

Uma Abordagem Multilíngue para Análise de Sentimentos

Julio C. S. Reis, Pollyanna Gonçalves, Matheus Araújo
Adriano C. M. Pereira, Fabrício Benevenuto

¹Universidade Federal de Minas Gerais (UFMG) – Brasil

{julio.reis, pollyannaog, matheus.araujo, adrianoc, fabricio}@dcc.ufmg.br

Abstract. *Sentiment analysis has been applied in many contexts, including analysis of user's opinions about products and services, prediction of elections and even market trending. Despite the large interest on this theme and amount of research efforts in the field, most of the existing methods are designed to work with English content. In this paper, we focus at filling this research gap by proposing an approach for the use of several state-of-the-art methods for sentiment analysis in 9 languages. In order to do it, we use many polarity labeled datasets of each language and develop a methodology to compare and validate them. Our results suggest that simply translating the input text on an specific language to English and then use one of the existing English methods has the potential to make sentiment analysis English independent.*

Resumo. *A análise de sentimentos tornou-se uma ferramenta essencial para aplicação em diversos contextos, incluindo análise de opinião do usuário sobre produtos e serviços, previsão durante campanhas políticas e até mesmo em tendências do mercado financeiro. Apesar do grande interesse neste tema e na quantidade de pesquisas na área, a maioria dos métodos foram projetados para funcionar com o conteúdo em inglês. Neste estudo, nos direcionamos a preencher esta lacuna propondo uma abordagem para o uso de determinados métodos estado-da-arte para análise de sentimentos em 9 diferentes linguas. Para isto, nós utilizamos bases de dados previamente rotuladas em cada idioma e uma simples tradução automática para o inglês e desenvolvemos uma metodologia para comparar e validar os resultados. Nossos resultados demonstram o potencial desta abordagem para tornar a análise de sentimentos independente da língua inglesa.*

1. Introdução

Desde o seu surgimento, as redes sociais online (OSNs - Online Social Networks) têm atraído vários usuários e se tornado um ambiente extremamente popular na web [Online 2009]. Esses sistemas fornecem mecanismos para que os usuários discutam sobre vários assuntos, expressando suas opiniões pessoais, visão política, orientação sexual e até mesmo conceitos subjetivos como humor e felicidade. Devido à imensa popularidade e à grande quantidade de informação compartilhada nestes ambientes, várias aplicações têm surgido com intuito de extrair opiniões e até mesmo inferir o sentimento público.

Neste contexto a análise de sentimentos que tem por objetivo identificar e extrair de forma automática, as opiniões, sentimentos e emoções, expressados em um texto [Narayanan et al. 2009], têm se tornado uma ferramenta bastante popular para a

mineração de dados em OSNs, com várias aplicações úteis. A partir desses dados, tais técnicas podem revelar por exemplo, como as pessoas se sentem sobre um determinado alvo, o que é essencial para que as empresas se tornem capazes de direcionar suas ações de marketing de forma mais embasada.

Desde que a análise de sentimentos se tornou um tema em voga em aplicações Web, considerando o crescente interesse por parte da indústria e da academia, muitas ferramentas e métodos têm sido propostos para a detecção de sentimentos em diversos ambientes. Estes métodos baseiam-se em diferentes estratégias, tais como a aprendizagem de máquina [Thelwall 2013, Wang et al. 2012], dicionários léxicos [Bradley and Lang 1999, Dodds and Danforth 2010, Tausczik and Pennebaker 2010], processamento de linguagem natural [Esuli and Sebastiani 2006], *crowdsourcing* [Mohammad and Turney], escalas psicométricas [Bollen et al. 2009] ou mesmo uma combinação de tais técnicas [Goncalves et al. 2013].

A diversidade de métodos e técnicas resulta e/ou tem impactos na diversidade de categorização de sentimentos utilizados em cada um dos métodos, nas suposições feitas e nos conjuntos de dados utilizados para validação. Contudo, grande parte destes recursos são disponibilizados apenas para a língua inglesa, considerando que este idioma é dominante no conteúdo disponibilizado. Alguns esforços têm sido realizados para o desenvolvimento de técnicas de análise de sentimentos em outros idiomas: Árabe [Abdul-Mageed et al. 2012], Alemão [Remus et al. 2010], Português [Souza and Vieira 2012], Russo [Yussupova et al. 2012], entretanto pouco se sabe sobre o desempenho e real necessidade ou viabilidade do desenvolvimento destas soluções.

Neste trabalho, endereçamos o problema da análise de sentimentos em diferentes idiomas com objetivo de preencher essa lacuna. Para isso, utilizando uma máquina de tradução de acesso aberto - Google Translate - propomos uma estratégia onde executamos métodos presentes na literatura em diferentes idiomas, apesar de serem originalmente criados para execução em textos no idioma Inglês. Em particular vamos analisar o desempenho de 13 métodos de análise de sentimentos - SentiWordNet, PANAS-t, SASA, SenticNet, Happiness Index, EmoLex, NRC Hashtag Sentiment Lexicon, OpinionLexicon, Sentiment140 Lexicon, VADER, LIWC, SentiStrength e Emoticons - em 9 idiomas diferentes do inglês - Português, Francês, Espanhol, Italiano, Turco, Russo, Árabe, Holandês e Alemão. De acordo com a *Internet World Stats*¹, isso corresponde a análise de 6 das *top 10* línguas mais utilizadas na internet.

Para validação da abordagem proposta contamos com a utilização de bases de dados com mensagens rotuladas como positivas e negativas em cada um dos diferentes idiomas que serão analisados nesse trabalho, compostas por *reviews* de filmes e produtos, tweets e mensagens associados a assuntos diversos. A partir de bases de dados reais, comparamos os métodos para análise de sentimentos em termos de abrangência (a fração de mensagens capturadas por cada método) e performance de predição (a fração de sentimentos corretamente identificados por cada método).

