	Team Control Number		
For office use only		For office use only	
T1 ________________	**91973**	F1________________	
T2 ________________		F2________________	
T3 ________________	Problem Chosen	F3________________	
T4 ________________	**B**	F4________________	

2018 MCM/ICM Summary Sheet

The Trends of Global Languages

In the last few years, the globalization of the economy has affected all aspects of the world and has also affected the trends of global languages. Moreover, it is important for a large multinational service company to investigate trends of global languages to seek appropriate locations for new offices. Aimed to accomplish this task, we employ **PG-M model** to filter factors and conclude the most influential criteria including population growth rate , Immigration rate, and GDP per capita , the cultural soft power and so on.

In order to calculate the weight of each factor, we use **Improved Markov Model** to investigate trends of global languages. Then we use our model to predict the variation of the numbers of native and total language speakers in the next 50 years. Ultimately, considering the global population and human migration patterns predicted for the next 50 years, we eventually conclude the changes of the geographic distributions of these languages over this same period of time.

Based on **FM model**, we can locate these offices and determine which kinds of languages should be used in the offices. We confirm 4 memberships of GDP, cultural soft power, trends of population and Infrastructure, then determine their evaluation weight. **The sensitivity analysis** of our model has pointed out that small alteration in our constrains affects the location of offices slightly. Finally conclude that we should open less than 6 offices.

Abstract: trends of language, geographic distributions of languages, population growth, migration pattern, Fuzzy Mathematical Analyses, Markov Chain.

To: the Chief Operating Officer of the service company
Date: Tuesday, February 13, 2018
Subject: The Trends of Global Languages

Through the establishment of two models, we imitate the trends of languages over time and the changes of geographic distributions of 13 languages. We consider several factors that could effect number of the total language speakers, but finally conclude that the main factors are economy and culture. Based on the calculation of the Fuzzy Mathematical Model, we can locate the new offices in the long term and in the short term. Moreover, we estimate the appropriate number of new offices.

The **Population growth-Markov Model** shows that the number of native and total languages speakers would increase in various degrees. With the increasing number of total speakers, the result of Model shows huge unstability of prosperity, population and economy in various countries. As a result, any of the languages listed could be replaced by another one. Our suggestions are as follows:

Short term: open new offices in countries whose native language, GDP per capita and CSP ranks well. By this means, it is more convenient to introduce qualified personnel and more client to serve, in the meanwhile, the profit will be larger. The cities we choose are New York, Barcelona, Jakarta, Moscow, Hamburg, Dar Es Slaam(capital of Tanzania).

Long term: mainly discuss the developing countries. Despite the fact that the GDP of these countries are not rank well, with the rapid development of their economy and population, these developing countries shows potential. So locate offices in the developing countries in the long term can get more profit. For example, India and China.

At last, we suggest that the number of new opening offices should be **less than 6**. Based on our model and considering the rapid development of communications techniques and internet, open 6 offices would not profit well.

But we need more information to analyze the profits of the company:
Historical data of the fix cost of opening a new office.
Historical data of the profit after opening a new office
The demonstration of the potential business we can take into account.

Contents

1 Overview

1.1 Background

There are currently about 6,900 languages spoken on Earth. But only twenty dominant languages are increasingly used more widely, on the other hand, many languages rapidly become endangered.

However, much of the world's population also speaks 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 Restatement of the Problem

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 Assumptions and Justifications

To simplify our problems, we make the following basic assumptions, each of which is properly justified.

- **The impact of policy on the second or third language is relatively small.** Because Policies do not often affect the second or third language, but also to simplify calculations.
- **The impact factors we choose have no mutation in the short term.** This is to simplify the problem and make it easier to calculate and predict.
- **The impact of migration on the native language of big countries is small.** Because of the huge population of big countries, this impact can be ignored
- **The impact of the birth and death rate on the second language is small.** To simplify the model, making it easy to calculate.

3 Notation

Abbreviation	*Description*
CSP	Cultural soft power
P	Transition probability matrix of the data of year 14-16

$x(t)$	The number of native language speakers of big nations.
r	The population growth rate
Y_i	The number of second language speakers after i years
A	The weight of the project in decision making
B	T comprehensive evaluation of the project

4 Model Theory

4.1 Analyses of Population Growth-Markov model

Because of the shortcut of the database, some data of the languages listed can not be found completely. So we discuss the 13 kinds of languages that have data more completely in order to better simulate the distribution of trends of languages.

To evaluate the influence level of the factors which have been described in the background paragraph above as well as other factors our group may identify quantitatively, that is, the correlation of various factors and language use, we examined variables include migration rate, population growth rate, economic factors(GDP), cultural soft power and other factors from two dimension: native language and second language, using SPSS.(table 3)[1]

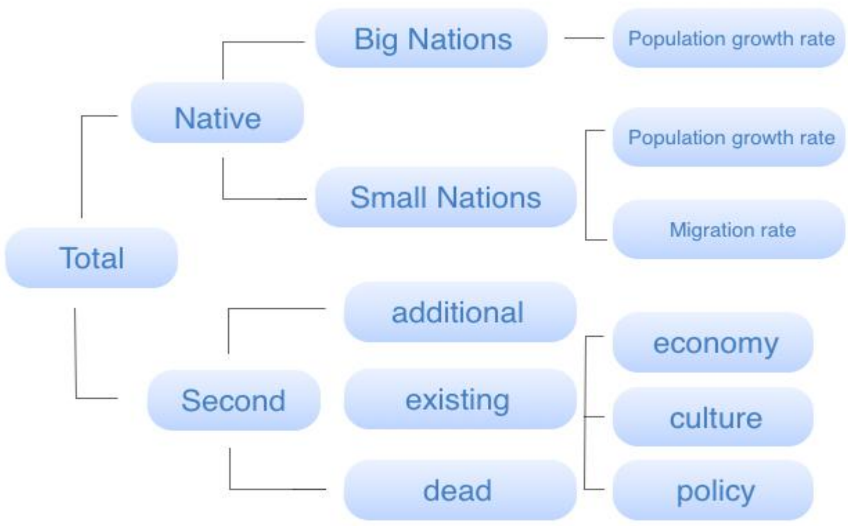

Figure 1 AHP of Total Language Speakers

- To discuss the variation of the language usage, we pick 13 kinds of languages as follows: English, Chinese, Indonesian, French, Hindustani, Urdu, Spanish, Swahili, German, Filipino, Thai, Russian, Bengali.

