Location: a feature for service selection in the era of big data

Zhiling Luo
Computer science and technology
Zhejiang University
Hangzhou, China
e-mail: luozhiling@zju.edu.cn

Ying Li
Computer science and technology
Zhejiang University
Hangzhou, China
e-mail: cnliying@zju.edu.cn

Jianwei Yin
Computer science and technology
Zhejiang University
Hangzhou, China
e-mail: zjuyjw@zju.edu.cn

***Abstract*—This paper introduces a service selection model with the service location considered. The location of a service represents its position in the network, which determines the transmission cost of calling this service in the composite service. The more concentrated the invoking services are, the less transmission time the composite service costs. On the other hand, the more and more popular big data processing services, which need to transfer mass data as input, make the effect much more obvious than ever before. Therefore, it is necessary to introduce service location as a basic feature in service selection. The definition and membership functions of service location are presented in this paper. After that, the optimal service selection problem is represented as an optimization problem under some reasonable assumptions. A shortest-path based algorithm is proposed to solve this optimization problem. At last, the case of railway detection is studied for better understanding of our model.**

Keywords-component; service selection; service composition; service location; shortest path; big data

I. INTRODUCTION

Service selection, one of the critical problems of service composition, is generating considerable interests in recent years in several computer science communities. Service selection determines the non-functional qualities of composite service in varieties directions, such as cost, reliability and reputation. A large number of methods for service selection are presented by different researchers and communities. There are a considerable amount of service selection approaches have been proposed by different researchers and communities. Some of these approaches have even already been used in industry.

However, the era of **big data** brings new challenges to service selection. The concept of big data comes from the research on Database, which represents that the size of data involved in computing is huge (Gigabyte, Terabyte, even Petabyte). Unlike being ignored or even considered useless before, in the last decades, the big data generating from producing and trading is considered as a treasure with of the development of data mining technology. As a result, the **big data services** which processes the big data are also brought to the fore now. An example of big data service is the customer mining service, which finds out those potential customers by analyzing the customer relations.

The data loads in big data services are typically very large, which brings a big challenge to classical service selection methods. In the example of customer mining service, all the known customer records should be read at the input, which can be as large as a few TB. When calling these big data services in a composite service, there are huge data should be transmitted in the network. The time cost in transmission process is the **transmission cost**, which is not fully taken into account in classical service selection approaches. Most classical service selection approaches use the response time to identify the transmission cost, which is not suitable in this context now. Because the response time represents the time recorded from user calls until the service responses. It is usually much smaller than the transmission time; especially the size of data is further huge.

In order to select the optimal services from candidate services with transmission time of big data considered, the service location is introduced in this paper. The location of a service is determined by the area the service belongs to. The area is an abstract concept which is defined as a set of services with high speed communication connected. That is to say, a sub network, a cloud, even a computer is suitable to be considered as an area. The transmission cost among two services in a same area is small enough to ignored comparing with those deployed in two different areas. By considering these definitions and assumptions, the optimal service selection turns out to be an optimization problem selecting the most concentrated services.

However, this optimization problem is quite difficult to solve. It is considerable to reduce the original problem to a shortest path selecting problem. An algorithm based on shortest path selection is proposed to solve the original optimization problem. At last, a case about railway detection is introduced in detail to study our model.

The contribution of this paper can be stated as follows:

- *A location aware service selection model.* Differing from classical service selection approaches, the service location is introduced as a critical feature to identify the transmission cost for big data services and the optimal selection problem is converted into an optimization problem.
- *An optimal service selection algorithm.* An algorithm based on the shortest path selection for location aware service selection is proposed. The

algorithm has a linear time complexity and easy to implement.

The rest of the paper is organized as follows. Section II reviews the classical service selection approaches. Section III proposes the system model with the location definition and a few considerable assumptions. Section IV discusses the problem of location aware service selection and introducing the selection algorithm. The case study is presented in section V. Finally, section VI concludes this paper and an outlook on possible continuations of our work.

II. RELATED WORK

There are lots of research achievements in service selection. However, the complexity of the problem makes it hard to find a general efficient method for all situations. In recent years, the quality of service (QoS) based selection is the major research direction. The QoS of a service is a set of features, including reputation, cost, reliability, etc.