Nossos resultados mostraram que existe uma consistência na acurácia e abrangência dos métodos de classificação de sentimentos mesmo quando submetidos em

¹<http://www.internetworldstats.com/stats7.htm>

bases de dados de diferentes idiomas, desde que previamente traduzidas para o idioma original do método, na maioria dos casos o inglês.

O restante desse artigo está organizado da seguinte forma. A Seção 2 descreve os trabalhos relacionados. Em seguida, na Seção 3, descrevemos os métodos utilizados nesse trabalho. Depois, na Seção 4, apresentamos nossa metodologia para tradução das bases e execução dos métodos. A análise e discussão dos resultados são apresentados na Seção 5. Por fim, concluímos e apresentamos direções para trabalhos futuros.

2. Trabalhos Relacionados

Grande parte dos métodos para análise de sentimentos em textos estão disponíveis apenas para o idioma inglês, e com crescimento do conteúdo Web disponibilizado em outros idiomas alguns trabalhos tem sido propostos com o objetivo de explorar a análise textos contendo opiniões ou sentimentos, independente do idioma no qual estão escritos, numa perspectiva multilíngue.

Neste contexto, a tradução automática do texto ou dicionário tem atraído vários pesquisadores. Em [Banea et al. 2008] é apresentada uma abordagem de classificação de polaridade onde os autores investigaram a geração automática de *corpora* para análise de subjetividade em idiomas que não possuem este recurso. Considerando traduções automáticas para romeno e espanhol, foram desenvolvidos experimentos para verificação de desempenho da classificação da polaridade (*i.e.* validade) a partir de um corpus em inglês, rotulado. A abordagem se mostrou simples e apresentou bons resultados quanto a classificação da polaridade.

Já no trabalho [Lin et al. 2011], os autores propuseram a utilização de um conjunto de palavras semente (advérbios) e treino de um classificador não-supervisionado. Um conjunto de dados rotulados para treinamento nos diversos idiomas foi elaborado automaticamente considerando o uso de características que independem de idiomas (*i.e.* emoticons) [Narr et al. 2012]. Foram efetuadas análises individuais e combinadas para os idiomas inglês, alemão, francês e português, e os resultados apresentados foram considerados satisfatórios em alguns cenários. De forma semelhante, no trabalho apresentado em [Abdel-Hady et al. 2014] é proposto um método não supervisionado para análise de polaridade em idiomas alvo, baseado em recursos em ferramentas disponíveis em inglês. A abordagem léxica com a realização de experimentos em espanhol e português proporcionou melhora no desempenho da classificação de polaridade em mídias sociais.

Um estudo de caso para mineração de opiniões multilíngues é apresentado em [dos Santos et al. 2014]. O trabalho sobre a classificação de emoções contidas em revisões de produtos investiga uma abordagem que combina a identificação de sentimentos ou emoções baseada em dicionário e tradução automática. O objetivo da proposta era identificar se é mais adequado traduzir o texto das revisões ou as palavras dos dicionários léxicos, além de investigarem, de forma complementar, a possível melhora dos resultados com a aplicação de um lematizador. A análise foi efetuada para o idioma português. Os resultados apresentados mostram que a tradução do texto alcança desempenho satisfatório enquanto a lematização não trouxe mudanças ou melhoras significativas.

Diferente dos trabalhos relacionados, nosso esforço toma outra direção. Nossa análise fornece uma avaliação mais abrangente de métodos de análise de sentimentos

desenvolvidos para o idioma inglês e aplicados à dados traduzidos em diferentes línguas. Não estamos limitados a um contexto ou conjunto específico de métodos, o que amplia a aplicabilidade das nossas descobertas. Além disso, a cobertura dos vários idiomas nos possibilita de forma vasta analisar o desempenho e limitações de cada um dos métodos nos diferentes cenários estudados.

3. Métodos de Análise de Sentimentos

Existem diversos métodos de análise de sentimentos na literatura. Esses métodos se diferenciam por suas técnicas de predição de sentimentos, com o uso de abordagem de aprendizado de máquina, dicionários léxicos, processamento de linguagem natural e escalas psicométricas. Neste trabalho estamos interessados em métodos que estão focados na detecção de polaridade em menor granularidade, assumindo que cada uma das frases possua um único sentimento [Liu 2012]. Esta seção apresenta uma breve descrição dos 13 métodos de análise de sentimentos que são aplicados nesse trabalho.

LIWC. Linguistic Inquiry and Word Count (LIWC) [Tausczik and Pennebaker 2010] é uma ferramenta para análise de texto que estima componentes emocionais, cognitivos e estruturais de um dado texto baseada no uso de dicionários contendo palavras e suas respectivas categorias, além de detectar pontuação de polaridade positiva e negativa em um texto. A ferramenta é unicamente comercial e fornece funções otimizadas, como a permissão para inclusão de dicionários personalizados. Para este trabalho, utilizamos a versão LIWC2007, a versão mais recente do sistema, e o dicionário padrão para o idioma Inglês. O software pode ser encontrado em <http://www.liwc.net/>.

SentiStrength. O SentiStrength [Thelwall 2013] é uma método de análise de sentimentos bastante conhecido que utiliza um dicionário léxico rotulado por humanos que é aprimorado por aprendizado de máquina. O trabalho utilizou para classificação uma versão expandida do dicionário do LIWC [Tausczik and Pennebaker 2010], com a adição de novas características para o contexto de redes sociais, como símbolos que dão entonação a um sentimento (ex.: emoticons e pontuações). Dada uma mensagem, o referido método a classifica numa escala de [-5, 5], onde -5 e 5 representam as pontuações mais negativas e positivas respectivamente. Neste trabalho utilizamos a versão 2.0 do método, que está disponível em [sen 2013].

