

Improving Social CRM through eletronic word-of-mouth: a case study of ReclameAqui

Gustavo R. T. de Almeida
Universidade Federal do Oeste do
Pará
Santarém, Pará
gr.almeida00@gmail.com

Douglas Rocha Cirqueira
Universidade Federal do Pará
Belém, Pará
douglas.cirqueira@itec.ufpa.br

Fábio M. F. Lobato
Universidade Federal do Oeste do
Pará
Santarém, Pará
fabio.lobatof@ufopa.edu.br

ABSTRACT

The ways of relationship between companies and customers have changed dramatically due to web-users engagement with Online Social Networks, impacting on the way consumers make purchase decisions. As a consequence, this phenomenon brought up a new concept, the Social Customer Relationship Management. In this context, Electronic Word-of-Mouth plays a significant role since customers are more exigent and having easy access to steadily more information about products, services and brands reputation. However, many reputation platforms do not provide an interface for automatic data collection. Thus, these data are not used in business decision making. Aiming to fill these gaps we propose an automatic data retrieval and analysis methods to knowledge extraction from Electronic Word-of-Mouth platforms, providing more reliability to the decision-making processes. The analysis consists in topic modeling using Latent Dirichlet Allocation to identify most frequent complaints and their correlations. Both, the data retrieval method and the analysis method are platform independent. A well-known reputation platform in Brazil was used to evaluate our proposal, the ReclameAqui.

KEYWORDS

eWoM, Social Media, Social CRM, Topic Modelling

1 INTRODUÇÃO

O engajamento dos usuários web com as redes sociais é notório [4], impactando significativamente na forma como os consumidores tomam decisões [14]. Como consequência, um novo conceito surgiu, o *Social Customer Relationship Management* (Social CRM), o qual representa a combinação entre as práticas de gerenciamento de clientes utilizadas nas empresas e os mecanismos de monitoramento e de mineração de dados de redes sociais [1]. Esse novo ambiente propiciou a troca de opiniões e experiências sobre os produtos/serviços entre consumidores. Neste contexto, percebeu-se que os dados gerados pelos usuários trazem consigo informações importantes sobre processos internos falhos ou problemas nos produtos [10]. Sendo assim, analisar estes dados cuidadosamente pode prover *insights* para melhorias, a fim de capturar *leads* e manter a fidelidade dos atuais clientes [9].

O fenômeno de troca de opiniões a respeito de produtos e serviços nas redes sociais é conhecido por eWoM (acrônimo de *Eletronic Word-of-Mouth*). O boca-a-boca virtual tem chamado atenção crescente das empresas por ser considerado mais efetivo que o boca-a-boca tradicional devido a sua rapidez, conveniência, amplitude, disponibilidade e engajamento [7]. Plataformas como ReclameAqui¹ e Bondfaro² são sistemas de reputação bem reconhecidos no Brasil. Entretanto, diferentemente de redes sociais tradicionais, tais plataformas não fornecem uma interface para a coleta de informações, tornando inviável a análise automática.

Por meio da pesquisa por trabalhos relacionados e de entrevistas com empresas parceiras do *Social CRM Research Center*³, percebeu-se que o desenvolvimento de uma ferramenta para coleta e análise automática de dados advindos de plataformas de eWoM é uma demanda mercadológica notória. Este trabalho tem por objetivo principal sanar esta lacuna no estado da prática. A fim de validar a abordagem proposta junto as empresas parceiras, delimitou-se o escopo de trabalho à plataforma ReclameAqui. Para extrair dados desta plataforma de reclamações, um Web Crawler foi desenvolvido baseando-se no *framework* arquitetural proposto por [11]. O método para modelagem de tópico *Latent Dirichlet Allocation* (LDA) foi aplicado. Como contribuições tangíveis deste trabalho, destacam-se: i) uma ferramenta para extração de dados da plataforma ReclameAqui; ii) conjuntos de dados de reclamações extraídas a partir do ReclameAqui; e iii) uma ferramenta para modelagem de tópicos. Os resultados obtidos são cuidadosamente analisados para responder as seguintes Perguntas de Pesquisa (PP):

- (1) É possível adotar o ReclameAqui como uma fonte de dados para estudos envolvendo eWoM?
- (2) Quais são os benefícios da adoção de análises automáticas dos dados provenientes do ReclameAqui?

