

Investigating Sentences Features for Subjectivity and Polarity Classification in Brazilian Portuguese

Miguel V. de Oliveira¹, Tiago E. de Melo¹

¹Grupo de Pesquisas em Sistemas Inteligentes
Laboratório de Sistemas Inteligentes
Escola Superior de Tecnologia
Universidade do Estado do Amazonas
Av. Darcy Vargas, 1200 – Manaus – Amazonas

{mvo.snfl8, tmelo}@uea.edu.br

Abstract. *Identifying subjective sentences and calculating the polarity of these sentences are two important tasks of sentiment analysis. Although they are not new problems, the most of solutions are focused on the English language. In this paper, we have proposed and evaluated an approach based on machine learning to handle both tasks in Portuguese. We investigated the use of two classifiers and also proposed a set of linguistic features extracted from the text. We experiment and evaluate the methods against a representative set of baselines and a diverse set of datasets. Our approach achieved the best results for both tasks and for all test datasets.*

Resumo. *Identificar sentenças subjetivas e calcular a polaridade destas sentenças são duas importantes tarefas de análise de sentimentos. Apesar de não serem problemas novos, a maioria das soluções são voltadas para o idioma inglês. Neste artigo, nós propusemos e avaliamos uma abordagem baseada em aprendizagem de máquina para lidar com as duas tarefas em português. Nós investigamos o uso de dois modelos de classificação e também propusemos um conjunto de características linguísticas do próprio texto. Nós experimentamos e avaliamos os métodos contra um representativo conjunto de baselines e em um conjunto diversificado de datasets. Nossa abordagem alcançou os melhores resultados nas duas tarefas e em todos os conjuntos de dados de teste.*

1. Introdução

Análise de sentimentos (AS) é o campo de estudo que investiga as opiniões, sentimentos, avaliações, atitudes e emoções das pessoas em relação a diversos tipos de entidades, tais como produtos, serviços, organizações, indivíduos, eventos e seus aspectos [Liu 2015]. AS comumente aplica análise de textos através de processamento de linguagem natural (PLN) e foca em duas principais tarefas: (1) identificação de sentenças factuais (objetivas) e opinativas (subjetivas) e (2) identificação da polaridade dos sentimentos das sentenças subjetivas como positiva, negativa ou neutra.

A tarefa de identificar a subjetividade envolve classificar um dado texto como sendo objetivo ou subjetivo. Por exemplo, a sentença “*eu fui no restaurante ontem*” é factual (objetiva). A partir dela, não é possível extrair qualquer sentimento do autor do texto. Já a sentença “*eu adorei a sobremesa*” é subjetiva. A partir desta sentença, é possível

identificar que alguém gostou de um determinado item de um restaurante. As sentenças factuais são aquelas que não possuem opinião e que, portanto, devem ser desconsideradas durante a análise de sentimentos. Além disso, há pesquisas que demonstram que a eliminação das sentenças objetivas contribui na performance de outras tarefas de análise de sentimentos, tais como extração de aspectos e cálculo de polaridade [Liu 2011].

A tarefa de identificar ou calcular a polaridade de um texto subjetivo envolve classificar um dado texto como sendo positivo, negativo ou neutro. Por exemplo, a sentença “*eu adorei a sobremesa*” mencionada anteriormente é positiva. Há também trabalhos que investigam a polaridade com mais profundidade e tentam identificar sentimentos como medo, raiva, felicidade e tristeza [Hassonah et al. 2020, Dosciatti et al. 2013]. Há ainda trabalhos que classificam a polaridade como uma escala, por exemplo, em um intervalo de 1 a -1 [Esuli and Sebastiani 2006, De Smedt and Daelemans 2012, Cambria et al. 2010].

Devido à importância das duas tarefas em análise de sentimentos, existem muitos trabalhos que tratam destas tarefas em textos escritos em inglês [Yue et al. 2019, Ravi and Ravi 2015, Schouten and Frasinca 2015]. Porém, ainda há necessidade de mais investigação de soluções para estas tarefas em outros idiomas, como, por exemplo, em português. A falta de ferramentas de processamento de textos e de dados anotados para a realização de experimentos são também desafios para a análise de sentimentos em português.

