BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Review of Educational Theory
http://ojs.bilpublishing.com/index.php/ret

REVIEW

Research Based on the Development Trend of World Language

Man Xiao*   Long Mao   Hongmei Meng   Jinping Liu   Huan Li   Hongzheng Yan
Chongqing University of Technology International College, Chongqing, 401135, China

ARTICLE INFO

Article history:
Received: 30th September 2018
Revised: 8th October 2018
Accepted: 15th October 2018
Published Online: 31st October 2018

Keywords:
Language communication model
Grey prediction
Efficiency model
The Bayes' theorem
Fussy comprehensive assessment

ABSTRACT

Multicoloured languages play an irreplaceable role in the whole world as a useful communication tool. With the development of technology and science, varieties of languages have an ideal prospective tendency to evolution during the long and wonderful history. Will they be thriving or decaying? To begin with, aimed to gain general tendency about the quantity of languages' speakers, we employ the Grey prediction to capture associative curve which can be seen in Figure 1. From the trend of this vivid figure, we not only can come to the conclusion that the number of English and Chinese users tend to increase but also find that Spanish development will reach the period of stagnation. Secondly, for further improvement, we take birth rate, death rate, economic factors and the immigration into consideration and establish the language communication model. This model is deduced from the population prediction model and virus transmission model. After data normalization, the eventual curve indicates that current top-ten languages seem to be replaced by other languages. This transformation phenomenon also occurs among such top-ten languages. For instance, Hindustani will replace Spanish in the future when seen from Table 1. What's more, after predicting the migration pattern, we can draw the conclusion that some range of languages' dissemination has obvious change. As show in vivid Figure 14, we know English will popularize widely among neighboring countries such as Canada, Mexico, Cuba and Russia. Moreover, with regard to how to manage international offices' quantity and locations in the world, we construct the efficiency model with combination of the Bayes' probability theory and Fussy comprehensive assessment. As a result, we obtain 9 optimal plans to establish the international offices. Intelligible result is showed in Table 4 and Table 5. In short, our model is reasonable and feasible, which can accommodate to different situation.

**Corresponding Author:*
Man Xiao,
Chongqing University of Technology International College,
No. 459 *Pufu Avenue, Longxing Town, Yubei District, Chongqing,* 401135*, China.*
E-mail: 1121952597@*qq.com.*

 DOI: https://doi.org/10.30564/ret.v1i4.110

1. Introduction

1.1 Background

From the perspective of languages' development, we are aware that multifarious languages have different tendency of development. However, what is the concrete quantity of language users in the world? Numerous researchers have been attracted by such issue. With the development of technology and economy in the world, the globalization becomes faster and faster. And we all know that languages are closely related to its environment, moreover, it plays an increasing important role in the process of globalization. At the same time, the positions of languages and languages themselves have also totally changed under the influence of globalization. Furthermore, the number of language users is going to change probably due to plentiful reasons. Through analyzing all relative influence factors, we know the first reason is the languages' popularization by government and its promotion. What's more, languages' popularization also has strong relationship with schools, social pressure, immigration and even emigration. In the era of globalization, there still exist plenty of essential factors which can connect some intermingled languages. These factors are international business relationship, the increase of global tourism, communication among electronic users, social media.

Figure 1. The distribution of languages

1.2 Mathematical Notation

See Table 1

2. Fundamental Assumption

(1) Suppose ignore unpredictable or high-impact, low probability events such as asteroid collisions and diseases.

(2) Suppose all language speakers won't forget their languages including mother language and the second languages.

(3) Suppose ignore the mass migration because of war.

Table 1. Mathematical notation

Symbol	Meas
p_1	The population of native people only can native language
n	The number of people only can native language
m	The population of can native language and can other languages
r	Mobility parameters
m	The coefficient of increase or decrease of birth rate
r	Birth rate
d	mortality rate
s	Economic factors
l	The number of immigrants entering
b	The social pressure coefficient
p_2	The population of foreigner who can use the special language

3. Model

3.1 Preparatory Regular Inquiry

3.1.1 Original Data Analysis

Aimed to gain the relative accurate population about a special language speakers' development tendency in the future, to begin with, we look up numerous literature and find that Grey Prediction model tend to have better performance than other predict model. With much intangibility, so we select English, Spanish and Chinese as our original sample test.

