O Que Tweets Contendo Emoticons Podem Revelar Sobre Sentimentos Coletivos?

Pollyanna Gonçalves[1], **Fabrício Benevenuto**[2], **Virgílio Almeida**[2]

[1]Universidade Federal de Ouro Preto (UFOP) – Ouro Preto – MG – Brasil

[2]Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte – MG – Brasil

{pollyannaogoncalves}@gmail.com

{fabricio,virgilio}@dcc.ufmg.br

Abstract. *Twitter has become an important social communication mean where users post messages about everything. Several social networking's users use emoticons, a pictorial representation of a facial expression, to try to represent the sentiment expressed in their messages. Potentially, when analyzed in large scale, messages containing emoticons could be used to measure public mood fluctuations at the level of populations. Such measurement could have several applications, such as to predict collective mood trends of populations related to specific topics as well as to understand the effects of social phenomena in people's feelings. In this paper, we give a first step on this direction by providing a deep understanding of how emoticons are used by Twitter's users, analyzing the most popular emoticons and how they appear associated to global scale events. In order to validate our approach, we used a unique dataset containing 1.8 billion tweets to evaluate public mood fluctuations on a series of events that span topics related to tragedies, releases, politics, health, as well sport events. Our results suggest that emoticons can correctly capture expected sentiments of messages of this social networking.*

Resumo. *O Twitter tem se tornado um importante meio de comunicação social onde usuários postam mensagens sobre tudo. Muitos usuários de redes sociais utilizam emoticons, uma representação pictorial de uma expressão facial, para tentar expressar sentimentos em suas mensagens. Potencialmente, quando analisadas em grande escala, mensagens contendo emoticons poderiam ser utilizadas para medir variações de humor do público coletivo. Esse tipo de medição pode ter diversas aplicações como prever o humor coletivo de uma população em relação a tópicos específicos com o objetivo de entender os efeitos de fenômenos sociais no sentimento das pessoas. Nesse artigo, damos um primeiro passo nessa direção, fornecendo um estudo profundo sobre como emoticons são utilizados pelos usuários do Twitter e analisando os emoticons mais populares e como eles aparecem associados a eventos de escala globais. Para validar nossa abordagem, utilizamos um conjunto de dados contendo 1,8 bilhões de tweets para analisar variações de humor dos usuários em uma série de eventos que cobrem tópicos relacionados a tragédias, lançamentos, política, saúde e eventos esportivos. Nossos resultados sugerem que emoticons podem capturar corretamente sentimentos em mensagens da rede, mas também revelam limitações dessa abordagem.*

1. Introdução

Recentemente, a rede social Twitter tem emergido como uma plataforma popular onde usuários postam mensagens curtas sobre tudo, incluindo conversas diversas, links interessantes, notícias e opiniões sobre acontecimentos marcantes. Celebridades e usuários comuns estão constantemente utilizando o Twitter para compartilhar informações em tempo real. Estimativas atuais sugerem que os 200 milhões de usuários ativos no Twitter postam cerca de 400 milhões de tweets[1] diariamente [Wickre].

Milhões de usuários compartilham no Twitter informações que vão descobrindo enquanto navegam na Internet, fazendo da rede social uma importante fonte de notícias durante eventos inéditos como revoluções, epidemias e desastres [Gomide et al. 2011, MacAskill 2009, Sakaki et al. 2010]. Não surpreendentemente, quando eventos importantes acontecem, usuários tendem a expressar seus sentimentos em relação a esses, mostrando como eles afetaram seu estado de humor. Para isso, diversos usuários tendem a utilizar emoticons, uma representação pictorial de uma expressão facial construída com números, letras e pontuações. Emoticons são frequentemente utilizados para reforçar o temperamento de um estado de humor, e podem melhorar ou até mudar a interpretação de um texto simples.

O uso de emoticons em conversas na Web tem se tornado tão popular que esse tipo de comunicação está sendo incorporada pelo conhecido Dicionário de Oxford [iran.twitter 2011]. Além de faces contendo expressões de tristeza, felicidade e outras expressões, há um incontável número de emoticons sendo utilizados na Internet. Por exemplo, "‘<3"’ representa um coração e é usualmente utilizado para expressar amor ou afeição, já o emoticon "‘\o/"’ é comumente utilizado para celebrações.

