Intelligent Decision Algorithm of Nutrients based on Big Data

Yungui Chen¹ and Hongbin Fan *,2

¹ 2016 grade Internet of things engineering professional school-enterprise cooperation class, College of software and communication engineering, XiangNan University,Chenzhou,423000, China
² College of software and communication engineering, XiangNan University,Chenzhou,423000, China
Email: fhb1208@163.com

Abstract. With the improvement of living standard, People's requirements are also improving for children's dietary nutrition, which fully illustrates the attention of people for healthy development of children. Children's dietary nutrition should not be solely based on advice of expert, because every child needs different dietary nutrients. In this paper, a large data computing framework MapReduce is studied, which fully considers the eating habits of children, the balance of dietary types and the interaction of food. A intelligent decision algorithm based on large data is proposed for the best children's dietary nutritional components. The algorithm is divided into two steps, the fuzzy GK algorithm is realized by the MapReduce framework, and the main nutrients are selected effectively; these nutrients are taken as parameters and used for multivariate linear regression model, so as to realize the rational allocation of nutrients. We design the algorithm, introduce the details of the algorithm in detail, and continuously optimize relevant parameters to produce a better dietary nutrition composition scheme. In the process of experiment, we analyze the performance of decision algorithm by testing examples, and make a lot of comparisons with the existing decision analysis algorithms. After a large number of arguments, it is proved that the algorithm is effective.

1. Introduction

According to the relevant statistics[1], the proportion of cereals and the proportion of vegetable foods are decreasing significantly in children's diet, the intake of vitamin and mineral are lower than 40% of the normal level, while the intake of meat food and high fat food exceed 130% of the normal level, which results in the proportion of overweight and obesity in children is above 10%. With the intake of high fat, children probably get some chronic diseases[2], such as diabetes and fatty liver. Some scholars have found that the phenomenon of childhood obesity is increasing year by year[3]. Children's irrational intake of dietary nutrients led to their diseases. Meanwhile, incorrect eating habits are also an important reason of childhood obesity.

In the process of eating breakfast, most children are used to monotonous food, or buy some fast food, a small part of children even do not eat breakfast; in the process of eating lunch, part of children are used to eat boxed lunch, there are few varieties of food, and very greasy; in the process of eating dinner, most children used to eat too much, but children can't digest easily because stomach is dormant at night. Many families prefer to do food processing in the way of frying, with more than edible oil of 45g and more than edible salt of 10g in the nutrient composition, all of which are over the relevant standards recommended by Chinese Nutrition Association. Long term consumption of children will lead to increasing of risk in the disease of heart disease and high blood pressure.
The rate of digestion, absorption and metabolism of different food are also different in the body of children, different nutrients have chemical reactions in the body of children[4], the combination of different nutrients may cause obstacles to nutritional absorption of children, and even to produce harmful substances to the body. When we chose the nutritional ingredients for children[5], especially for children of different physiques, we would avoid some unreasonable dietary nutrition components, which is not good for the healthy growth of children. Selection and composition of children's dietary nutrients involve several subjects, which include computer technology, nutrition, disease nutrition and data analysis. We study it and help relevant departments and children's parents develop a scientific nutrition decision strategy for children's dietary nutrition. Many researchers have put the theory of nutrition into making rational dietary strategies, but they do not classify the intake of different children's nutrients, and do not compare the actual intake of children's nutrients with the ideal supply either. In order to make more scientific decision of children's dietary intake components, we need to adopt relevant technology and algorithm of big data.

