Avaliação automática da utilidade de *reviews* usando Redes Neurais Artificiais no *corpus* do *Steam*

Jardeson L. N. Barbosa[1] **Raimundo S. Moura**[1]

[1]Departamento de Computação – Universidade Federal do Piauí (UFPI)
64.049-550 – Teresina – PI – Brasil

jardesonbarbosa@gmail.com, rsm@ufpi.edu.br

Abstract. *Through online reviews, consumers can communicate with product suppliers and influence the buying decision of other consumers over the Internet. However, thanks to the high number of published reviews daily, it is difficult to identify which texts to read. As a solution for that problem, some websites use a review evaluation system based on the users votes, which, while useful, is not always ideal. This paper proposes an automatic method to analyze the helpfulness of online user reviews from Steam using Multilayer Perceptron Artificial Neural Network. We found out that certain features of reviews affect the perception of helpfulness and we discuss applications and future researches.*

Resumo. *Através de reviews, consumidores podem se comunicar com os fornecedores de produtos e serviços e influenciar a decisão de compra de outros consumidores na Internet. Porém, com o alto número de reviews publicados diariamente, é difícil identificar quais textos devem ser lidos. Como uma solução para esse problema, alguns sites utilizam um sistema de avaliação de reviews baseado no voto dos usuários que embora útil, nem sempre é ideal. Este trabalho propõe um modelo de análise automática da utilidade de reviews online de usuários do Steam, usando Rede Neural Artificial Perceptron Multicamadas. Descobriu-se que certas características de reviews afetam a percepção de utilidade e discutimos aplicações e pesquisas futuras.*

1. Introdução

Com números que ultrapassam 12 milhões de lojas mundialmente, o *e-commerce* é, atualmente, um dos principais serviços da Internet[1]. Porém, graças à enorme quantidade de páginas e produtos disponíveis, essa pode ser uma atividade desgastante, porque, durante o processo de busca de informação, consumidores são geralmente submetidos a múltiplas escolhas [Vohs et al. 2014].

Para auxiliar o processo de decisão dos consumidores, algumas empresas de *e-commerce* introduziram alguns sistemas de recomendação automáticos que oferecem recomendações personalizadas aos consumidores. Entretanto, apesar do bom desempenho, esses sistemas ainda são menos eficientes do que recomendações de pessoas conhecidas, como familiares e amigos [Sinha and Swearingen 2001].

[Goldenberg et al. 2001] demonstraram que o processo de tomada de decisão dos consumidores é fortemente influenciado pela comunicação "boca-a-boca" (WOM, do

[1]http://blog.lemonstand.com/just-how-big-is-the-ecommerce-market-youll-never-guess

inglês "word of mouth"). Na Internet, redes sociais *online* concentram a comunicação WOM, uma vez que permitem que consumidores compartilhem suas experiências e recomendações através de opiniões, ou *reviews* – seja escrevendo, votando, comentando ou compartilhando [Kim and Srivastava 2007].

Consumidores que leem *reviews* online se atentam não somente às notas dos *reviews*, mas também às demais informações textuais e não-textuais [Hu et al. 2008]. Entretanto, dada a enorme quantidade de *spam*, textos de baixa qualidade e informações falsas na Internet, consumidores podem necessitar de uma grande quantidade de tempo para que encontrem *reviews* que afetem a decisão de compra.

Uma solução encontrada pela maioria dos *sites* de comércio eletrônico foi o estabelecimento de um padrão chamado de "utilidade do *review*". Em *sites* como *Amazon*[2], usuários podem votar em *reviews* que eles consideram úteis ou inúteis no processo de decisão de compra. Esses votos são utilizados para ordernar os *reviews* em função dos mais úteis. Dessa forma, os próprios usuários filtram os textos que devem ser lidos. Um dos problemas desse método é que *reviews* precisarão de muitos votos positivos para serem lidos e, consequentemente, postagens mais recentes deixarão de ser lidas por não possuírem votos suficientes, mesmo que sejam mais relevantes que as postagens mais votadas [Li et al. 2013].

Neste trabalho, estuda-se a percepção de utilidade de *reviews online* através da análise das descrições textuais e informações quantitativas presentes no texto. Este trabalho propõe um modelo capaz de prever o grau de percepção de utilidade de *reviews online* de forma automática utilizando uma Rede Neural Artificial *Multilayer Perceptron* (RNA MLP). Esta investigação é baseada em textos coletados da loja *Steam*[3], que distribui jogos para computadores pessoais. O site *Steam* foi escolhido por oferecer uma loja virtual e rede social com mais de 125 milhões de usuários ativos[4] e por permitir que seus usuários publiquem, leiam, compartilhem e votem em *reviews online*. Além disso, ainda são poucos os trabalhos que exploram a mineração de opiniões do *site Steam*.