Motivados pelas observações acima, neste artigo nós apresentamos um método que explora as características linguísticas dos textos escritos em português do Brasil para lidar concomitantemente com as tarefas de identificação de sentenças subjetivas e de cálculo de polaridade. Nossa principal contribuição é um estudo empírico de um método baseado em aprendizagem de máquina e que utiliza um conjunto de sete características (*features*) linguísticas para lidar com as duas tarefas simultaneamente.

Nós avaliamos o nosso método com um conjunto representativo de *baselines* aplicados em um conjunto de diferentes tipos de *datasets*. Nossa análise de resultados sugere que é possível utilizar o método proposto e dá algumas recomendações sobre o uso do método. Os resultados experimentais também revelaram a contribuição de cada uma *features* propostas.

O artigo está organizado da seguinte maneira. Na Seção 2, nós descrevemos os principais trabalhos relacionados. Na Seção 3, nós apresentamos o método proposto neste trabalho. Os experimentos são apresentados e discutidos na Seção 4. Finalmente, na Seção 5, nós apresentamos as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

Os dois problemas que investigamos neste artigo são importantes para a análise de sentimentos e despertam bastante interesse da academia [Moraes et al. 2016, Belisário et al. 2020, Britto and Pacífico 2019, Araújo et al. 2014, Ribeiro et al. 2016].

Os autores em [Moraes et al. 2016] fizeram a comparação de dois métodos para lidar com o problema de identificação de sentenças subjetivas. Um método é baseado em léxico e o segundo método propõe uma abordagem de aprendizagem de máquina. Belisário *et al.* [Belisário et al. 2020] reproduziram os experimentos do artigo citado acima, avaliando-os em outros *datasets* de domínios diferentes e comparando-os com duas abor-

dagens propostas, baseadas em aprendizagem de máquina e grafos.

Britto e Pacífico [Britto and Pacífico 2019] investigaram o uso de diversos métodos clássicos de aprendizagem de máquina, como Naïve Bayes, Árvores de Decisão, Máquina de Vetores de Suporte e Redes Neurais. Os autores também descrevem os passos para obtenção de um *corpus* de opiniões sobre aplicações móveis e a anotação automática baseada na nota de avaliação dada pelo usuário.

Em [Araújo et al. 2014], os autores apresentam uma ferramenta Web denominada *iFeel*¹ que dá acesso a vários métodos de análise de sentimentos presentes na literatura. De acordo com a nossa análise, estes métodos representam o estado da arte nos problemas enfrentados neste trabalho e decidimos usar estes métodos como referências (*baselines*) para avaliar a nossa abordagem. A comparação destes métodos apresentada em [Ribeiro et al. 2016] também nos ajudou a escolher os *baselines* que foram considerados neste estudo.

3. Métodos e Materiais

Neste trabalho nós lidamos com duas tarefas de análise de sentimento no nível de sentença. A primeira tarefa é a identificação de sentenças objetivas e subjetivas. A segunda tarefa é a identificação da polaridade de sentenças subjetivas e que podem ser classificadas como positivas ou negativas. Nós representamos as duas tarefas como um problema de classificação binária e para as duas tarefas, investigamos o uso de métodos de aprendizagem de máquina. As características do próprio texto foram codificadas como *features* em uma representação vetorial.

3.1. Modelos de Aprendizagem de Máquina

Nós consideramos em nossa investigação um modelo linear e outro não-linear de aprendizagem de máquina. O modelo linear adotado foi Máquinas de Vetores de Suporte (MVS) (em inglês, *Support Vector Machine*). Este é um modelo bastante efetivo na tarefa de classificação de textos [Cortes and Vapnik 1995]. O objetivo do MVS é tentar encontrar um hiperplano que forneça uma separação máxima entre as classes e os exemplos que estão mais próximos deste hiperplano são conhecidos como vetores de suporte. Já o modelo não-linear adotado foi *Gradient Boosting Trees* (GBT) [Quinlan 1996]. Este é um modelo também efetivo na tarefa de classificação de textos. O modelo é representado por um conjunto de árvores de decisão que categorizam os dados através de regras de decisão. Para a implementação dos dois modelos nós utilizamos a linguagem de programação Python e a biblioteca de aprendizagem de máquina de código aberto Scikit-Learn².