Data mining[6] is massive data analysis of knowledge in the face of database level and above, and it digs the hidden knowledge points in data. Different scholars have also improved the concept of data mining. Some scholars believe that data mining can extract hidden knowledge and model of potential practical mechanism from a pile of incomplete, unclear, large, noisy and stochastic data. The definition further illustrates the role of data mining, that is, it can analyze mass and irregular data, extract the hidden knowledge, generate useful knowledge, and provide valuable development suggestions for different industry decision-makers. Large data development environment can be used for analyzing massive nutrient data. As a large data development platform, Hadoop[7] has been get more and more applications. Many IT companies have built customer data scanning and log data processing based on Hadoop platform, which include Microsoft, IBM, Google, YAHOO and other companies. In the face of massive data, distributed development platform can cope with analysis of big data and extraction of useful knowledge. Hadoop is an integrated product of MapReduce computing framework. It is also a parallel computing platform based on object-oriented language. MapReduce[8] is the computing framework of Hadoop platform, data I/O can be in the form of key-value pairs, and each data is viewed as an object, and the frequency of different objects appearing is counted. MapReduce divides the data into key-value pairs, which can make data as a smaller unit, and the format between data becomes blurred. Therefore, fuzzy clustering and MapReduce can be well combined, used for classifying and selection of ingredients in the children's dietary nutrition.

Many scholars also put data mining methods into the study of children's nutritional components. Some scholars divide the intake of dietary nutrients into different nutritional grades, analyze the effects of these nutrients intake on different part of children's body, and we build a nonlinear model by computer technology and linear regression[9], the model uses multiple variables as parameters; some scholars put genetic algorithm into analyzing and solving of nutrition[10]; Some scholars analyzed data on energy metabolism, methionine and amino acids in different foods of children, and neural network is used for identifying and predicting the nutritional components of food[11]; some researchers tried to develop simple and easy hardware products for helping schools and parents care about children's nutrition. This paper, based on large data development framework MapReduce, realizes the selection of nutritional components in the children's diet, take nutrient components as variables and generate multiple linear regression models, thus, the decision-making results of children's dietary nutritional components are obtained. The demand can be met for the intake of every children's nutrition.

2. Related Knowledge of Decision Algorithm in The Dietary Nutrition Component
This part describes the data sources of nutritional components. From the perspective of decision algorithm design, we need to deeply study the large data parallel computing framework MapReduce, the distributed file system HDFS, and the linear regression model for composition of nutritional components.
2.1. Related Data of Nutrition

"The dietary nutrition reference intake of Chinese residents" that was revised by Chinese Nutrition Association can help us evaluate the dietary nutrition of different children and provide authoritative reference for the customizing of dietary nutrients[12,13]. We can see from this document that different children have a specific intake of nutrients, children of different constitution has a special need for intake of nutrients. All these data are universally acquired by the relevant departments of China. We can see from "Chinese food composition table", for different children, their amino acids, fatty acids, choline, vitamin K, vitamin D and other data, it is applicable to the decision analysis of dietary nutrients.

At the same time, we can also find a data form about food contradiction on the food science web site of china. The decision-making of dietary nutrients should avoid conflicting nutrients and ensure that different combinations of nutrients do not cause physical injury to children. Here we further enumerate the organization of children's recipes. As shown in table 1, it is arranged from the perspective of dietary intake.

| Intake       | Food                        |
|--------------|-----------------------------|
| <30g         | edible oil and salt         |
| >30g&&<50g   | Milk, soybean, nut         |
| >50g&&<100g  | Chicken, sausage, fish, shrimp, egg |
| >100g&&<300g | Vegetable, fruit           |
| >300g        | Rice, noodle, water        |

2.2. MapReduce

MapReduce is proposed by two engineers from Google, which abstracts the high order parallel functions, and the execution of data is reduced to two steps, it is, Map and Reduce. The process of Map is set by the user based on specific functions, data is classified based on the execution details of function, and a new key-value is given as intermediate value in each category; the process of Reduce is a summary for the results of Map classification.

MapReduce is put the large data as a whole, data can be divided into multiple parts, each part has a node to handle, multiple nodes are processed simultaneously, the process in this step is synchronized, then the results are collected together. From the internal framework perspective of MapReduce, the InputFormat function divides initial data set into multiple data blocks (Split), which is processed by Map administration, and the Map administration will assign a Map function to each data block; Map function classifies the initial key value pairs (k1, v1) into a series of intermediate key values pairs (k2, v2), after the Sort process and the Shuffle process, these intermediate keys will generate a set of values which have the same key, that is, (k3, list(v2)); Reduce phase, which will merge these intermediate sets (k3, list(v2)), and create (k3, v3), a framework flowchart of MapReduce is shown in figure 1.

