

Uso de Técnicas de Análise de Sentimentos em Tweets relacionados ao Meio-Ambiente

André Luiz F. Alves¹, Cláudio de Souza Baptista¹, Luiz Henrique de Andrade¹,
Roberta F. Cerqueira Paes²

¹Laboratório de Sistemas de Informação/Departamento de Sistemas e Computação
Universidade Federal de Campina Grande (UFCG) - Paraíba - Brasil

²Companhia Hidro Elétrica do São Francisco - Pernambuco- Brasil

andre@uepb.edu.br, baptista@dsc.ufcg.edu.br, luizha.cc@gmail.com,
rfcpaes@chesf.gov.br.

Abstract. *The proliferation of social media on the Web and the need for companies to measure the impact of their environmental initiatives by the affected population has driven the proposal of an automatic mechanism to analyze the population's opinions. This necessity becomes even more evident especially regarding companies that deal directly with high environmental risk. In this context, the goal of this work is to apply sentiment analysis techniques on tweets related to environmental issues aiming to help energy companies analyze the impacts of environmental actions taken over time through opinions within social media.*

Resumo. *A proliferação dos meios de comunicação na Web e a necessidade das empresas compreenderem os impactos de suas ações socioambientais junto à população afetada, torna necessária a adoção de um mecanismo automático para analisar as opiniões da população. Esta necessidade evidencia-se ainda mais no caso de empresas que tratam diretamente com alto risco ambiental. Neste contexto, o objetivo deste trabalho é a aplicação de técnicas de análise de sentimentos em tweets relacionados à temática ambiental para auxiliar empresas do setor energético na análise dos impactos das ações empreendidas ao longo do tempo através das opiniões contidas nas mídias sociais.*

1. Introdução

Nas últimas décadas, nota-se um crescimento na preocupação ambiental pelos diversos atores sociais - governantes, empresas e cidadãos - no tocante aos impactos ambientais decorrentes da ação humana. Tal preocupação tem resultado numa conscientização da sociedade, colocando-a como fiscal social com relação à degradação ambiental, e demandando leis ambientais cada vez mais rigorosas [Nicolella, Marques and Skorupa 2004]. Particularmente, no setor elétrico brasileiro, cuja base energética está nas hidroelétricas, que trazem grandes danos ambientais decorrentes das inundações geradas quando de suas implantações, há uma crescente preocupação de como mitigar estes impactos através de ações socioambientais junto à população afetada. Neste sentido, a política ambiental atual para o setor elétrico aduz que a gestão ambiental deve englobar do planejamento à implantação e manutenção dos empreendimentos energéticos.

Resultados em direção à melhoria no tratamento aos impactos ambientais causados por empresas de energia passam pela adoção de Sistemas de Gestão Ambiental

(SGA), que visam implantar políticas e procedimentos técnico-administrativos, possibilitando um melhor controle dos impactos ambientais causados. Neste processo de gestão ambiental, há uma necessidade de envolver a sociedade civil com o objetivo de implantar uma nova maneira de tratar as questões ambientais, promovendo, assim, a construção da cidadania e do fiscal social. Assim, compreender o que as pessoas estão pensando ou suas opiniões é fundamental para tomada de decisão, especialmente no contexto em que as pessoas expressam seus comentários de forma voluntária, principalmente através das diversas mídias sociais (blogs, fóruns de discussões, microblogging, redes sociais, etc.) proporcionados pelo surgimento da Web 2.0.

Analisar os comentários expressos nas mídias sociais é uma tarefa difícil de ser realizada manualmente, principalmente devido ao grande volume de dados. Neste contexto, a técnica de análise de sentimentos pode ser aplicada para compreender as opiniões da sociedade sobre a questão ambiental propagada em mídias sociais. De acordo com Liu [Liu 2012], a análise de sentimentos tem como principal objetivo obter e formalizar a opinião e o conhecimento subjetivo em documentos não estruturados (textos), para posterior análise dentro de um domínio específico. Visa, pois, analisar as opiniões, sentimentos, avaliações, atitudes e emoções de pessoas com relação a entidades, pessoas, produtos, serviços, eventos, dentre outros. Uma forma comum de compreender o sentimento geral de um conjunto de documentos é através da sumarização das opiniões realizada através das classificações das opiniões com base em categorias (Polaridade): positiva, negativa e neutra [Liu 2012], [Pang and Lee 2008].

