|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | |  |  |  |  |   The Prediction of Language Population  **Shanghai Experimental School**  **Class 3 Wang Chengkai**  Summary Sheet  Language is the basic of life. Language is the carrier of culture information and the approach for people preserving, delivering and comprehending historic experience and the method to understand achievements of science and art. More importantly, people use language to learn the world and to communicate.  Since more than a half people in the world speak top16 languages, the  development of these languages is more and more worth considering. |

In this essay, we construct two models in order to give solutions to the tasks listed following.

In Model A, a grey prediction is used to predict the fertility rate and the mortality rate of each individual country. From that we can calculate each country’s population growth and the population of native language speakers. In this model, we mainly consider the influence of natural growth of the population, including birth of newborns and mortality, which affect the number of speakers of each language.

In Model B, we divided the factors into 2 main parts: Active learning and Passive learning. Passive learning means language learning caused by immigrations. We ameliorated the G. K. Zipf’s Population gravity model by introducing the social resources, economic resources, environment resources, and technology resources to the equation. Active learning is divided into three parts: economy, technology and education. By using Partial derivation, we figured out the value of each weight.

And finally, through a series of census, we work out these two problems.

Contents

[The Prediction of Language Population 1](#_Toc12307147)

[1. Introduction 3](#_Toc12307148)

[1.1 background 3](#_Toc12307149)

[1.2 Our Work 4](#_Toc12307150)

[2. General assumptions 4](#_Toc12307151)

[3. Model A: Grey prediction on population 4](#_Toc12307152)

[3.1 Model overview 4](#_Toc12307153)

[3.2Model assumption 4](#_Toc12307154)

[3.3 Constants and variables 5](#_Toc12307155)

[3.4 Grey prediction 5](#_Toc12307156)

[4. Model B: Active and Passive learning 7](#_Toc12307157)

[4.1 Model overview 7](#_Toc12307158)

[4.2 Variables of Model B 8](#_Toc12307159)

[4.3 Passive learning 8](#_Toc12307160)

[4.4 Active learning 9](#_Toc12307161)

[5. Conclusions 11](#_Toc12307162)

[1.Languages in the current top-ten lists won’t be replaced in the next 50 years. 11](#_Toc12307163)

[2.About the human migration patterns: Some Africans will move to Europe and Asia.. 11](#_Toc12307164)

[6. Sensitive analysis 11](#_Toc12307165)

[6.1 Sensitivity of Q1 11](#_Toc12307166)

[7. Strengths and weaknesses 12](#_Toc12307167)

[7.1 Strengths 12](#_Toc12307168)

[7.2 Weaknesses 12](#_Toc12307169)

[8. References 13](#_Toc12307170)

1. Introduction

1.1 background

About 6,900 languages are currently spoken on Earth. The most important language are the following 16 kinds: Mandarin (incl. Standard Chinese), Spanish, English, Hindi, Arabic, Malay, Bengali, Portuguese, Russian, Punjabi, Hausa,

Persian, Swahili, French, German and Japanese. Meanwhile, a second language is widely spoken. When considering the total number of speakers of a particular language, the languages and their order change from the most spoken native 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 used by the government in a country, the language 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 Our Work

(1) Build a model to predict the numbers of native speakers and total language speakers in the next 50 years.

(2) Predict the total language speakers in the next 50 years.

1. General assumptions

**1.We do not consider the influence made by great changes such as natural disasters, major political changes, wars and etc.**

*For the GDP of the last decades of years has grown stably, we can predict that it will continue this stable trend of growth. Because of the unpredictability of the great changes such as natural disasters and the wars, we don’t take them into consideration.*

2.We don’t consider any change in current languages.

Language may change itself when time passes by, but this change is unpredictable. So, we don’t consider the change of any language. All the future language will remain at the current stage.

**3.We use a year(annual) as the minimum unit.**

*The minimum unit given in the subject is a year.*

1. Model A: Grey prediction on population

3.1 Model overview

In this model, we use grey prediction to predict the fertility rate and the mortality rate of each individual country. From that we can calculate the population of each individual language speaker. In this model, we mainly considered the influence of the natural growth of the population, including birth of newborns and mortality.

