

Um estudo de caso de mineração de emoções em textos multilíngues

Aline Graciela Lermen dos Santos¹, Karin Becker¹, Viviane Moreira¹

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

{aglsantos, karin.becker, viviane}@inf.ufrgs.br

Abstract. *Multilingual Opinion Mining deals with the analysis of opinions regardless of the language in which they are written. The vast majority of the work in this area focuses solely on classifying the polarity of the sentiment, overlooking the analysis of the emotions. In order to fill this gap, this work presents a case study about the classification of emotions present in product reviews, evaluating an approach that combines lexicon-based emotion classification and automatic translation. The case study aims at finding out if it is best to translate the text of the reviews or the dictionary. It also evaluates whether lemmatization can bring any benefits. The results of our experiments on real data show that translating the reviews yields better results and that lemmatization does not bring significant changes.*

Resumo. *O objetivo da Mineração de Opinião Multilíngue é extrair e analisar textos contendo opiniões, independente do idioma no qual estão escritos. A grande maioria dos trabalhos desta área foca apenas na classificação da polaridade do sentimento, sendo que a análise das emoções é pouco explorada. A fim de preencher esta lacuna, este trabalho apresenta um estudo de caso sobre a classificação das emoções presentes em revisões de produtos, avaliando uma abordagem que combina a classificação de emoções baseada em dicionário e tradução automática. O estudo de caso tem por objetivo identificar se é preferível traduzir o texto das revisões ou as palavras do dicionário, além de avaliar se a aplicação de um lematizador melhora os resultados. Os resultados dos experimentos em dados reais mostram que a tradução das revisões produz melhores resultados e que a lematização não traz mudanças significativas.*

1. Introdução

Análise de Sentimentos (também chamada de Mineração de Opinião) é o estudo computacional de opiniões, sentimentos e emoções expressos em textos [Liu 2012]. A partir de uma coleção de documentos, a Análise de Sentimentos se propõe a, automaticamente, identificar, classificar e agregar o sentimento a respeito de um alvo. Várias aplicações têm sido propostas para a Análise de Sentimentos, tais como [Liu 2012; Tsytarau and Palpanas 2012]: sumarização da opinião prevalecente em revisões de produtos; termômetros de popularidade para marcas, pessoas ou organizações, usadas em ações de marketing ou relações públicas; previsão de indicadores a partir de sentimentos (e.g. preços, resultados de eleições, movimentos da bolsa, etc.).

Um sentimento representa uma atitude, opinião ou emoção que o autor da opinião tem a respeito do alvo [Liu 2012]. A maioria dos trabalhos mensura este

sentimento na forma de polaridade, i.e. um ponto em alguma escala que representa a avaliação positiva, neutra ou negativa do significado deste sentimento. Já emoção é uma medida mais complexa, podendo abranger diversas categorias, cuja classificação não é conclusiva. Por exemplo, a surpresa pode ser considerada uma emoção positiva ou negativa, necessitando contexto para desambiguação. As abordagens para classificar o sentimento dividem-se em [Tsytarau and Palpanas 2012]: a) baseadas em dicionário, onde um léxico de sentimentos é utilizado; b) baseadas em aprendizado de máquina, onde algoritmos de classificação são treinados sobre um *corpus* rotulado; e c) estatísticas, usando medidas de co-ocorrência (e.g. PMI).

A Web tem sido muito explorada como fonte de opiniões, pela quantidade e riqueza do conteúdo que disponibiliza. Contudo, na prática o inglês é o idioma dominante no conteúdo disponibilizado. Duas consequências imediatas desta situação são: a) os recursos e técnicas disponíveis à Análise de Sentimentos são voltados prioritariamente à língua inglesa; e b) a oportunidade de processar indistintamente opiniões expressas em distintos idiomas, quer pela inexistência de conteúdo em uma dada língua alvo, quer como forma de contrastar opiniões expressas em diferentes contextos culturais. A Mineração de Opiniões Multilíngue propõe técnicas para classificação do sentimento sem uma linguagem alvo definida [Banea et al. 2008].

A maioria dos trabalhos em Mineração de Opinião Multilíngue trata da classificação do sentimento de acordo com o conceito de polaridade [Banea et al. 2008; Narr et al. 2012; Bader et al. 2011; Lin et al. 2011]. A Mineração de Emoções Multilíngue é uma área pouco explorada, e este trabalho se propõe a dar alguns passos iniciais através de um estudo de caso.