3.2. Conjunto de Características (*features*)

Nós propusemos um conjunto de sete *features* para os modelos de aprendizagem de máquina. Este conjunto de *features* foi usado pelos métodos para as duas tarefas. A seguir, nós detalhamos cada uma das *features* propostas:

- **Número de palavras.** Nós consideramos o número de palavras em uma sentença como uma *feature* do classificador. De fato, nós observamos que as pessoas usam

¹<http://blackbird.dcc.ufmg.br:1210/>

²<https://scikit-learn.org>

mais palavras para descrever sentenças subjetivas do que para descrever sentenças factuais. Já na identificação da polaridade, em geral, as pessoas usam mais palavras para descrever sentenças negativas do que sentenças positivas. Na média entre todas as sentenças da nossa coleção de dados, as sentenças subjetivas são 8,5% mais longas do que as sentenças factuais e as sentenças negativas são 7,4% mais longas do que as sentenças positivas.

- **Número de adjetivos.** O número de adjetivos é uma característica útil e aplicável para os dois problemas. De fato, nós observamos que as pessoas usam mais adjetivos para descrever sentenças subjetivas e positivas. Na média, as sentenças subjetivas possuem 37,7% mais adjetivos do que as sentenças factuais e as sentenças positivas apresentam 29,77% mais adjetivos do que as sentenças negativas.
- **Número de substantivos.** Nós consideramos o número de substantivos em uma sentença como uma *feature* do classificador. Nós observamos que as pessoas usam mais substantivos para descrever sentenças factuais. Na média, as sentenças factuais possuem 12,85% mais substantivos do que as sentenças subjetivas. Não foi possível observar uma diferença significativa no número de substantivos entre as sentenças positivas e negativas.
- **Número de termos superlativos.** Nós consideramos o número de superlativos usados em uma sentença como uma *feature* do classificador. De fato, ao analisarmos o conjunto de dados empregados nos nossos experimentos, nós observamos que os superlativos são usados com muito pouca frequência nas sentenças factuais e também nas sentenças negativas.
- **Número de advérbios.** Nós consideramos o número de advérbios usados em uma sentença como uma *feature* do classificador. De fato, ao analisarmos o conjunto de dados utilizados nos nossos experimentos, nós observamos que as pessoas usam mais advérbios para descrever sentenças subjetivas. Apesar de também usarmos essa *feature* para o problema de identificação de polaridade, nós observamos que essa característica não foi tão relevante para a identificação da polaridade.
- **Número de amod.** *amod* é todo adjetivo que modifica um substantivo em uma sentença. De fato, ao analisarmos o conjunto de dados utilizados nos nossos experimentos, nós observamos que as pessoas usam com mais frequência este tipo de modificador ao descrever sentenças que são subjetivas e também negativas. Na média, as sentenças subjetivas possuem 21% mais modificadores *amod* do que as sentenças factuais e as sentenças negativas apresentam 31,95% mais modificadores *amod* do que as sentenças positivas.
- **Número de termos comparativos.** Nós consideramos o número de termos comparativos usados em uma sentença como uma *feature*. De fato, ao analisarmos o conjunto de dados utilizados nos nossos experimentos, nós observamos que as pessoas usam mais termos comparativos para descrever sentenças subjetivas. Na média, as sentenças subjetivas apresentaram 23,8% mais termos comparativos do que as sentenças factuais. Apesar de também usarmos essa *feature* para o problema de identificação de polaridade, nós observamos que essa característica não foi tão relevante para essa tarefa.

Para a extração desse conjunto de características do texto, nós usamos o modelo para Português do Brasil do pacote spaCy³. Este pacote provê informações de classes

³<https://spacy.io/models/pt>

gramaticais e árvore de dependências.

3.3. Coleção de Dados

Nós utilizamos quatro diferentes *corpora* para avaliar os métodos investigados. O *dataset* TA-Restaurantes foi construído por nós e os outros três *datasets* foram obtidos na Web. Os *datasets* variam em domínio, tamanho e balanceamento. Como reportado na literatura [Belisário et al. 2020], os métodos de classificação são afetados por essas características nos dados. Portanto, avaliar os métodos em diferentes configurações permite uma avaliação da robustez dos métodos investigados. A seguir, nós apresentamos um resumo sobre cada um dos *datasets* considerados em nossos experimentos.

