

Estudo sobre Métricas para Definir Reputação do Autor de Comentários em Sites de Vendas de Produtos

Carlos Augusto de Sá¹, Raimundo Santos Moura¹

¹ Universidade Federal do Piauí (UFPI)
Departamento de Computação – Teresina, PI – Brasil

{carlos.sa, rsm}@ufpi.edu.br

Abstract. *Knowing the author's reputation for opinionated texts is of utmost importance for evaluating a comment on the Web. This paper presents a study on measures used in the process of evaluating the author's reputation on product sales sites. Two experiments were carried out with neural networks Multilayer Perceptron (MLP) and Radial Basis Function (RBF), and the results show that the MLP gave slightly better performance, but not significantly so. In addition, an experiment was carried out to compare the TOP(X) approach, which is used to infer the best comments, with the new approach that uses MLP in the author's reputation dimension. The results showed that the new approach obtained a gain in the classification of the importance of the comments.*

Resumo. *Conhecer a reputação do autor de textos opinativos é de suma importância para avaliação de comentários na Web. Este artigo apresenta um estudo sobre medidas usadas no processo de avaliação da reputação do autor em sites de vendas de produtos. Realizou-se dois experimentos com as redes neurais Multilayer Perceptron (MLP) e Radial Basis Function (RBF), sendo que a rede MLP obteve melhor desempenho. Comparou-se também a abordagem TOP(X) original, usada para inferir os melhores comentários, com um novo modelo que utiliza rede MLP na dimensão da reputação do autor. Considerando os comentários excelentes e bons, a nova abordagem apresentou resultados significativamente superiores.*

1. Introdução

As Redes Sociais Online (RSO) Twitter e Facebook, e serviços de troca de mensagens pelo celular como o Whatsapp, permitem interação social entre seus usuários. Além disso, sites de comércio eletrônico (*e-commerce*) permitem aos usuários deixar suas opiniões e comentários sobre o processo de aquisição de um produto ou realização de um serviço. Com o crescimento do volume de informações disponíveis e com o avanço da Computação, a área de Processamento de Linguagem Natural (PLN) ganhou bastante destaque por realizar análises de dados de maneira mais eficiente.

A Análise de Sentimentos ou Mineração de Opinião é uma subárea de PLN, que envolve Ciência da Computação, Linguística e Inteligência Artificial e tem atacado o problema de manipular grandes volumes de dados através de técnicas que analisam a linguagem escrita ou falada [Jackson and Moulinier 2007]. 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. Outro detalhe a ser considerado é a quantidade de *spam*, textos de baixa qualidade, erros ortográficos, *emoticons*, "internetês"¹ e informações falsas [Liu 2011]. Além dessas considerações, algumas vezes os comentários apresentam sarcasmos e ironias, que são difíceis de serem captados pelas técnicas atuais de PLN. No entanto, existem esforços no sentido de resolver esses problemas, como os trabalhos de [Hartmann et al. 2014, Carvalho et al. 2009, Gonçalves et al. 2015].

Destaca-se que avaliações positivas a respeito de um produto ou serviço trazem ao novo consumidor mais segurança no processo de compra, porém, avaliações negativas também podem auxiliar na escolha, gerando um impacto positivo nas vendas [Hamilton et al. 2014]. Geralmente, comentários negativos são escritos de forma mais crítica e apresentam boa legibilidade, sendo, algumas vezes, melhores do que comentários positivos. A Figura 1 apresenta um comentário negativo retirado do site da Amazon². Na parte inferior da figura, enfatiza-se 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 comentário possuir, melhor classificado ele será. Porém, uma desvantagem dessa medida é que comentários recentes e com alta significância ao consumidor, são ignorados por terem poucos votos [Li et al. 2013].



