

# Um Método Baseado em Grafos para Predição da Utilidade de Opiniões sobre Produtos

Rogério F. de Sousa<sup>1</sup>, Rafael T. Anchieta<sup>1</sup>, Maria das Graças V. Nunes<sup>1</sup>

<sup>1</sup> Instituto de Computação e Matemática Computacional  
Núcleo Interinstitucional de Linguística Computacional (NILC)  
São Carlos – SP – Brasil

rogerfig@usp.br, rta@usp.br, gracacn@icmc.usp.br

**Abstract.** *This paper presents a new approach to predict the helpfulness of opinions. Usually, researchers in this area use tables of attribute-value to aggregate the features that represent the evaluated texts. In this paper, this task is modeled as a network, considering the information of relations among objects in the network (comments, stars, and words). A regularization technique of graphs is used to extract the relevant features of graph structure and, after that, the comments are classified as helpful or unhelpful. We compared our network model with a baseline method based on fuzzy logic. Our model outperformed the fuzzy logic method in 0.13 of F1 measure.*

**Resumo.** *Este trabalho apresenta uma nova abordagem para predição da utilidade de opiniões. Usualmente, os trabalhos nessa área usam tabelas do tipo atributo-valor para agregar as características que representam os textos avaliados. Neste trabalho, essa tarefa é modelada em forma de rede, considerando, dessa forma, as informações de relacionamento entre os objetos da rede (comentários, estrelas e palavras). Uma técnica de regularização de grafos é utilizada para extrair características relevantes da estrutura do grafo, e, após isso, os comentários são classificados em duas classes: Útil ou Não Útil. Comparou-se a modelagem com um método baseado em lógica fuzzy e a modelagem apresentou resultados promissores, superando-o em 0,13 pontos na medida F1.*

## 1. Introdução

Uma prática comum entre as pessoas é buscar informações sobre produtos e/ou serviços pelos quais elas estão interessadas. A principal forma de aquisição dessas informações é por meio das opiniões de outras pessoas que possuem algum tipo de experiência com o item desejado. Segundo [Liu 2012], as opiniões são elementos centrais para a maioria das atividades humanas e são capazes de influenciar o comportamento humano.

Os conteúdos gerados por usuários fazem parte das principais fontes de conteúdo na Web. Os comentários sobre produtos e serviços formam uma grande parcela desses conteúdos. Porém, parte desse conteúdo pode ser considerado indesejado. [Kim et al. 2006] menciona que o conteúdo indesejado contempla textos mal-escritos, opiniões vagas, textos com conteúdo duvidoso, etc. Ou seja, os conteúdos gerados por usuários variam muito em qualidade e tais textos não ajudam na tomada de decisão dos leitores.

Um outro fator que dificulta na tomada de decisão é a imensa quantidade de comentários disponíveis na Web, tornando ainda mais complicado encontrar conteúdo relevante. Além disso, é impossível que os usuários leiam todo o conteúdo de boa qualidade disponível e, adicionalmente a esse fato, as informações existentes nos comentários podem não ser úteis para a tomada de decisão das pessoas.

Na área de Processamento de Língua Natural (PLN), a tarefa que se ocupa em definir formas e descobrir métodos para caracterizar conteúdo de boa qualidade e de utilidade para as pessoas é chamada de Predição de Utilidade de Opiniões. Seu objetivo é, portanto, identificar conteúdo relevante (útil) em comentários de usuários. Além disso, essa tarefa é importante para outras tarefas do PLN como sumarização de opiniões [Anchiêta et al. 2017].

A preocupação dos *sites* de *e-commerce* em apresentar conteúdo útil é grande, e por isso alguns deles pedem um *feedback* explícito ao usuário: esse comentário é útil ou não? Assim, os comentários apresentados são ordenados de acordo com os votos que receberam, os mais úteis primeiro. Um dos problemas decorrentes desse fato é que novos comentários úteis dificilmente estarão no topo do *ranking*. É necessário algum tempo para que várias pessoas votem e o comentário ganhe a devida visibilidade [Kim et al. 2006, Liu et al. 2007, Singh et al. 2017].

Para evitar esse tipo de problema, é necessário aprender características existentes em *rankings* de comentários já consolidados e, assim, avaliar automaticamente a utilidade de comentários gerados por usuários. A grande maioria dos trabalhos conhecidos realiza essa tarefa utilizando a representação atributo-valor. Mas, apesar de ser amplamente utilizada, ela não é capaz de capturar informações de relacionamento entre os objetos. As estruturas de dados que melhor representam relacionamentos entre objetos são as redes.