SentiWordNet. SentiWordNet [Esuli and Sebastiani 2006] é uma ferramenta muito utilizada em mineração de opinião, e é baseado no dicionário léxico WordNet [Miller 1995]. Esse dicionário agrupa adjetivos, verbos e outras classes gramaticais em conjuntos chamados *synset*. O SentiWordNet associa a cada *synset* do WordNet três valores de pontuação que indicam o sentimento de um texto: positivo, negativo e objetivo (neutralidade). Cada pontuação é obtida utilizando um método de aprendizagem de máquina semi-supervisionada, e variam de 0 a 1, com soma igual a 1. Nesse artigo utilizamos a versão 3.0 do SentiWordNet, disponível em <http://sentiwordnet.isti.cnr.it/>. Para associar polaridade baseados nesse método, consideramos a média da pontuação dos *synsets* e diremos que um texto dado é positivo se o valor resultante para positivo for maior que o valor encontrado para negativo. Pontuações para objetividade não foram consideradas nesse trabalho para determinar polaridade.

SenticNet. SenticNet [Cambria et al. 2010] é um método para mineração de

opinião e análise de sentimentos que explora técnicas de Inteligência Artificial e Web Semântica. O objetivo do SenticNet é inferir polaridade de textos em nível semântico, e não sintático. O método utiliza técnicas de Processamento de Linguagem Natural (PLN) para criar significados semânticos ou polaridade para aproximadamente 14.000 conceitos. Neste trabalho, consideramos pontuações menores que zero como negativas, iguais a zero como neutras e superiores a zero como positivas. Nós utilizamos neste estudo a versão 2.0 do SenticNet, disponível em <http://sentic.net/>.

SASA. Empregamos ao trabalho mais uma técnica baseada em aprendizado de máquina, o SailAil Sentiment Analyzer (SASA) [Wang et al. 2012]. O SASA foi originalmente proposto como um método para análise de 17.000 tweets rotulados associados as eleições norte-americanas de 2012. A ferramenta, de código aberto, foi avaliada no Amazon Mechanical Turk (AMT) [Amazon 2005], onde *turkers* rotularam tweets como positivos, negativos, neutros ou indefinidos. Esse método foi acrescentado a nossa análise por ser uma ferramenta aberta e ainda não ter sido comparado com nenhum outro método para análise de sentimentos da literatura. Utilizamos o pacote Python SASA na versão 0.1.3, disponível em <https://pypi.python.org/pypi/sasa/0.1.3>.

Happiness Index. Happiness Index [Dodds and Danforth 2009] consiste de uma escala de sentimentos que utiliza o popular *Affective Norms for English Words* (ANEW) [Bradley and Lang 1999]. O ANEW é uma coleção de 1.034 palavras associadas a dimensões afetivas de valência, excitação e dominância. Happiness Index foi construído baseado no ANEW e calcula pontuações com valores entre 1 e 9 para um texto dado, indicando a “quantidade” de felicidade que existe naquele texto. Autores calcularam a frequência em que cada palavra do ANEW aparece no texto e então computa o peso médio encontrado, levando em consideração apenas o sentimento de valência. Com objetivo de adaptar o Happiness Index para detectar polaridade, consideramos que qualquer texto classificado pelo método no intervalo [1..5) como sendo negativo e [5..9] como sendo positivo.

PANAS-t. O PANAS-t [Goncalves et al. 2012] é uma escala psicométrica para detecção de humor que captura flutuações de humor no Twitter. O método consiste de uma versão adaptada do *Positive Affect Negative Affect Scale* (PANAS) [Watson and Clark 1985], que é uma escala bastante conhecida na psicologia. O PANAS-t é baseado em um largo conjunto de palavras associadas a 11 sentimentos: jovialidade, autoconfiança, serenidade, surpresa, medo, tristeza, culpa, hostilidade, timidez, fadiga e atenção. O método foi desenvolvido para detectar qualquer acréscimo ou decréscimo de sentimentos ao longo de um período. Um método similar ao PANAS-t consiste de uma adaptação do *Profile of Mood States* (POMS) [Bollen et al. 2009], uma escala psicológica que mede 6 escalas de humor: tensão, depressão, raiva, vigor, fadiga e confusão. Entretanto, esse método não foi incluído nas nossas análises pois este não está disponível mesmo sob requerimento.

NRC Emotion Lexicon. O NRC Emoticon Lexicon [Mohammad 2012] é um método léxico que classifica textos em 8 categorias afetivas, definidas por [Plutchik 1980], são elas: alegria, tristeza, raiva, medo, confiança, desgosto, antecipação e surpresa. O método foi desenvolvido utilizando uma base de dados chamada pelos autores de EmoLex. O EmoLex consiste de palavras rotuladas no serviço Amazon Mechanical Turk service [Amazon 2005], assim como palavras do *General Inqui-*

rer [Stone et al. 1966] and *WordNet Affect Lexicon* (WAL) [Valitutti 2004]. O processo de tradução do NRC Emotion Lexicon foi possível com a tradução automática das palavras do dicionário, utilizando o serviço Google Translate [Google].

NRC Hashtag Sentiment Lexicon. O NRC Hashtag Sentiment Lexicon, ou EmoLex [Mohammad et al. 2013] é um dicionário léxico *hashtags* do Twitter também associadas a 8 categorias de afeto: alegria, tristeza, raiva, medo, confiança, desgosto, antecipação e surpresa. O dicionário foi criado com cerca de 775 mil tweets que contêm *hashtags* consideradas pelos autores como possuindo algum grau de positividade (ex.: #good e #excellent) e negatividade (i.e.: #anger e #disgust).

Sentiment140 Lexicon. O Sentiment140 Lexicon [Mohammad et al. 2013] é também um dicionário léxico de palavras associadas a sentimentos positivos e negativos. O método foi criado com uma base de dados 1ue consistia de cerca de 1,6 milhões de tweets rotulados como positivo ou negativo.

OpinionLexicon. OpinionLexicon [Hu and Liu 2004], também conhecido como Sentiment Lexicon, consiste de uma lista com cerca de 6.800 palavras rotuladas como positivas e 6.800 palavras rotuladas como negativas, incluindo gírias e abreviações no idioma Inglês.