Nos últimos anos, a ferramenta Twitter tem sido largamente utilizado para expressar opiniões diversas, tornando-se um repositório bastante atrativo para se implantar técnicas de análise de sentimentos [Kwak et al. 2010]. Todavia, devido à informalidade textual expressa pelos usuários, a tarefa de análise de sentimentos em tweets torna-se por demais complexa. Ademais, estes problemas são agravados pela limitação de 140 caracteres e pela complexidade gramatical da língua portuguesa.

Portanto, o principal objetivo deste trabalho é aplicar técnicas de análise de sentimentos em tweets escritos em português para auxiliar na atividade de compreensão do sentimento da sociedade relacionada à temática ambiental. Algoritmos de aprendizagem de máquina são utilizados para detecção da polaridade do sentimento. Para a sumarização do sentimento foram utilizados gráficos com variação temporal da polaridade detectada. A ferramenta desenvolvida com este estudo é aplicável dentro de um SGA - Sistema de Gestão Ambiental - para possibilitar uma observação detalhada de alto nível e global, auxiliando desta forma na tomada de decisões.

O restante deste artigo está organizado como segue. Na seção 2, são apresentados alguns trabalhos relacionados. Na seção 3, apresenta-se a metodologia considerada para o desenvolvimento do analisador de sentimentos. Na seção 4, apresentam-se os resultados obtidos. Por fim, na seção 5 são apresentadas as conclusões e os trabalhos futuros.

2. Trabalhos Relacionados

A análise de sentimentos tem sido uma das áreas de pesquisa mais ativas no campo de Natural Language Processing- NLP [Liu 2012]. Uma visão geral sobre análise de sentimentos pode ser encontrada em [Pang and Lee 2008]. Diversos domínios têm aplicado técnicas de análise de sentimentos tais como: mercado acionista, através da

identificação do humor do mercado baseado nas opiniões de especialistas [O'Hare et al. 2009]; opiniões dos consumidores acerca de produtos ou serviços [Eirinaki et al. 2012],[Hu and Liu 2004]; aplicações de turismo, através da análise dos comentários dos viajantes [Bjørkelund et al. 2012] análise de políticos e de política [Fang et al. 2012].

As principais técnicas para classificar a polaridade do sentimento em documentos não estruturados têm sido baseadas em aprendizado de máquina, análise semântica, estatística e análise léxica. No entanto, técnicas de aprendizagem de máquina têm sido mais exploradas na literatura [Sharma and Dey 2012], [Feldman, 2013]. Uma das principais limitações no uso de aprendizado supervisionado é a necessidade de dados rotulados para treinamento e teste (*dataset*). Para auxiliar a atividade de coletar os dados rotulados de forma automática, vários trabalhos fizeram uso de emoticons-caracteres que transmitem emoções. Em [Li and Li 2011] 87% dos tweets contendo *emoticons* possuem os mesmos sentimentos representados pelos *emoticons* no texto. Trabalhos que utilizam emoticons para treinamento dos classificadores têm apresentado excelentes resultados de acurácia (acima de 80%) [Pak and Paroubek 2010].

Por outro lado, ainda existem poucos trabalhos na literatura que realizam análise de sentimentos utilizando um corpus em língua portuguesa. Os trabalhos de [Chaves et al. 2012],[Sarmiento et al. 2009],[Tumitan and Becker 2013] utilizam técnicas de análise léxica baseada em dicionários e apenas o trabalho de [Nascimento et al. 2009] utiliza técnicas de aprendizagem de máquina.

