

Prediction of Language Development

Shuyao Liu

School of North China Electric Power University, BaoDing, China

Abstract. Since entering the 21st century, the factors influencing linguistic distribution such as electronic communication devices and social media have seen unprecedented growth. Therefore, we have selected the data on the number of users in each of the languages used in the past 10 years and have carried out these data with the year Fitting and selecting the 6th-order function with higher accuracy, we get the relationship between the number of language users and the year of the year. In the relation, each power of year x has its corresponding coefficient, which can be regarded as different factors (for example, social media. The use of the national population, education and social pressures, etc.) as a function of year relative change in language usage. I make prediction of the development of language in next few years based on the formula and data.

Keywords: Global; Language; influence; Changes.

1. Introduction

When you are asked about the question 'how many languages on earth?', what would you answer? 100? 1,000? Or just 'many'? Actually, there are approximately 6,900 languages on earth hitherto. However, only a few languages get the position at the top of the world hierarchy when there is devastation at the base. Such as Mandarin, Spanish, English, Hindi, Arabic, Bengali, Portuguese, Russian, Punjabi, Japanese, etc. Total number of speakers of a specific language is the most direct index to explore the changes of languages over time. Factors affecting the total number of people using the language from various aspects. In this paper, I propose six influencing factors, representing population, tourism, immigration, social media, social impact, and electronic communications respectively.

2. Overall Thinking

1. First, I need to predict trends in global languages and model the distribution of users in all languages. The various languages here should be included in the top ten list.

2. I will predict the quantitative change in the population of native speakers and general speakers over the next 50 years. In addition, I will make a predictable and reasonable explanation of the sequential changes in the top ten lists of languages in the next 50 years.

3. I will appropriately describe the changes in the geographical distribution of the ten languages mentioned in the background section during the same period. The change is based on the global population and migration patterns projected for the next 50 years.

2.1 Assumption

For simplicity, I will make the following assumptions as a start of construction of our models.

- Consideration The total number of people using a particular language is only affected by the population, tourism, immigration, social media, social aspects, electronic communications these six aspects. The reason I make this assumption is the limited ability to get the data.

- Ten languages I study only the ten languages listed in the title. The reason I make this assumption is that these ten languages have some representation because their native speakers make up half of the world's total population.

- Impact is constant I assume that the degree to which a particular language is affected by six aspects over the next five decades does not change significantly with the years. The reason I make this assumption is that the significant change in the degree of influence of the six factors is quite a small probability

2.2 Notations

Notations	Definitions
y	Total numbers of speakers of a particular language(unit: million)
x	Years
β_0	Other aspects of the impact
β_1	The influencing parameter of population
β_2	The influencing parameter of tourism
β_3	The influencing parameter of immigrant
β_4	The influencing parameter of social media
β_5	The influencing parameter of society
β_6	The influencing parameter of electronic communications
yM	The Derivative of Chinese Fitting Curve
yS	The Derivative of Spanish fitting curves
yE	The Derivative of English fitting curves
t0	Current time
P0	The population of the current time
r	Scale factor, a constant
M	The maximum capacity of the environment
t	Target time
Pt	The population of the target time

2.3 Models

In this section, I first introduce and analyze the LPM(Languages Prediction Model), and then the GBM(Geographical Distribution Model) .

2.3.1 LPM(Languages Prediction Model)

In order to investigate trends of global languages over time, I devise the Languages Prediction Model. Total numbers of speakers of a certain language serves as the sole index, which reflect the trend of the certain language intuitively and concisely. I utilize the method of spline fitting and consider six factors' influences in total numbers of speakers of a particular language. The equation I used in part one is as following:

$$y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4x^4 + \beta_5x^5 + \beta_6x^6$$

Population(β_1)

According to reliable statistics, there are about 7,057 million people on earth until 4 March 2013. In 2011, the growth rate of global population is 1.1%. It is estimated that by 2040, the world's population will reach 8 billion. Because of the large population in the world, the change of the variable y over time is small. Hence, I consider the impact of population on y as a first-order coefficient.

