

Desenvolvimento de um Sistema de Recomendação de Candidatos a Cargos Públicos Utilizando Técnicas de Aprendizagem de Máquina e Filtragem de Informação

Lucas Pereira Reis¹, Jailson Pereira Januário¹, Fábio Santos da Silva¹

¹Laboratório de Sistemas Inteligentes
Grupo de Pesquisas em Sistemas Inteligentes
Escola Superior de Tecnologia
Universidade do Estado do Amazonas (UEA), AM.

{lprs.eng, jpj.ads, fssilva}@uea.edu.br

Abstract. *Every four years the brazilians population is lost among the excessive number of candidates of political positions and ends up no finding a suitable candidate. This article presents a proposal for a Candidates Recommender System for political positions using machine learning and information filtering techniques aiming to ease voters' lives, presenting candidates that match their profiles.*

Resumo. *De quatro em quatro anos a população brasileira se perde na excessiva quantidade de candidatos a cargos políticos e acaba não conseguindo encontrar um candidato adequado, tornando votar uma tarefa árdua. Este artigo apresenta a proposta de um Sistema de Recomendação de candidatos a cargos públicos utilizando técnicas de aprendizagem de máquina e filtragem de informação visando facilitar a vida dos eleitores, apresentando candidatos que possuem seu perfil.*

1. Introdução

Eleição é todo processo onde um grupo de pessoas escolhe seus representantes de algum cargo público através de votações. As eleições no Brasil são realizadas através do voto direto, secreto e obrigatório.

No Brasil a cada eleição muitos políticos são candidatos para inúmeros cargos, gerando uma dificuldade dos eleitores em analisar todas as propostas de todos os candidatos para votar no mais apropriado. Esta dificuldade de escolha de uma opção diante de uma diversidade é conhecida na literatura como o problema da Sobrecarga de Informação. Uma das consequências deste problema é que o eleitor acaba escolhendo o candidato por meio de recomendações de amigos, por *marketing* através de TV, rádios, jornais, revistas etc.

Sistemas de recomendação estão presentes em nosso cotidiano em diversos tópicos, incluindo no mundo político. Normalmente vemos estes sistemas serem apresentados em páginas *web* ao invés do meio acadêmico como em [Época

Negócios Online 2014]. Em [Lim et al. 2017] temos um sistema similar que utiliza Filtragem Híbrida para recomendar candidatos da Coréia do Sul.

2. Fundamentação Teórica

Nesta seção serão descritos os conceitos e as técnicas utilizadas para o desenvolvimento deste trabalho.

2.1. Sistemas de Recomendação

Segundo [Adomavicius and Tuzhilin 2005], os Sistemas de Recomendação podem ajudar o usuário a lidar com o problema da sobrecarga de informação. Estes sistemas aplicam técnicas de análises de dados e filtragem de informação para recomendar itens relevantes de acordo com o perfil do usuário estabelecido. Sua construção é baseada em três técnicas da área de Recuperação de Informação: Filtragem Baseada em Conteúdo (FBC), Filtragem Colaborativa (FC) e Filtragem Híbrida (FH).

2.2. Filtragem Híbrida

A FH é uma técnica obtida a partir da combinação da FC com a FBC, onde a FC trabalha combinando as opiniões de usuários que expressaram inclinações similares para o usuário atual [Breese et al. 1998] e FBC trabalha analisando a descrição de itens para identificar quais são do interesse particular do usuário [Pazzani and Billsus 2007]. É possível combiná-las conforme visto em [Burke 2002].

3. Materiais e Métodos

Nesta seção serão descritos a forma de obtenção do *dataset* utilizado e como o sistema está arquitetado.

3.1. Dados Experimentais

Para implementar a FC, é necessário primeiramente desenvolver a FBC e a FC, as duas filtragens possuem seus próprios *datasets*. Para esta etapa da FBC foi obtido os dados dos ex-candidatos à presidência a partir de seus perfis no *Twitter*, onde foi realizado a extração de seus *tweets* durante a época das eleições de 2018 e feito a contagem de termos que são semelhantes ou sinônimos das seguintes áreas de interesse: cultura, economia, educação, meio ambiente, saúde, segurança e tecnologia, por fim utilizou-se a métrica TF-IDF para mensurar o grau de relevância do candidato para cada área dentro de sua proposta [Ramos et al. 2003].

Os dados da FC são obtidos a partir da coleta de opiniões de usuários, para isso criou-se um simples formulário no Google onde é solicitado uma avaliação de 1 a 5 dos ex-candidatos à presidência. Foi obtido ao todo 107 respostas e as avaliações desses usuários foram adicionadas ao *dataset*.

3.2. Arquiteturas de Sistemas de Recomendação

O protótipo desenvolvido para o sistema foi desenvolvido utilizando *Flask* com *SQLite*, as três filtragens de informação foram desenvolvidos à parte, assim o sistema apenas realiza a chamada do algoritmo. A seguir a arquitetura desenvolvida.

