

Analizador de Expressões Positivas e Negativas Aplicado em Comentários de Livros e Filmes.

Ubirajara Martin Coelho¹, Ana Carolina E.S.Lima², Nizam Omar³

(1) Universidade do Estado de Mato Grosso – UNEMAT, Brasil.

ubirajara@unemat.br

(2) Universidade Presbiteriana Mackenzie – MACKENZIE, Brasil.

aceslima@gmail.com

(3) Universidade Presbiteriana Mackenzie – MACKENZIE, Brasil.

nizamomar51@gmail.com

Resumo — As investigações em análise automática de documentos vêm permitindo grandes avanços permitindo no reconhecimento de aspectos subjetivos. Dentre estes, destaca-se a classificação da polaridade do texto, ou seja, o quão negativa ou positiva são as opiniões expressadas nele. Contudo, por ser uma área ainda em desenvolvimento, métodos criados para estas análises, na maioria, são para língua inglesa, o que dificulta sua utilização em textos escritos em português. Assim, esse trabalho tem como objetivo o estudo e a aplicação de uma ferramenta que, conterá um algoritmo de classificação de sentimentos, sendo ele capaz de avaliar a polaridade de comentário extraído de uma base de dados de livros da Amazon e outra do Twitter, baseando-se em técnicas de mineração de textos.

Keywords: Análise de Sentimentos, Classificação de Polaridade, Mineração de Opinião.

Abstract — The automatic analysis of documents has allowed great advances in the recognition of subjective aspects. Among these, we highlight the classification of the polarity of the text, that is, how negative or positive are the opinions expressed in it. However, because it is an area still under development, methods created for these analyzes are mostly for the English language, which makes it difficult to use them in texts written in Portuguese. Thus, this work has the objective of studying and applying a tool that will contain a classification algorithm of feelings, being able to evaluate the commentary polarity extracted from a database of Amazon and other books of Twitter, based In text mining techniques.

I. INTRODUÇÃO

As opiniões das pessoas são muito valiosas e importantes para a influência social e também como um efeito econômico. Essas informações tornam-se úteis para pessoas tomarem decisões satisfatórias com relação a um produto seja na compra realizada por um cliente ou pela venda feita pelo comerciante. As opiniões das pessoas, relacionadas com o produto, na maioria das vezes expressam vários aspectos positivos e negativos. Esses aspectos representam quaisquer pontos de vistas das pessoas, tratando-se como semelhantes ou desejáveis (positivo), antipatia ou indesejáveis (negativo) bem como ponto de vista neutro [1]. A polaridade das opiniões, podem ser extraídas em níveis diferentes como nível de documento, nível de sentença e nível de aspecto [1]. Essas

opiniões que acabam gerando bases de informações que após um tratamento especial disponibiliza conhecimentos ricos e preciosos que tornam uma fonte importante para empresas e consumidores de tal produto [2].

A disseminação de comentários de um produto encontra-se pela web ou em vários tipos de sites, desde blogs pessoais a site de compras. Comentários e opiniões postados em alguma rede social (e.g. *twitter*, *facebook*, *instagram*, etc...) seja formal ou informal, de acordo com o local de postagem, pode refletir o sentimento sem censura do usuário. Ao responder um questionário com perguntas entrevistadas por uma pessoa, esta poderia se sentir receio e não responder com a franqueza que se encontram nas postagens das redes sociais. Nesse contexto o Twitter é considerado como uma fonte ainda bastante explorada na área de análise de sentimento que é um serviço de micro-blogging permitindo postagens de mensagens de até 140 caracteres.

II. REFERENCIAL TEÓRICO

A. Análise de Sentimentos

Uma das áreas da Mineração de Texto também conhecida como mineração de opiniões e análise de sentimentos, busca classificar textos não por tópicos, e sim pelo sentimento ou opinião em documentos. Esses documentos em geral estão associados à classificação binária (0 ou 1) entre sentimentos (positivo e negativos), usa-se o termo de forma abrangente para demonstrar “*como são tratadas opiniões, sentimentos e subjetividade em documentos textos de forma computacional*” (Pang e Lee, 2008).

