

Approach to define Author Reputation in Web Product Reviews using Artificial Neural Networks

Carlos Augusto de Sá, Raimundo Santos Moura
Departamento de Computação
Universidade Federal do Piauí
Teresina, Piauí
Email: carlos.sa@ufpi.edu.br,
rsm@ufpi.edu.br

Abstract—Author reputation is a very important variable for evaluating web comments. However, there is no formal definition for calculating its value. This paper presents an adaptation of the approach presented by Sousa (2015) for evaluating the importance of comments about products and services available online, emphasizing measures of author reputation. The implemented adaptation consists in defining six measures for authors, used as input in a Multilayer Perceptron Artificial Neural Network. On a preliminary evaluation, the Neural Network presented an accuracy of 91.01% on the author classification process. Additionally, an experiment was conducted aiming to compare both approaches, and the results show that the adapted approach had better performance in classifying the importance of comments.

Index Terms—Author Reputation, Artificial Neural Networks, Fuzzy Systems.

I. INTRODUÇÃO

Com o advento da Web 2.0, a prática de interagir na grande rede de computadores cresceu muito nos últimos anos. As Redes Sociais Online (RSO) como *Twitter*¹ e *Facebook*², bem como serviços de troca de mensagens pelo celular como o *Whatsapp*³, permitem grande interação social entre os usuários da grande rede. Além das RSOs, *sites* de comércio eletrônico, *sites* de hotéis, bem como outros *sites* permitem aos usuários deixarem as suas opiniões e comentários sobre a compra de um produto ou serviço.

A grande quantidade de informação que precisa ser tratada pelos usuários é um problema que surgiu com o crescimento da Web 2.0 e as mídias sociais *online*. Com este crescimento e com o avanço da Computação, a área de Processamento de Linguagem Natural (PLN) ganhou bastante destaque por permitir analisar grandes volumes de dados e realizar processamento de maneira mais eficiente, com o desenvolvimento de novas soluções para trabalhar com a complexidade dessas redes.

Tendo em vista tal cenário, um dos tópicos de pesquisa da Ciência da Computação explorado nos últimos anos é o desafio de transformar textos não estruturados em conhecimento estruturado, de forma que um usuário possa se beneficiar melhor

desse volume de informações [1]. Neste contexto surgiu a área de pesquisa denominada Análise de Sentimentos ou Mineração de Opiniões, que se preocupa em extrair o sentimento a partir de descrições textuais [2].

A Análise de Sentimentos é uma subárea de Processamento de Linguagem Natural (PLN), que envolve Ciência da Computação, Linguística e Inteligência Artificial e tem atacado o problema de manipular grandes volumes de informações através de técnicas que analisam a linguagem escrita ou falada [3]. Um desafio da área se encontra na filtragem ou pré-processamento de comentários Web, já que existe uma tendência dos usuários de sistemas *online* a escrever com muitas gírias, o que dificulta o trabalho das ferramentas tradicionais de PLN. Por exemplo, no português temos a palavra "vc" indicando o pronome "você", "naum" e "nao" são usados como o advérbio "não". Além disso, a pouca preocupação com a grafia correta das palavras, ironias e sarcasmos dificultam a atuação dos sistemas baseados em PLN.

Dentro do contexto de comércio eletrônico, a comunicação boca-a-boca (do inglês "*word-of-mouth*" - WOM) tem uma grande influência no processo de aquisição de produtos e serviços no cotidiano dos consumidores. Anteriormente, as pessoas buscavam opiniões a respeito de um determinado produto ou serviço questionando amigos próximos ou parentes. Este tipo de comunicação envolve todas as formas de troca de informações entre os consumidores independente das características e utilidade de um produto ou serviço em particular [4].

No entanto, com o avanço da *Internet*, os serviços como redes sociais e *e-commerce* permitiram uma nova dimensão no processo de escolha por parte de novos consumidores. Com advento destes serviços, diversos usuários podem deixar os seus comentários sobre produtos das mais diversas formas, por exemplo, escrevendo, votando, comentando ou compartilhando opiniões [5]. Desta forma, tem-se uma enorme base de dados para que os novos consumidores possam pesquisar e se beneficiar das opiniões de outros.

Segue que, os usuários de *sites* de *e-commerce*, por exemplo, passaram a comentar mais, devido aos diversos recursos de avaliação disponibilizados. Com mais comentários escritos, tem-se a necessidade de mecanismos automáticos para inferir os comentários mais úteis, vez que é praticamente impossível

¹<http://twitter.com>

²<http://facebook.com>

³<http://whatsapp.com>

alguém ler todos os comentários disponíveis. Em uma pesquisa realizada pelo *site Trip Advisor*⁴, observou-se que 80% dos usuários leem de 6 a 12 *reviews* antes de reservarem um hotel e os mesmos estão interessados nos *reviews* mais recentes [6]. Porém, é importante notar que nem sempre esta quantidade *reviews* será realmente a mais significativa no processo de escolha do consumidor.

Assim, diante de tamanha disponibilidade de informações presentes nesses ambientes, o novo consumidor tem uma árdua tarefa de buscar os *reviews* mais significativos para auxiliá-lo no seu processo de escolha, tornando assim uma tarefa desgastante [7]. Diante desse cenário, um consumidor determinado a comprar um telefone celular, por exemplo, pode simplesmente desistir de realizar a compra devido a sobrecarga de informações presentes nesses sistemas *online* de avaliação.

De um lado, as avaliações positivas a respeito de um produto ou serviço trazem ao novo consumidor mais segurança no processo de compra, por outro lado, as avaliações negativas também podem auxiliar na escolha, gerando um impacto positivo na vendas [8]. Geralmente, os *reviews* negativos são escritos de forma mais crítica e apresentando boa legibilidade, algumas vezes, até melhor que os *reviews* positivos. A Figura 1 apresenta um *review* negativo retirado do *site da Amazon*⁵.

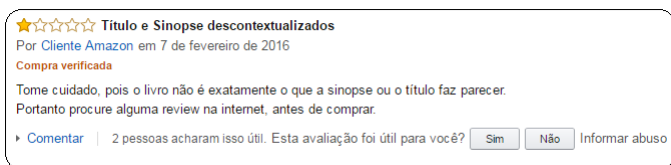


Figura 1. *Review* negativo no *site Amazon*

Na Figura 2 é apresentado um *review* positivo contendo um recurso bastante interessante para avaliar um comentário, conhecido como utilidade do *review*. Com este recurso, o usuário pode, ao terminar de ler, marcar a opção indicando se o comentário foi útil para ele, desta maneira quanto mais votos "Sim" um *review* possuir, melhor ranqueado ele será. Porém, uma desvantagem dessa técnica é que comentários recentes, e que teriam significância ao novo consumidor, podem ser ignorados por terem poucos votos [9].

