

The Application Research of Ant Colony Algorithm in Search Engine

Jian Lan Liu^{1, a}, Li Zhu^{2, b}

¹College of Mathematics and Computer Science, Jiangxi Science and Technology Normal University, Nanchang, China

²College of Mathematics and Computer Science, Jiangxi Science and Technology Normal University, Nanchang, China

^ajjlan0821@163.com, ^bbailan_1982@sina.com

Keywords: Search engine; Ant colony algorithm; Pheromone; Network; Personalized search

Abstract. The Internet will produce large amounts of information every day. How to effectively manage the information resources has become the key to improve the level of network application. Search engine is a software system applied in the web, which is an important tool for retrieving network information. It uses certain strategies to collect and find information, and process and organize the information, and provides users with network information query service, also can help users find valuable and meaningful information through computers. On the Internet, search engine collects different kinds of resources to navigation to facilitate a quick search. This article mainly discusses the concept of search engine. Through the study of ant colony algorithm, achieve further analysis of the application of search engine, and use pheromone binding characteristics to guide personalized search strategy in order to improve search efficiency. Search engine using ant colony algorithm theoretically expounds ant colony algorithm's openness and dynamic adjustment, and on the basis of it, analyzes the advantages of ant colony search engine algorithm and proves the validity and superiority of the algorithm.

I. INTRODUCTION

With the rapid development of Internet, network information presents exponential rise form with large quantity and more and more complex structure types. The information is to meet the diverse needs of different users. In this huge database, how to find the information resources people need becomes a difficult problem, and the emergence of search engine changes this situation. Search engine is an effective tool of access to information and important network application, which helps people find the needed information in the vast sea of information resources. It greatly improves the user retrieval, query and the ability to access to information resources [1]. Just as using filter, it is easy to filter out unwanted information resources and find the resources meeting the requirements. After a period of time people use it, they gradually found that the current general search engine is difficult to meet their individual needs. The retrieval result contains a lot of unnecessary information, and they hope that there is a smarter, faster and more accurate search engine, which can present personalized needs of different users. So search engines need to expand and improve and need to guide practice by theoretical knowledge improved with practice [2]; Search engine, as a kind of query tool for information resources, has become a hot object in the research of computer from the compositions of search engine: different groups, different point of views and different study focuses.

II. THE CONCEPT OF SEARCH ENGINE

Search engines mainly search for and find information through relevant strategies, and extract user query service through a web site after information resources organization and processing [3]; From the perspective of the user, the used web system software not only supplies users with related web interface, but also provides related knowledge or a word by using the browser to finally make a quick feedback to the Internet writing list. The list of each entry represents a class web page, and

each page contains elements such as title, url, keywords, abstract, etc., and some search engines provide more abundant information such as time, file type, file size, and page snapshot, etc.

A. The organization of search engine

Search engine comprises the following three aspects:

(1) Network spider. The main function is to repeat an application phase. Put various pages together to a repository, that is to query again the collected information from a specific web page or site, then according to all the URL of different pages, search the required page.

(2) Indexer function. It is to differentiate and select in different pages and makes a batch of mixed and disorderly calculation according to correlation algorithm. Select the desired web to query the relevant resources and access to the keywords in all page links and eventually connect resources into a web page index database.

(3) The query is to use web page index database to query the key web page to meet the demand of users. All of the information resources for this particular word have some traces in most similar web page index library. According to the review, similar number is got by combining similar information resources and individual pages to complete the screening. Finally feedback the users with the combination of system query sites as well as web information through web's every page links.

B. The working principle of search engines

(1) Scrap pages from the Internet. Repeated the process, and integrate it to the server based on the relevant pages checked and mainly uses Internet spider program integrated on the Internet to check the network, and query related pages according to web URL.

(2) Establish index database. Pages through system integration present the queried resources. The use of this algorithm is to realize batch mixed and disorder computing to get the same degree in links, and finally unite them into a page data.

(3) Search in index database. When a user types words retrieval in the search box, according to the system process, retrieve the required pages according to pages retrieval.

(4) Deal search results with whole sequence. According to the information needed of the particular term queried in index database, combine page rank and similar information data for filter, and feedback the information needed to the user and make detailed link url search and web document combination. Figure 1 is search engine's working principle diagram.

