| For office use only | Team Control Number | For office use only |
|---|---|---|
| T1 ____________ | **00000** | F1 ____________ |
| T2 ____________ | | F2 ____________ |
| T3 ____________ | Problem Chosen | F3 ____________ |
| T4 ____________ | **E** | F4 ____________ |

**2018**
**MCM/ICM**
**Summary Sheet**

# The measurement and prediction of water scarcity degree based on decision tree model


**summary**

We establish an measurement model aiming at assessing water scarcity situation and analyzing the water shortage causes. Meanwhile, combined a prediction method of influential factors, the evaluation model also can be utilized to forecast the future water scarcity condition of a region or nation. From the above, we design intervention plans of water treatment with comprehensive analysis.

We determine water scarcity degree(WSD) to evaluate the ability of a region to provide clean water to meet the needs of its population. And we establish a comprehensive measurement indexes system to evaluate WSD, which includes 16 influential attributes. Estimation of WSD can be regarded as a classification process, so we build the Decision Tree Model based on K-means-Clustering with ID3 algorithm. The method of K-means-Clustering is to classify each attribute into several different levels, and then produce the judgment results of WSD with Decision Tree. In the calculation process, through comparing the information gain of each attribute, we obtain six indicators, which exert significant impact on WSD. They are also the crucial reasons of serious water shortage.

Before using Decision Tree Model to predict the water shortage situation in 15 years, we are supposed to forecast the values of 16 attributes in ahead by Grey Prediction Model. Then, take these predicted values into Decision Tree and we can achieve the goal of forecasting. Take the India as an example, the future water condition is not optimistic at all.

How to solve water crisis in proper ways? We believe that the essential criterion of the intervention plan is acting appropriately to the situation. Therefore, we analyze the leading causes in accordance with the reasons listed in the above with. In addition, we also consider some comprehensive factors which don't represent in 16 attributes, making the plan more rational and integrated. As a result, developing the water-saving agriculture and waste water treatment technology are quite good suggestions. As for the influence of surrounding areas and the nation, we mainly consider it from the point of economy and ecology.



**Keywords**: Water Scarcity Degree ; K-means Clustering ; Decision Tree ; Gery Prediction

# Contents

# 1 Introduction

## 1.1 Background

The problem of water resources has increasingly attracted wide attention all over the world. With population expansion, industrial development, urbanization, intensive agriculture and improvement of people's life, the contradiction between water supply and demand is becoming more and more serious. [1]A report, published by the World Council for Environment and Development (WCED) in 1998, indicates that: âĂIJWater resources is replacing oil as the main problem in the global crisis.âĂİ How to provide the clean water needed for a growing population has been a global challenge.

However, the situation of worldwide water utilization and management is not optimistic. Due to the fact that water quality comes down every year in most parts of the world, and serious damage to forests and improper use of land, natural disasters related to water are on the rise. Now, two fifths of the population in the world live in areas which are prone to flooding, such as Bangladesh, China, India, the Netherlands. It is reported that there are 1.1 billion people still lack access to safe drinking water and 2.6 billion people lack basic sanitation in the world. [2]In addition, water management still has many problems. In many parts of the world, 30 to 40 percent of water resources are wasted or stolen due to pipeline and canal leaks and illegal connection of water pipes.

Although the water scarcity problem is related to natural geography factors, the more essential reasons should depend on our ability of effective water management and utilization. Now various strategies have been executed by most countries to deal with water scarcity problem. For example, Israel has congenitally severe water shortage because of the influence of geographical location and the weather. Nevertheless, Israel vigorously developed seawater desalination technology for many years and gradually established a successful water treatment system, which even has become a kind of export trade and promotes domestic economic development.

## 1.2 Analysis Of The Problems

- **Towards Task 1,2**

  For better illustration, we define a new indicator as Water Scarcity Degree(WSD) to measure the ability of a region to provide clean water to meet the needs of its population. It is necessary to establish a comprehensive measurement system to reveal the factors which have essential influence on the WSD. Thus, the problem of measuring the ability—qualified water resources satisfies the demand of all resident population—transforms into the problem of grading the level of WSD. There are many adoptable evaluation methods, like Analytic Hierarchy Process(AHP) and Fuzzy Comprehensive Evaluation(FCE). However, to be on the accuracy side, we consider to construct the WSD classifier based on Decision Tree and ID3 algorithm. In addition, seeking the major reasons which leads to present water shortage situation can be available in our model, which depends on comparing the entropy of every attributes.

- **Towards Task 3**

  Predicting the future situation of water resources can also utilize the Decision Tree Model, but the premise of this method is to determine the future value of attributes in 15 years and then make the prediction with Decision Tree. In order to solve this problem, we employ the method of Grey Prediction Theory to generate the data of indicators. Because Grey Prediction Theory is an effective method to analyze and forecast the future development trend based on existing data

- **Towards Task 4,5**

  Intervention plan should acts appropriately to the situation. The crucial step is to find the leading cause. From the above analysis, we can conclude the major causes from making the comparison of attributes' entropy. Then, we should design a series proper strategies around the encountered problems and especially focus on feasibility and effectiveness.

# 2 Notation

| Symbol | Description |
|---|---|
| $V$ | Sample set |
| $x$ | Impact factor $x$ on water scarcity degree |
| $Gain(V, x)$ | Information Gain of attribute $x$ of sample set $V$ |
| $Env(V)$ | Information Entropy of sample set $V$ |
| $\|V\|$ | The number of elements in sample set V |
| $k$ | Water scarcity degree |
| $p_k$ | The proportion of the number of $k$ degree water scarcity in sample set $V$ |
| $x_n$ | The value of impact factor $x$ in $n_{th}$ year |

Table 1: Symbol Description I

# 3 Water Scarcity Measurement Model Based on Decision Tree

## 3.1 Assumptions And Justifications

- We just select 16 representative dynamic indicators to explore how theses indexes influence on the ability of a region to provide clean water to meet the needs of its population. Water scarcity is a complex interdisciplinary issue. It is involved in various disciplines and fields like geology, economics, technology, culture, ecology, climatology and so on. It is impossible to give full consideration to every possible relevant indicator in our model.

- We don't consider the influence of the factors about history, religion, territory conflict and politic difference between different counties and regions. They are hard to quantify in the measurement model and difficult to change with artificial intervention.