3.1.2 Grey Prediction Model

Suppose $\partial^{(0)}(k)$ is series of n elements, according to Grey prediction we get Formula (1).

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

Through series of transformation, we get a white differential equation corresponding to the differential function about GM (1,1). Finally, we obtain three Formulas as below.

$$\frac{d\partial^{(1)}}{dt} + a\partial^{(1)} = b \tag{2}$$

$$\partial^{\wedge(1)}(k+1) = (\partial^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a}, k = 1,2...,n-1 \tag{3}$$

$$\partial^{\wedge(0)}(k+1) = \partial^{\wedge(1)}(k+1) - \partial^{\wedge(1)}(k), k = 1,2,...,n-1 \tag{4}$$

  DOI: https://doi.org/10.30564/ret.v1i4.110

Shown in the above Formulas, $\partial^{(0)}(k)$ is the Grey-scale coefficient, a means the develop coefficient, b stands the work effect.

3.1.3 Result Analysis

According to the population of Chinese, Spanish and English, we select data randomly to plot and predict the relative population curve in the future. Figure 2, Figure 3 and Figure 4 have been shown in the following:

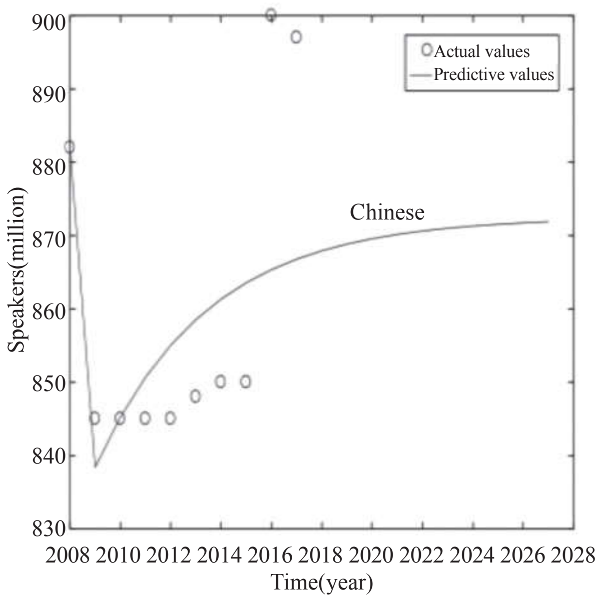


Figure 2

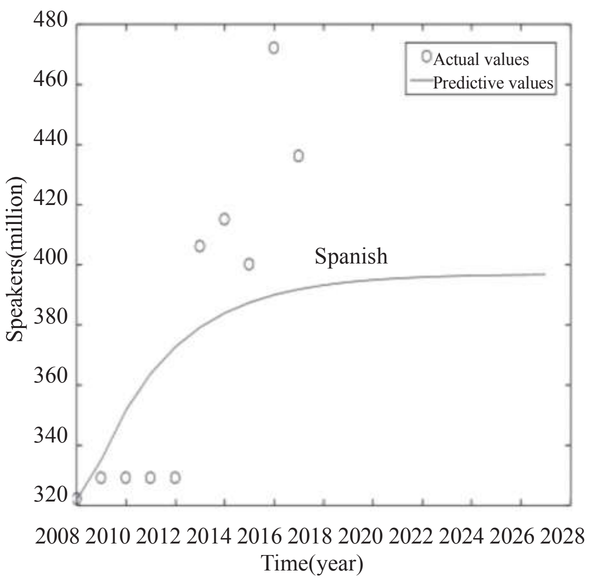


Figure 3

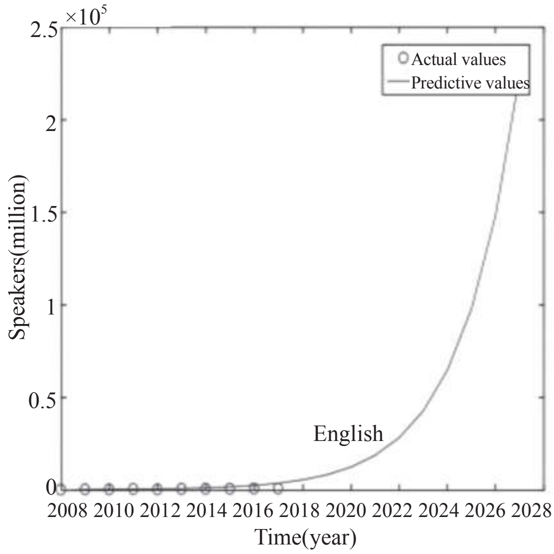


Figure 4

As show in the above figures, we can easily find that most of the predict result are away from the actual position. So we can find that Grey prediction model have worse efficiency. In order to deal with this issue which is mentioned above, we take many essential factors into account and construct a better model called language communication model in the bellow process.