Este trabalho apresenta um estudo de caso onde técnicas da Mineração de Opinião Multilíngue são empregadas para classificar emoções. O objetivo é verificar se propostas da Mineração de Opinião Multilíngue voltadas à classificação de polaridade baseadas em dicionário e tradução automática (como a de [Banea et al. 2008]) apresentam resultados satisfatórios quando aplicadas à classificação de emoções. Para isso, foram desenvolvidos experimentos que, além de investigar a possibilidade de tratar emoções de modo similar a polaridade, buscam estabelecer qual o melhor emprego da tradução: sobre o texto contendo o sentimento, ou sobre o dicionário usado na classificação. Ainda, buscou-se verificar o efeito da lematização na classificação da emoção, já que nem sempre normalizações sobre termos denotando sentimento têm levado a bons resultados [Liu 2012]. Um lema corresponde à forma canônica de um termo (e.g. "amor" é o lema de "amores" e "amor"), aumentando a probabilidade de sucesso na busca de termos em dicionários de sentimento. Comparou-se assim o resultado de classificação buscando no dicionário de sentimento o termo originalmente encontrado no documento, ou com base em sua forma canônica. Os experimentos mostram que a tradução do texto de entrada produz resultados superiores aos da tradução do dicionário, e que a lematização não melhora de forma consistente e sistemática a classificação da emoção.

O restante deste trabalho está estruturado como segue: a Seção 2 descreve os trabalhos relacionados a Mineração de Opinião Multilíngue e Mineração de Emoções; a Seção 3 aborda o Estudo de caso, descrevendo *corpus* e recursos utilizados, os métodos de classificação, e os resultados dos experimentos; a Seção 4 discute as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

2.1 Mineração de Opinião Multilíngue

A maior parte dos recursos para análise de textos se encontra disponível apenas no idioma inglês. Os trabalhos nesta área propõem-se a: a) criar recursos para análise em outros idiomas utilizando técnicas como tradução ou explorando *corpora* paralelos; ou b) desenvolver técnicas para fazer a análise sem a necessidade de recursos (avançados).

A tradução automática, que pode ser aplicada sobre o texto sendo analisado ou sobre os dicionários de sentimentos usados para a classificação, é utilizada para gerar recursos multilíngues (e.g. *corpora* anotados, dicionários especializados) que viabilizem a Mineração de Opiniões. A vantagem desta abordagem é sua simplicidade, visto que existem serviços de tradução automática gratuitos disponíveis na Web. Porém, os resultados dependem da qualidade da tradução, que pode estar muito aquém da ideal. Nesta categoria, o trabalho de Banea et al. [2008] tem como objetivo gerar automaticamente *corpora* anotados para análise de subjetividade em outros idiomas que não possuem este tipo de recurso. O trabalho considera como entrada um *corpus* em inglês anotado com rótulos de polaridade, e desenvolve experimentos para verificar se estes rótulos podem ser projetados (i.e. permanecem válidos) para as respectivas traduções automáticas em romeno e espanhol. Os experimentos avaliaram 3 cenários: (a) tradução para espanhol e romeno de um *corpus* em inglês manualmente anotado, seguida da projeção dos rótulos originais às respectivas traduções; (b) similar ao cenário anterior, exceto que o *corpus* em inglês é anotado por uma ferramenta de anotação automática; (c) os *corpora* em espanhol e romeno, não anotados, são traduzidos automaticamente para inglês, e então anotados automaticamente. O método proposto é simples e apresentou bons resultados quando os *corpora* resultantes foram classificados quanto a sua polaridade.

Bader et al. [2011] propõem uma abordagem espaço-vetorial para predição de sentimentos de documentos em múltiplos idiomas, sem a necessidade de tradução. A abordagem utiliza indexação semântica latente, e assume como entrada *corpora* paralelos multilíngue, nos quais alguns documentos estão anotados quanto à polaridade. Seu ponto forte é permitir que as traduções dos termos de cada idioma sejam comparáveis entre si, o que Banea et al. [2010] mostraram ser uma grande vantagem.

Com uma abordagem diferenciada, Lin et al. [2011] mostram um método que não depende de tradução, nem de *corpora* paralelos. Os autores propõem utilizar apenas algumas palavras-semente (advérbios), e treinar um classificador usando um algoritmo de aprendizado não-supervisionado. Outro método é usar características independentes de idiomas, tais como *emoticons*, para rotular *tweets* como positivos e negativos [Narr et al. 2012]. Usando este método, os autores criaram automaticamente *corpora* de treinamento para diversos idiomas. Os idiomas analisados foram inglês, alemão, francês e português, gerando um classificador para cada idioma e um classificador combinando todos os idiomas. O idioma português obteve os piores resultados de classificação, visto que a expressão de sentimentos nem sempre é baseada em *emoticons* (e.g. “rsrsrsr” e “kkkkk”).

Pode-se ver que a Mineração de Opinião Multilíngue tem apresentado bons resultados para polaridade, independente da abordagem utilizada. Entretanto, emoção não tem sido abordada, o que incentiva o estudo de caso proposto. Ele visa analisar

emoções em textos em português, utilizando um léxico de sentimentos disponível apenas em inglês. A abordagem escolhida foi a de tradução automática, pela simplicidade e disponibilidade de recursos.