## 3.2 Establish The Comprehensive Evaluation Index System

Water scarcity degree(WSD) will influenced by geology, society, economy, ecology, environment, and many other factors. For the convenience of grading and predicting the water scarcity degree, we artificially select 16 characteristic indexes from a large number of indexes published by Food and Agriculture Organization of the United States. According to the generating reasons and consumption reasons of water resources, we categorize 20 indexes into following four classes: Water Resources Storage, Social Development, Agricultural and Industrial Development , Waste water and Sanitation. The specific indexes system is showned in the Appendices.

## 3.3 Decision Tree Model

Decision tree is a tree-like structure in which each internal node represents a "test" on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules.

The construction process of decision tree is completed by studying the training data. The training set includes the data of 16 attributes (shown in the Figure 11) of 23 countries from 1970 to 2015. Every attribute has dynamic changes with the time flow and we select the statistics in 2010 to train the decision tree.

### 3.3.1 Kmeans-Cluster Based classification

In order to make the data able to apply to the decision tree model, we need to cluster the data and classify the data according to the result of the clustering. In the following rank dividing, we will use the K-means method to cluster, but the conventional clustering method requires us to enter the number of clusters. If we enter the same number of clusters for each indicators, data may be over fit. Due to too many indicators, the number of clusters that are artificially added to each indicator can be inevitably subjective. And it can lead to a heavy workload.

To solve this problem, we use a method based on information gain clustering. That is, the information gain corresponding to the number of each cluster is enumerated, the gains of these information are sorted, and the number of clusters corresponding to the minimum information gain is selected. Dividing the ratings of each attribute by this method can maximize the impact of each attribute on the final result and improve the accuracy of the decision tree decision.

$$Ent = Gain(V, x) \tag{1}$$

According to the equation 1 ,we can find the division of each level indicator

### 3.3.2 Decision Tree Based On Evaluation System

We have obtained a good level of data sample set $V$ by K-means Clustering method. According to different values of attribute $a$, we can obtain the subset of $V$, which contains

$V_1$,$V_2$,$V_3$,...,$V_n$. And then, the structure of the decision tree can be determined by calculating the information gain of attribute a. (We will discuss the calculation of the information gain in the below ) The information gain represents the purity of a node , which is the purity of the decision result for all sample sets of the node. The higher the purity of attribute is, the greater impact of this attribute has on the results of decision-making. Now we will discuss the definition of the Information Gain:

$$Gain(V, a) = Ent(V) + \sum_{v=1}^{m} (\frac{|V_v|}{|V|} Ent(V_v)) \tag{2}$$

$Gain(V, a)$ is the Information Gain of attribute $a$ . $V_v(v = 1, 2, 3, m)$ represent the subset of $V$ which obtained by dividing the sample set according the value of attribute $a$ . $|V|$ and $|V_v|$ respectively represent the number of samples in $V$ and $V_v$ . $Ent(V)$ is the Information Entropy of the sample set $V$ , which is defined as follows :

he decision tree decision:

$$Ent(V) = \sum_{k=1}^{|y|} (p_k log_2 p_k) \tag{3}$$

Pk represents the proportion of different decision-making results in sample set $V$ .

We can determine the maximal effect factors by comparing the Information Gains of different attributes . Then the set $V$ can be divided multiple sets $V_v(v = 1, 2, 3, m)$ by the different value of attributes . Throwing interating , we can get the decision tree .

## 3.4 Results

### 3.4.1 The Analysis Of K-means Clustering

Through the ranking of information gain, we select the sewage treatment capacity, the amount of sewage treatment, long-term rainfall, surface water entering in the country , the four indicators of India's actual water shortage compared to the situation. The purpose is to verify the clustering algorithm.

From Figure 1, we can clearly see that the long-term annual rainfall in India is moderate in our sample set. However, the inflow of surface water in India is the highest in the sample set and the inflow of surface water to the result is more influential and integrated, India is at the forefront of total external renewable water resources. Compared with other countries, India has greater surface water inflows, and not low long-term annual rainfall, which is consistent with the fact.

It can be clearly seen from Figure 2 that we classify three levels according to the sewage treatment capacity of each country by means of clustering. India is divided into "o", so India's sewage treatment capacity in these 24 countries in the medium. In similar countries, the Euclidean distance between India and the third type of country is the most recent. In addition, there are only a third type of countries with only a small proportion of them. Therefore, India's capacity for sewage treatment is relatively weak.

In summary, we can use Kmeans clustering method to classify the corresponding indicators into several different levels, which provides better data for the decision tree

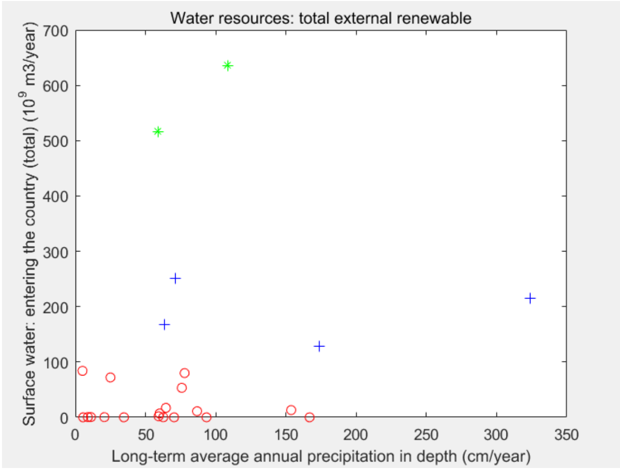


Figure 1: Water resource:total external renewable.The India location is (108.3,635.2).

