Emerging social search paradigms: can we learn from ants?

Alessio Malizia[1] Kai A. Olsen[2]

[1] Computer Science Department, Universidad Carlos III de Madrid, Spain
alessio.malizia@uc3m.es
[2] University of Bergen and Molde University College, Norway
kai.olsen@himolde.no

Abstract. In a world of an abundance of data, calculation of relevance is crucial for all search engines. Google's success with their PageRank algorithm proved this. A fundamental part of PageRank is to prioritize pages that are linked to by many other pages. But "relevance" is difficult to compute. While Google's definition works well and satisfy many users it has the disadvantage that it is in principle based on a static view of the Web. Important sites will be more important, and many new and interesting sites may get a low priority. To avoid this problem we aim at describing relevance in a social and dynamic way. The idea is taken from biology: from the way ants forage for food.

Keywords: Applications of Bio-inspired Algorithms, Bio-inspired Computational Paradigms, Emergent Behaviors, Web search algorithms.

1 Introduction

Traditionally, search engines determined relevance based on the number of occurrences of each search term in the document, for example in relation to document size. However, search term occurrence is not a good indicator for relevance. Google's PageRank algorithm [1] saved the day. Now relevance is determined by the number of links to a site, higher if these sites also have a high ranking.

While the PageRank algorithm functions well and offers a notion of "relevance" that is shared by many users, it has the disadvantage that it is static. It may be enhanced by other data and other techniques, but is in principle based on the current structure of the Web. The algorithm will even freeze this picture, as the high ranking will make the important sites more important. New interesting sites may get a low PageRank value and be presented further down on the search engine results page, and may therefore be noticed only by the most persistent users. Thus, the algorithm that is used to determine relevance may be self-fulfilling.

To make Web searching more dynamic we could try to exploit the experience of these "persistent" users. This could be done as easy as presenting an other-users-found-this list. That is, when other users have made a trail to a site, we can offer this as an aid when a user types in similar search terms.

To indicate that a Web site is interesting we could use a feature recently introduced by Google, the Google+1 service. This is similar to Facebook's "like" button. The next user giving the same or similar search terms could then go directly to this site, following the other-users-found-this link.

In this paper we shall explore this concept and present a model that describes a trend towards new Web-searching paradigms, which are both social and dynamic. The idea is taken from biology: from the way ants forage for food.

2 Ants

Ants offer two interesting ideas. The first is to make signals that define optimal trails towards resources, the second is the way they can organize supercolonies. Both of these aspects may aid us in designing better search systems.

Pheromone

Each day ants will leave the colony in search for food and building materials. They will exploit the surroundings in all directions in a somewhat random fashion. If an ant finds anything of interest, it will return to the colony depositing pheromone on the trail, a chemical substance that ants are able to detect. These signals define the trail between the colony and the food.

The quantity of pheromone deposited, which may depend on the quantity and quality of the food, will give other ants an idea of "relevance." Other ants in the colony may now use the pheromone as trail markers to reach the food. But these may also deposit pheromone. Since these markers evaporate over time, uninteresting and unused trails will disappear. Shorter trails will get a higher level of pheromone, thus inserting a notion of optimization.

Supercolonies

Normally, ants from different colonies exhibit aggression toward each other. However, some ants exhibit the phenomenon called unicoloniality. Here worker ants freely mix between different colonies, as described in [2]: "Unicolonial ants carry polydomy [multiple nests in a supercolony that all individuals rotate through] and polygyny [multiple queens in one nest] to extremes. Colonies are huge, each being a network of hundreds or thousands of nests, each with multiple queens. There is no worker aggression, and there is free movement among nests on a vast scale. The energy that might have been put into fighting and territoriality flows into the common good, more ants".

These ants clearly put group interests ahead of individual interests. Just one other specie have this ability, yes – us. This form of organization, known as supercoloniality, may be used to characterize social behavior on the Web, as shown in the following example.

3 Web searching – learning from ants

Let us assume that a set of users all start with the same query, for example "compact camera GPS". That is, they are all interested in finding Web sites that can offer a good bargain for such a camera ("food"). Our group may start with a Google query, and click on links to explore the results. These click streams will define our "pheromone" or virtual trails. This may, for example, be indicated by adding score values to each link, or visualized by representing the links by large fonts, stronger color, etc.

However, on the Web we can optimize, by leaving the trail metaphor and leading subsequent users directly to the "food" sources. We let our ants (the users) explore the Web, but we let them deposit the pheromone on the most interesting pages. The rest of the colony, other users with similar interest, can then go directly to these sites.

Users searching for a GPS camera may be characterized into different groups. For example, a hiker may have different camera interests than a real estate salesman or a professional photographer. Similarly, a Norwegian wanting to buy a camera may not be interested in links to US sites, where high shipping costs may apply. Thus the other-users-found this list may be presented for one colony, e.g., hikers, for a supercolony, e.g., Norwegians, or for all users.

4 The model

Within the ACO (Ant Colony Optimization) research field, the analogy with real world ants' colonies has been used to solve optimization problems, like the Travelling Salesman Problem as described by Dorigo et al. [3]. The ant metaphors have been used to create agents to explore the solution space before feasible solutions have been found with heuristics.

