Intelligent Information Search Mechanism Using Filtering and NFC Based on Multi-agents in the Distributed Environment

Subong Yi, Bobby D. Gerardo, Young-Seok Lee, and Jaewan Lee

School of Electronic and Information Engineering, Kunsan National University,
68 Miryong-dong, Kunsan, Chonbuk 573-701, South Korea
{subongyi, bgerardo, leeys, jwlee}@kunsan.ac.kr

Abstract. Intelligent search agent is popularly used for searching relevant information in the Internet and there are lots of tools that are used to satisfy the needs of the users. Since there is no sufficient cooperation among the agents and they are independent to each other, it is difficult to make an efficient search of information in the distributed environment. Therefore, a typical search agent is difficult to use and can contain irrelevant information for the users. To solve these problems, we use the CORBA architecture to create an agency in the broker agent and provide more reliable information to the users. Also, the proposed intelligent information search system use the NFC and filtering techniques through the multi-agents for fast and reliable search of information.

1 Introduction

Nowadays, we are already experiencing the high-speed networks which are the medium for the Internet and other distributed information systems. Accompanied by the existence of the Internet, lot of information has increased and the work of the users who search for relevant information also increased rapidly. There is much information we can gather on the Internet and some information are not relevant to the user's request, which makes it difficult for the user to search necessary information. It is common that users want to search for information on a fast and efficient way while avoiding the irrelevant information to be processed that can be time-consuming. Many researches proposed efficient search of information. Some developed an intelligent agent for searching of information on the Internet and introduced cooperative multi-agents for searching of information. A push and pull method of information provides an efficient utilization of network resources [4].

Moreover, existing researches use typical agent architecture for searching information from information resources that operate independently and are dependent to each platform [1][2]. The communication and cooperation of agent are inefficient. Consequently, the reliability of information that is gathered in the distributed environment becomes low. Many researches study neural network and it is widely used in research task. One example is an agent which uses neural network that improves the accuracy and reliability of information, but has a longer processing time [5].

In order to solve these problems, we proposed an information search multi-agent system. The proposed system use CORBA to provide relevant information to users

M. Gavrilova et al. (Eds.): ICCSA 2006, LNCS 3982, pp. 144 – 153, 2006.

and transparency of the system. The broker agent creates agency which communicates with the multi-agents to provide information. Internet Inter-ORB Protocol (IIOP) is used for the communications of multi-agent to send multiple packages of messages. In order to acquire accurate information, neural network and fuzzy technique are used. In addition, it uses filtering of the clustered data to reduce processing time. An agent which does the push service and manages the large resources in the distributed environment is also presented. The goals of the agent are to have an efficient resource management and provide fast processing of information.

2 Related Works

Information search in the Internet is a popular topic of research which includes reliability and efficiency. We gathered some related issues about neural networks and the multi-agents and discuss these in the following subsections.

2.1 Neural Networks

A neural network is an interconnected group of artificial or biological neurons. It is possible to differentiate between two major groups of neural networks. It is applicable in every situation in which a relationship between the predictor variables or inputs and predicted variables or outputs exists, even when that relationship is very complex and not easy to articulate in the usual terms of correlations. A few representative examples of problems to which neural network analysis has been applied successfully are in finance, medicine, engineering, geology and physics. There are a lot of researches which uses the neural network in information search. Two different approaches combining fuzzy genetic algorithms and the preprocessing stage of classification called feature selection is used for the information retrieval [1] where the processed document is rank by relevance of the information. A fuzzy clustering based on mobile agent is used for information search [5]. Here, a search mechanism based on ontology using a mobile agent to increase the accuracy rate of information is used. A cooperation of multi-agent is used to provide a reliable search mechanism in the distributed system [6].

