Problem Chosen	2021 MCM/ICM Summary Sheet	Team Control Number
B		**0000**

How Many Language?

Summary

Globalization affects all aspects of the world, and at the same time, it also affects the trend of global language, which changes the language environment all over the world. Therefore, it is very important for a large multinational service company to investigate global language trends in order to find a suitable location for a new office.In order to predict the growth of a language, we consider many factors such as birth rate, migration rate, national GDP and so on. It is difficult to determine the quantitative relationship between these factors and the number of native speakers. We assume that the population growth of a country or region is the growth of the number of native speakers in that country or region, and use the grey model GM (1,1) to predict the change of the number of native speakers and total speakers in each region of the world. Finally, considering the global population and human migration patterns in the next 50 years, we finally get the change of geographical distribution of these languages in the same period. Based on the FM model, we can confirm the location of the office and determine which language the office should use. We have determined five indicators of GDP, cultural soft power, language, global competitiveness and infrastructure development trend, and determined their evaluation weights. In conclusion, we should have less than six offices.

Keywords: Language distribution; Grey models, GM(1, 1);Fuzzy Mathematical Model

How Many Language?

LaTeX Studio

January 23, 2021

Summary

Globalization affects all aspects of the world, and at the same time, it also affects the trend of global language, which changes the language environment all over the world. Therefore, it is very important for a large multinational service company to investigate global language trends in order to find a suitable location for a new office.In order to predict the growth of a language, we consider many factors such as birth rate, migration rate, national GDP and so on. It is difficult to determine the quantitative relationship between these factors and the number of native speakers. We assume that the population growth of a country or region is the growth of the number of native speakers in that country or region, and use the grey model GM (1,1) to predict the change of the number of native speakers and total speakers in each region of the world. Finally, considering the global population and human migration patterns in the next 50 years, we finally get the change of geographical distribution of these languages in the same period. Based on the FM model, we can confirm the location of the office and determine which language the office should use. We have determined five indicators of GDP, cultural soft power, language, global competitiveness and infrastructure development trend, and determined their evaluation weights. In conclusion, we should have less than six offices.

Keywords: Language distribution; Grey models, GM(1, 1);Fuzzy Mathematical Model

Contents

1 Overview **1**

1.1 Background 1

1.2 Problem Restatement 1

2 Notation **1**

3 Model Theory **1**

3.1 Grey system theory-based models in Population growth prediction 1

3.1.1 Grey models, GM(1, 1) 2

3.1.2 Population growth-Grey Model 3

3.2 Fuzzy Mathematical Model 4

3.2.1 introduction of Fuzzy Mathematical Model 4

4 Model Implementation and Results **5**

4.1 Problem A in Part I 5

4.2 Problem B in Part I 6

4.3 Problem C in Part I 7

4.4 Problem A in Part II 7

4.5 Problem B in Part II 8

5 Strengths and Weaknesses **8**

5.1 Strengths 8

5.2 Weakness 8

Appendices **9**

1 Overview

1.1 Background

There are currently about 6900 languages spoken on Earth, of which about 25 are widely spread and used. at the same time, many languages are endangered. However, there are much of the worlds population also speak a second language. The total number of speakers of a language may increase or decrease over time because of a variety of influences to include, but not limited to, the language(s) used and/or promoted by the government in a country, the language(s) used in schools, social pressures, migration and assimilation of cultural groups, and immigration and emigration with countries that speak other languages. Moreover, in our globalized, interconnected world there are additional factors that allow languages that are geographically distant to interact. These factors include international business relations, increased global tourism, the use of electronic communication and social media, and the use of technology to assist in quick and easy language translation.

1.2 Problem Restatement

We are required to investigate trends of global languages and location options for new offices. Then based on projected trends, and some or all of these influences and factors, model the distribution of various language speakers over time. So we need to discuss the impact factors of native language and second language respectively. In addition, we should base on our modeling to locate these offices and determine what languages would be spoken in the offices. In order to solve those problems, we will proceed as follows:

- Build a model to to investigate trends of global languages.
- Base on our modeling to locate these offices and determine what languages would be spoken in the offices.
- Locate these offices using the consequences of our model and determine what languages would be used in the offices and the number of new offices.

