

Sistemas de recomendação e interação na web social

Fernando M. Figueira Filho¹

Paulo Lício de Geus²

Instituto de Computação - UNICAMP

{fmarques¹, paulo²}@ic.unicamp.br

João Porto de Albuquerque

Escola de Artes, Ciências e Humanidades - USP

j.porto@usp.br

RESUMO

Nos últimos anos tem-se percebido uma mudança de paradigma na *World Wide Web*. Novas funcionalidades nas aplicações *web* encorajam uma crescente mudança no papel do usuário, de um mero consumidor de informação para um ativo produtor de conhecimento. De fato, a colaboração do usuário na produção virtual de conhecimento tem sido considerada um importante processo social. O fenômeno despertou interesse científico para o que tem sido chamado de “*web social*”. Nesse contexto, esse artigo chama a atenção para os sistemas de recomendação. Esses sistemas têm por objetivo reduzir a sobrecarga de informação da *web* atual por meio da seleção de conteúdo baseada em preferências do usuário. Tradicionalmente, eles eram um objeto de estudo para pesquisa em recuperação de informação, mas alguns problemas demandam uma abordagem de pesquisa mais ampla. Modelos de recomendação baseiam-se na interseção dos contextos de usuários para estimar uma recomendação, o que tem sido feito através de dados semânticos ou pela análise de redes sociais. Entretanto, existe uma falta de modelos que unificam estrutura e semântica. Essa unificação é fundamental para que sistemas de recomendação levem em conta não apenas os elos entre os usuários, mas o significado desses elos. Baseado na perspectiva de contexto como um problema de interação social, o artigo faz uma análise crítica às abordagens de recomendação atuais e propõe uma nova direção de pesquisa no campo da *web social*.

ABSTRACT

In the last years, we noticed a paradigm change in the World Wide Web. New web application functionalities incite a growing change in the role of a user from a mere information consumer to an active knowledge producer. Indeed, user collaboration in virtual knowledge production has been considered an important social process. The phenomena leveraged scientific interest to what has been called “the social web”. In this context, this paper draws attention to recommender systems. These systems attempt to reduce the information overload by selecting content based on user preferences. Traditionally, recommender systems were an object of study for information retrieval

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

research, but some issues demands a broader research approach. Recommendation models are based in the intersection of user contexts to estimate a recommendation, what has been done by using semantic data or through social network analysis. However, there is a lack of models that unify semantics and structure. This unification is of great importance to recommender systems in order to take into consideration not only social ties, but the meaning of these ties. Based on the perspective of context as a social interaction problem, this paper critically analyzes the current recommendation approaches and proposes a new research direction in the social web field.

Author Keywords

Recommender systems, social networks, social web.

ACM Classification Keywords

H5.3 [Information Systems Interfaces and Presentation (e.g. HCI)]: Group and Organization Interfaces—collaborative computing, web-based interaction, theory and models.

INTRODUÇÃO

O paradigma da *web social* pode ser observado na manifestação de uma crescente tendência em explorar, de maneira explícita ou implícita, a riqueza dos elos que se firmam com a interação social mediada pelas tecnologias de informação. Essa mudança de paradigma revela-se também no desenvolvimento e no crescimento de popularidade de uma nova classe de aplicações para *web*, a exemplo dos sistemas colaborativos para produção de conhecimento (e.g. enciclopédias livres).

Dado o aumento vertiginoso do volume de informação produzido, sistemas de recomendação têm por objetivo reduzir a sobrecarga de informação, através da seleção de conteúdo baseado em preferências do usuário. Esses sistemas têm sido classificados em três categorias no que tange à abordagem utilizada para recomendação [1]: a) abordagem baseada em conteúdo, nos quais são recomendados itens similares com aqueles que o usuário mostrou preferência no passado; b) abordagem colaborativa, na qual são recomendados itens escolhidos por pessoas com preferências similares às do usuário e c) abordagens híbridas, que de alguma forma combinam técnicas na tentativa de solucionar alguns problemas potenciais das abordagens puras anteriores. Entretanto, conforme destaca [5], existem dois problemas com a investigação de sistemas de recomendação nos dias de hoje.