Zeng at [2, 3] focused on dynamic and quality-driven selection of services. A global planning method was used to find the best service components for the composition. Mixed linear programming (MIP) [4] was used to find the optimal selection of component services. Ardagna extended the mixed linear programming model to include local constraints [5, 6]. However, the model has the exponential time complexity, which performs poorly with high scalability. Alrifai decomposed the original quality constraint into local constraints [1]. Searching with local constraints has a linear time complexity. However, the decomposition process still has an exponential time complexity. Reference [7] introduces the Web Service Relevancy Function (WsRF) used for measuring the relevancy ranking a service based on QoS. Reference [11] ranks and selects service concerning the reputation and trust. Yu [12] extends the service selection problem with multiple QoS constraints. Shi [14] pays attention to users participating in QoS computation. Two kinds of users (experienced users and novice users) are discussed respectively. The technology of the skyline is also widely used in QoS-based service selection. WS-Sky [15] is a complete service selection framework proposed by Benouaret. His work in [16] pays attention on the Uncertain QoS. Lin concentrated on the context-aware factor that affect QoS attributes in service selection [17]. Meanwhile, many other QoS-based web service selection methods have also been proposed in [8-10].

Tang introduced a Location-aware collaborative filtering method for service recommendation [13]. They concentrate on the physical near of user and service, which is similar to our selection method. However, the majority problem of Tang is service recommendation instead of service selection. The differences between service selection and recommendation are presented in the next section.

These QoS-based selection methods solve the problem of selecting the optimal services for composition. However, these methods cannot solve the problem of location aware selection because the location has two important characteristics differing with the qualities mentioned in classical selection methods.

1. Location is not quantitative. A location of a service is not directly computable. The quantitative value extends from the location is the distance. The distance of any two services is determined when the locations of the two services are confirmed. It is better to understand distance as a relative value. Unlike location, other qualities such as reputation are really quantitative and it is an absolute value.
2. Distance is context-sensitive. There is not an optimal choice without considering the services selected before and after. The lowest transmission cost happens only with all nearby services have the smallest distance. Other qualities are not context-sensitive; it is possible to find an optimal service without considering the service before and after. Take the reputation as an example, the optimal service is the one with largest reputation.

Because of these different features, it is not possible to use the classical methods to solve the location aware selection problem. More details on the definition and membership functions of location are defined in the section III.

III. MODEL

A. Service Selection

Service selection is a critical part of service composition. Generally speaking, the process of service composition can be divided into three steps.

- Building the composite process with abstract services. Unlike the concrete service, the abstract service is the symbol representing a group of services with similar functions and interfaces. The abstract services are composed together by some control statements (such as assignment, switch and

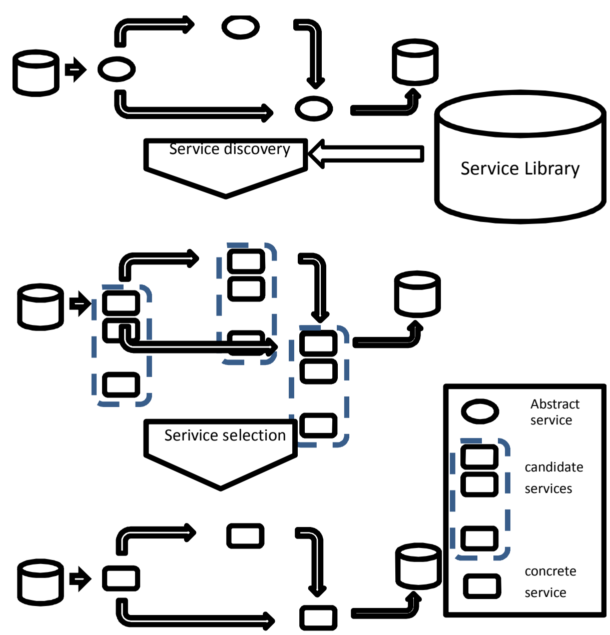

Figure 1: Process of service discovery and selection

loop).

- Finding some suitable services (namely candidate services) for each abstract service. These candidate services have the same functions and interfaces as the abstract service. This step is known as service discovery. The discovery process is mainly searching first k services for each abstract service in the service library with some functional constraints. Service recommendation is very similar, which selects the most suitable services and recommends to users.
- Selecting an optimal service from each candidate service group. Replace the abstract services with selected concrete ones in the composite process. This step is known as service selection. How to select the optimal services with service location considered is the major problem discussed in this paper.

Fig. 1 presents the process of service discovery and service selection. In this figure, the composite service process is presented at the top, which fetches data from a database, calls three abstract services (present as the ellipses) and writes the result into another database. After the step of service discovery, alternative services are picked up from service library and constructed as candidate service groups. The last step is selecting the optimal concrete services from each candidate service group.

In most composite services, the different control processes are used, including sequence process, loop process and switch process. The sample of Fig. 1 contains a switch process. However, the hybrid control processes make it too complex to analyses the optimal selection. To simplify the problem, the sequence process is majority discussed in the rest, and the analysis of switch process and loop process is the future work. An example of sequence process is present in Fig. 2. In order to simplify the following analysis, some math symbols are defined as follows.

