

Uma abordagem para análise de sentimentos de aplicações da Web em língua portuguesa

ABSTRACT

Sentiment analysis has been applied in many contexts, including user reviews analysis on products and services, trends and financial market moods. Established methods for sentiment analysis present a behavior that varies according to the application and its lexical base, generating different results among them. In this paper, a new sentiment analysis technique called SentiPipe will be presented, which takes the best of a set of methods, generating a less sensitive analysis based context. For such, it was used a real Web database of financial market news in Portuguese to which was applied the new sentiment detection proposal. The results were promising, showing improvement in all metrics analyzed in relation to the methods individually applied.

Categories and Subject Descriptors

J.4 [Computer Applications]: Social and behavioral sciences Miscellaneous; H.3.5 [Online Information Services]: Web-based services

General Terms

Human Factors, Measurement.

Keywords

Análise de Sentimentos, Redes Sociais, Classificação de Texto, Aplicações Web, Web 2.0, Língua Portuguesa

1. INTRODUÇÃO

Muitas das decisões tomadas por indivíduos passam, em algum momento, pela opinião de outros mais experientes ou influenciadores [23] e, em se tratando de escolha, acertar na decisão, é seu principal objetivo.

A *World Wide Web* (WWW) [9] tem sido um dos principais meios de divulgação. Sites de notícias, blogs e redes sociais concentram grande parte da informação produzida atualmente. Devido à imensa popularidade e à grande quantidade de informação compartilhada nesses sistemas, várias aplicações têm surgido na tentativa de extrair opiniões e até mesmo inferir o sentimento público.

O fato das redes sociais e microblogs, como Facebook¹, Instagram² e Twitter³, possuírem uma forma de publicação de informações mais livre e desestruturada que blogs e listas de e-mail tradicionais, está levando muitos usuários a migrarem para este novo formato que, apesar de limitar o conteúdo em textos mais curtos, permitem a inserção de *links* que apontam para notícias completas em outros sites [22].

A análise de sentimento é uma ferramenta popular para a mineração de dados de redes sociais *online* e tem muitas aplicações úteis. Além de possuir características como ser uma pesquisa de opinião de grande escala, não invasiva, rápida, autêntica, barata e automática [20], é comum encontrar opiniões ou comentários de produtos, serviços, eventos e marcas em dados sociais. As pessoas são fortemente influenciadas pelos formadores de opinião [17] que, apesar de serem minoria nas redes sociais, são leitores vorazes seguros de seus argumentos e compartilham o que sabem. Existem vários métodos de análise de sentimentos disponíveis, mas poucos para o idioma português, apesar deste ser o 5º mais usado na Web, de acordo com a *Internet World Stats*⁴ e 3º idioma mais utilizado no Twitter, atrás apenas do Inglês e Japonês [16].

Este trabalho tem como objetivo criar uma nova técnica de análise de sentimentos para o idioma português, extraindo o melhor de métodos já consagrados no idioma inglês. Para tal, são utilizados métodos adaptados para o idioma português, além de modificações em suas saídas para retornar resultados positivos, negativos e neutros além do resultado *miss*, quando o método não encontra qualquer ocorrência em sua base.

Foram utilizadas duas bases de dados reais da Web no idioma português, rotuladas nas classes positivo, negativo e neutro. A nova abordagem, aqui apresentada, gerou resultados superiores à maioria dos métodos individualmente, o que vem ao encontro da solução do problema descrito por Gonçalves [14], quando verificou que os resultados gerados pelos métodos de maneira individual dependem do tipo de base de dados à qual estes métodos são aplicados, já que os mesmos possuem uma base léxica diferente.

Para verificar a eficácia da técnica, foi testado o método de aprendizado de máquina supervisionado [19] Naïve-Bayes Multinomial, que apresenta um dos melhores resultados de classificação de sentimentos. Apesar do método supervisionado superar a nova técnica com resultados um pouco melhores, a proposta aqui apresentada se

¹<http://www.facebook.com/>

²<http://www.instagram.com/>

³<http://www.twitter.com/>

⁴<http://www.internetworldstats.com/stats7.htm>

mostrou superior à medida que a quantidade de dados para treino do método supervisionado diminuía, já que o método de aprendizado de máquina é supervisionado e demanda assim uma base de dados para treinamento. Além disso, a técnica permite facilmente que novos métodos sejam incorporados à mesma, melhorando ainda mais seu desempenho.

As principais contribuições deste trabalho são:

1. Criação de uma nova técnica de análise de sentimentos baseada em métodos adaptados para o idioma português;
2. Os experimentos conduzidos em base real demonstraram significativa melhora em relação aos métodos utilizados individualmente, além de superar o método referência quando disponíveis pouca ou nenhuma base para treino.

O restante deste artigo está organizado da seguinte forma: Na Seção 2 são detalhados alguns trabalhos sobre classificação de sentimento, comparação e combinação de métodos de análise de sentimento. A Seção 3 apresenta a base de dados, métodos de análise de sentimentos utilizados e formaliza a técnica proposta. A contextualização e descrição geral do estudo de caso estão dispostas na Seção 4.1. Os resultados dos experimentos e suas análises são demonstrados na Seção 4. Por fim, a Seção 5 apresenta as conclusões e proposta de trabalho futuro.