Este artigo encontra-se organizado como segue: a Seção 2 apresenta alguns trabalhos relacionados a eWoM, extração de dados de comunidades de eWoM e modelagem de tópicos. Na Seção 3 os estudos de caso utilizados nos experimentos são descritos. As Seções 4 e 5 descrevem as ferramentas para aquisição de dados e modelagem de tópicos, e os resultados experimentais, respectivamente. Por fim, as considerações finais são apresentadas na Seção 6.

2 TRABALHOS RELACIONADOS

O eWoM consiste no compartilhamento de avaliações sobre produtos e serviços por meio de plataformas on-line [2]. Geralmente,

In: XIV Workshop de Trabalhos de Iniciação Científica (WTIC 2017), Gramado, Brasil. Anais do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web: Workshops e Pôsteres. Porto Alegre: Sociedade Brasileira de Computação, 2017.

© 2017 SBC – Sociedade Brasileira de Computação.

ISBN 978-85-7669-380-2.

¹<https://www.reclameaqui.com.br/>

²<http://http://www.bondfaro.com.br/>

³<http://www.ssrc-leipzig.de/en/startpage/>

estas avaliações são baseadas em texto e vêm acompanhadas de uma pontuação que visa refletir o nível de satisfação do cliente. [3] demonstraram os efeitos do eWoM em uma compra e também em mudanças de atitude frente a uma marca. É importante destacar que eWoM não ocorre exclusivamente em plataformas de propósito específico. De acordo com [9], redes sociais de propósito geral, como Facebook ou Twitter, também são utilizadas e a maior parte dos trabalhos utilizam-nas pois elas fornecem *Application Programming Interfaces* (API) para aquisição de dados [13].

No contexto dos sítios web de propósito específico, o ReclameAqui é uma plataforma de reclamações popular no Brasil e que não disponibiliza API para obtenção dos dados. Para superar este obstáculo, investigou-se neste trabalho a adoção de um Web Crawler. [11] propôs um *framework* arquitetural para coletar grandes volumes de dados gerados por usuários de sítios web de eWoM. Como estudo de caso os autores utilizaram a plataforma Ciao⁴. Os resultados mostraram-se promissores e o método desenvolvido foi apto a coletar dados do sítio web em estudo. No presente artigo, o *framework* arquitetural proposto por [11] foi modificado para adaptar-se às particularidades do ReclameAqui.

No que tange à modelagem de tópicos, o LDA tem sido utilizado em uma variedade de aplicações, demonstrando seu potencial na detecção de tópicos e problemas de classificação de *Corpus* linguísticos. [6] aplicou LDA para recomendação de rótulos. Os resultados obtidos sugerem que o LDA alcança melhor precisão (*precision*) e revocação (*recall*) do que usando regras de associação. Em [5], os autores tentaram identificar os efeitos de tópicos escondidos em notícias do setor financeiro, como preços de ações, usando LDA. Em uma linha semelhante, o trabalho de [8] fez uma categorização de tópicos de notícias usando modelos de LDA e uma representação esparsa para aprimorar sistemas de recomendação de notícias. [12] usou uma abordagem LDA multi-modal supervisionada para auxiliar na classificação de eventos de um grande número de dados gerados por usuários, dispostos em mídias sociais.

3 ESTUDOS DE CASO

Com mais 90 mil empresas cadastradas, a plataforma ReclameAqui conecta estas empresas e consumidores para guiá-los na resolução de problemas. Uma lista não exaustiva de funcionalidades disponíveis inclui o ranqueamento das melhores empresas para se fazer negócios (categorizadas por setores de serviço/venda); disponibilização de estatísticas acerca da resolução das reclamações por cada companhia como taxa de solução, tempo de resposta e uma pontuação do atendimento; e, é claro, a lista de reclamações, com a resposta da empresa e um rótulo se o problema foi ou não solucionado. Por meio de uma navegação exploratória pelo conteúdo do site e das reclamações cadastradas, construiu-se um resumo do fluxo de informação e os dados disponibilizados em cada etapa conforme apresentado na Figura 1.