3.2 Language Communication Model

3.2.1 General Description of the Model

In order to predict the tendency about the quantity of native speakers and total language speakers in the next 50 years, we construct a language communication model to deal with this issue. In the process of modeling, we take amount of essential factors into consideration, such as international business relation, increased global tourism, the use of electronic communication and social media. However, aimed to gain a more accurate result, we simplify our model and ignore some unimportant elements. In our model, native language users include two parts. The first one is pure language users, and the second one is multilingual speakers. The simple and readable flow chart Figure 5 as below:

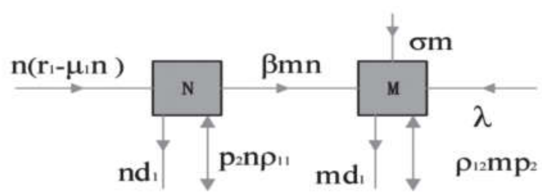


Figure 5. The distribution of native speakers

As show in the above-mentioned Figure 5, n is the pure language user, m is the multiple language user, r_1 is the pure language user birth rate,μ_1 is a relevant factor with the maximum environment capacity, d_1 is the pure language user death rate, p_2 is the foreigners who can use such language, ρ is the competition coefficient, λ is the emigration, δ is the economic element.

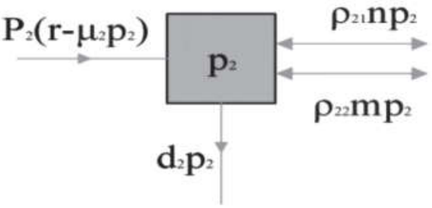


Figure 6. The distribution of foreign speakers

Moreover, for the second part of language speakers, which are total foreign language users, we consider such part's most important elements are birth rate, death rate, and migration. Finally, we simplify many processes and eventually gain the readable flow chart Figure 6 as above.

3.2.2 Model of Language Communication

(1) Main Model

Aimed to calculate the birth rate, we take numerous of influential elements into account. And under these conditions, we take advantage of the former research about logistic model, through further exploration and then we obtain the Formula (5).

$$\frac{dx}{dt} = r(1-\frac{x}{x_m})x \tag{5}$$

As Shown in the Formula (5), r stand the stable increasing rate, x_m is the maximum environment capacity.

From our common sense, the amounts of native language speakers will increase of reduce gradually with era's change. Such as death rate d_1n, the emigration rate $\rho_{11}np_2$, and social pressure βmn. For the social pressure, that is the motivation for child and deep learner to learn a second language. These people also have death rate d_1m, the migration rate $\rho_{12}mp_2$, economic influential factors and so on. Foreign language users are native language speakers but aren't native people. And such people are remarked as p_2 and have strong competition with native language speakers. However, there are also death rate d_2p_2, migration rate $\rho_{21}np_2$ and $\rho_{22}mp_2$. Considering about all essential highlighting point, we build corresponding differential equation as follows:

$$n' = n(r_1 - \mu_1 n) + \rho_{11}np_2 - d_1 n - \beta mn \tag{6}$$

$$m' = \beta mn + \lambda + \sigma m + \rho_{12}mp_2 - d_1 m \tag{7}$$

From the Figure 5 and Figure 6, we can obtain Formula (8) as follows:

$$p_2' = p_2(r_2 - \mu_2 p_2) + \rho_{21}np_2 + \rho_{22}mp_2 \tag{8}$$

Where p_1 stands the quantity of native speakers, n is the pure native language speakers, m means someone who can speak more than one language, ρ is the mobility parameter, μ stands the increase or decrease coefficient of birth rate, σ is the economic influential element, λ is the number of immigration, β means the social pressure coefficients.