Figura 1. Review negativo no site Amazon

No contexto de reputação em RSOs, os sistemas apresentam diversos recursos para avaliação de um comentário, a saber: quantidade de comentários emitidos por um usuário, número de favoritos, quantidade de estrelas atribuídas, pontuação positiva e negativa, número de seguidores e amigos, entre outros. Um problema que os sites de *e-commerce* apresentam é o fato de possuírem cadastros independentes para os seus usuários e o acesso a esses dados não ser permitido, dificultando a coleta de informações. Desta forma, esta pesquisa sugere que os sites de *e-commerce* possam rever as suas políticas de privacidade no futuro. Uma tentativa de solucionar o entrave é ligar o perfil dos usuários com a suas contas em RSOs populares como Twitter e Facebook.

As RSOs levam vantagem sobre os sites de comentários no que se refere as informações sobre os autores das postagens. O Facebook, por exemplo, permite que os participantes possam curtir um conteúdo ou mesmo comentar. Destaca-se que as empresas tentam explorar ao máximo esse novo tipo *marketing*, o que se confirma com a grande quantidade de perfis nesta rede social.

De maneira geral, com o objetivo de identificar os comentários mais relevantes, em [de Sousa et al. 2015] os autores propuseram uma abordagem para inferir os melho-

¹Neologismo (palavra: Internet + sufixo: ês) que designa a linguagem utilizada no meio virtual.

²<http://amazon.com>

res comentários sobre produtos ou serviços, denominada TOP(X), que 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" (ver Figura 2). No entanto, para definir a reputação do autor, os autores consideraram somente a quantidade de comentários publicados, ou seja, quanto mais comentários emitidos, melhor a reputação do autor. Destaca-se que essa hipótese é fraca e pode ser facilmente refutada pois um *spammer*³ será considerado um bom autor.

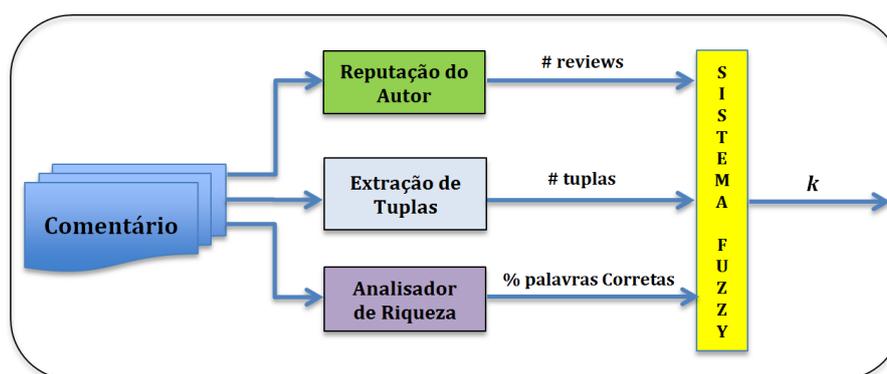


Figura 2. Abordagem TOP(X) proposta por [de Sousa et al. 2015]

Visando melhorar a análise da variável reputação do autor, este trabalho apresenta um estudo utilizando Rede Neural Artificial (RNA) para analisar um conjunto de medidas e definir quais são as mais relevantes no processo de avaliação. Outro aspecto mostrado é uma comparação entre a abordagem TOP(X) original com uma nova abordagem que utiliza RNA MLP na dimensão da reputação do autor. A hipótese é que o uso de uma RNA para inferir a reputação do autor no modelo original melhora o desempenho na classificação da importância dos comentários.

O restante deste trabalho está organizado de seguinte maneira: a Seção 2 apresenta alguns trabalhos relacionados com reputação de autor. A Seção 3 descreve a abordagem proposta para analisar o conjunto de medidas sobre reputação do autor, usando uma RNA. A Seção 4 explora a coleta e a preparação do *Córpus* utilizado na pesquisa. A Seção 5 apresenta e discute os resultados dos experimentos realizados. Por fim, a Seção 6 destaca as principais contribuições e trabalhos futuros.

2. Trabalhos Relacionados

Diversos autores têm investigado sobre avaliação de reputação de autor na Web e nas redes sociais, com destaque para os ambientes *Wiki* e para o *Twitter*. Os ambientes *Wiki* se caracterizam por permitir a colaboração mútua entre os usuários na produção de artigos dos mais variados temas. Um problema inerente desta liberdade é a possibilidade de se ter artigos de baixa qualidade, especialmente pela atuação de vândalos⁴. As principais formas de avaliar a reputação do autor nos ambientes *Wiki* são:

³Usuários que postam muitas propagandas sem a permissão dos demais usuários

⁴Usuários que editam os artigos com informações fora do contexto.