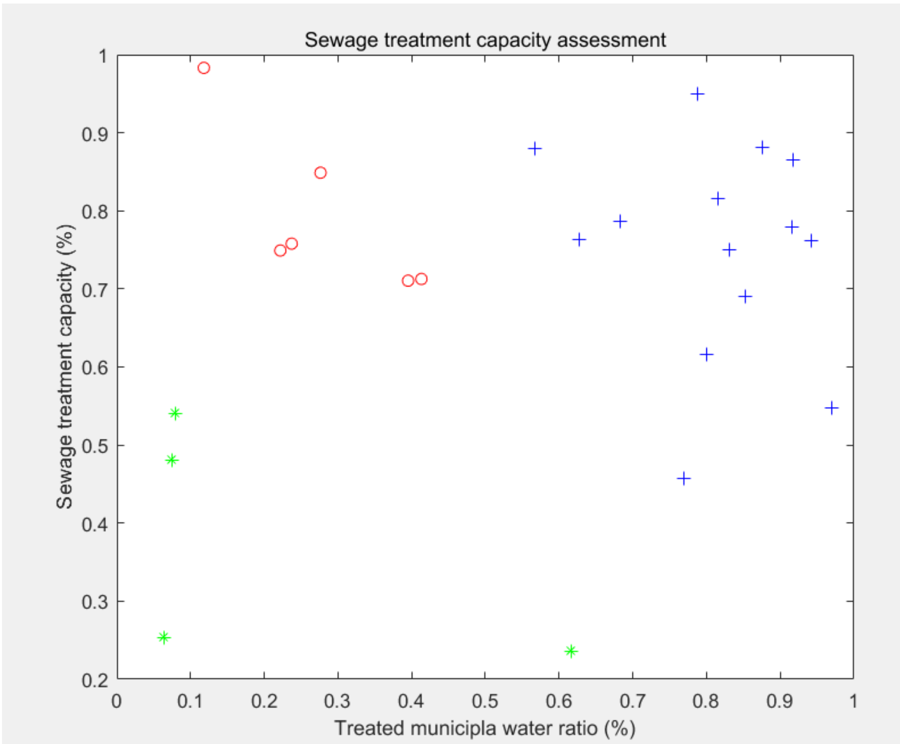


Figure 2: Sewage treatment capacity assessment system. * stands for weak ability to deal with sewage, o stands for medium ability to deal with sewage, + stands for strong ability to deal with sewage.

assessment and prediction model pre-treatment. 3.4.2 The analysis of decision tree

### 3.4.2 The Analysis Of Decision Tree

We got the decision-making path of India's water scarcity degree and result by bringing the rankings of India's indicators into the decision tree:

From figure 3, it can be seen that the primary reason for India's water shortage is that

India has a very low population with access to clean water.Secondly India has abundant external renewable water resources, which is good.Thirdly India has a particularly high proportion of agricultural water use.Fourth , India has alow urbanization rate.Lastly, the Treated municipal waste water ratio is very low. Drawn from the above five indicators of India,we can get the result that India belongs to Economical severe water scarcity.

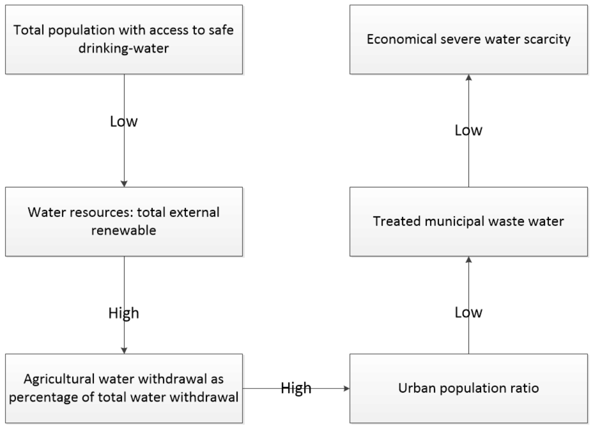


Figure 3: The result of Decision Tree.

## 3.5 Evaluation Of The Model

### 3.5.1 Strengths

- Compared with other evaluation model, the decision tree model can reflect the relationship between attributes through a tree structure. When analyzing the water scarcity problem which affected by multiple factors, it can directly obtain the reasons of water shortage and the relationship between these causes through the decision tree model.
- Decision tree model is not sensitive to missing value. The lack of data will not affect the decision tree for the results of judgment.
- Time complexity of the algorithm is low, so the decision tree needs only one time building and can be used repeatedly. Every prediction does not exceed the maximum calculation times of the depth of the decision tree.
- In the K-means clustering process, we do not need to give the number of clustering. It reduces the workload largely and avoids the interference of subjective factor.

### 3.5.2 Weaknesses

- We don't take account into the influence of political factors, such as government policy about management of water resources and the war.
- Data sample set we obtained is not big enough, which causes that some attributes are not used in the process. That is to say, the analysis of water shortage data is prone to misfit.

- K-means clustering will make results unstable. If we classify the same level for many times, it may keep the result of the decision tree variable. But the difference of each result is not too large, because the offset of every clustering center is not too big.

# 4 Water Situation In India

## 4.1 Current Situation Of Water Resources

According to the latest communique from the Indian Central Water Commission, in 2016, the water level of 91 major reservoirs in India was at its lowest in 10 years.

The domestic average water level in India is 71 percent of that in 2015. According to a study from McKinsey Company, India's total water supply is expected to fall by 50 percent by 2030, and nearly 100 million people will suffer from a severe shortage of drinking water. According to the latest data from the website âĂIJWaterâĂİ, 770 million people in India lack the access to safe drinking water and 769 million people are sort of sanitation equipment.[3]

Currently, India is one of the countries who face the most severe water crisis, the specific performances are the serious shortage of water resources, extremely uneven distribution, low utilization efficiency and severe water pollution, etc. [4]Although the Indian government has adopted a number of measures, the results are less significant. So far the future development prospects are still not optimistic.

## 4.2 The Reasons Why India Has Serious Water Scarcity

Based on comparing the information gain of attributes in the decision tree, we can select the greater one to be the attributes to divide the data set. And then repeat similar operations until all attributes have been traversed. And the higher information gain an attribute has, the more significant the attribute is to water scarcity degree(WSD). So we compare the value of information gain of 16 indicators in Figure 4.

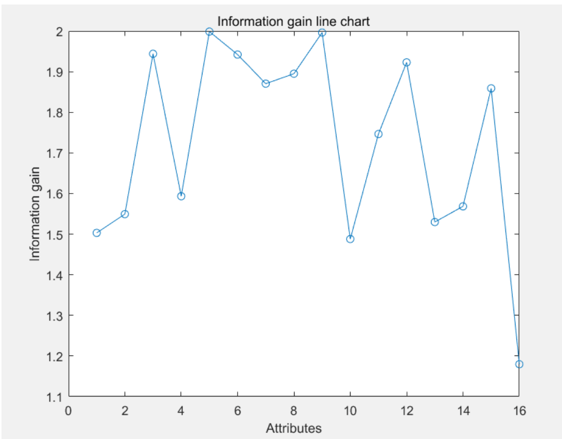


Figure 4: Information gain of 16 indicators.

Combined with the actual meaning of factors, we select six high-information-gain factors which mainly lead to the water shortage of India, including total population with access to safe drinking-water (JMP), agricultural water withdrawal as percentage of total water withdrawal, urban population ratio, treated municipal waste water, GDP per capita, industrial water withdrawal as percentage of total water withdrawal. Then we will analyze the reasons why India has water scarcity from two perspectives: One is from the point of the six factors, the other is from actual comprehensive factors, which donâĂŹt reveal apparently in our models.

