

UM MODELO DE REPUTAÇÃO PARA IDENTIFICAR NÍVEIS EMOCIONAIS EM REDES SOCIAIS

William Sousa da Silva¹; Richardson Ribeiro²; Marco A. de Castro Barbosa³; Fabrício Enembreck⁴
^{1,2,3} Universidade Tecnológica Federal do Paraná – UTFPR – Pato Branco – Paraná
⁴ Pontifícia Universidade Católica do Paraná – PUCPR – PPGIa – Curitiba – Paraná
{williamsousadasilva; prof.richard; marco.cb, fenembreck}@gmail.com

Resumo

Neste trabalho é apresentado um modelo de reputação para identificar o nível emocional de um indivíduo usando os conteúdos postados em suas redes sociais. A proposta do modelo é caracterizar um indivíduo emocionalmente, como forma de estabelecer o quanto confiável um indivíduo pode estar em problemas de tomada de decisão, coletivos, etc. Para validação dos resultados, utilizou-se um conjunto de palavras afetivas definidas pelo Anew-br, como forma de detectar sentimentos e caracterizar um indivíduo emocionalmente. Desse modo, o trabalho procurou identificar e comparar as informações compartilhadas pelos usuários com as palavras afetivas definidas pelo Anew-br e através de um modelo matemático emergiu um modelo de reputação capaz de identificar níveis emocionais.

Palavras-chave: Modelo de reputação, mineração de texto, redes sociais.

1. Introdução e contextualização

O crescimento cada vez maior das redes sociais e seus usuários tem proporcionado à comunidade científica realizar pesquisas em diversas áreas como psicologia comportamental, *marketing*, sociologia e ciência da computação. Dentre esses vários tipos de pesquisas destacam-se os modelos de reputação e confiança criados com o intuito de identificar a forma com que os indivíduos se comportam mediante a rede social, ou seja, mediante o convívio com outros indivíduos.

Em termos da ciência da computação, um modelo de confiança e reputação permite identificar se uma informação recebida de outro indivíduo é ou não confiável (Sabater e Sierra, 2001). Além disso, também é possível saber se uma informação deve ser aceita e quanta atenção dedicar a essa informação. Na literatura, um modelo de reputação é formado por duas dimensões (Mui, 2009): a) em nível local, na qual a confiança atribuída por um indivíduo para outro indivíduo é medida através das interações e da troca de informações entre eles; e b) em nível global, no qual a reputação é atribuída não apenas por um único indivíduo, mas pela comunidade, ou seja, são conjuntos de valores de confiança transmitidos de indivíduo para indivíduo através de votações e testemunhos.

Um modelo de reputação também pode refletir o estado mental de um indivíduo, que está diretamente ligado à forma com que suas decisões são tomadas e de alguma forma refletem o seu estado emocional. Neste trabalho é apresentado um modelo de reputação que utiliza a dimensão emocional de um indivíduo identificada a partir dos conteúdos postados em seus *sites* de relacionamento. Neste trabalho será utilizada a rede social *facebook* (Facebook, 2013). Para a construção do modelo proposto, serão utilizadas listas de palavras afetivas como maneira de definir seu nível emocional.

2. Estado da Arte

Martinazzo (2010) propôs um método de identificação de emoções em bases textuais em língua portuguesa baseado em *latent-semantic-analysis*. A técnica proposta foi aplicada em textos reduzidos de manchetes e notícias extraídas de *sites* da internet. Em alguns experimentos, o método

obteve uma taxa média de identificação de emoções na ordem de 70%. Já Santos (2010) desenvolveu um protótipo capaz de realizar mineração de opinião em textos de redes sociais, tendo como estudo de caso o *Twitter*. O autor utilizou a linguagem Java para a implementação do protótipo e usou o método de aprendizagem de máquina SVM (*support vector machine*) para realizar a classificação de maneira binária entre sentimentos positivos e negativos. Para representar os documentos de texto de maneira estruturada, o autor optou pelo modelo conhecido como vetor de características, uma abordagem estatística para análise de textos. Os experimentos mostram que o classificador alcançou em média uma taxa de acerto de 80%. Esses são apenas alguns dentre vários trabalhos que se dedicam ao estudo de abordagens para mineração de texto e análise de sentimentos.

3. Metodologia

A primeira etapa realizada para aplicar o modelo de reputação foi obter os dados dos usuários do *facebook*. O *facebook* é uma rede social que permite aos usuários publicarem diversos tipos de conteúdos, dentre eles: fotos, vídeos, mensagens de *status*, etc. Para este trabalho focamos o estudo nos conteúdos das mensagens de *status*, pois é nesse tipo de publicação que geralmente as pessoas expressam suas opiniões e sentimentos sobre um determinado assunto. O *facebook* possui uma *application programming interface* (API) denominada *GraphAPI* que permite a integração da rede social com outros *softwares*. Para isso foi utilizado o *software Pentaho Data Integration* (PDI) (Pentaho, 2013). O PDI é um *software* de código aberto que permite a extração de dados, transformação e carga (*load*). A Figura 1 ilustra o processo de extração de dados utilizando essa ferramenta.