A análise de sentimentos busca por meio de documentos textuais, opiniões que se formam, mas sem se preocupar no que está sendo comentado, e sim com a opinião constituída no que refere-se a sua polaridade, ou seja, se uma menção por exemplo, é positiva ou negativa com relação a um determinado produto [3].

Aplicamos mineração de opiniões em documentos textos em qualquer tamanho e formato, como páginas *web*, *post*, comentários, *tweets*, etc. A opinião é composta por dois elementos chaves, um alvo e um sentimento sobre esse alvo. Dizemos que um alvo classifica-se em: uma entidade, aspecto de uma entidade ou tópico que pode ser representado por um produto, pessoa, organização, marca, evento, etc. Por outro lado um sentimento representa uma atitude, opinião ou emoção que o autor da opinião tem a respeito do alvo [4]. A

polaridade de um sentimento está ligada a um ponto em alguma escala que representa a avaliação positiva, neutra ou negativa do significado deste sentimento [5].

Uma opinião corresponde a uma quintupla $(e_i, a_{ij}, s_{ijkl}, h_k, t_j)$ [6], em que:

- e_i : é o nome de uma entidade;
- a_{ij} : é um aspecto da entidade e_i (opcional);
- s_{ijkl} : é a polaridade do sentimento sobre aspecto a_{ij} que tem como alvo a entidade e_i ;
- h_k : é o detentor do sentimento (i.e. quem expressou o sentimento), também chamado de fonte de opinião;
- t_j : é o instante no qual a opinião foi expressa por h_k .

Chamado também por características ou propriedades o conceito de *aspecto*, permite que uma entidade seja observada por diferentes perspectivas ou atributos, ou como uma hierarquia de partes e subpartes [6].

Podemos dizer que uma opinião ou avaliação de um indivíduo sobre algum aspecto ou objeto é um sentimento [4].

Níveis de Análise Textual

Os níveis de análise textual têm-se a decisão de acordo com o contexto e aplicação para a detecção do sentimento em um texto. Esses níveis de análise de sentimento podem ser [4]:

- *Documento*: se um documento em geral expressa um sentimento positivo ou negativo, então a tarefa é classificar esse documento.

- *Sentença*: decide o sentimento de uma sentença específica de um documento. Diferencia sentenças objetivas (fatos – e.g., “comparamos um celular ontem”) e subjetivas (opiniões – e.g., “adoramos o celular comprado ontem”). [9].

- *Entidade ou Aspecto*: nesse nível entende-se que o documento possui diversos aspectos e opiniões. É identificado atributos de uma entidade (produto, pessoas, empresas, etc.) do documento e determina uma opinião para cada atributo, bem como os sentimentos relacionados a ele [10].

Tipos de Opiniões e Análise Linguística

Opiniões são geralmente conteúdos subjetivos que se referem aos sentimentos da população, avaliações, ou sentimentos relacionados com as entidades e suas propriedades [11]. Uma opinião [20]:

- *Regulares*: geralmente o autor expressa um sentimento, atitude, emoção ou percepção sobre um alvo (e.g., “Esta novela é muito boa”).

- *Comparativas*: essas opiniões expressam o sentimento com base na relação de similaridade ou diferença entre duas ou mais entidades, ou preferência quanto a algum aspecto compartilhado (e.g., “O som do toque deste celular é muito melhor do que o celular antigo”).

- *Diretas*: essas opiniões podem ser diretas (e.g., “Esta laranja é muito boa”).

- *Indiretas*: expressam opiniões indiretas (e.g., “minha garganta piorou depois que tomei água gelada”).

- *Implícitas*: essas opiniões expressam diretamente o sentimento (e.g., “Abriu-se um buraco na parede”).

