

Análise de Sentimentos: Comparando o uso de ferramentas e a análise humana

Alternative Title: Sentiment Analysis: Comparing the use of tools and the human analysis

Vanessa de Sales Moreira, Sean Wolfgang Siqueira, Leila Andrade, Mariano Pimentel

Departamento de Informática Aplicada - UNIRIO
CEP: 22290-240 – Rio de Janeiro – RJ - Brasil
{vanessa.moreira, sean, leila, pimentel}@uniriotec.br

RESUMO

Análise de Sentimentos emergiu a partir da necessidade de se tratar e avaliar textos, opiniões e comentários realizados por usuários na Internet, no intuito de compreender como eles se relacionam com uma determinada entidade. Diversos métodos de análise têm sido desenvolvidos na tentativa de melhor traduzir as incertezas e a subjetividade dos sentimentos humanos. Esta pesquisa utilizou os dados de interações de alunos em uma rede social para identificar os sentimentos envolvidos em cada comentário. A partir dos textos coletados foram usadas análises de sentimento manual e automática e os resultados encontrados foram comparados com a avaliação realizada pelos autores dos comentários, a fim de identificar qual método mais aproximava dos reais. A comparação realizada indicou diferenças entre os resultados obtidos e o real: enquanto as ferramentas utilizadas classificaram mais de 40% dos comentários como neutros, a análise dos autores das mensagens indicou que 71% dos comentários eram positivos. Na classificação foram identificados *outliers* e divergências na categorização da intensidade dos sentimentos por cada método. Nenhuma abordagem, incluindo a realizada pelo pesquisador, foi considerada suficientemente eficiente, uma vez que o maior nível de acurácia obtido foi inferior a 70%.

Palavras-Chave

Análise de sentimentos, Redes Sociais Online, Emoções, Sistemas de Informação Cognitivos

ABSTRACT

Sentiment Analysis emerged from the need to treat and evaluate texts, opinions and comments made by users on the Internet, in order to understand how they relate to a given entity. Several analytical methods have been developed in an attempt to better translate the uncertain and the subjectivity of human feelings. This research used data from interactions in a social network to identify the sentiments involved in each student comment. From the collected text different methods of sentiment analysis were used in order to identify which method had better results compared to the real ones. The comparison showed differences between these results and the real ones: while the used tools classified more than 40% of the comments

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SBSI 2016, May 17–20, 2016, Florianópolis, Santa Catarina, Brazil.

Copyright SBC 2016.

as neutral, the analysis of the messages' author showed that 71% of the comments were positive. In the classification, the occurrence of outliers was identified as well as differences on the intensity of the sentiments acquired with each method. No approach, including the one performed by one of the researchers, was considered efficient enough as the highest level of accuracy obtained was less than 70%.

Categories and Subject Descriptors

K.3 [Computers and Education]

General Terms

Human Factors

Keywords

Sentiment analysis, Online Social Networks, Emotions, Cognitive Information Systems

1. INTRODUÇÃO

Com a evolução das redes sociais, os usuários deixaram de ser apenas receptores passivos de informações para se tornarem os reais responsáveis pela produção, compartilhamento, interação e avaliação de todo o conteúdo produzido na Web [11]. Apesar do grande volume de dados produzido por essas interações, quase 80% dos dados gerados em redes sociais são considerados não estruturados, o que torna difícil a análise e a obtenção de informações úteis a partir destes [2].

A análise de sentimentos surgiu na tentativa de melhor compreender de que forma essas interações ocorriam, através da análise de opiniões, sentimentos e emoções de pessoas com relação a produtos, serviços, organizações, indivíduos e seus atributos [20]. Redes sociais como o Twitter¹ e o Facebook² têm servido como base de dados para estudos de aplicação da análise de sentimentos na predição de resultados de eleições [9], na identificação de taxas de depressão entre os usuários [19], na medição de níveis de satisfação dos consumidores [10] e no monitoramento da alteração de sentimentos de estudantes a partir de suas interações [15]. Na Educação, a avaliação dos sentimentos dos alunos serve como insumo para que professores e tutores trabalhem de forma a melhor conduzir suas aulas e a disseminação do conteúdo disponibilizado a fim de atrair e manter a atenção dos alunos [17], [15].

Apesar dos diversos esforços feitos em prol da melhoria dos métodos de análise, muito ainda precisa ser realizado para que, de fato, o nível de acurácia se aproxime da forma humana de avaliar sentimentos. Grande parte das dificuldades atualmente existentes na análise de sentimentos diz respeito à linguagem dos textos avaliados [13], [12], [1], à detecção de ironia [5] e à

¹ <http://www.twitter.com>

² <http://www.facebook.com>

dificuldade de tratar a subjetividade dos textos [18]. Em avaliação comparativa dos diversos métodos de análise de sentimentos existentes também é possível identificar divergências e lacunas entre as formas de avaliação [4], levando a crer que nem sempre um determinado método é adequado para todos os tipos de domínio.

Este trabalho tem como objetivo principal a realização de uma análise dos dados extraídos de uma plataforma de rede social online (no caso, um grupo do Facebook) a fim de identificar o teor afetivo dos comentários de alunos durante as discussões no grupo. Os dados extraídos foram analisados de acordo com duas diferentes abordagens: (i) automática – ferramentas de análise de sentimentos e (ii) manual – análise por um dos pesquisadores. Nestas abordagens, os comentários foram analisados de acordo com sua polaridade e sentimentos expressados. A partir destes resultados, foi estabelecida uma análise comparativa com os resultados da análise manual dos autores das mensagens com o objetivo de identificar a quantidade de acertos de cada abordagem, bem como sugerir possíveis motivos para as falhas de avaliação.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta alguns dos trabalhos na área de análise de sentimentos envolvendo a aplicabilidade do estudo no campo da educação e a comparação entre os mais diversos métodos de análise. Na Seção 3 são descritos brevemente os métodos de análises manuais utilizados durante a execução do estudo de caso, enfatizando a forma como foi realizada a categorização das polaridades dos textos. A Seção 4 descreve como foi realizado o estudo de caso, quais as abordagens seguidas, as limitações da pesquisa e a metodologia utilizada. Na Seção 5 são apresentados os resultados e discussões e, por fim, na Seção 6 são apresentadas as conclusões e trabalhos futuros.