Figura 1. Processo de extração dos dados

Como pode ser observado na Figura 1 o processo de extração consistiu em três etapas. Na primeira etapa (*Generate Rows*) é inserida uma chave gerada por um aplicativo da rede social. Na segunda etapa (*HTTP client*) é especificado o local no qual será feita a extração das informações. É nessa etapa que acontece a integração entre a API do *facebook* e o *Pentaho*. Já na terceira etapa (*Text File Output*) é configurado o *Pentaho* para retornar os dados em arquivos. Após isso foram realizadas as etapas de limpeza, seleção e transformação dos dados, para então aplicar o processo de mineração. Essas etapas são conhecidas como *Knowledge-Discovery in Databases* (KDD) (Fayyad, et al., 1996)

A próxima etapa foi comparar os conteúdos das mensagens postadas pelos usuários do *facebook* com uma lista de palavras afetivas definidas pelo Anew-br. O Anew-br (*Brazilian norms for the affective norms for english words*) é uma base textual com mais de 1.000 palavras em que cada uma dessas palavras possui dois valores, sendo um valor de valência e um valor de alerta. Tal base textual é baseado em (Bradley, 1999). Segundo Kristensen (2011) a emoção pode ser definida como uma reação temporal breve de prontidão para uma determinada ação, sendo composta por ao menos duas dimensões ortogonais, uma de valência (do desagradável ao agradável) e outra de alerta (do relaxado ao estimulado), em uma concepção definida como a teoria dimensional da emoção.

De acordo com Kristensen (2011), estímulos que evocam uma emoção de tristeza, por exemplo, podem ser classificados como de valência desagradável e alerta baixo, já estímulos que evocam uma emoção discreta de felicidade podem ser classificados como de valência agradável e alerta alto; estímulos de raiva podem ser classificados como valência desagradável e alerta alto.

A base textual do Anew-br possui palavras com valores de valência e alerta compreendidas no intervalo de 1 a 9. Sendo assim palavras que possuem valores baixos, próximos de 1, por

exemplo, apresentam valência e alerta baixos, ou seja, desagradável e relaxado respectivamente. Já palavras com valores próximos de 9 apresentam valência e alerta altos, ou seja, agradável e estimulado respectivamente. Sendo assim as palavras do Anew-br foram agrupadas em classes de 1 a 9 da seguinte forma: palavras com valores de valência compreendidas entre 1 e 1.99 pertencem a classe 1, entre 2 e 2.99 pertencem a classe 2, entre 3 e 3.99 classe 3 e assim sucessivamente. O mesmo foi feito para o alerta. A Figura 2 ilustra esse processo.

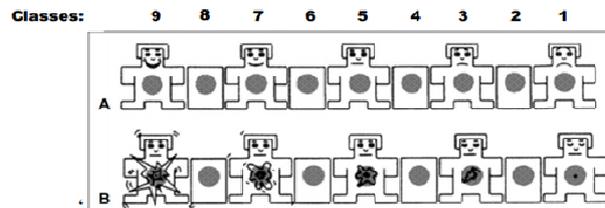


Figura 2. Classes de Valência (A) e Alerta (B). Fonte: (Kristensen, 2011)

Como pode ser observado na Figura 2 as classes 1, 3, 5, 7 e 9 possuem a imagem de um indivíduo que tenta representar as características físicas-sentimentais dessas classes, pois estudos correlatos psicofisiológicos indicam que julgamentos de alerta estão associados a variações na condutância elétrica da pele e julgamentos de valência estão associados a variações na contração de músculos faciais (Bradley, 1999). As demais classes são consideradas classes intermediárias. A partir disso, nós desenvolvemos modelos matemáticos que emergem um modelo de reputação capaz de estimar a média ponderada de valência e alerta, conforme equações (1) e (2):

$$M_V = \frac{q_1 * v_1 + \dots + q_n * v_n}{q_1 + \dots + q_n} \quad (1)$$

$$M_A = \frac{q_1 * A_1 + \dots + q_n * A_n}{q_1 + \dots + q_n} \quad (2)$$

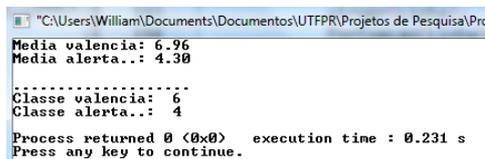
no qual: M_V é a média ponderada para a valência; M_A a média ponderada para o alerta; q_n a quantidade de vezes que uma palavra n é encontrada; v_n o valor de valência de uma palavra n ; e A_n é o respectivo valor de alerta de uma palavra n . As equações (1) e (2) foram implementadas na linguagem C, comparando as palavras do arquivo de saída da Figura 1 com a base textual definida no Anew-br, minerando as palavras afetivas em comum. O retorno do algoritmo é a média ponderada para a valência e para o alerta. Dessa forma foi possível identificar a classe de valência e alerta de um determinado indivíduo, usando como base a representação gráfica das classes definidas na Figura 2.