Em [Chaves et al. 2012] é apresentado um algoritmo que utiliza uma abordagem de análise léxica para classificação de sentimentos em comentários em português. Ontologias e lista de adjetivos polarizados (positivo, negativo e neutro) que expressam sentimentos são utilizados para definir a orientação semântica dos textos analisados. Os resultados indicam uma média do F-Measure de apenas 0,32 no reconhecimento da polaridade. Em [Tumitan and Becker 2013], os autores utilizam um dicionário de palavras (SentiLex-PT), que foi adaptado segundo o contexto, para analisar as opiniões em comentários sobre políticos realizados em jornais e estudar a correlação dos sentimentos expressos com as pesquisas de intenção de votos. Em [Nascimento et al. 2009] classificadores de sentimento são utilizados para avaliar as reações das pessoas no Twitter em relação as notícias vinculadas na mídia. Os resultados de acurácia variaram de 70% a 80% de acordo com o tipo de notícia e classificador utilizado.

A abordagem para análise de sentimento proposta neste trabalho difere dos trabalhos relacionados aqui analisados no tocante ao uso de dois classificadores, evitando o uso de PosTagger (Part-of-Speech) na identificação de um conteúdo opinativo. Destarte, o primeiro classificador identifica se um conteúdo é subjetivo ou objetivo; e o segundo classificador identifica a polaridade (positiva ou negativa) do conteúdo já identificado como subjetivo (opinativo).

3. Processo de Análise de Sentimentos

Nesta seção são abordadas a metodologia utilizada no processo de análise de sentimentos bem como a formalização do problema a ser endereçado.

3.1. Visão Geral da Metodologia

A abordagem de análise de sentimento proposta neste trabalho é caracterizada pela utilização de técnicas de Recuperação da Informação (RI) para prover suporte à tomada

de decisão através da mineração de informações contidas em microtextos, como os de microblogging e de redes sociais. Especificamente, este trabalho utiliza técnicas de análise de sentimentos para determinar a polaridade da opinião (positiva ou negativa) utilizando textos oriundos do microblogging Twitter. A Figura 1 apresenta a visão geral da metodologia utilizada no processo de análise de sentimento abordado neste trabalho.



Figure 1. Visão Geral do processo de Análise de Sentimento

O primeiro passo é a coleta dos dados a partir do Twitter. No segundo passo, os tweets são submetidos à uma fase de pré-processamento, que inclui a remoção de stopwords (preposições, artigos, etc.), termos especiais usados no Twitter (RT, via, etc.), remoção de nomes dos usuários e tratamento de hashtags (#) (separação de termos compostos de acordo com a capitalização das letras). No terceiro passo, contempla-se a etapa da classificação de polaridade de sentimento, que se constitui no principal componente e objeto de estudo de várias pesquisas da área de análise de sentimentos, cujo objetivo é a identificação dos sentimentos contidos nos textos. Neste trabalho, a partir dos resultados apresentados por [Alves et al. 2014] em estudos comparativos entre técnicas de classificação da polaridade, a definição da polaridade é realizada utilizando dois modelos de classificação: o primeiro classifica os tweets em opinativos (subjetivos) ou informativo (objetivo) e o segundo classifica a polaridade (Positivo ou Negativo) dos tweets opinativos. Finalmente, após a indexação dos sentimentos detectados pela etapa de classificação, atinge-se a última etapa da metodologia, na qual os resultados são explorados através de uma análise temporal dos sentimentos, obtendo assim a sumarização da análise de sentimentos.

3.2. Formalização do problema

Os microtextos explorados nesta pesquisa foram oriundos do microblogging Twitter, embora a proposta dessa pesquisa seja aplicável a outros tipos de microtextos, como os de redes sociais. Ademais, consideramos que uma mensagem do Twitter apresenta apenas um sentimento predominante, embora algumas pesquisas considerem múltiplas emoções associadas a uma mensagem.

Seja um tweet $t_i \in T$, tem-se que a principal atividade da classificação do sentimento é obter a opinião predominante sobre as expressões textuais do documento t_i . O problema de detecção de polaridade do sentimento pode ser tratado como uma tarefa de categorização de textos. Mais formalmente, uma tarefa de classificação é encontrar uma função que aproxima a uma função de classificação $f: T \rightarrow C$ com $f(t_i) = c_j$ tal que $C = \{c_1, \dots, c_n\}$ representa um conjunto de n categorias predefinidas. A função f descreve como os textos são associados às classes, atribuindo um texto $t_i \in T$ para sua categoria $c_j \in C$. Neste trabalho, o conjunto T representa todos os textos coletados e $c_j \in C = \{positivo, negativo\}$ é a polaridade predominante (orientação semântica do sentimento) relacionada ao *tweet* t_i .