To simplify our model, we look up numerous literatures, and we successfully find that many paper said the quantity of native speakers is equal to the addition of pure native language speakers and someone who can more than one language. We get Formula (9):

$$p_1 = n + m \tag{9}$$

We can receive a replacement Formula as Formula (10):

$$(r_1 - \mu_1 n)n \to (r_1 - \mu_1 p_1)p_1 \tag{10}$$

Through a series of calculate, we can also simplify the former Formula and get Formula (11):

$$p_1' - m' = (r_1 - \mu_1 p_1)p_1 + \rho_{11}p_1 p_2 - \rho_{11}mp_2 - d_1 p_1 + d_1 m - \beta m(p_1 - m) \tag{11}$$

From the former searchers' study, we know the immigration rate or emigration rate is same under a same condition. So we get Formula(12):

$$\rho_{11} = \rho_{12} \tag{12}$$

Simplify the above Formulas, we finally receive Formula(13):

$$p_1' = (r_1 - \mu_1 p_1)p_1 + \rho_1 \ p_1 p_2 - d_1 p_1 + \lambda + \sigma m \tag{13}$$

In conclusion, we take advantage of Formula(5) to Formula(13), and get the below Formula (14)

$$\begin{cases} p_1' = (r_1 - \mu_1 p_1)p_1 + \rho_{11}p_1 p_2 - d_1 p_1 + \lambda + \sigma m \\ m' = \beta m(p_1 - m) + \lambda + \sigma m + \rho_{12}mp_2 - d_1 m \\ p_2' = p_2(r_1 - \mu_2 p_2) + \rho_{21}(p_1 - m)p_2 + \rho_{22}mp_2 \end{cases} \tag{14}$$

(2) Data Normalization

As we can see, big data tend to be a huge dilemma during the process of calculating. Aimed to receive a series of desired and accurate result, we take advantages of data normalization. That's to mean, we should deal with the given data which aim to limit whole data in a relative range. As

  DOI: https://doi.org/10.30564/ret.v1i4.110

for our work, we adopted the linear function transformation, and then we analyze such function's thinking process in order to let all data in the range of [0,1]. We obtain Formula (15) as follow:

$$y_i = \frac{MaxValue - x_i}{MaxValue - MinValue}, (i = 1,2,...,n) \quad (15)$$

3.2.3 Result Analysis

(1) Post-processing Analysis

On the basis of the original data belong to every language, we deal with the top ten languages in the world, such as Chinese, Spanish, English, Hindi, Arabic, Bengali, Portuguese, Russian, Punjabi, and Japanese based on the language communication model, we extract three essential variables to analyze rigorously. These three p1 (the quantity of native speakers), p2 (the important elements are number of foreigner speakers,) and total (all the number of a special language speakers). The most highlighting point is that we use MATLB to calculate and predict, and last we plot the top ten languages in the world just is as same as the former mentioned.

(2) Image Analysis

Take advantage of the acquired data, see Figure 7, Figure 8.

First of all, we can easy get that two languages have relative biggest quantity's speakers, and there are Chinese and Hindi. On the contrary, however, the relative smaller quantity of language speakers is Bengali and Japanese. Moreover, the number of people who regard Chinese and Hindi as the first language have been in the state of increasing. And they are present by S curve generally. What's more, the population of regarding Hindi as the mother language is rising rapidly. What worth to pay attention to is that the two languages tend to be at a flat state in 2050? Obviously, we can find that Spanish speakers' quantity also has the same trend of increasing from 2010 to 2025, and after 2025 it will remain steady. Furthermore, when pay attention to English, Arabic, Bengali, Russian, Portuguese and Punjabi, we can easily find that they tend to rise between the year of 2010 and the year of 2010 but such increment quite marginal. During the same period, we can see that Portuguese's and Punjabi's users' quantity has slightly dropped. And the number of Japanese speakers is still being in the steady state.