- As for the factor of economy, we choose the GDP of 13 representative countries of 13 languages. (table 1)

- As for the factor of culture, we discuss the cultural soft power of the 13 countries above (table 2)， and generate the world ranking of them. From the table we can see that cultural soft power indeed has influence on the numbers of language speakers.

The calculation formula of cultural soft power(CSP) is given as follows:

$$CSP = \frac{\text{International Appeal} + \text{International Mobilization} + \text{Domestic Mobilization}}{3}$$

Table1 GDP Of 13 Representative Countries[1]

Country Name	2008 [YR2008]	2009 [YR2009]	2010 [YR2010]	2011 [YR2011]	2012 [YR2012]	2013 [YR2013]	2014 [YR2014]	2015 [YR2015]	2016 [YR2016]
China	3805.025999	4142.038286	4560.512586	4971.544929	5336.060143	5721.693819	6108.238775	6496.624013	6893.776361

United States	49364.64455	47575.60856	48373.87882	48783.46859	49497.58585	49976.62877	50782.5207	51855.91348	52262.78416
India	1156.932527	1237.339786	1345.770153	1416.403391	1474.967674	1550.14223	1646.781252	1758.043376	1861.491029
Saudi Arabia	19792.72038	18861.11	19259.58726	20575.49795	21056.34715	21005.01212	21183.46489	21507.95569	21395.35978
Spain	32303.24148	30874.12601	30736.62785	30321.70487	29414.85692	29008.02098	29496.37893	30530.56666	31532.81638
Russian Federation	11089.92995	10219.52131	10674.98771	11121.51353	11493.40204	11615.69592	11493.72654	11144.5969	11099.17306
Bangladesh	698.564892	725.76626	757.6717572	797.4116804	839.5136872	879.5819661	922.161128	971.641988	1029.578212
Portugal	22829.84787	22128.84552	22538.65408	22159.47542	21353.23026	21228.08961	21533.48999	22016.83692	22428.11468
Indonesia	2876.88504	2970.044131	3113.480635	3262.748613	3415.351267	3560.10658	3692.942875	3827.548307	3974.058485
Germany	42365.0975	40086.10476	41785.55691	44125.33141	44259.25991	44354.73689	45022.56535	45412.55681	45745.7899
Japan	45165.88716	42724.53491	44507.67639	44538.70809	45276.8287	46249.26351	46466.12297	47082.68859	47623.27093
France	41545.29345	40116.37953	40703.34279	41349.19094	41224.72903	41249.45128	41431.03961	41689.70798	42013.28629
Korea, Rep.	20803.50054	20843.1348	22086.95292	22724.70557	23123.76136	23685.40671	24323.57284	24870.7709	25458.88701

Table 2 CSP Of 13 Representative Countries[2]

Language	Represent Contry	culture soft power	Rank	Standard line
Chinese	China	1	19	20
English	the United States	2.139	7	20
Hindi	India	2.584	4	20
Arabic	Saudi Arabia	0.49	50	20
Spanish	Spain	1.565	13	20
Russian	Russia	0.91	24	20
Bengali	Bangladesh	0.297	77	20

Portuguese	Portugal	1.022	18	20
Malay	Indonesia	0.402	62	20
German	Germany	1.586	12	20
Japanese	Japan	1.564	14	20
French	France	2.13	8	20
Korean	Korea, Rep.	0.81	29	20

We only discuss the official language of a country instead the dialect such as Punjabi, Javanese.

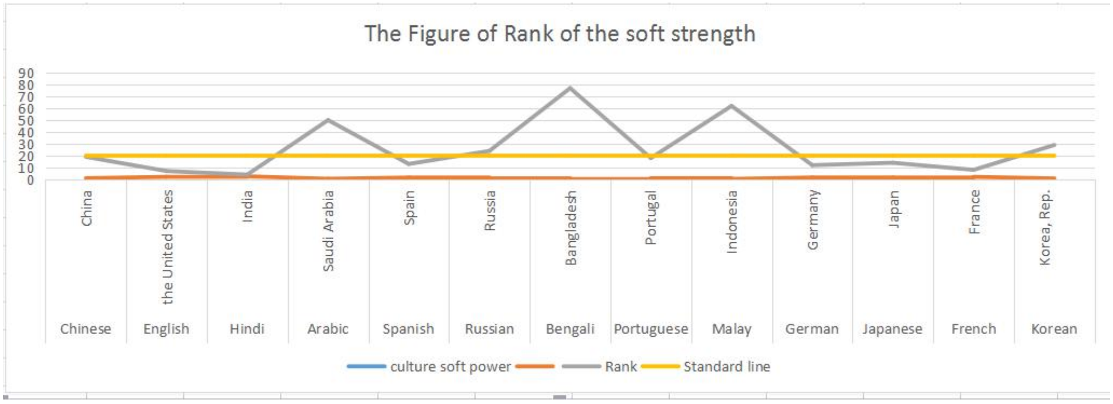

Figure 2 The Figure of Rank of The Soft Strength

From the figure above, most of the representative countries in the 13 languages we selected are all ranked in the top 20, indicating that there is a certain positive correlation between the number of their second languages and their cultural soft power.

table 3 Likelihood Ratio Tests Table

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	.000[a]	.000	0	.
GDP	20.683	20.683	24	.657

This table is a weight analysis of GDP per capita in China and the number of speakers in a second language. We can find that the correlation is 0.657, which is greater than 0.3. So it can be said that GDP per capita is significantly related to the number of people in a second language. In addition, we conducted a significant analysis of the selected 13 countries and found that both were greater than 0.3.

4.1.1 Markov chain model

The Definition of Markov Chain

A Markov chain is "a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.