2. TRABALHOS RELACIONADOS

Diversos trabalhos têm sido realizados em direção a um melhor entendimento sobre como deve ser feita a análise de sentimentos de textos oriundos de redes sociais. Em [2] foi realizada uma *survey* para avaliação do estado da arte do campo de análise de sentimentos onde foram identificados os níveis, técnicas e métodos atualmente existentes de categorização afetiva para as mais diversas aplicações. O estudo realizado se manteve apenas no âmbito descritivo sem, de fato, realizar uma comparação entre as técnicas de análise de forma que fosse possível estabelecer qual técnica possuiria a melhor eficácia.

Dentre os métodos para classificação de sentimentos, Araújo et al. [4] utilizaram 8 destes para realização de uma análise comparativa a fim de, através de diversas métricas, avaliar a eficiência e eficácia de cada um na determinação da polaridade dos textos. A partir desta avaliação, foi proposta uma ferramenta denominada *iFeel* que, através da combinação dos métodos, buscou uma melhora da abrangência na detecção dos sentimentos, alcançando um total de 95%.

O contexto da análise de sentimentos orientada à Educação também tem sido explorado. O entendimento da opinião dos alunos como pré-requisito para um desenvolvimento efetivo de sistemas de aprendizagem foi estudado através de métodos estatísticos, de ganho de informação (IG) e de informação mútua (MI), buscando formas de auxiliar desenvolvedores a melhorar e promover a qualidade dos serviços de aprendizagem oferecidos [14]. Em [15] foi proposto um método para análise de sentimentos no Facebook capaz de classificar as mensagens dos alunos de acordo com sua polaridade e detectar mudanças de humor através da combinação de técnicas de aprendizado de máquina e léxicas. A identificação dessas características foi realizada no intuito de auxiliar sistemas adaptativos de aprendizagem. A detecção dos sentimentos dos alunos também foi estudada por [3] que enfatizou a necessidade de compreender o comportamento dos

alunos e identificar sentimentos como tédio e confusão com o objetivo de aprimorar o ensino. Através de classificadores como SVM e *Naive Bayes*, os autores foram capazes de realizar uma melhor predição com determinados classificadores do que outros.

Ao realizar a análise dos trabalhos apresentados, é possível perceber que novos métodos e técnicas surgem com frequência para apoiar ou facilitar a tarefa de análise de sentimentos. Também é observável a importância de plataformas de redes sociais online como fonte de informações, sobretudo como um meio de apoio à educação. A ferramenta *iFeel* surgiu na tentativa de minimizar lacunas existentes entre os diferentes métodos de análise.

De forma semelhante, este trabalho busca realizar uma análise das divergências entre os métodos, diferenciando-se ao agir na identificação de lacunas tendo como base de comparação as avaliações reais, realizadas pelos próprios autores, através de um gabarito que aponte qual método mais aproximou seus resultados. A partir dessa análise, espera-se conseguir obter indícios que auxiliem pesquisadores na criação de melhorias para os métodos de classificação no intuito de poder, posteriormente, fornecer a professores e tutores informações sobre o sentimento e interesse dos alunos.

3. ANÁLISE AUTOMÁTICA

Essa seção busca descrever as principais ferramentas de análise automática para categorização dos sentimentos, listando suas principais características.

A definição de ferramentas aplicáveis ao contexto do estudo deu-se, sobretudo, a partir da avaliação de quais seriam mais eficientes no tratamento de detalhes relativos à escrita de textos em ambientes informais, como redes sociais, uma vez que estes tendem a possuir características específicas, como o uso de *emoticons* e uma forma de escrita peculiar. Também foi considerado o fato das ferramentas possuírem documentação para interpretação dos resultados obtidos e estarem disponíveis na Internet na forma de aplicativos ou páginas Web.

3.1 SenticNet

O SenticNet³ é uma ferramenta de análise semântica utilizada para tratar os sentimentos de um determinado texto a partir de conceitos oriundos da *Sentic Computing*. A partir de uma fonte de dados contendo mais de 30.000 entendimentos de senso comum, as características emotivas de um texto são analisadas de acordo com três tipos de categorização diferentes: (1) Polaridade, (2) *Sentic Level* ou Nível Sêntico e (3) Semântica.

Polaridade

A partir do SenticNet é possível categorizar uma palavra como positiva ou negativa de acordo com o contexto no qual ela se encontra. Dessa forma, cada palavra de um determinado texto que possui uma relação com algum dos conceitos de senso comum expressos pela Sentic Computing foi medida e sua polaridade definida: emoções classificadas com polaridades no intervalo de $[-1,0[$ foram consideradas como emoções **negativas**, ao passo que emoções no intervalo de $]0,1]$ foram consideradas como **positivas**. Emoções que possuíam um valor de polaridade igual a 0 foram tidas como **neutras**. O cálculo da polaridade de uma frase foi realizado através da média das polaridades de todas as palavras que a compõe.

Nível Sêntico

O nível sêntico (ou *sentic level*) de um texto pode ser obtido através da utilização do modelo denominado *Hourglass of Emotion* [7], um modelo de categorização afetiva que organiza todas as palavras do texto que possuem relação com um

³Disponível em <http://sentic.net/>

conceito de senso comum de acordo com 4 dimensões básicas: *Pleasantness*, *Attention*, *Sensitivity* e *Aptitude*⁴.

Seguindo esse modelo, a cada palavra ou expressão são atribuídos valores no intervalo de [-1,1] em cada uma das dimensões. A combinação desses valores e dos níveis alcançados através deles representou o tipo de sentimento expresso pelo texto. A tabela 1 apresenta os sentimentos mais básicos que podem ser inferidos a partir da análise do nível sêntico de um texto.