Dessa forma, neste trabalho, modelou-se a tarefa de predição de utilidade como uma rede heterogênea. Para avaliar essa abordagem, usou-se um *corpus* de 2.000 comentários extraídos da *Google Play* e, por fim, comparou-se com um *baseline* bem conhecido baseado em lógica *fuzzy* [de Sousa et al. 2015]. A abordagem baseada em rede superou o *baseline* em 0,13 pontos na medida F1, mostrando que ela é viável para predizer se um comentário é útil ou não.

É importante destacar que, até onde se sabe, esse é o primeiro trabalho que modela a tarefa de predição de utilidade como uma rede heterogênea. Além disso, as abordagens são aplicadas e avaliadas em comentários escritos na língua portuguesa a fim de fomentar pesquisas dessa área para o Português.

O restante do trabalho está organizado da seguinte forma. Na Seção 2, é descrito rapidamente os principais trabalhos relacionados. Na Seção 3, é apresentado o *corpus* e a modelagem desenvolvida, bem como as etapas para predizer a utilidade do comentário. Na Seção 4, é detalhado os experimentos realizados. Finalmente, Seção 5 conclui o trabalho.

## 2. Trabalhos Relacionados

Na literatura, existem diversos trabalhos que abordam a tarefa de predição de utilidade e, normalmente, essa tarefa é tratada como regressão, classificação ou ranqueamento modelados como uma tabela de atributo-valor [Diaz and Ng 2018]. Destacam-se a seguir

alguns trabalhos que realizam classificação para predição da utilidade de comentários.

[Kim et al. 2006] propuseram um método baseado em regressão. Como entrada para o método, os autores definiram um conjunto de características a fim de verificar quão bem cada característica captura a utilidade de um comentário. Por fim, eles avaliaram o método sobre um *corpus* da *Amazon.com*.

[Malik and Hussain 2017] trataram o problema como uma tarefa de classificação. Eles definiram um método para calcular o *score* de emoção de comentários considerando alguns sentimentos específicos como, confiança, surpresa, raiva, tristeza, etc. Eles utilizaram esses *scores* como uma *feature* além de outras mais gerais como, por exemplo, o ranking do produto na Amazon, o preço do produto, quantidade de verbos, substantivos, adjetivos e advérbios, dentre outras. Eles modelaram uma Rede Neural Profunda e avaliaram o método também sobre um subcorpus da *Amazon.com*. Os autores reportaram resultados em média de 89% de F1 usando emoções positivas e 87% de F1 usando emoções negativas.

Além dos trabalhos descritos anteriormente, vale mencionar os trabalhos realizados na área de utilidade com foco na língua portuguesa. Serão descritos quatro trabalhos: [Martins and Tacla 2015, de Sousa et al. 2015, Santos et al. 2016, Barbosa and Moura 2016].

O trabalho de [Martins and Tacla 2015] apresenta uma metodologia de avaliação de utilidade de opiniões com foco na identificação de *features* que exerçam maior influência sobre os votos de utilidade. Os experimentos são aplicados em opiniões do domínio de serviços (hotéis). Os autores propõem diversas *features* que sejam capazes de caracterizar os comentários, elas são divididas em duas categorias: textuais e semânticas. As *features* textuais são compostas principalmente por métricas de inteligibilidade. Para sua extração, eles usam uma versão adaptada por eles do *Coh-Matrix-Port* [Scarton and Aluísio 2010]. Além do índice de inteligibilidade, são usadas: contagem de sentenças, palavras, sílabas e entre outras. Para as *features* semânticas é utilizado o *LSA*. As contribuições incluem a confirmação do impacto positivo das *features* semânticas na avaliação de utilidade das opiniões também na língua portuguesa. O índice de inteligibilidade revelou que opiniões mais longas e mais complexas são mais úteis do que as opiniões mais curtas e mais fáceis de ler.

[de Sousa et al. 2015] apresentaram uma abordagem diferente para classificar a importância de comentários para a língua portuguesa. Os autores propuseram um Sistema de Inferência *Fuzzy* para classificar comentários do domínio de produtos em 4 classes: Insuficiente, Suficiente, Bom e Excelente, por meio de 3 *features*: reputação do autor, quantidade de pares do tipo (*característica, palavra opinativa*) e riqueza do vocabulário.

Os comentários são ranqueados de acordo com o valor expresso pelo sistema. Após a ordenação da lista, vários pontos de cortes foram definidos sucessivamente e em cada ponto de corte foi aplicado um método *baseline* para definição da polaridade dos comentários. Os autores compararam os resultados dos subconjuntos com a aplicação do mesmo método no conjunto completo. Os resultados mostraram que um ponto de corte considerando apenas 10% dos comentários obteve melhor resultado que analisar o conjunto completo. Os autores apresentaram 10% de aumento de *f-measure* para os comentários positivos e cerca de 20% de *f-measure* para os comentários positivos.

