

Anotando um Corpus de Notícias para a Análise de Sentimento: um Relato de Experiência

Mariza Miola Dosciatti, Lohann Paterno Coutinho Ferreira, Emerson Cabrera Paraiso

Pontifícia Universidade Católica do Paraná, Rua Imaculada Conceição, 1155 - Curitiba, PR, Brasil

{mariza, paraiso}@ppgia.pucpr.br, lohann.ferreira@pucpr.br

***Resumo.** Este artigo relata o processo de construção e anotação de um corpus de notícias para a Análise de Sentimento. Os textos, extraídos de jornais do Brasil, foram anotados com as emoções básicas (alegria, tristeza, raiva, surpresa, repugnância e medo) ou a ausência de emoção (neutro). O processo de anotação resultou em valor de concordância baixo ($\kappa = 0,38$). Apresentamos o processo de anotação e os resultados de alguns experimentos realizados durante e após a anotação, com o objetivo de entender os motivos da baixa concordância. O corpus anotado foi submetido a um método de identificação de emoções, sendo os resultados obtidos também apresentados.*

1. Introdução

O estudo da identificação de emoções em texto é uma área de pesquisa que ganhou impulso recente com pesquisadores buscando a avaliação automática de opiniões deixadas em sites Web ou nas Redes Sociais. A área de pesquisa que envolve o estudo e a identificação de emoções em informação textual é conhecida como Análise de Sentimento (AS). A AS requer o desenvolvimento de métodos e recursos que, integrados, possibilitam aos sistemas computacionais serem capazes de manipular significado afetivo no discurso. Estes recursos ainda são escassos para o Português do Brasil.

Neste artigo apresentamos o relato do processo de anotação de um corpus de textos em Português do Brasil. O corpus é composto por textos curtos (título e linha fina de notícias) extraídos de jornais online (exemplos na Tabela 1). Optamos por utilizar esse tipo de textos pois em geral, os trabalhos na área de AS utilizam pequenos textos, como tweets, posts, avaliações de produtos, etc. A necessidade da construção do corpus de notícias surgiu durante o desenvolvimento de um método para a identificação das seis emoções básicas de Ekman [Ekman 1992]: *alegria, tristeza, raiva, medo, repugnância e surpresa*.

O processo de construção e principalmente de anotação de um corpus costuma ser uma atividade bastante complexa e lenta, tendo em vista a subjetividade da rotulagem e a dificuldade em atingir um grau de concordância adequado entre os anotadores. Neste trabalho observamos um baixo grau de concordância entre os anotadores do corpus de notícias ($\kappa = 0,38$). Ao longo deste texto apresentamos os motivos que contribuíram para a obtenção deste resultado.

O artigo está organizado nas seguintes seções: a seção 2 apresenta alguns trabalhos que relatam corpora para a AS. A Seção 3 descreve a metodologia aplicada na anotação dos textos. Na Seção 4 é realizada uma análise dos resultados da concordância obtida durante o processo de anotação dos textos. A Seção 5 apresenta as conclusões e os trabalhos futuros.

2. Corpora em Análise de Sentimento

A literatura não é farta quanto a artigos que apresentem os resultados obtidos quando da construção de um corpus para a AS. Buscamos na literatura trabalhos que apresentassem os seguintes elementos: a existência de corpora para AS na língua Portuguesa do Brasil; a metodologia de construção e anotação de corpora utilizados em métodos de AS; o grau de concordância entre os anotadores nos corpora utilizados nestes trabalhos.

Alguns trabalhos envolvendo a construção de corpus em Português para a AS foram identificados na literatura. O trabalho de [Freitas et al. 2014] se refere a construção de um corpus composto por resenhas de livros publicadas na internet e anotado manualmente em relação à polaridade. O Reli (Resenha de livros), como é chamado o corpus, tem como objetivo identificar opiniões sobre entidades nos textos. As 1.600 resenhas totalizando 12.000 sentenças foram anotadas, considerando os aspectos linguísticos como as categorias morfossintáticas e a informação semântica. Os textos foram anotados por três anotadores e o estudo do acordo entre os anotadores foi realizado com 400 sentenças. Usando a métrica *Agr* (a mesma que foi usada em [Wiebe et al.]) o acordo de atribuição de polaridade alcançou quase a totalidade (100%) e a concordância na identificação de expressões que continham opiniões ficou em 80%.

