SA-Rough Sets K-means Resource Dynamic Allocation Strategy Based on Cloud Computing Environment

Meng Ying\textsuperscript{1}, Luo Ke\textsuperscript{1}, Liu Jian-hua\textsuperscript{2}, Jiang Fei\textsuperscript{2}
\textsuperscript{1}Institute of Computer and Communication Engineering, Changsha University of Sciences and Technology, Changsha 410114, China
\textsuperscript{2}Institute of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410114, China
*corresponding author, email: kf mengying@126.com, luok@csust.edu.cn

Abstract
Cloud computing has received increasingly attention from network computing model research, which can realize several kinds of resource sharing and dynamic resource allocation. However, how to effectively route storage resource in cloud, reduce dynamic load and cross data center data transmission, take into account global load balancing are important problems to be solved. This paper proposes a kind of SA-RS-K means resource dynamic allocation strategy based on cloud computing, which can reduce the time overhead, get the optimal computing resources and improve efficiency of algorithm. Simulation experiments in the end of paper verify the efficiency of this algorithm.

Keywords: resource dynamic allocation, K-means algorithm, cloud computing, dynamic load

Copyright © 2012 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction
Cloud computing [1] is connected through the Internet super computing mode, including distributed computing, parallel computing and grid computing related to technology. In other literature genetic algorithm (GA) technique is also used for loss tracing [2-3]. Cloud computing represents the IT field to the intensive, large-scale and specialization road development tendency, which is the profound change of IT field [4]. However, how to effectively route storage resource, reduce the routing between the dynamic load, considering the global load balancing, to obtain a set of optimum computing resources, which is one of the issues to be resolved. This paper proposes a kind of SA-RS-K means resource dynamic allocation strategy based on cloud computing, which can reduce the time overhead, get the optimal computing resources and improve efficiency of algorithm.

2. Relate Knowledge
IBM technology companies in the white paper" Cloud Computing" in the cloud computing makes a definition [5]: cloud computing is a term used to describe both a platform or a type of application According to the characteristics of cloud computing environment [6], providing equipment scale can according to user demand for resources. The large number of computing resources, storage resources and software resources are linked together, to form the enormous size of the shared virtual IT resources [7]. Cloud model [8] is a qualitative concept and quantitative representation conversion between models, which is assigned the most appropriate computing resources, to maximize the efficiency of the algorithm.

Simulated Annealing Algorithm (referred to as SA) is based on the Monte Carlo iterative strategy one kind of stochastic optimization algorithm [9]. The optimization process of SA algorithm and the physical annealing process similarity, according to Metropolis standards out of local optimal, to obtain the global optimal solution [10]. Rough Set (referred to as RS) is a kind of can deal with fuzzy and uncertain knowledge mathematical tool [11].

K-means algorithm [12-13] using mean error function formula (1) as a convergence criterion, this paper algorithm convergence criterion formula (2):
\[ J_c = \sum_{j=1}^{k} \sum_{i=1}^{n} \| x_i - m_j \|^{2} \]  

\[ RJ = \sum_{i=1}^{m} (w_s \ast \sum_{x \in c_a} d(X - c_j) + w_b \ast \sum_{x \in c_b} d(X - c_i)) \]  

In which: The first belongs to the class center of approximation set for objects to the distance weighted, and the second belongs to the class center of boundary set for the object of the boundary set distance weighted sum.

3. Algorithm Thought

3.1 Clustering center

Set object \( x_i \in C_j \), setting a radius of the circle \( r \), \( c(x_i, r) \) in \( x_i \) as the center for \( r \) the circle inner point number is defined as:

\[ r(x_i) = c(x_i, R) \geq \tau \quad \tau = \sqrt{m} \]  

If formula (1) is established, as the initial center point candidate point set \( X \), as much as possible to avoid the isolated point selected as the initial center of clustering quality. A new clustering center:

\[ c_j' = w_s \frac{1}{|C_s|} \sum_{x \in C_s} x + w_b \frac{1}{|C_b|} \sum_{x \in C_b} x, \quad j = 1, 2, \ldots, k \]  

In which: \( C_s \) and \( C_b \) are the class j approximation set and boundary set, \( w_s \) and \( w_b \) are the class j approximation set and boundary set of the weight, \( w_s + w_b = 1 \).

3.2 Convergence criteria

RS-K-means algorithm using convergence criterion [14]:

\[ RJ = \sum_{i=1}^{m} (w_s \ast \sum_{x \in c_a} d(X - c_j) + w_b \ast \sum_{x \in c_b} d(X - c_i)) \]  

In which: The first belongs to the class center of approximation set for objects to the distance weighted, and the second belongs to the class center of boundary set for the object of the boundary set distance weighted sum. Select the clustering objective function:

\[ F_{RJ} = \min RJ \]  

Selection of annealing:

\[ T(t) = T_0 \ast a^t \]  

3.3 Thought

SA-RS-K means algorithm [15] is the solid to be simulated as the objective function value, the maximum and minimum clustering method to get the K center as the initial solution; then, using rough set to improve convergence criterion for clustering, clustering and the center point is obtained; finally, according to the SA theory, clustering results generated repeated" new center - calculation of the target function - the accept or reject new center ".

