

Sistemas de recomendação de conteúdo em sites de notícias

Augusto Lohmann (ESDI/UERJ, Brasil)

augustolohmann@gmail.com

ESDI/UERJ

R. Evaristo da Veiga, 95 - Centro, Rio de Janeiro - RJ, 20031-040

André Monat (ESDI/UERJ, Brasil)

andresmonat@yahoo.com.br

ESDI/UERJ

R. Evaristo da Veiga, 95 - Centro, Rio de Janeiro - RJ, 20031-040

Sistemas de recomendação de conteúdo em sites de notícias

Resumo: O artigo apresenta um levantamento teórico e prático do cenário atual dos sistemas automatizados de recomendação de conteúdo, com ênfase nas páginas de matérias dos sites de notícias. Segundo Ricci (2011), os sistemas de recomendação automatizados, em qualquer campo de conhecimento, podem ser categorizados em seis (6) principais grupos. Serão apresentados exemplos de elementos presentes nesse sites que possam ser associados a cada uma das categorias definidas. Com isso, objetiva-se validar a aplicabilidade dessa categorização no contexto do jornalismo, bem como apresentar a questão da importância atribuída aos elementos de recomendação de conteúdo e entender os critérios utilizados na definição do que é recomendado para cada usuário nesses elementos.

Palavras-chave: sistema de recomendação; algoritmos; notícias.

Recommendation Systems on News Sites

Abstract: The paper presents a theoretical and practical survey of the current scenario of the automated content recommendation systems, with emphasis on materials of news sites pages. According to Ricci (2011), the automated recommendation systems, in any field of knowledge, can be categorized into six (6) major groups. Will be presented examples of elements that sites that may be associated with each of the defined above and detailed categories. Thus, the objective is to validate the applicability of this categorization in the journalism context and present the issue of the importance attributed to the content recommendation of elements and understand the criteria used in defining what is recommended for each user in these elements.

Keywords: recommendation system; algorithm; news.

1. Introdução

Aproximadamente três décadas após o início da transposição do conteúdo dos jornais impressos para o formato digital, os sites e aplicativos noticiosos oferecem maior variedade de fluxos de navegação e narrativas, se comparado ao modelo de leitura convencional do jornalismo impresso. Ao contrário de um único fluxo de consumo da informação (a leitura sequencial do conteúdo, baseada em critérios editorialmente definidos pela empresa noticiosa), na versão digital o usuário passa a ter múltiplas possibilidades de fluxos a seguir ao longo da experiência de uso do produto jornalístico. Nesse aspecto, a presença de elementos que ofereçam fluxos de navegação bem estruturados possibilita ao usuário percorrer os principais conteúdos e obter informações de interesse de diferentes maneiras.

Como elemento de oferta e/ou recomendação de conteúdo, entende-se qualquer componente que exiba para o usuário novos conteúdos disponíveis para a sequência do fluxo de navegação. A figura 1 apresenta um exemplo de página de conteúdo do site G1 (g1.com.br), com os elementos de oferta e/ou recomendação destacados em amarelo e ampliados para melhor visualização.

