For office use only
T1 _____
T2 _____
T3 _____
T4 _____

Team Control Number
74316

Problem Chosen
B

For office use only
F1 _____
F2 _____
F3 _____
F4 _____

2018
MCM/ICM
Summary Sheet

The purpose of this paper is to analyze the time and geographical distributions of various languages and forecast the development and distribution of the next 50 years, providing the theoretical basis and advice for the company's offices locating decision.

Firstly, we analyze the quantitative temporal distribution of language users. Various kinds of influences and factors that affect the number of languages users are considered, and concluded as ten indicators, such as GDP per capita, average years of schooling. By Principal Component Analysis, they are combined into four primary components: level of economic development, level of social equality, level of national welfare, and cultural exchanges. On this basis, the short-term difference models are established for native speakers and non-native speakers. First-order autoregressive model (AR(1)) is used to fit the time distribution of native speakers in order to reflect the autocorrelation characteristics. Most native speakers are consistent with non-stationary unit root process. Then, we construct the co-integration relationship between the principal components and the second language users. The error correction model is established and it is found that the random error and the error correction term all achieved stability. In the co-integration space, the influence of the principal component on L2 has a first-order differential stationary nature.

Based on the short-term model, the long-term differential model is further established. Considering the change process of native speakers as a logistic model similar to the natural population growth, the system is stable for the coefficients in the normal range, and the stable equilibrium is given maximum capacity under the current conditions. The time distribution largely synchronizes with the natural change process of the population. Besides, due to the differential smoothness of the linear combinations of the various factors, the influence on second-language users is regarded as a constant for a long time. Therefore, the model of L2 is a constant coefficient differential equation whose time path is determined by the strength of language influence. Therefore, we sum up L1 and L2 to calculate the total number of language speakers, which is a non-stationary dynamic system. It means that the driving forces of a particular language are the endogenous growth of native speakers and the external influence as a second language. The time distribution of the total number of languages is on the rise.

Secondly, we use the long-term model to predict the situation of each country in the next 50 years. The number of influential language speakers increase significantly, while the growth pattern is driven by the second language transmission. The number of less influential language speakers grow less obviously or even decreased. The growth pattern is endogenous to the native speakers. In Top 10, there is a possibility that the number of native speakers will drop significantly or the number of non-native speakers may not grow enough, thus the future rankings may be superseded. Sensitivity analysis and Monte Carlo robustness simulations show that our model is robust and predictable.

Thirdly, we build a Markov Model to analyze the geographical distribution of languages and their changes. This paper constructs a transition matrix of immigrants. Based on the information of population growth, natural growth rate and language distribution, the distribution of the total number of each language in each country is inferred. Then we center on and visualize the national capitals. The prediction shows the geographical distribution of languages tends to be intertwined and spread as second-language in the future.

Next, we locate the new offices by Cluster Analysis and think that the ability of speaking English and Chinese is needed and that the development of economy is considered as well. So, we construct the 4 indicators, the ratio of English speakers and Chinese speakers, GDP per capita and net immigrants. The short-term and long-term models and the Markov model of the forecast results calculate the value. According to the 4 indicators, we analyze the 224 countries by cluster analysis separately. Due to the quantified result of grades of each country, we use the Multiple Objective Decision Making (MODM) of Fuzzy Evaluation to calculate the grade and choose 6 national capitals with maximum grades, as the location of new offices.

Finally, Using MINE model, the latitude and longitude coordinate grids are meshed. And the grid density index is calculated to identify the areas with over-dense distribution of offices. Eliminating appropriate number of offices allowed for reduction costs by serving the largest global scale with minimal office locations. The result shows that it is more suitable to set up 4 new offices. Thus, as for the short term, we recommend to build 4 new offices in London, Singapore, Ottawa and Canberra. And Singapore, London, Canberra and Paris are recommended in the long run.

Key Words: Autoregressive Model, Co-integration, Differential System, Markov Chain, Clus[illegible]INE Model