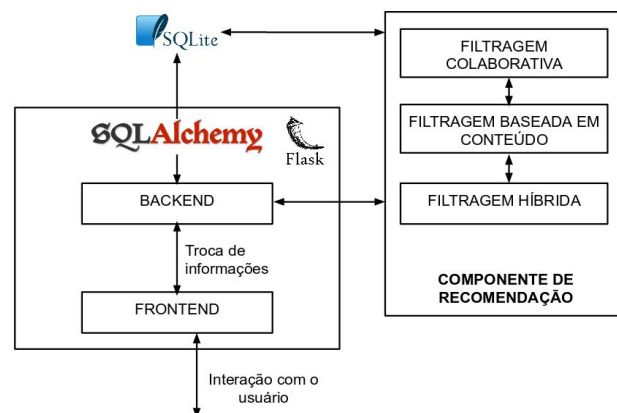


Figure 1. Arquitetura utilizada para o protótipo.

3.3. Descrição da Tarefa de Recomendação

Como pode ser observado na figura 1, o método utilizado para a FH é o *Cascade* [Burke 2002]. A FC encontra o(s) usuário(s) mais similar(es) ao usuário atual e recomenda candidatos analisando o perfil de cada um, a saída da FC é a entrada da FBC, e esta filtragem irá ordenar os candidatos pela similaridade com o usuário atual de acordo com o perfil especificado.

4. Resultados e Discussão

Para avaliar o protótipo desenvolvido para o sistema, foi convidado um grupo de pessoas para utilizá-lo, cada um especificou quais candidatos são relevantes e solicitou a recomendação do sistema. Utilizou-se métricas clássicas de sistema de recomendação para a avaliação conforme visto em [Shani and Gunawardana 2011].

Table 1. Desempenho do sistema por usuário e média.

Usuário	Precisão	Revocação	Medida-F	Nota
1	0.67	0.67	0.67	4.0
2	1.0	0.5	0.67	5.0
3	0.67	0.67	0.67	3.0
4	1.0	0.67	0.8	4.0
5	0.0	0.0	0.0	4.0
6	0.5	0.33	0.4	3.0
7	0.0	0.0	0.0	4.0
8	0.5	1.0	0.67	5.0
9	0.5	0.5	0.5	3.0
10	0.5	0.5	0.5	4.0
11	0.33	0.5	0.4	4.0
12	1.0	0.33	0.5	4.0
13	1.0	0.33	0.5	3.0
14	0.67	0.67	0.67	4.0
15	1.0	1.0	1.0	5.0
Média	62.22%	51.11%	52.88%	3.93

5. Considerações Finais

Podemos observar na seção anterior que os resultados obtidos não foram agradáveis. É importante ressaltar que em muitos casos os candidatos relevantes ao usuário são por questões ideológicas ou recomendações de amigos e mídias, portanto é possível que o perfil especificado pelo usuário seja equivalente a candidatos recomendados que não são relevantes, ocasionando uma insatisfação do usuário e uma queda nas suas métricas.

A continuação deste projeto irá se focar primeiramente em melhorar o algoritmo de recomendação e desmembrar as duas filtragens implementadas. Para a FBC será explorado diferentes abordagens utilizando a API do *Twitter* (como análise de sentimento) para enriquecer ainda mais o *dataset*, caso seja possível será realizado em conjunto com um modelo baseado em aprendizagem de máquina. Por fim, o protótipo em *Flask* será transformado em uma API para o *backend*, e o *frontend* será desenvolvido utilizando a linguagem *Javascript* para tornar a interface mais rica e agradável.

Agradecimentos

Os autores agradecem o apoio material e financeiro provido pela Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) por meio do PPP 04/2017.

Referências

- Adomavicius, G. and Tuzhilin, A. (2005). 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): 734-749.
- Breese, J. S., Heckerman, D., and Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 43–52. Morgan Kaufmann Publishers Inc.
- Burgarelli, R. (2016). Voto inconsciente: o que o brasileiro leva em conta para decidir seu candidato a prefeito?. <https://alias.estadao.com.br/noticias/geral,voto-inconsciente-o-que-o-brasileiro-leva-em-conta-para-decidir-seu-candidato-para-prefeito,10000077956>. Acessado em: 22 de abril de 2018.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370.
- Lim, K., Kim, C., Kim, G., and Choi, H. (2017). A recommender system for political information filtering. In *International Conference of Design, User Experience, and Usability*, pages 129–145. Springer.
- Pacievitch, T. Eleições no brasil. <https://www.infoescola.com/direito/eleicoes-no-brasil/>. Acessado em: 22 de abril de 2018.
- Pazzani, M. J. and Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer.
- Ramos, J. et al. (2003). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, volume 242, pages 133–142.
- Shani, G. and Gunawardana, A. (2011). *Evaluating Recommendation Systems*, volume 12, pages 257–297.
- Época Negócios Online (2014). 10 aplicativos e sites que ajudam a decidir em quem votar. <https://epocanegocios.globo.com/Informacao/Acao/noticia/2014/10/10-aplicativos-e-sites-que-ajudam-decidir-em-quem-votar.html>. Acessado em: 27 de abril de 2018.