O restante deste artigo é organizado da seguinte forma: a Seção 2 trata da utilidade de *reviews online* e apresenta uma revisão literária realizada na área. A Seção 3 apresenta o nosso modelo proposto. A Seção 4 apresenta os resultados dos experimentos. A Seção 5 apresenta as considerações finais e trabalhos futuros.

2. Revisão literária

2.1. Percepção de utilidade de reviews *online*

Em um sitema de votos como o proposto pelo *site Amazon*, a utilidade de um *review* pode ser definida como

$$H = \frac{n_p}{n_p + n_n} \tag{1}$$

onde n_p representa o número de votos positivos e n_n representa o número de votos negativos [Kim et al. 2006]. Entretanto, essa abordagem apresenta algumas tendências desvantajosas como, por exemplo, um cenário em que *reviews* novos ou com poucos votos

[2]https://www.amazon.com.br/
[3]http://store.steampowered.com
[4]http://store.steampowered.com/news/16000/?l=portuguese

nunca sejam lidos e, consequentemente, votados [Li et al. 2013]. Além disso, nem todo mundo que lê um *review* necessariamente vota ao final [Kim et al. 2006].

Outro problema dessa abordagem é que ela não leva em consideração a quantidade de votos recebidos por cada *review*. É impossível, por exemplo, diferenciar dois *reviews* com 80% de aprovação, mesmo que um deles possua 10000 votos e outro apenas 10. Por se tratar de uma distribuição Binomial, uma solução eficiente para contornar esse último problema é o uso do limite inferior do intervalo de confiança de Wilson para uma Binomial [Agresti and Coull 1998] apresentada em (2)

$$H = \left(\hat{p} + \frac{z^2_{\alpha/2}}{2n} \pm z_{\alpha/2} \sqrt{[\hat{p}(1-\hat{p}) + z^2_{\alpha/2}/4n]/n} \right) / (1 + z^2_{\alpha/2}/n) \tag{2}$$

onde $\hat{p}$ representa a porcentagem de votos positivos observados, n é o total de votos recebidos e z é o quantil da distribuição normal padronizada para o intervalo de confiança escolhido.

Essa equação considera a quantidade de votos de um *review* como uma amostra de uma população hipotética em que todos os *reviews* são votados em igual quantidade. Assumindo um determinado nível de confiança (95% neste trabalho), é determinada uma nova probabilidade de sucesso para a distribuição analisada. Esse mesmo cálculo é utilizado em alguns *sites* no processo de ordenação de *reviews*, tópicos e comentários, como *Reddit*[5] e *Yelp*[6].

Neste trabalho, a medida de utilidade de *reviews online* usada será a porcetagem de votos positivos do *review* com base na equação (2). Formalmente, entende-se a medida de utilidade de *reviews* de produtos *online* como a medida na qual consumidores percebem um *review* de produto como sendo capaz de facilitar o julgamento ou decisão de compra [Li et al. 2013].

2.2. Características formativas da percepção de utilidade de reviews *online*

Evidências suportam a conceitualização teórica de utilidade de *reviews* como uma construção formativa [Li et al. 2013]. Assim, entende-se que existe uma série de características que influenciam a percepção de utilidade de *reviews*.

Alguns trabalhos identificaram a influência direta do conteúdo do *review* na percepção de sua utilidade. Com a ajuda de um modelo de aprendizado de máquina supervisionado, [Kim et al. 2006] descobriram que o tamanho do texto, suas unigramas e a classificação do produto são características importantes na definição da percepção de utilidade de *reviews*. [Danescu-Niculescu-Mizil et al. 2009] realizaram um estudo sobre o *corpus* do *site Amazon* e identificaram que um *review* é avaliado como mais útil quando a sua nota (ou avaliação) é próxima ao consenso das demais notas para o mesmo produto. [Schindler and Bickart 2012] examinaram características do conteúdo do *review* e descobriram que o tamanho do texto e a quantidade de informações expressas influenciam positivamente a percepção de utilidade de *reviews*, mas somente até certo ponto. [Lee and Choeh 2014] descobriram que o tamanho do texto e o número de palavras com apenas uma letra são bons indicativos da percepção de utilidade do review.