2.2 Multi-agent

A multi-agent system (MAS) is a system composed of several agents, capable of mutual interaction. The agents are considered to be autonomous entities such as software programs or robots. Their interactions can be either cooperative or selfish. That is, the agents can share a common goal or they can pursue their own interests. An intelligent information retrieval (IIR) agent is one of a solution to the information overloading [2]. The multi-agent system identifies the desirable features of an IIR agent, including intelligent search, navigation guide, auto-notification, personal information management, personal preferred interface, and tools for easy page-reading. An information filtering approaches that are distributed with respect to knowledge or functionality, to overcome the limitations of single-agent centralized information filtering [3] is used. Large-scale experimental studies like in the Text Retrieval

Conference (TREC) are also presented to illustrate the advantages of distributed filtering as well as to compare the different distributed approaches.

3 Intelligent Information Search Agent (IISA) System

We proposed an Intelligent Information Search Agent (IISA) System, which is an information search system that produces more accurate information and process the information intelligently. Figure 1 presents the architecture of the proposed agent system which consists of three classes. These classes are interface agent, broker agent and resource agent.

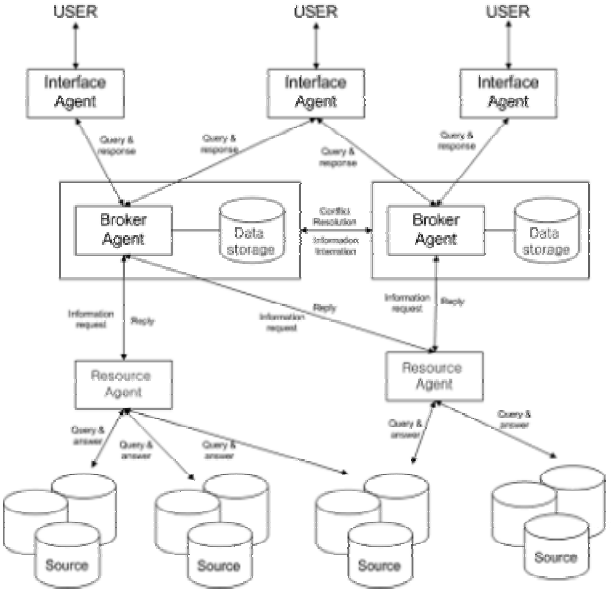

Fig. 1. Architecture of Intelligent Information Search Agent System

The interface agent which is a primary class takes charge of the input of the user for request of search. The resource agent manages the information resource contents. It collects information which is necessary and then sends it to broker agent for extracting the information. The broker agent acts as medium of communication between the interface and resource agent which manages the task of other agents and communicates by query language. In order to respond to the request of the user, the interface agent manages the request for information. The interaction of resource agent and broker agent leads to providing relevant information to the user.

3.1 Interface Agent of IISA

The interface agent interacts with the user and the IISA system. The inputs for the interface agent are the request of the user for information and then it stores the log

information of the user. After the interface agent process the request of the user then it shows the result of the process as output. The user's requests, interests and log information are delivered to broker agent and it processes the query of the user. The proposed system includes a resource agent that uses a data mining technique to gather the relevant data. The information which gets from the broker agent will be presented in a message format or graphical form to become easy to understand by the user.

3.2 Resource Agent of IISA

Resource agent manages the sources of information in the distributed environment. In order to gather the relevant and updated information, it uses the user log information of the interface agent. This log information is delivered to resource agent and it searches updated information about the previous log information. Then, the resource agent pushes the updated information to the broker agent. This is stored by the broker agent in the management table. Also, the resource agent does the resource monitoring in the dynamic distributed environment.

3.3 Broker Agent of IISA

Broker agent solves the heterogeneity and improves the transparency of the distributed environment by using CORBA architecture. One of the important goals of broker agent is to interact with the multi-agent to provide more reliable information. The broker agent creates an agency in order to communicate with the multi-agent agency (MAA). Figure 2 presents the broker agent architecture. After creating the agency, it performs the data mining using Fuzzy Neural Clustering (FNC). Figure 3 presents the process of extracting the information by using the FNC and filtering. In the first process, the fuzzy clustering is performed. The next process is filtering the clustered data by selecting the more interesting data with the user's request. This eliminates irrelevant data and improves the efficiency to extract information. The last process is using the neural network on the filtered data. This uses the self-organizing maps (SOM) to learn from data and then provide information. SOM is a good method to learn from data and to provide information but time processing is longer so we use filtering process to make it more efficient by not adding the irrelevant data in SOM. This approach reduces the processing time.