TA-Restaurantes⁴. Nós fizemos uma coleta de 350 comentários de usuários sobre restaurantes no Brasil no site TripAdvisor⁵, no período de janeiro a março de 2020. Os comentários foram divididos em 1.049 sentenças e cada sentença foi manualmente anotada por nós. Inicialmente, nós classificamos as sentenças em objetivas e subjetivas. Nós pudemos observar que existe um equilíbrio no número de sentenças em cada uma dessas classes. Em seguida, nós classificamos manualmente as sentenças subjetivas em positivas e negativas. As sentenças que eram subjetivas, mas com polaridade neutra, foram desconsideradas. Na prática, nós observamos que existem poucas sentenças que são subjetivas e neutras. Também foi possível observar que há uma quantidade bastante superior de sentenças positivas do que sentenças negativas.

ReLi Corpus. É um *corpus* composto por um conjunto de 350 resenhas de livros em português publicadas na Internet e anotado manualmente [Freitas et al. 2014]. As resenhas foram igualmente divididas em sentenças objetivas e subjetivas, sendo que as sentenças subjetivas foram igualmente divididas em relação à polaridade (positivas e negativas). Portanto, este é um exemplo de conjunto de dados bastante balanceado.

Computer-BR. O *corpus* é composto de 2.281 tuítes que foram coletados do período de janeiro a fevereiro de 2015 [Moraes et al. 2016]. As mensagens são comentários da área de tecnologia sobre notebooks. As mensagens foram divididas em sentenças e anotadas tanto para identificar a subjetividade como a polaridade. Porém, em ambos os casos, as classes não estão balanceadas.

Google Play. O *corpus* original é composto de 10.000 avaliações de aplicativos móveis publicadas na loja virtual Google Play Store⁶. As sentenças foram anotadas manualmente apenas para a identificação da polaridade [dos Santos and Ladeira 2014]. Nós consideramos um subgrupo de 1630 sentenças escolhidas aleatoriamente, igualmente distribuídas entre negativas e positivas [Stiilpen Júnior 2016].

⁴O conjunto de dados está disponível em <https://data.mendeley.com/datasets/hsn6g3dbsk>

⁵<https://www.tripadvisor.com.br>

⁶<https://play.google.com/store>

A Tabela 1 apresenta as estatísticas dos *datasets* onde as sentenças foram anotadas como objetivas e subjetivas, enquanto a Tabela 2 apresenta as estatísticas dos *datasets* onde as sentenças foram anotadas como positivas e negativas.

Tabela 1. Distribuição das sentenças objetivas e subjetivas nos *datasets*.

	ReLi	TA-Restaurantes	Computer-BR
Objetivas	175 (50%)	458 (43.66%)	1677 (73.52%)
Subjetivas	175 (50%)	591 (56.34%)	604 (26.48%)
Total	350	1049	2281

Tabela 2. Distribuição das sentenças positivas e negativas nos *datasets*.

	ReLi	TA-Restaurantes	Computer-BR	Google Play
Positivas	85 (50%)	505 (90.01%)	198 (33.12%)	815 (50%)
Negativas	85 (50%)	56 (9.99%)	400 (66.88%)	815 (50%)
Total	170	561	598	1630

4. Resultados e Discussão

Nós executamos um conjunto de experimentos para avaliar o desempenho da abordagem proposta ao lidar com as tarefas de identificação de sentenças subjetivas e cálculo de polaridade em português. Nós ainda comparamos a abordagem proposta com um conjunto de métodos usados nas duas tarefas. A seguir, nós apresentamos um resumo de cada um desses métodos usados como referências na avaliação dos resultados. Além disso, também fizemos um estudo comparativo da remoção das *features* propostas, afim de investigar a influência de cada uma nos classificadores.

4.1. Métricas

Nós utilizamos as bastante conhecidas métricas de precisão (P), revocação (R) e medida-F (F_1) para avaliar os métodos nas duas tarefas investigadas neste artigo [Baeza-Yates et al. 1999]. Seja A o conjunto de respostas corretas, de acordo com um conjunto de referência, e seja B o conjunto de respostas produzidas pelo método que está sendo avaliado. Nós definimos precisão (P), revocação (R) e medida-F (F_1) como:

$$P = \frac{|A \cap B|}{|B|} \quad R = \frac{|A \cap B|}{|A|} \quad F_1 = \frac{2 \times (P \times R)}{(P + R)}$$

4.2. Métodos de referência (*baselines*)

EmoticonsDS. Este método utiliza uma lista pré-definida de *emoticons* para calcular a polaridade das sentenças [Gonçalves et al. 2013]. Portanto, o método é focado em textos que sejam mais informais, como são as mensagens postadas no Twitter. Apesar do método ter, originalmente, a finalidade de calcular a polaridade de sentenças, nós também investigamos a aplicação deste método na identificação de sentenças subjetivas.