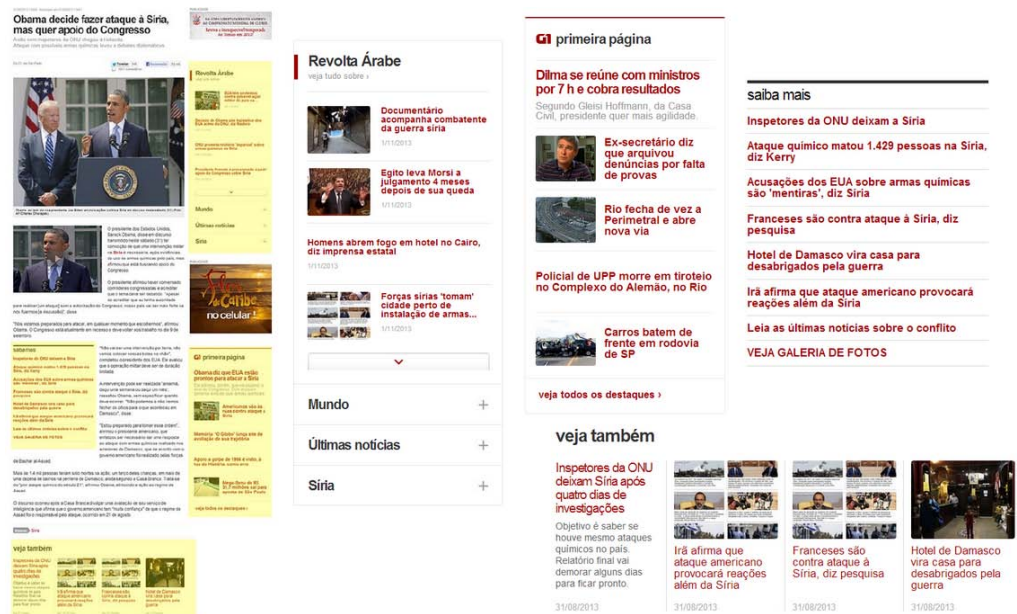


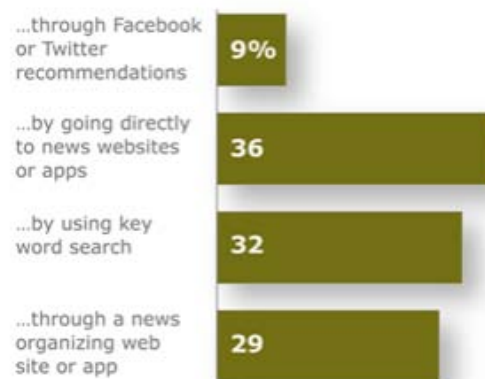
Figura 1. Exemplo de matéria do site G1 (g1.com.br) e dos elementos para consumo de novos conteúdos.

Além da importância como solução de navegação, esses elementos ganham mais relevância quando se observa o comportamento dos usuários em sites de notícias. Estatísticas de acesso dos mostram que as homepages

desses sites estão declinando em sua função de porta de entrada dos usuários, que acessam o site diretamente pelo conteúdo – normalmente a página de matéria – vindo a partir de buscadores, redes sociais e outras fontes externas. Em pesquisa realizada pelo The Pew Research Center's Project for Excellence in Journalism (figura 2), observa-se que apenas 36% dos usuários acessava diretamente um site de notícias a partir da url da página inicial do site em 2012 (Apesar de não haverem dados mais recentes, estima-se que este percentual esteja se reduzindo a cada ano). Os acessos originados a partir de buscadores, redes sociais ou outras fontes externas, quando consolidados, já representam a maioria dos acessos, e nesses casos, a página de conteúdo torna-se a landing page (primeira página que o usuário acessa em um site) para a maior parte dos usuários.

Social media is not an overwhelming driver of news (yet)

Percent of U.S. adults who get news on any digital device very often...



N=3,016

PEW RESEARCH CENTER'S PROJECT FOR EXCELLENCE IN JOURNALISM
2012 STATE OF THE NEWS MEDIA

Figura 2. Percentual de adultos americanos que acessa notícias em dispositivos digitais por diferentes origens.

Dessa forma, a navegação acontece menos orientada à organização de leitura proposta editorialmente pelas homepages e mais estruturada a partir dos elementos de oferta de conteúdo presentes na própria página da matéria noticiosa, de forma muito mais orgânica e contextual. Em função disso, os critérios utilizados para a recomendação de conteúdo em sites de notícias e suas respectivas taxas de assertividade são fundamentais para a sequência da navegação de seus usuários. A utilização de sistemas de

recomendação com algoritmos que automatizam esse processo vêm se apresentando como um dos modelos mais eficientes na otimização desse processo e na melhoria do engajamento do usuário com o conteúdo noticioso.