A symbol S represents the whole services used in selection, which is constructed by M groups of candidate services. $S = \{S_1, S_2, ..., S_M\}$. Each S_i not only represents an abstract service but also represents a group of candidate services having the same function as the abstract one. $S_i = (S_{i1}, S_{i2}, ..., S_{iN})$. N is the number of candidate services of each group. S can also be represented by a matrix.

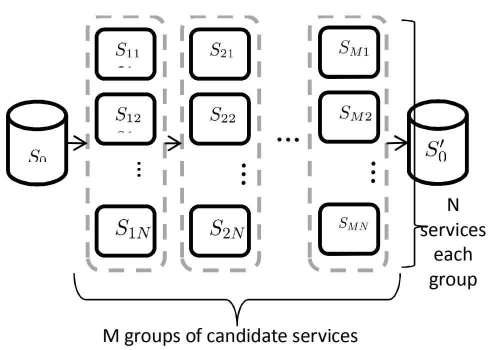

Figure 2: Sequence service process

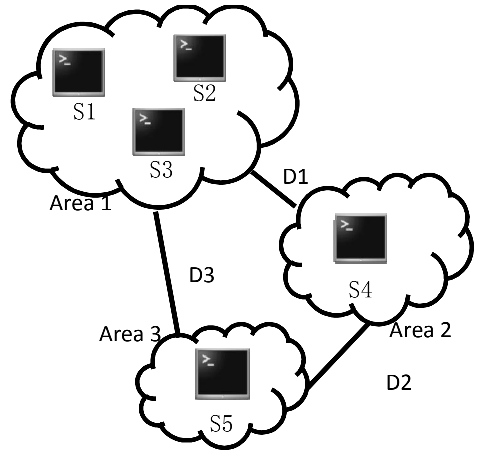

Figure 3: The relation with areas and services

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} \end{pmatrix} \quad (1)$$

There are two special services, namely S_0 and S'_0. They are the services providing the reading and writing operation of the database. S_0 and S'_0 are confirmed before selection.

B. Service Location and Distance

1) Distribution Assumption

The location of a service is the feature representing its network environment. Generally speaking, the distribution of services is not random in the network. A reasonable assumption of the service distribution is proposed.

Assumption I. (The gathering): *Services are not randomly distributed in the network. A certain number of services are naturally gathered as a service set. The services in one set are high speed connected. This services set is denoted as an area.*

The Assumption I can be explained as follows. On one hand, the services provided by the same provider are deployed in the same server or some servers in the same subnet. In this situation, the server and the subnet is an area. On the other hand, a Cloud is an area with lots of services deployed on it by different providers. By our definition, a server is an area, a subnet is an area and even a Cloud is also an area. The only constraint is that the services in the same area are high speed connected. Fig. 3 presents the relation of areas and services. In Fig. 3, S1, S2, S3, S4, S5 represent services. Area1, Area2 and Area3 represent areas. S1, S2, S3 belong to Area1, S4 belongs to Area2 and S5 belongs to Area3. By considering the definition of area, S1 S2 are high speed connect, namely the cost of transmitting data between them is quite low.

2) Location

Considering a context with O areas, $A = \{A_1, A_2, ..., A_O\}$ is used to represent the all areas. The map from the services S to the area A is confirmed before selection. In other words, the relationship of belonging is confirmed for any services. This belonging relationship represents the location of the service. A function (denotes $\mathcal{I}$) is defined to identify the index of the area which the service belongs to.

$$\mathcal{I}(s) = \sum_{i=0}^{n} i \times \{s \in A_i\} \quad (2)$$

A special index function is used above, whose value equals 1 when the proposition in the braces is TRUE, otherwise it equals 0:

$$\{p : boolean\} = \begin{cases} 1 & p \text{ is True} \\ 0 & p \text{ is False} \end{cases} \quad (3)$$

The location of a service is represented by the area it belongs to, in other words, the index of its area, namely $\mathcal{I}(s)$.

3) Distance of Areas

The distance of the two areas is the measurement of the speed of transmission data between the two areas. Since all the areas in the context are confirmed, the distances of each two of them are also confirmed. A matrix D is defined to represent these distances. D_{ij} represents the distance between A_i and A_j.

Table 1 is an example of D representing the distances of areas in Fig. 3. In fact D is just the Distance matrix of the graph which constructed by the areas. Generally speaking, $D_{ij} \neq D_{ji}$. Because the network transmission sometimes (differing from types of network) spends different time in uploading and downloading. The different types of network is not the majority of this paper, so it is considerable to generally use D to represent the distances rather than ensure D is a symmetrical matrix. While in order to simplify our computing, $D_{ij} = D_{ji}$ is assumed in the following samples. For the convenience of writing, a function $\mathcal{D}(i, j)$ is defined instead of using D_{ij} directly.

