CrossMark

# The Research on the Application of Qualitative Mapping in MapReduce

**Jiali Feng[1] · Guanglin Xu[2] · Xiaolin Xu[3]**

**Abstract** MapReduce is a mathematical tool handling the large-scale data sets through paralleling and distributive calculation. Currently the operations of MapReduce mainly include sorting, grouping and joining, etc. This paper undertakes a research on qualitative mapping and MapReduce, and finds that the solution procedure of qualitative mapping can be a new way of transforming data for MapReduce. Two examples are given to illustrate how to use qualitative mapping model to transforming semi-structured or unstructured data.

**Keywords** MapReduce · Qualitative mapping · Qualitative judgment

## 1 Introduction

In traditional relationship database, data processing can only be conducted by SQL language after loading of all the data. It must be structured with data in well-defined

✉ Guanglin Xu
glxu@outlook.com

Jiali Feng
jlfeng@shmtu.edu.cn

Xiaolin Xu
xlxu2001@163.com

[1] College of Information Engineering, Shanghai Maritime University, Shanghai 200135, China

[2] School of Mathematics and Information, Shanghai Lixin University of Commerce, Shanghai 201620, China

[3] College of International Vocational Education, Shanghai Second Polytechnic University, Shanghai 201209, China

schemas. Ever since the emergence of big data technology, semi-structured or unstructured data become accepted, and raw data is stored under the structure according to key/value. The processing method can be parallel and distributive, therefore extracting value from large-scale data becomes possible. MapReduce is based on the improvement of mathematics and computer technology handling and generating large-scale data sets.

MapReduce technology mainly includes two functions: mapping and reducing. "Mapping" provides a series of transformations from a source data to a set of intermediate key/value pairs. "Reducing" happens when all the intermediate values are merged in functions associated with the same intermediate key. To put it simply, the input data is fed into the mapping function and the resultant temporary data is put into a reducing function, then the data sets are sorted, grouped or joined.

In attribute theory, the attributes are defined as the characteristics and features of things, while quality is defined as the intrinsic attributes that are used to distinguish the differences of things. Because the elements or ingredients used to form things are different, the differences in the deep structure will affect the nature of things. Features are defined as the outward signs or the characteristics of things. The difference of things' nature will be manifest in their superficial structure difference too. Things have various characteristics. The attribute theory applies these methods like qualitative mapping, transformation function, and datum transformation, to study the qualitative and quantitative differences of things. Among which, qualitative mapping is not only a mathematical abstract expression, but also a mathematical model of transformation able to give a qualitative recognition and classification of things.

Qualitative mapping, in essence, is a mapping from quantity to conception, and meanwhile a mapping from conception to conception. Converse qualitative mapping, on the contrary, is a mapping from conception to quantity. In processing large-scale data, besides the operations of sorting, grouping and joining, there are some transformations done from qualitative to quantitative or vice versa. Based on the qualitative mapping theory, this paper seeks a new way of transforming data for MapReduce.

## 2 Related Works

In [1,2], there are detailed introductions on how to apply MapReduce in the large clusters to process data, Ekanayake et al. [4] proposes a method named "Twister" to shorten the time in MapReduce calculation. Ekanayake et al. [5] offers a practical example to apply MapReduce technique into two scientific data analyses. Abouzeid et al. [3] studies the feasibility of building a hybrid system that takes the best features from the hybrid of MapReduce and DBMS technologies. Hadoop [6] is an introduction to the current mainstream MapReduce software Hadoop. Ranger et al. [7] did a research on MapReduce high-performance calculation in multi-nuclear and multi-processor. Xu and co-workers [8–15] are the latest main achievements in the research in attribute theory and data processing.

## 3 The Simplest Qualitative Mapping Model Based on Simplest Qualitative Judgment

In medical science, doctors judge a patient's fever by his/her temperature(t); and judge whether the patient is a diabetic by the fasting blood glucose (FBG); and judge whether a patient has the disease of hyperlipemia by the three data of his cholesterol, triglyceride, high-density lipoprotein (HDL). This is to say, if the expert wants to determine the nature of the object(u), he should judge on the basis of the qualitative criterion ([$\alpha$, $\beta$]) of the quality (p(u)), and the quantity(x) of one of the attribute(a(u)) of the object(u). If x falls within [$\alpha$, $\beta$], then the object has the quality p(u); if not, the conclusion is that the object does not have the quality of p(u). This operation can be summed as: "p(u), if and only if, x $\in$ [$\alpha$, $\beta$]."