In our model, we first discuss the impact factors of native language and second language respectively. And model the distribution of various language speakers over time.Besides, we use our model to predict what will happen to the numbers of native speakers and total language speakers in the next 50 years. Then we consider the global population and human migration patterns predicted for the next 50 years, giving the change of the geographic distributions of these languages over this same period of time. we should base on our modeling to locate these offices and determine what languages would be spoken in the offices. And in an effort to save our client company resources, we suggest that the company should open less than six international offices.

2 Notation

3 Model Theory

3.1 Grey system theory-based models in Population growth prediction

As time goes by, the distribution of various language users is an abstract grey system. Population migration rate, birth rate, mortality rate, GDP and so on are all factors that will have

an impact. It is difficult to determine the quantitative relationship between these factors and the distribution of various language users in the future. Therefore, we use the grey model GM (1,1) for quantitative analysis to predict the distribution of various language users in the next 50 years. There are a large number of regions or countries using a certain language, we select the major countries or regions that use the language most, and think that the language will only increase in that country or region in the next 50 years, and other regions or countries will be ignored.

3.1.1 Grey models, GM(1, 1)

The Definition of GM(1,1)

Suppose $X^{(0)}$ is a Sequence with n elements

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n)) \tag{1}$$

The AGO generation sequence of $x^{(0)}$:

$$x^{(1)} = (x^{(1)}, x^{(1)}(2), ..., x^{(1)}(n)) \tag{2}$$

where,

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)(k = 1, 2, ...n). \tag{3}$$

Let $z^{(1)}$ be the nearest mean sequence of sequence $x^{(1)}$:

$$z^{(1)} = (z^{(1)}(2), z^{(1)}(3), ..., z^{(1)}(n)) \tag{4}$$

where,

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1), k = 2, 3, ..., n \tag{5}$$

So the grey differential equation model of GM (1,1) can be defined as

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{6}$$

a,b,are the coefficients; in Grey System theory terms, a is said to be a developing coefficient and b the grey input,$z^{(1)}(k)$is the whitening value ,$x^{(0)}$ is a grey derivative which maximizes the information density for a given series to be modelled. Then we have

$$\hat{a} = (a, b)^T \tag{7}$$

By means of least square method, we have

$$\hat{a} = (B^T B)^{-1} B^T Y_n \tag{8}$$

where,

$$B = \begin{bmatrix} -z^{(1)}(2), & 1 \\ -z^{(1)}(3), & 1 \\ \cdots, & \cdots \\ -z^{(1)}(n), & 1 \end{bmatrix} \qquad Y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \cdots \\ x^{(0)}(n) \end{bmatrix}$$

Then the whitening equation of grey differential equation is established as follows:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{9}$$

The result of whitening equation(time response function):

$$\hat{x}^{(1)}(t) = [x^{(1)}(0) - \frac{b}{a}]e^{-at} + \frac{b}{a} \tag{10}$$

So we have the time response series of corresponding GM(1,1) grey differential equation

$$\hat{x}^{(1)}(k+1) = [x^{(1)}(0) - \frac{b}{a}]e^{-at} + \frac{b}{a} \tag{11}$$

Let $x^{(1)}(0) = x^{(0)}(1)$

$$\hat{x}^{(1)}(k+1) = [x^{(0)}(1) - \frac{b}{a}]e^{-ak} + \frac{b}{a}, \quad k = 1, ..., n-1 \tag{12}$$

By inverse accumulated generating operation, we have prediction equation

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = [x^{(0)}(1) - \frac{b}{a}](1-e^{a})e^{-ak} + \frac{b}{a}, \quad k = 1, ..., n-1 \tag{13}$$

3.1.2 Population growth-Grey Model

The establishment of the Population growth-Grey Model includes the following 3 steps:
Step 1:
data class ratio validation:In order to ensure the feasibility of grey prediction, we need to do class ratio validation on the original sequence. Calculate the class validation of the original sequence $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(3))$, we can get

$$\lambda(k) = \frac{x^{0}(k-1)}{x^{(0)}(k)}, \quad k = 2, ..., n \tag{14}$$

If all class ratios satisfy capacitive coverage $\Theta = (e^{-2/(n+1)}, e^{2/(n+2)})$, then we can do the grey prediction; otherwise we need to do a translation transform on $x^{(0)}$, $Y^{(0)} = X^{(0)} + c$, let $Y^{(0)}$ meet the requirement of class ratio.
Step 2:
Establish GM(1,1) model and calculate the predicted value. Input data into the prediction equation(13) to get the predicted value.
Step 3:
Test predicted value
1.Absolute residuals test, We need to calculate that

$$\varepsilon = (k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}, \quad k = 2, ..., n \tag{15}$$