Tabela 1. Níveis Sênticos (nível 01)
Fonte: [24]

Intervalo	Pleasantness	Attention	Sensitivity	Aptitude
[G(1), G(2/3)]	Ecstasy	Vigilance	Rage	Admiration
[G(2/3), G(1/3)]	Joy	Anticipation	Anger	Trust
[G(1/3), G(0)]	Serenity	Interest	Annoyance	Acceptance
[G(0), -G(1/3)]	Pensiveness	Distraction	Apprehension	Boredom
[-G(1/3), -G(2/3)]	Sadness	Surprise	Fear	Disgust
[-G(2/3), -G(1)]	Grief	Amazement	Terror	Loathing

Semântica

Na abordagem proposta pela Sentic Computing, a semântica de determinada palavra ou expressão representa o conjunto dos possíveis significados desta palavra ou dos contextos nos quais ela se aplica. Nesse trabalho, não foi realizada a análise a respeito da semântica dos textos, uma vez que esta não se enquadra diretamente na proposta da avaliação comparativa.

3.2 SentiStrength

O SentiStrength⁵ é uma ferramenta de mineração de opinião e análise de sentimentos que realiza uma análise automática de textos utilizando como base para classificação mais de 16.000 textos previamente computados e classificados.

A forma de categorização dos sentimentos pela SentiStrenght baseia-se na teoria psicológica de que a distinção entre o positivo e o negativo ocorre através de um processamento paralelo realizado pelo ser humano, o que faz com que seja possível que um mesmo texto possua tanto um teor negativo quanto positivo. Dessa forma, a análise pela ferramenta classifica os textos através de atribuição de valores tanto no intervalo de [-5,-1] para a categoria negativa quanto no de]1,5] para a categoria positiva, onde os valores -1 e 1 correspondem à classificação neutra.

Nesse estudo a obtenção da polaridade total da frase foi realizada a partir da soma dos valores obtidos. Além da classificação usual feita pela ferramenta, o estudo também considerou que em casos onde a média das polaridades positiva e negativa fosse igual a zero, o texto seria considerado neutro.

3.3 LIWC

O LIWC⁶ é uma ferramenta de análise textual que realiza uma avaliação de cada uma das palavras contidas em um texto tendo como base um dicionário pré-definido pelo usuário. Essa avaliação consiste na tentativa de relacionar palavras a pelo menos uma das categorias definidas, tendo como base teorias da esfera psicológica, onde cada relação entre uma palavra e uma categoria indica uma maior tendência do usuário a uma determinada personalidade ou atitude. Na versão paga da ferramenta são disponibilizadas para avaliação mais de 100 categorias de palavras, mas nesse trabalho foi utilizada

somente a versão livre, que listava os sentimentos de acordo com apenas as 5 categorias apresentadas na tabela 2.

Com o objetivo de identificar apenas a polaridade dos textos, essa pesquisa optou por analisar exclusivamente as categorias referentes à expressão de emoções positivas (*Positive Emotion Words*) e negativas (*Negative Emotion Words*). Os valores atribuídos pelo método nestas categorias variaram entre 0 e 100, considerando que a soma entre os valores obtidos nestas duas categorias deveria respeitar sempre o valor máximo de 100. A polaridade total do texto foi obtida através da diferença entre os dois valores, onde valores muito próximos ou iguais a 0 foram considerados neutros e valores abaixo de 0 negativos.

Tabela 2. Categorias de análise do método LIWC

Fonte: <http://www.utpsyc.org/TAT/LIWCTATresults.php>

Categoria	Descrição
<i>I- Words (eu, meu)</i>	Pessoas que usam um grande número de autorreferências tendem a ser mais inseguras, nervosas, honestas e possivelmente depressivas
<i>Social Words (nós, conversar)</i>	Palavras sociais são palavras que fazem referência a outras pessoas
<i>Positive Emotion Words (feliz, amor, bom)</i>	Quanto mais as pessoas usam palavras positivas, mais otimistas elas tendem a ser
<i>Negative Emotion Words (triste, matar, medo)</i>	O uso de palavras negativas é fracamente relacionado aos níveis de ansiedade e neurose das pessoas
<i>Cognitive Words (pensando, porque, imaginando, conhecimento)</i>	Essas palavras refletem o quanto as pessoas estão ativamente pensando sobre o tópico que estão escrevendo

4. ESTUDO DE CASO

Essa seção define as etapas seguidas durante a aplicação do estudo de caso que teve como objetivo principal a extração de comentários, a avaliação dos sentimentos e suas polaridades e a comparação entre os diferentes resultados obtidos.

4.1 Definições Preliminares

O estudo de caso foi realizado com alunos do grupo do Facebook intitulado Sistemas Colaborativos - 2015.2, utilizado como apoio à disciplina homônima ministrada em um curso de pós-graduação stricto sensu. Foram coletadas informações a respeito das interações textuais (publicações e comentários) realizadas pelos alunos nas datas de 14 de agosto de 2015 a 1º de outubro do mesmo ano. No total foram analisados 84 comentários de 7 autores diferentes. Comentários realizados pelo professor, tutores e alunos ouvintes não foram considerados nesse estudo.

Os comentários extraídos foram analisados a partir de duas abordagens diferentes: (i) análise automática (ferramentas para análise de sentimentos) e (ii) análise manual pelo pesquisador.

4.2 Limitações da Pesquisa

O estabelecimento de um método comparativo entre as abordagens anteriormente definidas possui uma série de limitações. Dentre as dificuldades encontradas a que foi considerada mais limitante foi a impossibilidade de, através das ferramentas, analisar os dados em sua forma bruta, sendo então necessário primeiramente traduzi-los para a língua inglesa para que as informações se mantivessem de acordo com o tipo de dado de entrada esperado pelas ferramentas. A tradução foi realizada de forma manual na tentativa de

⁴Neste trabalho, optou-se por não realizar a tradução das dimensões a fim de manter o contexto semântico

⁵Disponível em <http://sentistrength.wlv.ac.uk>

⁶Disponível em <http://liwc.wpengin.com/>

minimizar possíveis erros e perda da coerência das frases, porém o impacto da perda de contexto e da perda de significância de características linguísticas como expressões, gírias e regionalismos, se manteve como um fator limitante.

