

**Ranqueamento de atributos de produtos com base na  
opiniões de consumidores**

**Orientador(a): Tiago Eugenio de Melo  
Bolsista: João Victor Silva de Souza**

## 1. Introdução

Ao tomar decisões de compra de um produto, os consumidores geralmente confiam em informações de dois tipos de fontes: especificações dos produtos fornecidas pelos fabricantes e comentários postados por outros consumidores em sites de comércio eletrônico, tais como Amazon.com ou Saraiva.com.br. Porém, em alguns competitivos segmentos de mercado, tais como o segmento de aparelhos de celular, muitos fabricantes produzem aparelhos com características muito parecidas. Por exemplo, os aparelhos “Moto Z Play” e “Samsung Galaxy S7 Plus” têm o mesmo tamanho de tela, sistema operacional, resolução de câmera, entre outras características. Portanto, quase não existe diferença entre esses produtos. Especialmente em casos como este, os comentários de outros usuários têm um papel importante na tomada de decisão do consumidor.

Infelizmente, não é viável para um consumidor examinar o grande volume disponível de comentários para manualmente comparar produtos similares. Por exemplo, existem mais de 2.000 comentários sobre o aparelho Samsung Galaxy S7 Edge postados no site da Amazon. Este tipo de problema tem sido tratado por técnicas que identificam sentenças comparativas de comentários [1,2,3]. Porém, essas técnicas possuem algumas limitações. Primeiramente, esses trabalhos não são capazes de identificar os alvos nas sentenças comparativas [3]. Além disso, sentenças comparativas representam apenas 10% do total de sentenças subjetivas em comentários de usuários [4]. Finalmente, sentenças comparativas são mais comuns em fóruns de discussão do que em comentários de sites de comércio eletrônico [2].

Outra abordagem para comparar as opiniões de produtos é considerar as avaliações das notas publicadas pelos usuários. Por exemplo, além dos comentários publicados, os sites permitem que os usuários dêem uma nota para os produtos. Essas notas costumam ser representadas pelo número de estrelas. Porém, a utilidade dessas notas postadas por usuários é limitada para potenciais consumidores, pois a nota representa uma avaliação para o produto como um todo e pode combinar tanto avaliações positivas como negativas relativas a diferentes atributos do produto. Além disso, esse tipo de avaliação não traz qualquer informação do por quê os usuários gostaram de um produto ou quais características do produto que eles mais gostaram. Um usuário que queira comprar um aparelho celular pode querer saber a opinião dos demais usuários sobre a bateria ou a tela, e não apenas uma nota geral.

Uma maneira tradicional de organizar o grande volume de comentários de usuários é criar um resumo das opiniões. Esse resumo fornece uma lista de aspectos dos produtos e as suas correspondentes opiniões, onde a abordagem mais comum é a sumarização de opinião baseada em aspectos [5]. Porém, essas técnicas não são adequadas para suportar consultas de usuários sobre características específicas dos produtos. A razão é que nos métodos atuais, as opiniões são agrupadas aleatoriamente por aspectos, mas esses aspectos não são, necessariamente, associados aos atributos dos produtos. Por exemplo, ao usar esse tipo de técnica, muitos grupos de opiniões relativas a tela podem surgir: um grupo de opiniões sobre a resolução e a cor, outro grupo que mistura brilho e tamanho, e assim por diante. Dessa maneira, o potencial consumidor ainda teria que executar uma tarefa não-trivial de identificar quais grupos de aspectos se referem ao mesmo aspecto de interesse. De fato, Zha et al. [6] reportou que para o aparelho “iPhone 3GS” existem mais de 3.000 aspectos encontrados nos comentários. Sumarizar esse grande volume de informações irá gerar centenas de grupos, sem identificar quais grupos são relativos a tela do produto.

Para lidar com o problema descrito acima, particularmente na categoria de aparelhos de celular, nós projetamos e implementamos um sistema chamado *Contender*. Esse sistema sumariza comentários de usuários alinhando-os às especificações dos atributos desses produtos. O sistema é capaz de lidar com comentários reais de usuários de sites de comércio eletrônico como fonte de dados. O sistema extrai as opiniões dos comentários e então mapeia essas opiniões para os atributos definidos pela especificação dos produtos. As tarefas de extrair opiniões de comentários e de mapear essas opiniões para atributos de produtos é decorrente de uma pesquisa que publicamos nos periódicos [7] e [8].