2.2 Mineração de Emoções

Pode-se avaliar o sentimento através de diversas medidas, como polaridade, emoção e força [Bravo-Marquez 2013]. Quando se usa emoção, o objetivo é classificar o sentimento em categorias como tristeza, alegria, surpresa, entre outras. Força define níveis de intensidade para a medida de sentimento, podendo ela ser emoção ou polaridade. O uso de polaridade é bem mais simples, já que tende a posicionar o sentimento em uma escala cujos sentimentos variam do negativo ao positivo, o que justifica sua popularidade. No entanto, alguns trabalhos relatam que o uso de emoções melhora significativamente os resultados da mineração de opiniões [Bollen et al. 2011; Asur and Huberman 2010].

Não existe consenso quanto às categorias básicas de emoção, e cada autor escolhe o grupo de emoções que mais se adéque aos seus objetivos [Ortony and Turner 1990]. Muitos trabalhos utilizam a classificação de Ekman [1992], que considera como básicas as emoções *alegria, surpresa, medo, tristeza, raiva e repugnância*¹. A classificação de Plutchik adiciona a esse conjunto duas emoções: *confiança e expectativa*². O léxico de sentimentos NRC (*word-emotion association*) [Mohammad and Turney 2013] é voltado ao idioma inglês, e associa a termos um peso (0 ou 1) para cada uma das 8 emoções de Plutchik, como ilustrado na Figura 1. O dicionário contém termos de diversas classes gramaticais e formas, incluindo lemas.

love	positive:1	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:1	sadness:0	surprise:0	trust:0
lovely	positive:1	negative:0	anger:0	anticipation:1	disgust:0	fear:0	joy:1	sadness:1	surprise:1	trust:1
lovmaking	positive:1	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:1	sadness:0	surprise:0	trust:1
lover	positive:1	negative:0	anger:0	anticipation:1	disgust:0	fear:0	joy:1	sadness:0	surprise:0	trust:1
loving	positive:1	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:1	sadness:0	surprise:0	trust:1
lower	positive:0	negative:1	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:1	surprise:0	trust:0

Figura 1. Trecho do dicionário de sentimentos NRC.

O presente trabalho contribui à Mineração de Emoções com um estudo que associa trabalhos de Mineração de Opinião Multilíngue com um dicionário de sentimentos que inclui emoções, o NRC, para classificação de sentimentos em textos em português.

3. Estudo de caso

Este trabalho apresenta um estudo de caso que se propõe, através de experimentos, a adaptar técnicas de Mineração de Opinião Multilíngue baseadas em tradução para classificar emoções em revisões de produto escritas em português. São utilizadas as 8 categorias de emoções disponíveis no NRC, o qual inclui termos exclusivamente em inglês.

¹ Termos originais: *joy, surprise, fear, sadness, anger e disgust*.

² Termos originais: *trust e anticipation*.

O objetivo do estudo de caso é verificar se traduzir textos contendo opiniões para um idioma em que haja recursos para tratar de emoções (no caso o inglês) e projetar os rótulos de emoções obtidos com as traduções para os textos no idioma original é uma abordagem viável para a classificação de emoções em Mineração de Opinião Multilíngue. No tocante ao uso da tradução, visa-se analisar qual abordagem produz o melhor resultado: a tradução da revisão ou a tradução do dicionário usado para a classificação.

Os idiomas envolvidos são português (idioma no qual estão escritas as revisões) e inglês (idioma do dicionário de sentimentos utilizado). O português foi escolhido por não haver muitos recursos disponíveis neste idioma, e o inglês por ter os recursos necessários disponíveis; pode-se expandir a quantidade de idiomas para o estudo de caso, bastando haver anotação para os outros idiomas.

Com o primeiro experimento, buscamos responder a questão “*É melhor traduzir a revisão ou o dicionário?*”. Para isto, avaliamos dois cenários: (a) a tradução do texto das revisões para inglês e (b) a tradução dos termos do dicionário para português. No primeiro cenário, a revisão em português passa pelo processo de tradução automática para inglês e o texto resultante é usado como entrada para um classificador, que atribui rótulos de emoção ao texto; os rótulos de emoções são então projetados para o texto original em português. No segundo cenário, o dicionário usado para classificação é traduzido para português, e então a revisão é diretamente rotulada pelo classificador. Neste experimento, as palavras das revisões foram procuradas diretamente no dicionário, sem nenhum pré-processamento.

O estudo de caso também investiga se a lematização do texto melhora a qualidade da classificação da emoção. Desta maneira, com o segundo experimento, procuramos responder a questão “*A lematização auxilia na classificação correta das emoções presentes nas revisões?*”. Nesta avaliação, as palavras do texto foram lematizadas. Com isso, intuitivamente, espera-se aumentar o número de rótulos de emoção encontrados. O lematizador foi aplicado sobre o texto das revisões traduzido para inglês, visto que não encontramos um lematizador para o português.

No restante da seção são descritos os recursos utilizados, o método de classificação empregado e discutidos os resultados obtidos.