Markov chain can be described as a random time series $\{\xi_n, n\in T\}$, where T is a finite set$\{1,2,\ldots,N\}$, state space $I=\{1,2,\ldots,M\}$ (N is the number of discrete time periods, M is the number of the discrete state of the event). For any positive integer step m, and time $n, n-1,\ldots 2,1\in T$, there is a relation between the corresponding event state i, $i_{n-1},\ldots i_2, i_1\in I$:

$$P\{\xi_{n+m}=j \mid \xi_n = i, \xi_{n-1} = i_{n-1},\ldots, \xi_1 = i_1\} = P\{\xi_{n+m}=j \mid \xi_n = i\}$$

For any time step m and time $t \in T$, there is $P_{ij}(t,m)=P_{ij}(m)$ for the transition probability matrixes of the homogenous Markov chain, where $P_{ij}(t,m)$ is the transition probability of a event from state i at time period t to state j at time period t+m. Thus, the homogenous Markov chain $\{\xi_t\}$ is completely determined by its initial state distribution $\{P(1,i), i=1,2\ldots M\}$and its state transition probability matrix $P_{ij}(i,j= 1,2\ldots,M)$[4]

4.1.2 Population growth-Markov Model

The establishment of the Population growth-Markov Model includes the following 3 steps:

Step 1

We split the total number of language speakers into native language speakers and second language speakers as the following equation:

$$N^{[j]} = N^{[j]}_{native} + N^{[j]}_{second}$$

N[j]denote the number of the j-th language speakers，N[j]native denote the number of speakers who use the j-th language as native language，N[j]second denote the number of speakers who use the j-th language as second language.

Step 2: Discussion of $N^{[j]}_{native}$ based on the population increase model

Notice that $N^{[j]}_{native}$ =the original population+births-deaths-the net migration
Then split those countries into two types: big nations& small nations

Studies revealed slight influence on population of a big country affected by migration. As a result, when it comes to the $N^{[j]}_{native}$ of **big nations**, we only need to take over the population growth rate into account. The immigration rate can be ignored.

Now we get the equation:

$$x(t)=\frac{x_m}{1+(\frac{x_m}{x_0}-1)e^{-rt}}$$

Where x_m denote the maximum population that natural resources and environmental conditions can sustain. r denote the population growth rate. x(t) denote the $N^{[j]}_{native}$ of big nations.

The calculation of the $N^{[j]}_{native}$ of **small nations** can be described as:

$N^{[j]}_{native}$ of **small nations**=the original population+births-deaths-the net migration

Step 3: The establishment of improved Markov Chain Model

For the second language or the third language, we consider the two major factors of GDP per capita and CSP. As same as the discussion of native language, split it into 3 parts:

$N^{[j]}_{second}$ =the original population+births-deaths-the net migration

More exactly:

$$N^{[j]}_{second}(t)=\sum_{i=1}^{n}(a_{ij}*N^{[i]}_{pop}*b_i+Y_i(t-1)*P_{ij}-N^{[i]}_{pop}*d_{ij})$$

Where aij denote the probability of new birth in the i-th country learning the j-th language as second language, d_{ij} denote the death rate of the j-th country using the i-th language as second language, b_i denote the birth rate of the i-th country, $N^{[i]}_{pop}$ denote the population of the i-th country.

[5]Suppose the number of the speakers who use the j-th language as second language in the t-th year is . $y_j(t), j=1,2,3$

The probability of that speakers switch from using the i-th language to the j-th language as second language is $p_{ij}, i, j = 1,2,3$.

Note that n is the total number of countries, t is the number of samples we use, then the transition probability matrix of the 3-year data of year 14-16 is

$$P=\begin{pmatrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \ddots & \vdots \\ p_{n,1} & \cdots & p_{n,n} \end{pmatrix}_{n*n}$$

Note that n is the total number of countries, t is the number of samples we use, then the transition probability matrix of the 3-year data of year 14-16 is

$$P=\begin{pmatrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \ddots & \vdots \\ p_{n,1} & \cdots & p_{n,n} \end{pmatrix}_{n*n}$$

the Markov prediction model is $Y_2 = Y_1 P$

Where $$Y_1=\begin{pmatrix} Y_1(0) & \cdots & Y_n(0) \\ \vdots & \ddots & \vdots \\ Y_1(t-1) & \cdots & Y_n(t-1) \end{pmatrix}_{t*n}$$

$$Y_2=\begin{pmatrix} Y_1(1) & \cdots & Y_n(1) \\ \vdots & \ddots & \vdots \\ Y_1(t) & \cdots & Y_n(t) \end{pmatrix}_{t*n}$$

Estimate p_{ij} by the least square method: take the primary industry for example, let e be the fitting error of the primary industry.

Then

$$e=\begin{pmatrix} y_1(1) \\ y_1(2) \\ \vdots \\ y_1(t) \end{pmatrix} - Y_1\begin{pmatrix} P_{11} \\ P_{21} \\ P_{31} \end{pmatrix}$$

Then

$$Q=e^{'}*e=\left(\begin{pmatrix} y_1(1) \\ y_1(2) \\ \vdots \\ y_1(t) \end{pmatrix} - Y_1\begin{pmatrix} P_{11} \\ P_{21} \\ P_{31} \end{pmatrix}\right)$$

Evaluating P_{ij} is equivalent to find the minimum of Q.
Then

$$\frac{\partial Q}{\partial P_{i1}}=-2Y_1^{'}\begin{pmatrix} y_1(1) \\ y_1(2) \\ \vdots \\ y_1(t) \end{pmatrix} - 2Y^{'}{}_1Y_1\begin{pmatrix} P_{11} \\ P_{21} \\ P_{31} \end{pmatrix}$$

Let the derivative be 0,then

$$Y^{'}{}_1Y_1\begin{pmatrix} P_{11} \\ P_{21} \\ P_{31} \end{pmatrix}=Y_1^{'}\begin{pmatrix} y_1(1) \\ y_1(2) \\ \vdots \\ y_1(t) \end{pmatrix}$$

Solving this equation system, the least squares estimation of the transition probability matrix of the primary industry p_{ij} is

$$\hat{P}_{i1}=\left(Y^{'}{}_1Y\right)^{-1}Y_1^{'}\begin{pmatrix} y_1(1) \\ y_1(2) \\ \vdots \\ y_1(t) \end{pmatrix}$$

So the result can be written as

$$\hat{P}=\left(Y^{'}{}_1Y\right)^{-1}Y_1^{'}Y_2$$

Input data to get solutions .(The result of *P* in Appendix 1)

Predict by model

$$Y_{50} = Y_1 P^{50}$$

We can get the data after 50 years.