Esse artigo tem por objetivo apresentar os sistemas de recomendação de conteúdo a partir da categorização proposta por Ricci (2011), e buscando exemplos nos sites de notícias atuais que possam ser relacionados com cada uma dessas categorias. Para aplicação da metodologia proposta, foram observados os elementos de oferta de conteúdo presentes nos seguintes sites de notícias: Globo.com, G1, O Globo, Terra, Uol, Folha, R7, NY Times, CNN, El País, BBC, Google News e Yahoo News. Tal amostragem foi definida de maneira arbitrária pelo autor, em função da relevância editorial das empresas e do estudo não se propor a fazer um levantamento geral dos sites, mas apenas como caráter ilustrativo do uso de algoritmos de recomendação no ambiente noticioso.

2. Sistemas de Oferta e/ou recomendação de conteúdo e exemplos práticos de aplicação

Segundo Hendrix (2014), a curadoria de conteúdo em produtos de notícia sempre foi de responsabilidade do corpo editorial. Entretanto, o crescimento exponencial no volume e na variedade de conteúdos publicados nos sites de notícias tornou cada vez mais difícil a realização dessa curadoria de forma manual ou não-automatizada, e em função disso, a tarefa de apresentar conteúdos verdadeiramente relevantes para o usuário se torna a cada dia mais complexa.

Empresas de outras áreas, como Amazon, Spotify e Netflix, também fazem uso de sistemas de recomendação e outras ferramentas algorítmicas, buscando oferecer conteúdo de maneira precisa para seus usuários. Para se ter uma ideia da eficiência desses mecanismos, no Netflix cerca de 75% da atividade dos usuários se dá a partir das recomendações feitas por esses sistemas. Entretanto, no âmbito da notícia, os sistemas desenvolvidos ainda não possuem a mesma eficiência. Para Hendrix, isso se deve a duas principais razões: Primeiro, para construir um bom sistema de recomendação é necessário ter uma grande quantidade de dados a respeito dos seus usuários. Ao contrário do contexto das empresas citadas anteriormente, a relação dos usuários com as empresas noticiosas é um pouco diferente. Normalmente a navegação não se dá de maneira logada, e os usuários tendem a consumir notícias de múltiplas fontes, não apenas de um único site. Em segundo lugar, o interesse por notícias é variável (a

mesma pessoa pode se interessar muito por futebol durante a Copa do Mundo mas no resto do tempo não ter interesse em esportes), ao passo que o gosto por música e filmes tende a ser muito mais homogêneo. Isso faz com que o investimento (financeiro e intelectual) em sistemas de recomendação de notícias tenha que ser elevado, o que dificulta sua implementação.

Do ponto de vista conceitual, em sistemas de recomendação de conteúdo baseados em algoritmos, o objetivo é estimar, a partir de um conjunto de princípios e regras, quais itens serão potencialmente interessantes para um determinado usuário, antes que ele tenha manifestado ativamente seu interesse pelo conteúdo recomendado (através de uma interação prévia com o site, por exemplo). O modelo da figura 3 representa o funcionamento básico dos sistemas de recomendação de conteúdo. Na figura, a matriz representa a relação entre usuários e itens, onde as pontuações equivalem à classificação dada ao interesse de um usuário específico por um item determinado. O trabalho dos sistemas de recomendação é, portanto, prever qual a classificação do interesse do usuário nos itens onde ainda não foi possível aferir um grau de classificação (as lacunas em branco na matriz da figura). Em um sistema de recomendação tradicional, os itens com maior classificação de interesse tendem a ser os itens recomendados para o usuário.

		<i>Items</i>					
		<i>1</i>	<i>2</i>	...	<i>i</i>	...	<i>m</i>
<i>Users</i>	<i>1</i>	5	3		1	2	
	<i>2</i>		2				4
	:			5			
	<i>u</i>	3	4		2	1	
	:					4	
<i>n</i>			3	2			
		<i>a</i>	3	5		?	1

Figura 3. Matriz de avaliação de usuários em sistemas de recomendação, onde cada célula corresponde a nota de classificação do item “i” para o usuário “u”. A tarefa é prever as classificações faltantes, baseando-se em classificações previamente atribuídas (MELVILLE, 2010).