2. TRABALHOS RELACIONADOS

Existem muitas técnicas de análise de sentimentos que se aplicam a diferentes propósitos [13]. Algumas com abordagem a nível de palavra, sentença ou documento. À medida que a análise se torna mais genérica, como análise de sentimento a nível de documento, é interessante fazer a análise de aspecto ou objeto, onde a tentativa se volta para um objeto central referenciado em todo texto. Diversas pesquisas em análise de sentimentos [3], [1], [25] utilizam de métodos léxicos, que permitem que sejam aplicadas melhorias como, por exemplo, o uso de conhecimento de linguística. Neste trabalho, o foco será em torno da análise a nível de palavras, por métodos que utilizam técnicas de Processamento de Linguagem Natural (NLP) para classificação de polaridade e técnicas de aprendizado de máquina.

Avanço [3], em seu trabalho, apresenta resultados de classificação de sentimentos em avaliações de produtos tecnológicos na Web escritas em português do Brasil. Utilizou-se 3 métodos de classificação de sentimentos léxicos com diferentes abordagens. Uma delas prioriza a polaridade do sentimento das palavras, outra que considera contextos de negação e outra que acrescenta o tratamento de intensificação do contexto além da negação. Os resultados se mostraram compatíveis e até superiores com os de trabalhos relacionados para o mesmo tipo de base, utilizando os métodos puros.

Combinar métodos de classificação com o intuito de gerar resultados melhores que os obtidos em abordagens individuais tem sido uma técnica bastante utilizada, atingindo bons resultados. Em seu trabalho, Wan [31] combina resultado de técnicas de análise de sentimentos nos idiomas chinês e inglês. Opiniões em chinês são traduzidas para o inglês, extraídas as polaridades e então combinadas com as polaridades obtidas em um método de análise de sentimentos na língua chinesa.

O uso combinado de métodos é abordado por muitos autores [32], [25], [2], [1], onde os resultados de cada método são combinados

em saída única melhorando, por exemplo, a Precisão e a Medida-F1 médios comparados aos métodos utilizados individualmente.

Gonçalves *et al.* [15] utilizam de métodos não supervisionados já consagrados para análise de sentimentos no idioma inglês, como PANAS-t, Emoticons, SentiStrength, SentiWordNet, SenticNet, SASSA e Happiness Index, para a criação de um método combinado que analisa a média harmônica entre Precisão e Revocação para todos os métodos com o objetivo de aumentar a abrangência e melhorando a concordância para um conjunto de dados.

Em seu trabalho, Prabowo [25] propõe um método combinado fundamentado em classificação baseada em regras, técnicas de aprendizado de máquina (SVM) e abordagem de classificação híbrida, onde os classificadores são aplicados em uma sequência, definida pelo grau de Precisão do método quando aplicado à base de testes. Após definidas as regras de classificação, estas foram combinadas em 10 sequências. Os resultados mostraram melhoria nos índices de precisão e revocação.

Abordagem semelhante fez Augustyniak *et al.* [2], quando criaram um método léxico *Bag of Words* (BoW) a partir da combinação de outros léxicos, onde demonstraram que a precisão obtida a partir desta abordagem superou a de outras baseadas em léxicos.

Até onde foi investigado constatou-se que se carece de métodos eficientes para análise de sentimento para o idioma português. Além disso, os métodos existentes para o idioma inglês produzem resultados inferiores quando aplicados a conteúdos na língua portuguesa. Sendo assim, a técnica aqui proposta visa trazer benefícios na utilização combinada de alguns métodos e suas adaptações para conteúdos em português.

3. METODOLOGIA

A seguir serão discutidas a origem da base de dados, os métodos de análise de sentimentos, as métricas de avaliação de desempenho e a construção do arcabouço.

3.1 Métodos

Todos os métodos utilizados para a criação da nova técnica, além de estarem adaptados para o idioma português, foram revisados para garantir as saídas que representem as classes positivo, negativo e neutro. Como não foi desconsiderada a classe dos neutros, no momento que algum método não encontra ocorrência em sua base, o método retorna não ocorrência (*miss*). A seguir serão apresentados os métodos, sua origem e características principais:

SenticNet (SN) - O método SenticNet [30], [7] é um recurso semântico para análise de sentimentos cuja base de conhecimento de senso comum afetiva é gerada a partir da Sentic Computing [6], que é um paradigma que aborda ontologias afetivas e ferramentas de raciocínio de senso comum. A Sentic Computing também envolve técnicas de Web Semântica, Inteligência Artificial, Mineração de Grafos, Linguística, Psicologia, Sociologia e Ética.

SentiStrength (SS) - É baseado em abordagens de aprendizado de máquina, e classifica mensagens combinando métodos de classificação supervisionadas e não-supervisionadas [29].

Happiness Index (HI) - Consiste de uma escala de sentimentos que utiliza o popular *Affective Norms for English Words* (ANEW) [4] e calcula a “quantidade” de felicidade existente em um texto usando

a frequência em que cada palavra do ANEW aparece no texto, computando o peso médio encontrado [10].

SentiWordNet (SWN) - É uma ferramenta de mineração de opinião baseada no léxico WordNet [21]. O SentiWordNet associa agrupamentos de adjetivos, verbos e outras classes gramaticais, chamados de *synset*, pontuações que denotam a polaridade do texto em relação aos sentimentos: positivo, negativo ou neutro [11].