3.1 Corpus e Recursos

Os dados utilizados foram revisões de *ebooks* Kindle extraídas do *site* Amazon (amazon.com.br). As razões para a escolha de revisões de livros como objeto de estudo foram: (a) maior possibilidade de encontrar emoções variadas, visto que livros podem evocar diferentes emoções; (b) os *ebooks* Kindle são vendidos em diversos países, acarretando grande variedade de idiomas, possibilitando a posterior expansão do estudo de caso para outros idiomas.

As revisões foram extraídas a partir da estrutura html das páginas dos produtos no site da Amazon, que necessita de poucos ajustes caso deseje-se extrair os dados das páginas em outros idiomas. Os campos usados nesse estudo de caso foram: título da revisão e corpo da revisão.

Para a classificação da emoção, foi adotado o dicionário de sentimentos NRC, apresentado na Seção 2.2, que classifica diversos termos em inglês segundo polaridade e

emoções. Este dicionário contém termos de diversas classes gramaticais, sendo que a maioria dos termos correspondem aos lemas. A única modificação feita no dicionário foi retirar os termos que apresentavam valor 0 para todas as categorias de emoções, por não contribuírem com informação relevante para o objetivo do estudo. Esta ação reduziu o dicionário de 14177 para 4460 termos. A Figura 2 mostra a distribuição das emoções no dicionário através da quantidade de termos contidos. A emoção predominante no NRC é medo, enquanto tristeza, raiva, repugnância e confiança são igualmente distribuídas. É importante notar que um termo pode expressar mais de uma emoção, como é o caso da palavra *lovely*, mostrada na Figura 1.

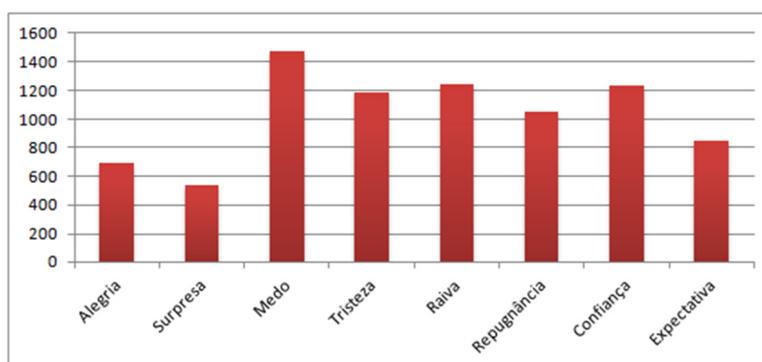


Figura 2. Distribuição das emoções no NRC.

O tradutor usado no experimento foi o Google Tradutor, através da biblioteca *goslate*³. Em alguns casos, um termo em inglês é traduzido para uma expressão em português. Isto não é um problema para a tradução da revisão, mas é para a tradução do dicionário, pois pode dificultar a busca. Sendo assim, optou-se por separar o dicionário traduzido em dois dicionários: a) dicionário de termos, incluindo apenas palavras simples (unigramas); e b) dicionário de termos compostos, contendo as traduções com mais de uma palavra.

Para a extração dos lemas no segundo experimento, foi utilizado o lematizador do pacote *NLTK*⁴, para processamento de linguagem natural. Foi utilizado também um etiquetador gramatical (*part-of-speech tagger*), contido no pacote.

O *corpus*, composto de 150 revisões, foi anotado manualmente, conforme as oito emoções do dicionário de emoções, por dois anotadores mestrandos em computação. A anotação foi intuitiva, com base na interpretação subjetiva do anotador para cada tipo de emoção, sem depender de conhecimento prévio do objeto da opinião. Para cada revisão, o anotador identificou a presença ou ausência de cada emoção no texto. Em outras palavras, a anotação foi binária, sem indicar intensidade de emoção no texto. Dessas 150 revisões, a maior parte é considerada positiva, em relação à quantidade de estrelas - em uma escala de 1 a 5 estrelas, cerca de 85% das revisões tinham 3 estrelas ou mais. A Figura 4 mostra as emoções anotadas nas revisões por cada anotador, e nela pode-se observar que, apesar dos anotadores terem uma boa concordância entre si, o *corpus* é desbalanceado: enquanto alegria está presente em mais da metade do *corpus*, algumas emoções estão pouco representadas, sendo raiva e tristeza as menos presentes. A

³ <https://pypi.python.org/pypi/goslate>

⁴ <http://www.nltk.org/>

justificativa para isso é que alegria é a emoção mais claramente positiva, e o *corpus* é predominantemente composto por revisões positivas. No conjunto das 150 revisões, os anotadores chegaram a um grau de concordância média de 55% por documento sobre a presença de emoções, com um desvio padrão de 33%.