Stanford Recursive Deep Model. O método consiste de um modelo de redes neurais chamado *Recursive Neural Tensor Network* (RNTN) que processa as sentenças levando em consideração a ordem das palavras [Socher et al. 2013]. São empregadas cinco classes para a classificação de polaridade (muito negativo, negativo, neutro, positivo, muito positivo). Apesar do método não ter sido criado com a finalidade de identificar sentenças subjetivas, nós também investigamos a aplicação deste método em ambas as tarefas. Nós consideramos as classes “muito negativo” e “negativo” como uma única orientação negativa, da mesma forma, as duas classes “positivo” e “muito positivo” foram agrupadas como uma única orientação positiva. Para a tarefa de identificação de subjetividade, consideramos a classe neutra como ausência de opinião (sentenças factuais).

Sentiment140. O método emprega um comitê de três classificadores (Máquina de Vetores de Suporte, Máxima Entropia e *Naïve Bayes*) treinados com dados do Twitter. O método foi criado para classificar polaridade de tuítes e utiliza *features* simples como unigramas, bigramas e *tags* de classe gramatical [Go et al. 2009]. Apesar do método não ter sido criado com a finalidade de identificar sentenças subjetivas, nós também investigamos a aplicação deste método em ambas as tarefas.

NRC Hashtag. O método usa um léxico de *hashtags* associadas às emoções consideradas básicas: alegria, tristeza, raiva, medo, confiança, nojo, surpresa, e antecipação. A pontuação é calculada verificando a frequência dos termos analisados dentro do dicionário [Mohammad and Turney 2010]. Apesar de ter sido desenvolvido para identificar emoções em tuítes, o método pode ser usado para identificação de polaridade, quando agrupadas as emoções em positivas e negativas.

SentiStrenght. O método descrito em [Thelwall 2017] possui um dicionário de 2,310 palavras associadas a sentimentos positivos ou negativos por meio de uma pontuação que varia no intervalo $[1, 5]$ para os termos positivos e $[-5, -1]$ para os termos negativos. Os termos do texto são comparados um a um com o léxico, e a resposta do método é sempre a maior pontuação positiva e a menor pontuação negativa que foram associadas ao texto após a comparação. O método pode ser usado tanto para medir a polaridade do sentimento, quanto a força de subjetividade, considerando que valores 0 são considerados ausência de sentimento.

SentiWordNet. O método descrito em [Esuli and Sebastiani 2006] apresenta pontuações para o quão objetivo, negativo e positivo é um termo com base na sua relação com um conjunto de sinônimos (*synset*). As pontuações estão no intervalo real $[0.0, 1.0]$ para cada classe, e a soma das três classes será sempre 1.0. O método foi criado para realizar as tarefas de identificação de subjetividade e cálculo de polaridade.

4.3. Resultados

Nesta seção, nós apresentamos e discutimos os resultados alcançados pelos métodos propostos e *baselines* nas tarefas de identificação de sentença subjetiva e cálculo de polaridade. O classificador baseado no modelo de Máquinas de Vetores de Suporte será no-

meado como *MVS* e o classificador baseado no modelo *Gradient Boosting Trees* será nomeado como *GBT*.

A Tabela 3 apresenta a performance dos métodos em termos de precisão (P), revocação (R) e medida-F (F_1) na tarefa de identificar sentenças subjetivas. Os valores mais altos para cada métrica e em cada *dataset* estão marcados em negrito. Neste experimento, nós executamos uma validação cruzada de 10 *folds* sobre o conjunto de dados.

Tabela 3. Resultados experimentais para a tarefa de identificação de sentenças subjetivas.

