

Sistema de Recomendação Baseado em Conteúdo Textual em Páginas Web

OMMITED

OMMITED
OMMITED

Resumo Recommendation Systems has emerged to assist users against information overload in a environment where not all information is relevant or of interest to the user. Recommendation Systems came with the proposal of being an independent area in Computer Science, rather than filtering information, the focus is to recommend information based on preferences, tastes and interests surprising with new information and helping in the pursuit of the real need. In this paper we present a recommendation system, that is compose by an middleware and a framework, for web environment based on textual content and the sequence of processes to arrive at a recommendation for the user, from the construction of profile (user interest) to the recommendation. This article also presents the strategies, technologies and areas related to Computer Science at each stage of the recommendation process.

Keywords: recommendation system, tag cloud, profile user

1 Introduction

2 Sistemas de Recomendação

Os sistemas de recomendação atuam em diferentes áreas, desde de Ciências Cognitivas, Teoria da Aproximação até Recuperação da Informação, com o intuito de prover indicação de conteúdo não classificado, somente com base em preferências do usuário [10].

Inicialmente, os sistemas de recomendação eram caracterizados como sistemas de filtragem colaborativa, em decorrência do primeiro sistema lançado, o *Tapestry* [4,9,7]. No entanto, foi pelo trabalho de Resnick e Varian [10] que foi proposto o termo mais genérico “Sistema de Recomendação”, sendo adotado por não consistir em um produto de colaboração entre usuários, mas sim como foco na recomendação de informações [1,7].

Segundo Gediminas e Alexander [1], o problema de recomendação pode ser simplificado ao problema da estimativa de classificação para as informações que ainda não foram visualizados pelo usuário.

Desse modo, novamente, Gediminas e Alexander [1] definem um sistema de recomendação pela combinação dos conjuntos de todos os usuários U e de itens a serem recomendados I , mediados pela função \tilde{u} , que mede quanto um item é

relevante, cujo resultado da combinação $U \times I$ é ordenado em um conjunto A . Resumidamente:

$$u \in U, i' u = \operatorname{argmax}_{s \in I} u(\ddot{u}, i)$$

Em geral, na maioria dos casos o conjunto de itens analisados é muito inferior ao total. De modo análogo, a mesma relação pode ser mapeada como uma matriz esparsa, conforme apresentado na Figura 1:

		Items					
		1	2	...	i	...	m
Users	1	5		1		3	
	2		2	4	2		1
	...	1	5	3		3	2
	u		2		1	4	1
	...	1		1	2	1	
	n	1	2			5	

Figura 1. Conjunto esparsa de recomendações.

A Figura 1 representa uma matriz de n usuários e m itens e cada célula $A(u, i)$ corresponde à uma avaliação do usuário u ao item i . O desafio é sempre superar a função \ddot{u} para todo o conjunto $U \times I$ para que o usuário u tenha novas experiências e conhecimento de todo o conteúdo em que ele está inserido [1].

De modo geral, os sistemas de recomendação atuam com base em um conjunto de atividades que iniciam na coleta de dados até a recomendação de itens, em que cada etapa é descrita brevemente nas Seções 2.1 a 2.3.

2.1 Coleta de Dados

A etapa inicial consiste na obtenção dos dados brutos, os quais atuam como base no processo de recomendação. Essa atividade pode impactar diretamente no restante do processo, pois quanto específico ou gerais esses forem, consequentemente os resultados atuam de modo semelhante, os quais podem ocorrer em imprecisões ou erros de classificação.

A atividade pode ser executada de modos distintos, seja explícita ou implícita ao usuário, dentre as possíveis estratégias tem-se o monitoramento de atividades, solicitação de cadastros, navegação em páginas com *cookies* e a exibição de

exemplos [3,8]. A partir disso, a cada usuário é definido um perfil com base nas interações conduzidas.

2.2 Obtenção de Informação

Como segunda etapa, a partir dos dados brutos é necessário conduzir duas principais análises, sob perspectivas gerais e específicas de seu conteúdo.

Em uma análise geral, o processo de identificação de que tipo de informação e contexto que os dados são oriundos. Uma abordagem comum é a aplicação de técnicas de visualização de informação, como exemplo, a utilização de *Tag Clouds*, a qual permite a identificação dos principais conceitos relacionados ao conteúdo analisado com base na frequência dos termos [6].

Posteriormente, uma análise mais específica deve ser conduzida, de modo a identificar a relevância do elemento analisado, para isso são utilizados modelos baseados em Recuperação de Informação, permitindo a inter-conexão entre o provedor de informação e o consumidor [12].

Diferentes estratégias possuem base em Inteligência Artificial, as quais utilizam análise dos dados (utilização de palavras-chave indexadas), modelos lógicos (aplicação de regras booleanas ou lógica *fuzzy*), vetoriais (frequência da ocorrência de elementos ou aplicação de redes neurais) [12] ou probabilísticos (métodos de inferência ou predição) [5,11]. Esses modelos possuem em comum a utilização do perfil do usuário como base na realização da consulta [2,12].

2.3 Tarefa de Recomendação

Por fim, cabe-se ao sistema de recomendação a condução na tomada de decisão pelo usuário. Nesse contexto, Silva [11] aponta alguns objetivos e tarefas que devem ser cumpridos:

- Anotações de Contexto: sugestão de opções que deveriam ser selecionadas.
- Seleção de Melhores Elementos: capacidade de identificar um conjunto limitado de potenciais elementos.
- Todos Elementos: recuperação de todos os elementos relacionados.
- Recomendação Sequencial: prover adequação nas sugestões com base no passo a passo conduzido.
- Recomendação Quantitativa: fornecer recomendação para uma análise exploratória.
- Recomendação Confiável: estabelecer graus de confiança com base na qualidade das recomendações.

Referências

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering* (6) (2005) 734–749

2. Baeza-Yates, R., Ribeiro-Neto, B., et al.: Modern information retrieval. Volume 463. ACM press New York (1999)
3. Cazella, S.C., Nunes, M., Reategui, E.: A ciência da opinião: Estado da arte em sistemas de recomendação. André Ponce de Leon F. de Carvalho; Tomasz Kowaltowski..(Org.). Jornada de Atualização de Informática-JAI (2010) 161–216
4. Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. *Communications of the ACM* **35**(12) (1992) 61–71
5. Khan, A., Baharudin, B., Lee, L.H., Khan, K.: A review of machine learning algorithms for text-documents classification. *Journal of advances in information technology* **1**(1) (2010) 4–20
6. Luis Dias Andreotti, A., de Medeiros Eler, D., Pola, V., Renê, I., Olivette Artero, A.: Análise visual da evolução de coleções de documentos utilizando tag cloud. In: *Colloquium Exactarum*. Volume 9. (2017)
7. Medeiros, I.: Estudo sobre sistemas de recomendação colaborativos. 30 f. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação)—Centro de Informática, Universidade Federal de Pernambuco, Recife (2013)
8. Reategui, E.B., Cazella, S.C., Osório, F.S.: Personalização de páginas web através dos sistemas de recomendação. *Tópico em Sistemas Interativos e Colaborativos*. São Carlos (2006)
9. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: Grouplens: an open architecture for collaborative filtering of netnews. In: *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, ACM (1994) 175–186
10. Resnick, P., Varian, H.R.: Recommender systems. *Communications of the ACM* **40**(3) (1997) 56–59
11. Silva, R.G.N., Loula, A.C.: Modern information retrieval (2016)
12. Souza, R.R., et al.: Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. *Perspectivas em ciência da informação* **11**(2) (2006) 161–173