[5]http://www.redditblog.com/2009/10/reddits-new-comment-sorting-system.html
[6]http://officialblog.yelp.com/2011/02/the-most-romantic-city-on-yelp-is.html

Alguns pesquisadores também exploraram a influência da declaração de autoria na percepção de utilidade de *reviews*. [Connors et al. 2011] examinaram alguns fatores básicos associados com a utilidade de *reviews*. Eles perceberam que a *expertise* do autor, ou seja, o conhecimento do autor a respeito do domínio do produto, possui influência positiva na percepção de utilidade de *reviews*. [Forman et al. 2008] descobriram que a presença de informações descritivas sobre a identidade do autor de *reviews online* possui impacto positivo na percepção de utilidade de *reviews*.

Este trabalho apresenta como diferencial o uso de RNA, a análise da utilidade de *reviews* escritos em Português Brasileiro e a exploração de características de *reviews* pouco estudadas em trabalhos anteriores, como o histórico de publicações dos autores de *reviews*, descrições sobre o uso do produto avaliado e a quantidade de relacionamentos dos autores dentro de uma rede social *online*. Essas característas puderam ser facilmente extraídas graças ao modelo de comunidade virtual proposto pelo *Steam*.

3. Modelo proposto

Para um dado *review*, o objetivo deste trabalho é encontrar o grau de percepção de utilidade *H*, definido pela equação (2), onde *H* é um número real na faixa de 0 a 1. Para desenvolvermos um modelo que prevê essa taxa, estudou-se as características formativas da percepção de utilidade de *reviews*. Com base no nosso estudo, hipotetiza-se que existam três importante grupos de características de *reviews* que modelam a sua percepção de utilidade: características a respeito do autor do *review*, características textuais e características relativas aos metadados do *review*. O modelo proposto é dado por uma função de regressão que tem como entrada um vetor X contendo essas características e como saída um valor escalar H, que é a medida de percepção de utilidade. Escolheu-se uma RNA MLP para aproximar essa função. Uma RNA MLP para aproximação de funções pode ser descrita da seguinte forma:

$$y(x_1, x_2, ..., x_n) = \sum_{i=1}^{n_1} \lambda_i g_j(u_i) \tag{3}$$

$$u_i = \sum_{j=1}^{n} W_{ij} x_j - \theta_i \tag{4}$$

onde a função y a ser mapeada pelo MLP será constituída pela superposição de funções de ativação do tipo logística, representadas pelos termos $g_j(u_i)$, as quais são ponderadas pelos fatores λ_i [da Silva et al. 2010]. A função logística é definida como

$$g(u) = \frac{1}{1 + e^{-\beta * u}} \tag{5}$$

onde β é uma constante real associada ao nível de inclinação da função logística frente ao seu ponto de inflexão.

As RNA são adequadas para esse problema porque são aproximadores de funções universais. O teorema da aproximação universal aplicado ao MLP fornece as bases necessárias para definir as configurações estruturais dessas redes com a finalidade de mapear funções algébricas [da Silva et al. 2010]. Além disso, o modelo de RNA permite

que qualquer forma funcional que relacione as variáveis independentes à variável dependente não precise ser imposto pelo modelo, ou seja, é possível capturar relacionamentos não-lineares nos dados sem a necessidade de especificação prévia [Lee and Choeh 2014].

Para definição das características no modelo proposto, analisou-se dados de *reviews* sobre jogos digitais do *site Steam* descritos nas subseções a seguir.

3.1. Coleta dos dados

Diferente de um grande número de *sites* de *e-commerce*, o *Steam* se comporta como uma comunidade online. Usuários do *Steam* possuem perfis com foto; podem adicionar amigos; trocar mensagens privadas e públicas entre si; publicar imagens, textos e vídeos, e podem participar de grupos.

O *site Steam* possui um detalhado sistema de *reviews* que inclui, além do texto do usuário, diversas outras informações. A Figura 1 mostra um típico *review* do *Steam*. A avaliação final do usuário é apresentada em destaque acima do texto (Recomendado ou Não Recomendado). Em um *review* também se encontram informações sobre o autor (nome, quantidade de produtos na conta e quantidade de análises), a quantidade de horas gastas com o produto, a quantidade de pessoas que consideraram a análise útil, além do sistema de votos que inclui a opção "engraçado", que não interfere no voto de utilidade.