- *Explícitas*: as opiniões expressam indiretamente o sentimento (e.g., “Abriu-se um buraco na parede após a chuva”).

Dependendo da forma expressa, elas influenciam diretamente a habilidade de processá-las corretamente acarretando em problemas e desafios comuns na linguística computacional [4]. Problemas como de co-referência tem sido enfrentado quando diferentes tipos de citações designam a uma mesma entidade. Por exemplo, as expressões “Collor”, “Presidente”, “Presidente Fernando Collor” faz referência a mesma pessoa, devendo ser reconhecidas e unificadas. Outro problema de mesmo contexto é a resolução de pronomes, com a finalidade de relacionar um pronome a uma determinada entidade. Por exemplo, no texto “Rio de Janeiro é uma cidade maravilhosa. Ela é um lugar maravilhoso para conhecer. Suas praias são bem conhecidas”, os pronomes “ela” e “suas” referem-se ao Rio de Janeiro. Esses dois problemas apresentados (co-referência e pronomes) é muito importante para análise de sentimentos nos níveis de aspectos e sentenças, uma vez que os mesmos tratam essa análise de forma isolada (i.e. cada sentença ou opinião), com resultado direto sobre a renovação [1].

Outra questão são as palavras de sentimentos (e.g. péssimo, ótimo). É comum na detecção de opiniões, mas em geral não é suficiente para identificar uma opinião e classificar sua polaridade [1]. Analisando o exemplo a seguir, a sentença “este relógio é muito caro”, a palavra de sentimento “caro” é negativa, por outro lado a sentença “este amigo me é muito caro”, ela é positiva, nessas duas sentenças as palavras de opiniões de acordo com seu contexto foram positivas e negativas. Outra questão é que nem toda opinião é expressa com palavras de sentimento (e.g. “comprei esta blusa de lã na semana passada, e já está toda desfiada”), ou vice versa (e.g. “se achar um bom relógio, vou usá-lo”). Deve-se observar também a negação é outra questão a ser tratada, já que inverte o sentido da opinião (“Esta comida não é nada boa”) [1].

Por fim, outro problema mais difícil de se tratar é a ironia/sarcasmo, por exemplo, “*Hoje assistindo o jornal do esporte, vi uma matéria sobre preços de jogadores: Kaká foi negociado por R\$20.000,00. Por que ninguém comentou isso antes?*”. Presente em alguns domínios como (discussões políticas e esportivas), opiniões sobre arte (filmes, bandas) é muito comum encontrar o uso de sarcasmo [12, 13, 14]. Na literatura trabalhos para identificação de sarcasmo/ironia fazem uso de artifícios: frequência do sinal de exclamação e interrogação, palavras capitalizadas, interjeições (e.g. “ah, oh, yeah”), emoticons (e.g. “;-)”) e superlativos [4,15].

B. Mineração de Opiniões

Em geral, o processo de mineração de opinião consiste basicamente em 3 etapas que demonstram estes processos (Fig.1). Como segue[16]:

- a) Identificação (tópicos, sentenças opinativas);
- b) Classificar a polaridade do sentimento;
- c) Sumarizar.

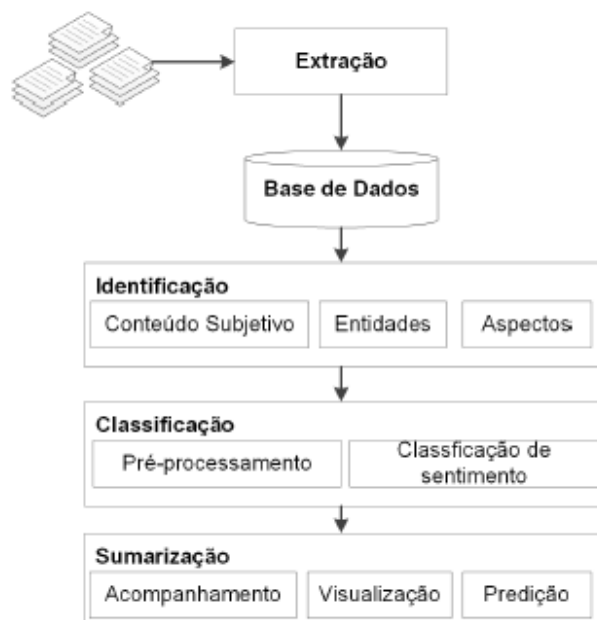


Fig. 1. Etapas da Mineração de Opinião [1]