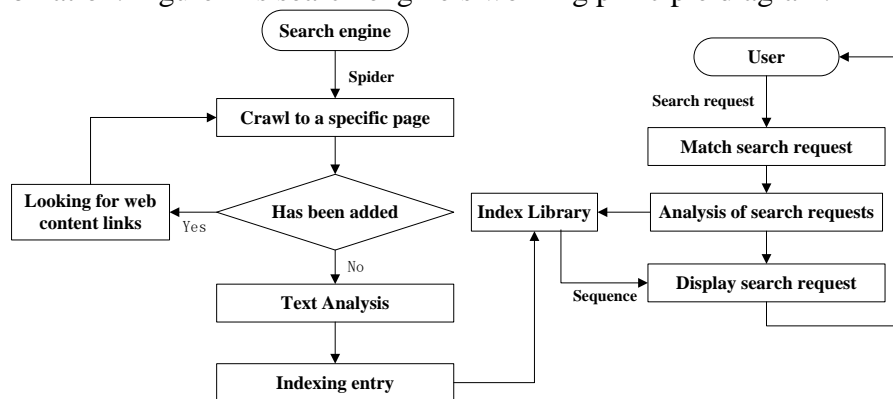


Figure 1 Search engine's working principle diagram

III. THE APPLICATION OF ANT COLONY ALGORITHM IN SEARCH ENGINE

For the application of search engine, a single server cannot meet the needs of users for efficiency, so search engine usually adopts multiple servers distributed services. Each server has its own index database, and different servers can visit each other for convenient retrieval [4]. Ontology search engine can realize the personalized search, but for large number of users, it is difficult to meet the personalized needs of different users. Users' search requires not only to return satisfactory retrieval results, but also to meet the requirements of efficiency. It is not practical for servers to store huge amount of user data, so in the process of retrieving information, optimize retrieval time to meet users' demand for efficiency. Ant colony algorithm has a good characteristic and can better solve

the problem.

A. Ant colony algorithm

Ant colony optimization, which is also called ant algorithm, is an algorithm got from natural phenomenon. It is a kind of probability type technology used for finding optimal path in the picture. Ant colony algorithm is a kind of positive feedback. It can be seen from the process of real ants foraging, the ants will eventually find the shortest path, which directly depends on the accumulation of pheromones in the shortest path, and the accumulation of pheromones is a positive feedback process [5]; for ant colony algorithm, in the initial time, there is exactly the same information hormone in the environment. Give the system a small perturbation to make the concentration trajectory on each side not identical. The solutions of ant structure are of advantages and disadvantages. The way of feedback the algorithm uses is to leave more information hormones in the path optimal solution goes through, and more information hormones attract more ants. This process of positive feedback makes initial difference get constantly expanding, and after a period of time, the whole ant colony will tend to use the path with high pheromone concentration. This path is the "shortest path" for the ants to the food [6].

B. The working principle of ant colony algorithm

Zoologists put forward that ants have no vision. They release pheromone in the process of looking for food, and the pheromone is an "important tool" of communication between ant colony, but the released pheromones will reduce with the increase of the paths; When the ants find food, they will inform other ants along the path to return to nest, and other ants will choose larger probability path to walk with more pheromones; Although residual pheromone on the path will evaporate as time pass by, but a lot of ants pheromone of choosing this path will be continuously accumulated, and other paths' pheromone will evaporate after a period of time, and this path becomes "the shortest path".

As shown in Figure 2: set A is the nest, "E" is food source, HC is an obstruction on the path of ant colony, and ants can only go from A to E from point H or point C, or from E to A. The distance between each point is shown in the Figure. If there are only 30 ants go from A to B in per unit time, there are 30 ants go from E to D, and ants left hormone content quality on the route is 1. Set pheromone volatilization time as 2, and ants move speed is 1. At the initial time, because there is no pheromone on the path of BH, BC and DH, DC, ants located in B and D can choose path randomly. As can be seen from the statistical point of view, they can determine BH, BC and DH, DC in the same probability as their own searching line; Due to the route distance of BCD is only half of that of BHD, after a unit time, there will be 30 ants to meet on H point, and there are 15 ants respectively in B and D and the pheromone density value on the path BCD is 30, while the density value of BHD is only 15; At this point, a group of ants comes to point B and point D, and there will be 20 ants getting to C respectively from B and D. Only 10 ants get to H respectively from B and D; with time goes by, the greater is the probability of ants choosing the path BCD, and finally complete path BCD so as to find the shortest path, namely the ants nest reaches food source passing the obstacles.

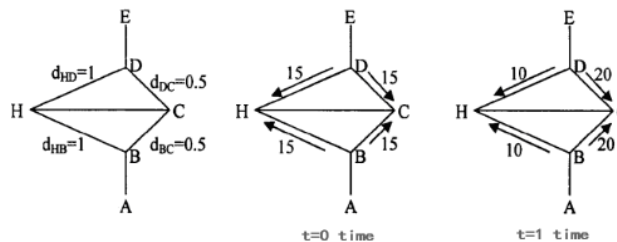


Figure 2 The working principle of ant colony algorithm

From the above analysis, ant colony method's basic idea is: if an ant is designated, it needs to decide how to get the food source from different lines and there is a bigger chance to select the route by leading ants' choice. The shorter is the route, the thicker is the pheromone. This also means to find an optimal solution.

