Ex	Export (dollar)

3.2 The model idea

Due to the lack of data for the number of native speaker, we consider the Fuzzy Synthetic Evaluation Model to simulate the growth of speakers in the next 50 years. As we have found the factor of native speaker's growth has a strong relation with population growth of the countries which take it as official language.

3.3 Fuzzy Synthetic Evaluation Model

We construct this model because it's a command evaluation method to reserve the mainly influence factor. Which affect the countries influences in the future. The factor can be divided at three part in this problem, native speaker, economy and social culture. Those factors are difficule defined as actual value directly. So as to evaluate, we have secondary indicator, using second-level fuzzy syntheic evaluation model. We divide each factors in futher detail. The table will be listed in the followiing:

Primary indicator	**Secondary indicator**
Native speaker	Number of native speakers
Economy	GDP(Gross domestic product) FDI(Foreign Direct Investment)
Social culture	Number of official language school Immigration Language on internet

Table 1: System of language development indicator of each country's influences

As we have constructed a model, there are 3 step to us to quantify the data shown above.

Step 1: Define the weight of each indicator that associate the influences of a language

Step 2:

Step 3: asd

(2)

$$a^2 \tag{2}$$

$$p_j = \begin{cases} 0, & \text{if } j \text{ is odd} \\ r!\,(-1)^{j/2}, & \text{if } j \text{ is even} \end{cases}$$

$$\arcsin\theta = \oiiint\limits_{\varphi} \lim_{x\to\infty} \frac{n!}{r!\,(n-r)!} \tag{1}$$

3.3.1 The model result

3.4 Markov Chain Model

Given that we have the data of population and emigration of each country, we are going to model the distribution and the trend of the shifting population by country.

3.4.1 Futher assumptions

In this model, we assumpt that the

3.4.2 Modelling

We can have the following model based on the data on 2017

$$Let \quad T = \begin{pmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_{1n} \\ \ell_{21} & \ell_{22} & \cdots & \ell_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{11} & \ell_{12} & \cdots & \ell_{nn} \end{pmatrix} \qquad \sum_{i=1}^{n} = \ell_{ij} = 1 \qquad (j = 1, 2, \cdots, n)$$

Matrix T refers to the distribution percentage of the emigration for each countries. And sum of each row is 1. Considering Markove Chain Model, notice that T^n is up-bounded for all integer n when the n is approaching to ∞, we obtain a stable Matrix P.

$$P = \lim_{n\to+\infty} T^n$$

P satisfied the following conditions.

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{11} & p_{12} & \cdots & p_{nn} \end{pmatrix} \qquad \sum_{i=1}^{n} = p_{ij} = 1 \qquad (j = 1, 2, \cdots, n)$$

And P is the patterns of distribution for emigration to each countries in future.

3.4.3 Prdiction in next 50 year

When we get the trend above, we construct the model to descriebe the change within 50 years. In the following,

$$\text{The set of shifting population of country i to others in one year } = (X_i \cdot Eg_i)P \qquad (3)$$

Where X_i and Eg_i refer to the set of country population and Emigration. Considering the death rate would affect the total shifting population, it is observe that can decrease the changing rate of population shifting.

For forecasting the 50 years, the model construction world be like this.
Let V_i = The set of shifting population of country i in next 50 years (each year)

$$V = \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{pmatrix} \qquad \text{n for the number of country}$$

$$V_i = \{(X_{i1} \cdot Eg_i)P, (X_{i2} \cdot Eg_{i2})P^2, \cdots, (X_{i50} \cdot Eg_{i50})P^{50}\} \qquad (4)$$

And we obtain the actual prediction on the distribution of emigration in next 50 year (each year). As for the immigration, we let $P' = P(P' is the transpose of P)$, P reviews the percentage distribution of the country which from others.we can take the same action to establish the model. We let the following Q for infinity of T', the reason is the same to the construction of P:

$$Q = \lim_{n \to \infty} T'$$

After taking the limitation, total coming population of country i in one year can be written as U_i vector form such as before we made.

$$U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}$$

By using the data of immigration of each country form each country, we can write down the following equation.

$$U_i = \{(X_{i1} \cdot Ig_i)Q, (X_{i2} \cdot Ig_{i2})Q^2, \cdots, (U_{i50} \cdot Ig_{i50})Q^{50}\} \qquad (5)$$

After finishing the model, we obtain Matrix V and U , which is direcely indicate the flow of distribution of language in next 50 years.

3.5 Decision-Making Model for choosing the place

The place to build up an office by variety of lanuanges.

4 Calculating and Simplifying the Model

We analysis how the factors affecting the influences of a countries, using the data from [**?**], we had let the weight of each factor in the evaluation model shown in the following.

Primary indicator	Weight of primary indicator	Secondary indicator	Weight of secondary indicator
Native speaker	0.5	Number of native speakers	0.2
Economy	0.3	GDP FDI	0.3 0.2
Social culture	0.2	Number of official language school Immigration Language on internet	0.1 0.05 0.09

Table 2: Elections in Götefrith province, 1900–1910. (Taken from [1], pg. 414.)

