

Recomendação de reviews personalizada para donos de estabelecimentos

Thiago R. P. Prado, Mirella M. Moro

Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brasil

{thiagorpp, mirella}@dcc.ufmg.br

Resumo. *Aplicativos online de avaliação geralmente recomendam as revisões (reviews) mais úteis para os usuários leitores de avaliações. Aqui, introduzimos um novo problema: avaliar a utilidade de uma revisão para o dono de um estabelecimento. Especificamente, propomos o uso de aspectos e sentimentos das revisões, e a geração de um ranking ordenado a partir das mais úteis para o gerenciamento e desenvolvimento do estabelecimento.*

Abstract. *Online review apps usually recommend the most useful reviews for consumers to read. Here, we introduce a new problem: to evaluate the usefulness of a review from the owner's perspective. Specifically, we propose using the review's aspects and sentiments, and generating a rank ordered by the most useful reviews from the establishment management and developing point of view.*

1. Introdução

Redes sociais baseadas em localização, como o Tripadvisor e o Foursquare, são hoje ferramentas importantes para os usuários escolherem hotéis, restaurantes e atrações turísticas. Esses locais são conhecidos como pontos de interesse (POI). Os conteúdos dessas redes são gerados pelos usuários, fornecendo assim acesso às opiniões de muitos indivíduos. Um usuário, ao contribuir com sua opinião, avalia um POI através de uma nota (*rating*) ou informando se gostou ou não e, opcionalmente, escrevendo uma revisão (*review*). Ao realizar uma pesquisa sobre uma região geográfica, o Tripadvisor e o Foursquare apresentam uma lista de POIs, classificadas com base nas avaliações realizadas.

Focando na perspectiva de *reviews* realizadas, por ser um serviço baseado em consumidores, há vários fatores que podem afetar o sentimento desses, por exemplo, nível de ruído, produtos de baixa qualidade, clima, estação do ano e até mesmo as suas expectativas. Nesse contexto, identificar e gerenciar esses fatores pode fornecer aos clientes e proprietários informações valiosas através da interpretação de grandes quantidades de dados [Maroun et al. 2016]. Para o proprietário do estabelecimento, o mais importante é ter uma forma rápida e confiável para identificar as *reviews* que contêm informações relevantes para a melhoria dos serviços prestados, enquanto que para o cliente, o ideal é identificar *reviews* que possuam informações com detalhes sobre o lugar que o ajudem a tomar uma decisão. Por exemplo, considerando a perspectiva do cliente, desenvolvemos a ferramenta *POIView* [Prado and Moro 2015] que utiliza técnicas de análise de sentimentos para visualizar mudanças nas opiniões sobre POI.

Assim como a POIView, os trabalhos relacionados são voltados para ajudar usuários (clientes) a identificarem *reviews* relevantes (ex. [Tang et al. 2013]). Especificamente, numa tentativa de ajudar usuários a identificar *reviews* úteis, muitos sites permitem que usuários votem na utilidade da revisão. Então, a maioria dos sites mostra a porcentagem de votos positivos, e poucos fornecem a nota que cada usuário deu para a revisão. Além disso, a avaliação da revisão tende a ser esparsa com muitas *reviews* sem receber nenhum *feedback* [O'Mahony and Smyth 2009]. Esse problema é devido ao efeito *rich-get-richer* onde as *reviews* do topo acumulam mais e mais *feedbacks*, e as *reviews* recentes raramente são lidas [Liu et al. 2007]. Mesmo que as notas dadas às *reviews* sejam muito esparsas para ajudar usuários a identificar *reviews* relevantes, elas podem prover dados importantes para prever automaticamente a qualidade de uma revisão [Deepak et al. 2014].