Tourism(β_2)

The world's tourism industry has entered a golden era of economic globalization and economic integration. Tourism has become the world's largest industry. Therefore, the impact of the development of tourism on the total number of language users is quite important. I use the polynomial quadratic coefficient β_2 to indicate the impact of tourism on the variable y

Immigrant(β_3)

The immigration of immigrants also has a considerable impact on the number of language users. During the independence of the United States, European immigrants who did not speak English as their native language accounted for a quarter of the total U.S. population. In Pennsylvania, two-fifths speak German.[1] I take β_3 as a parameter to show the effect of the number of immigrants on y.

Social Media(β_4) -

The widespread use of social media can also have a certain impact on the total number of people using language. I use this four-term coefficient β_4 to measure this effect.

Social Influencing(β_5)

Here is the impact of both social policies and social pressure. In China, adolescents are required to learn and use English in both middle and high schools, and in college, English becomes even more important. In addition, in some particularly developed cities in China, it is a general trend that most people use to communicate with each other. The β_5 belongs to the Social Influencing.

Electronic Communications(β_6)

I found that many useful software or some large-scale games, often in a language for the original version of the release. This has more or less the effect on those who are not native speakers of the language, although I have less impact but I can not ignore it. I use the highest coefficient β_6 to indicate the impact of electronic communications on language usage.

2.3.2 GBM(Geographical Distribution Model)

Based on the existing Logistic model, I construct a GDM(Geographical Distribution Model) model.

The title asks us to explore whether the geographical distribution of these languages changes over the same period, based on the projected pattern of population and population movements over the next 50 years. I subjectively judge the geographical distribution of a particular language as immigration shifts. However, due to lack of capacity, I have not been able to predict the pattern of population migration in the next 50 years, so the correctness of our results needs to be considered.

The Logistic model studies the change of population growth over time. I construct our GDM model based on the finite growth model. If the population is the ordinate and the time is the abscissa, I find that the curve of the population over time shows a 'S' shape. The formula is as follows:

$$P(t) = \frac{P_0 M e^{rM(t-t_0)}}{M - P_0 + P_0 e^{rM(t-t_0)}}$$

2.3.3 SSM(Site Selection Model)

First, I set up a site selection model, where I choose and determine the country. Then, I consider the influence of a country in terms of international status, international trade, international exchange, international influence and the principle that the number of users in a given language in a country is the highest. Our location will maximize the dispersion of the locations of six countries based on the world map.

2.3.4 SOM(Site Optimization Model)

The purpose of the model I built was to propose an optimization methodology for the six country offices set out in the title. My principle is that, taking into account the changing nature of global communications, it is even more important to save the company's resources. In fact, this is a question of how to determine the optimal distribution of resources to achieve the maximum benefit under the condition of limited resources. In this model, I consider the impact of five factors on the effectiveness of the company's international offices.

Factor1	Risk
Factor2	Prospect
Factor3	Policy
Factor4	Customer practical degree
Factor5	Social stability

2.4 The Model Results

From the website of Ethology, I find out the number of users in each of the ten languages from 2007 to 2017 and use polynomial spline fitting to get the curve of total number of users in a particular language over the years. According to the curves drawn by MATLAB software's curve spline toolbox,

the influence factors corresponding to each influencing factor are obtained. As shown in the following table:

Languages	Mandarin	Spanish	English	Hindi	Arabic
β_0	2.6E+17	2.2E+17	6.7E+15	1.5E+17	1.3E+16
β_1	7.8E+14	-6.7E+14	-2.3E+13	-4.4E+14	-3.9E+13
β_2	9.8E+11	8.5E+11	2.5E+10	5.5E+11	4.9E+10
β_3	-6.5E+08	-5.7E+08	-1.7E+07	-3.7E+08	-3.3E+07
β_4	2.4E+05	2.1E+05	6.3E+03	1.4E+05	1.2E+04
β_5	-4.9E+01	-4.3E+01	-1.3E+00	-2.8E+01	-2.4E+00
β_6	4.1E-03	3.6E-03	1.0E-04	2.3E-03	2.0E-04
Languages	Bengali	Portuguese	Russian	Punjabi	Japanese
β_0	1.2E+15	-4.7E+16	-2.2E+17	-2.3E+16	6.3E+15
β_1	3.6E+12	1.4E+14	6.6E+13	6.9E+13	-1.9E+13
β_2	4.4E+09	-1.8E+11	-8.4E+10	-8.6E+10	2.3E+10
β_3	-2.9E+06	1.3E+08	5.6E+07	5.8E+07	-1.5E+07
β_4	1.1E+03	-4.8E+04	-2.1E+04	-2.2E+04	5.7E+03
β_5	-2.1E-01	9.9E+00	4.3E+00	4.4E+00	-1.1E+00
β_6	1.7E-05	-8.4E-04	-3.6E-04	-3.7E-04	9.4E-05