4.2 Fuzzy Mathematical Model

4.2.1 introduction of Fuzzy Mathematical Model

To locate the offices, we developed this model based on the Analytic hierarchy process(AHP). Because the time is limited, we decide the countries to locate offices first. Then cite references to seek the ideal cities of these countries. Besides, when deciding cities, factors like economy, culture and the transportation must be considered.

4.2.1 establishment of Fuzzy Mathematical Model[4]

Step1: confirm the membership function:

Trends of language(TOL): $$y = \begin{cases} \frac{10x}{x_{\max}} if (x < 0.1x_{\max}) \\ 0 if (x < 0) \\ 1 else \end{cases}$$

GDP per capita: $$y = \begin{cases} 0 if (x < 0) \\ \frac{x}{0.8\overline{x}} else \\ 1 if (x > 0.8\overline{x}) \end{cases}$$

Cultural soft power(CSP): $$y = \frac{x}{x_{\max}}$$

Infrastructure(IF) $$y = x^2$$

Step2: calculate the degree of membership of each country using the membership function(table 4)

Table 4 The Degree of Membership of Countries

State code	USA	CHN	IDN	FRA	IND	PAK	..

GDP	1	0.682	0.442	1	0.193	0.143	..
CSP	1	0.165	0.066	0.352	0.427	0.023	..
TOL	0.205	0.198	0	0	0.223	0.066	..
IF	1	0.103	0.005	0.544	0.343	0.002	..

Then get the fuzzy relational matrix:

$$R=\begin{bmatrix} 1 & 0.682 & 0.442 & 1 & 0.193 & 0.143 & \cdots \\ 1 & 0.165 & 0.066 & 0.352 & 0.427 & 0.023 & \cdots \\ 0.205 & 0.198 & 0 & 0 & 0.223 & 0.066 & \cdots \\ 1 & 0.103 & 0.005 & 0.544 & 0.343 & 0.002 & \cdots \end{bmatrix}$$

according to the evaluation, the weight of each factor is

$$A=\begin{pmatrix}0.116 & 0.578 & 0.234 & 0.072\end{pmatrix}$$

the synthetic evaluation of the nations is:

$$B=AR=(0.908 \quad 0.463 \quad 0.272 \quad 0.816 \quad 0.262 \quad 0.096 \quad 0.666 \quad 0.078 \quad 0.689 \quad 0.188 \quad 0.437 \quad 0.758 \quad 0.097)$$

Step3: Determine evaluation weight

in order to confirm the value of a_{ij}, we suggest that use 1~9 and their reciprocal as scales. table 5 lists the meaning of the scales:

Table 5 The Meaning of The Scales

scales	meanings
1	the two factors have the same importance
3	the former factor is slightly more important than the later one
5	the former factor is evidently more important than the later one
7	the former factor is strongly more important than the later one
9	the former factor is extremely more important than the later one
reciprocal of 2, 4, 6, 8	The mid-range of the contiguous judgments above. If the ratio of the importance of the factor i to the factor j is a_{ij}, then the ratio of importance of the factor j to the factor i is $a_{ji}=1/a_{ij}$

According to the meaning of the scaling, we can construct the fuzzy judgement matrix $D=(d_{ij})_{n*m}$, and compute the feature vector $x=(x_1,....x_n)^T$ corresponding to the largest eigenvalue λ_{max} of D. Then normalization processing is carried out, getting the weight vector of the evaluation index:

$$A = \left(\frac{x_1}{\sum_{i=1}^{n} x_i}, \frac{x_2}{\sum_{i=1}^{n} x_i}, \ldots\ldots \frac{x_n}{\sum_{i=1}^{n} x_i}\right)$$

5 Model Implementation and Results

5.1 Problem A in Part I

We use the data to calculate the number of native languages speakers over 50 years and map changes of it by Python, the graph of the change shown in Figure 3