No trabalho de [Nascimento et al. 2012] foi construído e anotado um corpus de textos de tweets que se referem a comentários de notícias. As notícias (três no total) foram selecionadas por terem ampla repercussão na imprensa na época da coleta dos textos. Os documentos de tweets foram anotados manualmente por três pesquisadores envolvidos no trabalho, que poderiam atribuir a cada texto apenas uma categoria: positivo ou negativo. Ao final do processo foi criado um corpus composto de 850 documentos, divididos em 50% positivos e 50% negativos.

No trabalho de [Alves et al. 2014], foi construído um corpus com 17.000 tweets que foram colhidos durante a Copa das Confederações, em 2013. Dos 17.000 textos, 1.500 foram anotados por dez voluntários, que puderam atribuir a cada texto uma entre as três categorias possíveis, positivo, negativo ou neutro. A categoria final atribuída ao texto foi escolhida através de voto majoritário.

Alguns pesquisadores publicaram artigos apresentando os resultados do processo de construção e anotação dos corpora em outras línguas e mostraram o grau de concordância obtido entre os avaliadores por meio do índice kappa [Cohen 1960], que é um coeficiente que leva em conta a proporção de concordância que ocorre devido ao acaso. O kappa tem como valor máximo 1, que representa alta concordância entre os avaliadores e 0, que indica que não houve concordância.

Os textos de um corpus composto de 5.205 posts de blogs, escritos em Inglês, usado em [Aman e Szpakowicz 2007] e em [Ghazi et al. 2014], foram anotados em um nível de sentença por quatro anotadores. Cada anotador atribuiu, a cada texto, uma das

seis emoções básicas ou uma categoria chamada emoções mistas. Também classificaram o texto como emocional ou não emocional e avaliaram a intensidade das emoções atribuindo uma das categorias alta, média ou baixa. O valor kappa obtido entre os anotadores foi de 0,76 para textos emocionais e não emocionais, 0,65 (valor médio) para as categorias e de 0,52 (valor médio) para as intensidades. Em [Strapparava e Mihalcea 2008] um corpus composto por 1.250 textos de notícias, escritos em Inglês, foi anotado em um nível de documento por cinco anotadores sendo que em cada texto o anotador escolheu uma entre as seis emoções básicas de [Ekman 1992]. O valor kappa obtido entre os anotadores foi de 0,53. No trabalho de [Habernal et al. 2014] um corpus com 10.000 comentários extraídos do Facebook, escritos em idioma Tcheco, foi anotado em um nível de documento por dois avaliadores atribuindo uma entre três categorias possíveis: positivo, negativo ou neutro. O kappa obtido neste corpus foi de 0,66. Em [Alm et al. 2005] um corpus com 1.580 textos extraídos de 185 histórias infantis, escritas em Inglês, foi anotado em um nível de sentença por dois anotadores. Cada texto foi rotulado com: raiva, repugnância, medo, alegria, tristeza, surpresa positiva ou surpresa negativa. O grau de concordância kappa entre os anotadores desse corpus ficou entre 0,24 e 0,51.

Analisando o grau de concordância obtido nos trabalhos, percebe-se que corpora com seis ou mais classes tiveram um baixo grau de concordância. Para o Português do Brasil não foi encontrado nenhum corpus, que tenha sido anotado com as seis emoções básicas e estivesse disponível para ser utilizado em pesquisas de AS.