C. The realization process of ant colony algorithm

In ant colony algorithm, the cooperation between ants can better understand the chaotic optimal conditions, and any ant can find a solution. Ant colony can't be independent to contact, but they can use pheromones to lead ants to achieve the ant colony method process with a single ant function. To complete ant colony method, make ant colony function through pheromones indirect communication.

In the process of this algorithm, the following four special features must be finished by ants themselves:

(1) Local search strategy based on probability: in usual usage, if the ant can transfer to another point according to one point, each ant can realize the solution of a problem with restricted moving steps; each step in the process of ants moving, decide the position of the ants moving by the local search strategy of probability. This approach mainly adopts ants' memory, pheromone concentration and local information.

(2) Ants' memory: Ants' memory records a large number of information resources, which adopts some optimization combination, and clearly guides ants to set up relevant solution. According to the role of ant colony, in related applications, through the role of ant colony, memorize every place of the pace so as to revoke ant colony searching different address repeatedly, and can also grasp the relevant information.

(3) Released pheromone: ants mainly complete tasks by using pheromone. In ant group algorithm, the pheromone can produce the effect of positive feedback. Through concrete problem feature selection, we can know when ant colony can release pheromones and how many is the pheromone amount. When releasing pheromone amount, build a solution and form a positive relationship, namely pheromone amount ant colony released and ant colony found options' capability.

(4) Ants decision table: select table with the consistence of ants use pheromones function and heuristic information function and ants choose the direction by using the probability method. Through ants moving, take retrieval methods in order to improve pheromones play mechanism, namely forbidding ant colony soon convergence or premature.

Ant method mainly adopts ant colony single role to form solution. Release pheromone modulation, and finally find the optimal solution. The process of ant colony algorithm is shown in Figure 3:

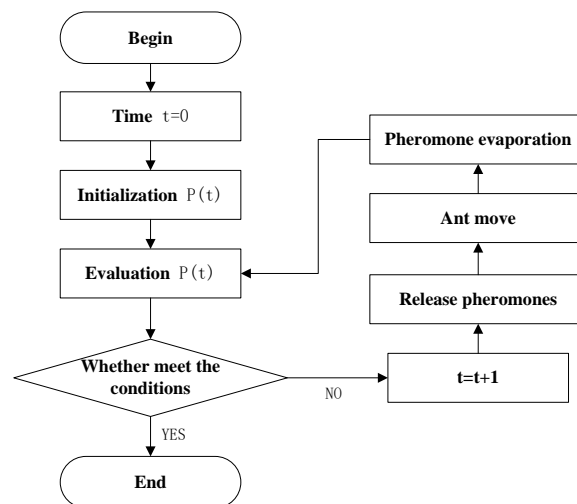


Figure 3 Ant colony algorithm's flow chart

(a) $t = 0$: at the beginning of ant colony algorithm, the initial time value is $t = 0$.

(b) Initialize $P(t)$: initialize the ant colony algorithm.

(c) Evaluate $P(t)$: evaluate each individual ant.

(d) $t = t + 1$: it does not meet the conditions, and get to the next time unit.

(e) Release pheromone: released pheromone according to the corresponding fitness, and fitness and the released pheromone amount is proportional.

(f) Ants move: the individual ant in the back chooses path to more according to the pheromones previous ants left behind.

(g) Pheromone volatilization: with the passage of time, the pheromones are violating.

D. Ant colony search algorithm performance analysis

If search engine is to provide the best results in the shortest possible time to the user, ant colony algorithm is adopted to research. Each optimization team throughout the Internet is constantly looking for "food (the most effective way to optimize)", and then finds the most effective way. Then share it with other Internet "ants", so we can always find a way to improve the site keywords ranking. Numerical simulation results show that ant colony algorithm is of effect and applied value of a new kind of simulated evolutionary optimization method [7].

In polling search algorithm, the server user requested get search connections with each server in the network system one by one, and request return to the corresponding search resources. Compared with the traditional polling searching algorithm, ant colony search engine algorithm has the following advantages: ant colony search algorithm avoids the requested server's needs for each server in the network search system for polling search, thus reduce the connection number the requested server need to establish at the completion of the search process and reduce the search load. In local circumstances, with a large number of search queries, the optimization effect has double ascension.

M is used for the number of servers in the system. For a single server, instant user request number is N. Compare the connections number NC of two local search algorithms.