Figura 1. Exemplo de *review* do *Steam* com 245 votos positivos.

A coleta dos dados para o treinamento e teste do modelo foi realizada de forma automática através de um *scraper*[7] escrito em *Python* utilizando a biblioteca *Scrapy*[8]. Definiu-se como a população de interesse, todos os *reviews* de jogos digitais do *site Steam* escritos em Português Brasileiro, com pelo menos 1 voto de utilidade e postados por usuários que possuem perfis públicos. Como os *reviews* do *Steam* são ordenados de forma que *reviews* com poucos votos ou com nota baixa sejam dificilmente acessíveis, a coleta concentrou-se apenas nos 100 jogos mais jogados no *Steam* no dia 26 de fevereiro. Essa lista é atualizada diariamente no próprio *site*[9].

Dos 100 jogos mais jogados no dia 26 de fevereiro, 5 não possuíam *reviews* em português brasileiro com pelo menos 1 voto. Inicialmente coletou-se uma amostra de 10.168 *reviews* retroativamente publicados entre 25 de novembro 2013 e 26 de fevereiro de 2016. Optou-se por eliminar *reviews* com menos de três votos para evitar assimetria e tendência nos dados, já que a maioria dos *reviews* possui poucos votos.

[7]Técnica computacional para a extração de informações de *websites*
[8]http://scrapy.org/
[9]http://store.steampowered.com/stats/

Como, neste trabalho, objetiva-se uma análise automática, optou-se por eliminar todos os textos com grau de corretude abaixo de um *threshold* definido arbitrariamente a fim de remover ruídos dos dados (como textos em outros idiomas). Modelou-se a corretude do texto como o número de palavras corretas divido pelo número total de palavras em um texto. Para determinar se a palavra pertence ou não ao Português, usou-se o dicionário *br.ispell* da USP[10]. Eliminando os *reviews* com menos de 3 votos e os reviews com corretude abaixo do *threshold*, definido como a média da amostra inicial (0.76), obteve-se uma amostra final contendo 5823 *reviews*.

3.2. Características de *reviews* analisadas

3.2.1. Características relativas à autoria

Em relação ao autor, destacam-se duas características: a reputação e a *expertise*. Estudos sugerem que *reviews* submetidos por autores com histórico positivo são vistos como mais úteis [Wathen and Burkell 2002]. A reputação também incorpora elementos referentes à credibilidade do autor. A *expertise* indica o nível de conhecimento de um autor a respeito do tópico abordado. Modelou-se a reputação de um autor a partir de três variáveis: a média de votos de *reviews* pela quantidade de *reviews* publicados pelo usuário, a razão dos votos positivos pela quantidade total de votos do usuário; e o número de amigos que o usuário possui na comunidade *Steam*. Por *expertise*, pode-se explorar a quantidade de horas que o autor do *review* dedicou ao jogo analisado. Um autor que jogou apenas 2 horas de um determinado jogo pode ter menos proficiência no tópico que foi analisado do que aquele que jogou 20 horas.

3.2.2. Características relativas ao conteúdo textual do *review*

Fundamentalmente, pesquisas anteriores se dedicaram a dois tipos de análise textual em *reviews online*: análise das características semânticas e análise das características estilísticas do texto [Kim et al. 2006, Cao et al. 2011]. Ao explorar o conteúdo textual de um *review*, pode-se atentar às opiniões e sentimentos expressos no texto. De acordo com o dicionário Oxford (tradução livre), opinião é uma visão ou julgamento formada a respeito de algo, não necessariamente baseada em fatos ou conhecimento.

Hipotetiza-se que *reviews online* com uma grande quantidade de opiniões são mais informativos. Nesta abordagem, opinões devem ser modeladas de forma númerica. Logo, é interessante saber apenas a quantidade de opinões expressas em cada *review*. Para tal, realiza-se um processo de extração de opiniões segundo uma adaptação do modelo proposto por [Sousa et al. 2015].

Na extração de opiniões, os textos são analisados de acordo com a estrutura frasal das sentenças. Adotou-se o padrão *"sujeito + verbo + predicativo do sujeito"*, onde o núcleo do sujeito representa a característica qualificada e o núcleo do predicado representa a palavra qualificadora. Por exemplo, "esse jogo é divertido" é extraído como *(jogo, divertido)*. Esses padrões linguísticos foram definidos através de uma análise manual realizada sobre uma outra amostra de 385 *reviews* coletadas no mesmo *site* (nível de confiança 95% e erro 5%). O resultado é apresentado na Tabela 1.