O processo descrito na Figura 1 inicia com uma reclamação por parte do usuário. O segundo passo fica a cargo da empresa, a qual deve prover uma possível solução ou esclarecimento. Em seguida, o usuário pode replicar e avaliar o atendimento, por meio de um questionário. Por fim, as interações descritas são utilizadas para computar as estatísticas supraditas, as quais são usadas no

⁴<http://www.ciao.com/>

Table 1: Resumo dos estudos de caso.

Denominação	Setor	Reclamações
Empresa A	Automotivo	525
Empresa B	Bens de Consumo	609

sistema de ranqueamento. Para avaliar a metodologia proposta, duas empresas foram selecionadas como estudo de caso, por seu alto número de reclamações atuais (Em um período de até 1 ano). Elas foram anonimizadas e são referenciadas aqui como Empresa A e B. A Tabela 1 sumariza as informações das empresas.

4 ABORDAGEM PROPOSTA

A ferramenta desenvolvida tem módulos. O primeiro faz a aquisição automática de dados. O segundo é responsável pela análise dos dados para identificar tópicos frequentes e tendências. Ambas as fases são descritas a seguir.

4.1 Data Extraction

O ReclameAqui é um sítio web baseado em *Javascript*, sendo que o *framework* arquitetural citado nos trabalhos relacionados não funciona apropriadamente. Por esta razão, adaptamos o *framework* considerando as particularidades do ReclameAqui. O novo esquema de funcionamento é apresentado na Figura 2.

No primeiro passo o usuário informa a página da empresa e os dados a serem extraídos. Então a ferramenta faz uma requisição da página ao ReclameAqui, que a envia em *Javascript*. Esta página é convertida para *HyperText Markup Language* (HTML) usando o navegador PhantomJS. O segundo passo (*crawling*) extrai os links de cada reclamação, sendo que todas as páginas são convertidas para HTML, a fim de permitir que a informação seja extraída na terceira etapa. A extração consiste na identificação de itens pré-definidos tais como o texto da reclamação, a localização do usuário e a pontuação associada ao atendimento, por exemplo. Sendo que cada um destes itens será armazenado como um atributo singular, possuindo um rótulo de identificação. Para concluir, os dados extraídos são organizados e armazenados em um arquivo no formato *Comma-Separated Values* (CSV). Este formato foi escolhido considerando a implementação do método de análise, o qual é descrito a seguir.

4.2 Modelagem de tópicos

Na ferramenta desenvolvida a modelagem de tópicos foi conduzida usando LDA. Esse modelo lida com cada documento como um vetor, o qual possui valores representando a frequência de palavras. Assim, o LDA assume cada tópico implícito como uma distribuição de probabilidade sobre algumas palavras de um *corpus* completo. O LDA foi aplicado nos dados sob o rótulo "reclamação" armazenados no arquivo CSV, e recebeu como parâmetro a busca por 10 tópicos dentre o corpo dos dados selecionados.

5 RESULTADOS EXPERIMENTAIS E DISCUSSÕES

A Tabela 2 apresenta alguns dos tópicos mais encontrados nas reclamações quanto às Empresas A e B, bem como os termos mais significativos dentro de cada tópico. Por meio da análise dos resultados é

Improving Social CRM through eletronic word-of-mouth: a case study of ReclameAqui

WebMedia'2017:Workshops e Pôsteres, WTIC, Gramado,Brazil

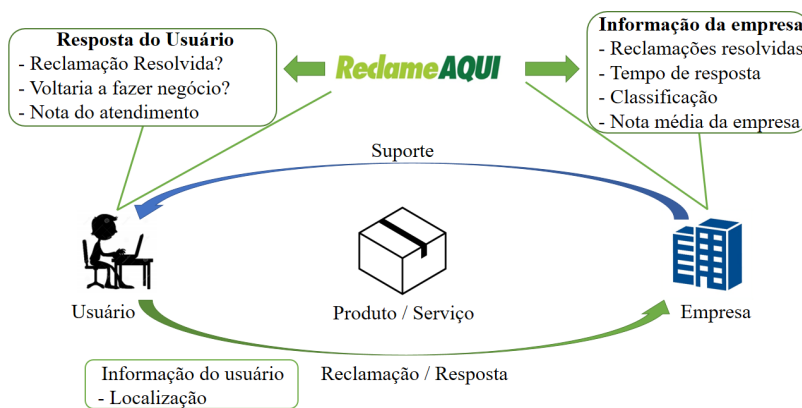


Figure 1: Esquema de interação das reclamações no Reclame Aqui.