Moreover, we will discuss relative conspicuous characteristics of population curve, which is presented base on the numerous data of the second language' population. Look at the Figure 9 and 10:

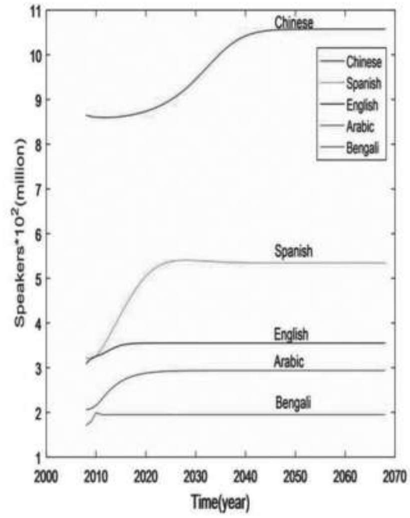


Figure 7. The prediction of first language

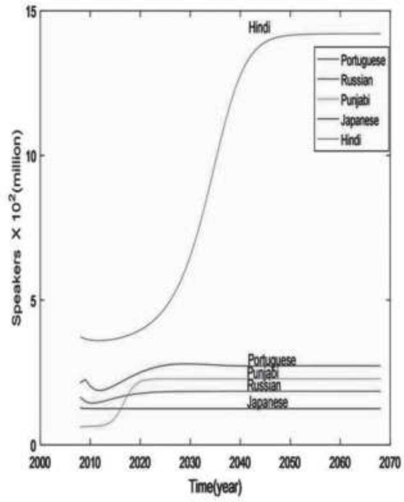


Figure 8. The prediction of first language

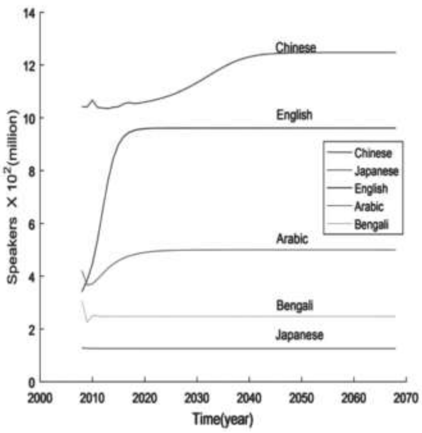


Figure 9. The prediction of total

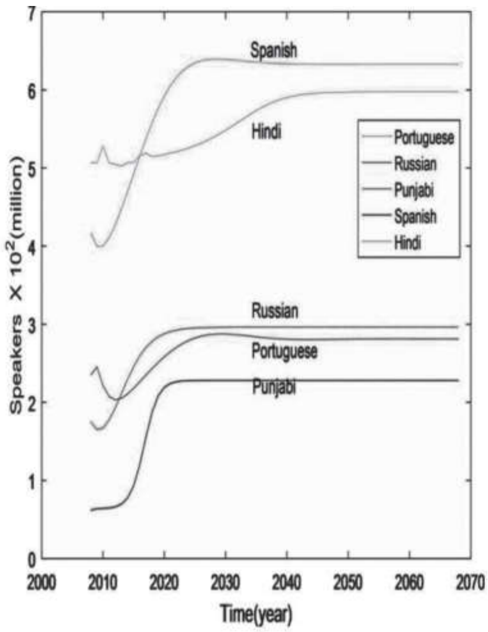


Figure 10. The prediction of total

From the figure, we know that the majority of Chinese, English, Japanese and Punjabi have the fewest speakers. Speakers of the majority of language were increasing between 2010 and 2020, except Japanese is be in a stable status. For Most languages, the number of speakers was declining, and the number of English speakers is increasing rapidly. Punjabi were slowly rising, all language speakers reach a steady state between 2020 and 2030 eventually.

(3) Compare with Now and Future

We take ten languages as our sample. They are Chinese, Spanish, English, Hindustani, Arabic, Bengali, Portuguese, Russian, Punjabi, and Japanese. In order to get more reliable rank, we select the year of 2017 to be the start position, and then we process a sequential steps to gain a relative dependable predict result in the latter 50 years. In the end, we design Table 2 and Table 3 as bellow. From the two tables we compare the rank with the year of 2017 and the year about 2067. What surprising us is that some languages' range of application will change? To start with, we can easy see that Chinese's rank is still stand at the first position. But the second language, Spanish's rank will become the third one in the latter time. On the contrary, the second rank will be Hindustani even though it is the fourth rank in the year of 2017. English tend to have a slightly decrease when compare to 2017. However, the Arabic and Japanese all keep the original level in 2067. Furthermore, it is obvious to find that Russian and Portuguese have an ideal rising trend, and so on.