Table 1: An example of D

D	A1	A2	A3
A1	0	D1	D2
A2	D1	0	D3
A3	D2	D3	0

4) Distance of Services

At last, in order to measure the time cost of transmission data between services, we need to define the distance of two services. Here is a proposition about service distance.

Proposition II (Service Distance): *For any service s, d_i represents the distance between service s and service s_i, which is in the same area as s and d_j represents the distances between service s and service s_j, which is in another area, we have $d_i < d_j, \forall i, j$.*

Proof. *Let's assume that existing a $\hat{j}$ and a $\hat{i}$, such that $d_{\hat{i}} > d_{\hat{j}}$. It contradicts with the definition of area. Because the Assumption 1 guarantees that the services in the same area have a high speed connection. If $d_{\hat{i}} > d_{\hat{j}}$, it is the service s_j that in the same area with s instead of service s_i. It contradicts the condition above.*

If exist a $\hat{j}$ and a $\hat{i}$, such that $d_{\hat{i}} = d_{\hat{j}}$, s_j and $s_i$should either both in the same area with s or both not in. It contradicts assumptions that service s_j is in another area. **Q.E.D.**

Proposition II guarantees $d_i < d_j$, which is a weak condition. Following assumption provides a strong condition.

Assumption III (Strong Service Distance Condition): *For any service s, $d_i$represents the distance between service s and service s_i, which is in the same area as s and d_j represents the distances between service s and service s_j, which is in another area, we have $d_i \ll d_j, \forall i, j$.*

The Assumption III is easy to understand because the transmission cost between the two services in a same subnet is always much smaller than the cost crossing subnets. The strong condition makes it possible to define the distance of any two services in a simple way.

The distance of any two services is defined by the distance of the areas they each belongs to. For service s_{ij} and s_{kl}, the distance function $dis(s_{ij}, s_{kl})$ is defined as follows:

$$dis(s_{ij}, s_{kl}) = \mathcal{D}(\mathcal{I}(s_{ij}), \mathcal{I}(s_{kl})) \quad (4)$$

Since the strong condition, 0 is used to represent the distance between services in the same area. The distance of S1 and S4 is just D1 in Figure 3, while the distance of S1 and S3 is 0.

IV. LOCATION AWARE SELECTION

A. Evaluation Function

With the definition of distance under the assumptions mentioned in section III, the problem of selecting optimal services with lowest transmission cost can be converted into a problem of selecting optimal services with the smallest distance between each two services. Considering the sequence process described in Fig. 4 and the candidate services S defined in Formula (1), the vector $\theta = (\theta_1, \theta_2, ..., \theta_M)$ is used to identify a considerable selected service group. For an element θ_i in θ, the service $s_{i\theta_i}$ is selected as the concrete service for the abstract S_i. ω is the weight coefficient, which is also a vector. The sum of the distances of the selected group θ is defined as an evaluation function as follows.

$$\mathcal{F}(\theta) = \sum_{i=1}^{m-1} dis(s_{i\theta_i}, s_{(i+1)\theta_{i+1}})\omega_i \quad (5)$$

Considering the distance between the first service and the database service S_0, and the distance between the last service and the database service S_0', the complete evaluation function is defined as follows.

$$\begin{aligned} \mathcal{F}'(\theta) \quad = \quad & dis(s_0, s_{1\theta_1})\omega_0 + \\ & \mathcal{F}(\theta) + \\ & dis(s_{m\theta_m}, s_0')\omega_m \end{aligned} \quad (6)$$

The Simple Additive Weighting (SAW) is used here. The ω is the weight coefficient. The weight coefficient really makes sense because the transmission cost is not only related to the transmission speed but also related to the size of data.

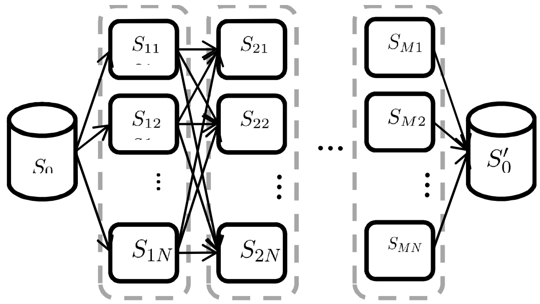

Figure 4: An example of the graph of selection problem

Generally speaking, the size of data in different part of composite process is different. An example is that S_1 read the whole data from the database service S_0 as the input, while output the statistic result of the origin data. The second service S_2 reads the statistic result as the input. In this condition, the distance between S_1 and S_0 has more influence on the whole transmission cost than that between S_2 and S_1. So we can initialize the weight parameter with $\omega_1 > \omega_2$.