	Reli			TA-Restaurantes			Computer-BR		
	P	R	F_1	P	R	F_1	P	R	F_1
<i>EmoticonsDS</i>	0.49	0.97	0.65	0.58	0.97	0.72	0.26	0.99	0.42
<i>Stanford</i>	0.49	0.94	0.65	0.57	0.87	0.69	0.27	0.94	0.42
<i>Sentiment140</i>	0.47	0.85	0.61	0.58	0.83	0.68	0.27	0.86	0.41
<i>NRC Hashtag</i>	0.48	0.77	0.59	0.56	0.60	0.58	0.27	0.79	0.40
<i>SentisStrenght</i>	0.53	0.29	0.38	0.64	0.26	0.37	0.23	0.30	0.26
<i>SentiWordNet</i>	0.51	0.46	0.49	0.63	0.25	0.36	0.28	0.56	0.37
<i>MVS</i>	0.73	0.65	0.68	0.74	0.78	0.76	0.42	0.68	0.51
<i>GBT</i>	0.76	0.68	0.71	0.71	0.91	0.80	0.39	0.34	0.36

De acordo com os resultados apresentados na Tabela 3, os métodos *MVS* e *GBT* alcançaram os melhores resultados gerais. Apenas o método *EmoticonsDS* conseguiu obter os melhores resultados na revocação nos três *datasets*. Porém, este método obteve um valor insatisfatório na precisão, que resultou em um valor baixo na métrica de F_1 . Ao compararmos a média de F_1 entre os métodos *MVS* e *GBT* nos três *datasets*, nós observamos que o método *MVS* alcançou um resultado superior. Porém, nós calculamos o desvio padrão (σ) de F_1 dos dois métodos e o método *GBT* apresentou um resultado (0.056) menor do que o *MVS* (0.074).

A Tabela 4 apresenta a performance dos métodos em termos de precisão (P), revocação (R) e medida-F (F_1) na tarefa de calcular a polaridade de sentenças subjetivas. Os valores mais altos para cada métrica e em cada *dataset* estão marcados em negrito. Neste experimento, nós executamos uma validação cruzada de 10 *folds* sobre o conjunto de dados.

Tabela 4. Resultados experimentais para a tarefa de cálculo de polaridade.

	Reli			TA-Restaurantes			Computer-BR			Google Play		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R_1	F
<i>EmoticonsDS</i>	0.75	0.50	0.34	0.90	0.87	0.83	0.10	0.32	0.16	0.72	0.49	0.35
<i>Stanford</i>	0.49	0.46	0.33	0.85	0.11	0.08	0.44	0.62	0.52	0.64	0.39	0.35
<i>Sentiment140</i>	0.55	0.47	0.50	0.84	0.68	0.75	0.63	0.50	0.54	0.73	0.60	0.62
<i>NRC Hashtag</i>	0.51	0.40	0.45	0.78	0.31	0.43	0.59	0.47	0.52	0.45	0.31	0.35
<i>SentisStrenght</i>	0.69	0.21	0.32	0.91	0.23	0.37	0.69	0.19	0.28	0.76	0.24	0.36
<i>SentiWordNet</i>	0.56	0.27	0.36	0.86	0.15	0.24	0.58	0.27	0.35	0.51	0.17	0.25
<i>MVS</i>	0.48	0.52	0.49	0.92	0.94	0.92	0.38	0.74	0.50	0.52	0.96	0.67
<i>GBT</i>	0.57	0.64	0.59	0.90	0.99	0.95	0.44	0.44	0.43	0.69	0.68	0.69

De acordo com os resultados apresentados na Tabela 4, os métodos MVS e GBT novamente alcançaram os melhores resultados gerais. Interessantemente, o método *Sentiment140* alcançou o melhor resultado de F_1 no *dataset* Computer-BR. Nós entendemos que isto aconteceu porque o método foi desenvolvido especificamente para a tarefa de cálculo de polaridade no Twitter. Neste caso, o método foi aplicado em um *dataset* de tuítes para a tarefa de cálculo de polaridade. Porém, o mesmo método não conseguiu obter bons resultados nos outros *datasets*, mesmo na tarefa de cálculo de polaridade. Finalmente, ao compararmos a média de F_1 entre os métodos MVS e GBT nos quatro *datasets*, nós observamos que o método GBT alcançou um resultado levemente superior ao método MVS. Nós calculamos o desvio padrão dos dois métodos e ambos ficaram resultados bastante próximos de $\sigma = 0.03$.

4.4. Impacto das *features* nos classificadores

Durante a nossa pesquisa, nós também avaliamos o impacto de remover cada uma das *features* propostas no desempenho dos classificadores. Nós queremos fazer isto para entender a importância de cada *feature* usada em cada uma das tarefas entre os vários *datasets*.

