数理モデル化と問題解決 40－8
（２００２．６．２６）

A Quantum Algorithm for Searching Web Communities

Mayumi Oto*,Masami Takata*,Hiroaki Yoshida** and Kazuki Joe*
*Nara Women's University, **Ochanomizu University

Abstract *With explosive growth of the Internet, the importance of the effective choice of web contents rises quickly. Currently the concept of web community is used for analysis of web contents and classification in general, but the search for web community is difficult with a simple keyword scheme. The search for web community is suitable for infinite parallelism that quantum computing potentially provides. In this paper, we apply a quantum algorithm to a conventional search to propose a web community search algorithm. In addition, we show the effectiveness of the algorithm through preliminary experiments. We describe an instruction set for a quantum computer architecture and show an example of quantum computing codes for the proposed algorithm.*

1 Introduction

In the Internet with a glut of information, web contents have been distributed more widely. The effective use of web contents becomes more and more difficult, and importance of effective choice rises. An idea of web community is widely used for an indicator to analyze and classify the circulation situation of web contents [1]. However, discovery of web communities is difficult by keyword search that is current dominant, and search methods of web communities by another approach is investigated as hot spot research topics. Quantum computing is known as a more suitable framework than Neumann type computing for the algorithms of which computational time increases exponentially by the size of data set.

In this paper, we propose a *web community search* (WCS) algorithm for the search of an arbitrary web page. The algorithm combines a typical search technique with a conventional quantum algorithm [2, 3]. We describe a prototype implementation of the WCS algorithm operating an adjacency matrix and evaluate the effectiveness. In the implementation, some part of quantum algorithm is encoded with a quantum computer model and an instruction set for a quantum computer architecture [4]. Finally, We describe the relevance of "hub and authority" and the algorithm.

2 WCS Algorithm and Preliminary Experiment

Using adjacency matrix, we propose an algorithm for searching web communities to which a given web page p_u belongs. We define for the WCS algorithm a web community as a set of web pages linking within a certain depth from a given web page. In the representation of an adjacency matrix of which element has binary (one or zero) value, we define the web community as a set of the elements with value one in the neighborhood of the diagonal elements. Next, we show the effectiveness of the algorithm by preliminary experiments.

2.1 WCS Algorithm

In this paper, we assume a condition that we remove the web pages which have been found as an authority of a web community from the target of the search in order to simplify the algorithm. By this assumption, we make the WCS algorithm to focus on a web page, and search a web community in a directed graph with tree structure defined by the web page as the root node. Among the row of the web page p_u, namely the uth row of the adjacency matrix, we search for authoritative pages. When the total number of web pages is N, the adjacency matrix A are expressed as $A_{ij} = 1$ $(i, j = 1, ..., N)$ where web page p_i has a link to web page p_j, otherwise $A_{ij} = 0$. In each page as well, we keep searching for authority pages by the depth limit given in advance.

When the depth limit is too small for the size of web community, some part of web communities are given. When the depth limit is large enough, all the web pages in a web community are given. We describe the WCS algorithm.

1. Load adjacency matrix A, and input initial web page p_u and the depth limit D.
2. Replace elements of initial web page p_u, namely the uth row and column, with those of p_1 and stitute 1 for i.
3. Using [3], search for j which is 1 from the ith row of adjacency matrix $\sum_{j=1}^{N} A_{ij}$. But except for the page whose flag bit is 1.
4. For all j which are found, exchange a matrix element upper as possible with the jth element. (except matrix elements which have already changed)
5. Set 1 to flag bits corresponding to j.
6. Substitute j for i.
7. Add 1 to variable expressing the depth of search.
8. Go to step10 if there is no value in all elements which follows the diagonal elements of searching the web page, or the search reaches to the depth limit D.
9. Go to step3.
10. Output all web pages which we found and exit.

2.2 Validation

The model of web community adopted by the preliminary experiments is: 1) The total number N of web pages is 1200. 2) There are three web communities and they are connected via links of about $N/20$. 3) Each web community consists of $N/3$, and each web community in the neighborhood of initial web page p_u is c_1, c_2, c_3.