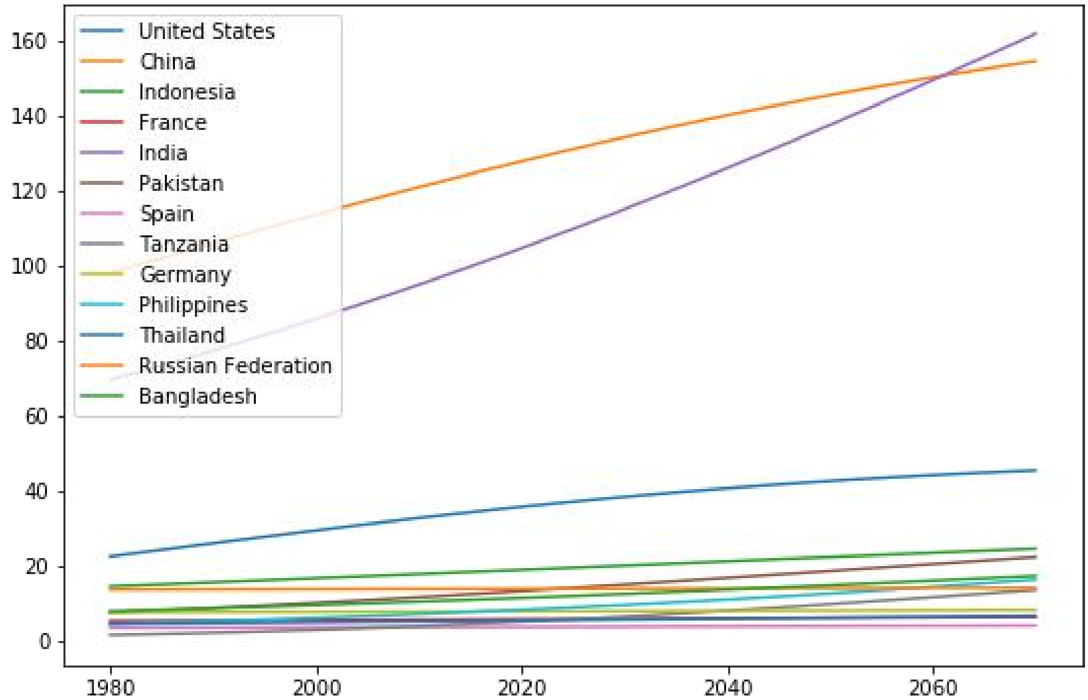

Figure 3 Changes of native language speakers in 50 years

By adding the changes of the number of second languages speakers we get the changes of total languages speakers.(Figure 4)

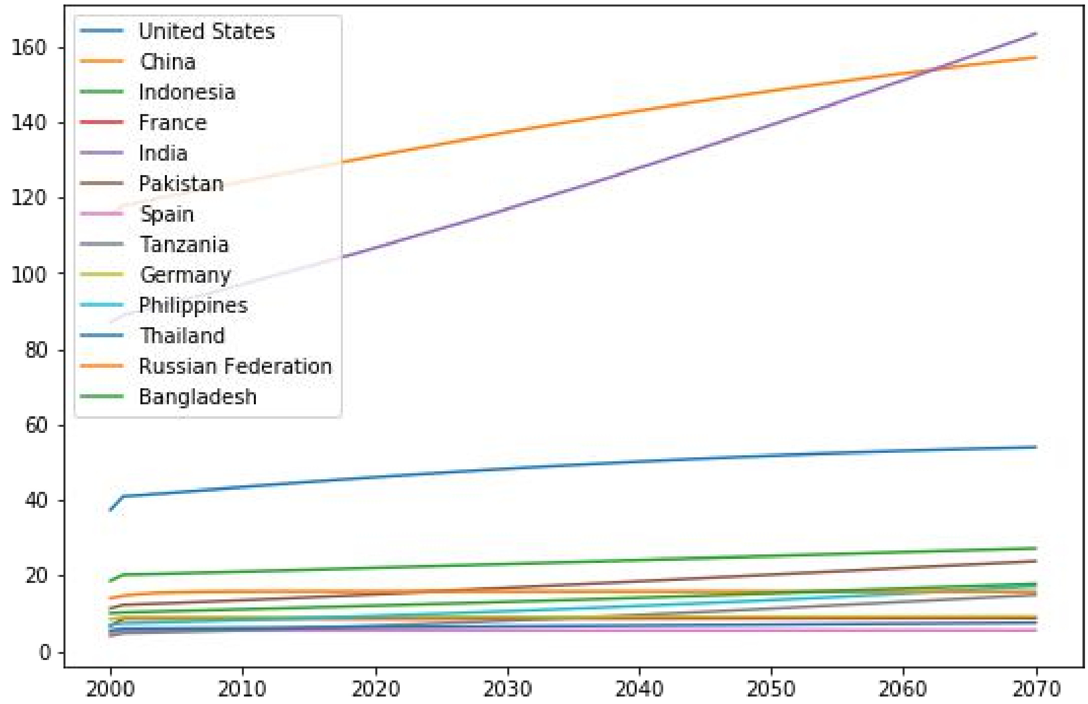

Figure 4 Changes of Total Language Speakers in 50 years

From the table we can see that the number of native and total languages speakers will increase in various degrees.

5.2 Problem B in Part I

Based on the result of the improved Markov model we get the population of the second language speakers

Table 6 Second Language Population After 50 Years

English	Chinese	Indonesia	French	Hindustani
490949418	156333833	141287077	123880799	976447868
Urdu	Spanish	Swahili	German	Filipino
758501056	736649541	663793234	424951760	363258973
Thai	Russian	Bengali		
322482272	837755136	154907934		

From the table we can see that, the population of Bengali speakers will increase, and the rest will decrease.The result shows huge unstability of the prosperity in different countries, so the languages listed would be replaced by any other languages.

5.3 Problem C in Part I

Given the global population and human migration patterns predicted for the next 50 years, we draw a chart showing the geographic distribution over this same period of time. Considering the geographic distribution of native language will not change unless the distribution of nations have changed, we only need to take second languages into account. The chart(mapped by ECharts) of distribution shown in Figure 5