B. Optimization Problem

Finding the smallest $\mathcal{F}'$ with a feasible parameter θ is an optimization problem, which can be described as follows:

$$\begin{array}{ll} \inf & \mathcal{F}'(\theta) \\ \text{subject to} & \theta_i \in [1, N] \\ & \theta_i \in \mathcal{Z} \end{array} \tag{7}$$

This formula means the feasible set of parameter θ is constrained by $1 \leq \theta_i \leq N$ and θ_i is an integer. The optimal solution $\hat{\theta}$ satisfies $\hat{\theta}$ belongs to feasible set and $\forall\theta, \mathcal{F}'(\hat{\theta}) \leq \mathcal{F}'(\theta)$. Following theorem presents the relationship with this optimization problem with our original service selection problem.

Theorem IV (Optimization problem): *A location aware service selection problem with service distribution $\mathcal{I}$ and area distance D confirmed, is equivalent to the optimization problem described in (7).*

The proof of this theorem is omitted because it is mentioned in previous sections. More concern is given on the solving of this optimization problem.

Formula (7) is an integer programming (IP) problem, which is a famous NP problem. Generally speaking, there is not an algorithm with Non-deterministic Polynomial time complexity. A famous IP problem is Traveling Salesman Problem (TSP). Because of the features of our model, an algorithm of polynomial time complexity is proposed later.

C. Shortest Path Problem and improved algorithm

A possible way of solving problem (7) is iterating with all the possible combination tried and finding the combination with $\mathcal{F}'(\theta)$ minimized, called the iterating algorithm. However, the time complexity is exponential, which is not usable in reality.

In order to fully take advantage of the structural features, the optimization problem (7) is converted into a shortest path selection problem. The transferring process can be described as the following steps.

1. Iterating all candidate services (including database services S_0 and S'_0) and doing step 2 and 3.
2. Append service s_{ij} in the vertex set V.
3. Iterating all candidate services in the next group. For each $s_{i+1,k}$, Append edge E_{ijk} which is the directed edge connecting from the service s_{ij} to the service $s_{i+1,k}$, in edge set E. The length of the edge E_{ijk} is $dis(s_{ij}, s_{(i+1)j})\omega_i$.
4. Find the shortest path from S_0 to S'_0.

The shortest path problem is a classical topic which has been studied in many researches. Once we can guarantee that the vertices of the shortest path are the optimal selection.

Lemma V. *The number of vertices of a path from S_0 to S'_0 is M+2. The i-th vertex is one of services in S_{i-1}.*

Proof. *The first conclusion is natural because the transferring process described above. If exist a $\hat{\imath}$, that the $\hat{\imath}$-th vertex does not belong to $S_{\hat{\imath}-1}$, then ($\hat{\imath}+1$)-th vertex must does not belong to $S_{\hat{\imath}}$. Because, the edges to vertex in $S_{\hat{\imath}}$ comes from $S_{\hat{\imath}-1}$. We can reduce that ($M+2$)-th vertex is not S'_0, which contradicts with the first conclusion.* **Q.E.D.**

With the lemma, it comes to a useful theorem which guarantees the optimal selection.

Theorem VI (Shortest path). *The vertices of the shortest path from S_0 to S'_0 is the optimal solution of the optimization problem (7).*

The proof is easy by applying the lemma above and the definition of evaluating function and it is omitted here.

With all the preparation theorems, we now can propose the location aware service selecting framework.

1. Finding out the areas of each candidate services belong to and the distances between each area.
2. Allocating the value of weight coefficients ω_i
3. Converting original process into a graph.
4. Finding out the shortest path from S_0 to S'_0.

Because of Theorem VI, the service invoking in the shortest path from S_0 to S'_0 is the optimal solution $\hat{\theta}$. The Theorem IV guarantees that the optimal solution $\hat{\theta}$ is the optimal selection with lowest transmission cost.

The fully algorithm is presented in Fig. 5. In this algorithm, M is the groups of candidate services, N is the number of candidate service each group. S stores not only the candidate services but also the two database services. The members in the first row of S are both S_0, and the members in the last row are both S'_0. D stores the distances of each area and Omega is the weight coefficient.

The first part of the algorithm converts S to a vertex set V. A considerable improvement is for any group contains two or more services in one area, any of them is feasible. So it is better to use only one vertex to represent these services as the code in line 4. The improvement reduces the number of vertices and promote the efficiency.

