

Agent Community based Peer-to-Peer Information Retrieval

Tsunenori Mine[†], Daisuke Matsuno[‡], Koichiro Takaki[‡], and Makoto Amamiya[†]
{Faculty[†], Graduate School[‡]} of Information Science and Electrical Engineering,
Kyushu University, Japan
6-1 Kasuga-koen, Kasuga, Fukuoka, Japan, 816-8580
{mine, matsuno, ktakaki, amamiya}@al.is.kyushu-u.ac.jp

Abstract

This paper proposes an Agent Community based Peer-to-Peer information retrieval method called ACP2P method, which uses agent communities to manage and look up information related to users. An agent works as a delegate of its user and searches for information that the user wants by communicating with other agents. The communication between agents is carried out in a peer-to-peer computing architecture. In order to retrieve information related to a user query, an agent uses a content file, which consists of retrieved documents, and two histories : a query/retrieved document history(Q/RDH) and a query/sender agent history(Q/SAH). We implemented this method with Multi-Agents Kodama[1], and conducted preliminary experiments to test the hypothesis. The empirical results showed that the method was much more efficient than a naive method employing 'multicast' techniques only to look up a target agent.

1. Introduction

The rapid increase of World Wide Web has made conventional search engines suffer from decreasing coverage of searching the Web. The Internet users meet information floods that are forced to filter out and choose the information they need. In order to deal with these problems, a lot of studies on distributed information retrieval, information filtering, information recommendation, expert finding, or collaborative filtering have been carried out. However most systems developed on those researches are, unfortunately, based on the server-client computational model and are often distressed by the fundamental bottle neck coming from its central control system architecture. Although some systems based on the peer-to-peer (P2P for short) computing architecture have been developed and implemented, each node of most those systems only deals with simple and monolithic processing things. Considering these issues, we propose an Agent Community based P2P information retrieval method (ACP2P method for short), which uses agent communities to manage and look up relevant information to a user query. An agent works as a delegate of its user and searches for information that the user wants by

communicating with other agents. The communication between agents is carried out based on a P2P computing architecture. In order to retrieve relevant information to a user query, an agent uses a content file, which consists of retrieved documents with their creator information, and two histories : a query/retrieved document history(Q/RDH for short) and a query/sender agent history(Q/SAH for short). The former is a list of pairs of query and the address of agent that replied relevant documents to the query, where the query was sent by the agent itself. The latter is a list of pairs of query and sender agents and shows "who sent what query to the agent". This is useful to find a new information source. Making use of the Q/SAH is expected to make a collaborative filtering effect emerge and to gradually create virtual agent communities, where agents with the same interests stay together. Our hypothesis is that a virtual agent community reduces communication loads to perform a search. As an agent receives more queries, then more links to new knowledge are achieved. From this behavior, a "give and take"(or positive feedback) effect for agents seems to emerge. We conducted preliminary experiments to test the hypothesis. The empirical results showed that the method was much more efficient than a naive method employing 'multicast' techniques only to look up a target agent.

2. Overview of ACP2P Method

The ACP2P method employs the three-type agents: user interface(UI) agent, information retrieval(IR) agent and history management(HM) agent. A set of the three-type agents (UI agent, IR agent, HM agent) is assigned to each user. Although a UI agent and a HM agent communicate only with the IR agent of their user, an IR agent communicates with other user's IR agents not only in the community it belongs to, but also in other communities, to search for relevant information to its user's query. A pair of Q/RDH and Q/SAH is managed by the HM agent.