VADER. Proposto por [Hutto and Gilbert 2014], VADER (Valence Aware Dictionary for sEntiment Reasoning) é um método para análise de sentimentos desenvolvido para o contexto de mídias sociais online. VADER é um método de análise de sentimentos validado por humanos desenvolvido com foco em avaliar mensagens no contexto do Twitter e outras mídias sociais sem requerer treinamento. Dado ao foco para mídias sócias, os autores desenvolveram técnicas para ir além do modelo de *bag-of-words*. Foram avaliados diversas pontos que envolvem sentimentos excesso de pontuação, uso de letras maiúsculas, conjunções que invertem a polaridade do sentimento, dentre várias outros pontos sintáticos na mensagem. VADER está disponível para download em <http://comp.social.gatech.edu/papers/>.

Emoticons. O método Emoticons consiste de um conjunto dos emoticons populares em sites como Yahoo! e MSN [Messenger 2014, MSGWeb 2006], construído por [Goncalves and Benevenuto 2013], e inclui as variações mais comuns para emoticons que expressam polaridade positiva, negativa e neutra. Uma das principais vantagens de um método para detecção de sentimentos baseados na detecção de emoticons é que esses são independentes de idioma já que um mesmo símbolo que representa um emoticon é utilizado em diversos países da mesma forma.

4. Metodologia

Depois de apresentarmos os 13 métodos de análise de sentimentos investigados neste trabalho, descreveremos como foi realizada a validação da abordagem multilíngue proposta, identificando as bases de dados e as métricas utilizadas para este fim.

4.1. Estratégia

A estratégia adotada para este trabalho consistiu na tradução das nossas base de dados rotuladas em cada um dos idiomas para o idioma Inglês. Dessa forma, poderíamos simplesmente executar cada método em sua versão original (desenvolvida e validada para o

| Idioma | Descrição/Conteúdo | # Pos | # Neg |
|-----------|---|-------|-------|
| Árabe | Tweets de contextos variados em árabe [Abdulla et al. 2013] | 250 | 250 |
| Holandês | Tweets de contextos variados em holandês ⁴ | 52 | 52 |
| Francês | Tweets de contextos variados em francês [Sascha Narr and Albayrak 2012] | 250 | 250 |
| Alemão | Tweets de contextos variados em alemão [Sascha Narr and Albayrak 2012] | 250 | 250 |
| Italiano | Tweets de contextos variados em italiano ⁵ | 250 | 250 |
| Português | Tweets de contextos variados em português [Sascha Narr and Albayrak 2012] | 250 | 250 |
| Russo | Tweets de contextos variados em russo ⁶ | 250 | 250 |
| Espanhol | Tweets de contextos variados espanhol ⁷ | 250 | 250 |
| Inglês | Tweets de contextos variados em inglês [Sascha Narr and Albayrak 2012] | 250 | 250 |
| Turco | Reviews de produtos em turco [Demirtas and Pechenizkiy 2013] | 250 | 250 |

Tabela 1. Descrição das bases de dados rotuladas

inglês) na base traduzida. O processo de tradução foi possível com o uso do pacote python *Goslate API*², que utiliza a API do *Google Translate*³ para realizar traduções de textos. A maturidade e confiabilidade desse sistema já foi investigada em trabalhos anteriores [Balahur and Turchi 2012].

4.2. Base de Dados

Neste trabalho contamos com bases de dados rotuladas em nove idiomas: Português, Francês, Espanhol, Italiano, Turco, Russo, Árabe, Holandês, Alemão e Inglês. A Tabela 1 apresenta uma breve descrição de cada base de dados em relação a proporção de mensagens positivas e negativas e contexto das mensagens rotuladas.

Como essas bases de dados são rotulada, foi possível analisar a capacidade de predição de cada método para identificar a polaridade das mensagens nos idiomas escolhidos para análise.

4.3. Métricas de Avaliação

Para medir o desempenho da classificação com múltiplas classes, utilizamos as medidas convencionais de Recuperação de Informação: revocação, precisão e *Medida – F1*. A revocação (R) de uma classe é a fração do número de mensagens corretamente classificadas pelo número de mensagens nessa classe. Precisão (P) da classe é a fração do número de mensagens corretamente classificadas pelo total de mensagens preditas como mensagens dessa classe. Com o objetivo de explicar tais métricas, construímos a matriz de confusão apresentada no quadro a seguir:

| | | <i>Observação real</i> | |
|--------------------------|----------|------------------------|----------|
| | | Positivo | Negativo |
| <i>Predição esperada</i> | Positivo | a | b |
| | Negativo | c | d |

Cada posição neste quadro representa o número de elementos em cada classe, e como essas foram previstas pelo modelo utilizado na classificação. Assim, conforme essa matriz, a precisão (P_{pos}) e revocação (R_{pos}) para a classe positiva podem ser calculadas da seguinte forma: $P_{pos} = a/(a + b)$ e $R_{pos} = a/(a + c)$. Já a acurácia e *Medida – F1*, podem ser calculados como: $A = (a + d)/(a + b + c + d)$ e $F1 = 2 \cdot (P \cdot R)/(P + R)$, respectivamente.

