

Uma abordagem via Colônia de Formigas para o problema de correspondência em redes sociais

Luziane Ferreira de Mendonça¹

¹Departamento de Ciência da Computação
Universidade Federal do Rio de Janeiro (UFRJ) – Rio de Janeiro, RJ – Brazil

luziane@dcc.ufrj.br

Abstract. *Personalised social matching systems can be seen as recommender systems that recommend people to others in the social networks, with desirable skills/characteristics. In this paper, 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, are taken into account their personal characteristics and preferences (and their relevances) relative to individual(s) who wants to interact.*

Resumo. *Sistemas de agrupamento sociais personalizados podem ser vistos como sistemas de recomendação, que sinalizam a uma pessoa específica outros membros de uma mesma rede social com habilidades e/ou características desejáveis por ela. Neste trabalho, um algoritmo baseado em Colônia de Formigas é aplicado para o problema de otimização de agrupamento/correspondência de indivíduos em uma rede social específica para esse fim; durante esse processo, são levadas em consideração suas características pessoais e preferências (e a relevância das mesmas) em relação ao(s) indivíduo(s) com quem quer se relacionar.*

1. Introdução

Sistemas de agrupamento sociais personalizados podem ser vistos como sistemas de recomendação, que sinalizam a uma pessoa específica outros membros de uma mesma rede social ao invés de recomendar produtos a uma pessoa [Alsaleh et al. 2011]. A recomendação de pessoas é algo muito mais complicado que a recomendação de produtos [Morgan et al. 2010], uma vez que nem sempre duas pessoas que possuem as mesmas características devem ser recomendadas uma a outra. Ou seja, as características pessoais podem não ser as preferências desse mesmo indivíduo, e a consequência imediata é que grande parte das técnicas de recomendação podem não ser adequadas [Terveen and McDonald 2005].

Além disso, o número de membros desse tipo de rede social vem aumentando muito nos últimos anos, o que gerando um considerável crescimento da complexidade computacional. Como exemplo, um dos maiores sites de relacionamento do Brasil possui mais de 30 milhões de usuários cadastrados¹ e uma média de 1000 novos membros todos os dias.

¹Site *ParPerfeito*, <http://www.parperfeito.com.br/>, 2011.

Mesmo se tratando de sites que não se destinam a relacionamentos amorosos, como os que buscam determinar grupos com interesse em comum, a complexidade computacional ainda assim é alta pois embora tenha um número menor de membros, é necessário realizar agrupamentos, onde o tamanho de cada grupo pode ser variável.

De uma forma geral, é utilizada alguma função que mensura a similaridade entre as características desejáveis por um indivíduo em questão e aquelas apresentadas pelos demais membros da rede; esse mesmo indivíduo recebe então a recomendação seguindo a ordem da similaridade: pessoas com maior similaridade recebem prioridade na recomendação [Oinas-Kukkonen et al. 2010]. Entretanto, mesmo considerando as pessoas classificadas nas primeiras colocações em similaridade em relação a um mesmo indivíduo, é possível que sejam similares entre si, gerando a recomendação sucessiva de indivíduos com as mesmas características. Se essas características não forem importantes para aquele que recebe a recomendação, ele se sentirá demotivado pois poderá realizar várias tentativas e não conseguir estabelecer um relacionamento com êxito.

Uma forma de aumentar as chances de sucesso durante o processo de recomendação é considerar a recomendação de vários membros da rede de uma única vez, formando um grupo e não apenas um par, considerando a similaridade entre as características pertinentes de uma pessoa e as desejáveis de outra, e a discrepância entre as características pertinentes entre duas pessoas. Essa forma de abordagem garante que o grupo formado seja altamente atraente, não só por contemplar as características desejáveis de cada membro, mas também por ser formado por pessoas com características heterogêneas. Dessa forma, caso a rede social esteja relacionada a um site de namoro, as chances de encontrar em menos tentativas um pessoa para iniciar um relacionamento aumentam; caso a rede social tenha por objetivo conectar pessoas com interesses em comum para debaterem sobre um assunto específico, o grupo de discussão gerado terá maior propensão a apresentar ideias variadas, gerando debates mais interessantes.

Neste trabalho é proposta uma modelagem para o problema de recomendação de uma ou mais pessoas para cada indivíduo de uma rede (formando assim grupos de 2 ou mais componentes), em que a escolha é baseada no ponderamento entre a compatibilidade dos indivíduos (discrepância entre as características pessoais de um e as desejáveis do outro) e a heterogeneidade do grupo formado (discrepância entre as características pessoais entre os indivíduos). Esses são os dois componentes principais do problema de otimização formulado, que é resolvido utilizando um algoritmo baseado na heurística de Colônia de Formigas [Dorigo and Stutzle 2004].

2. Modelagem do Problema

Considerando que uma rede social pode ser modelada como um grafo completo bidirecionado com N nós (indivíduos), para cada nó i é possível obter três conjuntos de NC informações/características (do próprio indivíduo, desejáveis no parceiro, e a importância que o indivíduo atribui a essas características²) a serem armazenadas nos vetores \bar{u}_i , \bar{v}_i e w_i .

²Nesse sentido, são adotados os valores de $(w_i)_p \in \{0, 1, 2, 3\}$, em que valores próximos 0 indicam indiferença em relação à característica p , enquanto que valores próximos de 3 indicam que a característica é relevante.

As componentes de \bar{u}_i e \bar{v}_i neste trabalho são números reais³ representando atributos como altura, peso, idade, escolaridade, sexo, etc., com valores normalizados.