Table 2. 2017 Year (Unit:10^2 million)

Language	L1	Rank	L2	Total
Chinese	8.97	1	1.93	10.9
Spanish	4.36	2	0.91	5.27
English	3.71	3	6.11	9.82
Hindustani	3.29	4	2.15	5.44
Arabic	2.9	5	1.32	4.22
Bengali	2.42	6	0.19	2.61
Portuguese	2.18	7	0.11	2.29
Russian	1.53	8	1.13	2.66
Punjabi	1.48	9	0	1.48
Japanese	1.28	10	0.01	1.29

Table 3. After 50 Years (Unit:10^2 million)

Language	L1	Rank	L2	Total
Chinese	20.38585453	1	3.56	23.9458
Hindustani	8.627293638	2	2.819876	11.4472
Spanish	5.3433	3	0.9852	6.3285
English	3.55	4	6.0608	9.6108
Arabic	2.9381	5	2.0652	5.0033
Portuguese	2.7322	6	0.0796	2.8118
Punjabi	2.2798	7	0	2.2798
Bengali	1.9509	8	0.5235	2.4744
Russian	1.8516	9	1.1122	2.9638
Japanese	1.2511	10	0.0100088	1.2611

3.3 Language Transformation Range Analysis

With the increment development of science and technology, the entire population belong to next 50 years will remain at a considerable level. However, what should we do to predict how about the geographic distributions of these languages change over this same period of time with utilizing the global population and human migration patterns for the next 50 years? In order to simplify the process, we design a selected plan to choose our text sample. Firstly, based on the rank of nowadays and next 50 years, we purposely select the top-ten language. Then we consult literature and numerous time-reports to comprehend how many countries is speaking a particular language. At the same time, we record the corresponding countries one by one. After this we select six counties which occur many times. Furthermore, we select tree main countries as the example, and they are American, China, and Singapore. Showing the Figure 11, 12, and 13:

  DOI: https://doi.org/10.30564/ret.v1i4.110

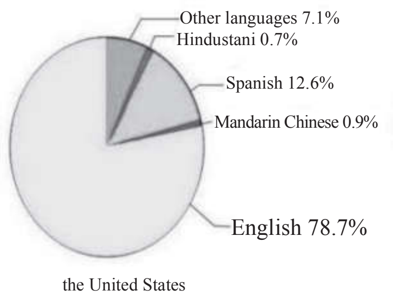


Figure 11

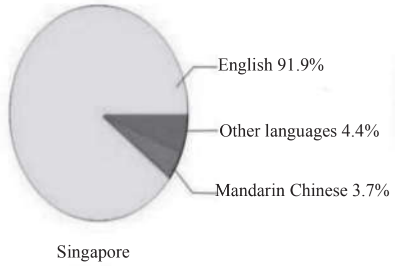


Figure 12

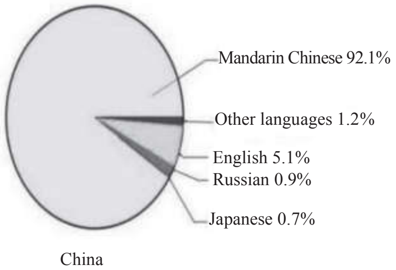


Figure 13

From the above vivid pictures, we can easy see that American have plenty of English speakers is as same as China. The number of English speakers will rapidly increasing when the Chinese users is only gradually aggrandize in the next 15 years. So we predict English tend to spread widely and fast. Based on our perspective, we believe this tendency is going to affect its' neighboring countries producing them to learn English. However, on one hand, Canada, Mexico, Cuba and many developing countries tend have an increasing number of English learner in the next 15 years. But Chinese is slightly increase. However, Chinese and English all grow but Chinese has more preponderance than English after 16 years. So we can come to the view that South Korea, Russia, Vietnam, India and Mongolia which is close to China are going to have more and more Chinese leaner. At the same time, English's spread will meet some stuff.

3.4 The Efficiency Model

3.4.1 General Description of Model

With the purpose of deciding where we would locate offices and what languages would be spoken in the offices, we construct the efficiency model to solve this issue. Taking the changing nature of global communication, and in an effort to save client company resources into consideration, we think over and make the best use of the Bayes' theorem of probability theory and Fussy comprehensive assessment of Fuzzy Sets.