Cada uma das oito emoções foi analisada e anotada de forma independente das outras. Foram feitos dois conjuntos de dados anotados: a) Dataset União, um conjunto usando a união das emoções anotadas por cada anotador, ou seja, as emoções identificadas pelo anotador 1 ou as emoções identificadas pelo anotador 2, fornecendo uma avaliação mais flexível, pois o classificador pode concordar com apenas um dos anotadores; e b) Dataset Intersecção, usando a intersecção das emoções anotadas por cada anotador, no qual entram apenas as emoções que foram identificadas por ambos anotadores.

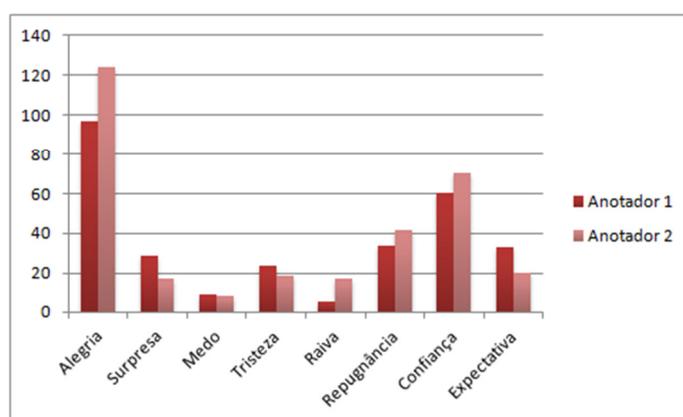


Figura 4. Distribuição das emoções anotadas por cada anotador nas revisões.

3.2 Classificação

A classificação de emoções nas revisões foi feita usando um classificador baseado em dicionário. As emoções presentes na revisão dependem das emoções associadas a cada termo contido no texto da revisão. A revisão inicialmente recebe valor 0 para cada uma das 8 categorias de emoção. Para cada termo encontrado no dicionário, seus valores de emoção são somados em cada um dos valores de emoção da revisão. Assim, ao término da busca, a revisão possui uma soma para cada uma das categorias de emoção, mostrando inclusive qual foi a emoção predominante na revisão.

No primeiro experimento, o texto das revisões em português foi traduzido automaticamente para inglês. A seguir, cada palavra da tradução foi procurada no dicionário em inglês. Ao final da classificação, os rótulos das emoções, compostos pela categoria de emoção e seu valor total, são atribuídos ao texto original em português - emoções com valor 0 indicam que não foram encontrados no texto termos que evoquem estas emoções.

Nos testes do mesmo experimento que avaliam a tradução do dicionário, a revisão permaneceu em português, e o dicionário foi traduzido do inglês para português automaticamente. Dois dicionários foram gerados: um dicionário de termos simples e um dicionário de termos compostos. Quando termos diferentes em inglês são traduzidos para um mesmo termo em português, o termo traduzido recebe os valores de emoções de todos os termos que geraram a tradução. A primeira parte da consulta ao dicionário

foi feita utilizando o dicionário de termos compostos procurando as expressões do dicionário na revisão. Assim, evitou-se a necessidade de utilizar uma função para a identificação de termos compostos na revisão. Se uma expressão do dicionário de termos compostos for encontrada no texto, somam-se os valores das categorias de emoções das expressões aos valores da revisão e então retira-se a expressão do texto da revisão para não contabilizar novamente as emoções relativas aos mesmos termos na consulta ao dicionário de termos simples. Após a busca com o dicionário de termos compostos, é feita a consulta com o dicionário de termos simples. Ao final, o texto em português obteve os rótulos de emoções para cada categoria.

No segundo experimento, que busca saber a utilidade da lematização, a revisão é traduzida automaticamente para inglês e passa por um etiquetador gramatical, onde cada termo é rotulado com sua classe gramatical (e.g. substantivos e verbos). Com a classe gramatical definida, cada termo é processado pelo lematizador para ficar na mesma forma em que se encontra no dicionário. A tradução lematizada passa então pelo classificador.

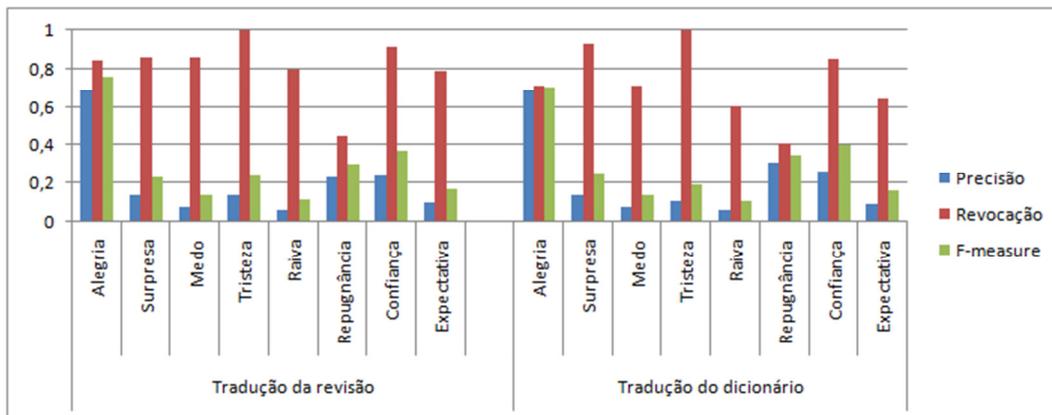
O classificador retorna os valores obtidos de cada emoção para a revisão analisada. Como mais de uma palavra pode indicar a mesma emoção, pode-se analisar a intensidade da emoção no texto, determinando a emoção predominante em cada revisão. A análise de intensidade da emoção será explorada em trabalhos futuros.