3. Construindo um Corpus de Notícias para a Análise de Sentimentos

A maioria dos textos usados para validar os métodos de AS costumam ser informais, com autores expressando livremente suas emoções. Os textos extraídos de notícias, por sua vez, possuem algumas características que os diferem dos textos comumente usados: são escritos usando uma estrutura formal, e as emoções não são explicitamente encontradas e, quando o são, normalmente apresentam-se contraditórias, como as identificadas no exemplo “*Mãe e bebê caem em rio do Recife e dupla consegue resgatar criança*” (alegria e tristeza).

Alguns pesquisadores da área de AS se interessaram em trabalhar com corpora de notícias. Gomes e colegas [Gomes et al. 2013] utilizaram um corpus de notícias para monitorar o estado da economia. Em [Balahur e Steinberger 2009], os autores destacam a importância de se aplicar a AS em notícias. Também destacam os três diferentes tipos de pontos de vista que devem ser levados em conta no momento da anotação quando se trata de textos de notícias: o ponto de vista do autor, do leitor e do texto. Do ponto de vista do autor e do leitor, os fatos transmitidos são interpretáveis pela emoção que emitem, porém algumas dessas emoções não são universais em seu significado e são determinadas por influências sociais e culturais. Os autores citam o exemplo do texto “*The results of the match between Juventus Torino and Real Madrid last night are 3-0*” (“Os resultados do jogo entre Juventus e Real Madrid ontem à noite foram 3-0”) que poderia ser interpretado como algo alegre para um jornal italiano ou uma notícia triste para um jornal espanhol.

Os textos de notícias que compõem o corpus apresentado neste trabalho de pesquisa têm, em média, 23 palavras em cada e foram extraídos automaticamente do site

www.globo.com por meio de uma ferramenta chamada FeedReader¹. Os textos do corpus de notícias pertencem a diferentes categorias, tais como: mundial, política, polícia e economia. O corpus contém 2.000 textos anotados e distribuídos da seguinte forma: 184 (9%) rotulados como *alegria*, 262 (13%) como *repugnância*, 222 (11%) como *medo*, 83 (4%) como *raiva*, 252 (13%) como *surpresa*, 455 (23%) como *tristeza* e 542 (27%) de textos neutros. Os textos foram avaliados considerando especificamente o ponto de vista do autor e foram anotados em nível de documento.

Uma primeira conclusão importante a se destacar é o alto grau de desbalanceamento entre as emoções (classes). Há uma grande dificuldade em encontrar textos jornalísticos com a emoção predominante *raiva*.

O processo de anotação foi realizado por cinco anotadores voluntários. Estabeleceu-se como regra que todos os textos do corpus fossem anotados por dois anotadores diferentes e, em caso de discordância, o texto deveria passar pela análise de um terceiro anotador. Uma das principais dificuldades do processo de anotação como um todo foi encontrar voluntários aptos e que executassem a atividade com comprometimento. O perfil esperado dos voluntários era de profissionais com experiência em linguística ou linguística computacional e que não estivessem envolvidos no projeto do método de AS. Na primeira etapa do processo de anotação, que consistia em anotar 2.000 textos, participaram cinco anotadores voluntários, todos profissionais com experiência mínima de 15 anos em linguística (professores no ensino superior). Cada anotador teve dois meses para que essa etapa fosse concluída. Ao final deste prazo apenas 1.540 textos foram anotados duas vezes e 460 textos tiveram que ser submetidos à análise de um sexto anotador.

A atividade de anotação consistiu em ler o texto e identificar a emoção (ou ausência dela) presente no mesmo. Os rótulos possíveis eram: *alegria*, *tristeza*, *raiva*, *medo*, *repugnância* e *surpresa*, e *neutro*. A cada texto foi atribuído um único rótulo, ou seja, aquele que representa a emoção predominante do texto. O anotador também atribuiu um rótulo de intensidade (ou neutralidade) da emoção no texto. Essa intensidade pôde ser escolhida entre *alta*, *média* ou *baixa*. O anotador tinha a possibilidade de escolher uma emoção secundária, para indicar um segundo rótulo e uma segunda intensidade. Apesar de não ser obrigatória, essa opção foi dada a fim de facilitar o processo de anotação de textos que possuem duas emoções na mesma proporção.

