

An Approach for Personalized Social Matching Systems by Using Ant Colony

Luziane Ferreira de Mendonça

Department of Computer Science, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
Email: luziane@dcc.ufrj.br

Received December 21, 2013; revised January 21, 2014; accepted February 11, 2014

Copyright © 2014 Luziane Ferreira de Mendonça. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for SCIRP and the owner of the intellectual property Luziane Ferreira de Mendonça. All Copyright © 2014 are guarded by law and by SCIRP as a guardian.

ABSTRACT

Personalized social matching systems can be seen as recommender systems that recommend people to others in the social networks, with desirable skills/characteristics. In this work, an algorithm based on Ant Colony is proposed to solve the optimization problem of clustering/matching people in a social network specifically designed for this purpose; during this process, their personal characteristics and preferences (and the degree of importance thereof) are taken into account. The numerical results indicate that the proposed algorithm can successfully perform clustering with a variable number of individuals.

KEYWORDS

Ant Colony; Social Matching Systems; Recommender Systems

1. Introduction

Nowadays in many situations we are faced with the need to relate people. For example, in recent years social networking sites have become very popular.

Many of them are designed to group your users (or create a recommendation list) according to their preferences and goals: sites for dating, language learning, exchanges of experiences in travel, research development, etc. For the last ones, strategy of grouping should take into consideration the goal of each member (or compatibility between a pair of them) and also all the desirable characteristics of the group members (as diversity of knowledge, experience and skills).

The performance of these sites depends on the users satisfaction, which is directly related to the quality of the recommendations and the number of attempts made by the user, until achieve success.

In this work a new formulation for the problem of recommending one or more people for each individual of a network is proposed, in order to reduce the number of tries made by users. This approach can be used in any situation that is necessary to create a recommendation list, not only for dating.

The main idea is to create groups of 2 or more components, taking into account compatibility between 2 people (discrepancy between personal characteristics and the desirable ones from the other person) and the heterogeneity of the group (discrepancy between the personal characteristics among group individuals) [1].

These are two components of objective function in the optimization problem proposed, which is solved using an algorithm based on Ant Colony heuristic (ACO) [2].

This paper is organized as follows. Section 2 describes the formulation that combines homogeneity and heterogeneity of the groups. Section 3 shows the proposed algorithm. Section 4 contains the results of numerical experiments. Finally, Section 5 presents some conclusions.

2. Recommendation Lists

Customized social grouping systems can be seen as recommender systems, which point to a specific person other members of a social network instead of recommending products [3].

The people recommendation is much more complicated than the products recommendation [4], since two people have the same characteristics not always be rec-

ommended to each other. That is, personal characteristics may not be the person's preference, and the immediate consequence is that most of the techniques of recommendation may not be appropriate [5].

Moreover, the number of members of such social networking has grown rapidly in recent years, which has generated a considerable growth of computational complexity. As an example, one of the largest social networking sites in Brazil has over 30 million registered users¹ and an average of 1000 new members every day.

In general, a function that measures the similarity between the desirable characteristics of an individual in question and those submitted by other members of the network is used to generate the recommendation list. Then, this same individual receives the recommendation following the order of similarity: people with greater similarity have priority in the recommendation [6].

However, even considering people classified in the first places of similarity (in relation to same person), they may be similar to each other. The consequence is a possible successive recommendation of people with the same characteristics; the user can feel unmotivated after several attempts without dating successfully [3].

One way to increase the chances of success during the recommendation process is to consider the recommendation of several network members at once (a group), taking into account the similarity between the relevant characteristics of a person and desirable another one, and the discrepancy between the relevant characteristics of two people from the group.

Thus, if the social network is related to a dating site, the chance of finding a suitable person to start a relationship increases. If the site intends create discussion groups, these groups are more likely to generate a high level debate, with different points of view.

Formulation

A social network can be represented as a complete graph with N nodes (individuals). Each node i has three sets of CN informations/features: the individual's ones (vector u_i), the desirable in the partner (vector v_i), and the importance of these features (vector w_i , with values² in the range $[0, 1]$).

The components u_i and v_i in this work represent attributes such as height, weight, age, education, gender, etc., with normalized values.

Without loss of generality, consider that the recommendation will be made to the person 1. A recommendation list with NR individuals can be represented by a closed path³ that passes by at node 1 and has $NR + 1$

nodes. Therefore, at each edge (i, j) we can associate a variable x_{ij} with value 1 if the edge belongs to the path X , and 0 otherwise.