Para geração do gabarito, foi disponibilizado um questionário para avaliação manual dos autores das mensagens (alunos). Neste questionário foi definida uma escala de valores onde para cada comentário era necessário atribuir um valor dentro do intervalo [-5,5], que classificaria o texto como muito negativo ou muito positivo, respectivamente. A escolha desta categorização foi feita no intuito de simplificar a atribuição de notas pelos alunos aos seus comentários e facilitar a normalização dos resultados obtidos pelas ferramentas automáticas para a escala do questionário.

Ao realizar a análise através do método SenticNet, observou-se que, ao tratar os comentários em sua forma bruta, a quantidade de palavras relacionadas a níveis sênticos chegava a mais de 50, a depender do tamanho do texto. Julgou-se também que comentários muito extensos poderiam dificultar a avaliação pelos próprios autores dos comentários, uma vez que durante a escrita de textos muito longos, a expressão dos sentimentos pode se tornar confusa e misturada. Dessa forma, foi implementada uma regra de subdivisão dos comentários de forma que todos os textos que possuíssem mais de três frases ou 100 palavras fossem divididos em novos, na tentativa também de reduzir o tempo gasto na leitura, considerando que cada pessoa gaste de um a dois minutos na leitura de 100 palavras. A partir dessa regra, os 84 comentários foram subdivididos, conforme necessário, totalizando 109 comentários.

4.3 Métricas Comparativas

A comparação entre os resultados obtidos foi realizada através do cálculo de quatro diferentes métricas. A avaliação do desempenho dos resultados foi realizada utilizando uma Matriz de Confusão [21], cuja forma de comparação entre polaridades está demonstrada na Tabela 3. Para a avaliação dos comentários de acordo com todas as categorias de polaridade (positiva, negativa e neutra), foi necessária a criação de uma matriz de confusão para cada execução, o que totalizou 4 matrizes criadas, uma para cada ferramenta e uma para avaliação manual do pesquisador.

A tabela 3 representa um exemplo de matriz de confusão para as polaridades atribuídas pela ferramenta LIWC em comparação com as atribuídas pelos autores dos comentários em suas avaliações, buscando representar quantos dos comentários foram preditos corretamente. Dessa forma, de acordo com o exemplo, na primeira linha estão representados todos os comentários classificados como negativos, neutros ou positivos de acordo com a ferramenta e que, na verdade, deveriam ter sido classificados como negativos (seguindo o gabarito dos autores dos comentários). A mesma lógica é aplicada também na comparação dos comentários neutros e positivos. Comentários corretamente classificados dentro de cada categoria são representados na tabela como “Corretos”.

Tabela 3. Matriz de Confusão das polaridades

Avaliação dos autores	Ferramenta de Análise - LIWC		
	Predito negativa	Predito neutra	Predito positiva
Polaridade negativa	Corretos	Incorretos	Incorretos
Polaridade neutra	Incorreto	Corretos	Incorretos
Polaridade positiva	Incorretos	Incorretos	Corretos

A partir dessa tabela e da obtenção da quantidade de comentários, para cada método é possível calcular a taxa de comentários classificados como uma determinada categoria e

que realmente pertence a ela (*true positives*), comentários classificados em uma categoria e que não pertencem a ela (*false positives*), comentários classificados como outras categorias e que pertencem a outras categorias (*true negatives*) e comentários classificados como outras categorias, mas que pertencem à categoria avaliada (*false negatives*). As fórmulas das métricas utilizadas podem ser definidas da seguinte maneira:

$$\text{Precision (P)} = \frac{T. P.}{(T. P. + F. P.)}$$

$$\text{Recall (R)} = \frac{T. P.}{(T. P. + F. N.)}$$

$$\text{Accuracy (A)} = \frac{(T. P. + T. N.)}{(T. P. + F. P. + T. N. + F. N.)}$$

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

4.4 Metodologia

Essa seção descreve as etapas do método seguidas para a realização da comparação entre os resultados obtidos, através da definição das ações necessárias para tratar os dados a serem trabalhados e da forma como essas informações foram avaliadas a partir de cada abordagem.

4.4.1 Extração da base de dados

Após a definição da base de dados a ser trabalhada, foi utilizada a ferramenta NodeXL⁷, plugin de licença livre, para extração das informações referentes as publicações, comentários e autores dos mesmos.

4.4.2 Pré-processamento dos dados

No pré-processamento de dados foi realizada a subdivisão dos comentários de acordo com o tamanho de seu texto e a tradução de cada uma dessas partes para a língua inglesa.

O principal intuito da subdivisão foi poder dar aos autores dos comentários a oportunidade de identificar mais de uma única polaridade em um mesmo comentário, considerando que é comum que um texto possa defender um determinado ponto de vista (positivo) ao mesmo tempo que realiza uma crítica a outro (negativo). Devido a este motivo, o estudo desconsiderou o possível impacto semântico negativo das regras de divisão utilizadas que poderiam fazer com que um comentário em sua forma bruta fosse considerado positivo ao mesmo tempo que, talvez, sua subdivisão indicasse polaridades positivas e negativas, por exemplo.

4.4.3 Abordagens

A avaliação do desempenho da abordagem automática a partir dos resultados das três ferramentas e da análise manual feita por um dos pesquisadores foi realizada através da comparação dos resultados das polaridades e dos valores aproximados com os resultados apontados pelos autores dos comentários.

Dessa forma, de acordo com cada uma das abordagens listadas, buscou-se obter uma categoria de polaridade (positiva, negativa ou neutra) e um valor numérico de polaridade para cada texto analisado. No tratamento dos comentários através da ferramenta SenticNet também foi avaliado o sentimento inferido pelas frases, tendo como base de análise os níveis sênticos definidos pela Sentic Computing, que posteriormente também foram comparados aos sentimentos apontados pelos autores para cada um de seus comentários.

Preparação do Gabarito - Autores

Os questionários criados foram personalizados de forma que cada autor pudesse avaliar suas próprias interações. Para cada

⁷ Disponível em <https://nodexl.codeplex.com/>

autor foi disponibilizado um documento relacionando todos os comentários de sua autoria incluindo a data, o conteúdo do tópico, nome do autor e conteúdo do comentário. A etapa mais importante dessa pesquisa foi a de categorização destes textos pelos autores através da atribuição de valores de [-5,5] para cada comentário. Os participantes também foram estimulados a avaliar os comentários realizados através do estabelecimento de relacionamentos entre cada comentário e, no mínimo, um sentimento de nível 1 da Sentic Computing. Em ambas as avaliações, foi possível que o autor concluísse que não houve tentativa de expressar nenhum sentimento ou emoção, classificando o texto como neutro.