3.3 Resultados

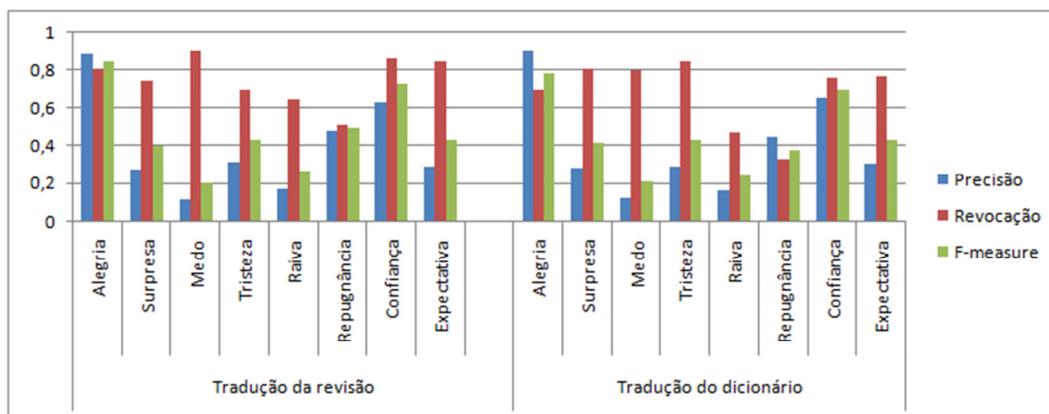
A Figura 5 mostra os resultados do primeiro experimento avaliados com a intersecção e com a união das emoções identificadas pelos anotadores. Nota-se que o classificador tem uma boa revocação em geral, indicando que classificou a emoção como presente na maioria das revisões em que ela efetivamente estava presente. Entretanto, a precisão menor mostra que também atribuiu emoções que não estavam presentes na revisão.

A comparação entre resultados considerando a intersecção e a união das anotações (Figuras 5(a) e 5(b), respectivamente) mostra que o segundo é muito superior ao primeiro. Um teste-t examinando o grau de concordância entre o classificador e os anotadores mostrou que há uma diferença estatisticamente significativa entre os dois cenários ($p = 4,5 \times 10^{-10}$). Isto era esperado, pois é mais fácil concordar com um dos anotadores do que com ambos. O ganho em precisão e medida-f veio acompanhado de uma pequena queda em revocação, uma vez que o conjunto de emoções anotadas é maior.

Ainda analisando a concordância entre o classificador e os anotadores, observou-se que os resultados são melhores com a tradução do texto da revisão do que com a tradução do dicionário. A diferença é estatisticamente significativa ($p = 0,01$). Isto pode ser explicado pelo fato de que, ao traduzir um texto, o tradutor automático produz melhores resultados do que ao traduzir palavras individuais. O texto fornece maior contexto, o que possibilita que o tradutor tenha mais informações para embasar suas escolhas nos casos de ambiguidade.



(a) Dataset Intersecção



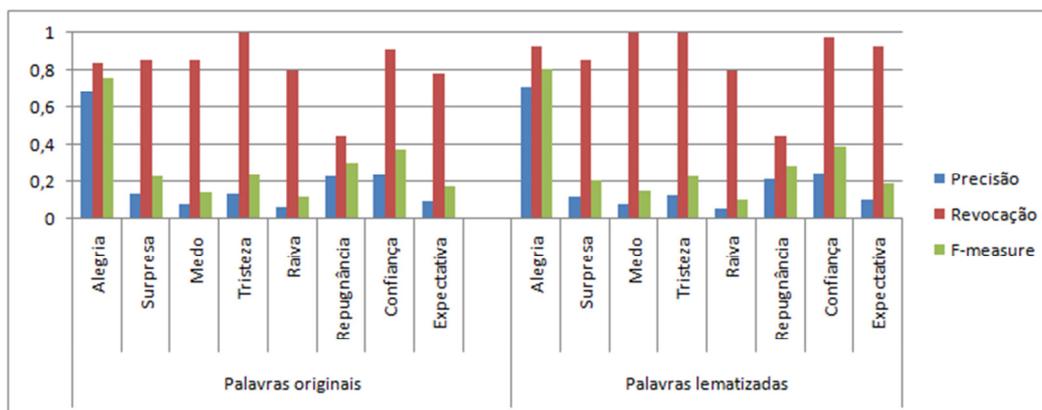
(b) Dataset União

Figura 5. Comparação entre a tradução da revisão e a tradução do dicionário com a intersecção (a) e com a união das anotações dos anotadores (b)