Seguindo a mesma linha, o trabalho de [Santos et al. 2016] estendeu o trabalho de [de Sousa et al. 2015]. Eles melhoraram as definições de algumas características e propuseram um estudo experimental para comparar a abordagem anterior de sistemas *fuzzy* com o uso de Redes Neurais Artificiais. Foram propostas duas topologias de redes neurais artificiais. Após os experimentos os autores reportam que não conseguiram melhorar os resultados anteriores, mas argumentam que o resultado se deve a alguns fatores: as amostras obtiveram resultados esperados dispersos, o que tornou difícil a generalização da rede; e nenhuma das topologias candidatas atingiu a acurácia mínima. A acurácia mínima serve como limite mínimo para considerar a rede treinada. Eles apresentaram 52,48% de *f-measure* para os comentários positivos e 62,53% de *f-measure* para os negativos.

Por fim, [Barbosa and Moura 2016] avaliaram a utilidade de opiniões no domínio de jogos. Os autores coletaram comentários da *Steam*<sup>1</sup> e usaram as *features* de autoria das opiniões, características textuais e metadados existentes no site como entrada para uma rede neural artificial do tipo MLP (*Multi-layer Perceptron*) para inferir a utilidade dos comentários. Após os experimentos, eles reportaram bons resultados e mostraram que as métricas relativas à autoria foram mais relevantes juntamente com o tamanho do texto. Já a data de postagem dos comentários não apresentou forte impacto na avaliação.

Todos os trabalhos mencionados abordam a modelagem da tarefa com o uso de uma tabela atributo-valor. Mas, acredita-se que modelar esse problema com o uso de redes pode gerar bons resultados, principalmente em virtude do acréscimo da informação de relacionamento entre objetos, inerente a elas. Portanto, neste trabalho, propôs-se uma modelagem baseada em rede heterogênea visando mostrar que é possível aproveitar a capacidade das redes para criar alternativas para esse problema.

Uma contribuição adicional é a produção de recursos e resultados para futuras pesquisas em língua portuguesa. No que segue, serão detalhados o *corpus* e a modelagem proposta para a tarefa de predição de utilidade.

### 3. Predição de Utilidade

#### 3.1. Córpus

Para avaliar nossa modelagem, extraiu-se 2.000 comentários de usuários de 10 aplicativos da *Google Play* da categoria de comunicação (ver Tabela 1). Escolheu-se uma única categoria a fim de facilitar a tarefa de predição, pois acredita-se que os comentários possuam termos e/ou tópicos semelhantes [Anchiêta and Moura 2017]. Além disso, essa abordagem é bastante utilizada na literatura [Diaz and Ng 2018]. Na Tabela 2, são apresentadas algumas estatísticas do *córpus* de trabalho.

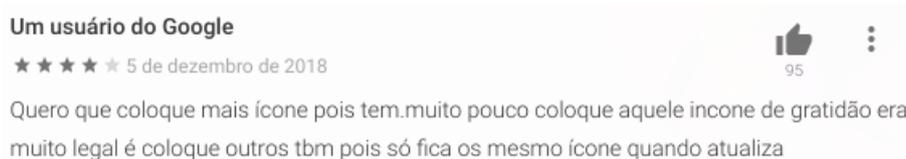
A extração do *corpus* foi necessária pois, para o Português, existe apenas um *corpus* que contém informações de utilidade [Hartmann et al. 2014], porém, o domínio é diferente do proposto no momento. Logo, existe a necessidade de geração de novos *corpus* e como mencionado anteriormente uma das contribuições deste trabalho é disponibilizar os recursos produzidos para trabalhos futuros.

Um comentário na *Google Play* contém o próprio texto do comentário, o autor do comentário, o número de estrelas, a data do comentário e o número de *likes* que o

---

<sup>1</sup><http://store.steampowered.com>

comentário recebeu (ver Figura 1). O número de estrelas e o número de *likes* estão no intervalo de  $[0, 5]$  e  $[0, +\infty)$ , respectivamente. A base de comentários utilizada neste trabalho é apenas um subconjunto de uma base maior. O código fonte a base de dados serão disponibilizados mediante solicitação por e-mail.