If $\varepsilon = (k) < 0.2$, it is considered to meet the general requirements; or $\varepsilon = (k) < 0.1$,it is considered to meet the higher requirements.
2.Class ratio deviation value test, based on the class ratio $\varepsilon = (k)$ and developing coefficient a calculated above, we can calculate the corresponding deviation value:

$$\rho(k) = 1 - (\frac{1-0.5a}{1+0.5a})\lambda(k) \tag{16}$$

If $\rho(k)$<0.2, it is considered to meet the general requirements; or $\rho(k)$<0.1, it is considered to meet the higher requirements.

Predict by model, we can get the data after 50 years.

3.2 Fuzzy Mathematical Model

3.2.1 introduction of Fuzzy Mathematical Model

1.establishment of Fuzzy Mathematical Model
Step1: confirm the membership function:

trends of language(TOL):

$$y = \begin{cases} 0 & (x < 0) \\ 1 & else \\ \frac{10x}{x_{max}} & (10x < x_max x) \end{cases}$$

GDP per capita:

$$y = \begin{cases} 0 & (x < 0.5x) \\ \frac{x-0.5x}{2.5x} & else \\ 1 & (x > 4x) \end{cases}$$

cultural soft power(CSP):

$$y = \frac{x}{x_{max}} \tag{17}$$

infrastructure:

$$y = \frac{x}{x_{max}} \tag{18}$$

global competitiveness:

$$y = \frac{x}{x_{max}} \tag{19}$$

Step2:
After getting the fuzzy relational matrix, we have judgment matrix D

$$D = \begin{bmatrix} 1 & , & 0.14285714 & , & 0.33333333 & , & 3 & , & 3 \\ 7 & , & 1 & , & 5 & , & 5 & , & 5 \\ 3 & , & 0.2 & , & 1 & , & 3 & , & 3 \\ 0.33333333 & , & 0.2 & , & 0.33333333 & , & 1 & , & 1 \\ 0.33333333 & , & 0.2 & , & 0.33333333 & , & 1 & , & 1 \end{bmatrix}$$

Then get the maximum eigenvalue λ_{max}

$$\begin{aligned} \lambda_{max} = (&5.39274992e+00+0.j, -5.20218640e-02+1.44404217j, \\ &-5.20218640e-02-1.44404217j, -2.88706196e-01+0.j, \\ &3.47156663e-32+0.j) \end{aligned} \tag{20}$$

After normalization, get the feature weight vector A

$$A = \left[0.1353678, 0.5314574, 0.19525851, 0.06895815, 0.06895815\right]$$

Then we need to check the consistency of judgment matrix:
(i) Calculate consistency index CI

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{21}$$

(ii) Average random consistency index RI

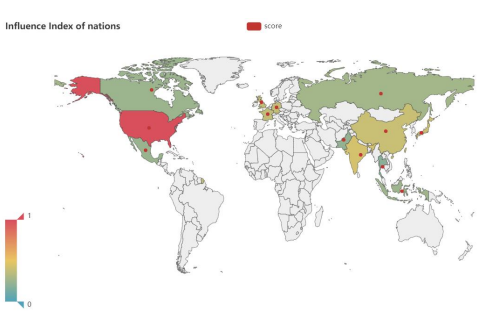
$$RI = \frac{\lambda'_{max} - n}{n - 1} \tag{22}$$

(iii) Calculate the consistency ratio CR:

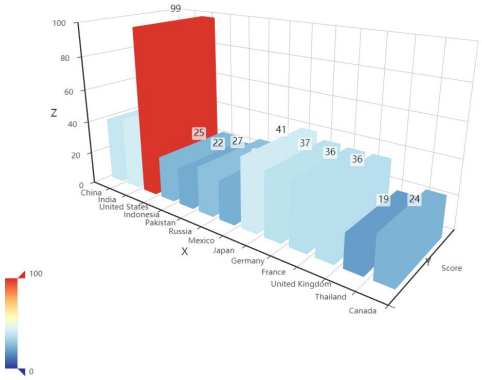
$$CR = \frac{CI}{RI} = 0.090982514640920120.091 \tag{23}$$