The second part of the algorithm is the Dijkstra Algorithm. Once the last vertex (namely S'_0) is reached in the iteration, the loop stops.

```
Algorithm 1: Selection(S, N, D, Omega)
    Input: M groups candidate services S, N is the number
of sevices each group, D is the distance matrix and Omega is
the weight coefficient
k=0;
for i = 1:M+2
  for j = 1:N
    if not ExistSameArea( S(i,j))
      V[k] = S(i,j);  //initialing the vertex set V
      V[k].j = j;
      V[k].dist = MAX; //initialing distance
      k++;
    end
  end
end
V[1].dist = 0;
si = FindMinDist(V);
while si > 0
  if IsLast(V[si])  //if V[si] is the last
    F'=V[si].dist;//F' is the shortest destance
    Break;
  else
    for i iterats next group
    if V[i].dist > V[si].dist + Omega[si] *
      D(Area(V[si]), Area(V[i]))
      V[i].dist = V[si].dist + Omega[si] *
              D(Area(V[si]), Area(V[i]))
      V[i].last=si;//recording the vertex in path
    end
    end
  end
  si=FindMinDist(V);
end
for i =m:-1:0 //generating the optimal selection as theta.
  theta[i]=V[si].j
  si=V[si].last;
end
```

Figure 5: Selection Algorithm

The last part is generating the optimal selection as theta.

V. CASE STUDY AND ANALYSIS

A. Case study

1) Introduction

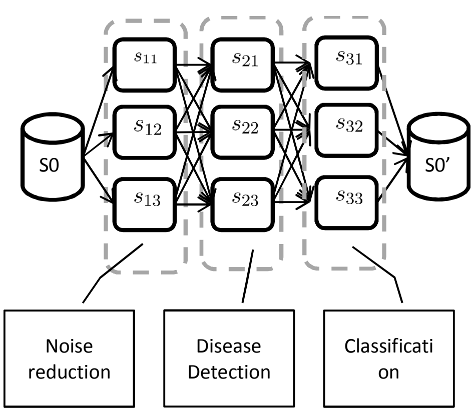

Figure 6: Process of railway detection

In order to understand our model, an example of railway status detection is studied. The railway status detection is a daily work to check the status of each part of the railway to find out the potential disease (the damage and the broken of the railway segment). Daily railway status detection ensures the safety of the railway.

We have studied the data of JGX (Beijing-Guangzhou railway line, China). The detecting train collects data (as a record) 4 times per meter. 24 kinds of data are recorded in a record. They are Meters, Flags, Event, Lprf (mm), Rprf (mm), Laln (mm), Raln (mm), Gage (mm), Can't (mm), Xlvl (mm) and etc. The whole length of JGX is 2300 kilometers. There are 2300*1000*4 = 9200000 records are collected in once detection. The size of data in once collection is nearly 200MB. The frequency of detection varies in different seasons. However the size of data collected a month is nearly 12GB. The whole railway detection data are nearly 450GB.

The data were analyzed manually in the past. It took much human labor. Recently some companies have started to develop professional data analyzing service for the railway department while different service providers are expert in different domain, e.g. some have provided Data filtering services, some have provided Data Encryption services, etc. it is necessary to compose these services together to fit the fully requirement. Railway status detection usually has these steps.

- Collection. Detecting the railway by a special detecting train, collecting variety kind of data, including gauges, warps and etc. this data is as large as a few Gigabytes for a long railway. This step does not belong to data analysis and it can only be done by the railway department itself.
- Noise Reduction. Original data contain many noises caused by the measurement error. It is necessary to reduce these noises before deep analysis. This service is S_1.
- Disease Detection. Analyzing the collected data to find out the potential diseases (the disease means the damage or broken in railway segment), by comparing the data with some special patterns of potential diseases. This service is S_2.

Table 2: Distribution

Service	Belonging Area	Service	Belonging Area	Service	Belonging Area
s_0	A_1				
s_{11}	A_1	s_{12}	A_2	s_{13}	A_1
s_{21}	A_2	s_{22}	A_3	s_{23}	A_4
s_{31}	A_4	s_{32}	A_4	s_{33}	A_1
s_0'	A_1				

Table 3: distance matrix of areas

Distance	A_1	A_2	A_3	A_4
A_1	0	12	22	200
A_2	12	0	24	400
A_3	22	24	0	432
A_4	200	400	432	0

Table 4: distance matrix (N stands for unreachable)

	S0	S11	S12	S13	S21	S22	S23	S31	S32	S33	S0'
S0	N	0	120000	0	N	N	N	N	N	N	N
S11	N	N	N	N	120000	220000	2000000	N	N	N	N
S12	N	N	N	N	0	240000	4000000	N	N	N	N
S13	N	N	N	N	120000	220000	2000000	N	N	N	N
S21	N	N	N	N	N	N	N	400	400	12	N
S22	N	N	N	N	N	N	N	432	432	22	N
S23	N	N	N	N	N	N	N	0	0	200	N
S31	N	N	N	N	N	N	N	N	N	N	200
S32	N	N	N	N	N	N	N	N	N	N	200
S33	N	N	N	N	N	N	N	N	N	N	0
S0'	N	N	N	N	N	N	N	N	N	N	N

- Classification. Different diseases may cause different problems. Some diseases can be ignored, while some diseases may even cause derailment. So it is necessary to classify these diseases into different levels. This service is S_3.