In figure 1, the MapReduce framework includes Split function, Map function, Sort function, Shuffle function, Merge function and Reduce function. The data set is divided by Split function; with the help of Map administration, a Map function is allocated to each sub block and used for calculating in block. The sub data block has an intermediate process from the Map function to the Shuffle function, that is, the Sort process, the function is used for sorting in the data of sub data block; The Shuffle function rearrange different sub data blocks, and merge data sets simply, so that the capacity of data sets is reduced.; In the Merge stage, the data can be further merged, so the data set can be further reduced; In the Reduce phase, sum up values of the same key in the data set by function and then output to HDFS.
2.3. HDFS

HDFS represents Hadoop Distributed File System, which is a Master/Slave architecture. The file is controlled by the master node (NameNode). The file does not have to consider data format in the file system. According to the size of file, file system provides multiple data blocks for storing, and the storage of file is controlled by data node (DataNode). The architecture of HDFS is shown in figure 2.

In figure 2, NameNode is the master node, and it is an administrator in the whole HDFS. It is responsible for managing name of files, related information of cluster configuration, at the same time, it also pays attention to the running state of every data node (DataNode). If a data node did not work, NameNode would allocate a new DataNode to store the data. SecondNameNode is the second master node, it is used for sharing the management pressure of NameNode. The namespace of NameNode is limited and the metadata that it can store is also limited, so it needs SecondNameNode to backup the data.

DataNode is used for storing data blocks, which store files in different Split, and the size of each Split is set to 64MB. According to the actual size of file, the Split value can be modified. The Client terminal can interact with HDFS, read /write file to HDFS. After the HDFS receives command of read /write, NameNode checks the data storage of internal DataNode node and give feedback of data storage state to NameNode. According to the storage state of data in the HDFS, it executes the operation of Client terminal, or returns the information of read /write errors.
2.4. Multiple Linear Regression

In the data mining technology, we need to do statistics from different perspectives. The data shows up in the form of field, and there may be a certain statistical law between different fields, and these fields are expressed in the form of function, which is called regression analysis. Regression analysis can help people speculate value of unknown fields from these fields [14,15]. Linear regression is a basic form of regression analysis. Least squares method is used for making linear fitting of different fields. The fields in linear regression may be affected by a variable, or may be affected by multiple variables. We assume that a field may be affected by multiple fields, so it is necessary to establish an unknown field solution by linear equations, which is called multiple linear regression.

We assume the dependent variable $r$ and a fields $\{e_1, e_2, ..., e_a\}$ constitute the expression of multiple linear regression equation, as the formula (1) shows:

$$ r = \theta_0 + \theta_1 e_1 + \theta_2 e_2 + \ldots + \theta_a e_a + \eta $$

In the formula (1), $\theta_0$ represents the intercept, $(\theta_0, \theta_1, \theta_2, \ldots, \theta_a)$ means the regression coefficient, which is the slope in the linear equation, and $\eta$ is expressed the difference between the observed values and the predicted values. Assuming that $r$ is expressed as $(r_0, r_1, r_2, \ldots, r_n)$, the linear regression model is further developed into a multiple linear regression model.

$$
\begin{align*}
    r_1 &= \theta_0 + \theta_1 e_{11} + \theta_2 e_{12} + \ldots + \theta_a e_{1a} + \eta_1 \\
    r_2 &= \theta_0 + \theta_1 e_{21} + \theta_2 e_{22} + \ldots + \theta_a e_{2a} + \eta_2 \\
    \ldots \ldots \\
    r_n &= \theta_0 + \theta_1 e_{n1} + \theta_2 e_{n2} + \ldots + \theta_a e_{na} + \eta_n
\end{align*}
$$

3. Overall Design of Dietary Nutrition Component Decision Algorithm

The food combination in some studies is the combination of user's own decision or the combination of expert recommendation, but the solution of food combination is also a process of data mining. We need to design the decision-making algorithm of nutrition components, at the same time, further describe selection of nutrients and matching of nutrients.