Entretanto, apesar de essencialmente a maioria dos sistemas de recomendação funcionar com base nesse princípio, os critérios pelos quais o interesse do usuário é mensurado podem variar muito, e diferentes algoritmos podem resultar em diferentes recomendações. Segundo Ricci et al (2011), os sistemas de recomendação podem ser categorizados em 6 principais grupos, determinados a seguir:

- Content Based (Recomendação por conteúdo): O sistema aprende a recomendar itens similares ao que o usuário está acessando naquele momento. A similaridade entre os itens é definida pelos critérios estabelecidos e atribuídos aos itens comparados. No exemplo da figura 4, ao acessar uma notícia sobre "Flamengo", outros conteúdos sobre o mesmo assunto são recomendados na coluna da direita.

The image shows a screenshot of a news article on the website Globoesporte.com. The main article is titled "Fla confirma contratações, e Ayrton e Alan Patrick serão apresentados sexta" (Flamengo confirms signings, and Ayrton and Alan Patrick will be presented on Friday). The article text mentions that the duo, previously from Palmeiras, has signed with Flamengo for the Campeonato Brasileiro and Copa do Brasil. Below the text is a photo of the two players in Flamengo jerseys. To the right of the article is a sidebar with a red advertisement for Brahma beer and a section titled "TUDO SOBRE Flamengo" (Everything about Flamengo). This sidebar contains several recommended articles, such as "BLOG: Alessandro grava vídeo para torcida do Fla e diz: 'Não é 'tchau', quem sabe 'até logo'" (Blog: Alessandro records video for Flamengo fans and says: "It's not 'goodbye', who knows 'until later'") and "Empresa confirma Vasco e Flamengo na Arena Pantanal, em Cuiabá" (Company confirms Vasco and Flamengo at Arena Pantanal, in Cuiabá).

Figura 4. Print de página de matéria do site Globoesporte.com, utilizando um elemento de oferta de conteúdo do tipo content-based na coluna à direita.

Ao longo da pesquisa, foi observado que os sistemas de recomendação por conteúdo são os mais utilizados em sites de notícias.

- Collaborative filtering (Filtragem colaborativa): O sistema recomenda ao usuário outros conteúdos que interessaram a usuários com características similares. O foco deixa de ser o conteúdo (item) e passa a ser o perfil de quem acessa. A similaridade de gostos entre usuários é calculada com base em avaliações comuns entre eles, acessos a conteúdos parecidos e padrões de comportamento comparativamente próximos.

Em sites noticiosos, a filtragem colaborativa ainda é usada de maneira muito incipiente. No exemplo da figura 5, o sistema de recomendação se utiliza das estatísticas de acesso de outros usuários do site, mas sem estabelecer uma relação direta entre estes e o usuário específico que acessa o conteúdo.

+ LIDAS	+ COMENTADAS	+ ENVIADAS	ÚLTIMAS
1	'Rei da soja', empresário Olacyr de Moraes morre aos 84 anos		
2	Governo e base aliada chegam a acordo por projeto de desoneração da folha		
3	Queda 'rápida e profunda' no comércio surpreende economistas		
4	Olacyr de Moraes, de rei da soja a súdito dos credores		
5	Bala Juquinha voltará a ser fabricada por empresário que comprou fórmula		

Figura 5. Elemento que recomenda as matérias "mais lidas" por todos os usuários do site (www.uol.com.br).

- Demographic (Recomendação Demográfica): O sistema recomenda conteúdos baseando-se no perfil demográfico do usuário. É assumido que pessoas de uma mesma região tendem a ter interesse em conteúdos similares, independente da navegação prévia de cada um dos indivíduos.

RIO DE JANEIRO

há 1 hora - Curte + de 2

pena que é uma trabalho acadêmico... eu sugeriria você colocar que isso é falta de sexo mesmo! (pra não dizer falta de pau)
há 1 minuto - Curte

Professor polemiza em post: 'falta de sexo'



Carta cita 'tão falada homossexualidade'

mais rio de janeiro > **ESCOLHA SUA REGIÃO**

Figura 6. No site Globo.com, o usuário tem a opção de receber conteúdos específicos da sua região.