A procedure of validation is: 1)Trade p_u with the top elements of the adjacency matrix. 2)Create three web communities and generate an adjacency matrix so that each community assembles in diagonal elements. 3)Expose elements of the matrix at random. 4)Apply the WCS algorithm. 5)Output the number of pages which we found. Note that states of adjacency matrix from step2 to step4 correspond to ones from state1 to state3.

Next, on the basis of data in states 1 and 3, we show the discovery rate of web community c_1. We calculate the ratio with the number of pages n of c_1 in state1 and the number of pages n_k in state3 of the kth trial in 100 trials and inspect the average value. A value of n is 400 from a model of web community. We define the discovery rate E_k of web community in the kth trial as follows;

$E_k = n_k/n \; if \; n_k < 400, 1 \; otherwise.$

Therefore, average value $\langle E \rangle$ of 100 trials is

$\langle E \rangle = \sum_{k=1}^{100} E_k.$

2.3 Analysis Result

On the basis of data in state3 in once trial, we show the evolution of the number of pages which we found. In Fig.1, we plot the number of pages which we found by depth of search in x-y coordinate till the trial exits and the total number of web pages is 1200.

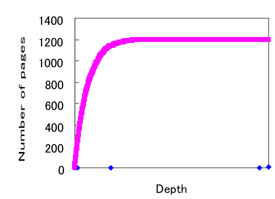

Figure 1: Relation between depth and number of pages

We can find more pages when we make the depth limit larger. However, the number of pages which we can find does not increase too much even if we make the depth limit larger than certain degree.

Next, we show three adjacency matrices from step2 to 4 (from state1 to 3) in Fig.2. We plot a point of 1 in each element of adjacency matrix. The depth limit is 10 in this trial.

As shown in state1, when a set of web pages which is close to the page number forms each web community, elements with value in the adjacency matrix assembles to diagonal elements. State3 is after applying WCS algorithm and web community c_1 gathers in the left corner.

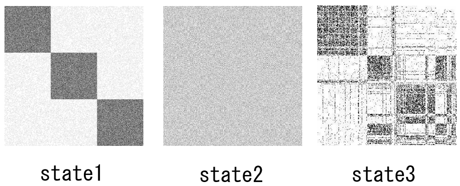

Figure 2: Tidied up web community, Random one, After applied algorithm

In addition, we show the result that executed a procedure of the above trial. Variables which we input are the number of web pages: 1200 and the depth limit: 8. $\langle E \rangle = 100.00...$

The value of this trial is insufficient as an experiment so that setting of the total number of web pages is small. If we give the enough depth limit to the total number of web pages, the algorithm can be effective for the web community search. But when the total number of web pages increases and we set the depth limit larger, the number of steps increases exponentially. From the result of the preliminary experiments, we can predict the effective depth limit considering from the computational amount and resource.

3 Quantum Algorithm

In this section, we show the part of the algorithm that can be applied to quantum algorithm in step3 of 2.1. The algorithm for database search needs $O(N/2)$ steps in Neumann type computing for amount of data N and $O(\sqrt{N})$ steps in quantum computing.

3.1 Quantum Unit Model and Instruction Set

Quantum computer architecture which we already proposed [4] consists of Neumann type units and quantum units. A quantum unit model constitutes quantum register files, quantum initialize register files, QR Selector and quantum ALUs. Before executing the algorithm, we assume all quantum registers are initialized. A quantum register file is a set of quantum registers and each has some qubits, and each qubit has some parameters. We express an arbitrary unitary matrix of two by two with four parameters. QR Selector controls quantum register numbers. The number of qubits that a quantum register has is variable-length. In case of initialization, QR Selector gets the same size area as a quantum register which is specified an operation land to a quantum initialize register and exchanges each register number. A unitary matrix is executed by phase rotation or phase rotation with condition and the product of a unitary matrix is executed by taking the sum of parameters. The value that we observe a quantum register is sent to a Neumann type register. We show an instruction set for quantum computer architectures in Table1 that we define above.

3.2 Quantum Algorithm and Coding

In the WCS algorithm, quantum computation is required in the step3. Quantum algorithm [3] uses [2] and enhances only the amplitude of states to satisfy condition $T[i] = x$ from N elements that the number of solutions is unidentified. We can get one arbitrary solution by observation and all solutions by repeating it. This quantum algorithm is used in the step of searching for 1 from a certain row. The below program is an implementation of [3]. Some parts of the algorithm are not written with instructions because of luck of space.