Emoticons (EMO) - Estão cada vez mais presentes em sistemas de mensagens eletrônicas como SMS (mensagens de celulares), microblogs (Twitter), programas de conversa online (WhatsApp⁵) e muitos outros. Neste artigo, são utilizados os mesmos *emoticons* descritos em Araújo [1]. O uso deste recurso se mostrou muito eficiente, gerando resultados de acurácia em torno de 70% em suas predições [26], por isso a escolha deste método para compor o arcabouço.

Para garantir total abrangência da técnica de análise de sentimentos aqui proposta, foi utilizado o classificador de aprendizado de máquina supervisionado **Naïve-Bayes Multinomial (ML)** [24]. Classificadores de aprendizado de máquina constroem um modelo baseado nas probabilidades de cada ocorrência das palavras por classe. Para que seja possível o uso deste classificador, é necessário que a base seja dividida em treino e teste. Após o treino, em cada instância do teste, é calculada uma probabilidade para cada classe e a vencedora é a que faz a rotulação. Este método sempre retorna uma classe, ou seja, sendo assim a abrangência é de 100%.

Com o objetivo de validar a escolha deste método, o mesmo foi treinado e testado em bases distintas, diversas vezes, com validação cruzada, para avaliar sua performance.

3.2 Métricas de Desempenho

Para melhor representação das métricas utilizadas neste trabalho, foi criada a matriz de confusão representada na Tabela 1. As colunas representam classificações realizadas por especialistas.

Para a classe Positivo, são separadas as instâncias corretamente classificadas pelo método (TP - *True Positive*), as classificadas como Negativo (FNpos - *False Negative - Positive*) e como Neutro (Fnpos - *False Neutral - Positive*). O mesmo se repete nas outras colunas para as classes Negativo (FPneg - *False Positive - Negative*), (TN - *True Negative*), (Fnneg - *False Neutral - Negative*) e Neutro (FPneu - *False Positive - Neutral*), (FNneu - *False Negative - Neutral*) e (Tn - *True Neutral*).

		Rotulação humana		
		Positivo	Negativo	Neutro
Rotulação Método	Positivo	TP	FPneg	FPneu
	Negativo	FNpos	TN	FNneu
	Neutro	Fnpos	Fnneg	Tn

Tabela 1: Modelo de matriz de confusão para as classes positivo, negativo e neutro

Aqui são utilizadas medidas tradicionais de desempenho de classificação para múltiplas classes: abrangência (AB), acurácia (A), precisão (P), revocação (R) e medida F1 (F1). Seguindo a notação da Tabela 1, a seguir serão demonstradas as fórmulas para a classe positivo, de cada uma das medidas [27].

⁵<http://www.whatsapp.com/>

A abrangência mostra a fração de mensagens classificáveis por cada método $AB = (\text{mensagens classificáveis} / \text{todas mensagens})$. A acurácia mede o desempenho dos métodos considerando os acertos realizados em relação a todas as classificações e não considera as classes positivas, negativas ou neutras $A = (TP + TN + Tn) / (TP + FP + TN + FN + Tn + Fn)$. A precisão mostra a quantidade de acertos em uma classe, considerando todas as instâncias classificadas como pertencentes à mesma classe $P = (TP) / (TP + FPneg + FPneu)$. Já a revocação mede a relação entre as instâncias classificadas como pertencentes a determinada classe e o total de instâncias que realmente fazem parte desta classe $R = (TP) / (TP + FNpos + FNpos)$. A medida F1 é a média harmônica entre precisão (P) e revocação (R), esta medida é importante para avaliar o desempenho dos classificadores em medida única $F1 = (2 * P * R) / (P + R)$. A medida F1 final é dada pelo macro-F1, que é a média da medida F1 para a classe de mensagens positivas e negativas.

3.3 SentiPipe

Foi construída uma técnica composta por 5 (cinco) métodos de análise de sentimentos que utiliza uma sequência de execução de classificações de acordo com a ordem decrescente da precisão (P) dos métodos escolhidos. Para confirmar a tese de que a precisão representa a melhor alternativa de escolha de quais métodos serão executados primeiro, foram geradas todas as 120 possibilidades de combinações entre os 5 métodos e calculadas as métricas descritas na Seção 3.2.

Desta forma foi possível confrontar os melhores resultados obtidos com o sequenciamento determinado pela ordem de precisão. Assim ficou constatado e definido que, após a análise de desempenho da precisão dos métodos junto à base, os mesmos seriam utilizados, em ordem decrescente, para gerar as classificações em conjunto, técnica chamada SentiPipe.

Como pode ser visto na Tabela 2, dificilmente um dos métodos selecionados para esta versão do SentiPipe possui abrangência de 100%, é comum encontrar instâncias que, apesar de passar pela análise de todos os métodos, não obtiveram sucesso na classificação. Nesta situação, quando todos os métodos retornam *miss* para uma instância, é utilizado um método de abrangência total para finalização das classificações, neste caso o método Aleatório.

A fonte de dados é separada em treino e teste, tratada com a remoção de *stopwords* (palavras que não contribuem com o significado do texto) e pontuações desnecessárias.