Identificação

Antes da identificação temos a etapa de coleta de dados na qual é realizado uma busca em fontes diversas (e.g. jornais, redes sociais, plataformas de revisão de produtos/serviços), na sequência é executada a etapa de identificação, que consiste achar os tópicos existentes (seus aspectos), e associá-los com o respectivo conteúdo subjetivo. O procedimento de identificação das entidades, aspectos e sentimentos estão condicionados da granularidade escolhida para análise [16,17].

Classificação de Polaridade

Na medição de polaridade ou orientação da opinião acontece a determinação de uma opinião quanto a valência positiva, negativa. Essa medição de polaridade, conhecida também por *classificação de polaridade*, pode-se considerar como sendo frequentemente um problema de classificação binário, onde a classificação de dados textuais resulta em uma das duas classes já mencionadas anteriormente quanto a *positivo ou negativo*. Para obtermos um nível de detalhes nesses resultados, pode-se ocorrer um desdobramento nas classificações com um grau de intensidade diferente (e.g. muitoPositivo, moderadamentePositivo), ou para um intervalo numérico que seja representado por esse grau de intensidade [16]. Algumas técnicas como *aprendizagem de máquina*, *seleção de palavras* e *análise sintática* ajudam nessa etapa de classificação [1].

Sumarização

Quando obtemos uma única opinião expressa de uma pessoa, não é suficiente para saber da opinião média ou predominante de um grupo de pessoas sobre um determinado tópico/entidade, o mais indicado, é que seja, uma grande quantidade de opiniões necessária para serem analisadas [4]. Para essa questão a importância em termos métricas e

sumários torna-se necessário para que quantificamos as mais diversas opiniões das pessoas referente a um mesmo alvo, sendo assim, o objetivo desta etapa está na criação de métricas para que possam representar o sentimento geral e visualizar os resultados ou disponibilizar para outras aplicações na forma de entrada [1].

Em nosso trabalho proposto, com base nos comentários dos consumidores que realizou a compra do livro “*Cinquenta Tons de Cinza*” e comentários espectadores que assistiram ao filme com o mesmo título, buscamos reconhecermos características importantes de ambos comentários, em seguida classificamos os sentimentos sobre esse aspecto com o uso do *framework PAFRA* desenvolvido por [10]. No capítulo seguinte uma explicação do *framework*.

III. METODOLOGIA

Nosso trabalho foca num estudo exploratório empregando técnicas de mineração de mineração de textos e análise de sentimento, no estudo de caso sobre comentários postados no Twitter sobre o filme “*Cinquenta Tons de Cinza*” e uma base de dados da Amazon de comentários postados do livro. Aplicando essas técnicas, buscou-se investigar as opiniões dos usuários se ele mostra ou não interesse sobre o livro e o filme.

Para realização dos experimentos foi utilizado o *framework* de Análise de Sentimento (*Polarity Analysis Framework – PAFRA*). A ferramenta PAFRA, atua na classificação de polaridade e estados emocionais. Para [10] o entendimento dessa técnica temos que: dado um *tweet* $t_i \in \mathbf{T}$, $\forall i$, onde \mathbf{T} é o corpora (conjunto de todos os *tweets*), t_i é associado a uma classe que pertence ao conjunto $C = \{c_1, c_2, \dots, c_k\}$ de classes, também chamadas rótulos ou categorias. Por meio de um método ou algoritmo de aprendizagem, um classificador aprende uma função γ que mapeia cada *tweet* em uma classe $\gamma: \mathbf{T} \rightarrow C$.