**Figura 1. Exemplo de comentário da Google Play**

**Tabela 1. Número de aplicativos e comentários extraídos**

Aplicativo	# Comentários	Aplicativo	# Comentários
Facebook	200	Skype	200
Google Allo	200	Snapchat	200
Hangouts	200	Telegram	200
Mensagens	200	Viber	200
Messenger	200	WhatsApp	200
<b>Total</b>		<b>2.000</b>	

**Tabela 2. Estatísticas da Base**

Classe	# Tokens	Types
Úteis	20588	3712
Não Úteis	13161	2709

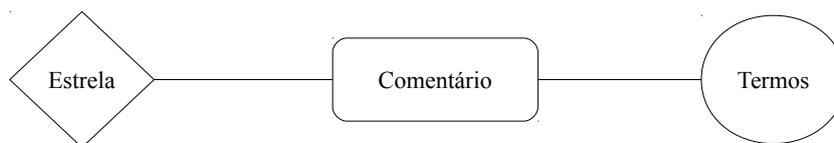
### 3.2. Modelagem proposta

A tarefa de predição de utilidade tem sido vista como um problema de regressão, classificação ou ranqueamento, modelado como uma tabela de atributo-valor. Entretanto, neste trabalho, modelou-se essa tarefa como uma rede heterogênea. Na Figura 2, é mostrado o esquema da rede modelada onde Estrela, Comentário, e Termos são os nós da rede.

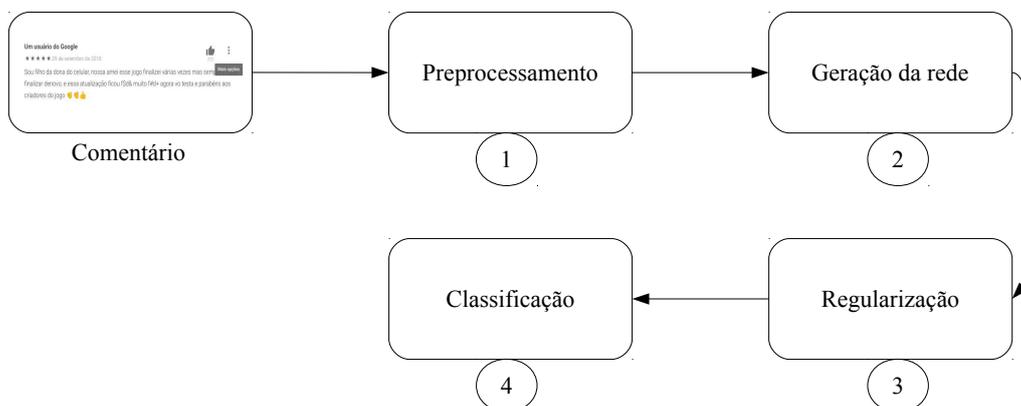
A fim de instanciar a rede heterogênea, desenvolveu-se uma metodologia organizada em 4 etapas a partir dos comentários extraídos. Na Figura 3, é apresentada uma visão geral da metodologia.

Na primeira etapa, realizou-se um pré-processamento nos comentários (subseção 3.3). Em seguida, gerou-se a rede heterogênea (subseção 3.4). Após a geração da rede, aplicou-se um algoritmo de regularização sobre a rede (subseção 3.5). Finalmente, classificaram-se os comentários em úteis ou não úteis (subseção 3.6).

É importante deixar claro que a abordagem como um todo é independente da língua. A única etapa que faz uso de componentes linguísticos é a de pré-processamento, e, nela é possível realizar a substituição das ferramentas utilizadas para que todo o processo seja aplicado a diversas outras línguas. Decidiu-se realizar o trabalho inicialmente para



**Figura 2. Esquema da rede heterogênea para predição de utilidade**



**Figura 3. Metodologia para instanciar a rede heterogênea**

esta língua, pois, o Português carece de técnicas e trabalhos para a tarefa de modelagem e predição da utilidade de comentários.

### 3.3. Preprocessamento

Nessa etapa, preprocessaram-se os comentários dos usuários com o objetivo de normalizá-los, de obter os termos mais relevantes do comentário e extrair o número de estrelas que cada comentário obteve.

Os comentários da *Google Play* são similares a *tweets*, normalmente curtos não obedecendo às regras gramaticais e de pontuação. Dessa forma, aplicou-se um normalizador [Bertaglia and Nunes 2016] desenvolvido para a língua portuguesa sobre cada comentário do *corpus* a fim de corrigir as palavras mal escritas. Em seguida, usou-se o *Part-Of-Speech Tagger* da NLPnet [Fonseca and Rosa 2013] nos comentários normalizados, com o objetivo de extrair palavras de classe aberta, como substantivo, verbo, adjetivo e advérbio. Por fim, aplicou-se um algoritmo de *stemming* [Orengo and Huyck 2001] a fim de obter os *stems* de cada palavra dos comentários.