Após escolha da ordem de execução dos métodos, definida pela ordem de precisão, cada instância é repassada ao *pipeline*. Assim que o primeiro método recebe a instância, ocorre a tentativa de classificação que, caso ocorra com sucesso, é excluída do processo e armazenada a saída em uma das classes. Caso não seja possível a classificação, o mesmo repassa a instância para o próximo método do *pipeline* que repete o processo até que a instância seja classificada ou terminem os métodos de classificação. Não sendo possível a classificação por nenhum dos métodos da técnica, a classificação é realizada pelo método Aleatório (ALE), o qual garante 100% de abrangência, retornando sempre uma das classes Não Negativo (1) ou Negativo (-1), conforme demonstrado na Figura 1.

A proposta permite facilmente que novos métodos sejam acrescentados e que o método Aleatório seja substituído, por exemplo, por um método de aprendizado de máquina como Naïve-Bayes ou ou-

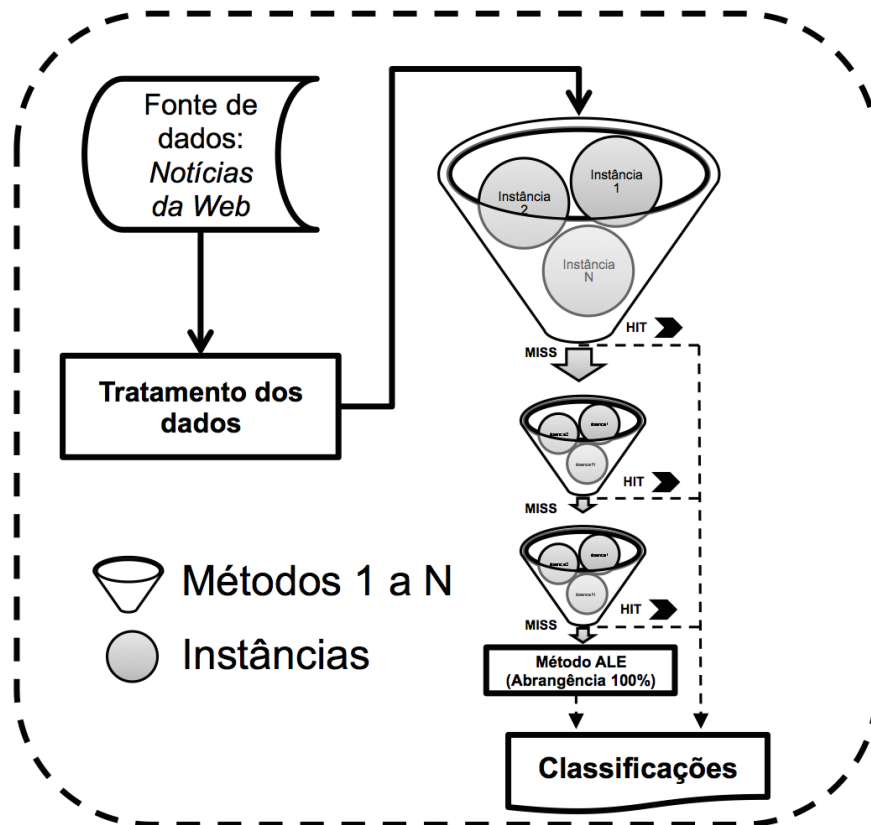


Figura 1: Representação gráfica da técnica de análise de sentimentos SentiPipe

tro método léxico que melhor contribua para as classificações na base em estudo.

4. EXPERIMENTOS

4.1 Estudo de Caso: Aplicações Web em língua portuguesa

Neste trabalho são utilizadas bases de dados reais de notícias do mercado financeiro e tweets sobre automóveis. As notícias foram subdivididas em duas bases, uma contemplando o título e o corpo das notícias e outra base formada somente pelo títulos da mesmas. Foram extraídas 2.132 notícias do mercado financeiro de diversos sites de notícias e 11.027 tweets, notícias relativas ao ano de 2014 e tweets de 2012, rotulados por especialistas no domínio da aplicação como pertencentes a uma das classes positivo (POS), negativo (NEG) ou neutro (NEU).

A base de Tweets foi extraída diretamente da aplicação utilizando a API gratuita (Interface de Programação de Aplicativos) de acesso da mesma, filtrando somente o intervalo de data desejado e as mensagens dos Tweets relacionadas à marca “Fiat” que, na sequência, rotuladas entre uma das 3 classes estudadas.

Os dados de Notícias utilizados foram obtidos através de um processo de coleta e tratamento realizado por uma equipe da UFMG do projeto INWeb (Instituto Nacional de Ciência e Tecnologia para a Web). Este conjunto de dados faz parte de uma base do mercado financeiro, que estão relacionados às empresas que fazem parte do principal índice de ações da BM&FBovespa, o Índice Bovespa.

Os especialistas rotularam as notícias sob o ponto de vista do investidor. Assim, mensagens que influenciam positivamente o mercado financeiro, ou ativos específicos, são classificadas como positivo, as que deixam o investidor inseguro ou demonstre algum revés que afete o mercado recebem o rótulo negativo e as que não representam qualquer polaridade, neutro.