Using these vectors is possible to build a function that will dock compatibility and heterogeneity of a group, taking into account the result of the interaction between each pair of individuals in this group.

The compatibility between node 1 and node j is inversely proportional to the weighted discrepancy ($C(X)$) between the desirable characteristics of 1 and features of j (and *vice versa*).

$$C(X) = \max_{i,j} \left\{ x_{i,j} \left\| \omega_1 * (v_1 - u_j) \right\|; x_{i,j} \left\| \omega_j * (v_j - u_1) \right\| \right\}$$

Thus, a suitable list of recommendation will be build when function $C(X)$ is minimized.

The heterogeneity of a recommendation group can be measured as the discrepancy between the characteristics u_i and u_j of all group's members, and can be estimated as $-H(X)$, where

$$H(X) = \max_{i,j} \left[-x_{i,j} \left\| v_1 - u_j \right\| \right]$$

Accordingly, to generate the recommendation group, the following minimization problem must be solved.

$$\begin{aligned} \min \quad & \theta C(X) + (1-\theta)H(X) \\ \text{s.t.} \quad & \sum_{i=1}^N \sum_{j=1}^N x_{ij} = NR + 1 \\ & \sum_{j=2}^N x_{1j} = 1 \\ & \sum_{i=1}^N x_{ip} = \sum_{j=1}^N x_{pj}, \quad \forall p = 1, \dots, N \\ & \sum_{i=1}^N x_{ip} \leq 1, \quad \forall p = 1, \dots, N \\ & x_{ij} \in \{0, 1\}, \quad \forall i, j = 1, \dots, N \end{aligned} \tag{1}$$

where θ is a weighting factor, which reflects the importance of each portion of the function.

The constraints ensure that: the group has NR people (besides the node 1); node 1 belongs to group; generated path is connected and passes through each node at most once (respectively).

The proposed model suits different types of recommendation. For example, by adopting $u_i = v_i$ and $\alpha = 1$, the solution will bring together nodes with the same characteristics. If the social network is aimed to connect people with common interests to discuss about a specific subject, the discussion group will have greater propensity to present diversity of ideas ($\alpha = 0$), generating more interesting debates.

Such formulation ensures that the group produces a highly attractive list recommendation, not only to con-

¹Site ParPerfeito, <http://www.parpfeito.com.br/>, January 2013.

²Values close to 1 indicate high relevance of the feature.

³An immediate consequence is that there are several different representations for the same group.

template the desirable characteristics of member 1, but also being formed by people with heterogeneous characteristics.

However, it should be noted that a same set of nodes can be connected in many different ways. Hence, the proposed formulation has several global minimizers. Furthermore, problem (1) is a constrained min-max problem, non differentiable, which has many local minimizers [7].

Because of these characteristics and the combinatorial nature of the problem (NP-complete [8]), it is necessary to use heuristics for its resolution.

3. Proposed Algorithm

3.1. Ant Colony Optimization

The proposed algorithm for solving the problem (1) is based on the search for the optimal using an adapted Ant Colony heuristic (*Ant Colony Optimization*, ACO) [9,10].

Ant colony optimization algorithm (ACO) meta-heuristic (which is inspired by the behavior of real ants) is one of the best-known examples of swarm intelligence systems, that can be used to find approximate solutions and near-solutions to intractable discrete optimization problems [2,9]; this probabilistic technique can be reduced to finding good paths through graphs.

The artificial ants incrementally build solutions by moving on the graph. The solution construction process is stochastic and is biased by a pheromone model (a set of parameters associated with graph components, whose values as modified at runtime by the ants).

3.2. Adapted Heuristic

A social network can be represented as a complete graph.

To solve the problem (1), each artificial ant k walks the graph, from node i to node j according to a probabilistic rule. The artificial ant moves successively, culminating in a closed path that corresponds to the connections among members of the same group. Thus, each path (with node 1) corresponding to a potential solution to the problem (1).

Ants deposit pheromone on the path in a quantity proportional to the quality of the solution represented by that path. This indirect form of communication (stigmergy) focuses de search around the most promising parts of the search space.