### 4.2.1 Six high-information-gain factors

- **Severe water contamination**

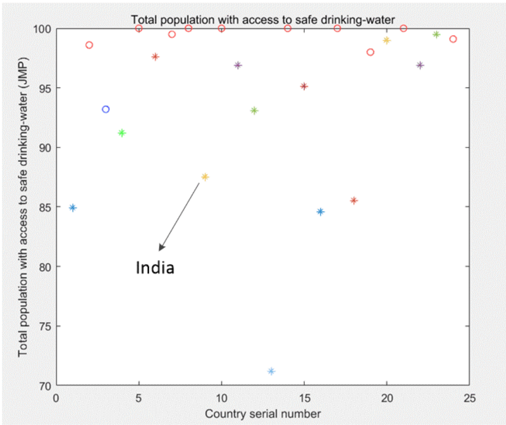


Figure 5: Total population with access to safe drinking-water

From Figure 5, total population with access to safe drinking-water (JMP) of India lists fifth from the bottom among 24 counties, which demonstrated that India has serious pollution. In 2010, Indian officials reported that 908 large cities and towns in India were seriously polluted, where the 5.8 billion people live in.

- **High agricultural water withdrawal ratio**

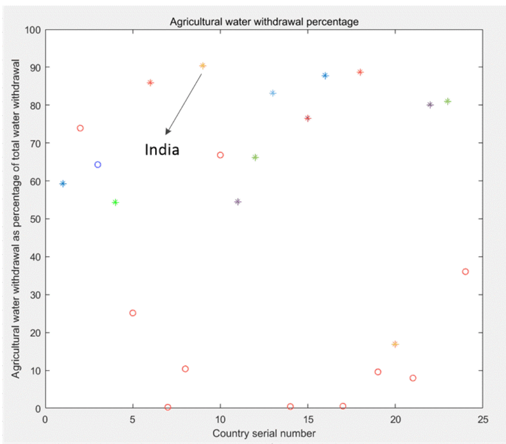


Figure 6: Agricultural water withdrawal percentage

From Figure 6, agricultural water withdrawal as percentage of total water withdrawal is highest among 24 countries, which indicates that agriculture accounts for a large proportion in the national economic structure. India's agriculture is dominated by rice,which needs a large amount of water to irrigate. In addition, irrigation method is backward and the effective utilization rate of irrigation water is extremely low. And India's most extensive and inefficient irrigation method still remains an absolute majority in all irrigation ways.

- **Low industrial water withdrawal ratio**

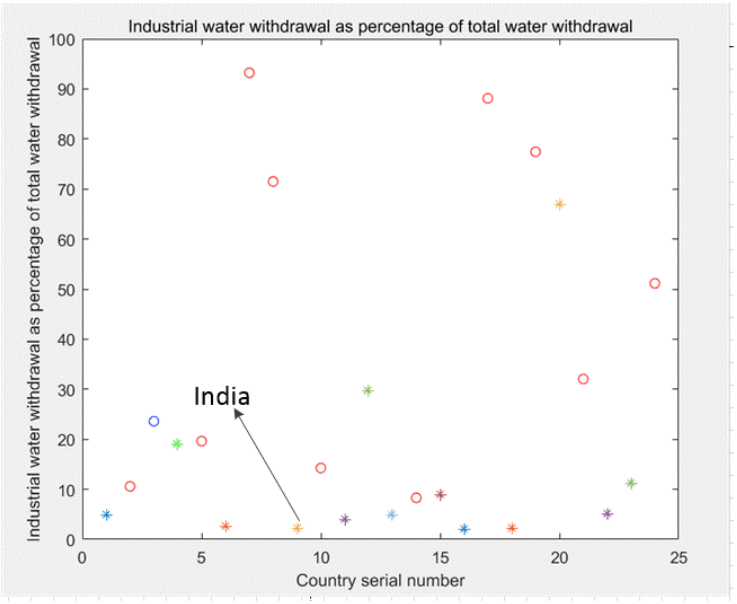


Figure 7: Industrial water withdrawal as percentage of total withdrawal

From Figure 7, industrial water withdrawal as percentage of total water withdrawal is lowest among 24 countries, which indicates that India is not a well-developed industrial country. Through contrast, we find that India develops extremely imbalance between agriculture and industry and the industrial structure is inappropriate. This may also contribute to water scarcity.

- **Low urbanization level**

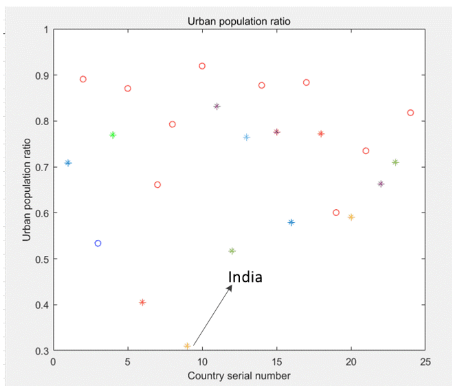


Figure 8: Urban population ratio

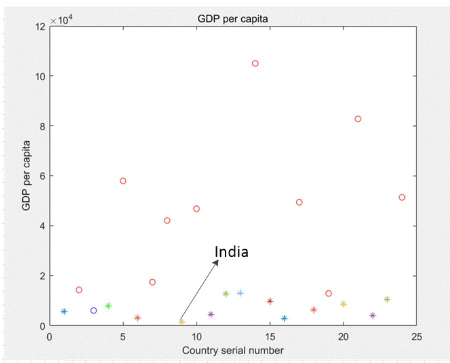


Figure 9: GDP per capita

From Figure 8 and Figure 9, urban population rate and GDP per capacity of India is the lowest among all countries, which reflects that the urbanization rate and economic development degree of India are low. The improvement of urbanization accelerate the industrial upgrading and indirectly reduce the demand of water on agriculture. With the development of economy, the technology of sewage treatment may enhance and the consciousness of saving water gain popularity. The Thus, the higher the economy level is, the less water shortage will be.