Existem divergências entre diversos autores quando se trata de avaliações neutras, alguns a consideram [5], [2], [11], [30], [8], [12], [18], [28] e outros a ignoram [1], [32].

No trabalho de Cabral [5], que analisou comentários de compradores da plataforma de vendas online eBay⁶, as opiniões classificadas como neutras não foram vistas como tal pelos compradores. Grande parte dos usuários as percebem como opiniões negativas, ou seja, apesar do classificador não ter detectado qualquer polaridade no texto, verifica-se que estas classificações são relevantes e não devem ser descartadas, apesar dos resultados melhorarem com somente duas classes.

Mesmo sabendo que os resultados são um pouco piores utilizando o neutro, já que esta classe é a mais difícil de ser identificada pelos métodos, esta não foi descartada. De acordo com a experiência dos especialistas que rotularam a base de notícias deste estudo de caso, agrupou-se os neutros à classe positivo, formando uma nova classe **não negativo** (NNEG). Além do peso para notícias negativas ser maior que as positivas para o cenário do mercado financeiro, uma notícia neutra pode contribuir mais como sendo uma notícia não

⁶<http://www.ebay.com/>

negativa do que negativa. Mantendo os neutros na base, garante-se maior confiabilidade nos resultados gerados pelas classificações dos métodos em estudo. Retirar os neutros pode mascarar os resultados, já que as opiniões neutras fazem parte do universo das classificações estudadas.

4.2 Resultados

Após aplicar todos os métodos individualmente e a técnica SentiPipe à mesma base de dados, foram calculadas as métricas de avaliação definidas na Seção 3 e seus resultados expostos na Tabela 2.

Para complementar as análises, foi utilizado o resultado do método **Aleatório**, que classifica automaticamente um texto distribuindo igualmente as classes não negativo e negativo dentre o número total de mensagens. As métricas para este método foram obtidas após 30 execuções do mesmo, para cada base, e gerada a média dos resultados para que fique mais confiável.

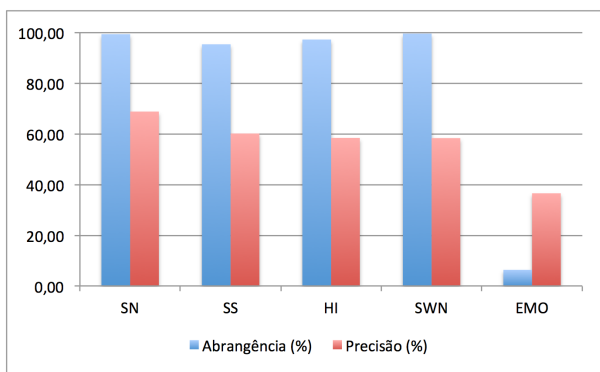


Figura 2: Abrangência dos métodos que compõem a técnica SentiPipe - Notícias

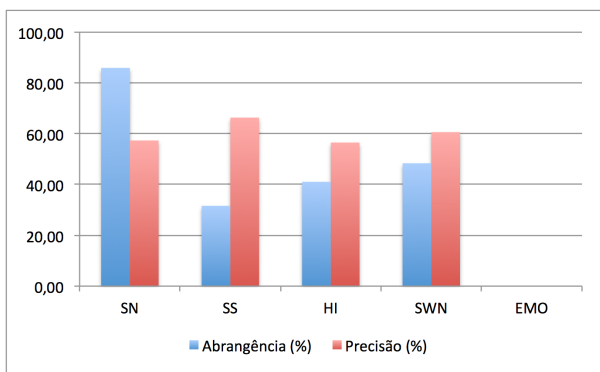


Figura 3: Abrangência dos métodos que compõem a técnica SentiPipe - Títulos das Notícias

Analisando a Tabela 2 é possível verificar que o desempenho dos métodos analisados variam de base para base e, em alguns casos, ficam abaixo até mesmo do método Aleatório. Isto ocorre devido às características da base analisada e da base léxica associada ao método. A base de Títulos possui a mesma quantidade de instâncias da base de Notícias, pois desta última retirou-se o conteúdo (Corpo), reduzindo a quantidade de termos a serem classificados pelos métodos, piorando a Abrangência, como pode ser visto nas Figuras 2, 3 e 4.

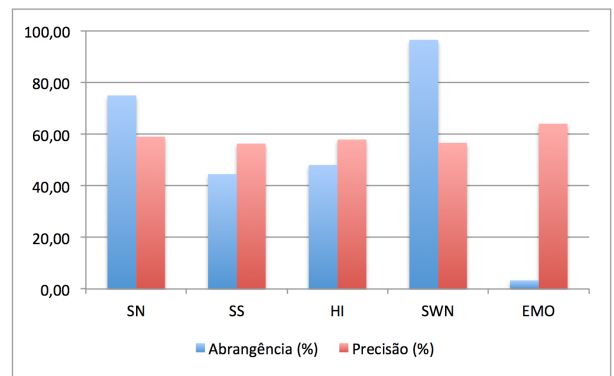


Figura 4: Abrangência dos métodos que compõem a técnica SentiPipe - Tweets

Método SN		Não Negativo	Negativo	Média
Acurácia		-	-	66,19%
Revocação		99,28%	3,42%	51,35%
Precisão		66,11%	71,43%	68,77%
Medida F1		79,37%	6,54%	42,95%
Predição/Rótulo	NNEG	NEG		
	275	141		
	2	5		

Tabela 3: Matriz de confusão para base de notícias - Método SN

Os resultados da Medida F1 da técnica aqui proposta, Tabela 2, se mostraram superiores aos métodos individuais na base de Tweets e Notícias (Títulos), com exceção do método Emoticons que possui boa Precisão, apesar da baixa Abrangência (Figura 2).