Neste trabalho, introduzimos um novo problema: identificar a utilidade de uma revisão para o *dono de um estabelecimento*. A relevância de uma *review* para o estabelecimento difere da relevância para o cliente pois agora o objetivo não é ajudar a decidir entre qual produto comprar ou qual lugar visitar. O foco é identificar comentários que abordem aspectos importantes, principalmente que tratem pontos negativos, relacionados ao estabelecimento, os quais podem ser melhorados de forma a aumentar a satisfação dos clientes e ajudar na tomada de decisões administrativas e estratégicas. Para abordá-lo, propomos o uso de aspectos e sentimentos das *reviews* [Jo and Oh 2011].

2. Trabalhos Relacionados

Determinar a qualidade de conteúdo gerado por usuários de redes sociais online é um problema difícil. Por exemplo, Bigonha et al. [2012] definiram métricas de qualidade para a análise de usuários influentes no *Twitter*. Pang e Lee [2005] estudaram a predição das avaliações de produtos, o qual pode ser relevante devido à correlação entre avaliação de um produto e a utilidade de uma *review*; porém, a avaliação geral de um produto já é conhecida. Ghose e Ipeirotis [2011] têm focado em determinar automaticamente a qualidade de *reviews* através do uso de atributos textuais ou aspectos sociais [McAuley and Leskovec 2013]. Atributos textuais incluem estatísticas do texto, como tamanho do texto da revisão, porcentagem de adjetivos, entre outros. Atributos sociais são relacionados ao autor da revisão e são extraídos de seu contexto social, como o número de avaliações já feitas pelo autor, grau de conexão do autor na rede social, entre outros.

A análise de sentimentos é muito utilizada para obter informações relevantes sobre *reviews*. Trabalhos anteriores que estudam aspectos e sentimentos de *reviews* podem ser classificados como mineração de opiniões. A mineração de opiniões opera sobre porções de texto, tais como páginas web, *tweets*, entre outros. Toda opinião é composta de pelo menos dois elementos: um alvo (tópico, pessoa, etc) e um sentimento (atitude, opinião) sobre esse alvo [Liu and Zhang 2012]. O processo de mineração de opiniões temporais envolve definir a opinião média sobre um determinado tema em dois ou mais momentos distintos no tempo. Alterações na opinião podem então ser identificadas e utilizadas para encontrar padrões ou sumarizar a opinião em relação a um aspecto específico [Sápiras and Becker 2014]. Lourenço et. al [2014] estudaram uma forma eficiente para analisar a opinião das pessoas sobre tópicos e entidades em redes sociais como o *Twitter*.

Em análise de opiniões e sentimentos, o foco é diferenciar entre declarações de fatos e opiniões e detectar a polaridade do sentimento expressado. Muitos pesquisadores

têm trabalhado nessa área com focos diferentes. Turney [2012] classificou a polaridade do sentimento de *reviews* em nível de documento. Wiebe et al. [1999] classificaram a subjetividade em nível de sentenças usando classes como adjetivos, verbos e atributos. Riloff e Wiebe [2003] extraíram expressões subjetivas das sentenças usando o processo de aprendizado *bootstrapping pattern*. Essas técnicas foram aplicadas e examinadas em diferentes domínios, como em *reviews* de usuários e artigos de notícias. Na solução proposta, a análise de sentimentos é utilizada para identificar opiniões e a polaridade do sentimento expressado nas *reviews*.

3. Solução Proposta

O problema de automaticamente identificar a relevância de *reviews* sobre POIs para o dono de um estabelecimento é muito importante para que se possa entender melhor a dinâmica de suas ações de negócio. Embora existam vários trabalhos que tentam prever a utilidade de uma revisão para um usuário (cliente) e outros que abordam os aspectos e sentimentos atrelados a uma revisão [Maroun et al. 2016], até onde sabemos, nenhum abordou o problema de criar um ranking com as avaliações mais relevantes para o dono em vez do cliente. A relevância de uma *review* para o estabelecimento difere da relevância para o usuário (cliente) pois agora o objetivo não é ajudar a decidir entre qual produto comprar ou qual lugar visitar. O foco é identificar comentários que abordem aspectos importantes, principalmente que tratem pontos negativos, relacionados ao estabelecimento, os quais podem ser melhorados de forma a aumentar a satisfação dos clientes e ajudar na tomada de decisões administrativas e estratégicas. Neste trabalho, propomos a criação de um ranking de avaliações para o dono de um estabelecimento de forma que elas sejam ordenadas de acordo com sua relevância para a tomada de decisões do local. Esse ranking é baseado nos aspectos descritos nas *reviews* de um POI e o sentimento relacionado a ele.