Primeiramente, a tradicional dicotomia entre abordagens colaborativas *versus* baseadas em conteúdo não mais comporta a evolução das técnicas de recomendação atuais, que ultrapassam os limites da área de recuperação de informação, trazendo para o plano frontal questões relevantes a outras áreas da computação, incluindo a área de interação humano-computador. Em segundo lugar, o enfoque de investigação em sistemas de recomendação tem negligenciado o fato que a recomendação *per se* é, antes de tudo, fruto de um processo social e tem influência dos elos sociais estabelecidos ao longo da atuação humana nesse processo.

De fato, a análise estrutural em redes sociais [8] ganhou um novo campo de aplicação no estudo de sistemas de recomendação para *web* social. Estudos similares já foram amplamente aplicados à estrutura de documentos da *web* [4], entretanto pouco foi revelado a respeito das estruturas sociais que emergem com o uso das ferramentas da *web* social. Com esse objetivo, o presente artigo apresenta um novo aporte teórico para abordar o problema da recomendação. A investigação aqui conduzida pode ser caracterizada em duas direções: a) o estudo das propriedades estruturais derivadas a partir da análise das relações entre produtores de conhecimento em ambientes colaborativos da *web* social e b) a relação dessas propriedades estruturais com o contexto semântico dos artefatos produzidos colaborativamente.

O presente artigo contempla essas duas direções de investigação baseado em um referencial teórico formado pela perspectiva de contexto como um problema de interação social, proposta por Dourish [2]. Após uma introdução da abordagem colaborativa de recomendação e da caracterização de alguns dos principais problemas enfrentados por essa abordagem, o artigo então traz uma breve revisão do referencial teórico e faz uma análise crítica das abordagens colaborativas de recomendação atuais, propondo uma nova direção de pesquisa na área baseada em modelos de recomendação que unificam propriedades estruturais com propriedades semânticas. O artigo termina com uma descrição dos trabalhos futuros nessa direção a serem realizados no contexto da *web* social.

SISTEMAS DE RECOMENDAÇÃO COLABORATIVOS

Abordagens colaborativas de recomendação têm suas origens nas abordagens colaborativas de filtragem de informação introduzidas na década de 90 [6]. Desde então, a abordagem colaborativa de recomendação mantém suas características principais, e consiste fundamentalmente em recomendar um artefato (e.g. um livro, um filme, uma página da *web*) que foi preferido por usuários similares ao usuário que recebe a recomendação.

A diferença entre os diversos sistemas de recomendação que usam a abordagem colaborativa está em como a similaridade entre os usuários é efetivamente calculada. A técnica de filtragem colaborativa baseia-se em avaliações explícitas (*ratings*) dos conteúdos. Dessa forma, usuários

que avaliam de forma semelhante os mesmos conteúdos são considerados usuários com preferências similares e, portanto, estima-se que um conteúdo avaliado positivamente por esse subconjunto de usuários do qual o usuário x faz parte deva também ser bem avaliado pelo usuário x (e.g. Paulo e Roberto gostam de Goethe. Paulo gosta de Molière, então Roberto também deverá gostar do escritor francês). Entretanto, essa técnica sofre da posição relutante dos usuários ao serem submetidos à avaliação explícita de artefatos. Essa resistência natural é resultado da falta de tempo e paciência dos usuários em colaborar com o sistema, e pode até mesmo ser relacionada com a questão da privacidade (i.e. a relutância dos usuários em revelar explicitamente suas preferências). Como consequência, sistemas de recomendação que se baseiam na técnica de filtragem colaborativa estão inevitavelmente sujeitos a problemas de eficiência gerados pela falta de um número suficiente de avaliações no sistema.