3.1. The Overall Design of Algorithm

The decision algorithm of nutrient composition mainly consists of two steps, one is the selection of nutritional components, and the other is the combination of nutrients. In the process of nutrient selection, we put whole process in the MapReduce framework. We take user's diet as the input of Map stage. We install the food conflicting combination as a filtering condition. In the different times of a day, we set the weight of children's food as a screening condition. In the Reduce stage, we choose the main nutrients from multiple nutrients of user recipes, and output result to HDFS.

From HDFS, we get limited nutrients. Diet data that involves nutrients of choosing can be taken as the input, and we find out the best combination of nutrients by multiple linear regression methods.
We test the algorithm in Windows XP system, we deploy Hadoop0.20.205 as development environment, and 3 Computers are built for distributed cluster. Programming is carried out in the Eclipse environment. Java is the main programming language, and the overall design of algorithm is shown in Figure 3.

### 3.2. Selection of Nutrient Components
We choose the main nutritional components from user recipes. The nutrition we choose is to take full account of children's physical health factors. The diet should be divided into breakfast, lunch and dinner, and fruit, nuts, and edible oil should be removed from three meals, then we analyze these foods by the way of independent samples. In the Map function, we need to take into account physical indexes and disease factors of child. If food caused child's physical discomfort, it would be removed from the combination of nutrients; if food induced the appearance of children's disease, it would be removed from the combination of nutrients either.

### 3.3. Composition of Nutrition
Most foods may contain a variety of nutrients, and one of the most nutritious ingredients of children is carbohydrate. If the child only took carbohydrate, it can cause malnutrition; if a child had a variety of food, it would cause some nutritional components to exceed the normal intake, and even affect the normal eating habits of children. In the process of designing multiple linear regression models, we combine food intake with nutritional needs. In the process of solving model, we should take children's three meals intake, nutrition, food induced diseases, and some food's influence on children's physique as constraints.

The main nutrients and intake needs in the children's diet are shown in table 2:

| Nutritional Component | carbohydrate | fat  | protein | vitamin B |
|-----------------------|---------------|------|---------|-----------|
| Intake                | 300g          | <60g | 60g     | 1.5mg     |

Children's food combination includes rice, eggs, fruits and milk, nutritional components that correspond with the proportion of different foods can be shown in table 3.
Table 3. The Proportion of Nutrients In Different Foods (%)

| Food  | Carbohydrate | Fat  | Protein | Vitamin B |
|-------|--------------|------|---------|-----------|
| rice  | 59.7         | 3.8  | 4.7     | 1.8       |
| egg   | 0.7          | 17.5 | 12.5    | 2.7       |
| fruit | 3.7          | 0.2  | 0.6     | 0.05      |
| milk  | 3.4          | 17.6 | 15      | 0.3       |

In the process of solving food combination, we take the intake of nutrients in different foods as parameter variables of the equation and the normal intake of nutrients can be as the value of equation. We select the main nutrients, then select the main food from the children's diet. According to the nutritional components of different foods, we establish multiple linear regression equation, so as to figure out the best combination which is suitable for different children's recipes.

4. Implementation Details of Nutritional Component Decision Algorithm

Fuzzy GK algorithm is a fuzzy clustering algorithm. The distance between two points is not fixed, the adaptive way is adopted. Each fuzzy cluster is in the scope of covariance matrix, which can better contain clusters of different shapes, therefore, it has a better clustering effect.

The objective criterion function of fuzzy GK algorithm is defined as:

$$
\min_Q(T, V, P) = \sum_{i=1}^{e} \sum_{j=1}^{f} \mu_{ij}^{a} De_{ij}^{2} \\
\sum_{i=1}^{e} \mu_{ij} = 1, \mu_{ij} \in [0, 1]
$$

(3)

In the formula (3), $m_{ij}$ denotes the correlation degree of jth data in the ith cluster center, $a$ denotes the fuzzy factor, $V$ denotes number of clusters, $f$ denotes the number of data object, $De_{ij}$ denotes the distance between data object and cluster center, and it can better explain the distribution in the cluster.