- **Weak ability of treating waste water**

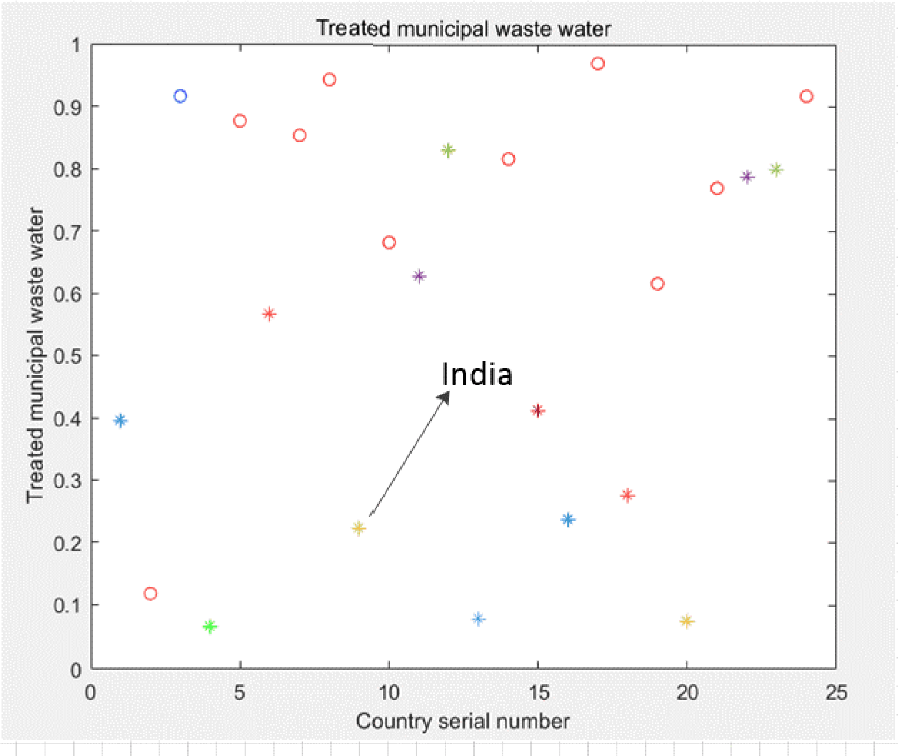


Figure 10: Treated municipal waster water

From Figure 10, the value of treated municipal waste water in India is low, which demonstrates that the ability of treating sewage is not satisfactory. The sewage produced in 2008 reached 38.2 billion liters, but only 11.8 billion liters of sewage were treated, whose quality of treatment was not fully up to standard. It also can prove that the ability of treating sewage of government is relatively low.

### 4.2.2 Comprehensive Factors

- **Population explosion**

  Rapid population growth is an important cause of India's water crisis.India's population is growing and it will reach 1.7 billion by 2050, putting much pressure on India's water supply. said Kskar, a water expert at the university of Delhi in India. The problems brought by population growth are numerous, including the increasing grain demand, deep development of forest, which will inversely weaken the ability of water conservation in the soil.

- **Natural environment**

  The distribution of water resources has seasonal variation characteristics. For example, the Deccan plateau rivers originate in the mainland or the south east and west mountain area, where water supply is limited. It mainly relies on monsoon rainfall, so water yield is unstable: In the drought period, rivers even appear to dry up. In the flood period, it always appears flooding. In recent years, the Deccan plateau rivers have also caused severe drought due to delayed monsoon or less rainfall.

- **Weak water management**

  In order to alleviate water shortages, the government also intensifies drinking water infrastructure construction. But Indian officials corruption is serious, it causes many facilities idle or disrepair. In 2004, The Times of India showed that 98% of people thought politicians and ministers were corrupt. In addition, water price in India is very low. Since neither local nor central political parties want to lose the popular vote, no one is willing to put forward a policy of raising water prices to control the waste water problem. The Hindustan Times reported that new Delhi daily wastes as much as 40 percent of its water due to poor management.

## 4.3 Forecast The Future Based On Decision Tree

### 4.3.1 Grey Prediction Model

- **Description of Model**

  We have constructed the decision tree model to weigh the ability of a region to provide clean water to its population, which also can predict India's water situation in the future. Before using decision tree model, we should predict the value of influential attributes in ahead. Because the indicators of water scarcity degree are of variability,uncertainty and ambiguity, which satisfies the working condition of grey prediction theory, we take this approach to forecast the future water condition of India in 15 years.

  Grey prediction theory is a method to predict the system with uncertain factors. Through identifying development trend of dissimilarity degree between the system factors, namely, correlation analysis, grey prediction can generate strong regularity of data sequences, and then establish the corresponding differential equation model to predict the future development trend of things.

  In order to achieve this goal, we have collected the data of 16 influential factors in India over the past 40 years, analyzed the inherent laws of values and inferred the prediction values in mathematical way. The following is an example of how JMP(Total population with access to safe drinking-water) can be predicted through grey prediction theory.

- **Establishment of Model**

  First, we construct $a$ sequence $x_n$, which is used to record the value of JMP. We will place our predicting outcomes in $x_n$.

  First we define two factors $a$, $b$. $a$ is the development grey number. $b$ is endogenetic control grey number. Both are important indicators of JMP. According to the grey theory, there are the following relations:

$$x_2 + az_2 = b \tag{4}$$

$$x_3 + az_3 = b \tag{5}$$

$$x_n + az_n = b \tag{6}$$

$$z_n = 0.5x_n + 0.5x_n \tag{7}$$

$$x'_n = \sum_{i=1}^{n} x_n \tag{8}$$

$$\frac{dx'_n}{dt} = x_n \tag{9}$$

build the equation:

$$\frac{dx'_2}{dt} + az_2 = b \tag{10}$$

$$\frac{dx'_3}{dt} + az_3 = b \tag{11}$$

$$\frac{dx'_n}{dt} + az_n = b \tag{12}$$

so,we can get:

$$x_n = (x_1 - \frac{a}{b})e^{-a(n-1)} + \frac{a}{b} - (x_1 - \frac{a}{b})e^{-a(n-2)} + \frac{a}{b} \tag{13}$$

Now we have obtained predicted values of all attributes, and they can be categorized according to Table n. Then utilize the constructed decision tree to achieve the water scarcity degree in 15 years in India.