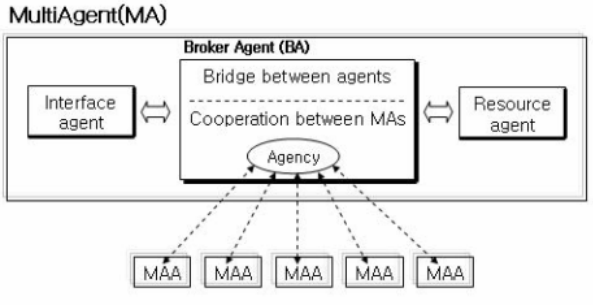

Fig. 2. Architecture of Broker Agent

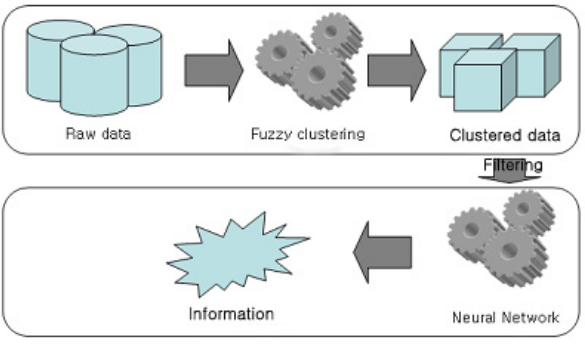

Fig. 3. NFC and filter processing of data

Agency. The agency is used for interaction of the multi-agent to provide data and mining process to produce information. The multi-agent interacts with the agency and it uses NFC and filtering processes. All information is stored in a stack which is the management table. This information will be accessed by the interface agent in case the user request for information.

FNC and Filtering. We present a technique using FNC and filtering of data. Our proposed algorithm enhances the speed of processing compared with the other existing information search technique that uses neural network. Figure 4 illustrates the architecture of the FNC and filtering. First, we use the Fuzzy C-Means (FCM) which is the second layer of the architecture of FNC. After the FCM, we filter the clustered data which occurred between the second layer and third layer. SOM is presented from the third layer until the fifth layer which is used to learn from the data and process it into information. In Equation 1 until Equation 5, we present the formula for fuzzy clustering. Equation 6 presents the formula of filtering the clustered data to be processed in SOM.

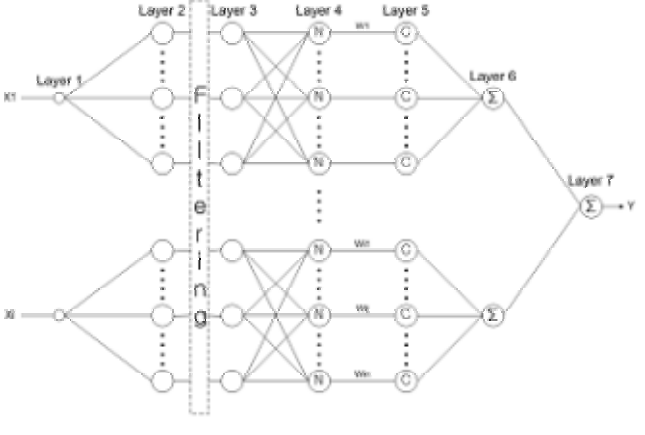

Fig. 4. Architecture of FNC

$$J(u_{ik}, v_i) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}(d_{ik})^2 \tag{1}$$