There is also a degree of pheromone evaporation, which allows some past history to be forgotten, to diversify the search. Therefore, when all the ants have completed a solution, the trails are updated by

$$\tau_{ij} \leftarrow (1-\rho)\tau_{ij} + \sum_k \Delta\tau_{ij}^k \quad (2)$$

where $0 \leq \rho \leq 1$ is the amount of pheromone deposited

on a state transition (i,j) , $0 \leq \rho \leq 1$ is user-defined pheromone evaporation coefficient. The amount of pheromone deposited is obtained by

$$\Delta\tau_{ij}^k = \begin{cases} Q/L_k & \text{if ant } k \text{ uses edge } (i,j) \text{ in its tour} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where L_k is the cost of the k -th ants tour, and Q is constant.

Besides the pheromone, the probability of linking node i to node j depends on the attractiveness η_{ij} on edge (i,j) . The probability is obtained deterministically or randomly according to a pre-fixed parameter $0 \leq \gamma_1 \leq 1$.

$$p_{ij}^k = \begin{cases} \frac{(\tau_{ij})^\alpha (\eta_{ij}^k)^\beta}{\sum_{p \in R_i} (\tau_{ip})^\alpha (\eta_{ip}^k)^\beta} & \text{if } \gamma_2 \leq \gamma_1 \\ \gamma_3 & \text{otherwise} \end{cases} \quad (4)$$

where R_i is the set of nodes that can be connected to i (complement of a tabu set I_k); the parameters α and β control the influence of pheromone (trail) and attractiveness, respectively; γ_2 e γ_3 are random numbers in $[0,1]$. The construction of the tabu set I_k takes account the maximum total tour length of an ant.

The attractiveness is the factor responsible for the grouping nodes taking into account the goal; it has high value when the pair of nodes promotes an decrease in objective function, and low otherwise.

In order to solve this problem, the attractiveness depends on the current path of the ant. In this way, the same edge (i,j) has different values of attractiveness for different ants

$$\eta_{ij}^k = \theta e^{-\max\{\|\omega_1^k(v_j - u_j)\|, \|\omega_2^k(v_j - u_i)\|\}} + (1-\theta)\|u_i - u_j\| \quad (5)$$

if j belongs to I_k (the set formed by the nodes already chosen by ant k), and 0 otherwise. The second term is inspired by heterogeneity and the first one is inspired bu compatibility (with an adjustment to result in positive values, since the smaller value for the attractiveness is 0).

The ant system performs an outer iteration where m ants construct in parallel their solutions, thereafter updating the trail levels. The performance of the algorithm depends on the correct tuning of all parameters, including the initial trail level τ_{ij} . The algorithm is the following.

Algorithm 3.1

1. (Initialization)

Initialize τ_{ij} , for all i,j .

2. (Construction)

For each ant k (state 1 initial) do

repeat

compute η_{ij}^k for all j (by means of (5)).

choose in probability the state to move into (by means of (4)).

append the chosen move to the k -th ant's set tabu I_k .
 until ant k has complete its solution.
 end for
 3. (Trail update)
 Compute $f_k = \theta C(X_k) + (1-\theta\alpha)H(X_k)$ for all $k = 1, \dots, m$, where X_k is the path described by ant k .
 Set p the ant which has the lowest image (current best solution).
 For each ant p move (i,j) do
 compute $\Delta\tau_{ij}$ (by means of (3)).
 update the trail matrix.
 end for
 4. (Terminating condition)
 If not(end test) go to step 2.
 The stop criterion is given by a pre-fixed number of outer iterations without find a new best solution.

4. Numerical Results

Tests were performed for $N = 20, 50$, and 100 users; preferences and their relevance (as well as personal data) to 5 features were generated randomly:

Feature 1: height (145 to 220 cm);

Feature 2: weight (45 to 150 kg);

Feature 3: age (18 to 90 years);

Feature 4: gender (0—female; 1—male);

Feature 5: education level (0—elementary school, 1—high-school, 2—undergraduate, 3—graduate, 4—post-graduate).

For each data set, tests were performed by using 11 values of θ (0, 0.1, 0.2, ..., 1), which is a weighting factor in the objective function. This factor is responsible for balance the compatibility with node 1 and the hetero-

geneity of the list generated.

The parameters used during implementation were $\gamma_1 = 0.05$ (rate-deterministic probability), $\alpha = 1$ (exponent of pheromone—function probability), $\beta = 2$ (attractiveness exponent—function probability), and $\rho = 0.1$ (evaporation rate of the pheromone). Maximum number of outer iterations was 10000, and the stopping criterion was a maximum number of outer iterations (100) without changing current optimal solution. Number of ants per inner iteration was 100.