### 3.4. Geração da rede

O processo de geração da rede é bem simples. Após o preprocessamento, um nó é criado para cada valor de Estrela, Comentário e Termo existente no comentário. Os nós do tipo Comentário e Estrela são conectados por meio de uma aresta de acordo com a quantidade de estrelas que o autor do comentário forneceu ao publicar o comentário, o mesmo acontece entre Comentário e Termo sempre que o texto de um comentário mencionar aquele termo. Nós do tipo Estrela não podem ser conectados entre si e nem com os nós do tipo Termo. Da mesma forma, Termos não podem ser conectados entre si. A rede não é direcionada e as arestas não contêm pesos. Na Figura 4, é ilustrada uma pequena instância da rede proposta.

A modelagem da rede foi realizada dessa forma pois, acredita-se que exista relação entre a utilidade e a quantidade de estrelas fornecidas pelo autor do comentário. Da mesma forma espera-se que exista relação entre os termos mencionados nos comentários e as suas utilidades.

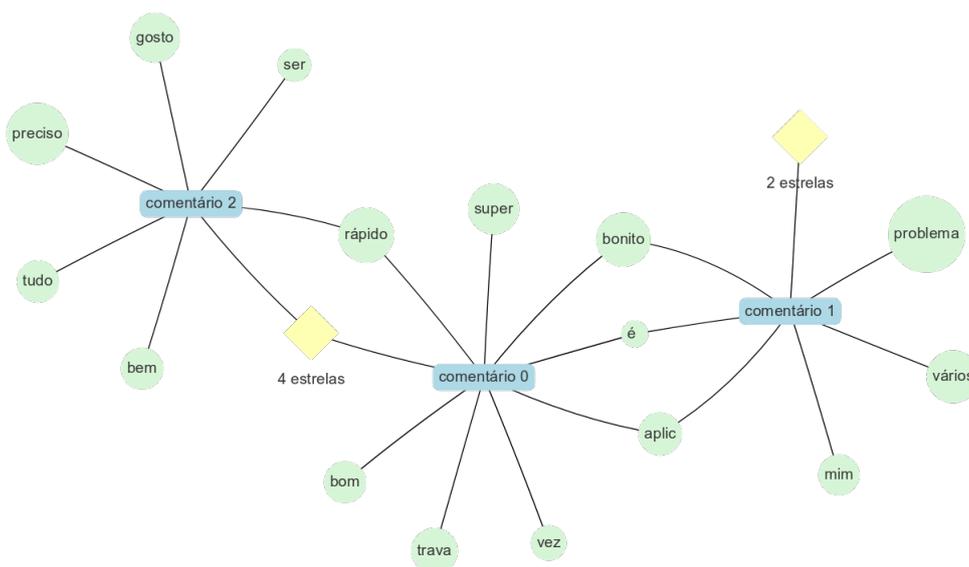


Figura 4. Exemplo de Instância da Rede

### 3.5. Regularização

A geração de redes normalmente considera informações de similaridade entre objetos da rede, por exemplo, usando uma representação do tipo *Bag-of-Words (BoW)* e calculando a similaridade de documentos por meio da métrica de similaridade de cosseno. Mas, nem sempre é possível criar uma rede dessa forma.

Por outro lado, para outros problemas, pode ser difícil ou pouco natural sua modelagem em redes, ou ainda, pode não ser simples considerar e agregar informações de domínio na medida de similaridade entre os nós. Nesses casos os algoritmos mais simples de classificação em redes não alcançam bons resultados.

A tarefa alvo deste trabalho é um exemplo dessa dificuldade. Em experimentos anteriores, a similaridade entre vetores de documentos não atingiu bons resultados. Portanto, optou-se por criar uma rede heterogênea considerando as informações de relacionamento entre os elementos existentes nos comentários. O uso de nós de diferentes tipos aumenta a informatividade da rede, porém, dificulta a aplicação de alguns métodos mais simples de aprendizado diretamente na rede.

Nesse contexto, optou-se por aplicar um método de regularização para realizar a extração de características sobre as classes dos objetos da rede. A regularização é um tipo de algoritmo de classificação transdutiva [Rossi 2016]. O objetivo da regularização é encontrar um conjunto de rótulos que respeite as seguintes condições: (i) ser consistente com o conjunto de rótulos anotados manualmente e; (ii) ser consistente com a topologia da rede, ou seja, considera que os vizinhos mais próximos tendem a ter os mesmos rótulos.