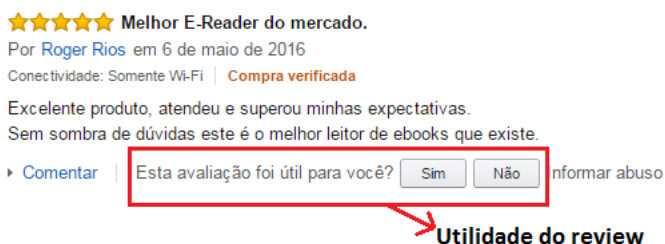


Figura 2. Utilidade do *review* no *site Amazon*

Desde o início de suas operações como *site de e-commerce*, a *Amazon* sempre permitiu a introdução de *reviews* por parte

⁴Disponível em <http://tripadvisor.com>

⁵<http://amazon.com>

dos usuários do serviço (inclusive alguns usuários que leem o comentário no próprio *site da Amazon* podem realizar a compra do produto desejado em outra loja). No entanto, com a introdução desta simples questão junto ao *review*: "Este *review* foi útil para você?", a *Amazon* conseguiu um aumento nos seus lucros de 2,7 bilhões de dólares anuais [10].

A reputação do autor, em geral, implica na confiança que outras pessoas atribuem a um indivíduo em particular baseado nas contribuições intelectuais dentro de um grupo de interesse. Em diversas áreas, quando um autor tem alta credibilidade, as suas opiniões são altamente respeitadas pelos membros ao redor, e podem influenciar o modo de agir e pensar dos outros. Este artigo foca em considerar diversas medidas sobre reputação do autor.

Dentro do contexto de reputação, os sistemas *online* apresentam diversos recursos para avaliação de um comentário, a saber: quantidade de comentários de um usuário, número de favoritos, quantidade de estrelas atribuídas, pontuação positiva e negativa, número de seguidores e amigos, entre outros. A partir de tais recursos um usuário pode considerar a reputação de um certo autor melhor do que a de outro. O valor dessa influência pode ser calculado baseado em recursos disponibilizados por esses sistemas *online*, que é um dos focos deste artigo.

Com o objetivo de identificar os comentários mais relevantes, [11] implementou a abordagem Top(x) para inferir os melhores comentários sobre produtos ou serviços. A abordagem Top(x) utiliza um Sistema *Fuzzy* com três variáveis de entrada: reputação do autor, número de tuplas <característica, palavra opinativa> e riqueza de vocabulário; e uma variável de saída: grau de importância do comentário, representado pela variável "k" (Figura 3). No entanto, para definir a reputação do autor, [11] considerou somente a quantidade de comentários emitidos, ou seja, quanto mais comentários emitidos, melhor a reputação do autor. Destaca-se que essa hipótese é considerada fraca e pode ser refutada pois um *spammer*⁶ é considerado um bom autor. Portanto, este artigo tem como objetivo analisar outras medidas para inferir a reputação de um autor de comentários na Web.

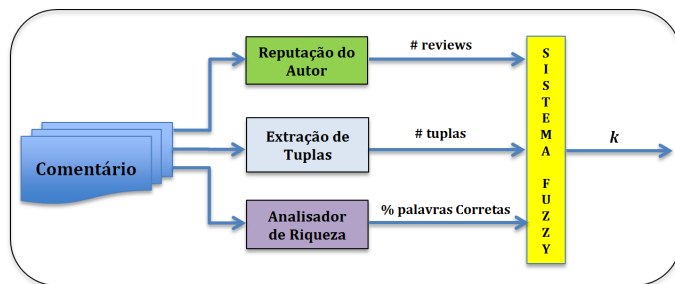


Figura 3. Abordagem TOP(X) proposta por [11]

Visando melhorar a análise da variável reputação do autor, este artigo apresenta um estudo utilizando uma Rede Neural

⁶Usuários que postam muitas propagandas sem a permissão dos demais usuários

Artificial *Multilayer Perceptron* (RNA MLP) para analisar um conjunto de medidas e definir quais são as mais relevantes no processo de avaliação da reputação do autor. Outro aspecto mostrado é uma comparação entre a abordagem Top(x) original com uma nova abordagem utilizando a Top(x) com uma RNA MLP na dimensão da reputação do autor. A hipótese é que aplicando esta RNA MLP com um conjunto de medidas eficazes para a reputação do autor se obtenha um desempenho superior na classificação da importância dos comentários, auxiliando os usuários em suas pesquisas para a aquisição de produtos ou serviços.

O resto deste artigo é organizado da seguinte forma. A Seção 2 apresenta os trabalhos relacionados na área. A Seção 3 explora a coleta e preparação do *Cópus* de pesquisa. A Seção 4 apresenta a abordagem proposta. A Seção 5 apresenta os resultados dos experimentos. Por fim, a Seção 6 apresenta as conclusões e trabalhos futuros.

II. TRABALHOS RELACIONADOS

Diversos autores têm investigado sobre avaliação de reputação de autor na Web, em ambientes *Wiki*, *blogs* e fóruns e em RSOs.

Com relação aos ambientes *Wiki*, eles se caracterizam por permitir a colaboração mútua entre os usuários na confecção de artigos dos mais variados temas. Nestes ambientes, existe uma liberdade muito grande no sentido de permitir a qualquer usuário, cadastrado ou anônimo, colaborar da forma que achar conveniente. Um problema inerente desta liberdade é a possibilidade de se ter artigos de baixa qualidade, especialmente pela atuação de usuários considerados "vândalos".

As principais formas de avaliar a reputação do autor nos ambientes *Wiki* são:

- **Histórico das edições:** baseado nos *edits* que os usuários realizam nos artigos, Wohner et al., em [12], argumentam que contribuições persistentes na *Wiki* duram, em média, 14 dias sem sofrer modificações e eles classificam um autor como bloqueado ou vândalo e regular. Halfaker et al., em [13], definem um juiz para classificar um artigo como aceito ou rejeitado pela comunidade *Wiki* baseado em três características: qualidade dos colaboradores, experiência e no conteúdo postado. Adler et al., em [14], indicam que autores *Wiki* ganham reputação quando seus *edits* são preservados por autores subsequentes e perdem reputação quando seus *edits* são desfeitos em um período curto de tempo. Adler e seus amigos definiram, então, o sistema de reputação *WikiTrust* [15], baseado em três características: qualidade do *edit*, reputação do autor e reputação do conteúdo.
- **Contexto social:** baseado em interesses comuns, Zhao et al., em [16], definiram a *SocialWiki*, um protótipo de sistema *Wiki*, que aproveita o poder das redes sociais para gerenciar automaticamente reputação e confiança para os usuários *Wiki*, baseado no conteúdo que eles contribuem e nas avaliações que eles recebem de outros usuários. Os autores consideram como colaboradores de um artigo, os usuários com interesses em comum.