Análise Automática

As análises realizadas a partir da categorização automática dos sentimentos respeitaram as diversas limitações impostas por cada uma das ferramentas. Após a tradução dos textos para a língua inglesa, os comentários traduzidos foram aplicados às ferramentas SenticNet, SentiStrenght e LIWC para a obtenção de valores de polaridade e inferência de qual a categoria (positiva, negativa ou neutra) os comentários se enquadravam.

Análise Manual - Pesquisador

Na análise manual realizada por um dos pesquisadores foi disponibilizado um questionário semelhante ao da análise realizada pelos autores dos comentários sem, contudo, haver a diferenciação por autoria. O pesquisador foi responsável pela atribuição de valores para cada um dos 109 comentários.

A avaliação realizada pelo pesquisador buscou pela existência de palavras claramente positivas ou negativas, como 'amor', 'gostar', 'sucesso', 'esperança', 'raiva', 'tristeza' ou 'odiar', as quais foram atribuídos os valores +5 ou -5 de acordo com a polaridade. Da mesma forma, conforme as palavras fossem sendo identificadas como tentativas de expressar sentimentos positivos/negativos, mas em menor intensidade, os valores atribuídos foram sendo alterados. Como no exemplo: "Bem legal os 'cartazes'. Eu também me identifico mais com o número 5." As palavras *legal* e *identifico* foram consideradas como de cunho positivo, dentro do contexto, tornando a frase como um todo também positiva, de forma que os valores numéricos atribuídos foram +5 e +3, tornando a polaridade total da frase sendo +4 (média das polaridades).

A realização de uma comparação entre as diferentes abordagens teve como objetivo não somente identificar o percentual de erros entre a análise automática realizada pelas ferramentas e o real sentimento que se buscou expressar, mas de também observar se um percentual de erros igual ou semelhante é obtido mesmo quando se trata de um humano avaliando os textos. A partir dessa verificação é possível a criação de hipóteses que apontem que, para uma avaliação mais próxima do real, seja necessária a utilização de formas de análise que auxiliem não somente durante a classificação automática, mas também na classificação manual por terceiros.

5. ANÁLISE DOS RESULTADOS

A partir dos resultados de cada abordagem, foram realizados quatro tipos de avaliação:

A1: Avaliação comparativa entre a classificação da **polaridade** por meio das análises automática e manual realizada pelo pesquisador para identificar o percentual de acertos destes métodos quando em comparação com os resultados reais definidos pelos autores dos comentários

A2: Avaliação comparativa entre a classificação dos **valores** por meio das análises automática e manual realizada pelo pesquisador para identificar o percentual de acertos destes métodos quando em comparação com os resultados reais definidos pelos autores dos comentários.

A3: Avaliação comparativa entre a classificação dos **sentimentos** por meio do método SenticNet para identificar a similaridade entre os resultados obtidos através deste método e os resultados reais definidos pelos autores dos comentários.

A4: Avaliação do desempenho geral das abordagens de análise automática e manual realizada pelo pesquisador através do cálculo de **métricas**.

Dessa forma, o estudo buscou identificar o número de classificações relevantes realizadas pelas ferramentas e pela análise do pesquisador na tentativa de apontar qual destes obteve melhores resultados.

5.1 Análise das Polaridades

A primeira avaliação realizada comparou os resultados obtidos por cada um dos métodos na categorização da polaridade dos comentários. A definição de a qual polaridade um texto pertenceria foi estabelecida de acordo com as formas de atribuição de valores utilizadas por cada abordagem. A tabela 4 representa o percentual de comentários em cada categoria.

De acordo com a avaliação dos autores dos comentários, a maior parte dos comentários realizados foi considerada positiva, o que sugere uma tendência dos usuários do grupo do Facebook de Sistemas Colaborativos em considerar o ambiente mais propício para o compartilhamento de conteúdo positivo, acadêmico ou não, do que como ambiente para disseminação de críticas, pesares e/ou reclamações.

Tabela 4. Comentários classificados por polaridade

	Polaridade		
	Negativo (<0)	Neutro (=0)	Positivo (>0)
LIWC	11 (10,1%)	44 (40,4%)	54 (49,5%)
SentiStrenght	23 (21,1%)	46 (42,2%)	40 (36,7%)
SenticNet	20 (18,3%)	45 (41,3)	44 (40,4%)
Avaliação Pesquisador	25 (23%)	20 (18,3%)	64 (58,7%)
Avaliação Autores	11 (10,1%)	20 (18,3%)	78 (71,56%)

Em contrapartida, pode-se observar que as ferramentas de análise automática classificaram a maior parte dos comentários como sendo de polaridade neutra, o que significa dizer que em mais de 40% dos textos analisados, as ferramentas não foram capazes de identificar nenhum sentimento relacionado ou que os sentimentos identificados anularam uns aos outros. A avaliação realizada pelo pesquisador foi a que mais se aproximou dos resultados indicados pelos autores dos comentários. Porém, apesar de os resultados apontarem uma maior facilidade humana para a classificação correta dos textos, mesmo esta avaliação não foi capaz de identificar corretamente todos os comentários.

5.2 Análise dos Valores

Em todas as abordagens, a polaridade dos comentários foi calculada tendo como base valores numéricos atribuídos pelos métodos a cada um dos textos; a partir dessa atribuição foi possível identificar a intensidade dos sentimentos em um comentário, que poderia variar de muito negativa a muito positiva. Para fins de comparação, após a classificação realizada pelas ferramentas, todos os valores foram transformados de forma a atingirem um nível de proporcionalidade à escala de valores utilizada nos questionários. A figura 1 representa a quantidade de comentários classificada de acordo com cada valor.