In this example, S_0 is the service providing database containing the collected data. S_0' is the service maintaining database containing different levels of diseases. In reality, S_0 and S_0' are usually the same. But we still assume they are different without loss of generality. The whole composite process of railway detection is represented in Fig. 6.

1) Preparing Service Locations

There are nine candidate services discovered from 4 service providers. They deployed their service on their server cluster. These server clusters are distributed in different areas. The collected database is in one of the four areas.

The distribution of the night candidate services mentioned in Fig. 6 is as follows in Table 2. The distances between each area are listed in Table 3.

1) Allocating ω_i

The value of ω_i is decided by the size of data transmitted between services. In the sample of railway detection, data transferring from S_0 to S_1 are the original data. The data transferring from S_1 to S_2 have the same size as the original data. The data used in S_3 and written into S_0' is just the data of disease points, whose size is only near 0.01% of the whole data. A possible allocation of ω_i is ω_0=ω_1=10000, ω_2=ω_3=1.

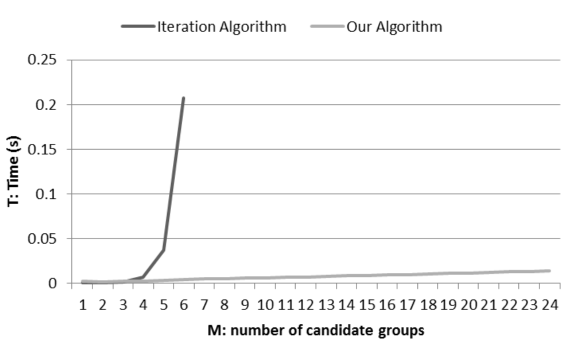

Figure 7 Comparison of algorithms

2) Conversion

It is needed to convert original sequence process into a graph. The vertices of the graph are the candidate services. The edges are the Connection of service s_{ij} with each candidate service $s_{i+1,k}$ in the next candidate service group. The length of the edge is $dis(s_{ij}, s_{(i+1)k})\omega_i$. There are 24 edges: ($s_0$, s_{11}), (s_0, s_{12}), (s_0, s_{13}), (s_{11}, s_{21}), (s_{11}, s_{22}), (s_{11}, s_{23}), (s_{12}, s_{21}), (s_{12}, s_{22}), (s_{12}, s_{23}), (s_{13}, s_{21}), (s_{13}, s_{22}), (s_{13}, s_{23}), (s_{21}, s_{31}), (s_{21}, s_{32}), (s_{21}, s_{33}), (s_{22}, s_{31}), (s_{22}, s_{32}), (s_{22}, s_{33}), (s_{23}, s_{31}), (s_{23}, s_{32}), (s_{23}, s_{33}), (s_{31}, s_0'), (s_{32}, s_0'), (s_{33}, s_0'). The distance matrix the graph is described in follows in Table 4, in which, N represents unreachable.

3) Finding Shortest Path

The algorithm is presented in Fig. 5. The M here is 3 and N is also 3. The shortest path is $s_0->s_{11}->s_{21}->s_{33}->s_0'$. The shortest distance from vertex S_0 to S_0' is 120012. So the optimal choice plan is selecting the first service in group1, the first service in group2 and the third service in group3.

B. *Performance study*

The time complexity of our algorithm is $\mathcal{O}(MN^2)$, where $M$is the number of the groups of candidate services, namely the number of rows in Formula (1), and N is the number of candidate services each group, namely the number of columns in Formula (1). In most conditions the number of candidate services is confirmed, which is decided by the service discovery. Therefore, the time complexity is $\mathcal{O}(M)$, which means it is linear time complexity.

We compared the time complexity of our algorithm with the original iteration algorithm mentioned in part C of section IV. In order to consider the time complexity of the different sequences process, we evaluated the calculating with M changes from 5 to 5000. At the same time, The number of candidate services each group (namely N), the number of areas and the distance between each two areas are fixed. The comparison result is presented in Fig. 7. The dark line represents the iteration algorithm. It is exponential time complexity. The light line represents our algorithm and it is linear time complexity.