#### 4.3.2 Result

In our grey prediction model, we achieved the mathematical relation between the known indicatorsâĂŹ values and predicted values, so we can conclude all predicted values of 16 attributes in the next 15 years. Considering that there are too many influencing factors, we only display the predicted results of JMP :

| Year | Predict of JMP(%) | Year | Predict of JMP(%) |
|---|---|---|---|
| 2018 | 92.8 | 2026 | 94.6 |
| 2019 | 92.9 | 2027 | 95.2 |
| 2020 | 93.1 | 2028 | 95.8 |
| 2021 | 93.2 | 2029 | 96.4 |
| 2022 | 93.8 | 2030 | 96.7 |
| 2023 | 94.0 | 2031 | 97.2 |
| 2024 | 94.2 | 2032 | 97.8 |
| 2025 | 94.5 | | |

Table 2: Predict of JMP's value

| Year | Grade | Year | Grade |
|---|---|---|---|
| 2018 | B | 2026 | A |
| 2019 | B | 2027 | A |
| 2020 | B | 2028 | A |
| 2021 | B | 2029 | A |
| 2022 | B | 2030 | A |
| 2023 | A | 2031 | A |
| 2024 | A | 2032 | A |
| 2025 | A | | |

Table 3: Predict of JMP's grade

According to the predicted values of various influential factors, we give the classification of JMP in the next 15 years, which is shown in the Table 3:

According to the classification of all influencing factors and the decision tree model, we obtain the final predicted results, which is shown in the Table 4:

| Year | Grade | Year | Grade |
|---|---|---|---|
| 2018 | C2 | 2026 | C2 |
| 2019 | C2 | 2027 | C2 |
| 2020 | C2 | 2028 | C2 |
| 2021 | C2 | 2029 | C2 |
| 2022 | C2 | 2030 | C2 |
| 2023 | C2 | 2031 | C2 |
| 2024 | C2 | 2032 | C2 |
| 2025 | C2 | | |

Table 4: Predict of Water scarcity degree

By predicting each indicator over the next 15 years, we come to the conclusion that: India's total freshwater withdrawal is on the rise; with no change in India's surface water and long-term average annual precipitation in depth; meaning India's remaining renewable water resources will be less and less. In addition, India's produced municipal waste water is growing rapidly and its growth rate far exceeds that of sewage treatment. As a result, water pollution in India will become more serious.

In summary, India's water scarcity will not improve and will be worsen in the next 15 years.

#### 4.3.3 Evaluation Of The Model

In the process of predicting the water situation in India in the next 15 years, we used the grey prediction model and decision tree model based on clustering. We evaluate the grey prediction model under the assumption that decision tree model is correct. The rationality and accuracy of grey prediction can be verified by comparing between predicted value and real value. In addition, considering that there are too many indicators we select, we can only show our inspection process using JMP. The predicted values and real values are shown in the Table 5:

| Year | JMP(%) | Predict of JMP(%) |
|---|---|---|
| 1977 | 64.5 | 64.5 |
| 1982 | 68.6 | 67.6115 |
| 1987 | 70.4 | 70.8426 |
| 1992 | 72.6 | 74.2288 |
| 1997 | 77.6 | 77.7765 |
| 2002 | 82.5 | 81.4936 |
| 2007 | 87.4 | 85.3887 |
| 2012 | 87.5 | 89.4697 |
| 2017 | 94.1 | 93.7459 |

Table 5: JMP and predict of JMP in 1977-2017

It can be seen from the above table that the theoretical values of JMP is in good agreement with the predicted values, indicating that the expected results of prediction model is satisfactory.

# 5 Intervention Plans For India

## 5.1 Plan Statement

According to all causes of IndiaâĂŹs water scarcity problem we mentioned in the above, we put forward a series of intervention plans to solve these issues.

- Strengthen the ability of waste water treatment and improve the reuse rate of water.
- According to the structure of the water resources, reasonably adjust agriculture and industry scale and properly reduce planting area of high water-consumption crops.
- Vigorously develop water-saving agriculture, and constantly popularize sprinkler irrigation and drip irrigation technology to reach maximum utilization of water resources.
- Control population growth
- Increase input in water conservancy projects and strengthen infrastructure construction to prevent flood and drought.
- Construct the effective water resources management system, such as preventing the corruption of government, determining the reasonable water price by central power and improving the ability of water treatment.

## 5.2 Impact On Surrounding Areas

- Waste water treatment will decrease the pollution of interior rivers and enhance the quality of water resources for local population.

- Optimizing the proportion of agricultural and industrial scale of domestic economy promotes upgrading of the industrial structure. This measure contributes to IndiaâĂŹs sustainable economic development in long term.
- Government can construct artificial channels and dams to control water resources better. Not only it greatly enhances the water storage ability of India in the period with abundance precipitation, but also it prevents to suffer extreme drought.
- mproving the technology of seawater desalination will increase the possibility of water supply.
- Adopting a more effective irrigation method will reduce agricultural water consumption to a large extent, because agriculture water withdrawal attaches much importance to total water consumption towards India.

## 5.3 Impact On Water Availability For India

- Improvement of quality of water resources will refresh that of lower branches of rivers—the India Ocean. At the same time, this measure purifies the water resources in the neighbor regions. Ecologic environment will also be advanced through waste water treatment.
- Upgrading industrial accelerates to consume excessive products through exporting them to other counties. Trade is beneficial to both of the exporters and importers. Thus, surroundings nations and regions are possible to be profitable.
- Pollution treatment of water resources will improve the soil structure, the nutritional value of crops and increase the yield of agriculture.

## 5.4 Evaluation Of The Prediction

### 5.4.1 Strengths

- We design intervention plans based on the reasons of IndiaâĂŹs water shortage, which are analyzed with nearly all-sided consideration. So our intervention plan is of strong reliability.
- Through mathematically determining the information gain of attributes, the six indicators we select have relative high correlation with WSD. In this way, our intervention plan can be highly targeted.

### 5.4.2 Weaknesses

- India is a country with strong religious attributes. However, we donâĂŹt take the religion factors into account. Some religious beliefs and customs have bad effect on the water resources in the river. On the one hand, people think the Ganges can wash away evil, so a large number of hindus bathe in the Ganges everyday. On the other hand, many hindus consider that they can go to heaven after death if blending in the Ganges. So thousands of hindus are cremated on the edge of the Ganges after the death, and then directly spread the ashes and residue into the river. In the long run, it is only a matter of time before the river is completely polluted.

# 6 Conclusions

According to Decision Tree Model based on Kmeans-Clustering, we firstly set the level of 16 indicators of 24 countries relatively, which avoids the error appearing in the process of artificial classification and enhance the rationality of measurement. Then we construct the decision tree by studying the training set with ID3 algorithm, and calculate the information gain of 16 attributes to determine most influential indicators. As a result, there are six indexes attach more importance on water scarcity of India, including total population with access to safe drinking-water (JMP), agricultural water withdrawal as percentage of total water withdrawal, urban population ratio, treated municipal waste water, GDP per capita, industrial water withdrawal as percentage of total water withdrawal.