(When CR<0.10, it holds that the consistency of the judgment matrix is acceptable.)
Finally, we get the synthetic evaluation of the nations:

$$B = AR = [0.39720680, 0.42471536, 0.87934924, 0.13741900, 0.10286798, 0.14205876, 0.12013810, 0.29880819, 0.30182129, 0.30543563, 0.30810833, 0.12927160, 0.21145447]$$



(a) Influence Index of nations



(b) the synthetic evaluation of the nations

Figure 1: the synthetic evaluation of the nations

4 Model Implementation and Results

4.1 Problem A in Part I

We use the data to calculate the number of native speakers in the next 50 years, and use Python to plot the growth curve and the corresponding geographical distribution, as shown in the figure 2 and Figure 3.

It can be seen from the figure that the distribution of various languages around the world at different times

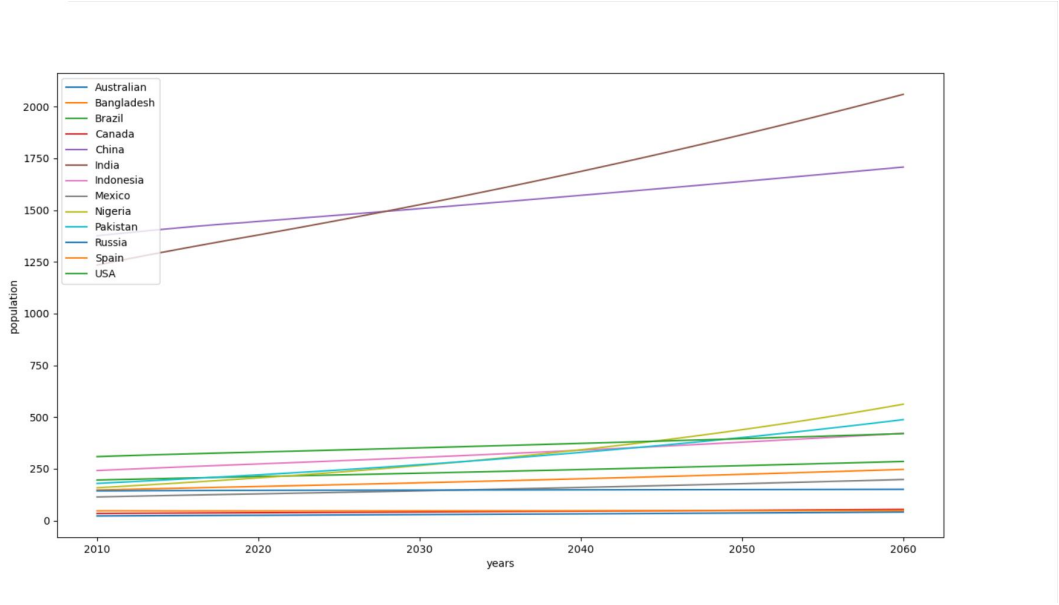
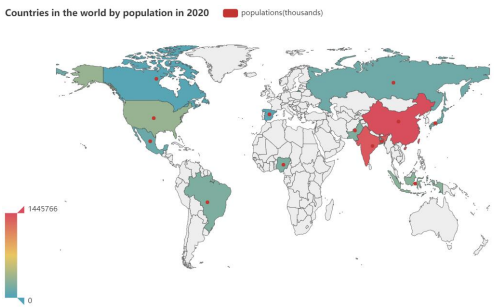
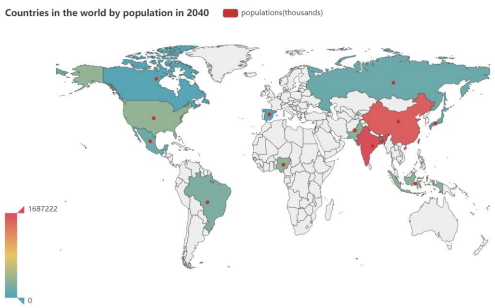