²<http://pythonhosted.org/goslate/>

³<https://translate.google.com>

| | Alemão | Árabe | Espanhol | Francês | Holandês | Inglês | Português | Russo | Turco | Italiano |
|-----------------|--------|-------|----------|---------|----------|--------|-----------|-------|-------|----------|
| Emolex | 50% | 56% | 58% | 51% | 72% | 51% | 58% | 47% | 56% | 12% |
| Emoticons | 19% | 1% | 5% | 10% | 0% | 13% | 5% | 6% | 5% | 1% |
| Happiness-Index | 59% | 50% | 52% | 57% | 66% | 61% | 48% | 59% | 63% | 2% |
| NRC-Hashtag | 92% | 76% | 92% | 95% | 93% | 93% | 93% | 692% | 94% | 64% |
| LIWC | 60% | 49% | 61% | 64% | 73% | 71% | 63% | 21% | 64% | 77% |
| Opinion-Lexicon | 51% | 48% | 49% | 55% | 66% | 54% | 51% | 46% | 57% | 5% |
| PANAS-t | 8% | 3% | 5% | 8% | 13% | 11% | 4% | 8% | 8% | 4% |
| SASA | 64% | 64% | 51% | 61% | 57% | 69% | 60% | 56% | 64% | 17% |
| SentiWordNet | 90% | 77% | 91% | 89% | 93% | 87% | 89% | 90% | 91% | 27% |
| SentiStrength | 77% | 62% | 76% | 82% | 78% | 89% | 77% | 81% | 79% | 14% |
| SenticNet | 90% | 82% | 95% | 92% | 89% | 91% | 94% | 91% | 91% | 50% |
| Sentiment140 | 95% | 79% | 91% | 95% | 91% | 94% | 92% | 92% | 95% | 49% |
| Vader | 42% | 32% | 38% | 47% | 60% | 58% | 39% | 45% | 45% | 3% |

Tabela 2. Abrangência dos métodos nos 9 idiomas analisados

4.4. Linha de base

Com o objetivo de verificar a qualidade das abordagens propostas nesse trabalho para detecção de sentimentos no contexto dos diversos idiomas fixamos como linha de base 50% de acurácia devido às 2 classes possíveis, positiva e negativa. Idealmente, uma abordagem de tradução deveria possuir capacidade de predição superior a esta linha de base.

Nas próximas seções apresentaremos os resultados da capacidade de predição e abrangência dos 13 métodos utilizados nesse trabalho, comparando nossa abordagem de tradução com a linha de base de referência. Além disso, será possível ter um referencial de comparação com os resultados do idioma Inglês.

5. Resultados Experimentais

Com o objetivo de identificar vantagens, desvantagens e possíveis limitações da abordagem multilíngua para métodos de análise de sentimentos discutidos nesse trabalho em cada um dos 9 idiomas considerados, apresentamos os resultados das comparações e análises realizadas. A ferramenta iFeel [Araújo et al. 2014] – implementada por parte dos autores deste trabalho em outra pesquisa – disponibilizada publicamente em www.ifeel.dcc.ufmg.br foi utilizada para executar os métodos presentes neste trabalho, com exceção do método LIWC.

5.1. Abrangência

Para cada subconjunto de dados descrito na Tabela 1, computamos a abrangência dos métodos analisados. Como dito anteriormente, consideramos como abrangência de um método a porcentagem de mensagens na base em que cada método conseguiu detectar algum sentimento (positivo ou negativo). Isso é de fundamental importância para compreender a aplicabilidade de cada método em diferentes contextos das bases de dados nos diversos idiomas. Como forma de analisar a abrangência da nossa abordagem multilíngua para métodos de sentimentos nas bases de dados nos idiomas considerados, apresentamos a Tabela 2.

Como podemos observar, os métodos SentiWordNet, Sentiment140 Lexicon e SenticNet alcançaram as 3 mais altas abrangências em todos os idiomas considerados. Tal resultado mostra uma recorrente vantagem desses métodos em relação aos outros em relação a porcentagem de dados de uma base de dados que esses conseguem cobrir.

5.2. Capacidade de Predição

Com o objetivo de analisar a qualidade da abordagem multilíngua proposta, apresentamos o resultado da acurácia de predição dos 13 métodos analisados nas bases dos 9 idiomas

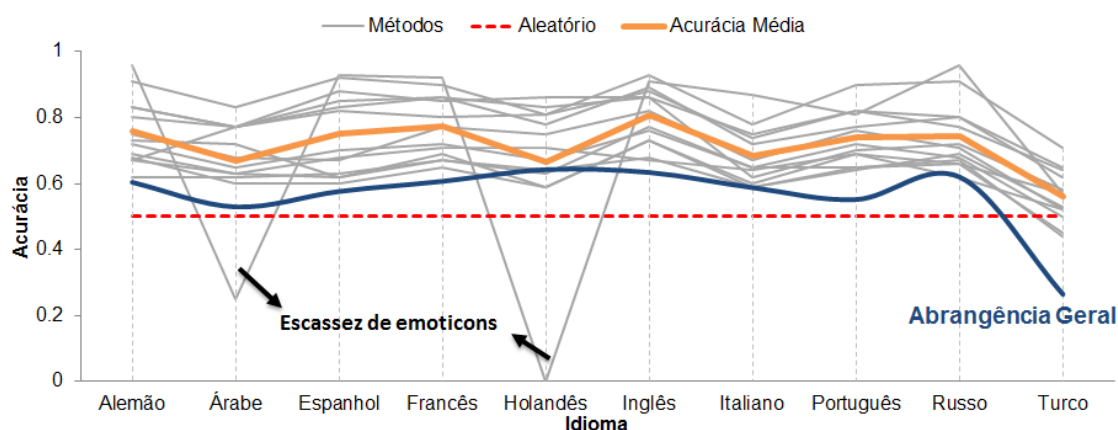


Figura 1. Gráfico destaca acurácia e abrangência dos métodos para a abordagem multilíngue proposta.

considerados. Na Figura 1, além dos resultados de abrangência discutidos anteriormente, podemos verificar a performance de predição dos métodos de modo geral e comparar com a linha de base tracejada no gráfico. Observamos que em média (destaque em laranja) sempre alcançamos resultados acima da linha de base.