[10]http://www.ime.usp.br/ ueda/br.ispell/

Tabela 1. Padrões linguísticos

SUBST ADV? V? (ADV? ADJ — ADV V)*
ADJ SUBST
ADV ADJ

Foram considerados como elementos estilísticos a legibilidade do texto e o tamanho do texto. Legibilidade é a facilidade em que um texto pode ser compreendido. A legibilidade de um texto em inglês pode ser analisada utilizando o teste de legibilidade de Flesch-Kincaid, um método matemático que avalia o quão legível é um texto por meio da sua média de palavras por sentenças e média de sílabas por total de palavras [Kincaid et al. 1975]. Uma adaptação para o Português foi feita pelo jornalista Alberto Dines [Squarisi and Salvador 2005]

$$\left(\frac{n_w}{n_s} + n_p\right) * 0.4 \tag{6}$$

onde n_w é o número total de palavras do texto, n_s é o número total de sentenças e n_p é o número total de polissílabos. A constante 0,4 representa a média de letras da palavra na frase do Português. Quanto maior o *score*, menos legível é o texto. Um texto com *score* 1 pode ser facilmente lido por qualquer pessoa.

Baseado em descobertas de [Kim et al. 2006] e [Lee and Choeh 2014], modelou-se o tamanho do texto a partir do número de palavras, número de sentenças e a quantidade de monossílabos do texto, já que o Português não possui grande quantidade de palavras semanticamente relevantes com apenas uma letra.

3.2.3. Metadados do *review*

Consideram-se como metadados a avaliação final do *review* e a data de postagem do *review* em dias. Avaliação final refere-se à avaliação de produto binária presente em cada *review*, que pode ser "recomendado" ou "não recomendado". É importante notar que leitores podem ser influenciados pela média das avaliações dos produto. Esse ideia foi explorada por [Danescu-Niculescu-Mizil et al. 2009].

Considerando uma avaliação de *review* como sendo 1, para "recomendado" (ou positivo), e 0, para "não recomendado" (ou negativa), modelou-se a avaliação como uma expressão, onde p é porcetagem de *reviews* com avaliações positivas (em representação decimal) para um produto e x é a avaliação individual do *review* para o mesmo produto:

$$xp + (1 - x) * (1 - p) \tag{7}$$

Partindo das descobertas de [Otterbacher 2009], modelou-se o tempo de postagem de um review como a diferença entre a data de lançamento do produto e a data de postagem do *review* em dias.

3.3. Validação

A RNA MLP para aproximação de funções é composta de três camadas, sendo uma camada de entrada, apenas uma camada escondida e uma camada de saída. A camada de

entrada mapeia cada uma das variáveis de entrada estudadas. A camada de saída possui apenas um neurônio que mapeia a variável de saída H, ou utilidade do *review*. A validação cruzada permite definir a quantidade mais adequada de neurônios na camada escondida. Utilizou-se a função logística como ativação para os neurônios da camada escondida (5). Já para a camada de saída, utilizou-se uma função de ativação linear (3), já que o neurônio de saída realiza apenas uma combinação linear das funções de ativação logística implementadas nos neurônios da camada escondida [da Silva et al. 2010]. Portanto, após o processo de treinamento da rede MLP, a matriz de pesos referentes ao neurônio de saída corresponderá aos próprios parâmetros λ_i da equação (3), isto é $\lambda_i = W_{1,i}$. A ilustração dessa rede é apresentada na Figura 2.