$$d_{ik} = d(x_k - v_i) = [\sum_{j=1}^{1}(x_{kj} - v_{ij})^2]^{1/2} \tag{2}$$

$$v_i = \{v_{i1}, v_{i2}, \ldots, v_{ij}, \ldots, v_{iL}\} \tag{3}$$

$$v_{ij} = \frac{\sum_{k=1}^{N}(u_{ik})^m x_{kj}}{\sum_{k=1}^{N}(u_{ik})^m} \tag{4}$$

$$u_{ik} = \frac{(1/\|x_k - v_i\|^2)^{1/m-1}}{\sum_{j=1}^{c}(1/\|x_k - v_i\|^2)^{1/m-1}} = \frac{1}{\sum_{j=1}^{c}(\frac{\|x_k - v_i\|}{\|x_k - v_j\|})^{2/m-1}} = \frac{1}{\sum_{j=1}^{c}(\frac{d_{ik}}{d_{jk}})^{2/m-1}}$$

In this point, $d_{ik} = d(x_k - v_i) = [\sum_{j=1}^{1}(x_{kj} - v_{ij})^2]^{1/2}$ (5)

$$DF_i = |V_i - Ix_i| \le F_w(0.2 \sim 0.8) \tag{6}$$

set a number of cluster
c(2<=c<n), m(1<m<∞)
Initialize Make U-matrix

$u_{ik}^{(r+1)} = \frac{1}{\sum_{j=1}^{c}(\frac{d_{ik}^{r}}{d_{jk}^{r}})^{2/m-1}}$ for $I_k = 0$

Calculate the center v
Calculate distance data and cluster
//Update the partition matrix U
Goes through all input sample and calculates membership function
according to the cluster centers.
Calculates the fuzzy objective function values given center and U
Filtering
$F_i = V_i - Ix_i \le F_w(0.2 \sim 0.8$
input a result:
Initialize all weights and biases in network;
while terminating condition is not satisfied {
for each training sample X in samples { //Propagate the inputs forward for
SOM unit j{
$I_j = \sum_i w_{ij}O_i + O_j$;
// compute the net input of unit j with respect to the previous layer, I
$O_j = \frac{1}{1+e^{-I}}$;} // compute the output of each unit j
// for each unit j in the hidden layers, from the last to the first layer
$Err_j = O_j(1-O_j)\sum_k Err_k w_{jk}$;
// compute the error with respect to the next higher layer, k
for each weight w_{ij} in network {
$\Delta w_{ij} = (l)Err_j O_i$; // weight increment
$w_{ij}^{r+1} = w_{ij}^{r} + \Delta w_{ij}$;} // weight update } }
While terminating condition is not satisfied {
similarity set of A and B();
calculate distance from near set();
calculate errors(); }
// compute the similarity using domain
function similarity set of A and B(){
if y≤(Min($U_A(X)$, $U_B(X)$))
$sim_i = U_{A,B}(x,y)$
else }
function calculate distance from near set() {
$S=Max(Value_{A,1}, Value_{B,1}, \cdots\cdots, Value_{c,1})$ }
function calculate errors() {
err = $Value_{c,1}$ $X(A_c \cap A_1 / A_c)$ }

Fig. 5. Algorithm for FNC and filtering

Procedures for FNC and Filtering:

1. Input value of the user by requesting the information
2. Fuzzy clustering using FCM
3. Filtering of the clustered data
4. Result of third procedure is gathered to be processed in SOM for learning of data
5. Calculates the weight value in the hidden layer of the SOM
6. Repeat the fifth procedure in the next hidden layer
7. Update the weight value from the result of the sixth procedure
8. Repeat fifth to seventh procedure until the weight value is more than the critical value. If the weight value is lesser than the critical value then end the procedure.

4 Performance Evaluation

The simulation environment of the proposed system consists of multi-agent and it has object modules for agency. These multi-agents provide information processing through the interaction to each other. To perform the heterogeneity of the system, we use the Solaris 9 and Windows XP operating systems. The information processing is done in the heterogeneous environment by using different operating systems and uses the Visibroker software by Borland which follows the OMG CORBA standards. The programming languages used in the development of the multi-agents are Java and C++. Figure 6 presents the system simulation environment.