Here is the fitting curve of Mandarin, Spanish and English (Figure 1 to 3). I derive the derivative of the fitted curve for the three languages above. As follows:

$$y_M = 0.0246x^5 - 245x^4 + (9.6e + 0.5)x^3 - (1.9e + 0.9)x^2 + (19e + 12)x - (7.8e + 14)$$

$$y_S = 0.0216x^5 - 215x^4 + (8.4e + 05)x^3 - (1.6e + 06)x^2 + (1.6e + 12)x - (6.7e + 14)$$

$$y_E = (0.6e - 3)x^5 - 6.5x^4 + (2.5e + 4)x^3 - (5.1e + 7)x^2 + (5.e + 10)x - (2.3e + 13)$$

In the next 50 years, the population will increase slowly. By 2040, the world population will reach 8 billion. Therefore, I judge that the number of native speakers and the total number of users in each language will increase by a certain margin on the current base.

Tourism development will be in full swing. Based on the current status quo of domestic tourism booming, tourism and cultural exchanges between countries and continents will become more frequent. Therefore, I conclude that some of the languages' native speakers and the total number of users will have a certain degree of increase in the next 50 years. Social media will remain at a certain level, and people will have access to national cultures and languages through social media, so the dependent variable y will also grow. Electronic communication is very likely to rapid development in China, the fifth generation of communication technology is being developed, I believe in the near future will