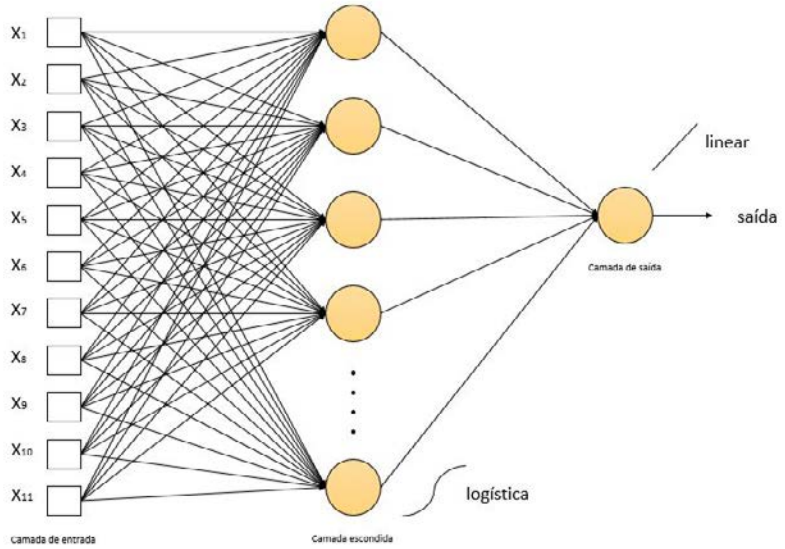

Figura 2. Ilustração da rede MLP aplicada em aproximação funcional

As variáveis de entrada, definidas na especificação de hipóteses são: i) a quantidade de votos de *reviews* pela quantidade total de *reviews* de um usuário; ii) a razão dos votos positivos pela quantidade total de votos de um usuário; iii) o número de amigos do usuário na comunidade *Steam*; iv) a quantidade de horas (em horas) que o autor dedicou ao jogo analisado; v) a quantidade de opiniões expressas no texto; vi) a legibilidade do texto; vii) o tamanho do texto em palavras; viii) o número de sentenças no texto; ix) o número de palavras monossílabas no texto; x) a diferença entre a avalição média do produto e a avaliação do usuário; xi) a diferença entre a data do lançamento do produto e a data da publicação do *review* em dias. A variável de saída esperada é a utilidade do *review*.

Pode-se calcular o relacionamento entre cada variável de entrada e a variável de saída utilizando o cálculo de *Relative Strength* (RS) [Lee and Choeh 2014] [Yoon et al. 1994]:

$$RS_{ji} = \frac{\sum_{k=0}^{n}(W_{ki}W_{jk})}{\sum_{i=0}^{m}\sum_{k=0}^{n}|W_{ki}W_{jk}|} \tag{8}$$

onde W_{ki} indica o peso entre a k-ésima unidade escondida e a i-ésima unidade de entrada. W_{jk} denota o peso entre a j-ésima unidade de saída e a k-ésima unidade escondida. RS_{ji} calcula o peso relativo entre a i-ésima variável de entrada e a j-ésima variável de saída. O numerador do cálculo de RS_{ji} calcula a relação entre a i-ésima variável de entrada e a j-ésima variável de saída e pode ser tanto positivo quanto negativo, dependendo dos pesos. Já o denominador calcula o relacionamento total entre todas as variáveis de entrada e saída.

4. Resultados

A especificação da topologia de rede MLP mais apropriada para mapear um problema específico é usualmente efetuada de forma empírica [da Silva et al. 2010]. O processo de seleção de topologias candidatas geralmente se dá por tentativa e erro [Lee and Choeh 2014]. Para a seleção da melhor topologia dentre as candidatas, optou-se pela validação cruzada *k-fold*, ou validação cruzada *k-partições*. Nesse processo, a amostra total é dividida em k partições, sendo $(k-1)$ delas usadas para compor o subconjunto de treinamento enquanto a partição restante constui o subconjunto de teste [da Silva et al. 2010]. O processo de aprendizado se repete k vezes até que todas as partições tenham sido utilizadas como subconjunto de teste. Neste trabalho, definiu-se k como 10. O desempenho global de cada topologia candidata é obtido em função da média entre os desempenhos individuais observados quando da aplicação das k partições.

A RNA implementada neste estudo tem valores randômicos como pesos iniciais, taxa de aprendizado 0.1 e *momentum* 0.3. Optou-se por valores pequenos para garantir que a convergência ocorra, mesmo que o processo necessite de mais épocas. Para evitar *overfitting*, cada topologia foi treinada até que fosse encontrado a menor raiz quadrada do erro quadrático médio (RMSE) local ou atingisse 10^4 épocas, o que ocorrer primeiro.

A Figura 3 apresenta um gráfico com os RMSE para cada uma das topologias testadas. Iniciamos com uma topologia de 1 neurônio na camada escondida e aumentamos essa camada em 1 neurônio a cada nova topologia. Observamos que o erro aumenta a cada novo neurônio incluído na camada escondida, mas passa a diminuir assim que atinge 5 neurônios. Escolheu-se como melhor topologia, aquela que apresenta o menor número de neurônios na camada escondida e o menor RMSE. O modelo final de RNA MLP escolhido é uma rede de três camadas que consiste em 11 nós de entrada, 5 nós na camada escondida e 1 nó de saída. Dada a complexidade de realizar predições baseadas no comportamento humano, o RMSE de 0.1929 foi considerado aceitável.