- **Histórico das edições:** os autores utilizam o histórico das páginas em busca de padrões de *edits* por parte dos usuários. [Wöhner et al. 2011] destacam que as contribuições persistentes de usuários na Wikipedia duram em média 14 dias sem sofrer modificações. Eles classificam os usuários como autores *vândalos* ou *regulares*. [Halfaker et al. 2009] 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 and de Alfaro 2007] 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* [Adler et al. 2010] baseado em três características: qualidade do *edit*, reputação do autor e reputação do conteúdo.
- **Contexto social:** [Zhao et al. 2010] 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, porém eles não descreveram a fórmula para calcular a reputação.
- **Mecanismos de recompensa:** [Hoisl et al. 2007] focaram sobre mecanismos de recompensa social, tais como aceitação, poder e *status*, para ranquear autores que mais colaboram com boas contribuições. Os autores concluíram que a abordagem de recompensas baseada em motivação pode produzir artigos de alta qualidade.

É importante destacar que a contribuição dos trabalhos que exploram o ambiente *Wiki* está na persistência das colaborações, ou seja, *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, usuários suspeitos e *spammers*. [Kwak et al. 2010], 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 *Page-Rank*. 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. 2014] 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. 2010] 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 and Sastry 2012] desenvolveram o algoritmo *IARank*, que observa o poten-

cial que um usuário possui de ampliar uma informação dentro do *Twitter*. Eles consideram dois fatores de entrada no algoritmo: a tendência de um usuário ser retuitado ou mencionado e o tamanho da audiência desses retuitos ou menções. Destaca-se que essas duas medidas não substituem o algoritmo *PageRank*, utilizado pelo *Twitter*.

No trabalho desenvolvido por [Aggarwal and Kumaraguru 2014], 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. Os autores relatam uma precisão de 88,2% no mecanismo de aprendizagem de máquina supervisionado usado para predizer seguidores suspeitos.

No que se refere a detecção de *spammers*, [Wang 2010] 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*, existem soluções que 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 azul.

Este artigo investiga se a reputação do autor de comentários Web pode ser calculada a partir de seis medidas, abordadas com mais detalhes na Seção 3.

3. Abordagem Proposta

A abordagem proposta neste trabalho visa analisar um conjunto de medidas para definir a reputação do autor de comentários em sites de vendas de produtos. De forma geral, a proposta representa uma adaptação da abordagem Top(X) original, com ênfase na dimensão reputação do autor. O estudo foi conduzido através da aplicação de redes neurais artificiais para inferir a reputação dos autores dos comentários e descobrir a importância de cada medida da entrada.

A Figura 3 mostra a visão geral da abordagem proposta, considerando apenas a dimensão reputação do autor. Na figura, a variável 'a' representa a saída da RNA e indica a reputação do autor normalizada para o intervalo de 0 a 10.

Considerando sites de *e-commerce*, definiu-se seis medidas para avaliar a reputação do autor dos comentários de produtos. Tais medidas foram extraídas levando em conta informações disponíveis em sites de lojas virtuais e comparadores de preços. Nossa pesquisa examinou também outras medidas utilizadas em ambientes Wiki, RSOs, Fóruns e Blogs. O estudo realizado sugere, então, que os sites de *e-commerce* devem expandir suas funcionalidades no sentido de reforçar a importância dos autores e seus relacionamentos. Uma maneira de realizar essa expansão é permitir a integração dos perfis de usuários dos sites com os seus respectivos perfis em redes sociais.