Because in judging p(u), the rules not only should be clear and simple, but also can be described by the arithmetic symbol "$\leq$" or "$\geq$", so it is called "simplest qualitative judgment". simplest qualitative judgment can be described in the mapping form below:

**Definition 1** If a(u) is a certain attribute of the object u, x $\in$ X $\subseteq$ R is one of its quantity value, and p(u) $\in$ P$_u$ is one quality of the attribute a(u), [$\alpha$, $\beta$] $\in$ $\Gamma$ is the qualitative criterion of p(u), then the mapping is: $\tau$ : X $\times$ $\Gamma$ $\rightarrow$ {0, 1} $\times$ P$_u$, so:

$$\tau(x, [\alpha, \beta]) = x \underset{?}{\in} [\alpha, \beta] = \begin{cases} p(o) & x \in [\alpha, \beta] \\ -p(o) & x \notin [\alpha, \beta] \end{cases} \tag{1}$$

Fomula (1) becomes the judgment mapping or qualitative mapping of the simplest quality p(u). Among which, $\in_?$ is the operator of the question: "Is x in[$\alpha$, $\beta$]?", or "Is x an operator to satisfy the qualitative criterion of quality p(u)?"

Because for every attribute a(u), the input is one quantity value (x), and the output is one quality p(u) or q(u) (not p(u)). Therefore, if we are to emphasize that (1) is a transformation process from quantity (x) to quality p(u), then the symbol $\in_?$ can also be called "Quality–quantity Feature Transformation Operator" or "Feature Extraction Operator" of the attribute a(u).

*Example 1* Suppose FPG(Tom) is Tom's blood sugar value, them the normal qualitative criterion for "blood sugar" is [3.9, 6.1 mm/l]. The process that the doctor gives a diagnosis whether the FPG(Tom) is normal can be described as:

$$\tau(FPG(John), [3.9_n, 6.1_n] = FPG(John) \underset{?}{\in} [3.9_n, 6.1_n]$$

$$= \begin{cases} Normal(FPG(John)) & FPG(John) \in [3.9_n, 6.1_n] \\ No - Normal(FPG(John)) & FPGx(John) \notin [3.9_n, 6.1_n] \end{cases} \tag{2}$$

Because x stands for a quantity, and p(u) is a quality feature, so qualitative mapping can be deemed as a transformation process from quantitative x to qualitative p(u).

In the law of mutual change between quality and quantity of things, there are two kinds of change: the qualitative change and the quantitative one, and there is also a law governing the change between the two of them. In other words, " if the change

of things is within the extent of degree, although its scale, quantity will change more or less, the quality of things will not be altered. The change is not a fundamental one. Only when the quantity accumulates to a certain degree or certain critical point, the quantitative change will escalate into a qualitative one. The change of things' attributes can be classified as follows: the qualitative change or quantitative change, the mutual transformation between the qualitative change and the quantitative change. There is also "degrees" or "qualitative criteria" regulating or adjusting their mutual transformation.

## 4 The Application of the Qualitative Mapping Model of "Simplest Qualitative Judgment" in MapReduce

Based on the simplest qualitative judgment mapping model, we get the following calculation method, i.e., MapReduce. If $x_1, x_2, \ldots, x_n \in [\alpha, \beta]$ are a set of quantity attributes corresponding to p(u), and it can be deemed as a key, and the quality of attributes p(u) can be deemed as a value, then $\{x_1/p(u)\}$ can correspond to the key/value pairs in the basic data units of large data. If there are n data pairs, respectively $\{x_1/p(u)\}, \{x_2/p(u)\}, \ldots, \{x_n/p(u)\}$, then after shuffling the data pairs, the result will be $\{p(u)/x_1\}, \{p(u)/x_2\}, \ldots, \{p(u)/x_n\}$, then after mapping the data, the result is $\{p(u)/(x_1, x_2, \ldots, x_n\ \}$. An attribute theory explanation will be that if things have several qualitative attributes $x_1, x_2, \ldots, x_n$, its corresponding qualitative attribute p(u), i.e., $(x_1, x_2, \ldots, x_n) \in [\alpha, \beta]$, if $x_1, x_2, \ldots, x_n$ are ordered, then $[x_1, x_n]$ are the qualitative criterion of p(u), that is to say, the mapping process is the solution process of qualitative criterion. And after getting $[x_1, x_n]$, the initial data pairs can be written as $\{p(u)/[x_1, x_n]\}$, which is a reduction of the initial data pairs. Therefore, the solution process of qualitative criterion is a calculative method called MapReduce.