We learn the idea of fuzzy GK algorithm, put it into MapReduce computing framework. The input data is the diet for children. The Map part and the Reduce part of algorithm are described as follows:

**map function:**

1. Input the data of Children' recipes;
2. Convert different foods into the composition of nutrients, $e_1, e_2, e_3...e_n$;
3. Split into <key: line, value: Food($e_1, e_2, e_3...e_n$)>;
4. Count the frequency of nutrients, install a ID of every nutrient;
5. Output <frequency of nutrient, food with this nutrient ($x_1, x_2, x_3...x_n$)>;

In the map function, we store children's recipes into HDFS; according to the proportion of nutrients, different foods are divided into different combinations of nutrients ($k_1e_1+k_2e_2+k_3e_3...+k_ne_n$), and the split combination of different foods is used as a value of each food; every nutritional component of different foods is distributed a ID. According to ID number, we count the occurrence frequency of nutrients, the food of frequency that corresponds with is used for the output of value in <k, v>.

**reduce function:**

1. Input <frequency of nutrient, food with this nutrient ($x_1, x_2, x_3...x_n$)>;
2. Chose several biggest frequencies of nutrient;
3. From foods of the nutrient, find food that corresponds to the largest proportion of nutrient;
4. Output <nutrient, corresponding food>;

In the reduce function, we take <k, v> as a new input in the output of map function, we select the most frequently occurring nutrients; from these nutrients, we find out the food that has the biggest proportion, that is, to find the food of the greatest correlation, which corresponds food with nutrient; we put <nutrient, corresponding food> as output, the result is finally stored in the output directory of
Multiple linear regression model is used for determining the linear relationship between multiple fields and predict fields, and it does not guarantee that each field has a direct impact on the prediction field. We put fuzzy GK algorithm, and MapReduce framework into children's selection of main food, and each of the major foods contains a relatively large amount of nutrition for children. We try to denote the main food of children as \( (w_1, w_2, w_3, \ldots, w_n) \), we choose \( TA \) kinds of foods \((w_{TA+1}, w_{TA+2}, w_{TA+3}, \ldots, w_{TA+TB})\) as lunch, and the rest of the food are as dinner. Each kind of food includes main nutritional components, as well as many trace nutrients. In the process of nutrient composition, we need to consider two factors, one is the total intake of nutrients, one is the amount of heat allocation for each meal, and the corresponding equation is as follows:

\[
1 \times u_1 + 2 \times u_2 + \ldots + u_n \times \frac{N_u}{u} = \frac{N_u}{u}
\]

\[
(J_{1}x_{1} + J_{2}x_{2} + \ldots + J_{TA}x_{TA}) : (J_{TA+1}x_{TA+1} + J_{TA+2}x_{TA+2} + \ldots + J_{TA+TB}x_{TA+TB}) : (J_{TA+TB+1}x_{TA+TB+1} + J_{TA+TB+2}x_{TA+TB+2} + \ldots + J_{n}x_{n}) = 3 : 4 : 3
\]

In formula (4) (5), \((x_1, x_2, \ldots, x_n)\) denotes the coefficient of parameters in the equation, \(u_i\) denotes that the intake of different foods that correspond to the same nutrient, \((J_1, J_2, \ldots, J_n)\) denotes the heat of different food consumption. In formula (5), the heat consumption of breakfast, meal and dinner is solved, which run up to 3:4:3.

5. Examples and Performance Analysis in The Decision Algorithm of Dietary Nutrition

In the process of fuzzy experiment, we do all the tests by MapReduce computing framework and HDFS, put children's basic information and recipes as the input. We build the experimental environment, 3 computers are installed Ubuntu10.04 operating system, respectively, create the user of hadoop, their permissions are equivalent to Root. Hadoop and eclipse are connected, the JDK, SSH version of 3 computers are in line with the version of hadoop. At the same time, 3 computers need to achieve login of no-password, the main node can access the subsidiary nodes, the distribution of corresponding roles in these nodes is shown in table 4.

| Host Name        | IP           | Distribution | HDFS Role  | MapReduce Role |
|------------------|--------------|--------------|------------|----------------|
| hadweibw         | 128.192.168.101 | master       | namenode   | 1.8            |
| hadweibw01       | 128.192.168.102 | slave        | datanode   | 2.7            |
| hadweibw02       | 128.192.168.103 | slave        | datanode   | 0.05           |