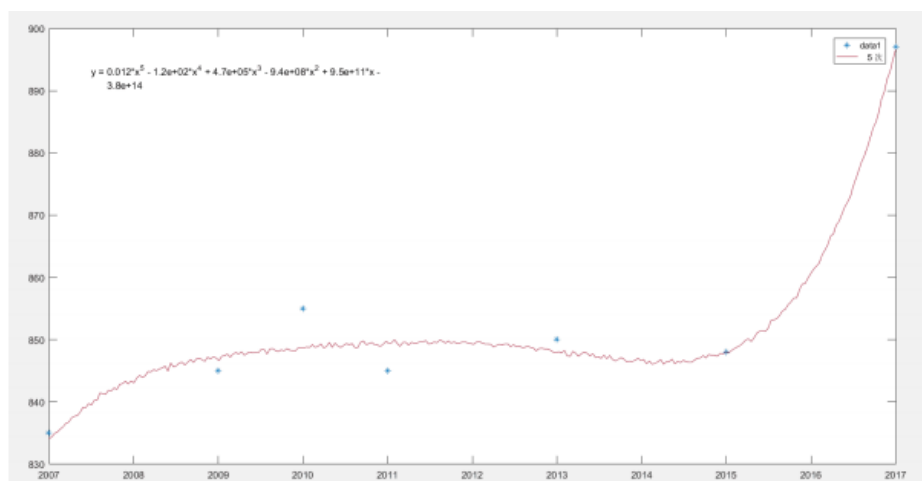


Figure 1. The fitting curve of Mandarin

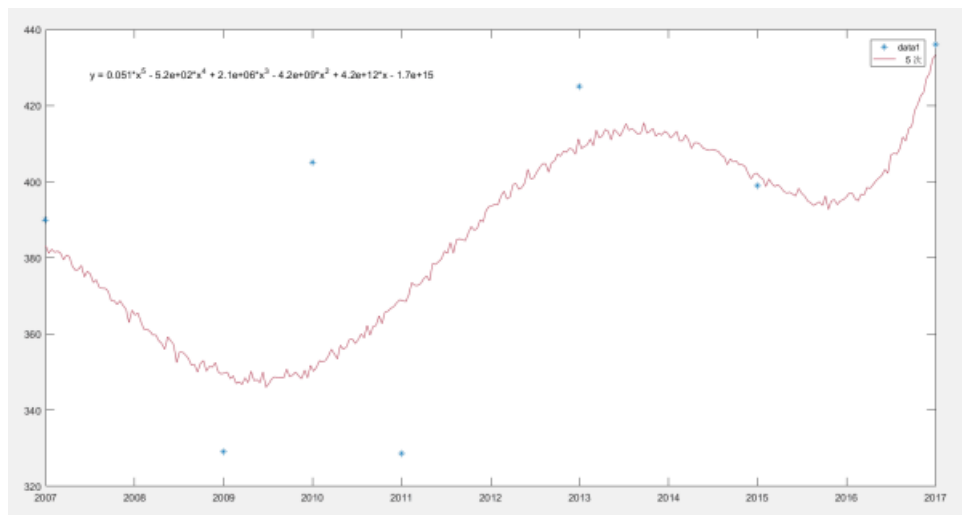


Figure 2. The fitting curve of Spanish

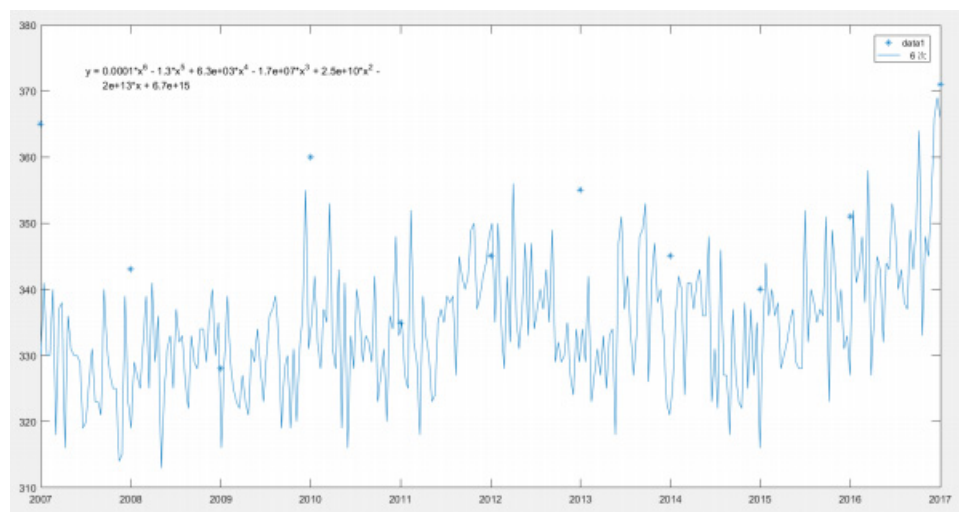


Figure 3. The fitting curve of English

be able to put into use.

In terms of the impact of migration, I find that the population of developing countries tends to move toward developed countries. As shown below:

As a result, the total number of Chinese, Spanish, Arabic, Portuguese, and Arabic speakers in the top 10 languages will continue to grow steadily in the future. However, the total number of users in languages such as English and Russian will first decrease and then increase; for those The total number of users and the number of native speakers of languages in the countries and regions where the number of migrants and the emigrants remained unchanged were quite small, like Hindi, Bengali, Punjabi and Japanese. First-generation immigrants will have less use of their mother tongue in their next generation, and in the third to fourth generations their mother tongue usage is almost zero.

3. Summary

Based on lots of data, I analyse the factors influencing the development of language, include social.

media, population etc. I set up the multivariate regression model for imitating the distribution of language users and forecast its changes in next 50 years. I also analyse the influence of global population and immigrant. Based on data, I set up logistic population growth model in every continent

and use regression analysis to make the model of immigrant. I get a conclusion that Chinese, Spanish, Arabic, Hindi , Native speakers in Portuguese and total language users steadily increased; English and Russian fell in the short term and then slowly increased; native speakers of other languages such as Bengali, Punjabi, Japanese and total language users maintained in a relatively stable level.

References

- [1]. Immigration and Language Diversity in the United States; Ruben G. Rumbaut and Douglas S. Massey.
- [2]. The Future of Language; David Graddol, et al.; Science 303, 1329 (2004).