A dificuldade em estimar a similaridade entre os usuários baseando-se em colaborações explícitas deu lugar a técnicas que procuram inferir as preferências do usuário através de suas demonstrações implícitas de interesse com relação a um dado artefato. A grande vantagem dessa técnica é que as manifestações de interesse dos usuários são silenciosamente deixadas como um resultado natural da manipulação dos artefatos, em vez de serem explicitamente solicitadas aos usuários. Isso leva a solução de um grande problema enfrentado pelas técnicas de filtragem colaborativa, que é a “partida a frio” (*cold-start problem*), isto é, o sistema perde em eficiência nas estimativas até que um número suficiente de avaliações seja realizado. No caso da técnica implícita, as avaliações estão implícitas nos dados gerados pelos usuários e são coletadas principalmente com o uso de heurísticas, como por exemplo no sistema PHOAKS [7], um dos primeiros a implementar a técnica. O sistema recomenda URLs (*uniform resource locators*) baseado na inclusão desses endereços em mensagens trocadas entre usuários na *Usenet*. Usando essa heurística, dois usuários são considerados similares se uma mesma URL foi trocada entre eles em uma mensagem. Um peso é dado ao elo que é implicitamente formado entre os usuários, baseado no número de usuários que recomendaram a mesma URL. Essa métrica foi depois estendida, formalizada e nomeada como “peso de autoridade” em [4].

A grande inovação introduzida por projetos como o PHOAKS está na exploração dos dados produzidos através da interação social que ocorre mediada pelas tecnologias de informação. Isso significou uma mudança de direção na pesquisa em sistemas de recomendação, que passou a considerar as associações estabelecidas entre as pessoas como dado de entrada para os mecanismos de recomendação. Em outras palavras, sistemas de recomendação passaram a considerar dados estruturais, em vez de apenas dados semânticos, como os *ratings* antes coletados na filtragem colaborativa. Entretanto, abordagens que consideram dados puramente estruturais também

apresentam um problema fundamental. É computacionalmente simples inferir que existe uma relação r entre duas pessoas $p1$ e $p2$, mas é impossível qualificar r . Dessa forma, a pergunta “como determinar o significado de r ?” requer, portanto, uma investigação mais extensa, que ultrapassa os limites das áreas mais tradicionais da computação e solicita soluções de natureza multidisciplinar.

SISTEMAS DE RECOMENDAÇÃO E O CONTEXTO

Coletar e processar um conjunto de dados para encontrar uma relação entre duas pessoas é certamente um problema que pode ser investigado com heurísticas, análise estrutural, teoria dos grafos etc. Extrair algum significado da relação entre duas pessoas é certamente um problema semântico que nos remete a pergunta: “Em que contexto as duas pessoas estão relacionadas?”. De fato, a questão do contexto reaparece em diversos momentos ao longo da história da área de interação humano-computador (IHC).

Conforme foi visto na seção anterior, sistemas de recomendação que aplicam a técnica de filtragem colaborativa estimam a similaridade entre dois usuários baseado no contexto que esses usuários compartilham. Esse contexto é formado pelo histórico de avaliações (*ratings*) conferido aos artefatos por cada usuário. Uma métrica de similaridade baseia-se, portanto, no cálculo da sobreposição dos históricos dos usuários para estimar a similaridade e fazer uma recomendação.

Nessa abordagem, o contexto que une dois usuários pode ser capturado através da avaliação de um artefato e representado através de estruturas de dados que guardam as avaliações de cada usuário para um dado artefato. Para isso é necessário, portanto, que esse contexto seja modelado *a priori*, através de um esquema de categorias, que podem ser notas num intervalo pré-definido (e.g. um filme pode receber uma nota de 0 a 5) ou mesmo uma classificação binária (e.g. um dado filme é bom ou ruim). Essa forma de modelar o problema de recomendação resolve o problema do contexto através da padronização de um esquema de representação único para avaliação de artefatos. Segundo Dourish [2], essa é uma das formas de entender o problema do contexto.