3.3. Identificação da polaridade do sentimento

Para definir a função de classificação $f: T \rightarrow C$ para a identificação do sentimento nos tweets utilizamos uma abordagem de aprendizagem supervisionada de máquina, que necessita de um conjunto de dados (*dataset*) já rotulados formado por dois conjuntos disjuntos: treinamento e teste (validação). Então, dados dois conjuntos T_t e T_v , respectivamente de treinamento e validação, tem-se que $T_t \cap T_v = \emptyset$. Assim, a classificação supervisionada começa com o conjunto de treinamento ($T_t = t_1, \dots, t_n$) com os textos já marcados com as categorias $c_j \in C = (c_1, \dots, c_n)$ e a tarefa é determinar o modelo de classificação capaz de correlacionar corretamente um novo texto $t_w \in T_v$ à sua categoria, ou seja, $f: T \rightarrow C$ com $f(t_w) = c_j$. Neste caso, o conjunto de treinamento é utilizado para a construção de um classificador que aprenderá automaticamente as regras e características gerais dos documentos classificados. Já o conjunto de teste faz-se necessário para validar o classificador treinado.

Na tarefa de classificação, devem-se considerar as características (*features*) que o modelo de classificação observa nos dados de treinamento no processo de aprendizagem. Na classificação de textos, as porções (*tokens*) dos textos são extraídas e analisadas, e o classificador seleciona as características relevantes, representando-as na forma de um vetor de termos (*bag of words*). Neste trabalho, apenas a característica de frequência de termos foi utilizada nos classificadores implementados.

4. Estudo de Caso

Nesta seção é apresentado um estudo de caso junto a empresa do setor elétrico Companhia Hidroelétrica do São Francisco - CHESF, que visa validar a metodologia aqui proposta.

4.1. Seleção do Corpus

O objetivo desse trabalho é analisar as opiniões da sociedade relacionadas à temática ambiental. Desta forma, tomou-se como estudo de caso o Plano de Ação Socioambiental (PAS), idealizado pela Chesf. O PAS é um conjunto de ações que envolve toda a sociedade civil com o objetivo de implantar uma nova maneira de tratar as questões ambientais, promovendo, assim, a construção da cidadania. O PAS abrange a área que fica circunscrita à área de influência do complexo hidrelétrico de Paulo Afonso, constituída por cinco municípios: Paulo Afonso - BA, Glória - BA, Delmiro Gouveia - AL, Pariconha - AL e Jatobá-PE.

O Twitter foi utilizado como fonte de dados para realização da análise de sentimentos. Para coletar os dados, foi criado um motor de busca, que utiliza a API (Application Programming Interface) do Twitter, e diariamente, durante o período de coleta, armazenava tweets que continham termos previamente selecionados. A Figura 2 apresenta o quantitativo de tweets coletados durante o mês de Março de 2015. No total, foram coletados cerca de 107.000 tweets.

Os termos utilizados na coleta dos dados foram definidos segundo três parâmetros: 1) cidades de atuação do PAS; 2) Termos relacionados ao meio ambiente; e 3) Termos relacionados à CHESF. Os gráficos da Figura 3 apresentam a distribuição dos quantitativos de tweets associados com os termos utilizados na coleta segundo estes três critérios mencionados.