3.2Model assumption

1. We think the natural growth of the population is only related to fertility rate and the mortality rate.
2. Those two factors play the most important role in the growth, while other factors don’t.

3.3 Constants and variables

In the following table are the constants and variables used in the Grey prediction.

|  |  |
| --- | --- |
| Constants | Definition |
| A | Development coefficient |
| B | grey actuating quantity |
| Variables | Definition |
|  | The fertility rate at (t) time point |
|  | The mortality rate (t) time point |
|  |  |

**Table 1** Constants and Variables for model

3.4 Grey prediction

Based on the known current fertility rate and mortality rate, we use grey prediction to get following equation:

This is a set of incomplete information of gray, with great randomness, which is generated and processed to provide more useful information. The following additive generation is used, and the result of the cumulative generation of M times is

So, we get the prediction model GM (1,1)

Among them

Among them, the labefaction differential equation of （1）is

a is for Development coefficient, b is for grey actuating quantity.

According to the least square method,

We can create a matrix

Discrete solution of the differential equation is

We also used grey prediction to predict the mortality rate

in the same way.

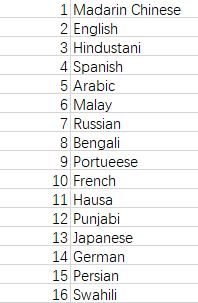


Figure 1 the number of native speakers of each language in year 2067(unit: people) [1] [2]

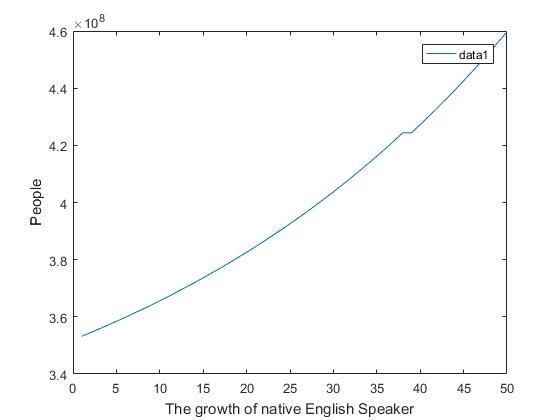


Figure 2 the number of native English speakers in the next 50years

1. Model B: Active and Passive learning

4.1 Model overview

In Model B, we divided the factors into 2 main parts: Active learning and Passive learning. Passive learning means language learning caused by immigration. We ameliorated the G. K. Zipf’s Population gravity model by introducing the social resources, economic resources, environment resources, and technology resources to the equation. Active learning is divided into three parts: economy, technology and education. By using Partial derivation, we figured out the value of each weight.

4.2 Variables of Model B

|  |  |
| --- | --- |
| Constants | Definition |
|  | The weight of technical resource in the gravitational model |
|  | The weight of environmental resource in the gravitational model |
|  | The weight of social resource in the gravitational model |
|  | The weight of economic resource in the gravitational model |
| Variables | Definition |
|  | The population density of the emigrant areas (i point) |
|  | The population density of the immigrant areas (p point) |
|  | The distance between point i and point j |
|  | coefficient of population migration between i and j |
|  | The papers released in year t in place i |
|  | Per capita green area in year t in place i |
|  | unemployment rate in year t in place i |
|  | GDP per capita in year t in place i |
| E | The economical index |
| T | The technical index |
| N | The difficulty index of a language |
|  |  |

**Table 2** Constants and Variables

4.3 Passive learning

According to the G. K. Zipf’s Population gravity model, the coefficient of population migration between i and j. () is directly proportional to the population of i and j, and is inversely proportional to the distance between i and j.

However, this model needs similar population between i and j, and it has a lot of limitations. Therefore, we introduced economic benefits, natural environment which can attract population, so we change the equation to

Since in large countries, the sum of population can’t present the real population density, we use , to describe the real population density.

is the index for i place. We introduced the social resources ,economic resources, environment resources, and technology resources.