The nutritional components of children's health include synthetic organic compound and non-synthetic organic compound. Non-synthetic products must be obtained by breakfast, lunch and dinner. The number of nutrients' kinds is more than 40, and record the nutritional components in every kind of food. Such operations are time-consuming. In the process of experiment, we choose protein, fat, carbohydrate, calcium, vitamin B, vitamin C and vitamin E. We recorded the children's standard intake of different nutrients by checking related information.

The fuzzy GK algorithm can be used for iterating data again and again, and the central point in the different clusters will change continuously. The extreme values may appear in the execution process. It is difficult for us to make sure that the result is the optimal solution of equation. GK algorithm needs to set the number of clusters, which different parameter settings will lead to different clustering results, thus affects the result of nutrient selection. The nutritional components of food are different in different seasons, that is, nutrients are easily affected by seasons. Therefore, in the process of statistics, we apply the average value into the express of food nutrient content, so the error of calculation is small. We choose the nutritional ingredients for different recipes, and the results are shown in table 5:
Table 5. Selection of Nutritional Components In Different Recipes

| Recipes               | Nutritional Components          |
|-----------------------|---------------------------------|
| rice, noodles         | carbohydrate                   |
| eggs, bread           | protein                         |
| grain, fresh vegetables, fruit | carbohydrate, vitamin B |
| chicken, sausages     | protein, fat                    |
| broccoli, tomato      | carotene                        |

From table 5 we can see that, if the diet is rice and noodles, its main nutrients are carbohydrates; if the diet is egg and bread, its main nutrient is protein; if the diet is cereal, fresh vegetables and fruit, its main nutrients are carbon and water compounds, vitamin B; if the recipe is chicken and Sausage, its main nutritional components are protein and fat. If the recipe is broccoli and tomato, its main nutritional component is carotene.

We adopt the decision algorithm, randomly choose 5 children, and calculate the output of heat. The calculated results are used as the actual results, the ideal heat output of children is defined by some literature, and the errors are analyzed with different results, as shown in table 6.

Table 6. Error Analysis of Children' Output Heat

| Number  | Ideal Heat Output | Actual Heat Output | Error Percentage |
|---------|-------------------|--------------------|------------------|
| 014501  | 785.2J            | 777.8J             | 0.94%            |
| 014507  | 774.6J            | 773.4J             | 0.15%            |
| 014508  | 781.5J            | 776.8J             | 0.6%             |
| 014512  | 776.4J            | 754.1J             | 2.8%             |
| 014527  | 772.6J            | 762.8J             | 1.2%             |

From table 6, we can see the results by decision algorithm of dietary nutrition we propose, the actual heat output is very close to the ideal heat output. The error result is between the 0.15%-2.8% and the average error is 1.1%, which proves that the algorithm can reasonably regulate the distribution of nutritional components in children.

6. Discussion
Children's dietary nutrients should not only focus on the intake of nutrients, but also consider the physical conditions of different children. In this paper, a large data parallel computing framework (MapReduce) is used for implementing GK algorithm. Thus, the main nutritional components of children are selected and these nutrients are used as parameters and they can be solved in the multiple linear regression equation. Experiments are proved that the algorithm is effective.

7. Acknowledgements
This work was supported by Hunan Province Colleges and Universities Teaching Reform Research Project "research and practice of application-oriented talents training of computer major in local colleges and Universities Based on the perspective of subject competition" (xiangjiaotong [2017] No.452, 449), and Xiangnan University Laboratory Open Project "Computer Application and Software Development" (Xiangnan University xiao fa [2018] No.159.12), and Xiangnan University students' research study and innovative experiment plan project "Design and implementation of the old man's bodyguard system based on Android platform and Beidou navigation" (Xiangnan University xiao fa [2018] No.46, 34).