Figure 1 shows an example of agent community structure which the ACP2P method is based on. A portal agent in the figure is the agent which is a representative of a community and manages the addresses of all member agents there, where each member agent of a community designates an IR

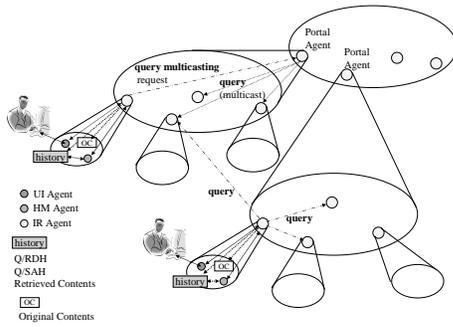


Figure 1: Agents and their Community Structure

agent. When a member agent wants to find any target agents to be sent a query, the agent looks up them using a content file and both history files : Q/RDH and Q/SAH, by calculating a similarity between the query and those files. If the target agents are found, the query is directly sent to them, and their retrieved results are also directly replied to the query sender IR agent. If the request number of such agents are not found, the agent asks the portal agent to send the query to the all member agents in the community by a multicast technique. More details will be shown in elsewhere.

3. Experiments

We use the Web pages of Yahoo! JAPAN (<http://www.yahoo.co.jp>) for the experiments. The Web pages used are broadly divided into five categories: animal, sport, computer, medicinal, and finance. Each of them consists of 20 small categories, which are selected in descending order of the number of Web pages recorded in the category. An IR agent is assigned to each selected category, and thus 100 IR agents are created and activated in the experiments. A category name is used as the name of an IR agent, and the Web pages in the category is used as the original documents of the agent. All agents are realized by describing their functions into plug-in modules of Kodama's application unit.

In the experiments, each IR agent sends 10 queries, each of which consists of a noun extracted from the category assigned to the agent in descending order of their occurrence frequency in the category. To send a query, the agent selects the request number(RN) of target agents. The experiments were carried out to show how both histories help to reduce communication loads between agents to look for relevant information to a user query, and how Q/SAH helps to search for new information sources. We compared the three methods: wQ/SAH (ACP2P with Q/SAH), woQ/SAH(ACP2P without using Q/SAH) and MulCST(Simple method always employing a 'multicast' technique). The results are shown in figure 2. In both cases, the number of exchanged messages in MulCST did not change for every query input, while that of in wQ/SAH and woQ/SAH were being re-

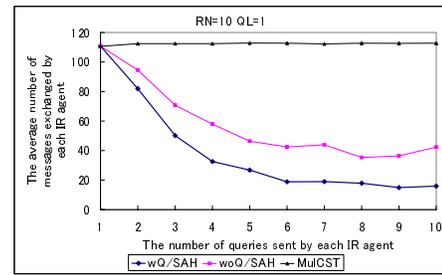


Figure 2: The average number of messages exchanged by each IR agent for each query input

duced. In addition, wQ/SAH showed its better performance than woQ/SAH's.

We also compared wQ/SAH with woQ/SAH from the following point of views: the average number of messages exchanged by each IR agent (30050 vs. 46112, the less is better), that of retrieved documents achieved by each IR agent (14014 vs. 9223, the more is better), that of agents which performs one-way communication with other agents (206 vs. 307, the less is better), that of agents which communicate with each other (326 vs. 144, the more is better). RN was set to 5. The result of wQ/SAH is shown in the left side of the parentheses and woQ/SAH in the right side. The results clearly shows wQ/SAH's better performance.

4. Conclusion

We presented the ACP2P method. The method was implemented with Multi-Agent System Kodama. We conducted the experiments to show how both histories, Q/RDH and Q/SAH, helped to reduce communication loads between agents when searching for relevant information to a query, and how Q/SAH helped to look up new information sources and to create virtual communities. The results showed the efficiency of ACP2P method and the usefulness of both histories for looking up new information sources and for increasing virtual communities, i.e., increasing the number of agents exchanging query messages together.

Acknowledgment

This research was supported by the Telecommunication Advancement Organization of Japan, under the grant to "the Research on Management of Security Policies in Mutual Connection", and by the Japan Society for the Promotion of Science under the Grant-in-Aid for Scientific Research (A) No. 15200002 and (C) No. 16500082.

References

[1] G. Zhong, S. Amamiya, K. Takahashi, T. Mine, and M. Amamiya. The design and application of kodama system. *IEICE Transactions INF.& SYST.*, E85-D(04):637-646, 4 2002.