- **Mecanismos de recompensa:** Hoisl et al., em [17], focaram sobre mecanismos de recompensa social, tais como aceitação, poder e *status*, para ranquear autores que mais colaboram com boas contribuições.

Destaca-se que uma contribuição importante dos trabalhos que exploram o ambiente *Wiki* está na persistência das colaborações, a saber: quanto mais tempo um *edit* persistir, melhor a reputação do autor.

Com relação a rede social *Twitter*, destacam-se os trabalhos para identificar os usuários mais influentes, bem como usuários suspeitos e *spammers*. Kwak et al., em [18], utilizam os dados coletados nos "*trending topics*" (assuntos do momento, em tradução livre) para criar *ranking* dos usuários de acordo com o número de seguidores e o algoritmo de *PageRank*. Eles notaram que esses dois *rankings* são similares. Os autores criaram um terceiro *ranking* baseado nos *retweets*, que é o processo de propagar na rede o *tweet* de outro usuário. Eles concluíram que um *retweet* possui alcance de, no mínimo, 1000 usuários, devido a forma de propagação instantânea e que mais de 85% dos tópicos classificados se referem a manchetes de provedores de conteúdo.

Weitzel et al., em [19], definiram medidas baseadas nos *retweets* para calcular a reputação dentro do *Twitter*, abordando informações no domínio da medicina. Os autores concluíram que a maioria dos perfis no *Twitter* são individuais ou de *blogs* e que a aplicação das medidas baseadas em *retweets* conseguem identificar os usuários mais populares dentro da rede.

Weng et al., em [20], propuseram a medida *TwitterRank*, baseada no número de seguidores e seguidos do usuário. De acordo com a abordagem dos autores, dados três usuários A, B e C, sendo que C segue A e B; se A e B publicam, respectivamente, 500 e 1.000 *tweets* sobre um dado tópico, então, a influência que B exerce sobre C é duas vezes maior que a influência de A. Ainda sobre medidas de ranqueamento, Cappelletti & Sastry, em [21], desenvolveram o algoritmo *IARank*, que observa o potencial que um usuário possui de amplificar uma informação dentro do *Twitter*. Eles consideram dois fatores de entrada no algoritmo: a tendência de um usuário ser retuitado (*retweeted*) ou mencionado (*@mention*) e o tamanho da audiência destes retuites ou menções. Destaca-se que essas duas medidas não substituem o algoritmo *PageRank*, utilizado pelo *Twitter*.

No trabalho desenvolvido por Aggarwal & Kumaraguru [22], os autores identificaram um "mercado negro" que vende/compra contas fraudulentas, curtidas no *Facebook* e até mesmo seguidores no *Twitter* para, artificialmente, melhorarem a reputação social dos usuários. Foram exploradas quatro características com o intuito de detectar usuários suspeitos:

- **Perfil do usuário:** informações do perfil do usuário disponibilizadas pelo *Twitter*;
- **Rede:** descreve o relacionamento do usuário com seus amigos e seguidores;
- **Conteúdo:** natureza dos *tweets* postados pelo usuário;
- **Comportamento do usuário:** define o padrão de postagens, bem como a dinâmica de "seguir" do usuário.

Os autores desenvolveram um mecanismo de aprendizagem de máquina supervisionado, usando as características supracitadas, que é capaz de prever seguidores suspeitos com uma precisão de 88,2%.

No que se refere a detecção de *spammers*, Wang, em [23], definiu reputação do autor como sendo uma relação entre o número de amigos e o número de seguidores. Os resultados obtidos demonstram que o sistema de Wang consegue detectar comportamentos anormais de usuários.

Por fim, no contexto dos *sites* de *e-commerce*, eles criam *rankings* e filtros dos comentários sobre os produtos para auxiliar os consumidores no momento da compra. Os *rankings* podem ser ordenados por data ou número de estrelas. Adicionalmente, podem existir filtros para listar apenas os comentários positivos, negativos, de compradores verificados ou de produtos de uma determinada característica, por exemplo, produto da cor preta.

III. COLETA E PREPARAÇÃO DO CÓRPUS

Para o entendimento geral da proposta e para a avaliação da abordagem, se fez necessária a construção de um *Córpus* de comentários. O processo de coleta dos dados foi realizado nos dias 28 e 29 de setembro de 2016 no *site* do Buscapé⁷. Foram coletadas informações referentes a diversos *smartphones* comercializados em que os usuários realizavam comentários escritos acerca dos mesmos, podendo indicar também a quantidade de estrelas do produto. Inicialmente foram coletados 2433 comentários no total, porém foi realizado um ajuste nos mesmos eliminando os duplicados e vazios para evitar inconsistências. Ao final restaram 2000 *reviews*, sendo 1000 de orientação positiva e 1000 negativa. A orientação semântica dos comentários é definida pelo próprio autor e são apresentados aos leitores de forma separada, através de guias com comentários positivos e negativos. Destaca-se que os trabalhos na área normalmente tem utilizado apenas 2000 comentários nos experimentos e avaliações, devido a dificuldade do processo de anotação do *Córpus*.

No entanto, ao serem analisadas com mais atenção, as orientações dos comentários nem sempre apresentavam a realidade do comentário. Assim, alguns comentários marcados como positivos poderiam ter na verdade orientação negativa ou vice-versa. Além disso, percebeu-se que muitos comentários não apresentavam orientação positiva e nem negativa, sendo neutros com relação ao produto ao qual se destinavam.

Para solucionar esse problema, decidiu-se por criar um *Córpus* anotado com a intenção de avaliar as abordagens. Neste processo, selecionou-se quatro especialistas da área de Análise de Sentimentos para avaliarem os comentários, sendo 1000 comentários para cada par. Eles avaliaram os comentários de acordo com as seguintes orientações: positivo, negativo, neutro ou lixo, que são comentários sem conteúdo. Ao final desta etapa, um quinto especialista ficou responsável por resolver as discrepâncias entre os quatro primeiros. Todos os especialistas são alunos graduados em Ciência da Computação e atuam como pesquisadores na área de PLN.