4. Algorithm Strategies

4.1 Transmission time overhead strategy

In a cloud computing environment, flow of each task to begin must meet two conditions [16]: (1) the task is dispatched to a data center; (2) the task of all input data sets is located at the local node.
Transmission time overhead strategy of data:

\[
\text{TimeCost}(d_i, d_j) = F_{sm} \times \frac{d_k}{\text{bandwidth}(e)} + C_{ij} \tag{8}
\]

Time overhead can be approximated as:

\[
\text{TimeCost}(d_i, d_j) = F_{sm} \times \frac{d_k}{\text{bandwidth}(e)} \tag{9}
\]

### 4.2 Resource dynamic allocation strategy

The node domain (Slave) as an undirected graph \(G(V, E)[17]\), \(V\) regional area all of a collection of Slave nodes, \(E\) connected to each slave node network collection.

Execution time: time-cost \((e)\), the diameter of \(e\) at the end of the computational resources to process the job expected consumption time.

Network bandwidth: bandwidth \((e)\), the diameter of \(e\) provided by the network bandwidth.

Network delay: delay \((e)\), the diameter of \(e\) produces a maximum network delay.

Resource selection and path is the process of finding satisfies the constraints as small as possible path and resources.

Resource selection constraint function:

\[
\text{Res}(e) = \frac{\text{time - cost}(e) + C_{\text{delay}}(e)}{\text{bandwidth}(e)}
\]

\[\text{s.t.: bandwidth}(e) < TL \]
\[\text{delay}(e) < DL \]

when the various computational resources to complete the job execution, its speed prediction:

\[
EV_{m}^{n}(k+1) = \frac{a_{k+1}}{a_{n}} \cdot ((1-\theta) \cdot EV_{m}^{n}(k) + \theta \cdot RV_{m}^{n}(k)) \tag{11}
\]

In which: A, B, C three constraint conditions of weight; TL, EL and DL as its boundary restriction conditions, \(RV_{m}^{n}(k)\) is M computing resources K the predicted speed of execution. \(\theta\) is adjusting parameters.

### 4.3 Algorithm step

**Step 1:** initialization parameters: setting the initial temperature \(T_0\), by the formula (7) to select the next step of annealing temperature, termination \(T_{\text{end}}\), cycle number \(S\), temperature parameter \(\xi\);

**Step 2:** the variable loop count LOOP=0;

**Step 3:** by the formula (3) no-isolated point to obtain the initial center of the candidate node set \(X\), at the same time by the maximum and minimum distance method and rough clustering to identify clustering number \(K\), and the \(K\) cluster center. by the formula (6) for calculating the current state value of the objective function \(F(K)\);

**Step 4:** compute new cluster center. Calculates new state objective function \(F(k+1): \Delta F = F(k+1) - F(k)\)

**Step 5:** if LOOP<S, else LOOP= LOOP+1, Step3; other Step6;

**Step 6:** if \(T(K) < T_{\text{end}}\), Step7; other \(T(k+1) = \xi \cdot T(k), \text{Step2} \);

**Step 7:** the algorithm terminates, clustering output end, global optimal solution;

**Step 8:** according to formula (9) output time overhead minimum set of nodes, and by using the formula (11) on its velocity prediction, to assign user optimal computing resource.

### 5. Experiment

Experiment environment: operating system Windows XP, Inter(R)Core(TM)2 Duo 2.93GHzRAM3GB, hard drive 320GB, network bandwidth 100MB.
Experiment result: respectively by the literature of [18], [16] proposed algorithm and the SA-RS-K means resource dynamic allocation strategy for comparative analysis, through the comparison across data center data transmission time, time cost and effective nodes distribution rate to compare the superiority of experimental result. Algorithm parameter settings as shown in Table 1, the algorithm simulation results as shown in Figure 1-Figure 4.

Table 1 Algorithm parameters table

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA</td>
<td>Temperature coefficient $\xi$</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>In front of number cool</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>The cooling step °C</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Initial temperature °C</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Adjustable parameter $\theta$</td>
<td>0.05</td>
</tr>
<tr>
<td>GA</td>
<td>Population size</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Cross rate</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>Mutation rate</td>
<td>0.05</td>
</tr>
</tbody>
</table>

![Figure 1. Node sets the effect of time cost](image1.png)

![Figure 2. Clustering center the effect of time cost](image2.png)

![Figure 3. Node sets the effect of transmission number](image3.png)

![Figure 4. Clustering center the effect of transmission number](image4.png)

Figure 1 and Figure 2 show that along with the input node sets and the number of cluster centers number increases, the other two strategies corresponding to the cross data center data transmission caused by the time overhead showing a rising trend; SA-RS-K means resource dynamic allocation strategy corresponding to the time spending growth slowly,
significantly better than the other strategies. Figure 3 and Figure 4 show that along with the input node sets and the number of cluster centers number increases, the other two strategies corresponding to the data centers data transmission times across clearly ascendant trend; SA-RS-K means resource dynamic allocation strategy corresponding to the data transmission times very slightly.

6. Conclusion
This paper proposes a kind of SA-RS-K means resource dynamic allocation strategy based on cloud computing. Compared with other strategies, this paper’s strategy has better performance than others. The SA-RS-K means resource dynamic allocation optimization algorithm can effectively complete the computing resource search and improve the efficiency of the algorithm in cloud computing environment. However, it hasn’t been discussed the node distribution of multidimensional data intensive, which will be the next focus of research.

Acknowledgement
Supported by: National Natural Science Foundation of China under Grant No. 11171095, 10871031, Science and Technology Planning Project of Hunan Province under Grant No. 2011FJ3051, Key Project of Hunan Province Education Department under Grant No. 10A015, and Natural Science Hengyang Mutual Fund of Hunan Province under Grant No. 10JJ8008

References