As principais características do framework são [10]:

- Estrutura modular que permite a configuração de cada módulo de acordo com a escolha do domínio de aplicação;
- Combinação de abordagens baseadas em dicionários (conhecimento) e aprendizado de máquina em um regime único para realizar a análise de polaridade;
- Geração automática do conjunto de treinamento para os algoritmos de aprendizado de máquina;
- Foco em mensagens curtas com classificação automática por verificação contextual;

A. Estrutura do Framework PAFRA

A estrutura básica do *framework* PAFRA possuem 5 módulos que são:

- Aquisição de dados – esse módulo obtém textos de duas formas (e.g. arquivos estruturados no modelo da ferramenta Weka (.arff) ou por meio de um *social media crawler*);
- Base de conhecimento – a base de conhecimento é utilizada para classificação a priori de algumas mensagens formando assim um conjunto de treinamento. Essa base pode ser alterada para o idioma que se deseja efetuar a classificação. Isto é, o classificador pode ser configurado para diferentes idiomas de acordo com a base de conhecimento fornecida.

- Classificação contextual – esse módulo recebe as mensagens do módulo de aquisição de dados e o conteúdo da base de conhecimento para iniciar a classificação automática das mensagens;

- Pré-processamento de mensagens – cria uma matriz de dados a partir dos textos, ou seja, transforma um documento não estruturado em uma base estruturada;

- Módulo classificação de sentimento – é responsável por rotular as mensagens não classificadas durante o processo contextual, isto é, classificar os *tweets* cujos rótulos são desconhecidos.

O *framework* também disponibiliza um módulo para que possamos medir os resultados do sistema quando se achar necessário chamado de “módulo de avaliação”. Esse módulo possui um sistema de seleção aleatória (amostragem) de mensagens com objetivo de obter uma amostra da base. Com base nessa amostra classificamos e avaliamos manualmente resposta do sistema. As medidas que o *framework* utiliza para classificar as polaridades são: *acurácia, precisão e medida F*. Essa estrutura pode ser visualizada em (Fig.2).

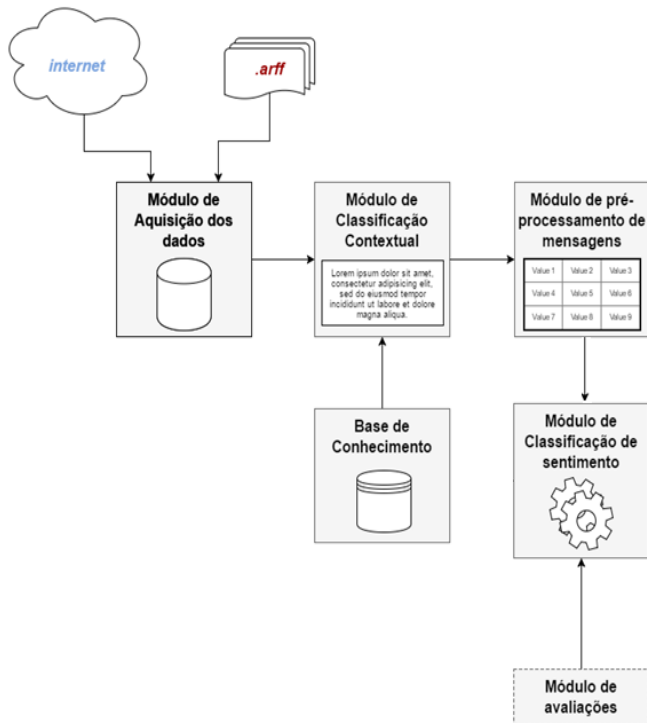


Fig. 2. Estrutura do framework PAFRA [10]