According to our description of language communication model, we can get $p_1=m+n$, there p_1 is the population of a special native language speakers, n is the pure language speakers, and m is someone not only can native language but also can other second languages. In this part, we take the population of English speakers as example. And we gain Formula (16) as follow:

$$q_1^e = n_e + m_e \tag{16}$$

We record $q_1^e = n_e + m_e$ the native ne stand for the number of people who only can use English, m_e present someone not only can speak English but also can use other language. Use m_i (i=1,2,...,9) represents someone who can other languages including the ith languages. However, $p(m / q^e)$ stand the probability of mi which is based on the number of people regard English as the i 2 second the language.

So based on the Bayes' theorem, we can get the Formula (17):

$$p(m_i / q_2^e) = \frac{p(m_i q_2^e)}{p(q_2^e)} \tag{17}$$

Through a series of processing, we get the detail Formula (18), which we call it the efficiency formula.

$$p(m_i / q_2^e) = \frac{p(m_i)p(q_2^e / m_i)}{\sum_j^n p(m_j)p(q_2^e / m_i)} \tag{18}$$

Suppose m_i is the someone who can the i_{th} languages and other second languages. And m_i^e is a part of m_i and these people can only English, m_i^{other} means can use the i_{th} languages and other second languages except English. So we get Formula (19).

$$m_i = m_i^e + m_i^{other} \tag{19}$$

According to Fussy comprehensive assessment, we can get the operator M(•, ⊕). So Formula (20) and Formula (21) is :

$$s_k = \min(1, \sum_{j=1}^n \mu_j r_k), \ k = 1,2,...,n \tag{20}$$

$$q_1^e = n_e + m_e \quad (21)$$

Through accumulating every $p(m_i / q_2^e)$, we can obtain the efficiency $\sum_{i=1}^{n} p(m_i / q_2^e)$. The last but not least, after calculating the efficient we personally set a critical comparison valueζ. If the accumulation of () is bigger than the critical comparison on value, then we can get a series of language' combinations. Formula (22) is:

$$\sum_{i=1}^{n} p(m_i / q_2^e) > \xi \quad (22)$$

3.4.2 Result Analysis

According to the model of our design, we gain the efficiency between 0.6 and 0.9. We divide them into three parts. Next, every part divides into three options. Every option has a different efficiency value. Finally, we find those concentrated financial and economic city.

In the Table 4, we intercepted three different options when 'Efficiency' satisfies 0.6, every option all has six different languages, option one is the most effective.

In the Table 5, we intercepted three different options when 'Efficiency' satisfies 0.7. Option two just has two options with six language and cities. But option two has five language and cities. Option one is the most effective.

In the Table 6, we intercepted three different options between 0.8 and 0.9. Option two just has two options with six language and cities. But option two has five language and cities. Option one is the most effective.

Table 4. 0.6<Efficiency<0.7

								Subtotal
Option One	Language	Punjabi	Chinese	Hindustani	Arabic	Japanese	Punjabi	0.6816
	City	Toronto	Beijing	Mumbai	Dubai	Tokyo	Toronto	
Option Two	Language	Russian	Hindustani	Japanese	Spanish	Portuguese	Russian	0.6761
	City	Moscow	Mumbai	Osaka	Madrid	St Paul	Moscow	
Option Three	Language	Russian	Chinese	Japanese	Spanish	Bengali	Russian	0.6663
	City	Moscow	ShenZhen	Tokyo	Madrid	Dhaka	Moscow	