- **DataReview:** a data de escrita do comentário, convertida para dias em comparação com a data inicial de coleta do *Córpus*. Esta informação é importante pois quanto mais recente, mais atualizado o comentário e, hipoteticamente,

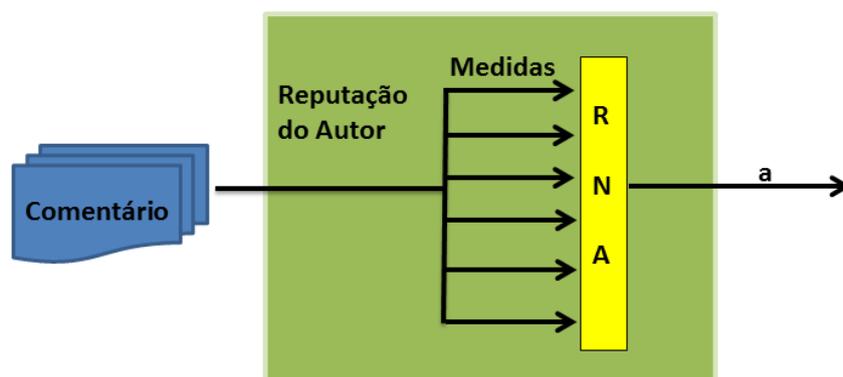


Figura 3. Abordagem proposta

deve ser melhor avaliado. No entanto, os comentários que são muito recentes podem ser prejudicados no processo de avaliação geral por não ter tempo hábil para leitura pelos consumidores;

- **DataCadastro:** 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 imagina-se que a reputação dos autores experientes seja melhor do que a autores novatos;
- **VotosPositivos:** quantidade de votos positivos atribuídos por outros usuários. A hipótese é que quanto mais votos positivos um autor receber de outros usuários, melhor será a sua reputação;
- **VotosNegativos:** 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;
- **TotalVotos:** soma dos votos recebidos pelo comentário. De forma geral, imagina-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;
- **TotalReviewsAutor:** quantidade de comentários que o autor realizou no site. Esta informação é relevante pois indica a participação ativa do usuário dentro do ambiente.

4. *Córpus*: coleta e preparação⁵

Para avaliação da abordagem proposta, foi criado um *Córpus* com 2.433 comentários do site do Buscapé⁶. A decisão de trabalhar com comentários do site do Buscapé deu-se por três razões principais: i) ser o maior site comparador de preços da América Latina; ii) necessidade posterior de comparar o modelo proposto com a abordagem Top(x), que utiliza comentários sobre *smartphones* do referido site; e iii) os comentários serem disponíveis publicamente para coleta com rastreadores Web.

Os dados foram coletados no ano de 2016 e são referentes a comentários escritos em português sobre diversos *smartphones*. Após a exclusão de comentários duplicados e

⁵Córpus disponível em <https://goo.gl/g5nrwJ>

⁶<http://www.buscape.com.br>

vazios, definiu-se uma amostra de 2.000 comentários, sendo 1.000 de orientação positiva e 1.000 negativa. Destaca-se que no site do Buscapé, a orientação do comentário é definida pelo próprio autor e, são apresentados aos usuários em guias/abas separadas. No entanto, em uma análise mais detalhada, verificou-se que alguns comentários marcados como positivos eram, na verdade, negativos e vice-versa. Além disso, muitos comentários são considerados neutros.

Para solucionar esse problema, decidiu-se fazer uma análise manual do *Córpus* quanto à orientação semântica. Ao final do processo, o *Córpus* anotado ficou com 923 comentários positivos, 602 comentários negativos, 141 comentários neutros e 334 considerados "lixo", que são comentários de usuários que declaram "não possuir o produto" e comentários totalmente sem sentido. Esses comentários foram desconsiderados nas nossas avaliações.

Em um segundo momento, criou-se um *Subcórpus* anotado com a reputação do autor, analisando uma amostra de 323 comentários (nível de confiança de 95% e margem de erro de 5%). Adicionou-se 33 comentários ao *Subcórpus* (10% da amostra), totalizando 356, sendo 132 positivos, 131 negativos e 93 neutros. A anotação foi realizada por três alunos da pós-graduação em Ciência da Computação da UFPI, considerando informações referentes ao autor, como a quantidade de votos positivos em seus comentários, quantidade de votos negativos, total de votos, entre outras medidas. Em seguida, aplicou-se 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.