*Example 2* Now there are five initial data sets for five students, the students' scores are respectively: 60, 90, 69, 100, 93, their corresponding grades are C, A,C,A and A. First of all, we query the grades and get all the data containing their grades, then after mapping the grades, the data we get are: {grade: a, {score: 90, 93, 100}; grade: c, {score: 60, 69}}. Because the data for the scores are ordered, therefore by reduction, we conduct simplest qualitative mapping (1),so we get the qualitative criteria for Grade C and Grade A, which are [90, 100] and [60, 69] respectively. The process is shown in Fig. 1.

## 5 Multi-attributes Qualitative Mapping Model

In the real world, an object contains many attributes, just as the doctor needs the indicators of cholesterol, triglyceride, HDL to determine whether a patient's blood-fat level is normal or not. For $a(u) = \wedge_{i=1}^{n} a_i(u)$, which is an integrated attributes containing several attribute factors, its qualitative value judgment can be summed as the following qualitative mapping:

**Definition 2** If $a(u) = \wedge_{i=1}^{n} a_i(u)$ is the Integrative Attributes of n attribute factors $a_i(u), i = 1, \ldots, n$, of the object u, $x = (x_1, \ldots, x_n)$ is the quantity value of the
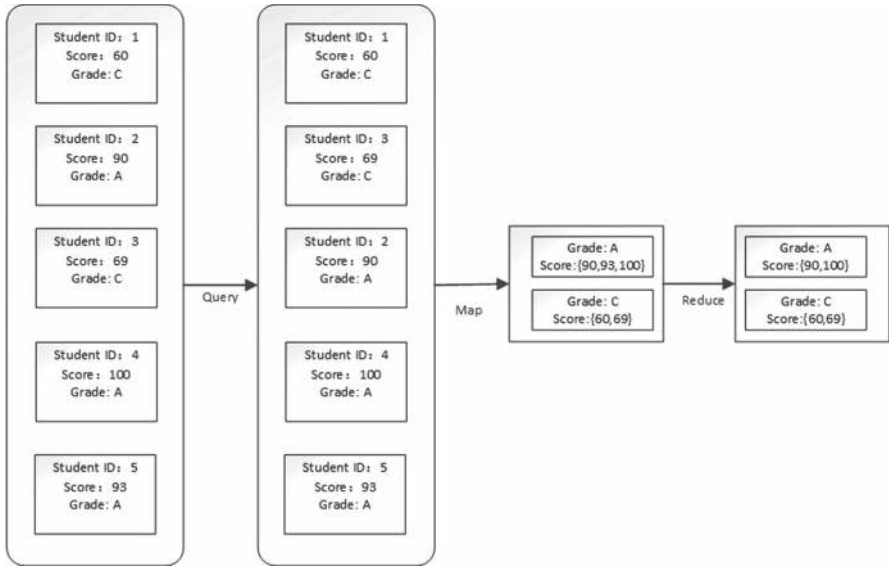
**Fig. 1** The application of simplest qualitative criterion on MapReduce

attribute a(u), among which, $x_i \in X_i \subseteq R$ is the quantity feature value of $a_i(u)$,and $p_i(u) \in P(u)$ is a certain quality of the attribute $a_i(u)$, $\Gamma = \{[\alpha_i, \beta_i] | [\alpha_i, \beta_i]\}$ is the quantitative criterion of the quality $p_i(u)$, then the hypercube $[\alpha, \beta] = [\alpha_1, \beta_1] \times \cdots \times [\alpha_n, \beta_n]$ is the qualitative crierion of the Integrated Quality $p(u) = \wedge_{i=1}^{n} p_i(u)$.
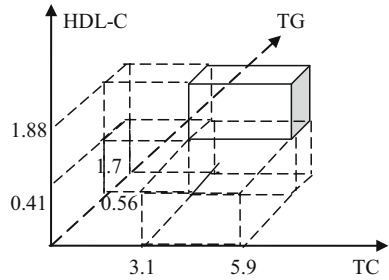
Then $\tau : X \times \Gamma \to \{0, 1\} \times P(u)$ is called a mapping based on the n-dimensional hypercube $[\alpha, \beta]$,thus it is a qualitative mapping given to $x = (x_1, \ldots, x_n)$. If for any $x \in X$, there exists $[\alpha, \beta] \in \Gamma$ and the quality $p(u) = \wedge_{i=1}^{n} p_i(u) \in P(u)$ with the qualitative criterion as $[\alpha, \beta]$,so:

$$\tau(x, [\alpha, \beta]) = x \underset{?}{\in} [\alpha, \beta] = \overset{n}{\underset{i=1}{\wedge}} (x_i \underset{?}{\in} [\alpha_i, \beta_i]) = \overset{n}{\underset{i=1}{\wedge}} \tau_{p_i(o)}(x_i) \qquad (3)$$

In medical science, cholesterol (TC), triglyceride (TG), high-density lipoprotein (HDL-C) are three indicators or parameters that a doctor uses to decide whether a patient's (John's) Lipoproteinemia is normal or not,and the qualitative criteria or diagnosis criteria of TC, TG and HDL-C (Normal)are respectively: $[3.1_n, 5.9_n]$ (TC), $[0.56_n, 1.7_n]$(TG) and $[0.4_n, 1.88_n]$ (HDL-C),with the unit as mmol/L,therefore, if and only if the value of the three laboratory indexes tc, tg and hdl-c fall within the normal value interval, Lipoproteinemia (John) can be judged as normal: LipedNormal (John). As indicated in Fig. 2, they form a 3-dimensional cuboid $[\alpha_{\text{Normal}}, \beta_{\text{Normal}}] = [3.1_n, 5.9_n]$ (TC)$\times[0.56_n, 1.7_n]$ (TG)$\times[0.4_n, 1.88_n]$ (HDL-C),that is to say, if and only if John's laboratory index vectors $x = $(tc,tg,hdl-c) fall within the cuboid of the three norms,his blood-fat level can be judged as normal. Therefore, the doctor's diagnosing operation for the patient John's blood-fat (tc,tg,hdl-c) can be described as:

$$\tau(tc(J), tg(J), hdl - c(J)), [\alpha, \beta])$$

**Fig. 2** The qualitative criteria of the normal (tc, tg, hdl-c)



$$= \tau \left( (tc(J), tg(J), hdl - c(J)), \begin{pmatrix} (3.1_n, 5.9_n](TC) \\ (0.56_n, 1.7_n](TG) \\ (0.41_n, 1.88_n](HDL - C) \end{pmatrix} \right)$$

$$= \{tc(J) \in (3.1_n, 5.9_n](TC)\} \underset{?}{\wedge} \{tg(J) \underset{?}{\in} (0.56_n, 1.7_n](TG)\}$$

$$\wedge \{hdl - c(J) \underset{?}{\in} (0.41_n, 1.88_n](HDL - C)\}$$

$$= \begin{cases} 1 & \{tc(J) \in [3.1, 5.9]_{TC}\} \wedge \{tg(J) \in [0.56, 1.7]_{TG}\} \\ & \wedge \{hdl - c(J) \in [0.41, 1.88]_{HDL\_C}\} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

*Example 3* A number of doctors' operations can be described by qualitative mapping (4) (used for the multi-attributes quantity–quality transformation), including the diagnosis based on laboratory tests, e.g., when the doctor diagnoses the hyperlipemia, the three-dimensional hexahedral Qualitative Criterion Grid drawn by the three indicators of cholesterol (TC), triglyceride (TG), high-density lipoprotein (HDL-C) can serve as a tool. It can be described as:

$$\tau \left( (tc(J), tg(J), hdl - c(J)), \right.$$
$$\left. \begin{pmatrix} [\alpha_L, 3.1_L](TC) & (3.1_N, 5.9_N](TC) & (5.9_H, \beta_H](TC) \\ [\alpha_L, 0.56_L](TG) & (0.56_N, 1.7_N](TG) & (1.7_H, \beta_H](TG) \\ [\alpha_L, 0.41_L](HDL - C) & (0.41_N, 1.88_N](HDL - C) & (1.88_H, \beta_H](HDL - C) \end{pmatrix} \right)$$

$$= \{[\tau_L(tc) \wedge \tau_L(tg) \wedge \tau_L(hdl - c)]\overline{\vee}, \ldots, \overline{\vee}[\tau_L(tc) \wedge \tau_H(tg) \wedge \tau_H(hdl - c)]\}$$