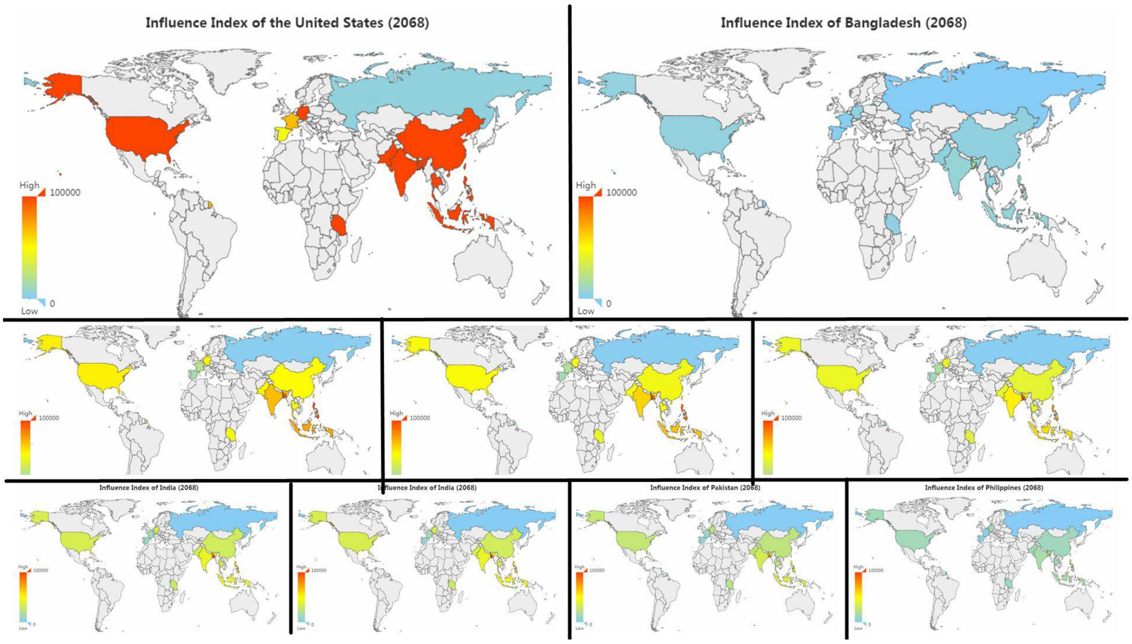

Figure 5 Influence Index of nations

From the chart we can see that, the geographic distributions of 13 languages have changed over time. What's more, the bigger the Integrated Impact Index of a language is, that is, the darker the color, the more countries which use this language as second language and the larger range this language have.

5.4 Problem A in Part II

We choose 4 factors and construct the evaluation system, rank the countries according to these 4 factors and get the final rank by running the Fuzzy Mathematical Model.

1 America 2 Indonesia 3 Spain 4 Tanzania 5 Germany 6 Russia

The rank of countries under our evaluation system shown in Figure 6

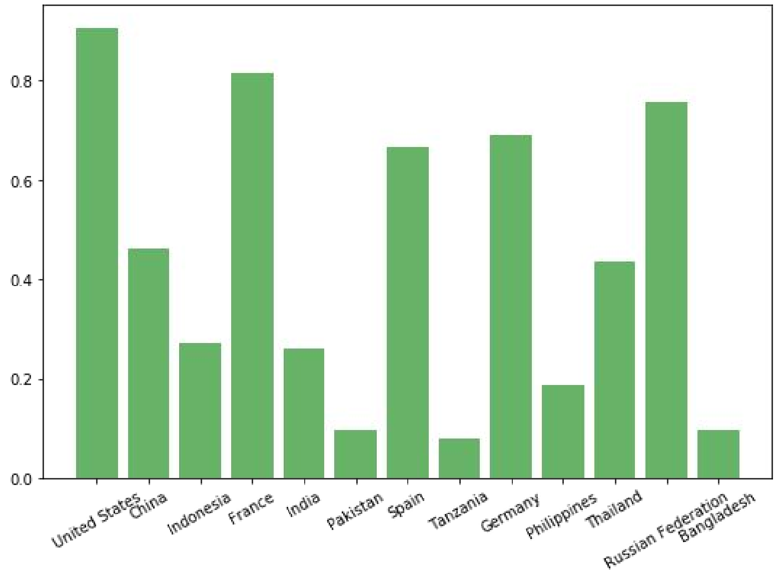

Figure 6 Rank of Countries Under Evaluation System

5.5 Problem B in Part II

Based on the result of the Population Growth-Markov Model 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 prosperity of countries would be unstable. So we suggest open less than 6 offices.

6 Sensitivity Analysis[4]

6.1 The reliability verification of Population growth-Markov Model

The Regression Square Sum: $S_{reg} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$

The Residual Square Sum: $S_{res} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

Where n denote the sample size, m denote the number of countries.

The total square sum (SST) can be divided into parts of the regression square sum (SSR) and the residual square sum (SSE)

MSR=SSR/(m-1)

MSE=SSE/(n-m-1)

$$F = \frac{MSR}{MSE} = \frac{SSR/(m-1)}{SSE/(n-m-1)}$$

If $F>F_{0.05}$ (m-1,n-m-1), then the regression equation is significant under the significance level $\alpha=0.05$

6.2 The sensitivity analysis of Fuzzy Mathematical Model

First calculate the Consistency Index CI:

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

Then find the mean Random Consistency Index RI(table 7)

Table 7 Mean Random Consistency Index of Each Scale

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

Construct 500 sample matrix using stochastic methods: choose the numbers of our scales randomly and construct Positive Reciprocal Judgment Matrix, then calculate the average of the largest eigenvalue $\overline{\lambda}_{\max}$.

Define:

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

Finally calculate the consistency ratio CR:

$$CR = \frac{CI}{RI} = 0.0877$$

(When CR<0.10, it holds that the consistency of the judgment matrix is acceptable.)

7 Strengths and Weaknesses

7.1 Strengths

- Our data comes from the official website such as wikipedia, Ethnologue, World Bank.

- Our models are robust while the parameters changes. That is to say, a slight change of parameter will not cause a significant change of the results.

- Our models can be well applied in other places and we just need to change a little the specific conditions.

- Based on different situation, we develop the corresponding modulating strategy and make a relative comprehensive plan. For a example, we discuss the impact factors of native language and second language respectively.

7.2 Weaknesses

- Our model didn't consider the impact of policy, increased global tourism, the use of electronic communication and social media.

- The data of some parameters is forecast data because we do not have real practical experience.

8 Conclusion

Through the establishment of two models, we imitate the trends of languages over time and the changes of geographic distributions of 13 languages. We consider several factors that could effect number of the total language speakers, but finally conclude that the main factors are economy and culture. Based on the calculation of the Fuzzy Mathematical Model, we can locate the new offices in the long term and in the short term.

Short term: open new offices in countries whose native language, GDP per capita and CSP ranks well. By this means, it is more convenient to introduce qualified personnel and more client to serve, in the meanwhile, the profit will be larger. The cities we choose are New York, Barcelona, Jakarta, Moscow, Hamburg, Dar Es Slaam(capital of Tanzania).

Long term: mainly discuss the developing countries. Despite the fact that the GDP of these countries are not rank well, with the rapid development of their economy and population, these developing countries shows potential. So locate offices in the developing countries in the long term can get more profit. For example, India and China.

At last, we suggest that the number of new opening offices should be **less than 6**. Based on our model and considering the rapid development of communications techniques and internet, open 6 offices would not profit well.