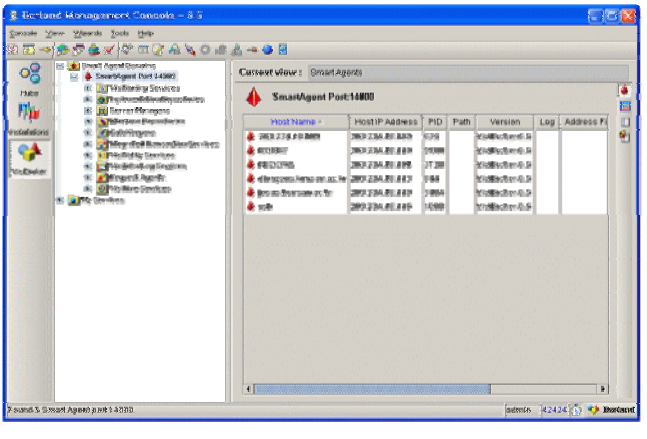

Fig. 6. System simulation environment presented in Borland management console

4.1 Results of Simulations

The simulation used 10 nodes for information processing and we calculate the processing time and accuracy of the information. In Figure 7 we present the result of processing time that has a filter variable set from 0.2 until 0.8. It is observed that when

increasing the filtering value, the processing time will also increase. In Figure 8 we present the result of the accuracy of the information by changing the filtering value in the same manner in Figure 7. To provide more accurate information and to process the information faster, we found the best value of filtering that has at least 80 percent accuracy of information and lesser processing time. Based on the graphical result, the filtering value of 0.5 has satisfied these conditions and will be used for comparison of the other algorithm.

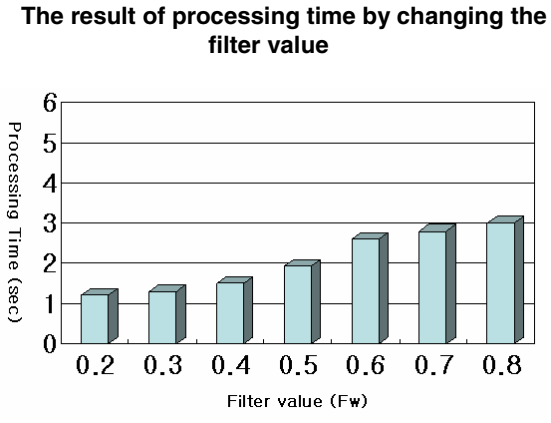

Fig. 7. Results of processing time

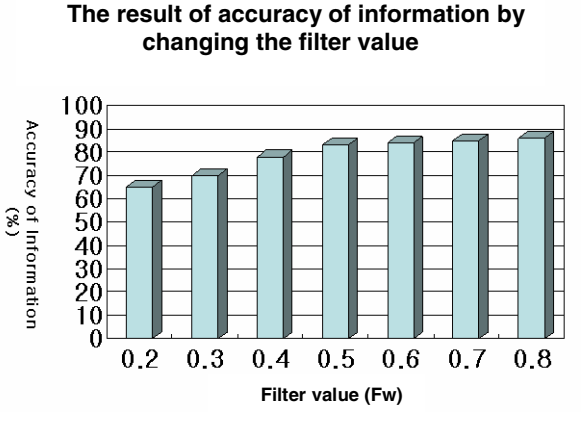

Fig. 8. Results of accuracy of information

We compared our proposed algorithm with SOM and F-NFC. We assumed that the number of nodes is the number of the multi-agents. The simulation used 100 nodes to process the information, setting up the learning rate to 0.5 and perform 100 times of looping value that was applied in all methods. We set the filtering value of the

FNC to 0.5. The result of the simulation is illustrated in Figure 9. The graphical results show that the FNC with filtering has the fastest processing time comparing with the three other methods. It was also observed that increasing the number of nodes will dramatically increase the processing time in SOM and F-NFC because it does not support filtering of data while NFC has lesser processing time because of filtering.