3.1. Coleta de Dados

Foram coletados dados de 834 hotéis e suas *reviews* das principais cidades dos Estados Unidos a partir da rede social *Tripadvisor*. Os usuários do *Tripadvisor* fazem *reviews* de um estabelecimento e podem dar notas para aspectos específicos do local (quarto, café da manhã, entre outros). Além disso, essas *reviews* podem ser respondidas por pessoas do estabelecimento de forma a dar um *feedback* para os usuários e tentar resolver possíveis insatisfações com o serviço prestado. A categoria *hotel* foi escolhida a partir de uma análise empírica dos dados do *Tripadvisor* onde constatou-se que é o seguimento que mais fornece *feedback* a seus clientes através de repostas de *reviews*.

3.2. Validação e Baseline

O gabarito utilizado para verificar a qualidade do ranking gerado foi criado com base nas respostas dadas pelos estabelecimentos a uma revisão de um usuário. Especificamente, quando o usuário faz uma *review* negativa apontando um problema sobre o estabelecimento, é comum (embora não obrigatório) que o dono do estabelecimento escreva uma resposta justificando tal problema ou apontando que a solução está a caminho. Através desse *feedback* pode-se determinar a relevância que uma revisão teve com base na similaridade de cosenos entre os aspectos mencionados na revisão e na resposta dada pelo estabelecimento. Similaridade alta significa que o estabelecimento se preocupou em responder os pontos abordados na revisão, indicando então que essa é uma revisão relevante para o estabelecimento. Por outro lado, similaridade baixa pode indicar respostas

Revisão	Sentimento	Tópico	Score
<i>"The air conditioning unit was very noisy and there was also a lot of noise from the lift at night, not only noise but vibrations in the room. Saying that the room although small was clean as was the rest of the hotel. Staff were friendly"</i>	clean	room	0.62
	friendly	staff	0.79
	very noisy	air conditioning unit	-0.778

Figura 1. Exemplo de uma revisão real com seus aspectos (score geral: 0.21)

genéricas dos estabelecimentos. É importante notar que é de interesse de um estabelecimento responder as reclamações de um cliente o mais detalhadamente possível, de forma a não perder o cliente e a sua reputação. Sendo assim, com base na similaridade entre revisão e resposta, é gerado um *ranking* de revisões para cada hotel, ordenado de forma decrescente pela similaridade, para ser utilizado como gabarito. O baseline utilizado são as revisões de um estabelecimento ordenado pelas mais recentes, por ser uma das formas mais comuns utilizadas por sites que permitem fazer revisões de locais.

3.3. Análise de Sentimentos e Identificação de Aspectos

A etapa de identificação de aspectos e análise de sentimentos é uma etapa complexa de processamento dos dados. Para evitar que seja necessário implementar todo um arcabouço como o descrito em [Jo and Oh 2011], foi utilizada a API de Análise de Sentimentos da *HP IDOL OnDemand*. Através dessa é possível processar revisões e obter uma lista de aspectos mencionados na revisão juntamente com o sentimento e o score associado a eles. A API de Análise de Sentimentos analisa um texto e retorna o sentimento como positivo, negativo ou neutro. A Análise de Sentimentos pode ser usada para obter informações valiosas sobre o que usuários, clientes, amigos, colegas estão dizendo. Automaticamente classificar o texto por sentimento permite que encontrar opiniões gerais de pessoas em sua área de interesse. Por exemplo, analisar os comentários de um produto pode ajudar a melhorar a experiência do cliente, ou para encontrar o produto mais ou menos popular. A Figura 1 mostra o texto de uma revisão real e seus aspectos.