4. Resultados Iniciais

Foram extraídos do *facebook* dados de dez indivíduos. Nos testes realizados observou-se que o modelo de reputação proposto é capaz de mostrar a dimensão emocional de um indivíduo conforme apresentado na metodologia. Nos casos no qual não houve uma classificação, devido não haver palavras em comum entre os dados extraídos do *facebook* e a base de palavras afetivas do Anew-br, os dados do indivíduo não foram considerados no modelo. A Figura 3 ilustra os resultados de uma extração e classificação.

Conforme ilustrado na Figura 3, observa-se que o indivíduo referente a essa análise foi classificado como valência e alerta 6 e 4 respectivamente, ou seja, a maioria das palavras publicadas

por ele pertenciam à essas classes. Logo, com base nas informações contidas na Figura 2 é possível observar seu aspecto emocional.



```
"C:\Users\William\Documents\Documentos\UTFPR\Projetos de Pesquisa\Prc
Media valencia: 6.96
Media alerta.: 4.30
-----
Classe valencia: 6
Classe alerta.: 4
Process returned 0 (0x0)   execution time : 0.231 s
Press any key to continue.
```

Figura 3. Janela de apresentação dos resultados (Classificação)

5. Conclusões

Este trabalho apresentou as etapas desenvolvidas em um projeto que utiliza a mineração de texto como base. Modelos matemáticos para emergir um modelo de reputação foram apresentados, como forma de identificar o nível emocional de um indivíduo usando os conteúdos postados em suas redes sociais. Essa abordagem tem sido pouco explorada na literatura e na maioria dos casos é utilizada apenas como forma de avaliação de produtos e serviços a partir de opiniões de usuários, como forma de identificar a reputação de um possível indivíduo ou grupo. Pragmaticamente, o trabalho procurou identificar e comparar as informações compartilhadas pelos usuários com as palavras afetivas definidas pelo Anew-br usando um modelo de reputação capaz de identificar o nível emocional de um indivíduo.

Apesar dos resultados iniciais serem satisfatórios e desafiadores, várias pesquisas estão em andamento para responder várias questões em aberto. Por exemplo, a base afetiva textual está sendo aumentada (enriquecida) conforme palavras associadas são identificadas nas redes sociais. Um sistema computacional capaz de obter os dados em tempo real está sendo desenvolvido, no qual o modelo de reputação apresentado está sendo usado como base. Usuários em cenários diferentes serão usados como forma de validar e melhorar o modelo. O modelo ajustado poderá ser usado em processos de decisão coletiva, ou ainda como forma de medir a confiabilidade dos usuários em processos avaliativos.

Referências

- BRADLEY, M. M.; LANG, P. J. Affective Norms for English Words (ANEW): instruction manual and affective ratings. Technical Report C-1. Gainesville: University of Florida; 1999.
- FACEBOOK, <https://pt-br.facebook.com/>, acesso em 15/05/2013.
- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. Artificial Intelligence Magazine, v. 17, n. 3, p. 37-54, 1996.
- KRISTENSEN, Christian. H. Normas brasileiras para o Affective Norms for English Words. Trends Psychiatry Psychother. 2011, 33(3) – 135-146.
- MARTINAZZO, Barbara. Um método de identificação de emoções em textos curtos para o português do Brasil. 2010. 82 f. Dissertação (Mestrado em informática) – Pontifícia Universidade Católica do Paraná, PPGIa, Curitiba, 2010.
- MUI, L. Computational Models of Trust and Reputation: Agents, Evolutionary Games, and Social Networks. MIT, 2002.
- PENTAHO, <http://www.pentaho.com/explore/pentaho-data-integration/>, acesso em 15/05/2013.
- SABATER, J.; SIERRA, C. Social regret, a reputation model based on social relations. SIGecom Exch., 3:44–56, 2001.
- SANTOS, Leandro. M. Protótipo para mineração de opinião em redes sociais: Estudos de casos selecionados usando o Twitter. 2010. 103 f. Monografia (Bacharel em ciência da computação) – Departamento de ciência da computação, Universidade Federal de Lavras, Lavras, 2010.