VI. CONCLUSION AND FUTURE WORK

In this paper, the service location is introduced as a new feature helping to select the optimal services with lowest transmission cost. The classical service selection methods mentioned in [1-3, 5] concerned with the general QoS evaluation in service selection. Data transmission cost is ignored by the classical service selection methods. However, the era of big data brings a new challenge for service selection. The huge size of data makes the transmission time been a majority cost. In this context, a service selection model with transmission time concerned is needed, which is also the motivation of our research. Location is introduced to describe the network context of a service with two distribution assumptions mentioned in part B of section III. The distance of any two services is introduced to represent the transmission speed between these two services. The evaluation function of service selection is the Simple Additive Weighting (SAW) of each distance between two nearby services in the sequence process. In order to solve this optimization problem, the original data are converted into a graph and the vertices invoked in the shortest path from the data reading service (namely s_0) to data writing service (namely s_0'), are the optimal selection of the original optimization problem, which is also the optimal selection. A fully case about the railway status detection is studied and the time complexity of our algorithm is analyzed. It is $\mathcal{O}(M)$ with the number of candidate services each group confirmed. Our future work will mainly focus on unifying the location aware selection model with the classical QoS aware selection model to propose a more general one and extending our model to other control process, such as loop and switch.

ACKNOWLEDGMENT

This work was supported by National Science and Technology Supporting Program of China (No.2012BAH06F02), National Natural Science Foundation of China under Grant (No.61272129).

REFERENCES

[1] M. Alrifai, T. Risse, “Combining Global Optimization with Local Selection for Efficient QoS-aware Service Composition” *in World Wide Web Conference (WWW 2009)* , pp. 881-890, 2009

[2] L. Zeng, B. Benatallah, M. Dumas, J. Kalagnanam, and Q. Z. Sheng. “Quality driven web services composition.” *In World Wide Web Conference(WWW 2003)*, pages 411–421, 2003.

[3] L. Zeng, B. Benatallah, A. H. H. Ngu, M. Dumas, J. Kalagnanam, and H. Chang. “Qos-aware middleware for web services composition.” *IEEE Transactions on Software Engineering*, 30(5):311–327, 2004.

[4] G. L. Nemhauser and L. A. Wolsey. “Integer and Combinatorial Optimization.” *Wiley-Interscience, New York,* NY, USA, 1988.

[5] D. Ardagna and B. Pernici. “Global and local qos constraints guarantee in web service selection.” *In International Conference on Web Services (ICWS 2005),*2005 pp 805–806,

[6] D. Ardagna and B. Pernici. “Adaptive service composition in flexible processes.” *IEEE Transactions on Software Engineering,* 33(6):369–384, 2007.

[7] E. Al-masri, Q. H. Mahmoud, “Discovering the best web service.” *in World Wide Web Conference Series (WWW 2007)* , pp. 1257-1258, 2007

[8] B. A. Soydan, M. P. Singh. “A DAML-based repository for QoS-aware semantic web service selection.” *In International Conference on Web Services (ICWS 2004), ACM, California, USA*, pp. 368-375, 2004

[9] S. P. Ran. “A model for web service discovery with QoS.” *ACM SIGecom Exchanges,* vol. 4, no. 1, 2003.

[10] Y. T. Liu, A. H. Ngu, L. Z. Zeng. “QoS computation and policing in dynamic web service selection.” *In World Wide Web Conference on Alternate Track Papers & Posters (WWW 2004), ACM, New York*, USA, pp. 66-73, 2004.

[11] L. Vu, M. Hauswirth, K Aberer. “QoS-Based Service Selection and Ranking with Trust and Reputation Management.” *OTM Workshops* , pp. 466-483, 2005

[12] T. Yu, K. Lin. “Service Selection Algorithms for Composing Complex Services with Multiple QoS Constraints.” *Service-Oriented Computing (ICSOC 2005)*, 2005

[13] M Tang, Y Jiang, J Liu. “Location-Aware Collaborative Filtering for Qos-Based Service Recommendation.” *in IEEE International Conference on Web Services (ICWS 2012),* 2012, pp. 202-209

[14] C Shi, D Lin, T Ishida. “User-Centerd QoS Computation for Web Service Selection.” *in IEEE International Conference on Web Services (ICWS 2012),* 2012, pp. 456-463

[15] K Benouaret, D Benslimane. “WS-Sky: An Efficient and Flexible Framework for QoS-Aware Web Service Selection.” *in IEEE International Conference on Web Services (ICWS 2012),* 2012, pp. 146-153.

[16] K .Benouaret, D Benslimane, A Hadjali. “Selecting Skyline Web Service from Uncertain QoS.” *in IEEE International Conference on Web Services (ICWS 2012),* 2012, pp. 523-530.

[17] D. Lin, C. Shi, T. Ishida, “Dynamic Serivce Selection Based on Context-Aware QoS” *in IEEE International Conference on Web Services (ICWS 2012),* 2012, pp. 641-648.