9 The Evaluation of The Model

Establish a Hierarchical Model

- **Cost Index:**

Direct: geographic factors(the rent of offices in prosperous area is higher)
tax (taxes differ from country to country)
production costs
property charges
Potential: staff benefits
bad account
inventory losses

- **Benefit Index:**

Economic effect: capital accumulation & industrial agglomeration
Social effect: policy
customer base
Environmental effect: transportation
Density of population

We should pay more attention on the economic factors, to establish a hierarchical model to insure the profits of the new opening offices.

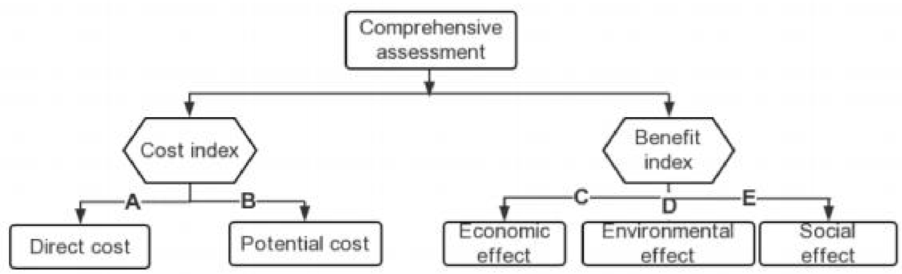

Figure 7 Hierarchical figure

Reference

[1] http://www.shihang.org/

[2] XueTing Wang , Xia Han, & Ma JianQiang Ma . (2015). Quantitative Assessment of National Soft Power. Statistics and Decision (13), 131-135.

[3] https://en.wikipedia.org/wiki/Markov_chain

[4] ShouKui Shi, XiJing Sun , & DeCun Zhang . (2015) Mathematical modeling algorithms and application of exercises to answer. National Defense Industry Press.

[5] XiaoXin Han(2009). Research on the Contribution Rates of Three Industries in China Based on Markov Chain. Cooperative Economy & Science (15), 24-25.

Appendix

1. The result of P

```
[[ 2.06909007e+00 6.64989288e-01 6.00856564e-01 5.27330804e-01
4.14059792e-01 3.22811732e-01 3.13534695e-01 2.82438339e-01
1.80807006e-01 1.54714299e-01 1.37145889e-01 2.28839725e-01
6.59219219e-02]
[ -1.92506049e+00 -6.18951593e-01 -5.58660689e-01 -4.89376670e-01
-3.84606612e-01 -3.00160247e-01 -2.91494053e-01 -2.62736977e-01
-1.68206042e-01 -1.43656834e-01 -1.27698319e-01 -8.94297927e-02
-6.12942048e-02]
[ -2.78817519e+00 -8.92809061e-01 -8.06671691e-01 -7.07533419e-01
-5.56471293e-01 -4.33259525e-01 -4.20789239e-01 -3.79128645e-01
-2.42710038e-01 -2.07552532e-01 -1.84153454e-01 -3.54703204e-01
-8.84793214e-02]
[ -8.09578932e-01 -2.53763605e-01 -2.27950384e-01 -1.97195779e-01
-1.57518189e-01 -1.22237123e-01 -1.18598504e-01 -1.07316342e-01
-6.87350597e-02 -5.79539112e-02 -5.24833785e-02 7.43742127e-02
-2.49636191e-02]
[ -2.59630840e+00 -8.38141323e-01 -7.58746853e-01 -6.68606322e-01
-5.22980070e-01 -4.07761716e-01 -3.96161904e-01 -3.56417673e-01
-2.28133008e-01 -1.96023027e-01 -1.72717090e-01 -5.14363531e-01
-8.32709152e-02]
[ 1.99970205e+00 6.36929480e-01 5.73414454e-01 4.99272723e-01
3.95048695e-01 3.07875267e-01 2.98853338e-01 2.69878431e-01
1.72814992e-01 1.46682093e-01 1.31563511e-01 -1.02996443e-01
6.28702679e-02]
[ 5.96305630e-01 2.10482803e-01 1.94838645e-01 1.80537496e-01
1.33396802e-01 1.05340378e-01 1.02732004e-01 9.09424212e-02
5.81016906e-02 5.25881750e-02 4.29172618e-02 6.91878884e-01
2.15099175e-02]
[ 1.75020103e+00 5.63732624e-01 5.08769012e-01 4.45704633e-01
3.50047976e-01 2.73393731e-01 2.65501914e-01 2.39303044e-01
1.53203037e-01 1.30854276e-01 1.16304028e-01 5.46842244e-02
5.58273500e-02]
[ -2.02511112e+00 -6.64942478e-01 -6.06814572e-01 -5.43697621e-01
-4.18901850e-01 -3.26472303e-01 -3.17577774e-01 -2.84216986e-01
-1.81810095e-01 -1.58915358e-01 -1.36563177e-01 -1.19493972e+00
-6.66757015e-02]
[ 3.89125400e+00 1.25106526e+00 1.12967889e+00 9.90330680e-01
7.77995449e-01 6.06950691e-01 5.89459650e-01 5.31181818e-01
3.40057194e-01 2.90651291e-01 2.58073687e-01 2.77137858e-01
```

```
1.23943826e-01]
[ 1.92127939e+00 5.98176595e-01 5.41796454e-01 4.75379943e-01
3.77600353e-01 2.90314819e-01 2.81963287e-01 2.54030271e-01
1.62623502e-01 1.39097340e-01 1.23376458e-01 7.84914572e-01
5.93059037e-02]
[ -3.64085655e-02 -3.45770594e-02 -2.95858567e-02 -2.59373625e-02
-1.53065959e-02 -1.68050030e-02 -1.63246985e-02 -1.46955227e-02
-9.40681823e-03 -8.06740236e-03 -7.12799075e-03 6.95257709e-01
-3.40748947e-03]
[ 6.19515954e+00 1.97698087e+00 1.77741009e+00 1.54415376e+00
1.22250579e+00 9.54499967e-01 9.26380310e-01 8.37134568e-01
5.36095337e-01 4.54007589e-01 4.08537668e-01 -8.66044152e-01
1.94905006e-01]]
```