B. Base de Dados

Utilizando o repositório *Machine Learning Repository*, coletou-se uma base de dados de opiniões de livros da Amazon. Dentre essas bases foi escolhido o livro “Cinquenta Tons de Cinza” como estudo de caso. A base possui 15.000 mensagens baixadas no dia 14 de fevereiro de 2017. Em seguida, coletamos do Twitter mensagens, sobre o livro/filme. Foram salvas 577 mensagens do *twitter* coletadas entre os dias 10 e 30 de janeiro de 2017 em horários alternados e com durações variadas de modo a não criar uma tendência na amostra coletada. As bases de dados que foram baixados e

utilizados nessa pesquisa, passaram por uma etapa de pré-processamento de acordo com nossa necessidade nessa pesquisa, mas que disponibilizamos a quem possa interessar.

C. Resultados

Os conjuntos de dados utilizados nos experimentos, contêm mensagens com rótulos positivos e negativos que foram usados para avaliação após a classificação das polaridades. A tabela a seguir apresenta os dados classificados pela ferramenta PAFRA (Tabela 1).

Tabela 1. Base de dados utilizada.

Base	Positivo	Negativo	Total
Amazon	5623	9377	15000
Twitter	529	48	577

Com essa disparidade entre as polaridades, os resultados podem ser direcionados para uma classificação onde a classe tenha mais objetos. As mensagens positivas e negativas das bases utilizadas pode-se observar a proporção na (Fig. 3a e 3b).

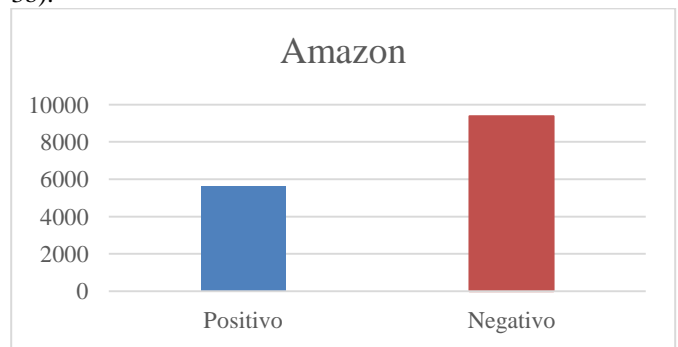


Fig. 3a. Proporção das bases de dados - Amazon (Fonte: O autor).

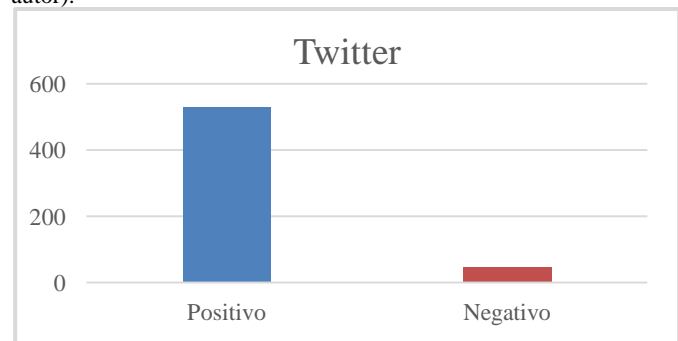


Fig. 3b. Proporção das bases de dados - Twitter (Fonte: O autor).

Para o experimento realizado, o *framework* classificou tudo que podia de forma automática, fez-se um pré-processamento da base e usou a base classificada automaticamente para o treinamento. Com esse treinamento criou-se o modelo de algoritmo e esse modelo foi usado para classificar o que sobrou. Para avaliação dos resultados nos acertos das polaridades dos comentários do livro, foi usada uma matriz de confusão como mostra a tabela 2.

Tabela 2. Polaridade positiva (1) e negativa (-1).

	1	-1
1	3266	2357
-1	3744	5633

Também para a avaliação dos acertos das polaridades positivas e negativas das mensagens postadas no *twitter* referente aos comentários do filme, utilizamos a matriz de confusão como mostra a tabela 3.