Se o contexto tem um caráter representacional, atividade e contexto são separáveis. Por essa razão, outros problemas surgem com as técnicas de recomendação baseadas na filtragem colaborativa e seus esquemas de avaliação. Usuários são considerados similares por seguirem padrões compatíveis de avaliação, entretanto essa forma de modelar parte do princípio que todos usuários possuem uma mesma sistemática de avaliação e que essa sistemática independe das situações as quais os usuários são submetidos. Em outras palavras, esse pressuposto assume, por exemplo, que a nota 5 dada por um usuário x para um artefato a tem o mesmo significado que a nota 5 dada para o mesmo artefato pelo usuário y . Segundo esse modelo de recomendação, x e y são usuários similares, mas do ponto de vista cognitivo, x e y podem ter concepções bem diferentes da relação entre

a nota 5 e o artefato a . As razões para essa diferença são diversas, e dentre elas podemos citar diferenças de expertise (e.g. a nota 5 para um especialista em filmes evoca um conjunto distinto de exemplares, refinados através de diferentes critérios, em comparação a concepção de nota 5 de um usuário leigo) e diferenças socioculturais que são fruto de diferentes retrospectos de vida e da inserção do usuário em diferentes grupos sociais - cada grupo, nesse sentido, possui sua própria sistemática de classificação (ver [3] para uma discussão mais aprofundada nessa temática). Dessa maneira, definir esquemas padronizados de classificação e assumir uma homogeneidade entre os usuários pode ter reflexos negativos na eficiência de um sistema de recomendação, gerando a insatisfação dos usuários ao criar estimativas que não condizem com a realidade.

Temos, portanto, um problema na área de sistemas de recomendação que permite contribuições da área de interação humano-computador. De um lado, temos as técnicas de filtragem colaborativa e todos os seus problemas que, até então, foram relacionados nesse artigo. Do outro lado, temos as técnicas que se baseiam somente nas propriedades estruturais dos elos que se formam através da análise implícita dos dados produzidos pelos usuários. Essas técnicas, contudo, não contemplam a questão da semântica e baseiam suas recomendações somente na existência de uma ligação entre os usuários, sem levar em consideração o significado dessas ligações. A seção seguinte aponta na direção de unificar a questão da estrutura e a questão da semântica, levando em conta uma perspectiva alternativa de contexto, baseada na interação social.

SISTEMAS DE RECOMENDAÇÃO E A WEB SOCIAL

Em contraste com a perspectiva representacional de contexto, Dourish [2] propõe uma alternativa baseada na interação social. Nessa perspectiva, o contexto entre dois usuários é constituído a partir de uma situação de interação entre eles e, por essa razão, não pode ser definido *a priori*, baseado em convenções de representação, como na técnica de filtragem colaborativa apresentada anteriormente nesse artigo.

Definir o contexto entre dois usuários a partir da interação social entre eles é uma tarefa compatível com o paradigma das ferramentas que compõem hoje a *web* social. Situações de interação que promovem um elo de ligação entre duas pessoas são abundantes nesses ambientes e despertam, portanto, grande interesse para a pesquisa em sistemas de recomendação. O restante dessa seção apresenta dois exemplos de aplicação nessa direção.

Exemplos de aplicação

A enciclopédia livre Wikipedia¹ beneficia hoje milhares de usuários com a produção colaborativa de conhecimento. A

¹ <http://www.wikipedia.org>

interação social nesse tipo de ambiente é fundamental não somente para produção, mas para a manutenção de níveis aceitáveis de qualidade e para coordenação das atividades colaborativas.

Uma rede social pode ser facilmente derivada a partir da heurística que um usuário está diretamente ligado a outro se ambos contribuíram em algum momento na composição de um mesmo artigo na enciclopédia, o que compõe um contexto estrutural comum entre dois usuários. Já o contexto semântico pode ser extraído por várias fontes, dentre elas as categorias as quais um artigo é associado (e.g. dois usuários contribuíram em um artigo sobre filosofia contemporânea). Podemos dizer, portanto, que o contexto semântico qualifica a relação r entre dois usuários $p1$ e $p2$. Dessa forma, é possível criar um modelo de recomendação que unifica propriedades estruturais com propriedades semânticas.