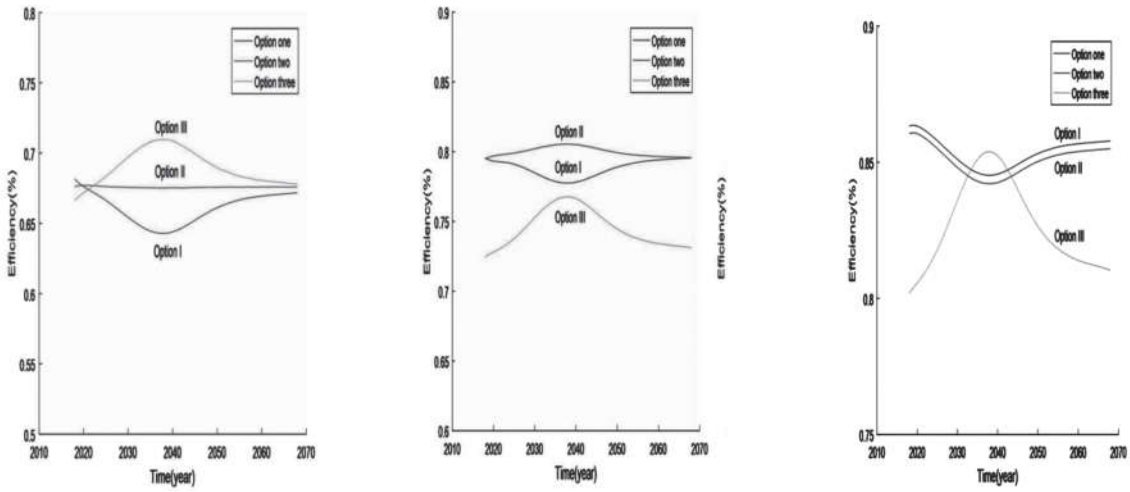
Table 5. 0.7<Efficiency<0.8

								Subtotal
Option One	Language	Spanish	Bengali	Chinese	Arabic	Hindustani	Punjabi	0.796
	City	Madrid	Dhaka	Tianjin	Dubai	Mumbai	Toronto	
Option Two	Language	Arabic	Hindustani	Punjabi	Japanese	Russian		0.795
	City	Dubai	Mumbai	Toronto	Osaka	Moscow		
Option Three	Language	Japanese	Punjabi	Arabic	Russian	Portuguese	Chinese	0.725
	City	Tokyo	Toronto	Dubai	Moscow	St Paul	Beijing	

Table 6. 0.8<Efficiency

								Subtotal
Option One	Language	Russian	Chinese	Punjabi	Japanese	Hindustani	Arabic	0.8833
	City	Moscow	ShenZhen	Toronto	Osaka	Mumbai	Dubai	
Option Two	Language	Russian	Hindustani	Chinese	Punjabi	Arabic		0.8605
	City	Moscow	Mumbai	Tianjin	Toronto	Dubai		
Option Three	Language	Arabic	Bengali	Russian	Portuguese	Punjabi	Chinese	0.8018
	City	Dubai	Dhaka	Moscow	St Paul	Toronto	Beijing	

  DOI: https://doi.org/10.30564/ret.v1i4.110



(a) 0.6<Efficiency<0.7 (b).0.7<Efficiency<0.8 (c) 0.8<Efficiency

Figure 14

4. Conclusion

In this dissertation, through the establishment of two models, we predict the multifarious languages' developmental tendency in the future and evaluate how these languages tend to spread in the world. Before establish our models, firstly, we try to use Grey Prediction model to preliminary analysis relative data about the number of native speakers, muti-language speakers, and foreigner speakers. When to come to the language communication model, that's creative since we combine the virus transmission model and population prediction model. Additionally, we also make best use of the data normalization when modeling. What's more, after determining main languages and their general distribution in the world, we construct the second model remark it as the efficiency model. Which take advantage of the Bayes' theorem and Fussy comprehensive assessment? During the process of modeling, based on different economic condition, we receive the efficiency and ultimately design 9 combined plans. To sum up, we make an objective analysis about our models' advantages and disadvantages, and then try an attempt to further expand the model aim to make our model more successful.

References

[1] Mingqiang Dai, Yaxin Zhu. Mathmatical Model and Its Application[M]. Science, 2015, 2:41-53.

[2] Feng Lin, Xiulan Zhang. Mathematical Modeling Experiment[M].Chemical Industry Press.

[3]https://en.wikipedia.org/wiki/List_of_languages_by_total_number_of_speakers[EB/OL]. January 17, 2018.

[4] "Summary by language size"[M]. Ethnologue, Retrieved 2016-04-06.

[5] Juliette Stehl©, Nicolas Voirin et al. Simulation of an SEIR infectious diseasemedel on the dynamic contact network of conference attendees[J]. BMC Medicine, 2011, 9:87.

[6] Zheng Lie, Yongsong Zhu et al. Mathematical Model[M]. Science, 2013,7:189-191.

 DOI: https://doi.org/10.30564/ret.v1i4.110