In all tests, the solution is a recommendation to user 1 of 5 others users.

In **Tables 1-3**, as the value of θ increases, the value of $H(X)$ increases (heterogeneity decreases), while the $C(X)$ decreases (compatibility increases).

The solutions $\theta = 0$ and $\theta = 1$ correspond to users with greater heterogeneity among themselves and greater compatibility with user 1, respectively.

The proposed algorithm had a good performance, in the robustness and efficiency sense: the algorithm converged in all proposed scenarios, always finding a global minimize with a low execution time.

The algorithm has good performance even considering a larger number of users. In this case, one can use a parallel implementation of the Ant Colony adapted to reach fast execution time [11].

5. Conclusions

In this work a new formulation was proposed for the problem in order to provide a recommendation list to a user, in a social network. An Ant Colony algorithm was adapted to solve the problem; this algorithm was robust and it had a good behavior in all performed simulations, finding global minimizers.

Table 1. Numerical results— $N = 20$.

| θ | Recommendation list | C(X) | H(X) | Image | Outer iter. |
|----------|---------------------|----------|-----------|-----------|-------------|
| 0 | 2, 3, 5, 8, 10 | 2.772806 | -0.984773 | -0.984773 | 1176 |
| 0.1 | 2, 3, 5, 8, 10 | 2.772806 | -0.984773 | -0.609015 | 1341 |
| 0.2 | 2, 3, 5, 8, 10 | 2.772806 | -0.984773 | -0.233257 | 1027 |
| 0.3 | 8, 10, 13, 15, 17 | 2.078833 | -0.806916 | 0.058808 | 1045 |
| 0.4 | 6, 10, 13, 15, 17 | 2.078833 | -0.806916 | 0.347383 | 1121 |
| 0.5 | 6, 10, 13, 15, 17 | 2.078833 | -0.806916 | 0.635959 | 1056 |
| 0.6 | 8, 10, 13, 15, 17 | 2.078833 | -0.806916 | 0.924534 | 1068 |
| 0.7 | 8, 10, 13, 15, 17 | 2.078833 | -0.806916 | 1.213109 | 1067 |
| 0.8 | 8, 10, 13, 15, 17 | 2.078833 | -0.806916 | 1.501684 | 1024 |
| 0.9 | 6, 10, 13, 15, 17 | 2.078833 | -0.806916 | 1.790259 | 1102 |
| 1 | 8, 10, 15, 16, 17 | 2.078833 | -0.528589 | 2.078834 | 1001 |

Table 2. Numerical results— $N = 50$.

| θ | Recommendation list | C(X) | H(X) | Image | Outer iter. |
|----------|---------------------|----------|-----------|-----------|-------------|
| 0 | 8, 9, 21, 43, 50 | 3.809949 | -1.019148 | -1.019148 | 1103 |
| 0.1 | 9, 10, 21, 28, 33 | 3.337757 | -1.044343 | -0.606133 | 2003 |
| 0.2 | 8, 9, 21, 28, 43 | 3.337757 | -1.058345 | -0.179125 | 2302 |
| 0.3 | 9, 21, 28, 31, 45 | 3.337757 | -1.058345 | 0.260486 | 2010 |
| 0.4 | 8, 9, 19, 28, 30 | 3.337757 | -1.012808 | 0.727418 | 1379 |
| 0.5 | 7, 9, 16, 21, 28 | 3.337757 | -1.058345 | 1.139706 | 1508 |
| 0.6 | 8, 16, 17, 23, 38 | 3.259867 | -0.958202 | 1.572640 | 1861 |
| 0.7 | 8, 19, 26, 30, 44 | 3.259867 | -0.936443 | 2.000974 | 1302 |
| 0.8 | 8, 16, 17, 26, 45 | 3.259867 | -0.958202 | 2.416254 | 1108 |
| 0.9 | 8, 16, 17, 23, 38 | 3.259867 | -0.958202 | 2.838061 | 1086 |
| 1 | 19, 22, 31, 35, 49 | 3.259867 | -0.546347 | 3.259867 | 1001 |

Table 3. Numerical results— $N = 100$.