Apesar da atividade de anotação ser uma tarefa completamente subjetiva, é preciso encontrar uma forma de padronizá-la. Assim, um manual do anotador foi escrito contendo informações a respeito dos textos, como tipo, categorias, como usar o sistema web de anotação, além de uma lista de 40 textos já anotados pela equipe do projeto. Esses textos-modelo foram escolhidos por serem textos difíceis de serem analisados. A maioria deles continha várias emoções por texto ou emoções contraditórias. Assim, os anotadores foram incentivados a: primeiramente identificar a emoção predominante em cada sentença do texto; em seguida identificar a emoção que obteve o maior número de ocorrências no texto como um todo e, por fim, determinar essa emoção como sendo a predominante do texto. Em muitos textos, porém, ocorreu um empate no número de

¹ <http://feedreader.com/>

emoções encontradas nas sentenças e nessas situações, optou-se pela escolha intuitiva da emoção analisando o documento em sua totalidade. Na Tabela 1 podem ser visualizados dois dos 40 textos que foram fornecidos previamente aos anotadores a título de ilustração do processo. Para gerenciar o processo de anotação, um sistema Web foi implementado.

Tabela 1. Exemplos de anotação de textos

Texto	Emoções em cada sentença	Emoção predominante	Intensidade
Mãe e bebê caem em rio do Recife e dupla consegue resgatar criança. Mulher ainda está desaparecida e bombeiros trabalham nas buscas. No momento do acidente, chovia muito e nível do Rio Tejiptió havia subido.	<i>Sentença 1:</i> tristeza e alegria <i>Sentença 2:</i> tristeza <i>Sentença 3:</i> repugnância	tristeza	alta
Estudante queimada em sessão de bronzamento recebe alta, em Goiás. Mãe comemora recuperação: 'Ela está bem emocionalmente, animada'. Treze mulheres se queimaram ao passar óleo de coco com canela, em Jataí.	<i>Sentença 1:</i> tristeza e alegria <i>Sentença 2:</i> alegria <i>Sentença 3:</i> tristeza	alegria	baixa

Na segunda etapa do processo de anotação foi necessário que um novo anotador decidisse o rótulo dos textos que não tiveram concordância na primeira etapa do processo. Nos casos em que os textos haviam recebido um segundo rótulo e/ou grau de intensidade em uma ou em ambas as anotações, o anotador analisava essas informações antes de escolher o rótulo final. Nos textos que não continham essa informação, o anotador era obrigado a escolher intuitivamente um entre os dois rótulos possíveis.

4. Avaliação da Concordância entre os Anotadores

Segundo [Klebanov e Beigman 2009], para a tarefa de classificação de textos, a prática corrente é usar o valor de um coeficiente de concordância inter-anotador, como o kappa, para verificar se o conjunto dados é adequado para treinar e testar um classificador. Um valor de concordância alto entre os anotadores indica que o conjunto, como um todo, é bom para treinar e testar algoritmos de classificação. Caso o valor de concordância seja baixo, o conjunto de dados é considerado pouco confiável.

O percentual de casos em que dois anotadores concordam em relação à classificação de um conjunto de textos com um dado número de categorias é a forma mais simples de se atribuir confiabilidade a um processo de anotação de textos realizado em um determinado corpus. Porém, este método não considera a proporção dessa concordância que é devido ao acaso. O coeficiente kappa leva em conta no cálculo a proporção de concordância que ocorre devido ao acaso e por isso é bastante utilizado para medir a concordância entre anotadores em corpora usados em sistemas de AS.

Em linguística computacional, o limite de aceitabilidade do grau de concordância de um corpus anotado pode variar de pesquisador para pesquisador. [Krippendorff 1980] defende que só pode ser considerado aceitável um corpus anotado com um valor kappa superior a 0,67. Em [Artstein e Poesio 2005] verificou-se que apenas valores acima de 0,8 sugerem uma anotação de qualidade. Di Eugenio e Vidro [Di Eugenio e Vidro 2004] sugerem que os pesquisadores devem apresentar detalhadamente a metodologia que foi seguida na coleta e anotação dos textos, como por exemplo, número de anotadores, se os textos foram anotados independentemente, se a anotação se baseou em um manual de anotação, dentre outros detalhes.