Na base de Notícias (Título + Corpo) esta métrica obteve resultados inferiores a quase todos os métodos individuais, mas isto se justifica quando se observa o algoritmo da técnica, que primeiro calcula a ordem de execução do SentiPipe levando em consideração os maiores valores de Precisão dos métodos individuais e em seguida executando-o nesta ordem.

Na Tabela 2 verifica-se que a ordem do método SentiPipe para a base Notícias (Títulos + Corpo) é: 1º SN com 68,77%, 2º SS com 60,12%, 3º HI com 58,38%, 4º SWN com 58,31% e por último o método EMO com 36,54%, como pode ser visto na Tabela 4. Ainda na Tabela 2 constata-se que o método SenticNet (SN) possui Abrangência que se aproxima da totalidade, gerando resultado semelhante a este, já que para 99,30% das instâncias são classificadas pelo primeiro método da série, restando apenas 0,70% da base para ser classificada pelos outros 4 métodos.

Aprofundando ainda mais a análise, na Tabela 3, pode ser observado o comportamento de classificação do método SN dividido em classes. Embora o resultado da Medida F1 da classe NNEG seja próximo a 80% a classe NEG não passa dos 7%, gerando uma média inferior aos demais métodos e, como dito anteriormente, resultando em dados não satisfatórios para a técnica SentiPipe.

Os métodos de aprendizado de máquina dependem de uma base treino para criação de um modelo de classificação. Sem este modelo é impossível que a classificação ocorra [26].

Para verificar o grau de dependência de base treino pelo método ML foi feita a redução da proporção entre o treino e teste aplica-

	Base	Abrangência(%)	Acurácia(%)	Revocação(%)	Precisão(%)	Medida F1(%)
SenticNet	Tweets	74,93	71,81	57,06	58,98	57,51
	Notícias	99,30	66,19	51,53	68,77	42,95
	Títulos	85,85	65,11	54,45	57,31	53,31
SentiStrength	Tweets	44,47	52,29	57,85	56,30	51,17
	Notícias	95,31	46,80	56,59	60,12	45,17
	Títulos	31,60	62,69	63,89	66,30	61,65
Happiness Index	Tweets	48,01	72,05	55,74	57,86	56,05
	Notícias	97,18	65,22	53,08	58,38	49,29
	Títulos	41,04	59,20	54,98	56,50	53,89
SentiWordNet	Tweets	96,46	69,41	57,50	56,60	56,88
	Notícias	99,53	60,85	58,86	58,31	58,34
	Títulos	48,35	30,43	30,89	60,59	40,45
Emoticons	Tweets	3,35	67,57	70,26	63,99	63,00
	Notícias	6,34	70,37	47,50	36,54	41,30
	Títulos	0,00	0,00	0,00	0,00	0,00
Aleatório	Tweets	100,00	53,39	50,03	50,04	49,82
	Notícias	100,00	53,33	48,10	47,99	47,59
	Títulos	100,00	58,35	52,32	52,53	52,24
Naïve-Bayes	Tweets	100,00	62,51	68,19	62,07	70,88
	Notícias	100,00	71,36	68,83	68,61	68,71
	Títulos	100,00	76,65	72,66	74,54	73,36
SentiPipe	Tweets	100,00	74,07	57,74	58,94	58,17
	Notícias	100,00	66,20	51,99	71,86	44,06
	Títulos	100,00	66,75	63,02	63,23	63,12

Tabela 2: Métricas de todos os métodos nas 3 bases

dos ao método. Por padrão foi escolhido 80% de treino e 20% de teste para execução do método referência e o mesmo teste aplicado aos demais métodos. Em seguida a relação treino/teste foi reduzida para (20%/80%), (05%/95%) e (01%/99%), como demonstrado na Tabela 4. Verifica-se que a Medida F1 do método ML decai sensivelmente à medida em que é reduzida a disponibilidade de treino para efetuar as classificações, enquanto o SentiPipe permanece estável. O mesmo pode ser observado nas Tabelas 5 e 6, representadas nas Figuras 5, 6 e 7.