A Tabela 3 apresenta uma descrição detalhada dos resultados da acurácia visualizada na Figura 1. Os dois pontos mais baixos do gráfico correspondem a acurácia do método Emoticons nas bases dos idiomas Árabe e Holandês, que continham poucos ou nenhum emoticon, no entanto, o método Emoticons obteve as maiores taxas de predição nos outros idiomas devido a sua independência em relação ao idioma da base de dados e sua capacidade de extrair o sentimento quase rotulado pelo autor da mensagem através dos emoticons próprios de textos de mídias sociais.

| | Alemão | Arabe | Espanhol | Francês | Holandês | Inglês | Italiano | Português | Russo | Turco |
|-----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------------|-------------|-------------|
| Emolex | 0.75 | 0.68 | 0.67 | 0.77 | 0.75 | 0.82 | 0.67 | 0.76 | 0.71 | 0.52 |
| Emoticons | 0.96 | 0.25 | 0.93 | 0.92 | 0 | 0.91 | 0.87 | 0.81 | 0.96 | 0.57 |
| Happiness-Index | 0.62 | 0.62 | 0.63 | 0.67 | 0.64 | 0.68 | 0.59 | 0.65 | 0.67 | 0.44 |
| NRC-Hashtag | 0.73 | 0.64 | 0.64 | 0.67 | 0.63 | 0.75 | 0.6 | 0.66 | 0.64 | 0.53 |
| LIWC | 0.8 | 0.77 | 0.83 | 0.86 | 0.78 | 0.89 | 0.72 | 0.77 | 0.8 | 0.62 |
| Opinion-Lexicon | 0.83 | 0.77 | 0.82 | 0.8 | 0.81 | 0.88 | 0.74 | 0.82 | 0.8 | 0.65 |
| PANAS-t | 0.67 | 0.77 | 0.88 | 0.85 | 0.86 | 0.86 | 0.62 | 0.7 | 0.72 | 0.58 |
| SASA | 0.68 | 0.6 | 0.6 | 0.65 | 0.59 | 0.73 | 0.6 | 0.69 | 0.62 | 0.52 |
| SentWordNet | 0.67 | 0.63 | 0.68 | 0.71 | 0.71 | 0.67 | 0.64 | 0.69 | 0.66 | 0.57 |
| SentiStrength | 0.83 | 0.77 | 0.85 | 0.86 | 0.83 | 0.86 | 0.75 | 0.82 | 0.77 | 0.64 |
| SenticNet | 0.69 | 0.63 | 0.62 | 0.69 | 0.59 | 0.73 | 0.59 | 0.64 | 0.69 | 0.53 |
| Sentiment140 | 0.72 | 0.63 | 0.68 | 0.7 | 0.69 | 0.76 | 0.61 | 0.67 | 0.62 | 0.51 |
| Vader | 0.91 | 0.83 | 0.92 | 0.9 | 0.81 | 0.93 | 0.78 | 0.9 | 0.91 | 0.71 |

Tabela 3. Acurácia dos métodos nos idiomas considerados

A Tabela 4 apresenta o resultado da métrica F-measure separado em capacidade de predição de positivo e negativo de cada método em cada idioma considerado. Observamos que existe uma tendência de 12 dos 13 métodos em analisar com mais qualidade textos positivos do que negativos. Apenas o método NRC Hashtag Sentiment Lexicon apresenta um comportamento fora deste padrão sendo capaz de avaliar melhor sentenças com caráter negativo em todos os idiomas. Estes resultados mostram como o problema de análise de sentimentos pode ser dividido em duas etapas: análise de sentimentos positivos e análise de sentimentos negativos, a fim de atribuir diferentes metodologias para uma classificação otimizada.

| | Alemão | | Árabe | | Espanhol | | Francês | | Holandês | |
|----------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | Pos | Neg | Pos | Neg | Pos | Neg | Pos | Neg | Pos | Neg |
| SentiWordNet | 0.70 | 0.62 | 0.71 | 0.48 | 0.73 | 0.60 | 0.77 | 0.65 | 0.74 | 0.67 |
| Emoticons | 0.97 | 0.92 | 0.40 | 0.00 | 0.96 | 0.67 | 0.95 | 0.75 | 0 | 0 |
| PANAS-t | 0.70 | 0.63 | 0.67 | 0.82 | 0.91 | 0.84 | 0.82 | 0.88 | 0.83 | 0.88 |
| SASA | 0.71 | 0.63 | 0.58 | 0.62 | 0.72 | 0.36 | 0.73 | 0.51 | 0.68 | 0.43 |
| SenticNet | 0.72 | 0.47 | 0.73 | 0.44 | 0.70 | 0.47 | 0.75 | 0.52 | 0.67 | 0.47 |
| Happiness Index | 0.70 | 0.45 | 0.75 | 0.24 | 0.73 | 0.47 | 0.75 | 0.52 | 0.73 | 0.47 |
| SentiStrength | 0.84 | 0.81 | 0.77 | 0.77 | 0.86 | 0.84 | 0.88 | 0.84 | 0.82 | 0.83 |
| Emolex | 0.78 | 0.72 | 0.74 | 0.58 | 0.70 | 0.64 | 0.79 | 0.77 | 0.78 | 0.7 |
| NRC-Hashtag | 0.68 | 0.75 | 0.62 | 0.67 | 0.57 | 0.69 | 0.64 | 0.71 | 0.74 | 0.76 |
| OpinionLexicon | 0.85 | 0.80 | 0.81 | 0.73 | 0.84 | 0.79 | 0.82 | 0.77 | 0.83 | 0.79 |
| Sentiment140 Lexicon | 0.65 | 0.74 | 0.65 | 0.61 | 0.66 | 0.7 | 0.67 | 0.73 | 0.67 | 0.72 |
| VADER | 0.94 | 0.85 | 0.89 | 0.66 | 0.93 | 0.88 | 0.92 | 0.85 | 0.84 | 0.75 |
| LIWC | 0.84 | 0.73 | 0.83 | 0.67 | 0.87 | 0.76 | 0.89 | 0.82 | 0.82 | 0.70 |
| | Inglês | | Italiano | | Português | | Russo | | Turco | |
| | Pos | Neg | Pos | Neg | Pos | Neg | Pos | Neg | Pos | Neg |
| SentiWordNet | 0.71 | 0.62 | 0.69 | 0.55 | 0.72 | 0.64 | 0.69 | 0.61 | 0.64 | 0.46 |
| Emoticons | 0.94 | 0.81 | 0.92 | 0.57 | 0.82 | 0.79 | 0.97 | 0.91 | 0.67 | 0.4 |
| PANAS-t | 0.89 | 0.8 | 0.64 | 0.6 | 0.71 | 0.68 | 0.73 | 0.72 | 0 | 0.73 |
| SASA | 0.75 | 0.71 | 0.66 | 0.5 | 0.73 | 0.64 | 0.61 | 0.64 | 0.39 | 0.61 |
| SenticNet | 0.78 | 0.66 | 0.69 | 0.4 | 0.72 | 0.51 | 0.74 | 0.63 | 0.67 | 0.19 |
| Happiness Index | 0.76 | 0.54 | 0.69 | 0.4 | 0.73 | 0.5 | 0.73 | 0.57 | 0.29 | 0.55 |
| SentiStrength | 0.86 | 0.85 | 0.74 | 0.75 | 0.83 | 0.81 | 0.77 | 0.77 | 0.71 | 0.55 |
| Emolex | 0.82 | 0.82 | 0.65 | 0.69 | 0.77 | 0.76 | 0.7 | 0.72 | 0.25 | 0.64 |
| NRC-Hashtag | 0.71 | 0.78 | 0.48 | 0.68 | 0.62 | 0.68 | 0.53 | 0.71 | 0.6 | 0.45 |
| OpinionLexicon | 0.89 | 0.87 | 0.76 | 0.71 | 0.83 | 0.82 | 0.79 | 0.81 | 0.71 | 0.57 |
| Sentiment140 Lexicon | 0.76 | 0.76 | 0.66 | 0.63 | 0.73 | 0.71 | 0.66 | 0.7 | 0.58 | 0.4 |
| VADER | 0.94 | 0.91 | 0.84 | 0.67 | 0.92 | 0.87 | 0.91 | 0.9 | 0.8 | 0.5 |
| LIWC | 0.91 | 0.88 | 0.78 | 0.63 | 0.84 | 0.56 | 0.81 | 0.79 | 0.72 | 0.41 |