O algoritmo de regularização utilizado neste trabalho é o *Gaussian Fields and Harmonic Functions (GFHF)* [Zhu et al. 2003], que busca minimizar a função dada na Equação 1. As informações de classe de cada objeto são dadas pela média ponderada das informações de classe dos seus vizinhos pelos pesos das conexões entre os objetos, como mostrado na Equação 2, sendo que os termos são os seguintes:

- $O$  é o conjunto de nós da rede.
- $O^L$  é o conjunto de nós pré-rotulados da rede.
- $F \in \mathbb{R}^O$  é a saída do regularizador. Representa um vetor (ver Figura 3) que representa as coordenadas referentes ao comentário no plano.
- $w$  é o peso referente à aresta entre os nós  $o_i$  e  $o_j$ .
- $y$  é o vetor de informações referente aos nós pré-rotulados.

$$Q(F) = \frac{1}{2} \sum_{o_i, o_j \in O} w_{o_i, o_j} (f_{o_i} - f_{o_j})^2 + \lim_{\mu \rightarrow \infty} \mu \sum_{o_i \in O^L} (f_{o_i} - y_{o_i})^2 \quad (1)$$

$$f_{o_i} = \frac{\sum_{o_j \in O} w_{o_i, o_j} f_{o_j}}{\sum_{o_j \in O} w_{o_i, o_j}} \quad (2)$$

Para execução da regularização alguns nós devem ser rotulados manualmente com suas classes específicas. O *GFHF* não prevê a modificação das classes pré-rotuladas para a aplicação do algoritmo (o processo de rotulação é apresentado na Seção 4). Dessa forma, a Equação 2 somente é aplicada aos nós não rotulados. A execução do algoritmo gera valores referentes às coordenadas para cada objeto da rede (Ver Tabela 3). Esses valores podem ser usados por diversos algoritmos de classificação para prever os resultados. Na subseção 3.6, são detalhados os algoritmos utilizados.

**Tabela 3. Exemplo de saída do algoritmo de regularização**

Id	Coordenada 1	Coordenada 2
0	0.13884248	0.11291029
1	0.13011554	0.12082376
2	0.12334355	0.13454545
3	0.12324345	0.12324455
...	...	...

### 3.6. Classificação

Nesta etapa, cada comentário é classificado como útil ou não útil a partir das características extraídas do método de regularização. Ou seja, a partir das coordenadas geradas para cada comentário para as classes *útil* e *não útil*, é possível tratar o problema de predição de utilidade como um problema de predição estruturada.

Com isso, avaliaram-se diversos algoritmos de classificação implementados pela ferramenta *scikit-learn* [Pedregosa et al. 2011], tais como: *Support Vector Machine (SVM)*, *Naive Bayes*, *C4.5* e *Multi-layer Perceptron (MLP)*.

Na Seção 4, serão detalhados os experimentos e os resultados obtidos pela nossa abordagem. Além disso, é apresentada uma comparação com um *baseline* bem conhecido da área.

## 4. Experimentos e Resultados

Para a aplicação da proposta deste trabalho, durante a experimentação, o conjunto de comentários foi anotado de acordo com os votos de utilidade extraídos do *Google Play*. Cada comentário recebeu um valor de utilidade calculado de acordo com a Equação 3, onde a utilidade de um comentário é igual ao número de votos que o comentário recebeu dividido pela soma dos votos de todos os comentários. Após o cálculo, os comentários são ordenados do maior valor de utilidade para o menor. Considerou-se os 40% comentários iniciais da lista e os 40% comentários finais da lista como úteis e não úteis, respectivamente. Os 20% de comentários restantes (do meio da lista) são desconsiderados por existir a possibilidade de sobreposição de classes. Então, efetivamente são considerados 1.600 comentários (800 úteis e 800 não úteis).

$$utilidade_{c_i} = \frac{votos_{c_i}}{\sum_{c_j \in C} c_j} \quad (3)$$

Com o objetivo de avaliar a abordagem baseada em redes, utilizou-se quatro algoritmos de classificação sobre o resultado da regularização dos 1.600 comentários considerados para a experimentação. Além disso, comparou-se essa abordagem com uma adaptação da abordagem baseada em lógica *fuzzy* de [de Sousa et al. 2015], que classifica os comentários em 4 classes: *Insuficiente*, *Suficiente*, *Bom* e *Excelente*. Porém, como no presente trabalho são utilizadas duas classes, a adaptação consistiu em considerar as classes *Excelente* e *Bom* como úteis e as classes *Suficiente* e *Insuficiente* como não úteis.

O trabalho de [de Sousa et al. 2015] foi usado para comparação pois, dentre os trabalhos para a língua portuguesa foi o mais simples em termos de adaptação e recuperação. Ele não necessita de muitas variáveis dependentes de domínio para ser replicado.