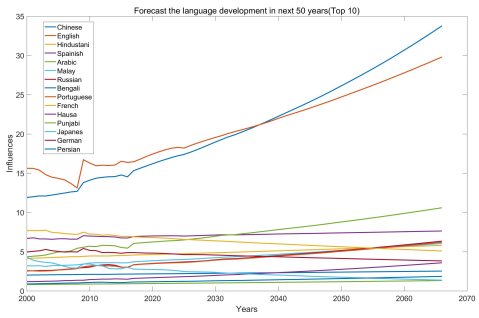

Figure 1: Influences ranking prediction

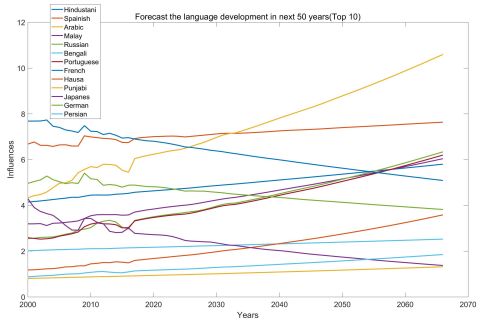

Figure 2: Influences ranking prediction(ignore top 2)

Figure 1 and 2 show how the influence change in the next 50 years. Refer to the figure, Chinese and English would dominate the futher, and Chinese transcend English to be the world comment language in 2038 .

5 Validating the Model

5.0.1 The influences of the model

We use the contral variable method

After the calculation of the Fuzzy Synthetic Evaluation Model, we obtain the global distribution of all total different languages speakers. Then we established economic model to choose the place to open a office. The model cosider the business effect and the profits that the company may receive.

5.1 Growth of influences model

We build up this model based on Fuzzy Synthetic Evaluation Model, growing number of

	1900		1906		1910	
Party	% of Vote	Seats Won	% of Vote	Seats Won	% of Vote	Seats Won
			Provincial Assembly			
Conservative	35.6	47	26.0	37	30.9	52
Socialist	12.4	18	27.1	44	24.8	39
Christian Democrat	49.2	85	41.2	68	39.2	59
Other	2.8	0	5.7	1	5.1	0
Total	100.0	150	100.0	150	100.0	150
			National Assembly			
Conservative	32.6	4	23.8	3	28.3	3
Socialist	13.5	1	27.3	3	24.1	2
Christian Democrat	52.0	7	42.8	6	46.4	8
Other	1.8	0	6.1	0	1.2	0
Total	100.0	12	100.0	12	100.0	13

Table 3: Elections in Götefrith province, 1900–1910. (Taken from [1], pg. 414.)

6 Conclusions

7 Evaluate of the Model

8 Strengths and weaknesses

8.1 Strengths

- **Applies widely**
  This system can be used for many types of airplanes, and it also solves the interfer-

ence during the procedure of the boarding airplane,as described above we can get to the optimization boarding time.We also know that all the service is automate.

- **Improve the quality of the airport service**
  Balancing the cost of the cost and the benefit, it will bring in more convenient for airport and passengers.It also saves many human resources for the airline.
-

[1]

8.2 Weaknesses

- **Policy never change**
  The model works only depand on no any outside force disturb, for instance: Policy won't change, and wherever is stable.
- **Data insufficient**

9 Memo

MEMORANDUM

To: Chief Operating Officer
From: Team #84802
Subject: The best location to open office
Date: February 13,2018

References

[1] John Grossman, editor. *The Chicago Manual of Style*. University of Chicago Press, Chicago, IL, 14th edition, 1993.

[2] Jonas Homqvist. *Language Influence in Services*. PhD thesis, Publication of Hanken School of Economics, 2009.

[3] Gary F. Simons and Charles D. Fennig (eds.). How many languages are there in the world? *Ethnologue*, 2017.

[4] Dazhi Sun Xuqian. Application of markov chain model on environmental fate of phenanthrene in soil and groundwater. Technical report, Jilin Institute of Chemical Technology, 45 Chende Street, Jilin City, Jilin Province, 132022 China College of Environmental and Resource, Jilin University, 6 Ximinzhu Street, Changchun City, Jilin Province, 130026 China, 2010.

Appendices

Appendix A First appendix

Here are simulation programmes we used in our model as follow.

Input matlab source:

```
clear;
A=load('native_speakers.txt');

x=2013:2017;

for i=1:22
    res(i,:)=polyfit(x,A(i,:),1);
end

data = A;
```

```
xx=2018:2068;

for i=1:22
    data(i,6:56)=polyval(res(i,:),xx);
    data(i,57)=i;
end

data=sortrows(data,-56);

for i=1:10
    num(i)=data(i,57);
end

colorspec = {[1 0 0];[0 1 0];[0 0 1];[1 1 0];[1 0 1];[0 1 1];[0 0 0];[0.1 0.5 0.9];[0.9 0.5 0.1

figure;
for i=1:10
    plot(2013:2068,data(i,1:56),'Color',colorspec{i});
    hold on;
    p(i).Color=colorspec{i};
end


hold off;

title('Forecast the total speakers of the different languages in next 50 years(Top 10)');
xlabel('Years');
ylabel('Population(million)');
legend('Chinese','Spanish','Hindustani','Arabic','English','Bengali','Punjabi','Portuguese','Pe
% legend(Country(num(1)),)
set(gca,'FontSize',15);
set(findall(gca, 'Type', 'Line'),'LineWidth',2);
legend('boxoff');
```

Appendix B Second appendix

some more text **Input C++ source:**

```
#include <iostream>
#include <cstdlib>
#include <ctime>

using namespace std;

int table[9][9];

int main() {

    for(int i = 0; i < 9; i++){
        table[0][i] = i + 1;
    }

    srand((unsigned int)time(NULL));

    shuffle((int *)&table[0], 9);

    while(!put_line(1))
```

```
    {
        shuffle((int *)&table[0], 9);
    }

    for(int x = 0; x < 9; x++){
        for(int y = 0; y < 9; y++){
            cout << table[x][y] << " ";
        }

        cout << endl;
    }

    return 0;
}
```