8. References
[1] Leve L D, Neiderhiser J M, Shaw D S, Ganiban Jody, Natsuaki M N and Reiss David. “The Early Growth and Development Study: A Prospective Adoption Study From Birth Through
Middle Childhood[J].”, Twin Research & Human Genetics the Official Journal of the International Society for Twin Studies, Vol 22 No 1, pp. 412-423, 2013.

[2] Dieffenbach S, Stein A D. “Stunted child/overweight mother pairs represent a statistical artifact, not a distinct entity[J].”, Journal of Nutrition, Vol. 142 No. 4, pp: 771-773, 2012.

[3] Brandao E, Soares S, Mateus N, Freitas VD. “Human saliva protein profile: Influence of food ingestion[J].”, Food Research International, Vol. 64, pp: 508-513, 2014.

[4] Snyderwine E G, Turesky R J, Turteltaub K W, Davis C D, Sadrich N, Schut H A, Nagao M, Sugimura T, Thorgeirsson U P, Adamson R H, Thorgeirsson S S. “Metabolism of food-derived heterocyclic amines in nonhuman primates.[J].”, Mutat Res, Vol. 376 No. 1-2, pp: 203-210, 2015.

[5] Damay C, Ezan P, Gollety M, Nicolas-Hemar V. “Social interactions as support for learning about food: the case of the French school cafeteria[J].”, Young Consumers Insight & Ideas for Responsible Marketers, Vol. 12, No. 2, pp: 110-120, 2013.

[6] Anitha J, Rangarajan R. “A technique to mine clusters using privacy preserving data mining[J].”, International Review on Computers & Software, Vol. 8 No. 10, pp: 2316-2326, 2013.

[7] Aji A, Wang F, Vo H, Lee R, Liu Q, Zhang X, Saltz J. “Hadoop-GIS: A High Performance Spatial Data Warehousing System over MapReduce[J].”. Proceedings Vldb Endowment, Vol. 6 No. 11, pp: 1009-1020, 2013.

[8] Lizeh Wang, Jie Tao, Rajiv Ranjan, Holger Marten, Achim Streit, Jingying Chen, Dan Chen. “G-Hadoop: MapReduce across distributed data centers for data-intensive computing[J].”, Future Generation Computer Systems, Vol. 29 No. 3, pp: 739-750, 2013.

[9] Salimans T, Knowles D A. “Fixed-Form Variational Posterior Approximation through Stochastic Linear Regression[J].”, Bayesian Analysis, Vol. 8 No. 4, pp: 837-881, 2013.

[10] Sakawa M, Nishizaki I, Uemura Y, Hitaka M. ‘Interactive fuzzy programming for two-level nonconvex nonlinear programming problems through genetic algorithms[J].’, Electronics & Communications in Japan, Vol. 10 No. 1, pp: 70-77, 2015.

[11] Wong C, Versace M. “CARTMAP: a neural network method for automated feature selection in financial time series forecasting[J].”, Neural Computing & Applications, Vol. 21 No. 5, pp: 969-977, 2012.

[12] Zhang J, Xue H, Cheng X, Wang Z, Zhai F, Wang Y, Wang H. “Influence of proximities to food establishments on body mass index among children in China[J].”, Asia Pacific Journal of Clinical Nutrition, Vol. 25 No. 1, pp: 134-141, 2016.

[13] Guo C, Wei J, Gao W, Pu L, Tang Z, Li L. “Plasma riboflavin is a useful marker for studying riboflavin requirement in Chinese male adults[J].”, Nutrition Research, Vol. 36 No. 6, pp: 534-540, 2016.

[14] Allahbakshianfarsani M, Abdian N, Ghsemidehkordi P, Sadeghiani M, Saffari-Chaleshtori J, Hashemzadeh-Chaleshtori M, Khosravi-Farsani S. “Cytogentic analysis of human dermal fibroblasts (HDFs) in early and late passages using both karyotyping and comet assay techniques[J].”, Cytotechnology, Vol. 66 No. 5, pp: 815-822, 2014.

[15] Alapati S, Kabala Z J. ‘Recovering the release history of a groundwater contaminant using a non–linear least–squares method[J].’. Hydrological Processes, Vol. 14 No. 6, pp: 1003-1016, 2015.