| θ | Recommendation list | C(X) | H(X) | Image | Outer iter. |
|----------|---------------------|----------|-----------|-----------|-------------|
| 0 | 13, 67, 83, 93, 98 | 3.530401 | -1.013864 | -1.013864 | 1193 |
| 0.1 | 13, 35, 68, 83, 90 | 3.054877 | -1.013863 | -0.606989 | 2406 |
| 0.2 | 13, 50, 64, 71, 97 | 2.581903 | -1.019357 | -0.299105 | 1759 |
| 0.3 | 31, 79, 87, 88, 98 | 2.280773 | -0.868041 | 0.076603 | 1472 |
| 0.4 | 35, 38, 79, 87, 100 | 2.267081 | -0.898944 | 0.367466 | 1840 |
| 0.5 | 23, 25, 28, 29, 96 | 1.617326 | -0.450159 | 0.583584 | 1900 |
| 0.6 | 11, 22, 25, 56, 71 | 1.592375 | -0.434746 | 0.781527 | 3320 |
| 0.7 | 11, 19, 25, 29, 71 | 1.627995 | -0.452723 | 1.003780 | 1760 |
| 0.8 | 22, 23, 28, 29, 52 | 1.617326 | -0.379522 | 1.217957 | 2317 |
| 0.9 | 25, 28, 52, 71, 73 | 1.665473 | -0.412739 | 1.457652 | 2548 |
| 1 | 11, 29, 73, 88, 96 | 1.695475 | -0.472928 | 1.695475 | 1107 |

In this formulation, the objective function is a weighted average of compatibility and heterogeneity between two users belonging to the list (both interactions, node i to j and *vice versa*, are checked). This approach increases the user's chance to find a partner quickly, improving the success of social network.

Future tests will be performed with several homogeneity, heterogeneity and attractiveness functions in order to evaluate and improve the quality of the generated clustering.

REFERENCES

- [1] L. F. Mendonça, "An Approach for Personalized Social Matching Systems by Using Ant Colony," *Proceedings of the Brazilian Workshop on Social Network Analysis and Mining, XXXII Congress of the Brazilian Computer Society Computer Society*, Curitiba, 2012.
- [2] M. Dorigo and T. Stutzle, "Ant Colony Optimization," The MIT Press, Cambridge, 2004. <http://dx.doi.org/10.1007/b99492>
- [3] S. Alsaleh, R. Nayak, Y. Xu and L. Chen, "Improving Matching Process in Social Network Using Implicit and Explicit User Information," *Proceedings of the Asia-Pacific Web Conference, Lecture Notes in Computer Science*, Beijing, 2011, pp. 313-320.
- [4] E. M. Morgan, T. C. Richards and E. M. VanNess, "Comparing Narratives of Personal and Preferred Partner Characteristics in Online Dating Advertisements," *Computers in Human Behavior*, Vol. 26, No. 5, 2010, pp. 883-888. <http://dx.doi.org/10.1016/j.chb.2010.02.002>

- [5] L. Terveen and D. W. McDonald, "Social Matching: A Framework and Research Agenda," *ACM Transactions on Computer-Human Interaction*, Vol. 12, No. 3, 2005, pp. 401-434. <http://dx.doi.org/10.1145/1096737.1096740>
- [6] H. Oinas-Kukkonen, K. Lyytinen and Y. Yoo, "Social Networks and Information Systems: Ongoing and Future Research Streams," *Journal of the Association for Information Systems*, Vol. 11, No. 2, 2010, pp. 555-566.
- [7] R. Andreani, J. M. Martínez, M. Salvatierra and F. Yano, "Quasi-Newton Methods for Order-Value Optimization and Value-at-Risk Calculations," *Pacific Journal of Optimization*, Vol. 2, 2006, pp. 11-33.
- [8] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, "Introduction to Algorithms," The MIT Press, Cambridge, 2009.
- [9] M. Dorigo, G. Di Caro and L. M. Gambardella, "Ant Algorithms for Discrete Optimization," *Artificial Life*, Vol. 5, No. 2, 1999, pp. 137-172. <http://dx.doi.org/10.1162/106454699568728>
- [10] M. Dorigo, V. Maniezzo and A. Colomi, "The Ant System: Optimization by a Colony of Cooperating Agents," *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, Vol. 26, No. 1, 1996, pp. 29-41. <http://dx.doi.org/10.1109/3477.484436>
- [11] E. Alba, "Parallel Metaheuristics: A New Class of Algorithms," Wiley Series on Parallel and Distributed Computing, Wiley-Interscience, Hoboken, 2005.