Neste trabalho, o coeficiente kappa foi usado para avaliar o grau de acordo entre os anotadores. Para tal, vários experimentos foram realizados durante o processo de anotação. No primeiro experimento, o objetivo foi verificar o grau de concordância geral entre as duas anotações realizadas em cada um dos 2.000 textos, além do grau de concordância entre as duas anotações em relação a cada categoria (emoção). A Tabela 2 apresenta a matriz de confusão da concordância para os 2.000 textos anotados. A Tabela 3 apresenta os valores de concordância obtidos por emoção.

Tabela 2. Matriz de confusão: concordância entre anotadores para 2.000 textos

		Anotação 1						
		Neutro	Repugnância	Alegria	Medo	Raiva	Surpresa	Tristeza
Anotação 2	Neutro	294	34	51	12	8	64	43
	Repugnância	32	66	4	21	13	37	55
	Alegria	34	1	97	3	0	43	7
	Medo	4	28	1	73	4	27	78
	Raiva	2	13	1	8	15	5	24
	Surpresa	55	18	31	10	8	150	35
	Tristeza	23	38	5	50	29	47	299

Na Tabela 2, os valores destacados na diagonal representam o número de textos que tiveram concordância. O valor kappa obtido para o acordo geral das categorias entre os seis anotadores foi 0,38, um valor baixo considerando as metas de anotação comumente usadas em linguística computacional [Artstein e Poesio 2008].

Analisando os valores de kappa apresentados na Tabela 3 e o número de textos que tiveram concordância/discordância em cada categoria apresentada na Tabela 2, pode-se verificar que as maiores discordâncias ocorreram entre as categorias *medo*, *repugnância*, *tristeza* e *raiva* e entre as categorias *neutro*, *alegria* e *surpresa*. Isso faz bastante sentido visto que um texto cuja emoção predominante é *tristeza*, por exemplo, pode conter palavras que remetem o anotador a interpretar a emoção do autor como *raiva*, *medo* ou *repugnância*. O texto "*Francesa admite que matou afogados dois bebês encontrados congelados. A mulher, que mantinha o corpo de dois bebês congelados em sua casa no centro da França, declarou à polícia ter matado os dois recém-nascidos afogados*" é um exemplo disso, pois na primeira anotação este texto foi anotado com a emoção *raiva* e na segunda anotação com *tristeza*. Essa situação também ocorre frequentemente ao analisar textos das categorias *neutro*, *alegria* e *surpresa*. No texto "*Jornalista Merval Pereira recebe prêmio da Universidade de Columbia: Colunista do jornal "O Globo" receberá medalha e um prêmio de US\$ 5 mil. Premiação acontecerá em Nova York no dia 14 de outubro*" foi analisado como *alegria* na primeira anotação e como *neutro* na segunda anotação.

Tabela 3. Valores kappa por emoção

Categoria	Neutro	Repugnância	Alegria	Medo	Raiva	Surpresa	Tristeza
kappa	0,50	0,23	0,47	0,31	0,18	0,33	0,43

No segundo experimento o objetivo foi verificar se havia diferença entre as duas anotações quando um mesmo texto é analisado por um mesmo anotador em datas diferentes. Percebemos que o grau de subjetividade e o alto número de textos a serem anotados por anotador estavam gerando diferenças de "comportamento" nos anotadores. O sistema de anotação foi configurado para que o anotador anotasse 25 textos por sessão. Se ele quiser, poderia realizar várias sessões em sequência. Como o intervalo

entre sessões poderia ser curto (segundos) ou longo (semanas), os anotadores não perceberam que anotaram duas vezes alguns textos. O sistema de anotação foi configurado então para que, aleatoriamente, em torno de 20% do total de textos do corpus fossem anotados duas vezes por um mesmo avaliador. Assim, 438 textos foram anotados duas vezes pelo mesmo avaliador na primeira etapa do processo de anotação. Estes textos foram analisados no segundo experimento: a Tabela 4 apresenta a matriz de confusão da concordância entre as duas anotações realizadas pelo mesmo anotador. A Tabela 5 apresenta os valores de concordância kappa obtidos por emoção.