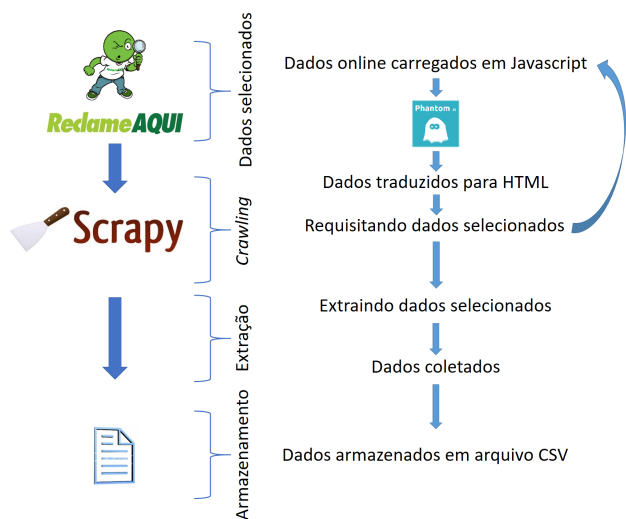


Figure 2: Esquema de funcionamento do módulo de aquisição de dados

possível observar que, para a Empresa A, os usuários da plataforma postam reclamações mais referentes a: a) setor de Atendimento; b) garantia de seus produtos; c) problemas com revisão de carros; d) e questões relacionadas a segurança dos veículos. Referente à Empresa B, as reclamações mais frequentes são quanto a: a) o setor de serviços prestados pela empresa; b) espera para conclusão do atendimento; c) o produto sabão em pó é um tópico frequente; d) o setor de atendimento da companhia também é mencionado.

Observa-se assim que a distribuição dos tópicos mais frequentes para a Companhia A é mais dispersa, dado que os conceitos encontrados não tem uma correlação forte quanto aos seus significados. Por outro lado, no caso da Empresa B, os tópicos de Serviço, Tempo de Espera e Atendimento, tem uma correlação entre si, caracterizando um setor mais amplo referente aos serviços providos pela

Table 2: Tópicos mais frequentes para os estudos de caso.

Empresa A				
Tópico	Termos mais significativos			
Atendimento	atendimento	informação	serviço	agendamento
Garantia	garantia	anos	mesmo	veículo
Revisão	revisão	problema	resolver	vez
Segurança	airbag	troca	segurança	passageiro
Empresa B				
Tópico	Termos mais significativos			
Serviço	pedido	contato	vendedor	entrega
Tempo de Espera	compra	consumidor	solução	aguardando
ProdutoSabão	roupa	mancha	MARCA	sabão
Atendimento	envio	reclamação	sac	email

empresa B. Como *insight* aos gestores, é possível indicar que a empresa precisa rever seus processos internos para melhorar o tempo de espera ou amenizar o desconforto que o mesmo causa aos seus consumidores.

Como análise final, é possível atestar que ambas as empresas podem realizar uma análise qualitativa mais profunda, referente a esses aspectos indicados pelos tópicos encontrados, dado que os mesmos servem como potenciais indicadores de quais produtos ou serviços devem ser melhorados. A Figura 3 contém uma visão geral da frequência dos tópicos mais relevantes encontrados, para ambas as empresas.

Analisando os resultados em relação aos dados coletados e análises feitas, é possível responder as Perguntas de Pesquisa (PP) levantadas. A resposta para a primeira PP, “É possível adotar o ReclameAqui como uma fonte de dados para estudos envolvendo eWoM?”, é sim, uma vez que a abordagem proposta contorna a necessidade de recuperação manual da informação, permitindo a aplicação de métodos inteligentes para extração de conhecimento. A possibilidade de aplicação destes dados em métodos de análise é mostrada na modelagem de tópicos realizada neste trabalho.