3.4. Recomendação

O processo de recomendação consiste em definir um novo *ranking* para as revisões de um estabelecimento, de forma que as revisões mais relevantes para o dono do estabelecimento fiquem no topo. Essa nova ordem das revisões visa facilitar a análise das principais revisões feitas por clientes e auxiliar na tomada de decisões do estabelecimento a partir do *feedback* fornecido por seus clientes. Para definir esse novo *ranking* é necessário analisar cada revisão e descobrir sua relevância para o estabelecimento. Acredita-se que uma revisão que aborde aspectos negativos do estabelecimento seja mais informativa no quesito *feedback* para o estabelecimento do que revisões genéricas ou que abordem apenas aspectos positivos da experiência que o cliente teve com o estabelecimento. A recomendação proposta se baseia em: (I) Identificação dos aspectos de uma revisão e o sentimento sobre ele, conforme descrito na Seção 3.3; (II) Métrica para definição da relevância de uma revisão; (III) Ordenação decrescente das revisões com base na sua relevância.

A relevância de uma revisão é feita com base nos aspectos e sentimentos obtidos na etapa anterior. Partindo-se da ideia de que uma revisão é mais relevante para um estabelecimento se ela aborda aspectos negativos sobre o local, propomos o seguinte processo

Tabela 1. Principais resultados

Revisões resp	nDCG médio - basel./recom.	nDCG máx - basel./recom.	Ganho
43,79%	0,199 / 0,256	0,336 / 0,774	73,4 %

para definir um *score* para uma revisão: (I) Dicionário com todos aspectos negativos, obtidos a partir das revisões de um local, com o soma dos k scores recebidos por aspecto: $aspects[i] = \sum_{j=1}^K score_i[j]$; (II) A relevância rel_{ru} , de cada revisão r do estabelecimento feita pelo usuário u obtida através de $rel_{ru} = \sum_{i=1}^R aspects[i] * score_u[i]$.

4. Resultados

A métrica de avaliação utilizada é a *Normalized discounted cumulative gain* (nDCG). Essa métrica avalia o desempenho de um sistema de recomendação baseado na relevância ponderada de cada entidade recomendada. O valor obtido varia entre 0.0 a 1.0, com 1.0 representando o *ranking* ideal das entidades. Essa métrica se baseia em duas regras: (I) Documentos extremamente relevantes são mais importantes que documentos com relevância marginal; (II) Quanto mais baixa a posição do documento no *ranking*, menor o valor deste documento para o usuário. O nDCG pode ser calculado a partir das equações: (I) $DCG_k = \sum_{i=1}^K 2^{rel_i} - 1 \log(i + 1)$ e (II) $nDCG_k = \frac{DCG_k}{IDCG_k}$, onde k é a quantidade de documentos que podem ser recomendados, rel_i é a relevância do documento i , $IDCG_k$ é o valor ideal do DCG para um conjunto de documentos e relevâncias.

A solução foi avaliada comparando o *nDCG* obtido para o baseline e o valor obtido para o *ranking* criado com o método de recomendação proposto. Nos dois casos o gabarito descrito na Seção 3.2 é utilizado como ordem ideal das revisões. Um *ranking* é considerado bom desde que ele tenha em seu topo avaliações que tiverem respostas que abordam os aspectos mencionados. A Tabela 1 apresenta os principais resultados.

O gabarito utilizado é diferente para cada caso de teste, uma vez que ele é construído com base nas respostas dos estabelecimentos. A técnica proposta consegue bons resultados para alguns casos de teste e resultados medianos para outros. Investigando-se o motivo dessa discrepância, descobriu-se que os casos com melhores resultados foram os hotéis que visam dar um *feedback* real para os usuários, comprovando nossa hipótese de que uma revisão com aspectos negativos é mais relevante do que apenas aspectos positivos. Os resultados não tão satisfatórios foram casos em que o estabelecimento dá respostas genéricas ou vagas sem focar nas reclamações dos usuários. Apesar dessa dependência das respostas dos estabelecimentos para construir o gabarito, o método proposto obteve resultado melhor do que o baseline em aproximadamente 3/4 dos casos de teste.