We use linear regression to figure out that:

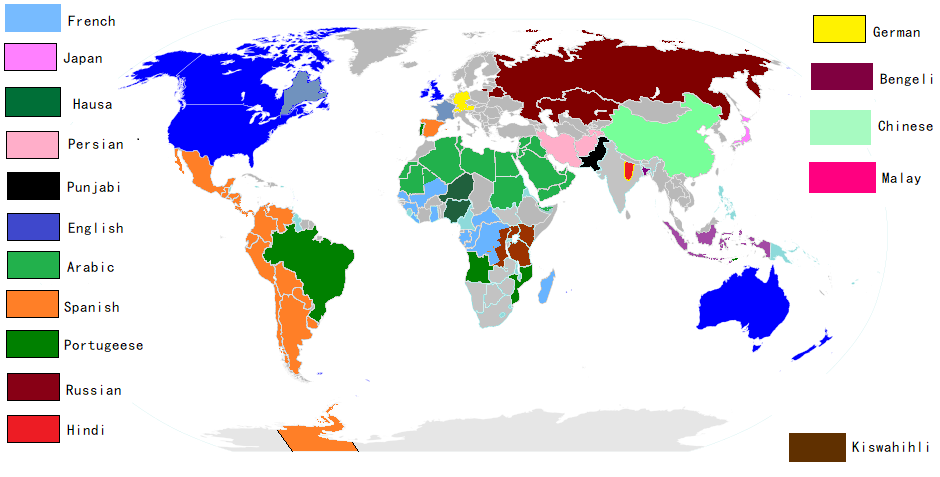
4.4 Active learning

The learning of a language is about the difficulty of learning the language, the status of its technology, and the economy of that area.

By using Partial derivation, we get that

After solving the differential equations, we get that:

After substituting data, we solved that



**Figure 3**the language which have top16 speakers in the world in 2017[3]

**The top10 language in 2067 will be:**

|  |  |
| --- | --- |
| Orders | Languages |
| No.1 | English |
| No.2 | Mandarin Chinese |
| No.3 | Arabic |
| No.4 | Spanish |
| No.5 | Hindustani |
| No.6 | Portuguese |
| No.7 | Bengali |
| No.8  No.9  No.10 | Malay  Punjabi  French |

**Table 3** Results

1. Conclusions

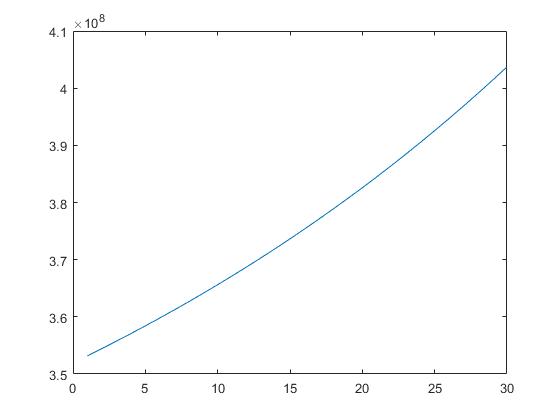
1.Languages in the current top-ten lists won’t be replaced in the next 50 years.

2.About the human migration patterns: Some Africans will move to Europe and Asia. Some Asians will immigrate to Europe.

1. Sensitive analysis

6.1 Sensitivity of

The curve fitted is just like a primary function, which is close to our prediction.



**Figure 4** the sensitive analysis

1. Strengths and weaknesses

7.1 Strengths

**1. Our model works steadily.**

Sensitivity analysis shows that our model is not easily disturbed by changes in its constants. Therefore, its results are relatively steady and reliable.

**2. Our model works ideally.**

The prediction given by our model ideally corresponds with mainstream points. In addition, there SA validate the stability of our model.

7.2 Weaknesses

**1.Our model involves a large operand.**

As our model involves huge amounts of data, the amount of calculation we need to conduct becomes enormous. This makes our model slow to operate and dependent on the quality of hardware.

**2. Our model relies on large amounts of data.**

To operate our model, great amounts of data are needed. When appropriate data is scarce, our model would be unable to function.

1. References

[1] https://www.ethnologue.com

[2] [https://data.worldbank.org](https://data.worldbank.org/)

[3] Crystal, David . Linguistics

[4] <http://mjl.clarivate.com/cgi-bin/jrnlst/jloptions.cgi?PC=K>

[5] [https://comtrade.un.org/data](https://comtrade.un.org/data/)

[6] [https://population.un.org/wpp](https://population.un.org/wpp/)

[7] Trudgill, P. Dialects. Florence, KY