2. The number of Calculate the number of native speakers

````
```python
import numpy as np
import pandas as pd
from sklearn import linear_model
import matplotlib.pyplot as plt

file = open("E:/Zeng Siwei/native.csv")
data = pd.read_csv(file)
n_features = data["Country"].count()
n_year = 91

print(data)

x_native = np.linspace(1980, 2070,n_year)#Linear interpolation
y_native = np.zeros((n_year))
coef_xm = np.array(data["xm"])
coef_r = np.array(data["r"])
x0 = np.array(data["x0"])

plt.figure(figsize=(9,6))
for j in range(n_features):
    for i in range(n_year):
        y_native[i] = coef_xm[j]/(1+(coef_xm[j] / x0[j] -1)*np.exp(-coef_r[j]*
(x_native[i]-1980)))
#       print(y_native)
    plt.plot(x_native, y_native, label = data["Country"][j])
plt.legend(loc = 2)
````
````

```
plt.show()
```

```
                Country        xm         r         x0
0         United States   50.0000  0.027982  22.722500
1                 China  180.9516  0.017700  98.123500
2             Indonesia   40.0000  0.011370  14.749036
3                France   10.0000  0.005040   5.534078
4                 India  400.0000  0.012990  69.678352
5              Pakistan   40.0000  0.018600   7.806814
6                 Spain   10.0000  0.002480   3.749116
7              Tanzania   25.0000  0.030290   1.868316
8               Germany   15.0000  0.001910   7.828858
9           Philippines   60.0000  0.016570   4.739697
10             Thailand   60.0000  0.004670   4.738532
11   Russian Federation   20.0000  0.001200  13.901000
12           Bangladesh   50.0000  0.011340   8.147086

![png](output_1_1.png)
```

3.Calculation of probability transfer matrix

```python
# for j in range(n_features):
#     clf.fit(x_data, y_data[j])
#     p[j] = clf.coef_
# print(p)

# # Transposition operation
# for j in range(n_features):
#     for i in range(n_samples):
#         y_train[j][i] = y_data[j][i] - x_data[i][j]
#         for k in range(n_features):
#             if k != j:
#                 x_train[i][k] = x_data[i][k] - x_data[i][j]
#             else:
#                 x_train[i][k] = 0
# #     print(x_train)
# #     print("==============================")
#     clf.fit(x_train, y_train[j])
```

````
#         p[j] = clf.coef_
#         p[j][j] = 1 - np.sum(p[j])
```



```python
Y1 = np.zeros((n_samples, n_features))
Y2 = np.zeros((n_samples, n_features))
for i in range(n_samples):
    Y2[i] = np.array(data[string[i+1]])
    Y1[i] = np.array(data[string[i]])
# print(Y1)
p = np.dot(np.dot(np.linalg.inv(np.dot(Y1.T, Y1)), Y1.T), Y2)

```



```python
p = p / 10
# print(p)
st = np.array(data["2014"])
print(np.dot(st, p))

```

    [   6.00864757e+08   1.93476979e+08   1.74723784e+08
1.53235287e+08
   1.20286905e+08   9.38862413e+07   9.11834787e+07
8.21576276e+07
   5.25956878e+07   4.49736837e+07   3.99077049e+07
4.18397753e+07
   1.91720754e+07]
````

4.Calculation of fuzzy mathematical model

````
```python
import numpy as np
import pandas as pd
from sklearn import linear_model
import matplotlib.pyplot as plt

file = open("E:/Zeng Siwei.refine.csv")
````

```
data = pd.read_csv(file)
n_samples = data["Country"].count()
n_features = 4
string = ["GROWTH", "AVG", "CSP", "FF"]
print(data)
```

FF	Country	GROWTH	AVG	CSP
0	United States	0.205491	1.000000	1.000000	1.000000
1	China	0.197583	0.681648	0.165207	0.103041
2	Indonesia	0.000000	0.442457	0.066413	0.004679
3	France	1.000000	1.000000	0.351892	0.543759
4	India	0.222710	0.192865	0.426896	0.343396
5	Pakistan	0.065537	0.142689	0.023294	0.001936
6	Spain	0.000000	1.000000	0.258549	0.384276
7	Tanzania	0.000000	0.098813	0.089542	0.000000
8	Germany	0.113426	1.000000	0.262019	0.511654
9	Philippines	0.025128	0.302386	0.043119	0.003493
10	Thailand	0.067168	0.696973	0.096811	0.041943
11	Russian Federation	1.000000	1.000000	0.150339	0.394887
12	Bangladesh	0.187528	0.109653	0.049067	0.000660

```python
r = np.zeros((n_features, n_samples))
for i in range(n_features):
    r[i] = data[string[i]]
print(r)
```

```
[[ 2.05490712e-01 1.97582679e-01 0.00000000e+00
1.00000000e+00
 2.22710471e-01 6.55368340e-02 0.00000000e+00
0.00000000e+00
 1.13426026e-01 2.51284410e-02 6.71680260e-02
1.00000000e+00
 1.87528009e-01]
 [ 1.00000000e+00 6.81647507e-01 4.42456969e-01
1.00000000e+00
 1.92864576e-01 1.42689165e-01 1.00000000e+00
9.88131510e-02
 1.00000000e+00 3.02385855e-01 6.96972848e-01
1.00000000e+00
```

```
            1.09653062e-01]
        [           1.00000000e+00              1.65207335e-01              6.64133490e-02
3.51891624e-01
            4.26895754e-01     2.32942340e-02     2.58549480e-01     8.95423760e-02
            2.62018834e-01     4.31191140e-02     9.68114980e-02     1.50338675e-01
            4.90665790e-02]
        [           1.00000000e+00              1.03041000e-01              4.67856000e-03
5.43758760e-01
            3.43396000e-01                 1.93600000e-03                 3.84276010e-01
0.00000000e+00
            5.11654090e-01     3.49281000e-03     4.19430400e-02     3.94886560e-01
            6.60490000e-04]]
```