Tabela 4. Textos anotados duas vezes pelo mesmo anotador

		Anotação 1						
		Neutr	Repugnância	Alegri	Medo	Raiva	Surpres	Tristez
Anotação 2	Neutro	75	4	2	1	1	7	8
	Repugnância	4	14	0	4	3	1	8
	Alegria	2	0	28	0	0	4	1
	Medo	0	5	0	36	2	2	24
	Raiva	1	2	0	1	7	0	2
	Surpresa	14	4	4	2	3	33	8
	Tristeza	5	5	2	11	4	9	85

O grau kappa de concordância obtido neste experimento foi de 0,55 e, dessa forma podemos concluir que mesmo quando um texto é anotado duas vezes pelo mesmo avaliador, ainda assim o grau de discordância é bastante alto.

Tabela 5. Valores kappa por emoção

Categoria	Neutro	Repugnância	Alegria	Medo	Raiva	Surpresa	Tristeza
kappa	0,68	0,36	0,77	0,51	0,40	0,46	0,52

Alguns experimentos também foram realizados submetendo o corpus, ou parte dele, a um método de identificação de emoções. O método treina um classificador SVM [Chang e Lin 2011] para identificar a emoção predominante nos textos [Dosciatti et al. 2013]. O SVM foi configurado com kernel RBF, $cost = 1$ e $gamma = 0$ e avaliado com validação cruzada com 10 partes. Assim, o terceiro experimento teve como objetivo verificar se os textos que tiveram total concordância obtiveram um resultado melhor ao serem submetidos ao método de identificação de emoções. Foram extraídos do corpus de notícias dois conjuntos de amostras, um composto de 994 textos, que tiveram concordância entre os anotadores e outro composto de 994 textos, em que os anotadores discordaram.

Tabela 6. Identificação de emoções no conjunto de textos sem concordância

	A	B	C	D	E	F	G	Precisão	Cobertura	F-Measure
A = Neutro	206	11	29	3	9	4	7	0,61	0,77	0,68
B = Alegria	39	18	6	2	5	6	8	0,34	0,21	0,26
C = Repugnância	45	9	83	9	13	9	11	0,51	0,46	0,48
D = Tristeza	12	2	13	113	16	9	7	0,66	0,66	0,66
E = Medo	16	3	9	29	61	6	4	0,57	0,48	0,52
F = Raiva	6	1	11	7	0	27	3	0,42	0,49	0,45
G = Surpresa	14	9	13	8	4	3	56	0,58	0,52	0,55

Acurácia: 56,7%

Nos resultados apresentados na Tabela 6, o classificador foi treinado e testado com textos que possuem um alto grau de discordância entre os anotadores, $\kappa=0,38$, e obteve uma taxa de acerto de 56,7%.

Tabela 7. Identificação de emoções no conjunto de textos com concordância

	A	B	C	D	E	F	G	Precisão	Cobertura	F-Measure
A = Neutro	248	13	8	8	3	0	14	0,72	0,84	0,78
B = Alegria	40	19	7	13	2	1	15	0,35	0,20	0,25
C = Repugnância	13	3	29	12	6	0	3	0,42	0,44	0,43
D = Tristeza	25	7	10	239	8	0	10	0,73	0,80	0,76
E = Medo	5	2	5	27	23	0	11	0,45	0,32	0,37
F = Raiva	1	0	4	8	0	1	1	0,33	0,07	0,11
G = Surpresa	12	10	6	20	9	1	92	0,63	0,61	0,62
Acurácia: 65,5%										