A Tabela 1 mostra o resultado da avaliação dos seres humanos para a reputação dos autores dos comentários. As 11 notas atribuídas aos autores foram generalizadas para o universo completo dos comentários através de uma RNA, como será descrito na próxima seção. 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*.

Tabela 1. Resultado da anotação do *Subcórpus* por seres humanos

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

É importante mencionar que a anotação manual de *Córpus* deu-se devido à ausência de recursos linguísticos para a língua portuguesa disponíveis publicamente. Sabe-se que o processo de anotação manual pode causar um viés no modelo proposto e, possivelmente, comprometer a viabilidade da solução. Porém, os riscos da anotação foram minimizados com o envolvimento de especialistas da área de linguística computacional.

5. Experimentos

Para avaliar o modelo de RNA proposto, realizou-se dois experimentos com as arquiteturas *Multilayer Perceptron* (MLP) e *Radial Basis Functions* (RBF), usando a ferramenta de análises estatísticas SPSS sobre o *Subcórpus* anotado. É importante mencionar que

nos dois experimentos foram usados 6 neurônios na camada de entrada da rede, correspondentes às medidas discutidas na seção anterior. Para comparar a abordagem TOP(X) original com a nova abordagem que usa RNA na dimensão da reputação do autor, fez-se um terceiro experimento discutido a seguir.

5.1. Experimento 1: RNA MLP

No primeiro experimento usamos uma RNA MLP e o melhor ajuste se deu com 8 neurônios na camada escondida e a função de ativação Tangente Hiperbólica. Na camada de saída utilizamos o atributo de supervisão "ReputacaoManual" 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 o ajuste é realizado pela ferramenta SPSS, através de diversos testes internos para atingir a melhor configuração possível para a performance geral da rede.

A Figura 4 mostra a disposição dos neurônios em cada camada na melhor topologia escolhida pela execução da RNA MLP. O processo de treinamento e teste foi aplicado sobre o *Subcórpus* utilizando o método de validação cruzada *10-fold cross validation*.

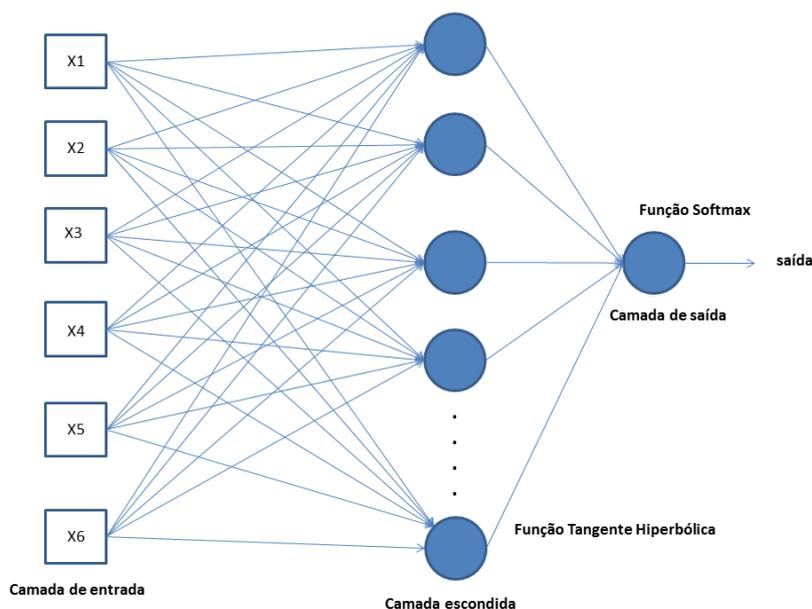


Figura 4. Topologia da RNA MLP

Em redes neurais, o valor da importância de uma variável de entrada é calculado levando em consideração o peso das conexões dos neurônios entre as camadas da rede. Já a importância normalizada é simplesmente os valores de importância divididos pelos maiores valores de importância e expressos em porcentagens.