5. Conclusões

Foi proposto um novo método para criação de um *ranking* das revisões de um estabelecimento, de acordo com sua relevância, focado no dono do local. O *ranking* é construído a partir da análise das revisões e obtenção dos aspectos abordados e o sentimento do cliente relacionado a cada aspecto. Um dos grandes desafios neste trabalho era definir uma forma de avaliar a solução proposta. O uso de dados provenientes do *Tripadvisor* resolveu esse problema já que ele permite que os estabelecimentos respondam as revisões recebidas de

forma a fornecer um *feedback* aos clientes. Sendo assim, com base nas respostas dos estabelecimento foi possível determinar quais revisões são mais importantes para definirmos um *ranking* e utilizar como gabarito. O baseline utilizado é a forma mais comum de se ordenar revisões que é pelas mais recentes primeiro.

Apesar da simplicidade da solução proposta obteve-se um resultado satisfatório ganhando do baseline em aproximadamente 3/4 dos casos de teste. Se considerarmos apenas estabelecimentos que visam responder os pontos abordados pelos clientes o método proposto obtém bons resultados e comprova nossa hipótese de que uma revisão que contém aspectos negativos é mais relevante do que uma que possui apenas aspectos positivos. Acredita-se que esse método pode ser melhorado para ser aplicado em cenários reais e substituir métodos de *ranking* simplistas utilizados pela maioria dos sistemas.

Agradecimentos. Trabalho parcialmente financiado por CNPq.

Referências

- Bigonha, C., Cardoso, T. N. C., Moro, M. M., Gonçalves, M. A., and Almeida, V. A. F. (2012). Sentiment-based influence detection on twitter. *JBCS*, 18(3):169–183.
- Deepak, A. et al. (2014). The promises and perils of mining github. In *MSR*.
- Ghose, A. and Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE*, pages 1498–1512.
- Janyce, W. et al. (1999). Development and use of a gold-standard data set for subjectivity classifications. In *ACL*, pages 246–253.
- Jo, Y. and Oh, A. H. (2011). Aspect and sentiment unification model for online review analysis. In *WSDM*, pages 815–824.
- Liu, B. and Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. In *Mining Text Data*, pages 415–463.
- Liu, J. et al. (2007). Low-quality product review detection in opinion summarization. In *EMNLP-CoNLL*.
- Lourenço Jr., R. et al. (2014). Economically-efficient sentiment stream analysis. In *SIGIR*, pages 637–646.
- Maroun, L., Moro, M. M., Almeida, J., and da Silva, A. P. C. (2016). Assessing review recommendation techniques under a ranking perspective. In *ACM Hypertext*.
- McAuley, J. and Leskovec, J. (2013). Hidden factors and hidden topics: Understanding rating dimensions with review text. In *RecSys*, pages 165–172.
- O’Mahony, M. P. and Smyth, B. (2009). Learning to recommend helpful hotel reviews. In *RecSys*.
- Pang, B. and Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *ACL*, pages 115–124.
- Prado, T. R. P. and Moro, M. M. (2015). POIView: Análise de Tendências a partir de Revisões Online. In *SBB D Demo Session*, pages 167–172.
- Riloff, E. and Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *EMNL*, pages 105–112.
- Sápiras, L. A. and Becker, K. (2014). Mineração da opinião sobre aspectos de candidatos a eleições em comentários de notícias. In *SBB D*, pages 117–126.
- Tang, J. et al. (2013). Context-aware review helpfulness rating prediction. In *RecSys*, pages 1–8.
- Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *ACL*, pages 417 – 424.