Nos resultados apresentados na Tabela 7, o classificador foi treinado e testado com textos que tiveram total concordância durante a anotação e obteve uma taxa de acerto de 65,5%. Percebe-se que existe um melhor desempenho do método de AS quando os textos submetidos a ele tiveram maior concordância durante a anotação. Aplicou-se um teste de hipótese para comparar duas proporções amostrais (teste Z [Palaniswamy e Palaniswamy 2006]) para verificar se a acurácia obtida com o conjunto de textos que tiveram concordância era melhor que a acurácia obtida com o conjunto de textos sem concordância, em um nível de significância de 5%. O resultado do teste apresentou p-valor igual a 0,00003. Portanto, conclui-se que o método de AS teve um desempenho significativamente superior ao ser treinado e testado com textos de notícias que tiveram total concordância entre os avaliadores durante o processo de anotação.

Os dois conjuntos de dados usados no terceiro experimento foram unificados para serem usados no quarto experimento, que teve como objetivo verificar o desempenho do método de AS ao ser treinado e testado com uma mesma quantidade de textos sem concordância e com concordância (Tabela 8).

Tabela 8. Identificação de emoções no conjunto de 1.988 textos

	A	B	C	D	E	F	G	Precisão	Cobertura	F-Measure
A = Neutro	435	28	20	22	11	1	23	0,65	0,81	0,72
B = Alegria	72	64	7	13	3	9	15	0,42	0,35	0,38
C = Repugnância	69	11	103	42	18	9	8	0,56	0,40	0,46
D = Tristeza	41	10	18	316	42	15	11	0,64	0,70	0,67
E = Medo	18	5	8	57	112	7	13	0,56	0,51	0,53
F = Raiva	7	2	12	22	1	33	5	0,42	0,40	0,41
G = Surpresa	24	32	16	24	13	5	136	0,65	0,54	0,59
Acurácia: 60,3%										

Ao comparar os percentuais de acurácia, verificou-se que a acurácia obtida com textos que tiveram concordância é maior que a obtida com o conjunto completo de textos. Para confirmar essa hipótese, também foi aplicado o teste Z, em um nível de significância de 5%. O resultado do teste apresentou p-valor igual a 0,003 e permitiu concluir que o desempenho do método foi melhor quando se usou textos com total concordância. Isso significa que os textos que tiveram discordância prejudicaram o aprendizado do classificador. Com base nos resultados obtidos no terceiro e no quarto experimento foi possível verificar que para o método de identificação de emoções em

questão, a taxa de concordância dos anotadores impacta diretamente no desempenho do mesmo.

5. Conclusões e Trabalhos Futuros

Neste artigo apresentamos o relato do processo de rotulação de um corpus de notícias. Experimentos foram realizados para entender o baixo grau de concordância entre os anotadores. Com base na análise dos resultados foi possível chegar a algumas conclusões. Inicialmente é importante destacar que textos jornalísticos não têm uma grande variabilidade de emoções expressas, em função da forma de escrita utilizada por partes de seus autores (jornalistas). Pôde-se concluir ainda que analisar emoções em um nível de documento contribui para se obter um baixo grau de concordância devido ao grande número de documentos que contém mais de uma emoção presente. Finalmente, o fato de nos interessar a identificação de seis diferentes emoções também colabora para que o grau de concordância entre os anotadores seja reduzido.

Foi possível observar o desempenho de um método de identificação de emoções quando do processamento do corpus. Percebe-se que ao testar o método com a porção de textos que tiveram total concordância obtém-se uma taxa de acerto de 65,5%, o que pode ser considerado um bom resultado visto que o método, para o Português do Brasil, identifica categorias de emoções usando exclusivamente uma abordagem sem léxicos. Porém, não se pode considerar um resultado obtido com um método no qual o classificador foi treinado somente com textos que tiveram total concordância. Dessa forma, a taxa de acerto de 60,3%, apresentada na Tabela 8, reflete um resultado mais realista.