A Tabela 2 apresenta a importância de cada variável de entrada na rede MLP. É possível observar que a variável mais importante para avaliar a reputação do autor foi "VotosPositivos", seguida de "TotalVotos" e "VotosNegativos". Tal resultado se mostra interessante porque a variável "VotosPositivos" é definida por outros usuários da rede, confirmando a boa reputação do autor daquele comentário que recebeu o voto positivo.

Por outro lado, a variável menos importante foi "DataCadastro" que indica o tempo, em dias, do cadastro do usuário no site.

Tabela 2. Importância das variáveis de entrada na RNA MLP

Variável de Entrada	Importância	Importância Normalizada
DataReview	0,106	42,10%
DataCadastro	0,086	34,40%
VotosPositivos	0,252	100,0 %
VotosNegativos	0,205	81,70%
TotalVotos	0,245	97,60%
TotalReviewsAutor	0,105	41,80%

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 as faixas de valores: 0-3 para **baixo**, 4-7 para **médio** e 8-10 para **alto**, que são normalmente usados em sistemas de reputação, a precisão da rede atingiu o valor de 91,01%.

5.2. Experimento 2: RNA RBF

Visando apresentar uma alternativa para a arquitetura MLP, executou-se o segundo experimento com uma RNA com funções de base radial. O melhor ajuste se deu com 11 neurônios na camada escondida e a função de ativação Softmax. Na camada de saída utilizou-se também o atributo de supervisão "ReputacaoManual" como variável dependente para testar a rede, classificando as 11 notas possíveis dos autores (0 a 10) e a função de ativação Identidade. Assim como na rede MLP, ressalta-se que o ajuste da rede RDF foi realizado usando a ferramenta SPSS. A topologia com a disposição dos neurônios é semelhante a topologia da rede MLP, com a diferença apenas da quantidade de neurônios na camada escondida e das funções de ativação usadas nas camadas escondidas e de saída. O processo de treinamento e teste utilizou o mesmo *Subcórpus* e o mesmo método de validação *10-fold cross validation*.

A Tabela 3 apresenta a importância de cada variável de entrada usando a rede RBF. Observa-se que os resultados são similares aos apresentados no experimento com a RNA MLP, porém existe uma mudança na ordem das variáveis mais importantes, sendo "TotalVotos" a mais importante, seguida de "VotosPositivos" e "VotosNegativos". Por outro lado, a variável menos importante foi "DataCadastro", assim como na rede MLP.

Tabela 3. Importância das variáveis de entrada na RNA RBF

Variável de Entrada	Importância	Importância Normalizada
DataReview	0,127	60,30%
DataCadastro	0,111	52,80%
VotosPositivos	0,210	99,60 %
VotosNegativos	0,210	99,50%
TotalVotos	0,211	100,0%
TotalReviewsAutor	0,132	62,60%

Com relação a precisão de inferência da rede dentro do conjunto de treinamento e teste, atingiu-se um valor de 52,25% no processo de classificação para os valores numéricos de 0 a 10 e 87,36% para as três faixas de valores baixo, médio e alto. Desta forma, devido a vantagem da rede MLP sobre a rede RBF, decidiu-se usar apenas a primeira arquitetura nos experimentos de comparação entre a abordagem TOP(X) original e a nova abordagem que utiliza uma RNA na dimensão da reputação do autor.

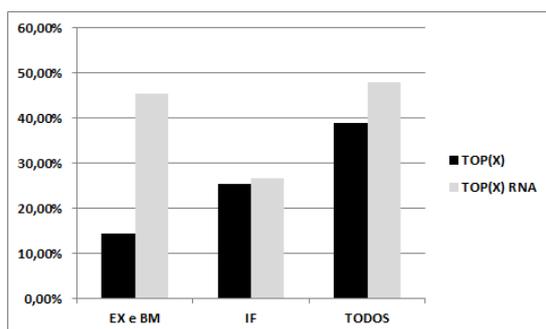
5.3. Experimento 3: Comparação entre as abordagens

Neste experimento 271 comentários foram selecionados aleatoriamente, sendo 100 positivos, 100 negativos e 71 neutros. Com relação a importância dos comentários, a amostra utilizada possui a seguinte anotação: 17 excelentes (EX), 24 bons (BM), 145 suficientes (SF) e 85 insuficientes (IF), conforme classificação proposta em [de Sousa et al. 2015].