$$\overline{\vee}\{[\tau_N(tc) \wedge \tau_L(tg) \wedge \tau_L(hdl - c)]\overline{\vee}, \ldots, \overline{\vee}[\tau_N(tc) \wedge \tau_H(tg) \wedge \tau_H(hdl - c)]\}$$

$$\overline{\vee}\{[\tau_H(tc) \wedge \tau_L(tg) \wedge \tau_L(hdl - c)]\overline{\vee}, \ldots, \overline{\vee}[\tau_H(tc) \wedge \tau_H(tg) \wedge \tau_H(hdl - c)]\}$$

$$= \underset{i_1 \in \{TC, TG, HDL-C\}}{\overline{\vee}} \{ \underset{j_i \in \{L,N,H\}}{\overline{\vee}} [ \underset{i_1 \in \{TC, TG, HDL-C\}}{\wedge} \underset{v(i_1 j_1, i_2 j_2, i_3 j_3)}{\tau} (b - lipld)] \} \tag{5}$$

In case of multi-dimension mapping, if $a_i(u)$ is one attribute of the object $u, i = 1, \ldots, n, x_i \in X_i$ is the quantity feature value of $a_i(u)$, $p_{ij}(u)$ is the $j$th quality feature of $a_i(u), j = 1, \ldots, m, |[\alpha_{ij}, \beta_{ij}] \subseteq X_i$ is the qualitative criterion of $p_{ij}(u)$, $\Gamma = \{[\alpha_{ij}, \beta_{ij}]\}$ is the cluster of the qualitative criteria, and space x is divided by sub qualitative criteria $[\alpha_{ij}, \beta_{ij}] \subseteq X_i$ into m divisions or dimensional sub-spaces: $X_i = \bigcup_{j=1}^{m} [\alpha_{ij}, \beta_{ij}]$ ,then the input space x is cut into a grid, as shown in Fig. 3.
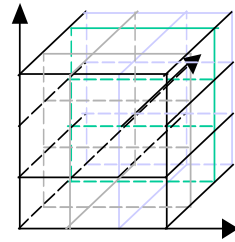
**Definition 3** If $a_i(u)$ is a certain attribute of the object u, $i = 1, \ldots, n$, $x_i \in X_i$ is the quantity feature value of $a_i(u)$, $p_{ij}(u)$ is the $j_{th}$ quality feature of $a_i(u)$, $j = 1, \ldots, m$, $|[\alpha_{ij}, \beta_{ij}] \subseteq X_i$ is the quality criterion of $p_{ij}(u)$, $\Gamma = \{[\alpha_{ij}, \beta_{ij}]\}$ is the cluster of its quality criteria, and $[\alpha_{ij}, \beta_{ij}] \bigcap [\alpha_{il}, \beta_{il}] = \varnothing$, $1 = 1, \ldots, m$, $1 \neq j$, and $X_i = \bigcup_{j=1}^{m} [\alpha_{ij}, \beta_{ij}]$ are satisfied. If $[\alpha_{i_k j_l}, \beta_{i_k j_l}]$ is the quality criterion of the $j_{lth}$ quality $p_{i_k j_l}(u)$ of the $i_k$ attribute of $a_{i_k}(o)$, $i_k \in \{1, \ldots, n\}$, $j_1 \in \{1, \ldots m\}$, $[\alpha_v, \beta_v] = \{x = (x_1, \ldots, x_n) | (x_1, \ldots, x_n) \in [\alpha_{i_1 j_1}, \beta_{i_1 j_1}] \times \cdots \times [\alpha_{i_k j_l}, \beta_{i_k j_l}] \times \cdots \times [\alpha_{i_n j_m}, \beta_{i_n j_m}]\}$ is an n-dimensional parallelepiped constituted by n different quality dimensions $[\alpha_{i_k j_l}, \beta_{i_k j_l}]$, hence, $(i_1 j_1, \ldots, i_k j_l, \ldots, i_n j_m)$ is a combination of subscript variables $i_k$ and $j_l$, $v = v(i_1 j_1, \ldots, i_k j_l, \ldots, i_n j_m)$ is the ordinal number of this combination, because for every $i_k, j_l$ there is m choices, so different combinations have $m^n$ kinds of combinations, therefore, $v \in \{1, \ldots, m^n\}$. If $p_v(u) = \wedge_{k=1}^{n,m}{}_{l=1} p_{i_k j_l}(u)$ is the integrated quality of the object based on the quality criterion of $[\alpha_v, \beta_v]$, $\Gamma = \{[\alpha_v, \beta_v]\}$ is the cluster of the grid of all the quality criterion $[\alpha_v, \beta_v]$, and if $([\alpha_v, \beta_v]) = \begin{pmatrix} [\alpha_{11}, \beta_{11}] & \ldots & (\alpha_{1m}, \beta_{1m}] \\ \vdots & (\alpha_{i_k j_l}, \beta_{i_k j_l}] & \vdots \\ [\alpha_{n1}, \beta_{n1}] & \ldots & (\alpha_{nm}, \beta_{nm}] \end{pmatrix}$ is the grid constructed by the $m^n$ disjoint n-dimensional parellelepiped $[\alpha_v, \beta_v]$, then the mapping $\tau : X \times \Gamma n \to \{0, 1\}$ based on the quality criterion of $([\alpha_v, \beta_v])$ (or $G([\alpha_v, \beta_v])$), if for any $x \in X$, there exists the quality $p_v(u) \in P(u)$ based on the quality criteria of $[\alpha_v, \beta_v] \in \Gamma n$ and $[\alpha_v, \beta_v]$, so:

$$\tau \left( (x_1, \ldots, x_n), \begin{pmatrix} [\alpha_{11}, \beta_{11}] & \ldots & (\alpha_{1m}, \beta_{1m}] \\ \vdots & (\alpha_{i_k j_l}, \beta_{i_k j_l}] & \vdots \\ [\alpha_{n1}, \beta_{n1}] & \ldots & (\alpha_{nm}, \beta_{nm}] \end{pmatrix} \right)$$

$$= \bigvee_{\substack{j=1 \\ l}}^{m} \bigwedge_{\substack{i=1 \\ k}}^{n} \{(x_1, \ldots, x_n) \underset{?}{\in} [\alpha_{i_1 j_1}, \beta_{i_1 j_1}]$$

$$\times \cdots \times (\alpha_{i_k j_l}, \beta_{i_k j_l}] \times \cdots \times (\alpha_{i_n j_m}, \beta_{i_n j_m}]\}$$

$$= \bigvee_{\substack{j=1 \\ l}}^{m} \{\cdots \{\bigwedge_{\substack{i=1 \\ k}}^{n} \tau_{v(i_1 j_1, \ldots, i_k j_l, \ldots, i_n j_m)}(x)\}\} \tag{6}$$

Among which,

$$\tau_{\nu(i_1 j_1, \ldots, i_k j_l, \ldots, i_n jm)}(x) = \begin{cases} 1 & iff \ x \in (\alpha_v, \beta_v] \\ 0 & iff \ x \notin (\alpha_v, \beta_v] \end{cases} \tag{7}$$

Then we call Formula (6)a qualitative mapping judging whether an object u with x vectors has the quality $p_v(u)$ or not, and we call Formula (7)a qualitative mapping based on the quality criterion of $[\alpha_v, \beta_v]$,or the factor mapping of Formula (7),written as:$\tau p(x, [\alpha_v, \beta_v])$

## 6 The Application of Multi-attributes Qualitative Mapping into MapReduce

The Large-scale data set in the real world is usually a Cascading Structure made up of a set or collection of key/value pairs. Confronted by the complex data of the Cascading Structure,we can conduct a MapReduce on the data by a multi-attributes qualitative mapping model. According to Definition 3, $a(u) = \wedge_{i=1}^{n} a_i(u)$ is n factor attributes $a_i(u)$ of the object u, i = 1,…, n, its integrated attributes, $x = (x_1, \ldots, x_n)$ is the quantity value of the attribute a(u), among which, $x_i \in X_i \subseteq R$, is the quantity feature value of $a_i(u)$, $p_i(u) \in P(u)$ is a certain quality of the attribute $a_i(u)$, $\Gamma = \{[\alpha_i, \beta_i]|[\alpha_i, \beta_i]$ is the quality criterion of the quality $p_i(u)$, $|[\alpha_i, \beta_i] \subseteq X_i$ is the quality criterion of $p_i(u)$. If $p_i(u)$ is the quality criterion of Key $p_i(u)$, $|[\alpha_i, \beta_i]$ is the corresponding value of $p_i(u)$, then the qualitative mapping for the key/value of the large-scale data sets can be described as: $\{[\alpha_i, \beta_i]/p_i(u)\}$. If $[\alpha_i, \beta_i$ can be broken down as $[\alpha_{i1}, \beta_{i1}] \cup [\alpha_{i2}, \beta_{i2}] \cup [\alpha_{i3}, \beta_{i3}]\ldots[\alpha_{in}, \beta_{in}]$, and after shuffling the pairs in the data, the result will be $\{p_i(u))/x_1\}, \{p_i(u)/x_2\},\ldots,\{p_i(u)/x_n\}$,and then after the mapping of the data, the result will be $\{p_i(u)/(x_1, x_2, \ldots, x_n\}$. According to the integrated attributes of the object, a further mapping of the data is conducted, and the following formula (8) is got. If an object has several attributes, then the solution process of the quality criterion of the integrated attribute of the object u is the method of MapReduce.