Fazendo uso desses vetores, é possível construir uma função que irá acoplar a compatibilidade e a heterogeneidade para um grupo, fazendo a média aritmética do resultado da interação entre cada par de indivíduos desse grupo. A compatibilidade entre o nó i e o nó r pode ser mensurada de forma inversamente proporcional à discrepância entre as características v_i desejáveis por i e as características u_r que r possui, ponderadas pela importância que as características possuem, enquanto a heterogeneidade como a discrepância entre as características u_i e u_r que i e r possuem.

Conseqüentemente, o melhor conjunto de grupos $G = \{G_1, G_2, \dots, G_{NG}\}$ em que os nós podem ser separados é obtido maximizando a função correspondente à média das imagens de cada grupo.

$$F(G) = \frac{1}{NG} \sum_{k=1}^{NG} \left\{ \frac{1}{|G_k|(|G_k| - 1)} \sum_{i \in G_k} \sum_{r \in G_k, i \neq r} \theta \left[\exp \left(- \sum_{p=1}^{NC} (w_i)_p ((v_i)_p - (u_r)_p)^2 \right) \right] \right. \\ \left. + (1 - \theta) \left[\sum_{p=1}^{NC} ((u_i)_p - (u_r)_p)^2 \right] \right\}, \quad (1)$$

onde $\theta \in [0, 1]$ é um fator de ponderamento, que reflete a importância de cada uma das parcelas na função.

A modelagem proposta pode ser utilizado em diferentes tipos de recomendação, selecionando adequadamente os vetores u_i e v_i . Por exemplo, ao adotar $v_i = u_i$ e $\theta = 1$ o algoritmo agrupará indivíduos com as mesmas características, enquanto que com $\theta = 0$ serão agrupados indivíduos com características díspares.

3. Heurística Proposta

O algoritmo proposto para a resolução do problema de maximizar (1) é baseado na busca pelo ótimo via Colônia de Formigas (*Ant Colony Optimization*, ACO) [Dorigo et al. 1996] que é uma heurística baseada em probabilidade, criada para solução de problemas computacionais que envolvem procura de caminhos em grafos.

Para o problema abordado neste trabalho, é considerado que cada formiga constroi iterativamente um grupo de nós, conectando-os. A cada iteração interna, cada formiga k se move do nó i para o nó r de um grafo de acordo com uma regra probabilística definida, sucessivamente, culminando em um caminho fechado que corresponde às conexões entre os membros de um mesmo grupo. Quando um grupo é formado, outra formiga é situada em um dos nós não selecionados, para que novas conexões sejam realizadas e um novo agrupamento seja formado. Esse procedimento se repete até que não haja nó fora de algum grupamento.

A probabilidade de associar o nó i para o nó r depende da combinação de dois fatores: a atratividade η_{ir} e o feromônio τ_{ir} da aresta (i, r) . O feromônio varia a cada

³Quando se tratar de componentes com atributos em formato de *string*, estas podem receber um pré-processamento para se converterem em escala numérica.

iteração externa em todas as arestas, com valores inicialmente iguais, sofrendo decaimento natural de iteração para iteração (correspondente à evaporação) exceto para as arestas que conectam nós de um mesmo grupo do conjunto ótimo (funcionando como a herança das boas informações já obtidas de uma geração de formigas para a outra). A atratividade é o fator responsável pelo agrupamento de nós que condizem com o objetivo, tendo valor alto quando o par de nós promove um aumento na função de compatibilidade, e baixo caso contrário,

$$\eta_{ir} = \frac{\theta}{NC} \sum_{p=1}^{NC} \exp \left\{ - \sum_{p=1}^{NC} (w_i)_p ((v_i)_p - (u_r)_p)^2 \right\} + \frac{(1 - \theta)}{NC} \sum_{p=1}^{NC} ((u_i)_p - (u_r)_p)^2$$

quando i e r pertencem ao mesmo grupo, e $\frac{1}{NC}$ caso contrário.

O trabalho se encontra em desenvolvimento, onde já foram realizados testes computacionais preliminares. Foram realizados testes numéricos para redes com tamanho variável geradas de forma aleatória, onde foi possível constatar que a discrepância das características entre os membros do grupo influencia na composição do mesmo não só pelos indivíduos selecionados, mas também na quantidade destes. Os testes numéricos indicam que conforme aumenta o tamanho da rede, há uma tendência maior de que os grupos formados sejam maiores (dentro do limite máximo estipulado).

References

- Alsaleh, S., Nayak, R., Xu, Y., and Chen, L. (2011). Improving matching process in social network using implicit and explicit user information. In *Proceedings of the 13th Asia-Pacific web conference on Web technologies and applications*, APWeb'11, pages 313–320. Springer-Verlag.
- Dorigo, M., Maniezzo, V., and Coloni, A. (1996). The ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, 26:29–41.
- Dorigo, M. and Stutzle, T. (2004). *Ant Colony Optimization*. MIT Press.
- Morgan, E. M., Richards, T. C., and Vanness, E. M. (2010). Comparing narratives of personal and preferred partner characteristics in online dating advertisements. *Computers in Human Behavior*, 26(5):883–889.
- Oinas-Kukkonen, H., Lyytinen, K., and Yoo, Y. (2010). Social networks and information systems: Ongoing and future research streams. *Journal of the Association for Information Systems*, 11(2).
- Terveen, L. and McDonald, D. W. (2005). Social matching: A framework and research agenda. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(3):401–434.