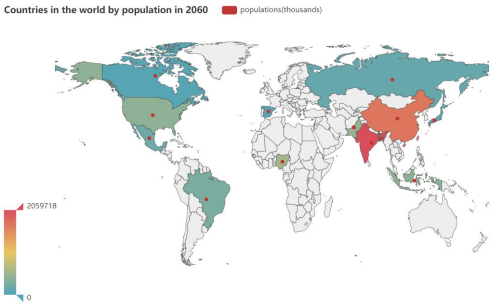
Figure 2: Changes of Native Language Speakers in 50 years



(a) Countries in the world by population in 2020



(b) Countries in the world by population in 2040



(c) Countries in the world by population in 2060

Figure 3: Countries in the world by population at different times

4.2 Problem B in Part I

We can see that the number of native and total languages speakers will increase in various degrees according to Figure2 and Figure4.

Based on the result of the Population growth-Grey Model, we get the top10 languages by total number of speakers from 2020 to 2070.From the wordcloud 5, we can see the proportion of various language users in 2070 According to the data(Appendices), over the next 50 years, Marathi will replace Indonesian, the fastest declining language, as the fastest rising language.

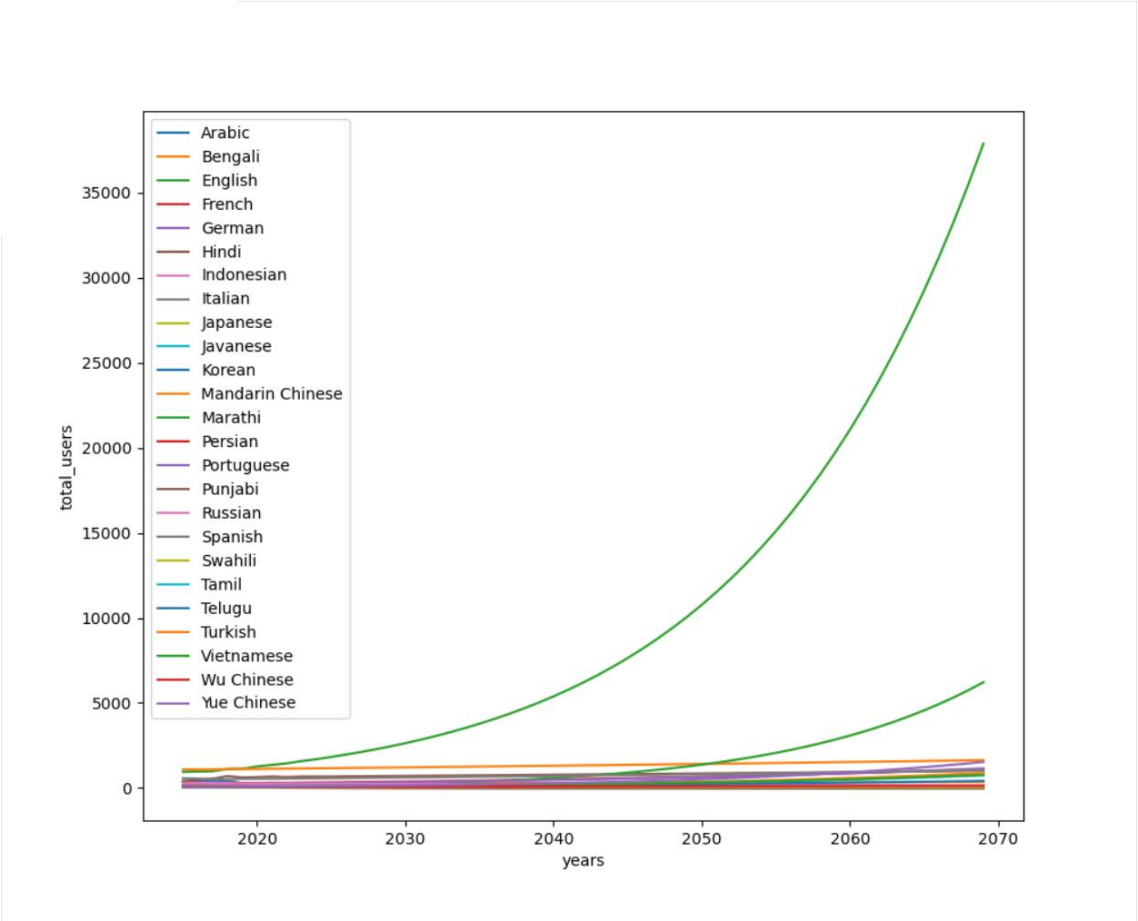


Figure 4: Changes of Total Language Speakers in 50 years

4.3 Problem C in Part I

Based on the prediction of population growth and human migration patterns in the next 50 years, we draw a chart. Considering that the geographical distribution of mother tongue will not change, and the number of users of a certain language in other place is relatively less than the mother tongue area, therefore we only consider the distribution of mother tongue in this area. As shown in the figure 4.

4.4 Problem A in Part II