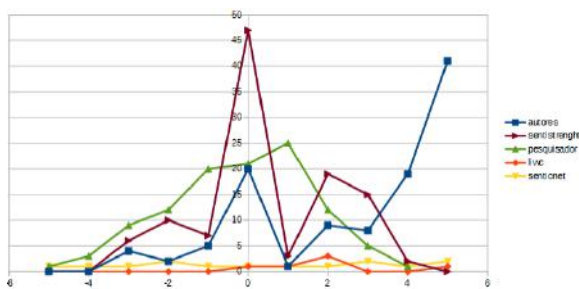


Figura 1. Comentários classificados por valor

Através do gráfico, pode-se observar uma série de divergências ao realizar uma comparação entre os resultados obtidos. Assim como na análise percentual, na avaliação dos níveis de intensidade de sentimentos dos comentários classificados pelos seus autores, houve ocorrência de um *outlier*, valor atípico na análise, devido a grande incidência de comentários classificados como muito positivos (de valor igual a 5). Da mesma forma, a análise automática usando a ferramenta SentiStrenght também apresentou um ponto de afastamento dos demais pontos do gráfico ao classificar quase 50 comentários como sendo neutros, ou seja, com valor igual a 0.

Os resultados das ferramentas SenticNet e LWC foram considerados os mais discrepantes do grupo. Diferente do resultado da ferramenta SentiStrenght, os resultados de ambos indicou que todos os comentários foram categorizados com valores muito próximos uns dos outros sem, contudo, classificar uma grande quantidade de comentários com o mesmo valor, gerando, assim, a criação de duas linhas quase retas no gráfico. Apesar de não ter sido possível identificar um padrão de concordância entre os resultados das ferramentas, a maioria destes demonstrou uma tendência ao não classificar os comentários nos extremos das escalas, sendo, então, poucos os categorizados como muito negativos (de valor igual a -5) ou muito positivos (de valor igual a +5).

Nenhuma das ferramentas da abordagem automática e nem a abordagem manual pelo pesquisador conseguiu realizar a classificação da intensidade dos comentários de forma igual ou similar ao gabarito fornecido pelos autores dos comentários, o que aponta que mesmo que os métodos consigam definir se um comentário é positivo ou negativo, dificilmente estes conseguem avaliar a intensidade afetiva do texto.

5.3 Análise dos Sentimentos

Através do SenticNet, além da classificação de acordo com a polaridade, também foi possível categorizar os comentários a partir do estabelecimento de relações entre os textos e emoções do nível 01 da Sentic Computing. Dessa forma, com a análise dos valores obtidos em cada uma das quatro dimensões de níveis sênticos, pôde-se inferir qual(quais) sentimento(s) o texto em questão procurou expressar. Uma categorização similar foi realizada pelos autores dos comentários que atribuíram um ou mais sentimentos aos comentários que julgaram possuir teor afetivo. Em ambos os casos, na ocorrência de comentários com mais de um sentimento relacionado, para facilitar a análise comparativa, apenas um destes, escolhido de forma aleatória, foi considerado.

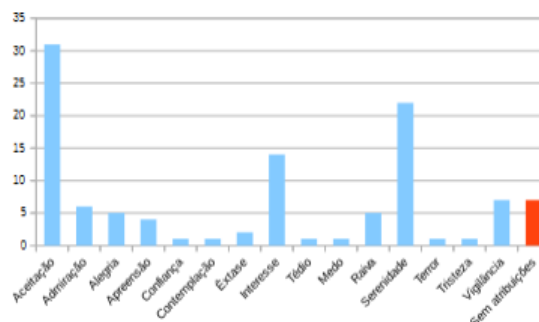


Figura 2. Classificação dos sentimentos usando SenticNet

A figura 2 representa a classificação afetiva dos comentários realizada pela ferramenta SenticNet, através da exibição de todos os níveis aos quais foram atribuídas quantidades de comentário. Pode-se observar que dentre os níveis de sentimentos, a maior parte dos comentários foi classificada como 'Aceitação' ou 'Serenidade', seguido de 'Interesse' e 'Vigilância'. Uma análise mais profunda destes resultados gera indícios de que as palavras utilizadas nos textos não possuísem, de fato, relação com sentimentos extremos como 'amor', 'ódio' e 'medo' e que por isso os níveis sênticos mais atingidos foram níveis um tanto quanto subjetivos.

Dentre todos os 109 textos avaliados, um número pequeno destes foi considerado pelo método como sem atribuições, ou seja, sem nenhum sentimento atrelado, apesar de na avaliação pela polaridade ter sido observado que muitos destes comentários foram considerados neutros. Não foi possível estabelecer uma regra clara sobre o porquê estes comentários em particular não foram considerados pela ferramenta como relacionados a nenhum conceito de senso comum.

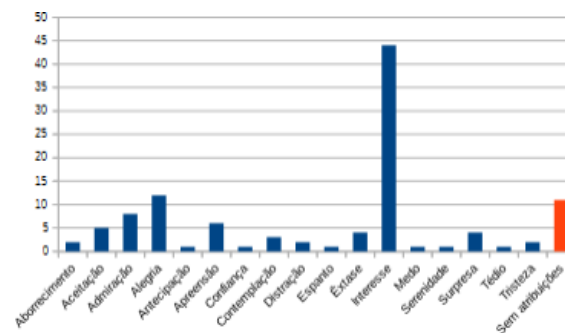


Figura 3. Classificação dos sentimentos pelos Autores

No gráfico da figura 3 é possível perceber que, novamente, o nível correspondente ao 'Interesse' obteve uma quantidade diferenciada dos demais ao agrupar mais de 40% dos comentários e que um número maior de comentários foi classificado de forma positiva na categoria de 'Alegria'. Também se observa a não ocorrência de um número expressivo de comentários em níveis sênticos que retratassem um sentimento negativo, como 'Tédio' e 'Tristeza'. Comparando as figuras 2 e 3 não é possível dizer que houve um equilíbrio nos resultados ou mesmo uma concordância entre as formas de análise, mais uma vez indicando que, por mais que as motivações dos alunos de um mesmo grupo ao compartilhar informações possam vir a ser semelhantes, a abordagem automática de avaliação encontrou dificuldades durante a categorização.