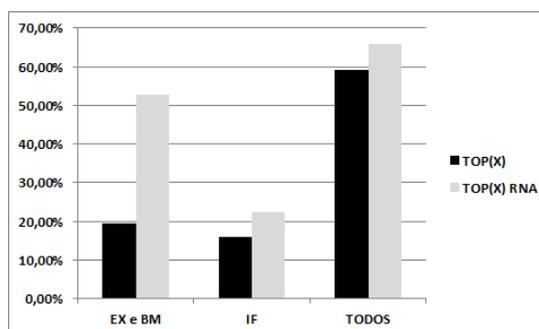
Para avaliar as abordagens, calculou-se as medidas de Precisão, Cobertura e a medida harmônica Medida-F para cada classe. Essas medidas são normalmente usadas em avaliação de abordagens na área de aprendizagem de máquina [Powers 2011].

A Figura 5(a) apresenta graficamente a comparação baseada na Medida-F entre as duas abordagens, relacionando os comentários positivos em termos de sua importância. É possível observar que abordagem TOP(X) com RNA supera com boa margem a abordagem TOP(X) original nos comentários excelentes e bons. Estes comentários são relevantes pois, normalmente, o usuário procura os melhores comentários para ler e decidir sobre a compra de um produto ou serviço. Desta forma, o usuário 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.

A Figura 5(b) apresenta o gráfico comparativo com relação aos comentários negativos e sua respectiva importância, também baseada na Medida-F. Novamente é possível observar que abordagem TOP(X) com RNA também supera com boa margem a abordagem original nos comentários excelentes e bons.



(a) Comentários positivos



(b) Comentários negativos

6. Conclusão e Trabalhos Futuros

Neste artigo foi apresentado um estudo sobre métricas para definir a reputação do autor de comentários em sites de comparação de preços de produtos. De forma geral, o modelo proposto representa uma adaptação da abordagem TOP(X) [de Sousa et al. 2015], com ênfase na dimensão reputação do autor. O estudo foi conduzido através da aplicação

de redes neurais para inferir a reputação dos autores dos comentários e descobrir a importância de cada medida de entrada.

Realizou-se dois experimentos com a aplicação de RNAs MLP e RBF sobre um *Subcórpus* para realizar o treinamento da rede. Os resultados obtidos apresentaram similaridades entre as redes quanto a indicação da importância das variáveis de entrada. Observou-se 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 analisado. Com relação ao desempenho dos modelos, a rede RBF atingiu 87,36% de precisão, enquanto que a rede MLP atingiu 91,01%. Portanto, a rede MLP foi escolhida para a evolução de nossas pesquisas.

Em um terceiro experimento comparou-se as abordagens TOP(X) original e TOP(X) com RNA MLP, utilizando como base a média harmônica Medida-F. Com foco nos comentários excelente e bons, a nova abordagem apresentou resultados significativamente superiores. Conclui-se, então, que tal abordagem pode auxiliar os usuários na busca por produtos ou serviços, reduzindo o tempo e esforço gastos no processo. No entanto, considera-se ainda abaixo do esperado, pois os resultados na classificação ficaram em torno de 50%.

Como trabalhos futuros, pretende-se: i) aplicar a nova abordagem em um *Cópus* maior, realizando um processo mais extenso de anotação manual; e ii) investigar o impacto de reputação do autor em notícias falsas (*fake news*). Sabe-se que existem vários artifícios utilizados para potencializar o alcance de uma notícia ou comentário, bem como impulsionar a reputação de um autor.