⁷<http://www.buscape.com.br>

Tabela I
CÓRPUS ANOTADO QUANTO A ORIENTAÇÃO

Orientação	Total
lixo	334
negativo	602
neutro	141
positivo	923

A Tabela I apresenta o *Córpus* anotado após o trabalho dos especialistas. Nota-se que muitos comentários foram marcados como lixo. Isto se deve ao fato de vários usuários postarem as suas colaborações nos *sites* sem possuir o produto ou mesmo postando algo sem o menor sentido.

A. Subcórpus I

Em um primeiro momento, foi realizada a avaliação por seres humanos para definir a reputação do autor dos comentários. Dos 2000 comentários do *Córpus*, separou-se uma amostra para análise e avaliação com 323 comentários, considerando o nível de confiança de 95% e margem de erro de 5%. Adicionou-se 33 comentários, que corresponde a 10% da amostra, totalizando 356, sendo 132 positivos, 131 negativos e 93 neutros, identificada como **Subcórpus I**. Um especialista observou informações referentes ao autor, como: a quantidade de votos positivos em seus comentários, quantidade de votos negativos, total de votos, entre outras medidas, e aplicou uma nota de 0 a 10 para cada um dos autores dos comentários dentro da amostra definida, sendo guiado unicamente pelas variáveis de entrada propostas (Mais detalhes na Seção IV). É importante destacar que um mesmo autor pode ter escrito mais de um comentário no *Córpus* coletado, porém a reputação foi calculada individualmente para o autor do comentário.

Tabela II
RESULTADO DA AVALIAÇÃO REALIZADA POR SERES HUMANOS QUANTO A REPUTAÇÃO DO AUTOR

Reputação	Quantidade
0	68
1	163
2	50
3	23
4	11
5	16
6	9
7	5
8	1
9	5
10	5

A Tabela II mostra o resultado da avaliação realizada por seres humanos quanto a Reputação do Autor dos comentários. As 11 notas aplicadas aos autores dentro da amostra foram generalizadas para o universo completo dos comentários através de uma RNA MLP. Esta generalização se dá pela rede neural que infere, a partir das medidas de entrada, qual a reputação do autor para qualquer comentário dentro do *Córpus*.

B. Subcórpus II

Em um segundo momento, foram selecionados aleatoriamente 271 comentários, sendo 100 positivos, 100 negativos e 71 neutros, para serem avaliados de acordo com a sua importância, identificados como **Subcórpus II**. Foi necessário tal procedimento para poder avaliar a precisão da abordagem original definida por [11] e também da nova abordagem que está sendo apresentada neste artigo. Esse número foi definido tendo em vista a distribuição estatística no Córpus principal em termos das orientações positivas, negativas e neutras, excluindo lixo.

Tabela III
AVALIAÇÃO REALIZADA POR SERES HUMANOS DA IMPORTÂNCIA DOS
COMENTÁRIOS

Importância	Quantidade
excelente	17
bom	24
suficiente	145
insuficiente	85

A Tabela III apresenta o resultado da avaliação realizada por um especialista quanto a importância dos comentários. Para realizar o procedimento de avaliação, o especialista levou em consideração três fatores: a quantidade de informações destacadas no comentário, a riqueza do vocabulário e a reputação do autor, esta última apresentada como uma nota de 0 a 10 para o especialista em cada comentário, conforme a reputação inferida pela RNA MLP.

C. Abordagem Top(x)

A abordagem Top(x) proposta por [11] foi definida para inferir a importância das opiniões de comentários a respeito de produtos e serviços em *sites* de comércio eletrônico. Ela utiliza três variáveis para calcular a importância dos comentários, a saber:

- **Reputação do Autor:** Opiniões de especialistas sobre questões pertencentes a suas áreas de pesquisa têm um peso maior em comparação a pessoas que não têm o mesmo grau de especialização;
- **Quantidade de Tuplas:** Quanto mais características e qualidades forem citadas pelos autores, mais importante é o comentário;
- **Riqueza de Vocabulário:** Quanto mais correto um comentário for escrito, mais útil será a opinião expressa.

Estas três variáveis são dadas como entrada em um Sistema Fuzzy que fornece como saída a importância calculada do comentário, representada por um valor real definido na escala [0..10].

IV. ABORDAGEM PROPOSTA

A abordagem proposta neste artigo visa analisar um conjunto de medidas para definir a reputação do autor de comentários em *sites* de vendas de produtos e serviços. Como descrito anteriormente, a proposta representa uma adaptação da abordagem apresentada por [11]. Destaca-se que uma RNA MLP foi aplicada com o objetivo de inferir a reputação dos

autores dos comentários do Córpus, bem como fornecer a importância de cada medida de entrada. A Figura 4 apresenta a abordagem proposta neste artigo. Na figura, a variável 'a' representa a saída da RNA e indica a reputação do autor normalizada para o intervalo de 0 a 10. Já a variável 'k' indica o grau de importância do comentário, como na abordagem original.

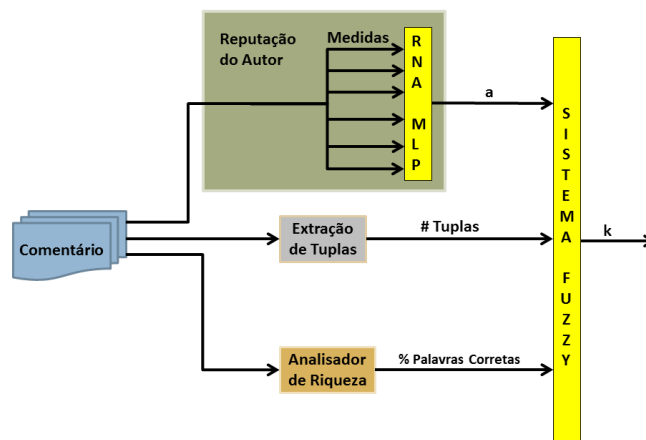


Figura 4. Abordagem proposta

Definiu-se seis medidas para avaliar a reputação do autor dos comentários de produtos no Córpus coletado. Tais medidas foram extraídas levando em conta o máximo de informações que puderam ser aproveitadas para avaliar a reputação do autor. Seguem as medidas definidas abaixo:

- **Data do review:** a data de escrita do *review*, convertida para dias em comparação com a data inicial de coleta do Córpus. Esta informação é importante pois quanto mais recente um *review*, mais atualizado o comentário e, conseqüentemente, deve ser melhor avaliado. No entanto, os *reviews* que são muito recentes podem ser prejudicados no processo de avaliação geral por não ter tempo hábil para leitura pelos consumidores;
- **Data de cadastro no site:** a data em que o autor fez o seu cadastro no *site*, convertida para dias em comparação com a data inicial da coleta do Córpus. Esta informação é importante pois considera-se a reputação dos autores experientes melhor do que de autores novos;
- **Votos positivos:** quantidade de votos positivos atribuídos por outros usuários. Quanto mais votos positivos um autor receber de outros usuários em seus comentários, melhor será a sua reputação;
- **Votos negativos:** quantidade de votos negativos atribuídos por outros usuários. A importância dos votos negativos é inversamente proporcional aos votos positivos, pois quanto mais votos negativos o autor receber em seus comentários, pior a sua reputação;
- **Total de votos:** soma dos votos recebidos para o comentário. De forma geral, hipotetiza-se que quanto mais

votos o usuário tenha em seus comentários, sejam positivos ou negativos, melhor a sua reputação pois o mesmo está sendo observado;

- **Total de reviews do autor do comentário:** quantidade de comentários que o autor realizou no Buscapé. Esta informação é relevante pois indica a participação ativa do usuário dentro do sistema. Destaca-se que [11] aplicou essa informação como única medida para avaliar a reputação do autor.

V. EXPERIMENTOS E RESULTADOS

Os experimentos detalhados nesta Seção são divididos em duas etapas, sendo a primeira com a aplicação de uma RNA MLP sobre o **Subcórpus I**, através da ferramenta SPSS [24]. Na segunda etapa, o experimento utilizou a saída calculada pela RNA MLP da primeira etapa como entrada para variável reputação do autor em um Sistema de Inferência *Fuzzy* baseado no trabalho de [11]. Este Sistema *Fuzzy* teve como objetivo extrair as tuplas e calcular a riqueza do vocabulário dos comentários coletados, tal experimento foi realizado sobre o **Subcórpus II** e utilizou a ferramenta NLTK [25].

A. RNA MLP

Com o objetivo de realizar uma comparação entre a abordagem Top(x) de [11] e a nova abordagem proposta neste artigo, foi executado um experimento aplicando uma RNA MLP para definir a reputação do autor em vez da contagem direta dos comentários. Após este primeiro experimento, um segundo foi realizado calculando a importância dos comentários na abordagem Top(x) original e na nova Top(x) com a RNA MLP. Nesta seção os experimentos são detalhados.

Tabela IV
INFORMAÇÃO DA REDE NEURAL

Camada de Entrada	
Número de neurônios	6
Método de escalabilidade das variáveis	Padronizado
Camada Escondida	
Número de neurônios	8
Função de ativação	Tangente Hiperbólica
Camada de Saída	
Variável dependente	reputacao_manual_int
Número de neurônios	1
Função de ativação	Softmax

A Tabela IV mostra informações referentes a topologia utilizada na Rede Neural proposta no primeiro experimento. A ferramenta utilizada para concepção e execução da rede neural foi o *software* de análises estatísticas SPSS. Na camada de entrada tem-se 6 neurônios, em que cada um representa uma variável de entrada, a saber: **dias_review**, **dias_buscape**, **votos_positivos**, **votos_negativos**, **total_votos** e **total_reviews_autor**. Tais medidas foram consideradas importantes para avaliar a reputação do autor dentro do contexto do Córpus coletado. É importante ressaltar que em RSOs, outras medidas podem ser utilizadas para calcular a reputação do autor, por exemplo, número de seguidores e número de retuites no *Twitter* [18], [20], [22], no entanto para *sites* de

compra e venda elas ainda não estão disponíveis, tornando um grande desafio a tarefa de integrar os perfis de usuários dos *sites* de *e-commerce* e RSOs. Na camada escondida, o melhor ajuste se deu com 8 neurônios e a função de ativação Tangente Hiperbólica. Por fim, a camada de saída utiliza o atributo de supervisão "reputacao_manual_int" como variável dependente para testar a rede, classificando as 11 notas possíveis dos autores (0 a 10) e a função de ativação *Softmax*. É importante relatar que este ajuste é realizado pela ferramenta SPSS, que realiza diversos testes internos para atingir a melhor configuração possível para a performance geral da RNA. O processo de treinamento e teste foi aplicado sobre o **Subcórpus I** utilizando o método de validação cruzada denominado *k-fold*.

Com relação à função de ativação tangente hiperbólica, a mesma assumirá sempre valores reais entre -1 e 1, tendo sua expressão matemática definida pela equação

$$g(\mu) = \frac{1 - e^{-\beta\mu}}{1 + e^{-\beta\mu}} \quad (1)$$

onde β é uma constante real associada ao nível de inclinação da função (parâmetro de inclinação). Já a função *softmax* [26] normaliza entre 0 e 1 as saídas, com o objetivo de definir a probabilidade de uma classe dentro de um problema multi-classes, cuja sua expressão matemática é dada pela equação

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^c e^{x_j}} \quad (2)$$

onde c é o número de camadas escondidas da RNA e x_i é o valor resultante do i -ésimo neurônio da camada escondida. O resultado da equação é interpretado como probabilidades, sendo úteis na classificação pois dá uma medida de certeza sobre as classificações do problema.

Tabela V
IMPORTÂNCIA DAS VARIÁVEIS DE ENTRADA

Variável de entrada	Importância	Importância Normalizada
dias_review	0,106	42,10%
dias_buscape	0,086	34,40%
votos_positivos	0,252	100,00%
votos_negativos	0,205	81,70%
total_votos	0,245	97,60%
total_reviews_autor	0,105	41,80%

A importância de uma variável independente é uma medida de quanto o valor predito pelo modelo da rede muda para diferentes valores da variável independente. O valor da importância de uma variável de entrada é calculado levando em consideração o peso das conexões dos neurônios entre as camadas da RNA. A importância normalizada é simplesmente os valores de importância divididos pelos maiores valores de importância e expressos em percentagens. A Tabela V apresenta a importância de cada variável de entrada na rede neural proposta. É possível observar que a variável mais importante para avaliar a reputação do autor foi "votos_positivos", seguida

de "total_votos" e "votos_negativos". Tal resultado se mostra interessante porque a variável "votos_positivos" é definida por outros usuários da rede, confirmando a boa reputação do autor daquele comentário que recebeu o voto positivo. No entanto, um dos problemas que podem surgir priorizando esta variável é que os *reviews* precisarão de muitos votos positivos para serem lidos deixando de lado os *reviews* mais recentes que não serão lidos por terem poucos ou nenhum voto positivo, mesmo que possuam relevância [9]. Por outro lado, a variável menos importante foi "dias_buscapse" que indica o tempo, em dias, do cadastro do usuário no *site* do Buscapé.