According to the fuzzy mathematical model, we can get the the synthetic evaluation of the nations and the final rank.
1 United States
2 India
3 China
4 United Kingdom
5 France
6 Germany
The rank of countries under our evaluation system shown in Figure 6

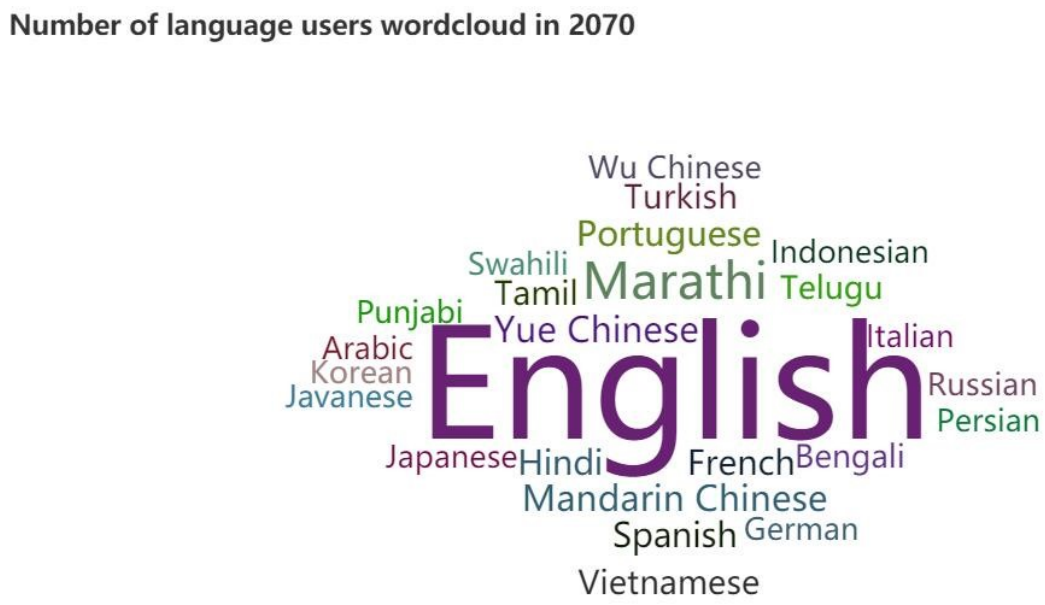


Figure 5: Number of language users wordcloud in 2070

4.5 Problem B in Part II

Based on the result of the Grey system theory-based models in Population growth prediction and the Fuzzy Mathematical Model, we conclude that the population and number of total language speakers should increase in the next 50 years. So the future development of countries would be unstable. And considering the rapid development of communications techniques and internet, we suggest open less than 6 offices.

5 Strengths and Weaknesses

5.1 Strengths

- Our data are from authoritative and reliable official websites, such as wikipedia, Ethnologue, World Bank.
- Our model is very flexible and stable. The parameters can be changed according to different conditions to adapt to the needs; at the same time, a slight change of parameter will not cause a significant change of the results.
- We have worked out corresponding solutions according to different situations.

5.2 Weakness

- In our model, in order to simplify our model, we only consider the development of a certain language in the area with the largest proportion of users. In fact, a language may be widely spread and used all over the world.