Na sequência, pretende-se anotar o corpus de notícias em um nível de sentença para verificar o desempenho do método de AS e comparar com os resultados obtidos no nível de documento. Pretende-se também estudar os graus de intensidade indicados pelos anotadores e, até o presente momento, não utilizados efetivamente.

Referências

- Alm, C. O. Roth, D. e Sproat, R. (2005) Emotions from text: Machine learning for text-based emotion prediction. In Proceedings of Human Language Technology Conference / Conference on Empirical Methods in Natural Language Processing.
- Alves, A. L. F., Baptista, C. S., Firmino, A. A., Oliveira, M. G. e Paiva, A. C. (2014) Uma comparação de SVM Versus Naive Bayes -Técnicas para Análise de sentimento nos tweets: Um Estudo de Caso com o 2013 Copa das Confederações. WebMedia 2014, 123-130.
- Aman, S. e Szpakowicz, S. (2007) Identifying expressions of emotion in text. In: Proc. 10th International Conf. Text, Speech and Dialogue. SpringerVerlag, 196-205.
- Artstein, R. e Poesio, M. (2005) Bias decreases in proportion to the number of annotators. In Proceedings of FG-MoL 2005, 141-150, Edinburgh.
- Artstein, R. e Poesio, M. (2008) Inter-coder agreement for computational linguistics, Computational Linguistics, vol. 34 n.4, 555-596.

- Balahur, A. e Steinberger, R. (2009) Rethinking Sentiment Analysis in the News: from Theory to Practice and back. Proceeding of WOMSA.
- Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37-46.
- Chang, C.-C. e Lin, C.-J. (2011) LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1-27:27.
- Di Eugenio, B. e Glass, M. (2004) The kappa statistic: A second look. *Computational Linguistics*, 30(1):95-101.
- Dosciatti, M. M., Ferreira, L. P. C. e Paraiso, E. C. (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.
- Ekman, P. (1992) An argument for basic emotions. *Cognition & Emotion* 6. 3-4: 169-200.
- Freitas, C., Motta, E., Milidiú, R. L. e César J. (2014) Sparkling Vampire... lol! Annotating Opinions in a Book Review Corpus. In Sandra Aluísio & Stella E. O. Tagnin (eds.), *New Language Technologies and Linguistic Research: A Two-Way Road*. Cambridge Scholars Publishing, 128-146.
- Ghazi, D., Inkpen, D. e Szpakowicz, S. (2014) Prior and contextual emotion of words in sentential context, *Comput. Speech Lang.*, vol. 28, no. 1, 76 -92.
- Gomes, H., Neto, M. C. e Henriques, R. (2013) Text Mining: Sentiment analysis on news classification. 8th Iberian Conference on Information Systems and Technologies, 1-6.
- Habernal, I. Ptáček, T. e Steinberger, J. (2014) Supervised sentiment analysis in Czech social media. *Inf. Process. Manag.*, vol. 50, no. 5, 693-707.
- Klebanov, B. B. e Beigman, E. (2009) From annotator agreement to noise models. *Computational Linguistics*, vol. 35 n.4, 495-503.
- Krippendorff, K. (1980) *Content Analysis: An Introduction to Its Methodology*. Chapter 12. Sage, Beverly Hills, CA.
- Nascimento, P., Aguas, R., Lima, D., Kong, X., Osiek, B., Xexeo, G. e Souza, J. (2012). Análise de sentimento de tweets com foco em notícias. *Brazilian Workshop on Social Network Analysis and Mining*.
- Palaniswamy, U. R. e Palaniswamy, K. M. (2006) *Handbook of Statistics for Teaching e Research in Plant e Crop Scienc*. 201-203, Publishing Food Products Press.
- Wiebe, J., Wilson, T. e Cardie, C. (2005). Annotating expressions of opinions and emotions in language, in: *Language Resources and Evaluation*, vol. 39, issue 2-3, 165-210.
- Strapparava, C. e Mihalcea, R. (2008) Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*. ACM, 1556-1560.