Uma possível aplicação para esse modelo está na recomendação de artigos da enciclopédia. Através da análise estrutural da rede social formada pelos usuários, é possível identificar usuários que possuem maior peso de autoridade [4] em determinados assuntos (e.g. usuários especialistas em filosofia contemporânea). Um possível tema para trabalho futuro constitui em verificar empiricamente se os artigos com a contribuição desses usuários possuem de fato uma maior qualidade. Um sistema de recomendação baseado nessa métrica seria interessante para sugestão de artigos relacionados e para a criação de novos critérios de ordenação em sistemas de busca.

Outra aplicação possível para sistemas de recomendação na *web* social é encontrada nos fóruns técnicos de discussão para solução de problemas. Uma característica desse tipo de ambiente colaborativo é a heterogeneidade dos usuários. Especialistas e novatos interagem com o intuito de produzir soluções para problemas de natureza diversa e o resultado é a produção de conhecimento na forma de *threads* de discussão. Entretanto, em meio a milhares de mensagens, achar a melhor solução para um dado problema usando apenas a busca por palavras-chave pode ser um trabalho bastante penoso. Esse problema motiva a criação de um modelo de recomendação de soluções baseado em aspectos estruturais (e.g. dois usuários são ligados quando participaram de uma mesma *thread* de discussão) e aspectos semânticos, que podem ser extraídos do próprio conteúdo textual que compõe as discussões. A unificação de estrutura e semântica também permite a identificação automática de potenciais comunidades de interesse em determinados assuntos, o que abre novas possibilidades de pesquisa na área dos sistemas de recomendação.

TRABALHOS FUTUROS

Os trabalhos futuros incluem a construção de um *web crawler* para a mineração de dados estruturais e semânticos em ferramentas da *web* social, a exemplo das aplicações apresentadas na seção anterior. Os autores estão

interessados nas relações entre propriedades estruturais e propriedades semânticas, a serem identificadas com um conjunto apropriado de heurísticas e aplicando técnicas de aprendizagem de máquina. A partir de estudo de caso a ser realizado, um modelo de recomendação será elaborado e avaliado através da construção de um protótipo.

BIOGRAFIA DOS AUTORES

Fernando Marques Figueira Filho possui graduação em Ciência da Computação pela Universidade Estadual de Campinas (2003). Atualmente faz doutorado pelo Instituto de Computação da mesma universidade. Tem experiência na área de Ciência da Computação, com ênfase em Sistemas de Informação. João Porto de Albuquerque é professor doutor da Escola de Artes, Ciências e Humanidades da Universidade de São Paulo (EACH/USP). Tem experiência nas áreas de Sistemas de Informação e Ciência da Computação. Paulo Lício de Geus é professor associado da Universidade Estadual de Campinas. Tem experiência na área de Ciência da Computação, com ênfase em Teleinformática.

REFERÊNCIAS

1. ADOMAVICIUS, G., AND TUZHILIN, A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING* (2005), 734–749.
2. DOURISH, P. What we talk about when we talk about context. *Personal and Ubiquitous Computing* 8, 1 (2004), 19–30.
3. FIGUEIRA FILHO, F., PORTO DE ALBUQUERQUE, J., AND DE GEUS, P. Analisando sistemas de classificação na web sob a perspectiva da interação social em comunidades de prática. In *IHC'08: Proceedings of the X Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais* (2008). to appear.
4. KLEINBERG, J. Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM* 46, 5 (1999), 604–632.
5. PERUGINI, S., GONÇALVES, M., AND FOX, E. Recommender Systems Research: A Connection-Centric Survey. *Journal of Intelligent Information Systems* 23, 2 (2004), 107–143.
6. RESNICK, P., AND VARIAN, H. Recommender systems. *Communications of the ACM* 40, 3 (1997), 56–58.
7. TERVEEN, L., HILL, W., AMENTO, B., MCDONALD, D., AND CRETER, J. PHOAKS: a system for sharing recommendations. *Communications of the ACM* 40, 3 (1997), 59–62.
8. WASSERMAN, S., AND FAUST, K. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.