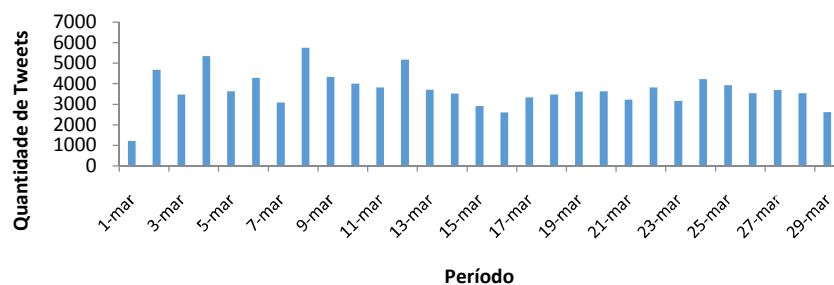


Figure 2. Quantidade de Tweets Coletados durante o período considerado

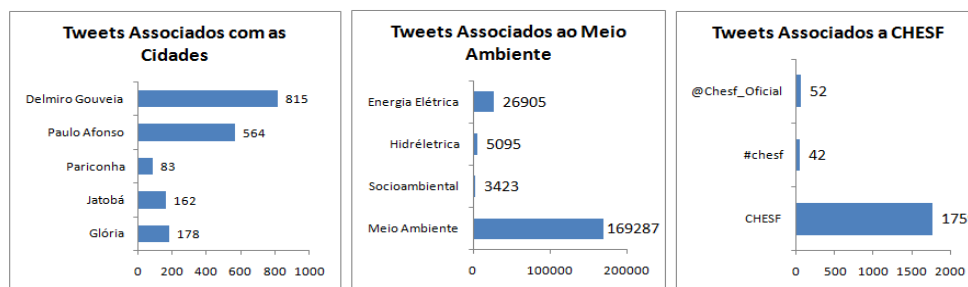


Figure 3. Quantidade de tweets associados com os termos da coleta

4.2. Métricas para Avaliação

As métricas utilizadas para avaliação dos resultados dos algoritmos de detecção de polaridade foram as utilizadas frequentemente na literatura em avaliações de sistemas de Recuperação da Informação (RI): acurácia, precisão, revocação e F-measure. Especificamente, na área de análise de sentimentos, são as métricas utilizadas na literatura para avaliação dos algoritmos de detecção de polaridade dos sentimentos.

Para a avaliação do algoritmo que utiliza técnicas de aprendizado de máquina supervisionado para identificação da polaridade do sentimento, uma fração dos *tweets* rotulados é reservada para treinar o modelo, não sendo utilizada para a obtenção das métricas. A outra fração é utilizada para aplicar o classificador de sentimento e comparar os resultados com os rótulos marcados. Dessa forma, adotamos duas abordagens para obtenção de tweets com sentimentos já rotulados:

- *Emoticons*: utilizou a abordagem de [Pak & Paroubek 2010] na qual considera-se que toda as palavras contidas na mensagem que contém os caracteres que expressam emoções, e.g. *emoticon* alegre - “:-)”, “:)”, “=)”, “:D”, etc - e *emoticon* triste “:-(”, “:(”, “=(”, “;(”, etc -, também estão associadas a emoção do caractere (*emoticon*). Assim se um tweet apresenta um *emoticon* alegre (“:-)”), por exemplo, sua polaridade considerada é positiva.
- Rotulação Manual: 1.778 tweets dentre os coletados foram escolhidos de forma aleatória e separados para rotulação manual da polaridade do sentimento.

A Tabela1 apresenta exemplos de twitters que foram rotulados manualmente. Os dois métodos de obtenção de rotulação de sentimentos foram utilizados na comparação e combinação de resultados dos classificadores. A Tabela 2 apresenta o quantitativo de tweets com sentimentos rotulados a partir de cada abordagem.

Tabela 1. Exemplo de Tweets Rotulados Manualmente

Rótulo	Tweet
Positivo	"a parte da hidreletrica foi uma das melhores"
Negativo	"É impressão minham ou a energia elétrica está tendo aumento de 15 em 15 dias? Isso é umavergonha!"
Neutro	"A competência para legislar sobre meio ambiente do trabalhonão é concorrente mas privativa da União. STF ADI 1893-MC"

Tabela 2. Quantidade de Tweets Rotulados (Conjunto de Treinamento e Testes)

Abordagem	Quantidade			
	Positivos	Negativos	Neutro	Total
<i>Emoticons</i>	126	73	-	199
<i>Rotulação Manual</i>	181	188	1409	1778

Para avaliar a capacidade de generalização dos modelos de classificação, foi utilizado o método de validação cruzada *k-fold*, com $k=10$, ou simplesmente *10-fold*.