$$\{p_1/(x_1, x_2, \ldots, x_{m1}) \cup p_2/(x_1, x_2, \ldots, x_{m2}) \cup \ldots \cup p_n/(x_1, x_2, \ldots, x_{mn})\}$$
$$= \{p_1/[\alpha_1, \beta_1] \cup p_2/[\alpha_2, \beta_2] \cup \ldots \cup p_n/[\alpha_m, \beta_m]\}$$
$$= \overset{n}{\underset{i=1}{\wedge}} a_i(u) = a(u) \tag{8}$$

The explanation given by attribute theory is that when things have several quantity features $x_1, x_2, \ldots, x_n$, their corresponding quality features are $p_i(u)$, i.e., $(x_1, x_2, \ldots, x_n) \in [\alpha_i, \beta_i]$, if $x_1, x_2,\ldots,x_n$ is ordered, then $[x_1, x_n]$ is the quality criterion of p(u),that is to say, the mapping process is the solution process of qualitative criterion. After getting $[x_1, x_n]$, the original data pairs can be written as $\{p(u)/[x_1, x_n]\}$,that is the reduction of the original data sets. Therefore the solution of the qualitative criterion of the multi-attributes qualitative mapping can be applied to most of the multi-attributes data in the large-scale data sets. This is the process of MapReduce.
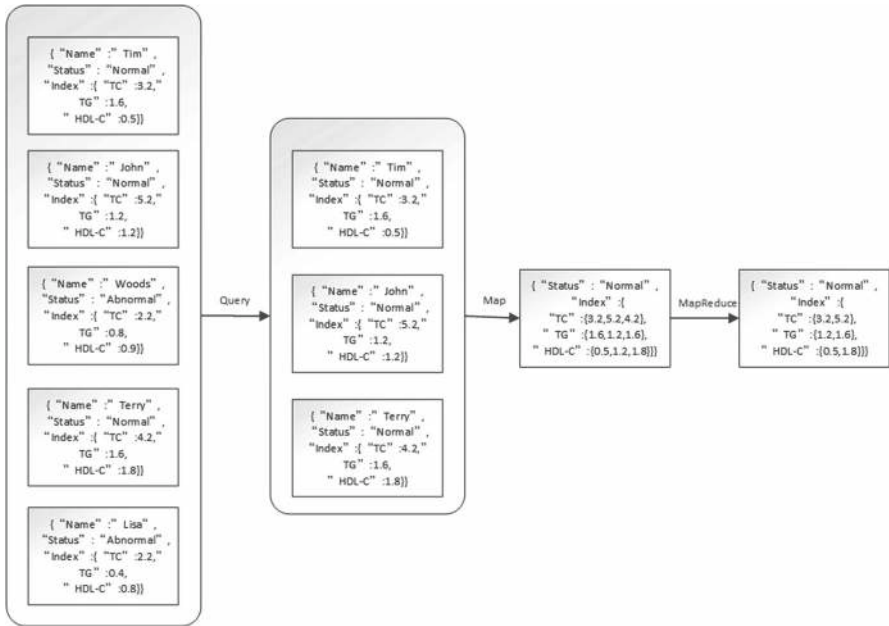
**Fig. 4** The application of integrated attribute qualitative criteria in MapReduce

Based on Example 2, Example 3 is an introduction on how to map-reduce based on the multi-attributes qualitative mapping model. As shown in Fig. 4, in the original data, there are "patients' name" and "blood-fat", among which "blood-fat "is the object attribute. Whether "blood-fat" is normal or not depends on its three sub-attributes, i.e., TC,TG and HDL-C. The three sub-attributes' quality and quantity feature attributes are: $\{p_{tc}(status)/\{x_1, x_2, \ldots, x_n\}\}$, $\{p_{tg}(status)/\{y_1, y_2, \ldots, y_m\}\}$, and $\{p_{HDL-c}(status)/\{z_1, z_2, \ldots, z_h\}\}$.