Our approach consists of modeling users' behavior as virtual ants looking for a solution space ("food") by using trails left by others. In the ACO field the ants make probabilistic decisions based on specific problem information, for instance the specific input data for the task. In our case the goal of users (the colony) is modeled by information-seeking behavioral theories such as information foraging, introduced by Pirolli and Card in [4]. The Information-seeking theory models how people seek, navigate and consume information; showing that there are many possible strategies to employ when searching for information on the Web. In our model users browsing for digital information replace the ACO ant agents.

Let us assume that a set of users start with the same query, e.g. a set of keywords, submitted to the Web page of a search engine. Their successive behavior can be modeled as if they were building a feasible solution to their information-seeking problem, i.e., optimizing the route to the information goal. We model this situation with virtual trails as we described in [5]. These trails may lead directly to the goal, or there may be false information, for example where previous users have followed dead-ends before making a decision. We can model this situation by a pheromone factor $p \in [0,1]$. Without trail information, which is the current situation in search engines, the forager will be presented with no cue. This corresponds to $p=0$, i.e., ignoring the information about trails. In the ideal situation we would have $p=1$, i.e. precise information about trails leading to users' goals. Information on trails will increase information scent and lead the user to feasible solutions.

Users exploring information over the Web can be modeled using a construction graph G. We can assume that users follow available information and makes rational choices depending on their information target. Each user can be visualized with branches and connections over the graph G with associated trails, each trail encoding a history of search processes. We assume that user picks out the next available branch by making a decision on alternative trails taken by others, e.g. maximizing the information scent.

When a user selects a branch (link) and thus adding it to the current solution, information on the trails associated with the branch or the corresponding connection will be updated. In ACO this is called online step-by-step pheromone update. In addition we can add pheromone to links that the user finds especially interesting, explicitly (by hitting a "like"-button) or implicitly (by noting her activities on the page, e.g, using time, exploring sub-links, printing or buying). So in our virtual trails model we assume users, if provided with an adequate information scent, to follow the meta-heuristic:

```
Procedure VirtualTrails metaheuristic
     ScheduledActivities
            ShowVirtualTrails()
            ManageUserActivity()
            EvaporatePheromone()
     End ScheduledActivities
End VirtualTrails metaheuristic
```

In the ACO-like meta-heuristic for virtual trails the first step of activities consists of providing information for minimizing false alarms and show virtual trails made by others with pheromone information. The second step manages users choices and add branches to the construction graph according based on user acitivity. The last step updates pheromone evaporation.

5 Discussion

Gone are the days when we browsed the Web, read HTML pages and sent links by email to friends. Today we click on buttons and decide that we "like it", “tweet/retweet it”, “recommend it”, and “tag it”. That is, we are leaving pheromone signs about a Web page or content.

Recently, Google introduced their Social Search service declaring: “with these changes, we want to help you finding the most relevant information from the people who matter to you”[1]. That is, in a way, “people who matter to you” may be defined as a colony. The mechanism is the Google+1 button, which let users share interesting pages with their contacts - a way of releasing pheromone.

Bing, Microsoft’s new search engine, employs Facebook’s social graph for each user to rank search results and to present search history. They define the colony as our own Facebook contacts. Again, we deposit pheromone by a click on the ‘”like” button. Twitter Search is another example fitting our model. Here the colony is made up by people we are “following” and by our “followers”. Twitter has introduced a pheromone-like mechanism called “Top tweets”, i.e. the tweets that are popular over a period of time. This can be seen as a way of implementing evaporation of pheromone.

However, as we suggest the ants may help us to get even further. By using both explicit and implicit information, and combining this information with the key words offered to a search engine, we may offer a general enhancement to Web searching. The advantage is that the user still initiates the query, that is, the information provided will be relevant to her needs. This is often not the case with “like” buttons, especially if these go outside the close colony of friends. Thus we may use information from supercolonies without adding to much noise or Spam.

6 Conclusion

We have presented an idea of enhancing Web searching by learning from ants. The pheromone that ants deposit on the trail to food or other resources can be emulated in computer systems by “like” buttons and similar techniques. Some of these ideas are already implemented by Google, Microsoft, Facebook, Twitter and other companies, where we can recommend Web contents to our friends. However, as ants we can employ a supercoloniality principle, and exploit also the suggestions and experiences from other users. These cases show the paradigm shift in Web searching that we are experiencing today. We hope that our approach can be a first step in modeling and describing these trends. The model is explanatory and descriptive but we think that it could also be predictive (when we started to work on it there was no Google+1, neither “Top Tweets”).

Acknowledgment

This work has been developed under Goodle (n.195538) project funded by The Research Council of Norway and UrThey (TIN2009-09687) project funded by Spanish Ministry of Science and Innovation.

References

1. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web, Technical Report, Stanford InfoLab, 1999.
2. Helanteräl, H., Strassmann, J.E., Carrillo, J., Queller, D.C. (2009) Unicolonial ants: where do they come from, what are they and where are they going?, Trends in Ecology and Evolution, Volume 24, Issue 6, June 2009, Pages 341–349.
3. Dorigo, M., Maniezzo, V., Colorni, A.: Ant system: optimization by a colony of cooperating agents, IEEE Transactions on Systems, Man, and Cybernetics-Part B , vol. 26, n. 1, pages 29-41, 1996.
4. Pirolli, P., Card, S.:Information Foraging, Psychological Review, Vol 106(4), Oct 1999, 643-675.
5. Olsen, K., Malizia, A.: Following virtual trails. Potentials, IEEE, 29(1):24-28, jan.-feb. 2010.

[1] http://googleblog.blogspot.com/2011/02/update-to-google-social-search.html