Based on the Grey Prediction Model, we obtain the relation between the known data and predicted values, so we conclude the predicted values of all 16 attributes and the varied trend of each indicator is shown in the 4.3.2. Totally speaking, India's water scarcity will not improve and will be worsen in the next 15 years.

Furthermore, combined some comprehensive factors, we analyzed the major causes which leading to the serious water shortage, such as weak ability of treating waste water, low urbanization level, high agricultural water withdrawal ratio, severe water contamination and bad water management,etc.

Finally, we design our intervention plan from the following six aspects of strengthening the ability of waste water treatment , reasonably adjust agriculture and industry scale and properly reduce planting area of high water-consumption crops, vigorously develop water-saving agriculture, control population growth and increase input in water conservancy projects. The impact on surrounding areas and India itself is demonstrated in economy, ecology and industrial upgrading.

# References

[1] Hou Chunmei, Zhang Zhiqiang, Zhi Sami. The United Nations world water development report calls for strengthening integrated water resources management [J]. Earth science progress,2006(11):1211-1214.

[2] Tide Wheel. The current situation of water resources in the world is extremely difficult to challenge [J]. Ecological economy,2012(05):8-13.

[3] Zeng Xiangyu, Zhu Yufan. Current situation and impact of water security in India [J]. International research reference,2015(01):16-22.

[4] Yuan Keirong. India's multistep response to water shortage [N]. People's Daily,2017-03-31(021)

[5] Yin shi-cong. A brief introduction to the management of water resources development in India [J]. Global science and technology economics, 1987(8):16-17

[6] Chen Xiaofeng, Huang Xianjin. Research on the lack of water resources and coping strategies in India [J]. Technology outlook, 2016, 26(18)

# Appendices

| First-class Indicator | Second-class Indicator | Third-class Indicator |
|---|---|---|
| Water Scarcity Degree | Water Resources Storage Indexes | Long-term average annual precipitation in depth<br>Total internal renewable water resources (IRWR)<br>Surface water: entering the country (total)<br>Water resources: total external renewable<br>Total renewable water resources per capita<br>Total exploitable water resources<br>Dam capacity per capita |
| | Social Development Indexes | Total population<br>Urban population<br>GDP per capita<br>Total freshwater withdrawal<br>Total water withdrawal per capita |
| | Agricultural and Industrial Development Indexes | Agricultural water withdrawal as percentage of total water withdrawal<br>Industrial water withdrawal as percentage of total water withdrawal<br>Cultivated area<br>Total agricultural water managed area |
| | Waste Water and Sanitation Indexes | Produced municipal waste water<br>Treated municipal waste water<br>Capacity of the municipal waste water treatment facilities<br>Total population with access to safe drinking-water (JMP) |