In the Querying process, we extract all the "normal" data, and in the mapping process, we integrate all the quantity feature data of the three sub attributes TC,TG and HDL-C, and their resluts are "TC": {3.2,5.2,4.2}, "TG": {1.6,1.2,1.6}, and "HDL-C": {0.5,1.2,1.8}}} respectively. Because the data of TC,TG and HDL-C are ordered, then from Formula (8), the integrated qualitative criterion for the normal "blood-fat" attribute can be expressed by Formula (9).

$$p(status) = p_{tc}(status) \wedge p_{tg}(status) \wedge p_{HDC-c}(status)$$
$$= [3.2, 5.2] \times [1.2, 1.6] \times [0.5, 1.8] \tag{9}$$

Since every individual attribute of the integrated attribute is equal, therefore in parallel calculation, the qualitative criteria of every attribute can be broken down into a thread, and Integrated qualitative criteria are got through integrating the solved qualitative criteria of all the attributes.

# 7 Conclusion

The data operation of MapReduce mainly includes sorting, grouping, and joining. After the analysis of Qualitative research and MapReduce, this paper supplies an approach to apply the solution procedure of qualitative mapping to the operation of MapReduce, and two intact examples are supplied in it as well.

# References

1. Dean J, Ghemawat S (2008) MapReduce: simplified data processing on large clusters. Commun ACM 51(1):107–113 **(50th anniversary issue)**
2. Dean J, Ghemawat S (2010) MapReduce: a flexible data processing tool. Commun ACM 53(1):72–77
3. Abouzeid A, Bajda-Pawlikowski K, Abadi DJ, Silberschatz A, Rasin A (2009) HadoopDB: an architectural hybrid of MapReduce and DBMS technologies for analytical workloads. In: Proceedings of the conference on very large databases, Lyon, France, 2009
4. Ekanayake J, Li H, Zhang B, Gunarathne T, Bae S-H, Qiu J, Fox G (2010) Twister: a runtime for iterative MapReduce. In: 19th ACM international symposium on high performance distributed computing, pp 810–818
5. Ekanayake J, Pallickara S, Fox G (2008) MapReduce for data intensive scientific analyses. In: The proceedings of the 2008 fourth IEEE international conference on escience
6. Hadoop Documentation and open source release. http://hadoop.apache.org/core/
7. Ranger C, Raghuraman R, Penmetsa A, Bradski G, Kozyrakis C (2007) Evaluating: MapReduce for multi-core and multiprocessor systems. In: High performance computer architecture, pp 13–24
8. Xu G, Liu N, Duan X (2014) Study of multi-attribute comprehensive evaluation method based on attribute theory. In: Management of e-Commerce and e-Government (ICMeCG), 2014 international conference, pp 6–10
9. Feng J (2006) Qualitative mapping orthogonal system induced by subdivision transformation of qualitative criterion and biomimetic pattern recognition. Chin J Electron 15(4A):850–856
10. Feng J (2008) Attribute network computing based on qualitative mapping and its application in pattern recognition. J Intell Fuzzy Syst 19(4):243–258
11. Feng J (2010) Topos induced by qualitative mapping. In: 2010 IEEE international conference on granular computing, pp 155–160
12. Feng J (2008) Attribute computing network based on qualitative mapping: a kind of model for fusing of artificial intelligent methods. J Comput Inf Syst 4(2):747–755
13. Xu G, Feng J (2007) Pattern recognition method based on the attribute computing network. Comput Sci 35(4):200–203
14. Xu G, Min S, Feng J (2012) Research on multi-agent comprehensive evaluation model based on attribute coordinate. In: Granular computing (GrC), 2012 IEEE international conference, pp 556–562
15. Wang S, Xu G, Du R (2013) Restricted Bayesian classification networks. China Sci 56(7):1–15