Referências

- [Adler and de Alfaro 2007] Adler, B. T. and de Alfaro, L. (2007). A content-driven reputation system for the wikipedia. In *Proc. of the Int. Conference on World Wide Web*, pages 261–270. ACM.
- [Adler et al. 2010] Adler, B. T., de Alfaro, L., and Pye, I. (2010). Detecting wikipedia vandalism using wikiptrust.
- [Aggarwal and Kumaraguru 2014] Aggarwal, A. and Kumaraguru, P. (2014). Followers or phantoms? an anatomy of purchased twitter followers. *CoRR*.
- [Cappelletti and Sastry 2012] Cappelletti, R. and Sastry, N. (2012). IARank: Ranking users on twitter in near real-time, based on their information amplification potential. In *SocialInformatics*, pages 70–77.
- [Carvalho et al. 2009] Carvalho, P., Sarmiento, L., Silva, M. J., and de Oliveira, E. (2009). Clues for detecting irony in user-generated contents: Oh...!! it's "so easy";-). In *Proc. of the Int. Workshop on Topic-sentiment Analysis for Mass Opinion*, pages 53–56.
- [de Sousa et al. 2015] de Sousa, R. F., Rabelo, R. A. L., and Moura, R. S. (2015). A fuzzy system-based approach to estimate the importance of online customer reviews. In *International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8.
- [Gonçalves et al. 2015] Gonçalves, P., Dalip, D., Reis, J., Messias, J., Ribeiro, F., Melo, P., Araújo, L., Gonçalves, M., and Benevenuto, F. (2015). Bazinga! caracterizando e

- detectando sarcasmo e ironia no twitter. In *Proc. of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*.
- [Halfaker et al. 2009] Halfaker, A., Kittur, A., Kraut, R., and Riedl, J. (2009). A jury of your peers: Quality, experience and ownership in wikipedia. In *Proc. of the Int. Symposium on Wikis and Open Collaboration*, pages 15:1–15:10. ACM.
- [Hamilton et al. 2014] Hamilton, R., Vohs, K. D., and McGill, A. L. (2014). 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*, 41(1):197 – 212.
- [Hartmann et al. 2014] Hartmann, N., Avanço, L., Balage, P., Duran, M., Nunes, M. D. G. V., Pardo, T., and Aluísio, S. (2014). A large corpus of product reviews in portuguese: Tackling out-of-vocabulary words. In *Proc. of the Int. Conference on Language Resources and Evaluation (LREC’14)*.
- [Hoisl et al. 2007] Hoisl, B., Aigner, W., and Miksch, S. (2007). *Online Communities and Social Computing: Second International Conference, OCSC 2007*, pages 362–371. Springer Berlin Heidelberg.
- [Jackson and Moulinier 2007] Jackson, P. and Moulinier, I. (2007). *Natural language processing for online applications: Text retrieval, extraction and categorization*. John Benjamins, Amsterdam.
- [Kwak et al. 2010] Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? In *Proc. of the Int. Conference on World Wide Web*, pages 591–600. ACM.
- [Li et al. 2013] Li, M., Huang, L., Tan, C., and Wei, K. (2013). Helpfulness of online product reviews as seen by consumers: Source and content features. *Int. J. Electronic Commerce*, 17(4):101–136.
- [Liu 2011] Liu, B. (2011). *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- [Powers 2011] Powers, D. M. W. (2011). Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1):37–63.
- [Wang 2010] Wang, A. H. (2010). Don’t follow me: Spam detection in twitter. In *Proc. of the Int. Conference on Security and Cryptography (SECRYPT)*, pages 1–10.
- [Weitzel et al. 2014] Weitzel, L., de Oliveira, J. P. M., and Quaresma, P. (2014). Measuring the reputation in user-generated-content systems based on health information. *Procedia Computer Science*, 29:364 – 378.
- [Weng et al. 2010] Weng, J., Lim, E.-P., Jiang, J., and He, Q. (2010). Twitterrank: Finding topic-sensitive influential twitterers. In *Proc. of the Int. Conference on Web Search and Data Mining*, pages 261–270. ACM.
- [Wöhner et al. 2011] Wöhner, T., Köhler, S., and Peters, R. (2011). Automatic reputation assessment in wikipedia. In *Proc. of the Int. Conference on Information Systems*.
- [Zhao et al. 2010] Zhao, H., Ye, S., Bhattacharyya, P., Rowe, J., Gribble, K., and Wu, S. F. (2010). Socialwiki: Bring order to wiki systems with social context. In *Social Informatics - Second International Conference, SocInfo*, pages 232–247.