Treino/Teste (%)	Naïve-Bayes (%)	SentiPipe (%)	Sequência Execução
80/20	68,71	44,06	SN->SS->SWN->HI->EMO
20/80	69,48	59,00	SWN->HI->SS->SN->EMO
05/95	63,35	59,35	SWN->HI->SS->SN->EMO
01/99	48,26	59,46	SWN->HI->SS->SN->EMO

Tabela 4: Variação da Medida F1 relativa à proporção Treino/Teste - Base Notícias Completo

Treino/Teste (%)	Naïve-Bayes (%)	SentiPipe (%)	Sequência Execução
80/20	73,36	63,12	SS->SWN->SN->HI->EMO
20/80	68,78	66,61	SS->SWN->HI->SN->EMO
05/95	61,47	65,47	SS->SWN->HI->SN->EMO
01/99	48,77	64,96	SS->SWN->HI->SN->EMO

Tabela 5: Variação da Medida F1 relativa à proporção Treino/Teste - Base Notícias Título

Treino/Teste (%)	Naïve-Bayes (%)	SentiPipe (%)	Sequência Execução
80/20	70,88	58,17	EMO->SN->HI->SWN->SS
20/80	68,39	63,92	SS->EMO->HI->SN->SWN
05/95	57,83	63,86	SS->EMO->HI->SN->SWN
01/99	52,46	63,81	SS->EMO->HI->SN->SWN

Tabela 6: Variação da Medida F1 relativa à proporção Treino/Teste - Base Tweets

Apesar da nova abordagem não superar o método de aprendizagem de máquina quando se dispõe de treino, ele melhora de forma significativa os outros métodos a partir do resultado combinado, permite que sejam analisados dados sem que haja necessidade de

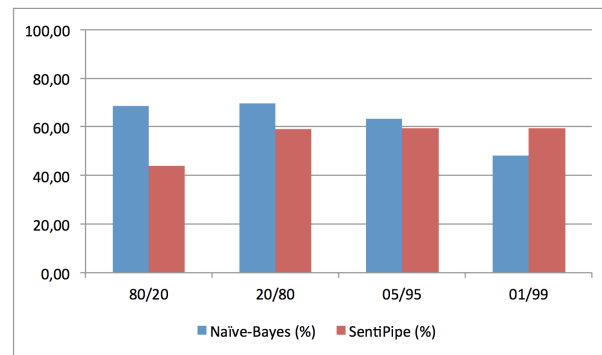


Figura 5: Desempenho da Medida F1 entre métodos Naïve-Bayes e SentiPipe - Notícias

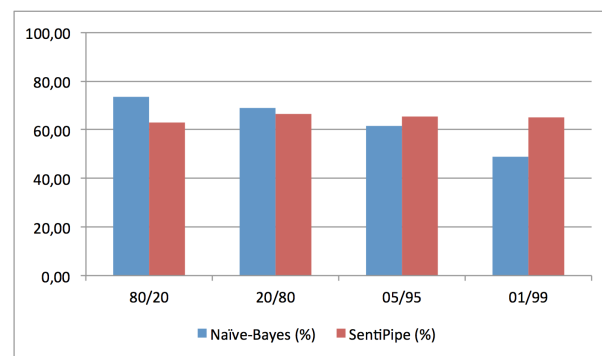


Figura 6: Desempenho da Medida F1 entre métodos Naïve-Bayes e SentiPipe - Títulos das Notícias

treino, abrindo a discussão para novos trabalhos que utilizem técnica semelhante.

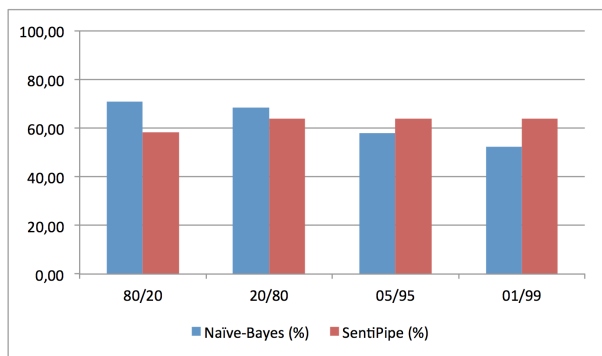


Figura 7: Desempenho da Medida F1 entre métodos Naive-Bayes e SentiPipe - Tweets

5. CONCLUSÕES

Recentes pesquisas utilizam métodos de análise de sentimentos para extrair informações relevantes de comentários e notícias na Web. Apesar dos bons resultados apresentados por esses métodos, ainda é limitada a detecção de sentimentos no idioma português.

Os métodos de análise de sentimentos possuem desempenho que varia de acordo com a base a ser classificada e a base lexical do método em questão. Neste trabalho, foi apresentada uma nova abordagem para a melhoria da detecção de sentimentos em textos no idioma português a partir da criação de uma técnica que utilize métodos consagrados de análise de sentimentos em conjunto.

Aplicando a técnica a bases da Web no idioma português, foram obtidos resultados superiores aos métodos individuais, originalmente criados para o idioma inglês. Em situações onde a nova abordagem não dispõe de treino, os resultados foram ainda melhores, superando inclusive um método de aprendizado de máquina, além de atingir 100% de Abrangência das classificações.

Foi visto que, utilizando métodos de base léxica individualmente, as classificações realizadas pelo mesmo método mudam drasticamente de uma base para outra. A técnica SentiPipe reduziu a sensibilidade na análise de sentimentos, gerando resultados mais homogêneos e de qualidade.

Como trabalho futuro, pretende-se investir na tentativa de identificação de palavras chave no corpo das notícias para melhorar a classificação e analisar questões de linguística específicas do idioma português.