    Store 1, N-R_m
    Store 6/5, N-R_λ
L7 {Choose k at random}
    Store j, N-R_k
    {Repeat the paper [2] k times.}
    {If the observe value i satisfies $T(i) = x$
    then output i and exit, otherwise set m
    to $\min(\lambda m, \sqrt{N})$ and back to L7.}
    Stop

The below program is an implementation of [2].

    Read N

Table 1: A quantum instruction set

Quantum instruction	Mnemonic	Operation land	Specification of a quantum instruction
Set a register length	QSetLength	Q-R$_i$,N-R$_j$	Set N-R$_j^{\star\star}$ as the length of Q-R$_i^\star$
Initialize a register	QExchange	I-Reg.,Q-R$_i$	Copy and initialize I-Reg.*** to Q-R$_i$
Phase rotate	QRP	Q-R$_i$, θ	Rotate the phase of all qubits θ in Q-R$_i$
Phase rotate with Cond.	QRPS	Cond.,Q-R$_i$,Q-R$_j$, θ	Rotate phase θ corresponding to Q-R$_i$ which only satisfies Cond. in Q-R$_j$
Observe	QObserve	Q-R$_i$,N-R$_j$	Store the observational result of Q-R$_i$ in N-R$_j$
(Neumann type instr.)	CPhase	Matrix,N-R$_i$	Compute the phase of matrix and store it in N-R$_i$

*quantum register i, **Neumann type register j, ***quantum initialize register

```
      Load N,N-R_N
      {Compute minimum n which satisfies 2^n ≥ N.}
      Store n, N-R_n
L1    QSetLength Q-R_1,N-R_n
      QExchange Q-R_1,I-Reg
      Read M
      CPhase M, N-R_M
      QRP Q-R_1,N-R_M
      Read D
      CPhase D, N-R_D
      {Repeat the following till L2 O(√N) times.}
      QRPS "C(Q-R_1)=1", Q-R_1, Q-R_1,π
L2    QRP Q-R_1,N-R_D
      QObserve Q-R_1, N-R_k
      {If C(N − R_k) ≠ 1 then back to L1.}
      Stop
```

4 Discussion and Conclusion

There are two methods for the classification of web communities, one is to arrest web as a set of web pages based on link information, and another is to regard it topologically as the graph which consists of web pages and links.

Our WCS algorithm is equivalent to the latter and forms a tree structure whose top is initial web pages. But we remove web pages from a target of the search if they are found once. If we assume a set of web pages which forms a closed path is a web community, when a certain web page is found for the second time we can find web community in all web pages which we found by expanding the WCS algorithm to memorize paths of web pages.

In addition, classification of authority related to arbitrary topic is difficult because the WCS algorithm does not mention anchor text of each page and weight of links, and trace just links. It is similar in case of classification of hub. With an instruction set for a quantum computer architecture we have already proposed, some part of the WCS algorithm which requires huge computational could be replace by a quantum algorithm.

Incidentally, it may not be tied closely between all communities that we found and the initial web page. It is our future work to calculate the degree of relevance of the initial web pages and web communities we found and to aim at expanding the algorithm to be applied in web communities of various types.

References

[1] Flake,G., Lawrence,S. and Giles,C.: Efficient Identification of Web Communities, 6th Int'l Conf. on Knowledge Discovery and Data Mining, pp.150–160 (2000).

[2] Grover,L.: A fast quantum mechanical algorithm for database search, 28th Ann. ACM Symp. on the Theory of Computing, pp.212–219 (1996).

[3] Boyer,M.,Brassard,G.,Høyer,P. and Tapp,A.: Tight Bounds on Quantum Searching, Fortschritte der Physik,Vol.46,pp.493–505 (1998).

[4] Oto,M., Nakajo,H. and Joe, K.:A Possible instruction set for quantum computer architectures, Int'l Conf. on Parallel and Distributed Processing Techniques and Applications, vol.3,pp.1221–1227 (2001).