Após o treinamento da rede neural, a reputação inferida foi usada como entrada na abordagem Top(x), substituindo a informação anterior, quantidade de comentários emitidos pelo autor. Com relação a precisão de inferência da RNA MLP dentro do conjunto de treinamento e teste, atingiu-se um valor de **62,08%** no processo de classificação para os valores numéricos de 0 a 10. No entanto, considerando os termos linguísticos usados na abordagem original e descritos a seguir, a precisão da RNA MLP atinge o valor de **91,01%**.

- **Baixo:** 0 a 3;
- **Médio:** 4 a 7;
- **Alto:** 8 a 10.

B. Comparação da abordagem original com a abordagem proposta

Para avaliação das abordagens, calculou-se as medidas de Precisão (P), Cobertura (R) e Medida-f (F) para cada classe. Essas medidas são normalmente usadas em avaliação de abordagens na área de aprendizagem de máquina [27].

É importante destacar que a Precisão (P) é calculada como a porcentagem de exemplos corretamente classificados em cada classe, por meio da seguinte fórmula:

$$P = \frac{TP}{TP + FP} \quad (3)$$

onde TP representa os comentários classificados corretamente (do inglês: true positive) e FP representa os comentários classificados incorretamente (do inglês: false positive). Por exemplo, na Tabela VI a classe EX possui 2 comentários que foram classificados de maneira correta mas 6 de maneira incorreta, sendo a precisão igual a $2/(2+6) = 25\%$.

Já a Cobertura (R) é calculada através da porcentagem de exemplos corretamente classificados em relação ao total de instâncias da classe do *Córpus* anotado, por meio da fórmula:

$$R = \frac{TP}{TP + FN} \quad (4)$$

onde TP representa os comentários classificados corretamente e FN representa os comentários classificados incorretamente em relação ao *Córpus* (do inglês: false negative). Por exemplo, na Tabela VI a classe EX possui 2 comentários classificados de maneira correta mas 9 de maneira incorreta em relação ao *Córpus*, sendo a cobertura igual a $2/(2+9) = 18,18\%$.

Por fim, a Medida-f é uma média harmônica entre Precisão e Cobertura derivada de [28] e calculada conforme a equação:

$$F = \frac{2 * P * R}{P + R} \quad (5)$$

onde P representa o valor da Precisão e R representa o valor da Cobertura. Por exemplo, com a precisão da classe EX igual 25% e a cobertura igual a 18,18% calculadas anteriormente, tem-se a Medida-f igual a 21,05%.

Tabela VI
MATRIZ DE CONFUSÃO: ABORDAGEM TOP(X) ORIGINAL - POSITIVOS

Valor Real	Valor Predito				Total
	EXC	BM	SF	ISF	
EXC	2	4	5	0	11
BM	2	1	9	2	14
SF	4	2	30	8	44
ISF	0	2	23	6	31
Total	8	9	67	16	100

Tabela VII
MATRIZ DE CONFUSÃO: ABORDAGEM TOP(X) COM A RNA MLP - POSITIVOS

Valor Real	Valor Predito				Total
	EXC	BM	SF	ISF	
EXC	6	3	2	0	11
BM	2	4	8	0	14
SF	0	4	32	8	44
ISF	0	0	25	6	31
Total	8	11	67	14	100

Tabela VIII
MATRIZ DE CONFUSÃO: ABORDAGEM TOP(X) ORIGINAL - NEGATIVOS

Valor Real	Valor Predito				Total
	EXC	BM	SF	ISF	
EXC	2	0	2	0	4
BM	2	1	7	0	10
SF	4	5	54	0	63
ISF	2	1	18	2	23
Total	10	7	81	2	100

Tabela IX
MATRIZ DE CONFUSÃO: ABORDAGEM TOP(X) COM A RNA MLP - NEGATIVOS

Valor Real	Valor Predito				Total
	EXC	BM	SF	ISF	
EXC	2	2	0	0	4
BM	2	8	0	0	10
SF	0	9	53	1	63
ISF	0	1	19	3	23
Total	4	20	72	4	100

As Tabelas VI, VII, VIII e IX apresentam as Matrizes de confusão das duas abordagens, tanto para os comentários

positivos quanto para os negativos. Nelas, é possível observar a relação entre os valores preditos pelas duas abordagens e o valor real anotado por um especialista.

Tabela X
COMPARAÇÃO DOS COMENTÁRIOS POSITIVOS NAS DUAS ABORDAGENS

	TOP(X) Positivos		
	Precisão	Cobertura	Medida-F
EX e BM	17,65%	12,00%	14,29%
IF	37,50%	19,35%	25,53%
TODOS	39,00%	39,00%	39,00%
	TOP(X) RNA Positivos		
	Precisão	Cobertura	Medida-F
EX e BM	52,63%	40,00%	45,45%
IF	42,86%	19,35%	26,67%
TODOS	48,00%	48,00%	48,00%

A Tabela X apresenta os resultados das medidas de Precisão, Cobertura e Medida-F para os comentários positivos na abordagem Top(x) e na nova abordagem Top(x) com a RNA MLP, utilizando o **Subcórpus II**. Levando em conta a Medida-F, média harmônica entre Precisão e Cobertura, observa-se que a abordagem Top(x) com a RNA MLP possui desempenho superior na identificação dos comentários excelentes e bons (EX e BM). Já nos comentários insuficientes (IF) a melhora é pouco significativa e, por fim, quando se agrupam todos os comentários há uma melhora razoável. É importante destacar que os comentários suficientes (SF) não foram apresentados, pois, normalmente, os usuários têm interesse nos melhores comentários ou nos piores (comentários mais críticos). Destaca-se ainda que os resultados baixos (Medida-F < 50%) podem ser justificados pela dificuldade do especialista em avaliar precisamente, seja na avaliação da reputação do autor (**Subcórpus I**), em que o especialista deve evitar o enviesamento da amostra para não comprometer a capacidade preditiva da RNA, seja na avaliação da importância dos comentários (**Subcórpus II**), que pode gerar inconsistências no processo de avaliação final das abordagens.