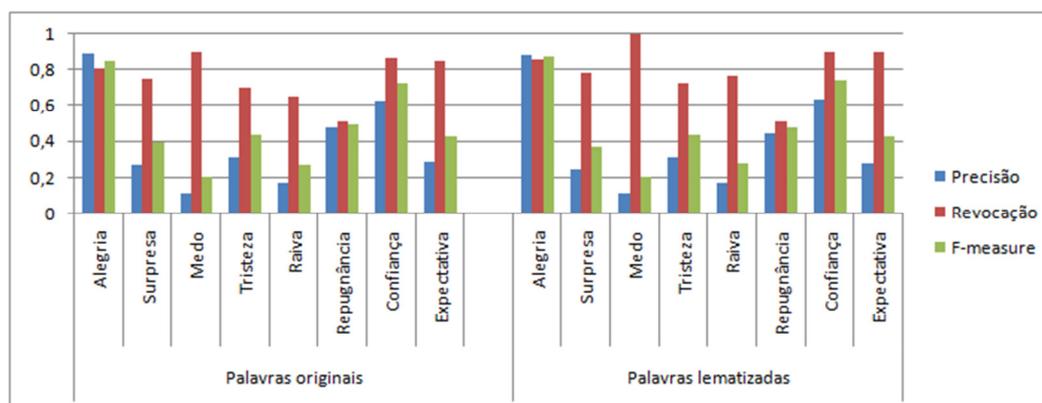
Os resultados do segundo experimento, envolvendo o uso de lemas, são mostrados na Figura 6 para os dois conjuntos de anotações (intersecção e união). O uso de lemas consegue identificar mais emoções na revisão, aumentando levemente a revocação. Com o aumento da quantidade de palavras encontradas no dicionário, mais emoções são identificadas. Se por um lado isto aumenta as emoções corretamente identificadas pelo classificador, também identifica emoções que não deveriam estar presentes segundo os anotadores, aumentando os falsos-positivos (emoções que não deveriam estar presentes). Consequentemente, reduz a precisão.

Além disso, observamos que, ao lematizar, houve uma perda ou mudança no significado da emoção. Por exemplo, o termo *annoying*, que no dicionário utilizado denota apenas raiva, foi lematizado para *annoy*, que no dicionário denota raiva e repugnância. Assim, os resultados quanto ao uso de lemas para melhorar a classificação das emoções, para este experimento, podem ser considerados positivos quanto à revocação, mas não afetam consistentemente outras medidas. Para algumas emoções houve melhora, e para outras, piora. Um teste-t mostra que não há ganhos estatisticamente significativos com a lematização ($p = 0,9$). A emoção que apresentou os melhores resultados para as medidas de avaliação foi a alegria. Isso se deve a ela ser a única emoção claramente positiva dentre as categorias disponíveis. Logo, quando uma

revisão contém sentimentos positivos, é anotada para alegria, por falta de outra emoção que possa expressar melhor o sentimento. Os termos no dicionário também seguem esse padrão: palavras como *brilliant* e *share* expressam *joy*, além de outras emoções.



(a) Dataset Intersecção



(b) Dataset União

Figura 6. Comparação entre o uso das palavras originais e dos lemas com a intersecção (a) e com a união das anotações dos anotadores (b).

O NRC contribuiu para a baixa precisão do classificador. Algumas palavras no dicionário estavam associadas a emoções, quando no texto tinham outra conotação. Por exemplo, este é o caso da palavra *author*, que o dicionário associa à emoção confiança, e que no contexto da revisão (livros), não expressa essa emoção. Outra palavra é *interesting*, presente em diversas revisões anotadas, mas que não é associada a emoções pelo dicionário. A anotação de emoções na revisão considerou a emoção expressa no texto como um todo, enquanto que o classificador identifica as emoções em cada palavra (ou expressão) em separado, o que leva a perder a informação geral, e a identificar emoções que não estariam presentes no atual contexto de determinado termo, segundo os anotadores.

Uma limitação do trabalho é a validação do tradutor automático, que não foi feita, assim como em [Banea et al. 2008]. Alguns erros de tradução fizeram com que os classificadores não identificassem corretamente algumas emoções. Outra limitação do trabalho é que foi utilizada apenas a primeira opção de tradução fornecida pelo tradutor

automático, perdendo-se possíveis traduções mais adequadas. Essa limitação afeta em especial a tradução do dicionário, visto que não há como verificar a tradução mais adequada, devido à falta de contexto. Ainda, certas palavras não constam no dicionário: *weak* (fraco) não se encontra no dicionário, sendo *weakness* (fraqueza) presente. Boa parte dos termos no NRC correspondem aos lemas das palavras, porém existem casos em que mais de uma forma da palavra está presente, como os termos *damage* e *damages*.