6. REFERENCIAS

- [1] M. Araújo, P. Gonçalves, F. Benevenuto, and M. Cha. Métodos para análise de sentimentos no twitter. In *Proceedings of the 19th Brazilian symposium on Multimedia and the Web (WebMedia'13)*, 2013.
- [2] L. Augustyniak, T. Kajdanowicz, P. Szymanski, W. Tuligłowicz, P. Kazienko, R. Alhajj, and B. Szymanski. Simpler is better? lexicon-based ensemble sentiment classification beats supervised methods. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, pages 924–929. IEEE, 2014.
- [3] L. V. Avanco and M. d. G. V. Nunes. Lexicon-Based Sentiment Analysis for Reviews of Products in Brazilian Portuguese. In *2014 Brazilian Conference on Intelligent Systems*, pages 277–281, 2014.
- [4] M. M. Bradley and P. J. Lang. Affective norms for English words ({ANEW}): Stimuli, instruction manual, and affective ratings. Technical report, Center for Research in Psychophysiology, University of Florida, Gainesville, Florida, 1999.
- [5] L. Cabral and A. Hortaçsu. The dynamics of seller reputation: Evidence from ebay. *Journal of Industrial Economics*, 58(1):54–78, 2010.
- [6] E. Cambria and A. Hussain. *Sentic computing*. Springer, 2012.
- [7] E. Cambria, A. Hussain, C. Havasi, C. Eckl, and J. Munro. Towards Crowd Validation of the UK National Health Service. In *ACM Web Science Conference (WebSci)*, 2010.
- [8] D. C. Cavalcanti, R. B. C. Prudêncio, S. S. Pradhan, J. Y. Shah, and R. S. Pietrobon. Análise de Sentimento em Citações Científicas para Definição de Fatores de Impacto Positivo. In *Proceedings of the IV International Workshop on Web and Text Intelligence (WTI)*, pages 1–10, 2012.
- [9] P. Dias. Hipertexto, hipermedia e media do conhecimento: representação distribuída e aprendizagens flexíveis e colaborativas na web. 2000.
- [10] P. S. Dodds and C. M. Danforth. Measuring the happiness of large-scale written expression: songs, blogs, and presidents. *Journal of Happiness Studies*, 11(4):441–456, 2009.
- [11] A. Esuli, F. Sebastiani, and V. G. Moruzzi. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. *Proceedings of LREC 2006*, pages 417–422, 2006.
- [12] T. R. Evangelista and T. P. P. Padilha. Monitoramento de Posts Sobre Empresas de E-Commerce em Redes Sociais Utilizando Análise de Sentimentos. 2013.
- [13] R. Feldman. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89, 2013.
- [14] P. Gonçalves and M. Araújo. Comparing and combining sentiment analysis methods. *Proceedings of the first ACM conference on Online social networks*, 2013.
- [15] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha. Comparing and combining sentiment analysis methods. In *Proceedings of the 1st ACM Conference on Online Social Networks (COSN'13)*, 2013.
- [16] L. Hong, G. Convertino, and E. H. Chi. Language Matters in Twitter: A Large Scale Study. In *Proceedings of the Fifth International AAI Conference on Weblogs and Social Media*, number 1, pages 518–521, 2011.
- [17] E. Keller and J. Berry. One American in ten tells the other nine how to vote, where to eat, and what to buy. *They are The Influentials, New York*, 25(5):1–8, 2003.
- [18] F. H. Khan, U. Qamar, and M. Y. Javed. Sentiview: A visual sentiment analysis framework. In *Information Society (i-Society), 2014 International Conference on*, pages 291–296. IEEE, 2014.
- [19] B. Liu, X. Li, W. S. Lee, and P. S. Yu. Text classification by labeling words. In *AAAI*, volume 4, pages 425–430, 2004.
- [20] Y. Malheiros and G. Lima. Uma Ferramenta para Análise de Sentimentos em Redes Sociais Utilizando o SenticNet. *Simpósio Brasileiro de Sistemas de Informação*, IX:517–522, 2013.
- [21] G. A. Miller. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.

- [22] A. Pak and P. Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In N. C. C. Chair, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, editors, *International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may 2010. European Language Resources Association (ELRA).
- [23] B. Pang and L. Lee. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval* 2, 1-2(1), 2008.
- [24] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, July 6-7, 2002, Philadelphia, Pennsylvania, USA*, pages 79–86, 2002.
- [25] R. Prabowo and M. Thelwall. Sentiment analysis: A combined approach. *Journal of Informetrics*, 3(2):143–157, 2009.
- [26] J. Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop*, pages 43–48. Association for Computational Linguistics, 2005.
- [27] M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009.
- [28] J. Spencer and G. Uchyigit. Sentimentor: Sentiment analysis of twitter data. In *CEUR Workshop Proceedings*, volume 917, pages 56–66, 2012.
- [29] M. Thelwall. Heart and soul: Sentiment strength detection in the social web with sentistrength. *Cyberemotions*, pages 1–14, 2013.
- [30] F. K. Uk, E. Cambria, R. Speer, and C. Havasi. SenticNet : A Publicly Available Semantic Resource for Opinion Mining. *AAAI Fall Symposium (FS-10-02)*, pages 14–18, 2010.
- [31] X. Wan. Using Bilingual Knowledge and Ensemble Techniques for Unsupervised Chinese Sentiment Analysis. In *Proceedings of the Conference on Empirical Methods*, number October, pages 553–561, 2008.
- [32] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu. Sentiment classification: The contribution of ensemble learning. *Decision Support Systems*, 2014.