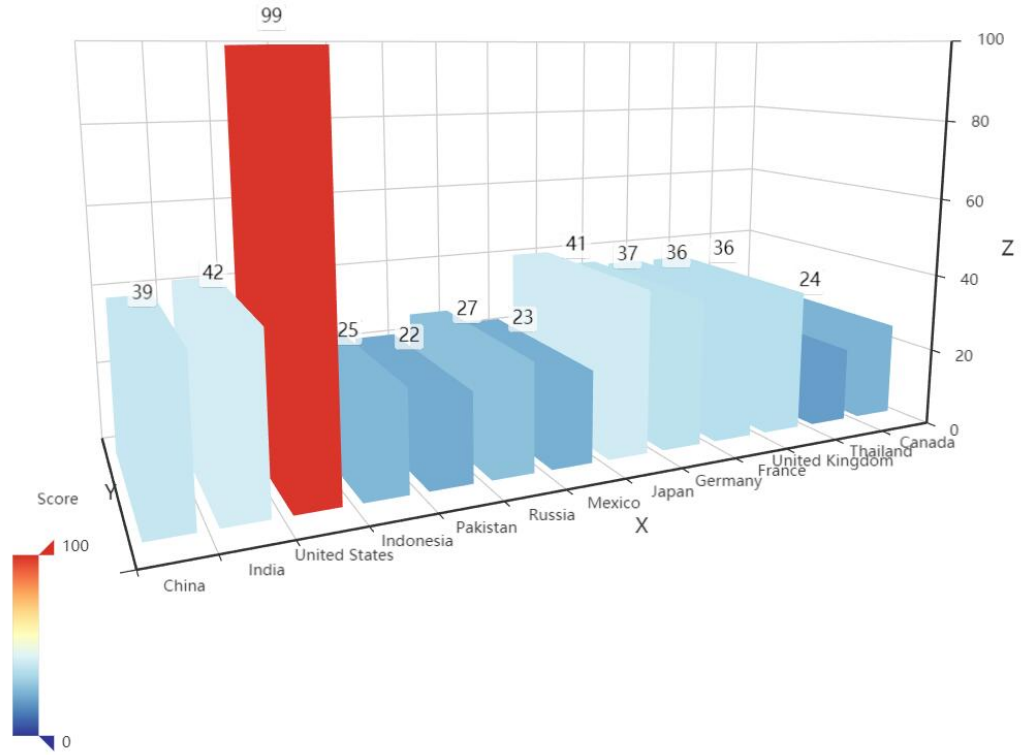


Figure 6: Changes of Native Language Speakers in 50 years

- The data of some parameters are only forecast data, because we don't practice and can't get the real data.