Figure 11: comprehensive measurement indexes system

```
//Kmeans

function result=kmeans(A,c)
[row,column]=size(A);
uc=fix(row/c);
u=zeros(c,column);
cluster=cell(c,1);

for i=1:c
    cluster{i,1}=zeros(1,column);
end

for i=0:c-1
    t=1+i*uc;
    u(i+1,1:column)=A(t,1:column);
end
d=zeros(c,1);
flag=0;
count=0;

while flag~=c
    flag=0;
    for i=1:c
        cluster{i,1}=zeros(1,column);
    end
    for i=1:row
        for j=1:c
            d(j)=norm(A(i,1:column)-u(j,1:column),2);
        end
        m=min(d);
        l=find(d==m);
        cluster{l,1}=[cluster{l,1};A(i,1:column)];
        d=zeros(c,1);
    end
    for i=1:c
        t=size(cluster{i,1});
        t=t(1);
        if u(i,1:column)~=sum(cluster{i,1})/t;
            u(i,1:column)=sum(cluster{i,1})/t;
        else
            flag=flag+1;
        end
    end
    count=count+1;
end
result=u;




//Decision Tree

package decisionTree.myTree;

import java.io.FileWriter;
import java.util.ArrayList;

import java.io.BufferedReader;
import java.io.File;
import java.io.FileReader;
```

```
import java.io.IOException;
import java.util.Iterator;
import java.util.LinkedList;
import java.util.List;
import java.util.regex.Matcher;
import java.util.regex.Pattern;

import org.dom4j.Document;
import org.dom4j.DocumentHelper;
import org.dom4j.Element;
import org.dom4j.io.OutputFormat;
import org.dom4j.io.XMLWriter;

public class ID3 {
    private ArrayList<String> attribute = new ArrayList<String>();
    private ArrayList<ArrayList<String>> attributevalue = new ArrayList<ArrayList<String>>();
    private ArrayList<String[]> data = new ArrayList<String[]>();
    int decatt;
    public static final String patternString = "@attribute(.*)[{](.*?)[}]";

    Document xmldoc;
    Element root;
    Element firstElement;

    public ID3() {
        xmldoc = DocumentHelper.createDocument();
        root = xmldoc.addElement("root");
        firstElement= root.addElement("DecisionTree").addAttribute("value", "null");
    }

    public static void main(String[] args) {
        ID3 inst = new ID3();
        inst.readARFF(new File("F:\\data\\code\\data\\data4.txt"));

        inst.setDec("result");

        LinkedList<Integer> ll = new LinkedList<Integer>();
        for (int i = 0; i < inst.attribute.size(); i++) {
            if (i != inst.decatt)
                ll.add(i);
        }

        ArrayList<Integer> al = new ArrayList<Integer>();
        for (int i = 0; i < inst.data.size(); i++) {
            al.add(i);
        }

        System.out.println("al:");
        System.out.println(al);
        System.out.println();
        System.out.println("ll:");
        System.out.println(ll);
        System.out.println();


        inst.buildDT(inst.firstElement,"DecisionTree","DecisionTree", "null", al, ll);
        inst.writeXML("F:\\data\\code\\data\\data4.xml");
        return;
    }

    public void readARFF(File file) {
```

```
        try {
            FileReader fr = new FileReader(file);
            BufferedReader br = new BufferedReader(fr);
            String line;
            Pattern pattern = Pattern.compile(patternString);
            while ((line = br.readLine()) != null) {
                Matcher matcher = pattern.matcher(line);
                if (matcher.find()) {
                    attribute.add(matcher.group(1).trim());
                    String[] values = matcher.group(2).split(",");
                    ArrayList<String> al = new ArrayList<String>(values.length);
                    for (String value : values) {
                        al.add(value.trim());
                    }
                    attributevalue.add(al);
                } else if (line.startsWith("@data")) {
                    while ((line = br.readLine()) != null) {
                        if (line == "")
                            continue;
                        String[] row = line.split(",");
                        data.add(row);
                    }
                } else {
                    continue;
                }
            }
            br.close();
        } catch (IOException e1) {
            e1.printStackTrace();
        }

    }

    public void setDec(int n) {
        if (n < 0 || n >= attribute.size()) {
            System.err.println("errorãĂĆ");
            System.exit(2);
        }
        decatt = n;
    }

    public void setDec(String name) {
        int n = attribute.indexOf(name);
        setDec(n);
    }

    public double getEntropy(int[] arr) {
        double entropy = 0.0;
        int sum = 0;
        double[] arrcopy = new double[arr.length];

        for (int i = 0; i < arr.length; i++) {
            arrcopy[i] = arr[i];
            sum += arrcopy[i];
        }
        double[] arrcopy2 = new double[arrcopy.length];

        for (int i = 0; i < arrcopy.length; i++) {
            arrcopy2[i] = arrcopy[i] / sum;
        }
```

```
        for (int i = 0; i < arrcopy2.length; i++) {
            if(arrcopy2[i]>0.0&&arrcopy2[i]<1.0){
                entropy -= arrcopy2[i] * Math.log(arrcopy2[i]) / Math.log(2.0);
            }
        }
        return entropy;
    }

    public double getEntropy(int[] arr, int sum) {
        double entropy = 0.0;
        for (int i = 0; i < arr.length; i++) {
            entropy -= arr[i] * Math.log(arr[i] + Double.MIN_VALUE) / Math.log(2);
        }
        entropy += sum * Math.log(sum + Double.MIN_VALUE) / Math.log(2);
        entropy /= sum;
        return entropy;
    }

    public boolean infoPure(ArrayList<Integer> subset) {

        String value = data.get(subset.get(0))[decatt];
        for (int i = 1; i < subset.size(); i++) {
            String next = data.get(subset.get(i))[decatt];

            if (!value.equals(next))
                return false;
        }
        return true;
    }


    public double calNodeEntropy(ArrayList<Integer> subset, int index) {

        int sum = subset.size();

        double entropy = 0.0;

        int[][] info = new int[attributevalue.get(index).size()][];

        for (int i = 0; i < info.length; i++)
            info[i] = new int[attributevalue.get(decatt).size()];

        int[] count = new int[attributevalue.get(index).size()];


        for (int i = 0; i < sum; i++) {
            int n = subset.get(i);
            String nodevalue = data.get(n)[index];
            String decvalue = data.get(n)[decatt];
            int decind = attributevalue.get(decatt).indexOf(decvalue);

            info[nodeind][decind]++;

        }


        for (int i = 0; i < info.length; i++) {
            entropy += getEntropy(info[i]) * count[i] / sum;
        }

        return entropy;
```

```
    }

    public void buildDT(Element element,String fullName,String name, String value, ArrayList<In
                        LinkedList<Integer> selatt) {
        Element ele = null;
        System.out.println("fullName:"+fullName);
        @SuppressWarnings("unchecked")
        List<Element> list = root.selectNodes( "//" +fullName);
        Iterator<Element> iter = list.iterator();
        while (iter.hasNext()) {
            ele = iter.next();
            if (ele.attributeValue("value").equals(value))
                break;
        }

        if (infoPure(subset)) {

            element.setText(data.get(subset.get(0))[decatt]);
            return;
        }


        int minIndex = -1;
        double minEntropy = Double.MAX_VALUE;
        for (int i = 0; i < selatt.size(); i++) {
            if (i == decatt)
                continue;
            double entropy = calNodeEntropy(subset, selatt.get(i));

            if (entropy < minEntropy) {
                minIndex = selatt.get(i);
                minEntropy = entropy;
            }
        }


        String nodeName = attribute.get(minIndex);
        selatt.remove(new Integer(minIndex));
        ArrayList<String> attvalues = attributevalue.get(minIndex);


        for (String val : attvalues) {


            ArrayList<Integer> al = new ArrayList<Integer>();
            for (int i = 0; i < subset.size(); i++) {
                if (data.get(subset.get(i))[minIndex].equals(val)) {
                    al.add(subset.get(i));
                }
            }
            LinkedList<Integer> ll = new LinkedList<Integer>();
            for (int i = 0; i < selatt.size(); i++) {

                ll.add(selatt.get(i));
            }
            if(al.size()==0)
                continue;

            Element element2= element.addElement(nodeName).addAttribute("value", val);
            String nextFullName=fullName+"/"+nodeName;
            buildDT(element2,nextFullName,nodeName, val, al, ll);
```

```
        }
    }

    public void writeXML(String filename) {
        try {
            File file = new File(filename);
            if (!file.exists())
                file.createNewFile();
            FileWriter fw = new FileWriter(file);
            OutputFormat format = OutputFormat.createPrettyPrint();
            XMLWriter output = new XMLWriter(fw, format);
            output.write(xmldoc);
            output.close();
        } catch (IOException e) {
            System.out.println(e.getMessage());
        }
    }
}


//Grey Prediction


function [a,b]=predict(data)
    dataLength=length(data);


    for i=1:dataLength
        x(i)=0;
        for j=1:i
            x(i)=data(j)+x(i);
        end
    end

    for i=2:dataLength
        z(i)=0.5*x(i)+0.5*x(i-1);
    end

    Y=data(:,2:dataLength);
    Y=Y';

    B=z(:,2:dataLength);
    B=B';
    B=-B;
    for i=1:dataLength-1
        C(i,1)=1;
    end

    B=[B, C];

    u=inv(B'*B)*B'*Y;
    a=u(1);
    b=u(2);


end


function result = allPredict(data,n0,n)
```

```
    [row,column]=size(data);

    result=[];
    for i=1:row
        [a,b]=predict(data(i,:));
        a
        b
        result=[result
            predictN(a,b,n0,n,data(i,1))];
    end

end

function result = predictN(a,b,n0,n,x0)
    for i=n0:n0+n-1
        y2=(x0-b/a)*exp(-a*(i-1))+a/b;
        y1=(x0-b/a)*exp(-a*(i-2))+a/b;
        result(i)=y2-y1;
    end
end

function data = prePropess(preData)
    dataLength=length(preData)
    flag=1;

    for i=2:dataLength
        q(i)=preData(i-1)/preData(i)
        if q(i)<=exp(-2/(i)) || q(i)>=exp(2/(i+1))
            flag=0
        end
    end
    flag

end
```