Tabela XI
COMPARAÇÃO DOS COMENTÁRIOS NEGATIVOS NAS DUAS ABORDAGENS

	TOP(X) Negativos		
	Precisão	Cobertura	Medida-F
EX e BM	17,65%	21,43%	19,35%
IF	100,00%	8,70%	16,00%
TODOS	59,00%	59,00%	59,00%
	TOP(X) RNA Negativos		
	Precisão	Cobertura	Medida-F
EX e BM	41,67%	71,43%	52,63%
IF	75,00%	13,04%	22,22%
TODOS	66,00%	66,00%	66,00%

A Tabela XI apresenta os resultados das medidas de Precisão, Cobertura e Medida-F para os comentários negativos na abordagem Top(x) e na nova abordagem Top(x) com a RNA MLP. A abordagem Top(x) com a RNA MLP mostra-se superior em desempenho em todos os casos, especialmente na identificação dos comentários excelentes e bons. Já na identificação dos comentários insuficientes e todos agrupados,

o desempenho é ligeiramente superior ao comparar com a abordagem Top(x) original.

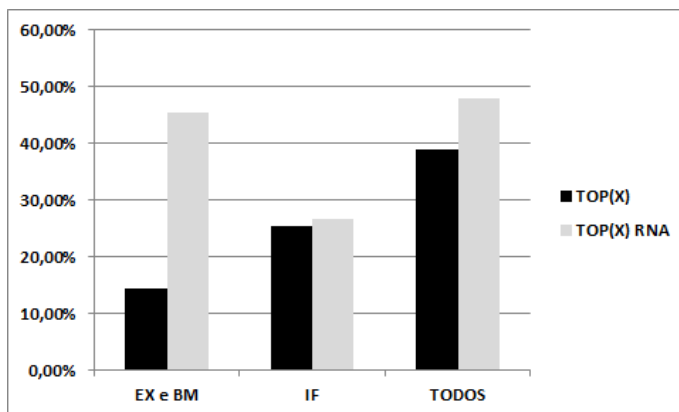


Figura 5. Gráfico comparativo entre os comentários positivos

A Figura 5 apresenta graficamente a comparação entre as duas abordagens relacionando os comentários positivos em termos de sua importância, utilizando a Medida-F. É possível observar que a abordagem Top(x) com a RNA MLP supera com boa margem a abordagem original nos comentários excelentes e bons, estes são relevantes pois o usuário final vai sempre buscar os melhores comentários para ler e decidir sobre a compra de um produto ou serviço. Desta forma, o usuário não vai precisar ler um número grande de comentários e poderá focar em um pequeno grupo de comentários selecionados pela abordagem, gerando um ganho de tempo e esforço na pesquisa pelo produto que deseja adquirir.

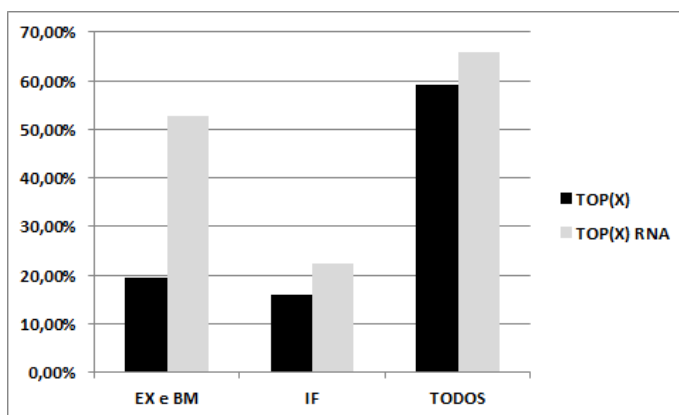


Figura 6. Gráfico comparativo entre os comentários negativos

A Figura 6 apresenta um gráfico comparando as duas abordagens com relação aos comentários negativos e sua respectiva importância, também com a Medida-F. Novamente é possível observar que a abordagem Top(x) com a RNA MLP também supera com boa margem a abordagem original nos comentários excelentes e bons.

VI. CONCLUSÃO

Neste artigo foi apresentada uma nova abordagem para avaliar comentários de produtos na Web baseado na abordagem

Top(x) de [11]. A nova abordagem propõe uma melhoria na dimensão da Reputação do Autor, que na abordagem original considera somente a quantidade de comentários para inferir a sua reputação, podendo beneficiar os *spammers*. Para resolver tal empecilho, este artigo apresentou um conjunto de seis medidas extraídas de um Córpus coletado no *site* Buscapé. O Córpus foi anotado quanto a orientação dos comentários, a importância dos mesmos e da reputação do autor, resultado em um conjunto de 271 comentários de um universo de 2000, denominado **Subcórpus II**.

Segue que, com as medidas extraídas relacionadas a reputação do autor, foi aplicada uma RNA MLP para generalizar os resultados da amostragem para o universo inteiro. Tal procedimento obteve uma precisão de **91,01%** na classificação. O experimento com a RNA MLP se deu sobre 356 comentários, denominado **Subcórpus I**.

Além de analisar as medidas para definir a reputação do autor nos *sites* de *e-commerce*, esta pesquisa examinou um conjunto de medidas utilizadas em outros sistemas web, como: ambientes *Wiki*, redes sociais, fóruns *online* e blogs. O estudo sobre esses sistemas web apresentou diversos recursos para a avaliação de autores que foram aplicados com sucesso e que poderiam ser utilizados em *sites* de *e-commerce*. De fato, este artigo sugere que os portais de comércio eletrônico possam expandir os seus métodos de avaliação de comentários no sentido de reforçar a importância do autor e sua reputação. Uma maneira de realizar essa expansão é permitir a integração dos perfis de usuários dos *sites* com os seus respectivos perfis em redes sociais *online*.

Com relação a relevância das medidas exploradas, esta pesquisa observou que a quantidade de votos positivos que um autor recebe tem um peso significativo em sua reputação, sendo considerada a principal medida para avaliar a reputação do autor no contexto trabalhado. Tal fato demonstra que informações atribuídas por outros usuários possuem mais peso na reputação do que informações intrínsecas ao usuários, como a quantidade de comentários postados pelo mesmo ou a sua data de cadastro no *site*. Porém, um problema inerente ao utilizar os votos para avaliar a reputação é que usuários novos na rede serão prejudicados, neste caso, informações textuais devem ser investigadas em pesquisas futuras.

Por fim, foram calculadas e comparadas as medidas de avaliação das abordagens, a saber, Precisão, Cobertura e Medida-F. Utilizando como base a média harmônica Medida-F, a nova abordagem obteve melhores resultados para classificar a importância dos comentários excelentes e bons, insuficientes e ao agrupar todos os comentários. Com foco nos comentários excelente e bons, a nova abordagem apresentou resultados significativamente superiores, auxiliando assim aos usuários finais na sua pesquisa por produtos ou serviços, reduzindo o tempo e esforço no processo. No entanto, considera-se ainda abaixo do esperado, pois os resultados na classificação ficaram em torno dos 50%.