5.4 Análise do Desempenho de Predição

Após a realização de todas as análises unitárias e comparativas, buscou-se avaliar o desempenho das abordagens através do cálculo das métricas anteriormente definidas. A tabela 5 representa os valores obtidos de precisão (*precision*), cobertura

(recall), acurácia (accuracy) e *f-measure* para as três ferramentas usadas na análise automática e o método de análise manual pelo pesquisador.

A métrica de precisão (*precision*) indicou quantas das avaliações realizadas a respeito dos comentários foi, de fato, relevante, ao considerar a taxa de comentários classificados de acordo com determinada categoria que foi ou não classificada corretamente. O método que atingiu maior precisão foi o SentiStrength na classificação de comentários positivos; dos 40 comentários classificados, 36 deles foram corretamente categorizados, o que fez com que o desempenho de precisão atingisse o valor de 0.9 sendo este mais alto, inclusive, do que a avaliação realizada pelo pesquisador. Através do cálculo do *recall* foi possível observar qual método menos classificou incorretamente comentários que eram de uma determinada categoria como pertencentes a outra. A análise manual realizada pelo pesquisador foi a que obteve os melhores resultados quando da avaliação dos comentários negativos, porém, ainda assim, os resultados não foram totalmente satisfatórios, uma vez que 18.18% dos comentários ainda foram classificados de maneira incorreta. A análise manual também se destacou por possuir a maior acurácia observada, no valor de 0.6972. Apesar do alto nível de acurácia alcançado é necessário ressaltar que, devido as diferenças de quantidade de comentários classificados pelos autores dos mesmos de acordo com cada polaridade, a utilização desta métrica por si só não reflete corretamente a eficácia de um método. Um alto nível de acurácia também teria sido alcançado se um determinado método classificasse aleatoriamente todos os comentários em uma única categoria de polaridade; os valores de *positivos verdadeiros* para esta categoria teriam 100% de acerto e

obteriam uma acurácia proporcional a quantidade de comentários da categoria.

Por fim foi realizada a medida geral de eficiência do desempenho dos métodos levando em consideração os erros de classificação em todas as três categorias de polaridade. Através do cálculo da F-measure observou-se que os melhores resultados foram obtidos pela análise manual pelo pesquisador na classificação de comentários neutros e positivos, seguido pela ferramenta LIWC, também na categorização dos comentários positivos, que obtiveram valores de 0.7821, 0.6972 e 0.6570, respectivamente.

6. CONCLUSÕES

O estudo realizado buscou avaliar os comentários de alunos em um grupo do Facebook a fim de categorizá-los de acordo com a polaridade dos textos e os sentimentos inferidos.

A abordagem baseada na análise automática através do uso das ferramentas LIWC, SenticNet e SentiStrenght e a abordagem manual foram utilizadas como formas de classificação dos sentimentos dos textos e os resultados obtidos foram comparados ao gabarito gerado pela abordagem manual de categorização realizada pelos autores dos comentários. O gabarito dos autores dos comentários apontou mais de 70% dos comentários como possuindo teor positivo, algo que os outros métodos não conseguiram captar. Em alguns casos, a eficiência da análise manual do pesquisador foi semelhante ou pior a dos métodos automáticos, indicando que apesar das falhas no processamento automático dos textos, nem mesmo um humano consegue avaliar com alto grau de certeza que sentimento outra pessoa quis expressar.

Tabela 5. Métricas de Desempenho dos métodos

Método	Polaridade - Negativa								
	T.P.	F.P.	T.N. (Neu.)	T.N. (Pos.)	F.N.	Recall	Precision	F-measure	Accuracy
LIWC	2	9	14	44	9	0.1818181818	0.1818181818	0.2733485194	0.5504587156
SentiStrenght	6	17	14	36	5	0.5454545455	0.2608695652	0.3460350154	0.5137614679
SentirNet	6	14	9	35	15	0.2857142857	0.3	0.3627569258	0.4587155963
Avaliação Pesquisador	9	16	10	57	2	0.8181818182	0.36	0.4748351267	0.6972477064
Método	Polaridade - Neutra								
	T.P.	F.P.	T.N. (Neg.)	T.N. (Pos.)	F.N.	Recall	Precision	F-measure	Accuracy
LIWC	14	30	2	44	15	0.4827586207	0.3181818181	0.5504587156	0.5504587156
SentiStrenght	14	32	6	36	6	0.7	0.3043478261	0.5137614679	0.5137614679
SentirNet	9	36	6	35	11	0.45	0.2	0.4587155963	0.4587155963
Avaliação Pesquisador	10	10	9	57	10	0.5	0.5	0.6972477064	0.6972477064
Método	Polaridade - Positiva								
	T.P.	F.P.	T.N. (Neg.)	T.N. (Neu.)	F.N.	Recall	Precision	F-measure	Accuracy
LIWC	44	10	2	14	34	0.5641025641	0.8148148148	0.6570433051	0.5504587156
SentiStrenght	36	4	6	14	42	0.4615384615	0.9	0.6541207008	0.5137614679
SentirNet	35	9	6	9	43	0.4487179487	0.7954545455	0.5818786367	0.4587155963
Avaliação Pesquisador	57	7	9	10	21	0.7307692308	0.890625	0.782161235	0.6972477064

A avaliação com base nas métricas de desempenho demonstrou que a precisão máxima obtida por um método foi através da ferramenta SentiStrenght, que alcançou o valor de 0.9 para precisão. A abordagem de avaliação pelo pesquisador atingiu bons níveis de *recall* e *F-measure* com os valores de 0.8181 e 0.78. Muitos comentários classificados como “Muito Positivo” não demonstraram possuir, de fato, palavras de característica positiva (como ‘amor’ ou ‘alegria’), sendo essencial a avaliação do próprio autor para identificar qual era o real sentimento deste ao escrevê-los. Posteriormente, os próprios autores definiram que o sentimento muitas vezes é expresso durante “o calor do momento”, sendo difícil até mesmo para eles categorizarem em um momento do futuro o que, de fato, haviam tentado expressar em determinado comentário.