4.3. Classificadores Utilizados na Definição da Polaridade

O processo de classificação da polaridade é realizado com a utilização de dois classificadores. O primeiro classificador é utilizado para classificar os textos em objetivos e subjetivos, funcionando como uma espécie de filtro para o próximo classificador. Já o segundo classificador foi treinado para classificar as polaridades das mensagens subjetivas em apenas duas classes: Positiva e Negativa.

Inicialmente, no planejamento do experimento deste estudo de caso, foi cogitado a utilização de classificadores SVM (Suport Vector Machine) (Ian H. Witten and Eibe Frank, 2005), considerando que este classificador tem apresentado bons resultados em problemas de classificação [Sharma and Dey 2012], [Pang and Lee 2008], [Read 2005]. No entanto, considerando o conjunto de tweets rotulados (*dataset*), conforme Tabela 2, observamos um desbalanceamento entre as classes dos tweets rotulados. Isto deve-se pelo fato de que no contexto aplicado, o número de notícias relacionadas ao tópico é notoriamente maior que o número de opiniões relatadas. Assim, ao escolher o classificador utilizado neste trabalho, considerou-se o desempenho do algoritmo com uma base de dados desbalanceada [Akbari et al. 2004]. Portanto, os classificadores Naive Bayes foram utilizados neste trabalho, uma vez que se aplicam perfeitamente com a premissa de dados independentes e com classes não balanceadas.

4.4. Resultados

A Tabela 3 apresenta, através do processo de validação cruzada *10-fold*, os resultados (média ponderada) obtidos pelo processo de classificação da polaridade. Os resultados são considerados excelentes para tweets no idioma português, principalmente pelo fato do desbalanceamento apresentado na base de dados rotuladas (*dataset*), onde a fração de

tweets objetivos (polaridade neutra) é muito maior quando comparado com os tweets com polaridades positivas e negativas.

Tabela 3. Resultados do Classificador de Sentimentos

Classificador	Acurácia	Precisão	Revocação	F-Measure
NaiveBayes- Classificação Dupla	0,721	0,730	0,721	0,725

Aplicando este classificador a todos os tweets coletados, as polaridades foram definidas e indexadas para análise na etapa de sumarização do sentimento. Do total de tweets, apenas 16.134 (cerca de 15%) foram considerados pelo classificador como tweets que apresentam sentimentos (positivos ou negativos). Os gráficos da Figura 4 ilustram o comportamento do sentimento detectado ao longo do período. Na Figura 4.a ilustra a distribuição temporal contendo o comportamento da quantidade de sentimentos positivos e negativos por dia. Este gráfico é utilizado quando deseja-se conhecer de forma global, considerando inclusive o fator quantitativo na análise, os sentimentos expressos pelos usuários ao longo do período analisado. Já a Figura 4.b ilustra as proporções de tweets positivos e negativos considerando apenas as quantidade de tweets por dia. Por exemplo, percebe-se que no dia 28 de Março, cerca de 80% de tweets do dia apresentam sentimentos positivos e apenas 20% com sentimentos negativos.