Tabela 4. F-measure das polaridades positivo e negativo dos métodos nos idiomas considerados

A partir dos resultados apresentados, observamos que não há uma relação direta entre o tipo do método (que usa abordagem de aprendizado de máquina ou que adota um dicionário léxico) com seu desempenho em termos das métricas de avaliação. Ou seja, um método que foi treinado utilizando técnicas de aprendizado de máquina não necessariamente tem melhor desempenho que os métodos que utilizam dicionários léxicos para a tarefa de análise de sentimentos. Isso pode ser observado pelos resultados dos métodos de aprendizado de máquina SentiStrength e SASA, que aparecem com frequência na lista dos melhores e dos piores métodos, respectivamente. O mesmo acontece com métodos léxicos, por exemplo, os métodos léxicos VADER e PANAS-t aparecem com frequência na lista dos melhores métodos, mas os léxicos NRC Hashtag Sentiment Lexicon e Happiness Index aparecem na lista dos métodos com piores performances.

No entanto, podemos observar a prevalência de alguns métodos no ranking de métodos com melhores performances. É o caso do Vader, SentiStrength e do Emoticons, que aparecem no Top-3 em 80% dos idiomas. Apesar da alta acurácia o Vader e o Emoticons possuem abrangência abaixo da média geral dos métodos.

6. Conclusões e Trabalhos Futuros

Pesquisas recentes em redes sociais online vêm adotando vários métodos para análise de sentimentos em conteúdo Web, onde o usuário é um ativo gerador de conteúdo. Vários desses métodos se tornaram populares e tem sido utilizados como ferramentas para medir o sentimento ou polaridade em redes sociais online. No entanto, poucos esforços vêm sendo feitos no desenvolvimento de métodos para detecção de sentimentos em mensagens em idiomas diferentes do inglês.

Nesse trabalho, apresentamos uma abordagem para detecção de sentimentos para mensagens compartilhadas em aplicações da Web. Traduzimos bases de dados rotuladas em 9 idiomas diversos para o idioma inglês e executamos 13 métodos de análise de sentimentos em sua versão original, proposta para a língua inglesa. Executamos experimentos com diferentes métricas para medir a eficácia de cada método.

A partir da análise de nossos resultados fundamentamos que a tradução de bases de dados para outras línguas não interfere de maneira significativa a acurácia e nem a abrangência de métodos para análise de sentimentos. Desta forma é viável a análise de sentimentos em textos escritos em outras línguas desde que a tradução para o idioma original do método não seja um problema.

Como contribuição, pretendemos disponibilizar a abordagem proposta para a comunidade científica bem como as bases rotuladas, facilitando que melhorias e investigações neste cenário possam ser aprofundadas. Contudo, percebemos que os métodos possuem variados graus de abrangência e acurácia competitiva, mas que não existe um método com melhores resultados para as diferentes aplicações e contextos.

Como trabalhos futuros pretendemos incluir nesta análise mais métodos de análise de sentimentos, assim como base de dados em outros idiomas. Pretendemos também comparar a distância linguística de cada idioma para identificar possíveis relações entre idiomas com mesma família terem resultados similares.

7. Agradecimentos

Esse trabalho foi apoiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela Fundação de Amparo à Pesquisa do estado de Minas Gerais (FAPEMIG), pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e também pelo Instituto Nacional de Ciência e Tecnologia para a Web (InWeb).