Appendices

years	Rank1	Rank2	Rank3	Rank4	Rank5	Rank6	Rank7	Rank8	Rank9	Rank10
2020	Mandarin Chinese	English	Spanish	Arabic	Hindi	Indonesian	Portuguese	Russian	Bengali	French
2021	Mandarin Chinese	English	Spanish	Hindi	Arabic	Indonesian	Russian	Portuguese	Bengali	French
2022	Mandarin Chinese	English	Hindi	Spanish	Arabic	Indonesian	Russian	Bengali	French	Portuguese
2023	English	Mandarin Chinese	Hindi	Spanish	Arabic	French	Indonesian	Russian	Bengali	Portuguese
2024	English	Mandarin Chinese	Hindi	Spanish	French	Arabic	Bengali	Russian	Portuguese	Indonesian
2025	English	Mandarin Chinese	Hindi	Spanish	French	Arabic	Bengali	Russian	Portuguese	Indonesian
2026	English	Mandarin Chinese	Hindi	Spanish	French	Bengali	Portuguese	Russian	Arabic	Indonesian
2027	English	Mandarin Chinese	Hindi	Spanish	French	Bengali	Portuguese	Russian	Arabic	Indonesian
2028	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Bengali	Russian	Arabic	Indonesian
2029	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Bengali	Russian	Marathi	Arabic
2030	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Bengali	Russian	Marathi	Japanese
2031	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Bengali	Russian	Marathi	Japanese
2032	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Bengali	Russian	Marathi	Yue Chinese
2033	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Bengali	Russian	Marathi	Yue Chinese
2034	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Bengali	Russian	Marathi	Yue Chinese
2035	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Bengali	Marathi	Russian	Yue Chinese
2036	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Bengali	Marathi	Russian	Yue Chinese
2037	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Marathi	Bengali	Russian	Yue Chinese
2038	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Marathi	Bengali	Russian	Yue Chinese
2039	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Marathi	Bengali	Russian	Yue Chinese
2040	English	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Marathi	Bengali	Russian	Yue Chinese
2041	English	Mandarin Chinese	Hindi	Spanish	French	Marathi	Portuguese	Bengali	Yue Chinese	Russian
2042	English	Mandarin Chinese	Hindi	Spanish	French	Marathi	Portuguese	Bengali	Yue Chinese	Russian
2043	English	Mandarin Chinese	Hindi	Spanish	Marathi	French	Portuguese	Bengali	Yue Chinese	Russian
2044	English	Mandarin Chinese	Hindi	Spanish	Marathi	French	Portuguese	Bengali	Yue Chinese	Turkish
2045	English	Mandarin Chinese	Hindi	Spanish	Marathi	French	Portuguese	Bengali	Yue Chinese	Turkish
2046	English	Mandarin Chinese	Hindi	Spanish	Marathi	French	Portuguese	Yue Chinese	Bengali	Turkish
2047	English	Mandarin Chinese	Hindi	Spanish	Marathi	French	Portuguese	Yue Chinese	Bengali	Turkish
2048	English	Mandarin Chinese	Hindi	Marathi	Spanish	French	Portuguese	Yue Chinese	Bengali	Turkish
2049	English	Mandarin Chinese	Marathi	Hindi	Spanish	French	Portuguese	Yue Chinese	Bengali	Turkish
2050	English	Mandarin Chinese	Marathi	Hindi	Spanish	French	Portuguese	Yue Chinese	Bengali	Turkish
2051	English	Mandarin Chinese	Marathi	Hindi	Spanish	French	Portuguese	Yue Chinese	Bengali	Turkish
2052	English	Mandarin Chinese	Marathi	Hindi	Spanish	French	Portuguese	Yue Chinese	Turkish	Bengali
2053	English	Mandarin Chinese	Marathi	Hindi	Spanish	French	Portuguese	Yue Chinese	Turkish	Bengali
2054	English	Mandarin Chinese	Marathi	Hindi	Spanish	French	Portuguese	Yue Chinese	Turkish	Vietnamese
2055	English	Mandarin Chinese	Marathi	Hindi	Spanish	French	Portuguese	Yue Chinese	Turkish	Vietnamese
2056	English	Marathi	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Yue Chinese	Turkish	Vietnamese
2057	English	Marathi	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Yue Chinese	Turkish	Vietnamese
2058	English	Marathi	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Yue Chinese	Turkish	Vietnamese
2059	English	Marathi	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Yue Chinese	Turkish	Vietnamese
2060	English	Marathi	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Yue Chinese	Turkish	Vietnamese
2061	English	Marathi	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Yue Chinese	Turkish	Vietnamese
2062	English	Marathi	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Yue Chinese	Turkish	Vietnamese
2063	English	Marathi	Mandarin Chinese	Hindi	Spanish	French	Portuguese	Yue Chinese	Turkish	Vietnamese
2064	English	Marathi	Mandarin Chinese	Hindi	Spanish	Yue Chinese	Portuguese	French	Turkish	Vietnamese
2065	English	Marathi	Mandarin Chinese	Hindi	Spanish	Yue Chinese	Portuguese	French	Turkish	Vietnamese
2066	English	Marathi	Mandarin Chinese	Yue Chinese	Hindi	Spanish	Portuguese	French	Turkish	Vietnamese
2067	English	Marathi	Mandarin Chinese	Yue Chinese	Hindi	Spanish	Portuguese	French	Turkish	Vietnamese
2068	English	Marathi	Mandarin Chinese	Yue Chinese	Hindi	Spanish	Portuguese	French	Turkish	Vietnamese
2069	English	Marathi	Mandarin Chinese	Yue Chinese	Portuguese	Spanish	Hindi	French	Turkish	Vietnamese
2070	English	Marathi	Mandarin Chinese	Yue Chinese	Portuguese	French	Spanish	Hindi	Turkish	Vietnamese

Figure 7: the top10 languages by total number of speakers from 2020 to 2070.