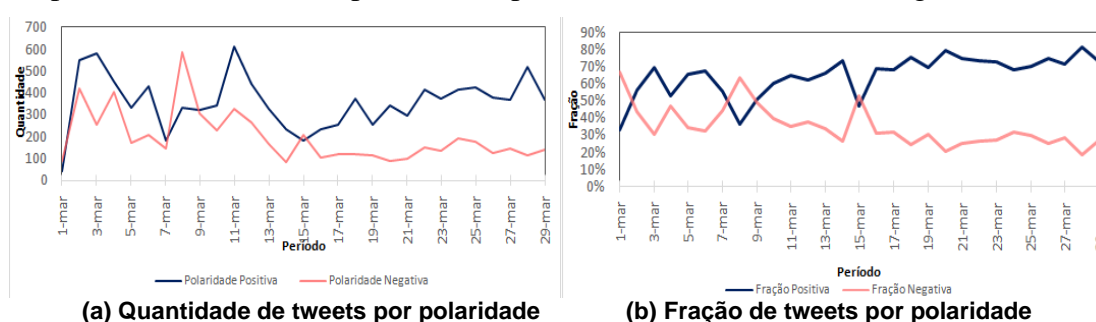


Figure 4. Comportamento da polaridade detectado sobre todos os tweets coletados

Uma forma de obter a orientação semântica geral do sentimento expresso nos microtextos é através da subtração do número de tweets com sentimentos positivos pela quantidade de tweets com sentimento negativo. A Figura 5 apresenta um gráfico que ilustra a orientação semântica. O objetivo deste gráfico é ilustrar a tendência do sentimento em relação ao tema analisado. Por exemplo, percebe-se que no período de 4 a 7 de Março há uma mudança de sentimento, e desta forma há possibilidade de analisar os motivos dessa variação de sentimentos. No gráfico, também utilizamos a média móvel para suavizar a curva, eliminando os efeitos de valores atípicos (*outliers*).

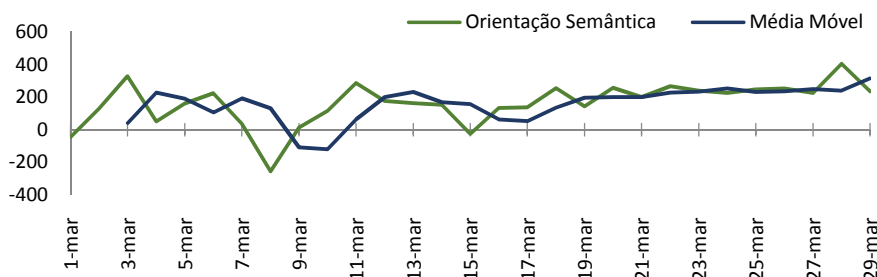


Figure 5. Orientação Semântica Geral

5. Considerações Finais

Neste artigo, foi apresentado um estudo de caso que utiliza técnicas de análise de sentimentos aplicadas a tweets relacionados ao meio-ambiente. Um modelo de classificação Naive-Bayes foi treinado e validado, obtendo excelentes resultados para tweets escritos no idioma português. Mesmo o conjunto de tweets coletados apresentando majoritariamente tweets objetivos, aproximadamente 85% dos tweets, foi possível definir um classificador que identifica a polaridade do sentimento de um tweet com uma acurácia de 72,1% e F-Measure de 72,5%.

Identificado os tweets opinativos de todo conjunto coletado e realizando classificação da polaridade dos sentimentos, através de técnicas de aprendizagem de máquina, foi possível sumarizar as opiniões através de uma análise temporal sobre os dados, possibilitando a identificação da orientação semântica dos sentimentos expressos pelos usuários. Embora a técnica de sumarização da opinião tenha sido aplicada a todo o conjunto de dados coletado, através dos termos das pesquisas, é possível realizar a sumarização individualmente sobre cada termo de interesse. Tomando por exemplo este estudo de caso, é possível realizar uma análise de sentimentos nos tweets das cidades de atuação do PAS, analisando o impacto ou a correlação das ações ambientais promovidas pela CHESF nos sentimentos expressados pela sociedade nas redes sociais ou microbloggings. Além do mais, utilizando técnicas avançadas para identificar o objeto avaliado em um tweet opinativo, a exemplo de Reconhecimento de Entidades Nomeadas (NER), é possível identificar o sentimento da sociedade em relação a empresa CHESF.

A abordagem desenvolvida neste estudo é aplicável dentro de um SGA para possibilitar uma observação detalhada de alto nível e global, auxiliando desta forma na tomada de decisões. Um trabalho futuro é aplicar essas técnicas em um ambiente corporativo, no qual são coletadas as opiniões dos agentes envolvidos no PAS, e assim possibilitara compreensões das opiniões de forma automática desses agentes em relação as ações empreendidas ao longo do tempo.