Polling search algorithm: $NC = N * M$

Ant colony search algorithm: $NC(ant) = N$

Obviously, ant colony search algorithm reduces partial connection number by M times, and in practice, when there are a lot of local users' requests; M times reduce the connection times produced locally, and reduce the connection burden for a single server. Better use the entire network and avoid too much local data processing; at the same time ant colony search algorithm dynamically optimizes search cost.

IV. SIMULATION AND TEST RESULTS

The purpose of using multiple pheromone ant colony algorithms is to improve the efficiency of personalized search. Improve efficiency is under enough high precision premise. With higher precision ontology search engine, improve efficiency and use ant colony algorithm to improve search time of ontology search engines, Baidu and Google and make comparative analysis of the keywords under "network" ontology. The result is shown in Figure 4:

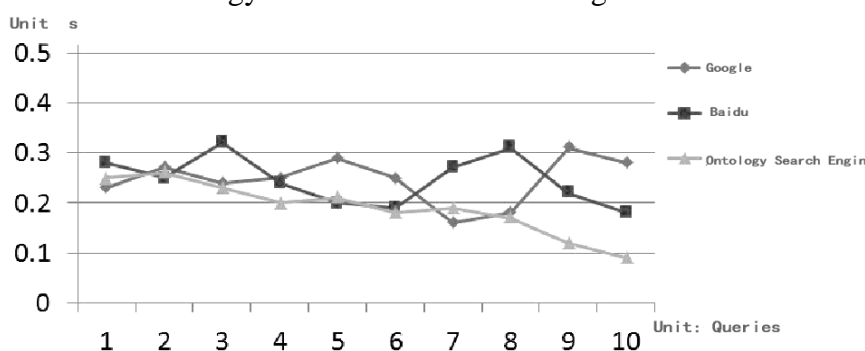


Figure 4 Keywords retrieval efficiency comparison graph in "network" ontology field

The graph shows the contrast of retrieval efficiency. Google and Baidu optimizations are doing very well, so in the Internet huge user group, the retrieval efficiency is also quite good. But because retrieval efficiency has relation with retrieval words and word processing, general search engines retrieval efficiency such as Google and Baidu have bigger difference for efficiency. The retrieval time is usually controlled between 0.2 s and 0.3 s, and the retrieval efficiency of ontology search engine is gradually increased with the increase of retrieval number with the use of pheromone ant colony algorithm, and finally reaches a stable value. It can be seen from the graph that for a keyword search with ontology search engine in a domain, the retrieval time will gradually reduce and contrast shows that multiple pheromone ant colony algorithm has an improved effect of

ontology search engine efficiency significantly.

V. CONCLUSION

The development and application of the Internet have caused profound changes to the way of human economic activities. As search engine service market expands, its users are more and more, and the application field is broader; Search engine accords with the requirement of users accurate finding information with its convenient use way and fast information positioning; Search engine system must have the basic functions such as collect information, process information, and provide inquiry services [8]. We study ant colony algorithm application in search engine, detailed introduce the concept and principle of ant colony algorithm, and the application in search engine, and verified based on simulation experiment platform, search engine system structure and search algorithm fundamentally determine the efficiency and service performance of the whole system. The growing demand of information and information retrieval in the network constantly challenge the efficiency and service performance of the search engine system. In order to solve users' looking for needed information from the information ocean, ant colony algorithm is applied to search engine to well realize retrieval efficiency improvement.

REFERENCE

- [1] Liu Bo. Ant colony algorithm improvement and applied research [D]. Yanshan University, 2010.6
- [2] Hu Hui. The robot three-dimensional path planning technology research based on improved ant colony algorithm [D]. Zhejiang Normal University, 2011.6
- [3] Wu Zexin. Introduction and advance of search engine optimization [M]. People's Posts and Telecommunications Press, 2014.07
- [4] Jin Nan. SEO search engine real explanation [M]. Tsinghua University Press, 2014.2
- [5] Geng Dongshan. Robot global path planning based on ant colony algorithm [D]. Zhengzhou: Zhengzhou University, College of Electrical Engineering, 2010
- [6] Shi Huayu. The application of improved ant colony algorithm in actual VRP research [D]. Shandong, Shandong University, 2012.11
- [7] Qiu Lili. Robot path planning based on improved ant colony algorithm [D]. Donghua University, 2015.1
- [8] Yuan Jinsheng, Li Qun. Search engine basic tutorial [M]. Tsinghua University Press, 2010.07
- [9] Zhou Likun, Liu Hongzhao. Adaptive ant colony algorithm application in three-dimensional wheeled robot path planning [J]. Journal of Mechanical Science and Technology, 2013
- [10] Tan Min, Wang Shuo. Robotics research progress [J]. Journal of Automation, 2013
- [11] Chen Yi. Enterprise search engine marketing decision model and its application research [D]. Henan, Henan University, 2014.5