Além da comparação com o trabalho de [de Sousa et al. 2015] realizou-se a implementação de dois métodos *baseline* para comparação de resultados. O primeiro *baseline* é um método que usa um *bag-of-words* como *features*. Já o segundo *baseline* é um método que utiliza três *features* simples e bem conhecidas na literatura (tamanho médio de sentenças, quantidade de *tokens* e quantidade de sentenças). Em ambos os métodos as *features* são aplicadas no algoritmo *Naive-Bayes*.

Como o algoritmo de regularização necessita de uma porção dos comentários pré-anotados, adotou-se a estratégia de anotar os comentários percentualmente e progressivamente, em: 0.5%, 1.0%, 1.5% e 2.0% de comentários como úteis e não úteis, escolhidos aleatoriamente. Por exemplo,  $0.5\% * 1600 = 8$ , ou seja, 4 comentários úteis e 4 comentários não úteis. Com isso, avaliou-se cada algoritmo de classificação sobre essas configurações. Para isso, aplicou-se a técnica de validação cruzada sobre os comentários com  $k = 10$ . Na Tabela 4, são mostrados os resultados para cada porção de comentários anotados e para cada algoritmo.

Pode-se ver que o algoritmo que obteve melhores resultado foi o MLP. Além disso, ele também atingiu melhor resultado com 2% de comentários anotados manualmente. Vale ressaltar que foram realizados experimentos um número maior de comentários pré-anotados, porém, não observou-se uma melhora nos resultados. Após os 2% de comentários o valor permaneceu estável para as demais quantidades de comentários. Adicionalmente, a Tabela 5 apresenta os resultados da execução do MLP em termos de Precisão,

**Tabela 4. Resultados dos algoritmos de classificação para cada porção de comentários anotados**

Comentários Pré-anotados	SVM	Naive Bayes	C4.5	MLP
	F1	F1	F1	F1
0.5 %	0.50	0.53	0.53	<b>0.58</b>
1.0 %	0.49	0.56	0.50	<b>0.57</b>
1.5 %	0.50	0.51	0.51	<b>0.58</b>
2.0 %	0.58	0.53	0.55	<b>0.60</b>

**Tabela 5. Valores por classe do algoritmo MLP para o melhor resultado**

Classe	Precisão	Recall	F1
Útil	0.62	0.57	0.61
Não Útil	0.59	0.63	0.60

**Tabela 6. Comparação entre as abordagens**

Abordagem	F1
<i>Baseline 1 (Textuais)</i>	0.58
<i>Baseline 2 (BoW)</i>	0.57
<i>Fuzzy</i>	0.47
Rede (nossa)	<b>0.60</b>

*Recall* e F1 para cada classe, e nela é possível observar que os resultados das classes possuem valores próximos. A Tabela 6 mostra, para comparação, os resultados para as execuções de todos os métodos analisados neste trabalho.

Pode-se observar que nossa abordagem em rede superou o método baseado em lógica *fuzzy* na medida F1 e aos dois métodos *baseline*. Esse experimento evidencia que uma abordagem baseada em rede tem potencial para modelar e prever a utilidade de comentários. No entanto, ainda resta investigar outras topologias de redes, algumas métricas específicas e ainda realizar a adaptação de *features* usadas em trabalhos em línguas estrangeiras para comparação.

Vale ressaltar que devido ao resultado muito próximo com os métodos *baseline* foi realizado um teste de significância estatística entre os resultados dos métodos. Encontrou-se que o P-valor para os resultados é menor que 0,05 o que indica que são estatisticamente significativas com 95% de nível de confiança. Quanto à diferença entre os resultados do método de redes com o *Fuzzy* o P-valor também é menor que 0,05 o que indica que são estatisticamente significativas com 95% de nível de confiança.

## 5. Conclusão e Trabalhos Futuros

Neste trabalho, modelou-se o problema de predição de utilidade de opiniões como uma rede heterogênea. A partir da rede, aplicou-se um algoritmo de regularização para extrair as características dos comentários de acordo com seus relacionamentos na rede. Por fim, avaliou-se diversos algoritmos de classificação sobre as características extraídas. Num experimento, essa abordagem superou uma abordagem baseada em lógica *fuzzy*. Para comparar as abordagens, extraiu-se 2.000 comentários sobre aplicativos da *Google Play*.

A principal contribuição deste trabalho é a apresentação de uma nova modelagem para a tarefa de predição da utilidade de opiniões. A capacidade de informatividade das redes é um grande diferencial em comparação às abordagens que utilizam tabelas atributo-valor. Apesar dessa modelagem ser relativamente embrionária, os resultados demonstram

que há benefícios em seu uso.

Como trabalhos futuros, além de construir um corpus maior e multi-domínio, pretende-se seguir em duas linhas. Num viés mais profundo, ou seja, linguisticamente motivado, serão avaliadas características linguísticas que possam indicar se um comentário é útil ou não. De outro lado, deverão ser adaptados trabalhos já desenvolvidos para a língua inglesa, de modo a investigar as características independentes de língua.