Agradecimentos

Os autores agradecem o suporte financeiro da ANEEL, sob o contrato de P&D+I N° ANEEL 0048-1119/2012.

Bibliografia

- Akbani, R., Kwek, S. and Japkowicz, N.. Applying support vector machines to imbalanced datasets. 15th ECML, 39-50 (2004).
- Alves, A. L. F., Baptista, C. S., Firmino, A. A., Oliveira, M. G., Figueirêdo, H. F.. Temporal Analysis of Sentiment in Tweets: A Case Study with FIFA Confederations Cup in Brazil. DEXA (1) 2014: 81-88.
- Bjørkelund, E., Burnett, T. H., and Nørvi K. A study of opinion mining and visualization of hotel reviews. In Proceedings of the 14th International Conference on IJWAS (New York, USA, 2012), ACM Press, p. 229.
- Chaves, M., De Freitas, L., Souza, M., Vieira, R.: PIRPO: An Algorithm to Deal with Polarity in Portuguese Online Reviews from the Accommodation Sector. Natural Language Processing and Information Systems 7337, 1–5 (2012).

- Eirinaki, M., Pisal, S., and Singh, J. Feature-based opinion mining and ranking. *Journal of Computer and System Sciences* 78, 4 (July 2012), 1175–1184.
- Fang, Y., Si, L., Somasundaram, N., Yu, Z.: Mining contrastive opinions on political texts using cross-perspective topic model. In: *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM'12*. p. 63. ACM Press, New York, USA (2012).
- Feldman, R.: Techniques and applications for sentiment analysis. *Communications of the ACM* 56(4), 82 (Apr 2013).
- Hu, M., and Liu, B. Mining and summarizing customer reviews. In *KDD '04* (New York, New York, USA, 2004), ACM, ACM Press, p. 168.
- Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: *Proceedings of the 19th international conference on WWW '10*. p. 591. ACM Press, New York (2010).
- Li, Y.M., Li, T.Y.: Deriving Marketing Intelligence over Microblogs. In: *2011 44th Hawaii International Conference on System Sciences*. pp. 1–10. IEEE (Jan 2011)
- Liu, B.: *Sentiment Analysis and Opinion Mining*. *Synthesis Lectures on Human Language Technologies* 5(1), 1–167 (May 2012).
- Nascimento, P., Aguas, R., Lima, D.D., Kong, X., Osiek, B.: Análise de sentimento de tweets com foco em notícias. In *I Brazilian Workshop on Social Network Analysis and Mining*, 2012.
- NICOLELLA, G.; MARQUES, J. F.; SKORUPA, L. A.; *Sistema de Gestão Ambiental: aspectos teóricos e análises de um conjunto de empresas da região de Campinas, SP. Embrapa*, 2004.
- O'Hare, N., Davy, M., Bermingham, A., Ferguson, P., Sheridan, P., Gurrin, C., and Smeaton, A. F. Topic-dependent sentiment analysis of financial blogs. In *Proceeding of the TSA '09* (New York, USA, 2009), ACM Press, p. 9.
- Pak, A., Paroubek, P.: Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Proceedings of the Seventh conference on International Language Resources and Evaluation LREC'10 (ELRA)*, 1320–1326 (2010).
- Pang, B., and Lee, L. *Opinion Mining and Sentiment Analysis*. *Foundations and Trends in Information Retrieval* 2, 2 (2008), 1–135.
- Read, J.: Using Emoticons to reduce Dependency in Machine Learning Techniques for Sentiment Classification (June), 43–48 (2005).
- Sarmiento, L., Carvalho, P., Silva, M.J., de Oliveira, E.: Automatic creation of a reference corpus for political opinion mining in user-generated content. In: *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion - TSA '09*. p. 29. ACM Press, New York (2009).
- Sharma, A., and Dey, S. A comparative study of feature selection and machine learning techniques for sentiment analysis. In *RACS '12* (New York, USA), ACM Press, p. 1.
- Tumitan, D., Becker, K.: Tracking Sentiment Evolution on User-Generated Content : A Case Study on the Brazilian Political Scene. *SBBD 2013* pp. 1–6 (2013).