- Knowledge-based (Recomendação por conhecimento de domínio): A recomendação é feita a partir de um conhecimento prévio do domínio do conteúdo em questão, independente de uma similaridade clara entre características dos itens comparados ou do interesse prévio de outros usuários. Por exemplo, se um usuário em um site de compras manifesta interesse em adquirir espetos para churrasco e carvão, o sistema pode identificar que o domínio de interesse trata da realização de um churrasco, e recomendar outros itens de possível interesse dentro do mesmo domínio, como carnes, cerveja, acendedores, etc.

Nos sites de notícias observados não foi possível identificar o uso desse tipo de sistema de recomendação.

- Community-based (Recomendação por Comunidade) : A recomendação é baseada no que os amigos do usuário manifestaram interesse. Se um amigo de um usuário leu uma determinada notícia, entende-se que essa notícia tenha grande probabilidade de interessar ao usuário. Na figura 7, o elemento "Seus amigos no G1" recomenda para o usuário conteúdos que já foram previamente acessados por seus amigos no Facebook.



Figura 7. No site G1, o elemento "Seus amigos no G1" recomendam conteúdos que foram previamente acessados por amigos do usuário.

- Hybrid recommender systems (Sistemas de recomendação híbridos): Sistemas que utilizam uma combinação de 2 ou mais sistemas citados anteriormente. A combinação dos sistemas permite o aproveitamento das vantagens dos sistemas e o descarte de possíveis desvantagens que cada um deles apresenta individualmente. Tendem a ser modelos mais complexos, mas comparativamente apresentam os melhores resultados de acerto na oferta de conteúdo relevante.

Apesar de sua relativa complexidade de implementação (pela combinação de múltiplos algoritmos), esse sistema vem se popularizando nos sites de notícias em função de empresas que vêm se especializando em terceirizar o serviço para sites de notícias, como no exemplo da figura 8, que utiliza um box de recomendação de conteúdo oferecido pela empresa Outbrain (outbrain.com).



Figura 8. Sistema de recomendação híbrido em matéria do site Terra (terra.com).

3. Considerações Finais

O artigo apresenta portanto um levantamento teórico e prático do cenário atual dos sistemas de recomendação de conteúdo nas páginas de matérias dos sites de notícias, apresentando exemplos reais de elementos presentes nesse sites que possam ser associados a cada uma das categorias definidas e detalhadas anteriormente.

A partir dessa análise, foi possível observar quais dos sistemas de recomendação de conteúdo são mais comumente utilizados nos sites de notícias atuais, e que a categorização proposta por Ricci (2011) atende bem ao cenário do jornalismo digital.

Referências

DIAKOPOULOS, N. **Algorithmic accountability reporting: on the investigation of black boxes.** 2013. Disponível em:

<http://towcenter.org/wp-content/uploads/2014/02/78524_Tow-Center-Report-WEB-1.pdf>.

HENDRIX, P. **5 ways startups are narrowing the content discovery gap**. 2014. Disponível < <https://gigaom.com/2014/03/15/5-ways-startups-are-narrowing-the-content-discovery-gap> >.

MELVILLE, P; SINDHWANI, V. **Encyclopedia of Machine Learning**. 2010. Disponível em: <<http://www.prem-melville.com/publications/recommender-systems-eml2010.pdf>>.

RICCI, F. et al. (eds.). **Recommender Systems Handbook**, 1. DOI 10.1007/978-0-387-85820-3_1, © Springer Science+Business Media, LLC 2011

Como citar

LOHMANN, Augusto; MONAT, André. **Sistemas de recomendação de conteúdo em sites de notícias**. Arcos Design. Rio de Janeiro: PPD ESDI - UERJ. Volume 9 Número 2 Dezembro 2016. pp. 65-76. Disponível em: [<http://www.e-publicacoes.uerj.br/index.php/arcosdesign>]

DOI



A Revista Arcos Design está licenciada sob uma licença Creative Commons Atribuição - Não Comercial - Compartilha Igual 3.0 Não Adaptada.