Referências

- (2013). Sentistrength 2.0. <http://sentistrength.wlv.ac.uk/Download>.
- Abdel-Hady, M., Mansour, R., and Ashour, A. (2014). Cross-lingual twitter polarity detection via projection across word-aligned corpora. *ICML WISDOM 2014 Conference*.
- Abdul-Mageed, M., Kübler, S., and Diab, M. (2012). Samar: A system for subjectivity and sentiment analysis of arabic social media. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pages 19–28. Association for Computational Linguistics.
- Abdulla, N., Ahmed, N., Shehab, M., and Al-Ayyoub, M. (2013). Arabic sentiment analysis: Lexicon-based and corpus-based. In *Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on*, pages 1–6.
- Amazon (2005). Amazon mechanical turk. <https://www.mturk.com/>. Accessed June 17, 2013.
- Araújo, M., Gonçalves, P., Cha, M., and Benevenuto, F. (2014). ifeel: A system that compares and combines sentiment analysis methods. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 75–78.
- Balahur, A. and Turchi, M. (2012). Multilingual sentiment analysis using machine translation? In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pages 52–60. Association for Computational Linguistics.
- Banea, C., Mihalcea, R., Wiebe, J., and Hassan, S. (2008). Multilingual subjectivity analysis using machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 127–135. Association for Computational Linguistics.
- Bollen, J., Pepe, A., and Mao, H. (2009). Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena. *CoRR*, abs/0911.1583.
- Bradley, M. M. and Lang, P. J. (1999). Affective norms for English words (ANEW): Stimuli, instruction manual, and affective ratings. Technical report, Center for Research in Psychophysiology, University of Florida.
- Cambria, E., Speer, R., Havasi, C., and Hussain, A. (2010). Senticnet: A publicly available semantic resource for opinion mining. In *AAAI Fall Symposium Series*.
- Demirtas, E. and Pechenizkiy, M. (2013). Cross-lingual polarity detection with machine translation. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining, WISDOM '13*, pages 9:1–9:8. ACM.
- Dodds, P. and Danforth, C. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *J. of Happiness Studies*, 11.

- Dodds, P. S. and Danforth, C. M. (2009). Measuring the happiness of large-scale written expression: songs, blogs, and presidents. *J. of Happiness Studies*, 11.
- dos Santos, A. G. L., Becker, K., and Moreira, V. (2014). Um estudo de caso de mineração de emoções em textos multilíngues. In *Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM'13)*.
- Esuli and Sebastiani (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proc. LREC*.
- Goncalves, P., Araujo, M., Benevenuto, F., and Cha, M. (2013). Comparing and combining sentiment analysis methods. In *Proc. COSN*.
- Goncalves, P. and Benevenuto, F. (2013). O que tweets contendo emoticons podem revelar sobre sentimentos coletivos? In *II Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*.
- Goncalves, P., Dores, W., and Benevenuto, F. (2012). Panas-t: Uma escala psicometrica para analise de sentimentos no twitter. In *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*.
- Google. Google translate. <https://translate.google.com>. Acessado em Março de 2015.
- Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. *Proc. KDD'04*, pages 168–177.
- Hutto, C. J. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *ICWSM*.
- Lin, Z., Tan, S., and Cheng, X. (2011). Language-independent sentiment classification using three common words. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1041–1046. ACM.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.
- Messenger, Y. (2014). Yahoo messenger emoticons. <http://messenger.yahoo.com/features/emoticons>.
- Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38.
- Mohammad, S. (2012). #emotional tweets. In *SEM*.
- Mohammad, S. M., Kiritchenko, S., and Zhu, X. (2013). Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. In *Proc. SemEval-2013*.
- Mohammad, S. M. and Turney, P. D. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon.
- MSGWeb (2006). List of emoticons in msn messenger. <http://messenger.msn.com/Resource/Emoticons.aspx>.
- Narayanan, R., Liu, B., and Choudhary, A. (2009). Sentiment analysis of conditional sentences. In *Proc. of the Conference on Empirical Methods in Natural Language Processing: Volume 1*, pages 180–189. Association for Computational Linguistics.
- Narr, S., Hulphenhaus, M., and Albayrak, S. (2012). Language-independent twitter sentiment analysis. *Knowledge Discovery and Machine Learning (KDML), LWA*, pages 12–14.
- Online, N. (2009). Social networks and blogs now 4th most popular online activity, ahead of personal email, nielsen reports. http://www.nielsen.com/us/en/press-room/2009/social_networks_.html. Accessed in May, 05, 2014.
- Plutchik, R. (1980). *A general psychoevolutionary theory of emotion*, pages 3–33. Academic press, New York.
- Remus, R., Quasthoff, U., and Heyer, G. (2010). Sentiws-a publicly available german-language resource for sentiment analysis. In *LREC*.
- Sascha Narr, M. H. and Albayrak, S. (2012). Language-independent twitter sentiment analysis. In *Workshop on Knowledge Discovery, Data Mining and Machine Learning (KDML-2012), Dortmund, Germany*.
- Souza, M. and Vieira, R. (2012). Sentiment analysis on twitter data for portuguese language. In *Computational Processing of the Portuguese Language*, pages 241–247. Springer.
- Stone, P. J., Dunphy, D. C., Smith, M. S., and Ogilvie, D. M. (1966). *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.
- Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *J. of Lang. and Soc. Psych.*, 29.
- Thelwall, M. (2013). Heart and soul: Sentiment strength detection in the social web with sentistrength. <http://sentistrength.wlv.ac.uk/documentation/SentiStrengthChapter.pdf>.
- Valitutti, R. (2004). Wordnet-affect: an affective extension of wordnet. In *In Proceedings of the 4th International Conference on Language Resources and Evaluation*.
- Wang, H., Can, D., Kazemzadeh, A., Bar, F., and Narayanan, S. (2012). A system for real-time twitter sentiment analysis of 2012 u.s. presidential election cycle. In *ACL System Demonstrations*.
- Watson, D. and Clark, L. (1985). Development and validation of brief measures of positive and negative affect: the panas scales. *J. of Pers. and So. Psych.*, 54.
- Yussupova, N., Bogdanova, D., and Boyko, M. (2012). Applying of sentiment analysis for texts in russian based on machine learning approach. In *IMMM 2012, 2nd International Conference on Advances in Information Mining and Management*, pages 8–14.