A Figura 1 apresenta o resultado da remoção de cada uma das *features* na tarefa de identificação de sentenças subjetivas em cada um dos *datasets*. O resultado apresentado foi alcançado através do método *MVS*. Nós experimentamos também com o classificador *GBT* e o resultado foi bastante próximo. De acordo com o gráfico, a quantidade de advérbios (*adv*) foi a *feature* mais relevante para o *dataset* ReLi, a quantidade de adjetivos foi a *feature* mais relevante para o *dataset* TA-Restaurantes e a quantidade de substantivos foi a *feature* mais relevante para o *dataset* Computer-BR. A distribuição de importância entre as várias *features* assinala que não existe uma característica que irá sempre se sobrepôr às demais e indica também que a combinação do conjunto proposto de *features* funciona melhor que uma específica, pois cada *feature* pode capturar aspectos do texto que as demais não conseguem.



Figura 1. Avaliação do impacto das *features* na tarefa de identificação de sentenças subjetivas.

A Figura 2 apresenta o resultado da remoção de cada uma das *features* na tarefa de cálculo de polaridade de sentenças subjetivas em cada um dos *datasets*. O resultado

apresentado foi alcançado através do método *MVS*. Nós experimentamos também com o classificador *GBT* e o resultado ficou bastante próximo. De acordo com o gráfico, a quantidade de adjetivos (*adj*) foi a *feature* mais relevante para os *datasets* Computer-BR e Google Play. Ao compararmos com as demais *features*, observamos que não houve uma grande diferença entre elas no *dataset* Google Play. Já a quantidade de *amod* foi a *feature* mais relevante para o *dataset* ReLi e a quantidade de termos comparativos (*comp*) e a quantidade de termos superlativos (*super*) foram as *features* mais relevantes para o *dataset* TA-Restaurantes. Assim como ocorreu anteriormente, a distribuição de importância entre as várias *features* assinala que não existe uma característica que irá sempre se sobrepor às demais e indica que a combinação do conjunto proposto de *features* funciona melhor que uma específica.

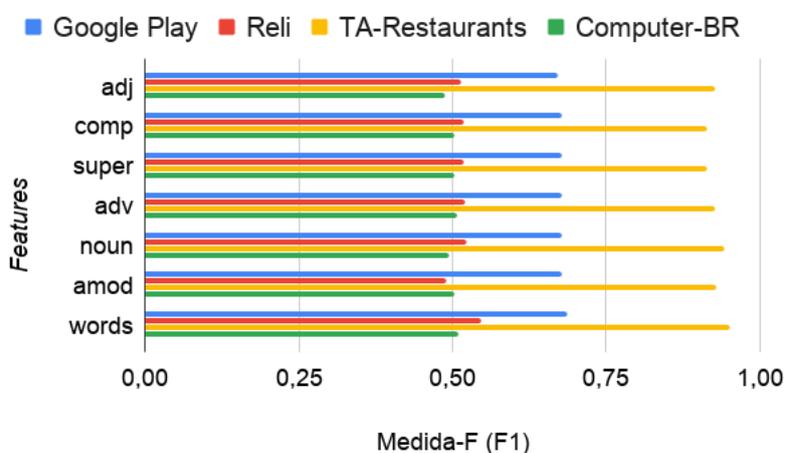


Figura 2. Avaliação do impacto das *features* na tarefa de cálculo de polaridade.

5. Conclusões e Trabalhos Futuros

Neste artigo, nós propusemos uma abordagem para identificar sentenças subjetivas e calcular a polaridade destas sentenças subjetivas. A abordagem proposta é baseada em aprendizagem de máquina e em sete características linguísticas dos textos escritos em português. Nós avaliamos a performance desta abordagem usando dois diferentes classificadores contra um conjunto representativo de *baselines* em diferentes tipos de *datasets*. De acordo com os resultados experimentais reportados, os classificadores Máquina de Vetores de Suporte (MVS) e *Gradient Boosting Trees* (GBT) alcançaram os melhores resultados gerais nas duas tarefas. Mais especificamente, o classificador MVS apresentou melhor resultado na tarefa de identificação de sentenças subjetivas e o classificador GBT apresentou o melhor resultado na tarefa de cálculo de polaridade. Isto indica que escolher o correto conjunto de *features* para a representação das sentenças é o fator mais importante em ambas as tarefas.