Tabela 3. Polaridade positiva (1) e negativa (-1).

	1	-1
1	382	14
-1	147	34

D. Métricas

Um ponto importante para avaliação dos métodos para análise de sentimentos diz respeito às métricas utilizadas. Essas métricas podem ser divididas em três que são bastante usadas na validação da eficiência de um método: acurácia, precisão e revocação [18].

		Predição	
		Positiva	Negativa
Correto	Positiva	a	b
	Negativa	c	d

Acurácia: é o número de classificações corretas dividido pelo número total de classificações.

$$A = \frac{a + d}{a + b + c + d}$$

Precisão: mede a quantidade de objetos recuperados que são relevantes.

$$P(\text{neg}) = d / (b+d)$$

Revocação: a quantidade de objetos relevantes que foram recuperados.

$$R(\text{neg}) = d / (c+d)$$

Outra medida de acurácia muito comum em mineração de dados conhecida por medida-F, ou *score-F*, trata-se de uma medida harmônica entre ambos precisão e a revocação.

$$F(\text{neg}) = \frac{2 * P(\text{neg}) * R(\text{neg})}{P(\text{neg}) + R(\text{neg})}$$

A tabela 4 apresenta os resultados do experimento 2-classes (positivo e negativo) para 2 datasets. O baixo valor de medida F para classe negativa indica que o classificador está com uma tendência a classe com maior proporção devido ao desbalanceamento entre as classes. Neste caso, a base do Twitter contém 529 mensagens positivas e 48 negativas. É importante como trabalho futuro trabalhar com técnicas para classes desbalanceadas.

Tabela 4. Resultados do experimento com 2-classes para 2 datasets

	Dataset	Amazon	Twitter
Sentimento Positivo	Acurácia	59,32	72,09
	Precisão	46,59	72,21
	Revocação	58,08	96,46
Sentimento Negativo	Medida-F	51,3	82,28
	Precisão	70,5	70,83
	Revocação	60,07	18,78
	Medida-F	64,61	28,63

IV. CONCLUSÃO

Este trabalho apresentou um *framework* de Classificação de Sentimentos por polaridade (*positivo, negativo*) para Mineração de Textos capaz de classificar automaticamente essa polaridade em comentários de textos extraídos do Twitter e de um dataset da Amazon. O trabalho também mostrou o diferencial do *framework* destacando suas principais características. O modelo de classificação de sentimentos do *framework* permitiu conhecer a opinião da rede social do usuário do Twitter e de usuário da Amazon sobre o livro e o filme “50 Tons de Cinza”. As postagens dos usuários foram categorizadas neste trabalho em um sentimento: positivo e negativo.

Além disso, nos experimentos realizados, a determinação do sentimento tanto para relação a uma entidade (lista de entidade pré-definidas), quanto para documento geral é permitido pelo *framework* sem fazer uma associação a uma entidade específica.

Trabalhos futuros envolvem aplicar a ferramenta no context de outros idiomas para validar a capacidade de generalização com a troca da base de conhecimento. O próximo passo é trabalhar com textos em português conforme visto em [21]. Nesse trabalho os autores apresentam a classificação de polaridade para *tweets* em português. A base avaliada possui 1700 mensagens e foi usado o OpLexicon como base de conhecimento (palavras anotadas como positivo, negative ou neutron) e Sentilex.

Esse *framework* faz parte da ferramenta que é composta dos seguintes módulos computacionais: Classificação de Sentimento, Classificação de Temperamento e Classificação de Personalidade. Esses três módulos juntos formaliza o conceito de *persona virtual e tríade da persona virtual*. No momento o código da ferramenta não está disponível.

O fruto para esse trabalho teve inspiração no artigo *A polarity analysis framework for Twitter messages, Applied Mathematics and Computation*, aceito em ACM Digital Library [19].