No que diz respeito à modelagem do problema como rede, novas topologias devem ser investigadas para que melhores resultados sejam alcançados. Por exemplo, topologias bipartidas (termos x comentários), ou ainda topologias com várias camadas (Uma camada comum e uma camada com informações semânticas do tipo, sinônimos, entidades nomeadas, palavras de sentimento, etc.). Além disso, as novas topologias podem permitir o uso de diversas métricas topológicas importantes. Como por exemplo, a descoberta de *hubs* e comunidades com o uso de medidas de centralidade (Coeficiente de agrupamento, *betweenness*, *closeness*, etc).

## 6. Agradecimentos

Os autores agradecem ao IFPI pelo apoio a este trabalho.

## Referências

- Anchiêta, R., Sousa, R. F., Moura, R., and Pardo, T. (2017). Improving opinion summarization by assessing sentence importance in on-line reviews. In *Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology*, pages 32–36.
- Anchiêta, R. T. and Moura, R. S. (2017). Exploring unsupervised learning towards extractive summarization of user reviews. In *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web*, pages 217–220. ACM.
- Barbosa, J. L. and Moura, R. S. (2016). Avaliação automática da utilidade de reviews usando redes neurais artificiais no corpus do steam. In *Anais do XXVI Congresso da Sociedade Brasileira de Computação: BraSNAM - 5º Brazilian Workshop on Social Network Analysis and Mining*. Brazilian Computer Society.
- Bertaglia, T. F. C. and Nunes, M. d. G. V. (2016). Exploring word embeddings for unsupervised textual user-generated content normalization. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, pages 112–120.
- de Sousa, R. F., Rabêlo, R. A., and Moura, R. S. (2015). A fuzzy system-based approach to estimate the importance of online customer reviews. In *Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on*, pages 1–8. IEEE.
- Diaz, G. O. and Ng, V. (2018). Modeling and prediction of online product review helpfulness: A survey. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 698–708.
- Fonseca, E. R. and Rosa, J. L. G. (2013). Mac-morpho revisited: Towards robust part-of-speech tagging. In *Proceedings of the 9th Brazilian symposium in information and human language technology*, pages 98–107.
- Hartmann, N. S., Avanço, L. V., Balage Filho, P. P., Duran, M. S., Nunes, M. D. G. V., Pardo, T. A. S., Aluisio, S. M., et al. (2014). A large corpus of product reviews in por-

- tuguese: Tackling out-of-vocabulary words. In *International Conference on Language Resources and Evaluation*. European Language Resources Association-ELRA.
- Kim, S.-M., Pantel, P., Chklovski, T., and Pennacchiotti, M. (2006). Automatically assessing review helpfulness. In *Proceedings of the 2006 Conference on empirical methods in natural language processing*, pages 423–430. Association for Computational Linguistics.
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.
- Liu, J., Cao, Y., Lin, C.-Y., Huang, Y., and Zhou, M. (2007). Low-quality product review detection in opinion summarization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.
- Malik, M. and Hussain, A. (2017). Helpfulness of product reviews as a function of discrete positive and negative emotions. *Computers in Human Behavior*, 73:290–302.
- Martins, A. C. S. and Tacla, C. A. (2015). Assesment of features influencing the voting for opinions’ helpfulness about services in portuguese. In *Proceedings of the annual conference on Brazilian Symposium on Information Systems: Information Systems: A Computer Socio-Technical Perspective-Volume 1*, page 21. Brazilian Computer Society.
- Orengo, V. and Huyck, C. (2001). A stemming algorithm for the portuguese language. In *String Processing and Information Retrieval*, pages 186–193.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Rossi, R. G. (2016). *Classificação automática de textos por meio de aprendizado de máquina baseado em redes*. PhD thesis, Universidade de São Paulo.
- Santos, R. L. d. S., de Sousa, R. F., Rabelo, R. A., and Moura, R. S. (2016). An experimental study based on fuzzy systems and artificial neural networks to estimate the importance of reviews about product and services. In *Neural Networks (IJCNN), 2016 International Joint Conference on*, pages 647–653. IEEE.
- Scarton, C. E. and Aluísio, S. M. (2010). Análise da inteligibilidade de textos via ferramentas de processamento de língua natural: adaptando as métricas do coh-metrix para o português. *Linguamática*, 2(1):45–61.
- Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., and Roy, P. K. (2017). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, 70:346–355.
- Zhu, X., Ghahramani, Z., and Lafferty, J. D. (2003). Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th International conference on Machine learning (ICML-03)*, pages 912–919.