Como resultados gerais da pesquisa foi possível criar suposições acerca das classificações como, por exemplo, o fato do contexto do estudo de caso ter exercido influência direta nos comentários realizados, uma vez que por se tratar de um ambiente onde os principais objetivos eram o conhecimento acadêmico e o apoio à aprendizagem, os comentários publicados possuíam, em sua maioria, uma conotação muito positiva, tornando mais fácil e tendenciosa a classificação automática dos comentários.

A partir da problemática observada pôde-se concluir que as ferramentas existentes, apesar de adequadas dentro de determinado contexto, não conseguiram, de fato, sanar todas as necessidades para execução das análises dos comentários dos autores, gerando uma série de limitações que impactaram no resultado final da avaliação. A análise dos comentários como um todo se faz importante por tornar possível que o professor avalie quais os sentimentos dos alunos com relação aos assuntos ministrados, de forma a melhor conduzir discussões ou direcionar temas de acordo com interesses dos alunos.

Como trabalhos futuros pretende-se explorar a análise de sentimentos levando em consideração o perfil do usuário, ou seja, adaptando as avaliações dos comentários de forma que sejam consideradas particularidades de cada um dos alunos. Também se tem como objetivo o tratamento dos textos sem a necessidade de traduzi-los previamente, minimizando a perda de contexto e de características linguísticas.

7. AGRADECIMENTOS

Este trabalho foi parcialmente financiado pela FAPERJ (projeto E-26-102.256/2013 – Associa: Explorando um Ambiente Semântico e Social de Ensino-Aprendizagem).

8. REFERÊNCIAS

- [1] Abbasi, A., Chen, H., & Salem, A. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Transactions on Information Systems (TOIS)*, Vol. 26, No. 3, p. 12.
- [2] Ahmed, K.; El Tazi, N.; Hossny, A. H. Sentiment Analysis Over Social Networks: An Overview.
- [3] Altrabsheh, N., Cocea, M., & Fallahkhair, S. 2014. Learning sentiment from students’ feedback for real-time interventions in classrooms. In: *Adaptive and Intelligent Systems*. Springer International Publishing. p. 40-49.
- [4] Araújo, M. et al. 2013. Métodos para análise de sentimentos no Twitter. In: *Proceedings of the 19th Brazilian Symposium on Multimedia and the Web (WebMedia'13)*.
- [5] Bosco, C.; Patti, V.; Bolioli, A. 2013. Developing corpora for sentiment analysis: The case of irony and senti-tut. *IEEE Intelligent Systems*, No. 2, p. 55-63.
- [6] Cambria, E.; Hussain, A. 2015. *Sentic computing: a common-sense-based framework for concept-level sentiment analysis*. Springer. Vol 1. Disponível em: <<http://www.sentic.net/sentic-computing.pdf>> Acesso em: 20

nov, 2015

- [7] Cambria, E; Livingstone A.; Hussain A. (2012) *The Hourglass of Emotions*. In: LNCS, Vol 7403. p. 114- 157. Springer. Disponível em: <<http://www.sentic.net/hourglass-of-emotions.pdf>>. Acesso em 21 nov, 2015.
- [8] Cambria, E.; Grassi, M.; Hussain A.; Havasi C. 2012. *Sentic Computing for social media marketing*. Springer.
- [9] Cervi, E. U., & Massuchin, M. G. 2012. Redes sociais como ferramenta de campanha em disputas subnacionais: análise do Twitter nas eleições para o governo do Paraná em 2010. *Sociedade e Cultura*, Vol. 15, No. 1.
- [10] Chamlerwat, W. et al. 2012. Discovering Consumer Insight from Twitter via Sentiment Analysis. *J. UCS*, Vol. 18, No. 8, p. 973-992.
- [11] de Oliveira, Aletheia Machado. *Redes Sociais Virtuais, Blog, Wiki e Moocs, como parte de uma arquitetura pedagógica*. 2016. Rehutec, Vol 5, No. 1, p12.
- [12] Duwairi, R. M. 2015. Sentiment analysis for dialectical Arabic. In: *Information and Communication Systems (ICICS)*, 2015 6th International Conference on. IEEE, p. 166-170.
- [13] Huangfu, Y. et al. 2015. An improved sentiment analysis algorithm for Chinese news. In: *Fuzzy Systems and Knowledge Discovery (FSKD)*, 2015 12th International Conference on. IEEE. p. 1366-1371.
- [14] Kechaou, Z.; Ben Ammar, M.; Alimi, A. M. 2011. Improving e-learning with sentiment analysis of users’ opinions. In: *Global Engineering Education Conference (EDUCON)*, 2011 IEEE. IEEE, p. 1032-1038.
- [15] Ortigosa, A.; Martín, J. M.; Carro, R. M. 2014. Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, Vol. 31, p. 527-541.
- [16] Liu, B. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, Vol. 5, No. 1, p. 1-167.
- [17] Mendes Gomes L et al. 2015. Facebook vs moodle: Surveying university students on the use of learning management systems to support learning activities outside the classroom. In: *Information Systems and Technologies (CISTI)*, 2015 10th Iberian Conference on. IEEE, p. 1-4.
- [18] Montoyo, A.; MartíNez-Barco, P. ; Balahur, A. 2012. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems*, Vol. 53, No. 4, p. 675-679.
- [19] Moreno, A. et al. 2011. Feeling bad on Facebook: Depression disclosures by college students on a social networking site. *Depression and anxiety*, Vol. 28, No. 6, p. 447-455.
- [20] Nasukawa, T.; Yi, J. 2003. Sentiment analysis: Capturing favorability using natural language processing. In: *Proceedings of the 2nd international conference on Knowledge capture*. ACM. p. 70-77.
- [21] Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. *Science*, Vol. 240, No. 4857, p. 1285-1293.
- [22] Thelwall, M.; Buckley K.; Paltoglou G. 2012. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*. Vol. 63, No. 1, p. 163-173.
- [23] Wang, H. et al. 2012. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In: *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, p. 115-120.
- [24] Waters, R. D. et al. 2009. Engaging stakeholders through social networking: How nonprofit organizations are using Facebook. *Public relations review*, Vol. 35, No. 2, p. 102-106.