REFERENCIAS

[1] Becker, K, Tumitan, D. Introdução a Mineração de Opiniões: Conceitos, Aplicações e Desafios. Simpósio Brasileiro de Banco de Dados (2013).
 [2] Santos, L.M., Esmín, A.A.A., Zambalde, A.L., & Nobre, F.M. (2011). Twitter, análise de sentiment e desenvolvimento de produtos: Quanto os usuários estão expressando suas opiniões? *Prisma.Com*, (13).

- [3] Cavalcanti, D. C.; Prudêncio, R. B. C.; Pradhan, S. S.; Shah, J. Y. Pietrobon, R. S. Análise de Sentimento em Citações Científicas para Definição de Fatores de Impacto Positivo. Proceedings of the IV International Workshop on Web and Text Intelligence (WTI), p. 1–10, 2012.
- [4] Liu, B. (2012) Sentiment analysis and opinion mining. Morgan & Claypool Publishers.
- [5] Tsytsarau, M. and Palpanas, T. (2012). Survey on mining subjective data on the web. Data Mining and Knowledge Discovery, 24(3):478-514.
- [6] Liu, B. (2010) Sentiment analysis and subjectivity. Handbook of natural language processing, 2:568.
- [7] Abbasi, A; CHEN, H.; Salem, A. Sentiment Analysis in multiple languages: Feature selection for opinion classification in web forums. ACM Transactions on Information Systems (TOIS), ACM, v.26, n.3, p.12, 2008.
- [8] Gonçalves, P. et.al. Comparing and combining sentiment analysis methods. In: ACM. Proceedings of the first ACM conference on Online social network. [S.L.], 2013.p.27-38.
- [9] Thet, T., Na, J., and Khoo, C. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of Information Science, 36(6):823-848.
- [10] Lima, E. S. Ana Carolina (2016). Mineração de Mídias Sociais como ferramenta para a análise da tríade da persona virtual. Tese de doutorado – Universidade Presbiteriana Mackenzie
- [11] Indurkha, Nitin; Damerau, Fred J. Handbook of natural language processing. 2ed. Florida: CRC Press, 2010.666 p
- [12] Sarmiento, L., Carvalho, P., Silva, M., and de Oliveira, E. (2009). Automatic creation of a reference corpus for political opinion mining in user-generated content. In Proceedings of the 1 st international CIKM workshop on Topic-sentiment analysis for mass opinion, pages 29-36. ACM.
- [13] Turney, P.D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th annual meeting on association for computational linguistics, pages 417-424. Association for Computational Linguistics.
- [14] Balahur, A., Kozareva, Z., and Montoyo, A. (2009a). Determining the polarity and source of opinions expressed in political debates. In Computational Linguistics and Intelligent Text Processing, pages 468-480. Springer.
- [15] Carvalho, P. Sarmiento, L., Silva, M.J., and de Oliveira, E. (2009). Clues for detecting irony in user-generated content: oh...!It's so easy;-). In Proceedings of the 1 st international CIKM workshop on Topic-sentiment analysis for mass opinion, pages 56-56. ACM.
- [16] Tsytsarau, M. and Palpanas, T. (2012). Survey on mining subjective data on the web. Data Mining and Knowledge Discovery, 24(3): 478-514.
- [17] Pang, B. and Lee, L. (2008) Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2):1-135.
- [18] Castro, L.N. e Ferreira, D.G. (2016). Introdução a Mineração de Dados: Conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva, 2016.
- [19] Lima, A C E S; de Castro L N; Corchado, JM (2015). A polarity analysis framework for Twitter messages. Applied Mathematics and Computation 270, 756-767
- [20] Liu, B. (2010) Sentiment analysis and subjectivity. Handbook of Natural Language Processing, 2:568.
- [21] Souza M; Vieira R (2012). Sentiment analysis on twitter data for portuguese language. Computational Processing of the Portuguese Language. PROPOR 2012. Lecture Notes in Computer Science, vol 7243. Springer, Berlin, Heidelberg