Como trabalhos futuros espera-se aplicar a nova abordagem em um Córpus maior, realizando um processo mais extenso de avaliação realizada por seres humanos, e explorar medidas

mais eficientes para definir a reputação do autor. Métodos de avaliação automatizado também poderão ser estudados e aplicados. De fato, pesquisadores do nosso grupos estão realizando trabalhos em outras dimensões da abordagem Top(x), por exemplo, explorando melhor a riqueza do vocabulário e também há um trabalho que substituiu o Sistema *Fuzzy* por uma RNA que identifica melhor os comentários mais importantes [29].

REFERÊNCIAS

- [1] SBC, "Grandes desafios da pesquisa em computacao no brasil 2006 - 2016," *Sociedade Brasileira de Computacao, Porto Alegre, RS, Brasil, 2006*, 2006.
- [2] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing, Second Edition*. Taylor and Francis Group, Boca, 2010.
- [3] P. Jackson and I. Moulinier, *Natural language processing for online applications: Text retrieval, extraction and categorization*. Amsterdam: John Benjamins, 2007.
- [4] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Marketing Letters*, vol. 12, no. 3, pp. 211–223, 2001. [Online]. Available: <http://dx.doi.org/10.1023/A:1011122126881>
- [5] Y. A. Kim and J. Srivastava, "Impact of social influence in e-commerce decision making," in *Proceedings of the Ninth International Conference on Electronic Commerce*, ser. ICEC '07. New York, NY, USA: ACM, 2007, pp. 293–302. [Online]. Available: <http://doi.acm.org/10.1145/1282100.1282157>
- [6] TripAdvisor, "Independent study of more than 12,000 global respondents reveals online travel review trends," https://www.tripadvisor.com/PressCenter-i6603-c1-Press_Releases.html, 2013, acessado em 04/04/2016.
- [7] K. D. Vohs, R. F. Baumeister, B. J. Schmeichel, J. M. Twenge, N. M. Nelson, and D. M. Tice, "Making Choices Impairs Subsequent Self-Control: A Limited-Resource Account of Decision Making, Self-Regulation, and Active Initiative," *Journal of Personality and Social Psychology*, vol. 94, no. 5, pp. 883–898, 2008. [Online]. Available: <http://www.apa.org/pubs/journals/releases/psp945883.pdf>
- [8] R. Hamilton, K. D. Vohs, and A. L. McGill, "We'll be honest, this won't be the best article you'll ever read: The use of dispreferred markers in word-of-mouth communication," *Journal of Consumer Research*, 2014.
- [9] M. X. Li, L. Huang, C. H. Tan, and K. K. Wei, "Helpfulness of online product reviews as seen by consumers: Source and content features," *Int. J. Electronic Commerce*, 2013.
- [10] J. Spool, "The magic behind amazons 2.7 billion dollar question," <https://articles.ue.com/magicbehindamazon/>, 2009, acessado em 05/04/2016.
- [11] R. F. de Sousa, R. A. L. Rabelo, and R. S. Moura, "A fuzzy system-based approach to estimate the importance of online customer reviews," in *Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on*, Aug 2015, pp. 1–8.
- [12] T. Wöhner, S. Köhler, and R. Peters, "Automatic reputation assessment in wikipedia," *Thirty Second International Conference on Information Systems, Shanghai 2011*, 2011.
- [13] A. Halfaker, A. Kittur, R. Kraut, and J. Riedl, "A jury of your peers: Quality, experience and ownership in wikipedia," in *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*. New York, NY, USA: ACM, 2009.
- [14] B. T. Adler and L. de Alfaro, "A content-driven reputation system for the wikipedia," in *Proceedings of the 16th International Conference on World Wide Web*. New York, NY, USA: ACM, 2007, pp. 261–270.
- [15] B. T. Adler, L. de Alfaro, and I. Pye, "Detecting wikipedia vandalism using wikitrust," in *CLEF 2010 LABs and Workshops, Notebook Papers*, 2010.
- [16] H. Zhao, S. Ye, P. Bhattacharyya, J. Rowe, K. Gribble, and S. F. Wu, "Socialwiki: Bring order to wiki systems with social context," in *Social Informatics - Second International Conference, SocInfo 2010, Laxenburg, Austria, October 27-29, 2010. Proceedings*, 2010.
- [17] B. Hoisl, W. Aigner, and S. Miksch, *Online Communities and Social Computing: Second International Conference, OCSC 2007, Held as Part of HCI International 2007, Beijing, China, July 22-27, 2007*.

- Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, ch. Social Rewarding in Wiki Systems – Motivating the Community.
- [18] H. Kwak, C. Lee, H. Park, and S. Moon, “What is twitter, a social network or a news media?” in *Proceedings of the 19th International Conference on World Wide Web*. New York, NY, USA: ACM, 2010.
 - [19] L. Weitzel, J. P. M. d. Oliveira, and P. Quaresma, “Measuring the reputation in user-generated-content systems based on health information,” *Procedia Computer Science*, 2014, 2014 International Conference on Computational Science.
 - [20] J. Weng, E. P. Lim, J. Jiang, and Q. He, “Twitterrank: Finding topic-sensitive influential twitterers,” in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*. New York, NY, USA: ACM, 2010.
 - [21] R. Cappelletti and N. Sastry, “Iarank: Ranking users on twitter in near real-time, based on their information amplification potential.” in *SocialInformatics*. IEEE Computer Society, 2012.
 - [22] A. Aggarwal and P. Kumaraguru, “Followers or phantoms? an anatomy of purchased twitter followers,” *CoRR*, 2014.
 - [23] A. H. Wang, “Don’t follow me: Spam detection in twitter,” in *Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on*, July 2010, pp. 1–10.
 - [24] SPSS, “Spss for windows, version 16.0,” Chicago USA, 2007.
 - [25] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. O’Reilly Media, Inc., 2012.
 - [26] J. S. Bridle, “Advances in neural information processing systems 2,” D. S. Touretzky, Ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990, ch. Training Stochastic Model Recognition Algorithms As Networks Can Lead to Maximum Mutual Information Estimation of Parameters.
 - [27] D. M. W. Powers, “Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation,” *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
 - [28] C. J. V. Rijsbergen, *Information Retrieval*, 2nd ed. Newton, MA, USA: Butterworth-Heinemann, 1979.
 - [29] R. L. de S. Santos, R. F. de Sousa, R. A. L. Rabelo, and R. S. Moura, “An experimental study based on fuzzy systems and artificial neural networks to estimate the importance of reviews about product and services,” in *2016 International Joint Conference on Neural Networks (IJCNN)*, July 2016.