As revisões que obtiveram os piores resultados são aquelas nas quais os anotadores identificaram emoções no texto, porém, o classificador não conseguiu encontrar nenhuma das palavras da revisão no dicionário, não retornando emoção. Também, quanto mais palavras tem a revisão, maiores as chances de elas estarem presentes no dicionário e identificarem uma emoção que a princípio não deveriam estar presentes, gerando ruído. As revisões com melhores resultados são aquelas mais curtas, com poucas palavras, o que diminui o ruído causado por palavras que não deveriam expressar emoções mas também estão presentes no dicionário.

4. Conclusão

Este estudo de caso se propôs, através de experimentos, a verificar se a adaptação de uma abordagem de Mineração de Opinião Multilíngue para identificação de emoções era válida, e os resultados mostraram que é possível, necessitando de mais experimentos incluindo outras técnicas para melhorá-los. A tradução do texto da revisão é uma abordagem que produz resultados melhores do que a tradução do dicionário usado. O uso de lemas não produz melhorias estatísticas nos resultados.

Uma vantagem da abordagem é que ela pode ser expandida para diversos idiomas, basta que exista suporte do tradutor automático para os idiomas a serem analisados. Ainda, pode-se usar apenas os dicionários de sentimentos em inglês, sem precisar de recursos nos outros idiomas, além da anotação dos textos nos idiomas originais (para validação).

Nota-se nestes experimentos uma baixa concordância entre anotadores quanto à presença de emoções. Comparada com a anotação para polaridade, a qual dificilmente atinge níveis de concordância maiores que 75% [Becker & Tuminan 2013], a anotação de emoções é ainda mais difícil, devido à necessidade de concordar sobre presença/ausência sobre um número maior de emoções não excludentes [Wiebe et al. 2005]. Adicionalmente, as emoções são ortogonais, i.e. a presença de uma não necessariamente exclui a presença da outra. A tradicional técnica de agregar um terceiro anotador para resolver conflitos não surte muito efeito neste caso, podendo aumentar ainda mais o nível de discordância. Isto sugere a investigação de técnicas adequadas à anotação de emoções, e o uso das anotações nos processos de classificação.

Em trabalhos futuros, pretende-se expandir a quantidade de idiomas a serem analisados, além de buscar outros tipos de *corpora* para usar, como *corpora* paralelos. Espera-se também verificar se a identificação de alguma emoção pode ser melhor que outra, fazendo uso de um *corpus* balanceado, o que não foi possível nesse estudo de caso.

Agradecimento

Este trabalho foi parcialmente financiado pelo CNPq.

Referências

- Asur, S., & Huberman, B. A. Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on* (Vol. 1, pp. 492-499), 2010, August.
- Bader, B. W., Kegelmeyer, W. P., & Chew, P. A. Multilingual sentiment analysis using latent semantic indexing and machine learning. In *Data mining workshops (icdmw)*, (pp. 45–52), 2011.
- Balahur, A., & Turchi, M. Comparative Experiments for Multilingual Sentiment Analysis Using Machine Translation. In *SDAD 2012 The 1st International Workshop on Sentiment Discovery from Affective Data* (p. 75), 2012.
- Banea, C., Mihalcea, R., Wiebe, J., & Hassan, S. Multilingual subjectivity analysis using machine translation. In *EMNLP* (pp. 127–135), 2008.
- Banea, C., Mihalcea, R., & Wiebe, J. Multilingual subjectivity: are more languages better? In *ACL* (pp. 28–36), 2010.
- Becker, K., & Tumitan, D. Introdução à Mineração de Opiniões: Conceitos, Aplicações e Desafios. Simpósio Brasileiro de Banco de Dados, 2013.
- Bollen, J., Mao, H., & Zeng, X. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8, 2011.
- Ekman, P. An argument for basic emotions. *Cognition & Emotion* 6.3-4: 169-200, 1992.
- Lin, Z., Tan, S., & Cheng, X. Language-independent sentiment classification using three common words. *CIKM* (pp. 1041–1046), 2011.
- Liu, B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167, 2012.
- Mihalcea, R., Banea, C. & Wiebe, J. Learning multilingual subjective language via cross-lingual projections. *ACL 2007*.
- Mohammad, S. M. & Turney, P. D. NRC Emotion Lexicon. *NRC Technical Report*, 2013, December.
- Narr, S., Hülfenhaus, M. & Albayrak, S. Language-independent twitter sentiment analysis. *KDML*, 2012.
- Ortony, A., & Turner, T. J. What's basic about basic emotions? *Psychological review*, 97(3), 315, 1990.
- Tsytarau, M., & Palpanas, T. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3), 478–514, 2012.
- Wiebe, J., Wilson, T., & Cardie, C.. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2-3):165–210, 2005.