A Tabela 2 apresenta o peso relativo dos determinantes da utilidade dos *reviews* utilizando a equação (8). Valores acima de 0.02 são bons indicativos de impacto na percepção de utilidade do *review* [Lee and Choeh 2014]. Os melhores índices foram da média de votos por review do usuário, nota média do usuário, quantidade de horas jogadas e tamanho do texto em palavras.

Os resultados desta análise corroboram pesquisas anteriormente realizadas por [Danescu-Niculescu-Mizil et al. 2009], [Schindler and Bickart 2012] e [Kim et al. 2006]. Notou-se que a avaliação do usuário e o tamanho do texto são caracteristicas que impactam positivamente a percepção de utilidade de *reviews*. Já a data de postagem dos *reviews* aparenta não ter forte impacto na percepção utilidade em comunidades *online* como o Steam, o que contraria descobertas realizadas por [Otterbacher 2009]. Uma das possíveis

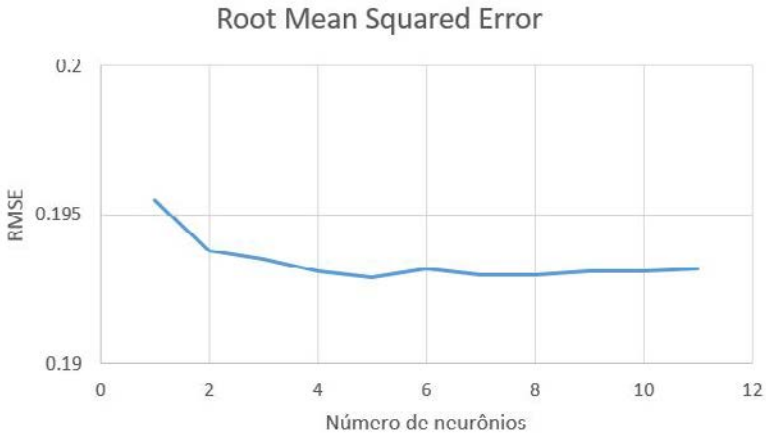

Figura 3. Variação do RMSE a medida que o número de neurônios na camada escondida aumenta.

Tabela 2. Peso relativo de cada variável.

Variáveis	Força
média de votos por review do usuário	0.14272
nota média do usuário	0.06112
quantidade de horas jogadas	0.02507
quantidade de amigos	0.01743
padrões linguísticos	-0.00442
legibilidade	0.00789
quantidade de palavras	0.02794
quantidade de sentenças	0.01224
diferença entre avaliação do usuário e média do produto	0.01302
quantidade de palavras monossílabas	-0.00962
diferença de dias entre a lançamento do produto e postagem	0.00122

justificativas poderia ser o fato de o *Steam* realizar promoções regulares que movimentam as páginas de produtos velhos.

Entre as novas descobertas, tem-se que a reputação do autor é fator importante na percepção de utilidade de *reviews*. Descobriu-se que autores com histório positivo de votos recebidos têm *reviews* melhor avaliados. A *expertise* do autor também aparenta ser uma característica importante. No caso do *Steam*, a quantidade de horas jogadas por um autor no momento da publicação do *review* tem impacto positivo sobre a percepção de utilidade. A quantidade de amigos do usuário na comunidade também possui algum impacto positivo sobre a percepção de utilidade de *reviews*.

5. Conclusões e trabalhos futuros

Este trabalho teve como objetivo estudar formas de quantificar e prever a percepção de utilidade de *reviews* de usuários *online*. Os resultados deste estudo foram positivos e dão indícios sobre quais características formam a percepção de utilidade de *reviews*. Para o setor de *e-commerce*, estes resultados dão pistas sobre quais estruturas de *website* podem ser utilizadas para que os melhores *reviews* sejam lidos pelos consumidores e, consequentemente, maximizar as vendas.

Para trabalhos futuros, considera-se substituir o uso de padrões linguísticos

por técnicas de PLN mais bem elaboradas como LSA, uma técnica de extração e representação de uso contextual de palavras utilizando computações estatísticas para um grande *corpus* de texto. O princípio da LSA é que o agregado de todos os contextos de palavras em que uma dada palavra aparece ou não aparece provém um conjunto de restrições mútuas que determina, em grande parte, a similaridade do significado de palavras e conjuntos de palavras [Landauer et al. 1998]. Espera-se que com técnicas mais robustas, seja possível identificar situações que não puderam ser analisadas em nosso modelo, como a ocorrência de sujeito oculto e o humor em textos.