A Figura 1 abaixo mostra os principais módulos do aplicativo. Quando o usuário abre o aplicativo, o sistema mostra duas caixas de texto onde o usuário pode digitar os nomes dos aparelhos de interesse. Como mostrado na Figura 1 (a), o usuário digita dois nomes: “Motorola Moto X4” e “Samsung Galaxy S7 Edge”. Nosso sistema mostra o resultado da consulta de entrada em três diferentes módulos: “Specs”, “Reviews” e “Charts”.



O módulo “Specs” mostra as especificações de cada produto fornecidas pelos fabricantes e disponibilizadas em sites de comércio eletrônico (Figura 1 (b)). O módulo “Reviews” mostra os comentários de usuários para cada produto. O diferencial, quando comparamos com os recursos oferecidos em sites de comércio eletrônico, é que os comentários podem ser filtrados pelos atributos dos produtos. Por exemplo, se o usuário selecionar “bateria”, o sistema mostrará apenas os comentários relativos a esse atributo (Figura 1 (c)). O módulo “Charts” mostra quatro diferentes tipos de gráficos para que o usuário possa realizar a comparação de produtos: gráfico de barras; gráfico radar; gráfico de pizza; e ranqueamento dos atributos. Por exemplo, a Figura 1 (d) mostra os atributos com uma avaliação calculada automaticamente. Isto, somente é possível porque o sistema permite associar (mapear) opiniões de usuários para os atributos dos produtos.

Contender foi apresentado recentemente como uma demo na 41ª Conferência Européia de Recuperação de Informação na cidade de Cologne na Alemanha [7]. Apesar do bom resultado alcançado, o método que propusemos para gerar o ranqueamento dos produtos é baseado na simples operação aritmética das notas das opiniões. Diante desse contexto, a proposta de projeto aqui apresentada tem o objetivo de estudar diferentes técnicas, tais como [10,11,12,13,14] que fazem o ranqueamento de aspectos de produtos e implementar a técnica mais apropriada no sistema Contender.

## 2. Objetivo(s)

O objetivo deste estudo é investigar o problema de ranqueamento de atributos de produtos com base nas opiniões publicadas pelos usuários em sites de comércio eletrônico e implementar um método que trate desse problema.

### Objetivos Específicos:

a) Estudar sobre trabalhos sobre o tema de mineração de opinião (análise de sentimentos). Duração: 1 (um) mês.

b) Estudar a versão atual de funcionamento do aplicativo e das técnicas usadas no Contender. Duração: 2 (dois) meses.

c) Pesquisar sobre trabalhos que façam o ranqueamento de atributos de produtos com base nas opiniões de usuários. Duração 2 (dois) meses.

d) Implementar métodos que permitam ranquear os atributos de produtos com base nas opiniões de usuários. Duração: 5 (cinco) meses.

e) Avaliar a performance dos métodos implementados e, eventualmente, adaptar esses métodos para obter resultados mais efetivos. Duração: 3 (três) meses.

f) Implementação do método mais efetivo de ranqueamento na ferramenta Contender. Duração: 3 (três) meses.

### **3. Metodologia**

O projeto de pesquisa será executado nas seguintes fases:

#### Fase 1:

Na primeira etapa, o aluno estudará sobre o tema de mineração de opinião, também conhecido como análise de sentimentos. Além disso, fará o estudo sobre o funcionamento do Contender e das técnicas que foram empregadas. O professor-orientador proponente é um dos autores desse método e irá auxiliar o aluno na compreensão do método.

#### Fase 2:

Na segunda etapa, o aluno fará uma pesquisa sobre possíveis métodos que realizem o ranqueamento de atributos de produtos. Essa etapa de investigação poderá envolverá implementação e/ou testes de possíveis métodos que possam ser usados como referências.