Avaliando a segunda PP “Quais são os benefícios da adoção de análises automáticas dos dados provenientes do ReclameAqui?” deve-se observar que um dos principais pilares do Social CRM é a integração de dados. Devido à relevância do ReclameAqui para o

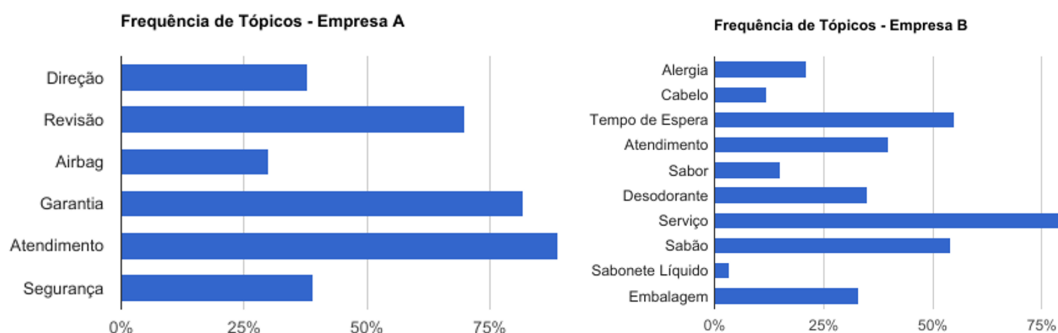


Figure 3: Gráfico com a frequência de tópicos mais reclamados para as Empresas A e B.

mercado brasileiro e, considerando que o comportamento dos consumidores é diferenciado dependendo do canal de comunicação, entende-se que quanto mais informações as empresas obtêm, melhor a confiabilidade dos processos decisórios, agregando valor aos negócios.

6 CONSIDERAÇÕES FINAIS

O relacionamento entre empresas e consumidores mudou drasticamente devido ao engajamento dos usuários Web com as redes sociais online. A integração entre a análise das redes sociais com o gerenciamento de relacionamento com os consumidores fez emergir o Social CRM, a qual traz consigo novos serviços e demandas do mercado. Neste âmbito, o boca-a-boca virtual ganha cada vez mais atenção dado seu impacto na forma com que os consumidores tomam suas decisões de compra. No Brasil, o ReclameAqui é um sistema de reputação bem-conceituado. Apesar de sua importância para o mercado brasileiro, a maior parte dos estudos e ferramentas voltadas para plataforma fazem a análise de dados manualmente e com um número reduzido de amostras.

Visando atender esta demanda mercadológica, descreveu-se neste artigo uma abordagem para se fazer automaticamente a extração e a análise de dados de reclamações feitas no ReclameAqui. O módulo de extração de dados foi desenvolvido usando o Web Crawler Scrapy. O módulo de análise consiste na análise exploratória e estatística dos dados e da modelagem de tópicos, por meio do Latent Dirichlet Allocation. Os resultados obtidos demonstraram que os dados do ReclameAqui podem ser utilizados para agregar valor aos negócios, auxiliando no aprimoramento dos atuais sistemas de Social CRM, por meio do incremento dos dados disponíveis para as análises. Como trabalho futuro pretende-se testar outros métodos de modelagem de tópicos e incorporar métodos para identificação de emoções, com um intuito de se construir um serviço de atendimento ao consumidor mais humano.

AGRADECIMENTOS

Os autores gostariam de agradecer a Universidade Federal do Oeste do Pará (UFOPA) e a Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação Tecnológica (Proppit) pelo suporte e concessão de uma bolsa de iniciação científica. E igualmente agradecer ao *Deutscher Akademischer Austauschdienst* (DAAD - *Germany*) pelo suporte nesta pesquisa.