Por fim, pretende-se explorar o impacto do humor na percepção de utilidade de *reviews*. O próprio *Steam* permite que usuários votem em *reviews* engraçados, opção pouco utilizada atualmente, mas que pode servir como base para estudos futuros.

Referências

[Agresti and Coull 1998] Agresti, A. and Coull, B. A. (1998). Approximate Is Better than ”Exact”for Interval Estimation of Binomial Proportions. *The American Statistician*, 52(2):119–126.

[Cao et al. 2011] Cao, Q., Duan, W., and Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2):511 – 521.

[Connors et al. 2011] Connors, L., Mudambi, S. M., and Schuff, D. (2011). Is it the review or the reviewer? a multimethod approach to determine the antecedents of online review helpfulness. In *In 44th Hawaii International Conference on System Sciences*.

[da Silva et al. 2010] da Silva, I. N., Spatti, D. H., and Flauzino, R. A. (2010). *Redes Neurais Artificiais: para engenharia e ciências aplicadas*, chapter Redes Perceptron Multicamadas. Artliber.

[Danescu-Niculescu-Mizil et al. 2009] Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., and Lee, L. (2009). How opinions are received by online communities: A case study on amazon.com helpfulness votes. *18th International Conference on World Wide Web,*.

[Forman et al. 2008] Forman, C., Ghose, A., and Wiesenfeld, B. (2008). *Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets*. INFORMS.

[Goldenberg et al. 2001] Goldenberg, J., Libai, B., and Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters, Volume 12, Issue 3*, pages 211–223.

[Hu et al. 2008] Hu, N., Liu, L., and Zhang, J. (2008). Do online reviews affect product sales? the role of reviewer characteristics and temporal effects. *Information Technology and Management Vol. 9 No. 3*, pages 201–214.

[Kim et al. 2006] Kim, S.-M., Pantel, P., Chklovski, T., and Pennacchiotti, M. (2006). Automatically assessing review helpfulness. In Jurafsky, D. and Gaussier, r., editors, *EMNLP*, pages 423–430. ACL.

[Kim and Srivastava 2007] Kim, Y. A. and Srivastava, J. (2007). Impact of social influence in e-commerce decision making. In *Proceedings of the Ninth International Conference on Electronic Commerce*, ICEC '07, pages 293–302, New York, NY, USA. ACM.

[Kincaid et al. 1975] Kincaid, J. P., Fishburne, R. P., Rogers, R. L., and Chissom, B. S. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Technical report.

[Landauer et al. 1998] Landauer, T., Foltz, P., and Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, 25:259–284.

[Lee and Choeh 2014] Lee, S. and Choeh, J. Y. (2014). Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Syst. Appl.*, 41(6):3041–3046.

[Li et al. 2013] Li, M., Huang, L., Tan, C.-H., and Wei, K.-K. (2013). Helpfulness of online product reviews as seen by consumers: Source and content features. *International Journal of Electronic Commerce*, 17:101–136.

[Otterbacher 2009] Otterbacher, J. (2009). 'helpfulness' in online communities: A measure of message quality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 955–964. ACM.

[Schindler and Bickart 2012] Schindler, R. and Bickart, B. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behaviour*, 11:234–243.

[Sinha and Swearingen 2001] Sinha, R. R. and Swearingen, K. (2001). Comparing recommendations made by online systems and friends. In *DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries*.

[Sousa et al. 2015] Sousa, R. F. d., Rabelo, R. A. L., and Moura, R. S. (2015). A fuzzy system-based approach to estimate the importance of online customer reviews. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*.

[Squarisi and Salvador 2005] Squarisi, D. and Salvador, A. (2005). A arte de escrever bem: um guia para jornalistas e profissionais do texto. pages 50–52. Editora Contexto.

[Vohs et al. 2014] Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., and Tice, D. M. (2014). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Motivation Science, Vol 1(S)*, pages 19–42.

[Wathen and Burkell 2002] Wathen, C. N. and Burkell, J. (2002). Believe it or not: Factors influencing credibility on the web. *Journal of the American Society for Information Science and Technology*, 53(2):134–144.

[Yoon et al. 1994] Yoon, Y., Guimaraes, T., and Swales, G. (1994). Integrating artificial neural networks with rule-based expert systems. *Decis. Support Syst.*, 11(5):497–507.