#### Fase 3:

Na terceira etapa, o aluno deverá implementar as possíveis técnicas investigadas na fase anterior.

#### Fase 4:

Na etapa final, o aluno deverá realizar testes para avaliar a performance dos métodos implementados e, eventualmente, comparar com possíveis *baselines*. Além disso, o aluno implementará o método investigado que obteve a melhor performance na ferramenta Contender.

#### 4. Resultados esperados

Espera-se como resultado do projeto de pesquisa proposto o desenvolvimento de novas tecnologias que tenham tanto relevância acadêmica, caracterizada por meio da publicação de trabalho científico de impacto, como também relevância econômica, caracterizada pelo desenvolvimento de protótipo que demonstrem a viabilidade dos métodos propostos no projeto.

#### 5. Cronograma de execução

Meta/Atividade	1	2	3	4	5	6	7	8	9	10	11	12
Revisão bibliográfica sobre o tema e sobre o problema proposto.												
Pesquisa sobre possíveis métodos para resolução do problema.												
Implementação dos métodos identificados na fase anterior.												
Realização de testes para avaliar a performance dos métodos e a eventual implementação de possíveis melhorias.												
Implementação de um protótipo que permita visualizar o funcionamento do método desenvolvido nessa pesquisa.												
Elaboração do relatório final e submissão de artigo científico para publicação com os resultados obtidos.												

## 6. Referências

- [1] Jindal, N., & Liu, B. (2006, August). Identifying comparative sentences in text documents. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 244-251). ACM.
- [2] Ganapathibhotla, M., & Liu, B. (2008, August). Mining opinions in comparative sentences. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1* (pp. 241-248). Association for Computational Linguistics.
- [3] Varathan, K. D., Giachanou, A., & Crestani, F. (2017). Comparative opinion mining: a review. *Journal of the Association for Information Science and Technology*, 68(4), 811-829.
- [4] Kessler, W., & Kuhn, J. (2013). Detection of product comparisons-how far does an out-of-the-box semantic role labeling system take you?. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1892-1897).
- [5] Moussa, M. E., Mohamed, E. H., & Haggag, M. H. (2018). A survey on opinion summarization techniques for social media. *Future Computing and Informatics Journal*, 3(1), 82-109.
- [6] Zha, Z. J., Yu, J., Tang, J., Wang, M., & Chua, T. S. (2014). Product aspect ranking and its applications. *IEEE transactions on knowledge and data engineering*, 26(5), 1211-1224.
- [7] de Melo, T., da Silva, A., & de Moura, E. S. (2018). An aspect-driven method for enriching product catalogs with user opinions. *Journal of the Brazilian Computer Society*, 24(1), 15.
- [8] de Melo, T., da Silva, A. S., de Moura, E. S., & Calado, P. (2019). OpinionLink: Leveraging user opinions for product catalog enrichment. *Information Processing & Management*, 56(3), 823-843.
- [9] de Melo, T., da Silva, A. S., de Moura, E. S., & Calado, P. (2019, April). Contender: Leveraging User Opinions for Purchase Decision-Making. In *European Conference on Information Retrieval* (pp. 230-235). Springer, Cham.
- [10] Zha, Z. J., Yu, J., Tang, J., Wang, M., & Chua, T. S. (2014). Product aspect ranking and its applications. *IEEE transactions on knowledge and data engineering*, 26(5), 1211-1224.

[11] Harper, F. M., Xu, F., Kaur, H., Condiff, K., Chang, S., & Terveen, L. (2015, September). Putting users in control of their recommendations. In *Proceedings of the 9th ACM Conference on Recommender Systems* (pp. 3-10). ACM.

[12] Wu, Y., & Ester, M. (2015, February). Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (pp. 199-208). ACM.

[13] Debnath, S., Ganguly, N., & Mitra, P. (2008, April). Feature weighting in content based recommendation system using social network analysis. In *Proceedings of the 17th international conference on World Wide Web* (pp. 1041-1042). ACM.

[14] Kagie, M., van Wezel, M., & Groenen, P. J. (2009). Determination of attribute weights for recommender systems based on product popularity.