REFERENCES

- [1] Rainer Alt and Olaf Reinhold. 2013. How Companies are Implementing Social Customer Relationship Management : Insights From Two Case Studies. *26th Bled eConference* (2013), 206–221.
- [2] A M Barreto. 2014. The word-of-mouth phenomenon in the social media era. *International Journal of Market Research* 56, 5 (2014), 631–654. citeulike-article-id: 13374933<http://www.mrs.org.uk/ijmr>
- [3] Wenjing Duan, Bin Gu, and Andrew B. Whinston. 2008. The dynamics of online word-of-mouth and product sales-An empirical investigation of the movie industry. *Journal of Retailing* 84, 2 (6 2008), 233–242. <https://doi.org/10.1016/j.jretai.2008.04.005>
- [4] Weiguo Fan and Michael D Gordon. 2014. The Power of Social Media Analytics. *Commun. ACM* 57, 6 (2014), 74–81. <https://doi.org/10.1145/2602574>
- [5] Stefan Feuerriegel, Antal Ratku, and Dirk Neumann. 2016. Analysis of How Underlying Topics in Financial News Affect Stock Prices Using Latent Dirichlet Allocation. In *49th Hawaii International Conference on System Sciences, HICSS 2016, Koloa, HI, USA, January 5-8, 2016*. 1072–1081. <https://doi.org/10.1109/HICSS.2016.137>
- [6] Ralf Krestel, Peter Fankhauser, and Wolfgang Nejdl. 2009. Latent Dirichlet Allocation for Tag Recommendation. In *Proceedings of the Third ACM Conference on Recommender Systems (RecSys '09)*. ACM, New York, NY, USA, 61–68. <https://doi.org/10.1145/1639714.1639726>
- [7] So-Hyun Lee, Seung-Eui Noh, and Hee-Woong Kim. 2013. A mixed methods approach to electronic word-of-mouth in the open-market context. *International Journal of Information Management* 33, 4 (2013), 687 – 696. <https://doi.org/10.1016/j.ijinfomgt.2013.03.002>
- [8] Y. S. Lee, R. Lo, C. Y. Chen, P. C. Lin, and J. C. Wang. 2015. News topics categorization using latent Dirichlet allocation and sparse representation classifier. In *2015 IEEE International Conference on Consumer Electronics - Taiwan*. 136–137. <https://doi.org/10.1109/ICCE-TW.2015.7216819>
- [9] Fábio Lobato, Márcia Pinheiro, Antonio Jacob, Olaf Reinhold, and Ádamo Santana. 2017. Social CRM: Biggest Challenges to Make it Work in the Real World. In *Business Information Systems Workshops: BIS 2016 International Workshops, Leipzig, Germany, July 6-8, 2016, Revised Papers*, Witold Abramowicz, Rainer Alt, and Bogdan Franczyk (Eds.). Springer International Publishing, Cham, 221–232. https://doi.org/10.1007/978-3-319-52464-1_20
- [10] M. R. Martínez-Torres, S. Toral, M. Olmedilla, and F. J. Arenas-Marquez. 2016. Features of Reputed Users in eWOM Using Evolutionary Computation. In *2016 9th International Conference on Developments in eSystems Engineering (DeSE)*. 223–227. <https://doi.org/10.1109/DeSE.2016.36>
- [11] M. Olmedilla, M.R. Martínez-Torres, and S.L. Toral. 2016. Harvesting Big Data in social science: A methodological approach for collecting online user-generated content. *Computer Standards & Interfaces* 46 (2016), 79–87. <https://doi.org/10.1016/j.csi.2016.02.003>
- [12] Shengsheng Qian, Tianzhu Zhang, Changsheng Xu, and M. Shamim Hossain. 2015. Social Event Classification via Boosted Multimodal Supervised Latent Dirichlet Allocation. *ACM Trans. Multimedia Comput. Commun. Appl.* 11, 2, Article 27 (Jan. 2015), 22 pages. <https://doi.org/10.1145/2659521>
- [13] Wendel Silva, Ádamo Santana, Fábio Lobato, and Márcia Pinheiro. 2017. A Methodology for Community Detection in Twitter. In *Proceedings of the International Conference on Web Intelligence (WI '17)*. ACM, New York, NY, USA, 1006–1009. <https://doi.org/10.1145/3106426.3117760>
- [14] Andrew T. Stephen and Donald R. Lehmann. 2016. How word-of-mouth transmission encouragement affects consumers' transmission decisions, receiver selection, and diffusion speed. *International Journal of Research in Marketing* 33, 4 (2016), 755–766. <https://doi.org/10.1016/j.ijresmar.2016.08.003>