Para trabalhos futuros, nós planejamos explorar novas características linguísticas para o nosso método. O conjunto de características proposto neste artigo não é necessariamente ótimo e nós esperamos melhorar os resultados explorando um novo conjunto, mais abrangente. Por exemplo, nós pretendemos investigar o uso de características específicas empregadas em textos curtos e de mídias sociais como o Twitter.

Agradecimentos

Os autores agradecem o apoio financeiro e material provido pela Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) por meio do Projeto PPP 04/2017 e pelo apoio do Laboratório de Sistemas Inteligentes (LSI) da Universidade do Estado do Amazonas (UEA).

Referências

- Araújo, M., Gonçalves, P., Cha, M., and Benevenuto, F. (2014). ifeel: a system that compares and combines sentiment analysis methods. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 75–78.
- Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463. ACM press New York.
- Belisário, L. B., Ferreira, L. G., and Pardo, T. A. S. (2020). Evaluating methods of different paradigms for subjectivity classification in portuguese. In *International Conference on Computational Processing of the Portuguese Language*, pages 261–269. Springer.
- Britto, L. and Pacífico, L. (2019). Análise de sentimentos para revisões de aplicativos mobile em português brasileiro. In *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*, pages 1080–1090, Porto Alegre, RS, Brasil. SBC.
- Cambria, E., Speer, R., Havasi, C., and Hussain, A. (2010). Senticnet: A publicly available semantic resource for opinion mining. In *2010 AAAI fall symposium series*.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.
- De Smedt, T. and Daelemans, W. (2012). Pattern for python. *The Journal of Machine Learning Research*, 13(1):2063–2067.
- dos Santos, F. L. and Ladeira, M. (2014). The role of text pre-processing in opinion mining on a social media language dataset. In *2014 Brazilian Conference on Intelligent Systems*, pages 50–54. IEEE.
- Dosciatti, M. M., Ferreira, L., and Paraiso, E. (2013). Identificando emoções em textos em português do brasil usando máquina de vetores de suporte em solução multiclasse. *ENIAC-Encontro Nacional de Inteligência Artificial e Computacional. Fortaleza, Brasil*.
- Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *LREC*, volume 6, pages 417–422. Citeseer.
- Freitas, C., Motta, E., Milidiú, R. L., and César, J. (2014). Sparkling vampire... lol! annotating opinions in a book review corpus. *New Language Technologies and Linguistic Research: A Two-Way Road*, pages 128–146.
- Go, A., Bhayani, R., and Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, 1(12):2009.
- Gonçalves, P., Araújo, M., Benevenuto, F., and Cha, M. (2013). Comparing and combining sentiment analysis methods. In *Proceedings of the first ACM conference on Online social networks*, pages 27–38.

- Hassonah, M. A., Al-Sayyed, R., Rodan, A., Ala'M, A.-Z., Aljarah, I., and Faris, H. (2020). An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on twitter. *Knowledge-Based Systems*, 192:105353.
- Liu, B. (2011). Opinion mining and sentiment analysis. In *Web Data Mining*, pages 459–526. Springer.
- Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.
- Mohammad, S. and Turney, P. (2010). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, pages 26–34.
- Moraes, S. M., Santos, A. L., Redecker, M., Machado, R. M., and Meneguzzi, F. R. (2016). Comparing approaches to subjectivity classification: A study on portuguese tweets. In *International Conference on Computational Processing of the Portuguese Language*, pages 86–94. Springer.
- Quinlan, J. R. (1996). Boosting first-order learning. In *International Workshop on Algorithmic Learning Theory*, pages 143–155. Springer.
- Ravi, K. and Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89:14–46.
- Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., and Benevenuto, F. (2016). Sentibench—a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5(1):1–29.
- Schouten, K. and Frasincar, F. (2015). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3):813–830.
- Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., and Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642.
- Stiilpen Júnior, M. (2016). Um arcabouço de processamento de textos informais em português brasileiro para aplicações de mineração de dados. Master's thesis, Universidade Federal de Ouro Preto.
- Thelwall, M. (2017). The heart and soul of the web? sentiment strength detection in the social web with sentistrength. In *Cyberemotions*, pages 119–134. Springer.
- Yue, L., Chen, W., Li, X., Zuo, W., and Yin, M. (2019). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, pages 1–47.