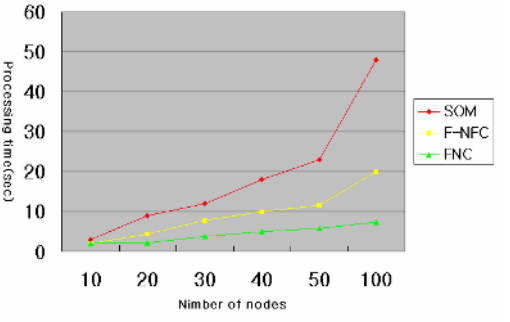

Fig. 9. Comparison result of the processing time by the methods

5 Conclusions

The multi-agent system provides transparency and accuracy service to the user through the interaction with multi-agent in the distributed environment. In addition to multi-agent interaction supports satisfaction of the users by providing more reliable information and faster information processing service.

In this paper, we used CORBA for the multi-agent interaction to solve the heterogeneity of the distributed environment. Also, we proposed a method of NFC with filtering that solved the problems of the existing research which has longer processing time using neural networks. The result implies that using our proposed algorithm can increase the reliability of the information processed and reduced the processing time of the search. The resource agent gathers resources in the distributed environment and performs the data mining. These data are pushed to the broker agent and it creates an agency which interacts with the multi-agent and process the data using the NFC with filtering method. All information is stored in the management table and the user can access the information needed from the management table.

The result of the proposed system simulation is evaluated over 80 percent accuracy of information, which is similar to other existing research but has reduced the processing time and is more effective in case the number of nodes increase.

This research uses an interaction of multi-agent and it applies to the information search system. The future works will be focused on the push service and management table.

References

1. Martín-Bautista, M., Vila, A., Sánchez, D., and Larsen, H.: Intelligent Filtering with Genetic Algorithms and Fuzzy Logic. Technologies for Constructing Intelligent Systems, (2002) pp. 351-362
2. Tu, H., and Hsiang, J.: An Architecture and Category Knowledge for Intelligent Information Retrieval Agents. Proceeding of the Thirty-First Annual Hawaii International Conference on System Sciences, Vol. 4. (1998) pp. 0405
3. Mukhopadhyay, S., Peng, S., Raje, R., Mostafa J., and Palakal, M.: Distributed Multi-agent Information Filtering - A Comparative Study, Journal of the American Society for Information Science and Technology, Vol. 56, Issue 8, pp. 834 - 842
4. Bhide, M.: Adaptive push-pull : Disseminating Dynamic Web Data, IEEE Transaction on Computer, Vol.51, No.6, (2002) pp.652-668
5. Ko J., Gerardo, B. D., Lee, J., and Hwang, J.: Information Search System Using Neural Network and Fuzzy Clustering Based on Mobile Agent, In the Proceeding of ICCSA (2005) pp. 205-214
6. Lee, J., Park, M., Gerardo, B., and Byun, S.: Reliable Information Search Mechanism through the Cooperation of Multiagent Systems in the Distributed Environment. 2nd ACIS International Conference on Software Engineering Research Management & Applications, (2004) pp. 184-188
7. Nürnberger, A., Klose, A., and Kruse, R.: Self-Organising Maps for Interactive Search in Document Databases, Intelligent Exploration of the Web (2002)
8. Park, S., and Wu, C. Intelligent Search Agent for Software Components, Sixth Asia-Pacific Software Engineering Conference (1999) pp. 154
9. Carmine, C., d'Acierno, A., and Picariello, A.: An Intelligent Search Agent System for Semantic Information Retrieval on the Internet, Proceedings of the 5th ACM International Workshop on Web Information and Data Management (2003) pp. 111 – 117
10. Yong S. Choi and Suk I. Yoo.: Multi-agent learning approach to WWW Information Retrieval Using Neural Network, Proceedings of the 4th International Conference on Intelligent User Interfaces, (1998) pp. 23 – 30