Potencialmente, quando analisadas em larga escala, mensagens contendo emoticons podem ser utilizadas para medir variações de humor do público à nível de população. Esse tipo de medição pode possuir diversas aplicações, como a previsão de humor coletivo de uma população relacionada a tópicos específicos com o objetivo de compreender como, quando e porque sentimentos variam de acordo com eventos sociais, econômicos, políticos, etc. Além das potenciais aplicações, há um número limitado de estudos nessa direção, onde a maioria está relacionada a extração de opiniões (não sentimentos) [Pang and Lee 2008], ou não foram propriamente desenvolvidas para esse contexto [Dodds and Danforth 2010].

Para responder a pergunta que dá nome ao artigo, fizemos diversas analises utilizando dados empíricos de um conjunto único de dados contendo 1,8 bilhões de tweets. Primeiramente definimos uma lista de 21 emoticons que estão associados a vários sentimentos. Para validar nossa abordagem, fornecemos análises sobre variações do humor do público em relação a uma série de eventos associados a tragédias, lançamentos, política, saúde e eventos esportivos. Nossa avaliação fornece evidências que o uso de emoticons como método para capturar sentimentos funciona corretamente para eventos em que sentimentos relacionados a eles são esperados.

O restante desse trabalho é analisado como se segue. Em seguida, fazemos um breve resumo das abordagem existentes na detecção de sentimentos na Web. Logo após, detalhamos como nosso método funciona e fornecemos evidencias experimentais que a mesma funciona. Finalmente, concluímos o artigo e oferecemos direções para trabalhos futuros.

[1]Mensagem de no máximo 140 caracteres que são postadas pelos usuários do Twitter.

2. Trabalhos Relacionados

Com o crescimento das redes sociais online, a análise de sentimentos e extração de opiniões tem se tornado assunto de estudo para muitos pesquisadores e diferentes técnicas tem sido propostas e aplicadas em diferentes contextos. A seguir, fornecemos um breve resumo das diferentes técnicas utilizadas para detecção sentimentos na Web e como são aplicadas.

Uma visão geral dessas abordagens foram apresentadas em um *survey* de Pang e Lee [Pang and Lee 2008]. Esse estudo cobre diversos métodos que utilizam processamento de linguagem natural para análise de sentimentos, uma técnica onde propriedades subjetivas de um texto são inferidas utilizando métodos estatísticos. Tais métodos são geralmente adequados na construção de aplicações para extração de opiniões que analisam o feedback de consumidores/usuários sobre produtos para entender o que os mesmos estão pensando [Airoldi et al. 2006, Aue and Gamon 2005].

Pak e Paroubek [Pak and Paroubek 2010] consideram para conhecer a subjetividade de um texto escrito por usuários no Twitter formas gramaticais como verbos e adjetivos. Como resultado dos experimentos, os autores concluíram que adjetivos superlativos, verbos na primeira pessoa e pronomes pessoais são frequentemente utilizados por usuários que expressam emoções e opiniões. Contrariamente, adjetivos comparativos e pronomes próprios são considerados um forte indicador de objetividade.

Uma outra direção consiste na medição de índices de felicidade em um texto. Dodds e Danforth [Dodds and Danforth 2010] propõem um método que computa o nível de felicidade de um texto estruturado. Os autores analisaram o índice de felicidade em letras de músicas famosas e em mensagens compartilhadas em blogs. Os resultados apresentados sugerem que o índice de felicidade nas letras de músicas foi decrescente entre a década de 60 e 90, ao contrário do blogs, que obteve um aumento do índice entre 2005 e 2009.

Kim e colaboradores [Kim and Gilbert 2009], propõem um método para detecção de sentimentos utilizando o *The Affective Norms for English Words* (ANEW), um conjunto de dados que contém uma classificação emocional normativa para 1034 palavras em inglês. Cada palavra da sua base de dados contendo uma amostra de tweets sobre a morte de Michael Jackson foi associada com uma escala de três dimensões emocionais, são elas: prazer, excitação e dominância. Como resultados, puderam mostrar que houve um aumento no sentimento de valência no período analisado. Outro estudo fez o uso da escala psicométrica *Profile of Mood States* (POMS) [Adamic et al. 2011]. O POMS consiste de um conjunto adjetivos que qualificam 6 sentimentos específicos: tensão, depressão, raiva, fadiga, vigor e confusão. Os autores aplicaram a escala para identificar sentimentos em uma amostra de tweets e avaliaram o humor dos usuários em relação a flutuações do mercado de ações e em eventos como as eleições nos Estados Unidos. No entanto, essa escala também deixa de fora sentimentos importantes que estão ligados a felicidade de um usuário, pois ela é capaz de detectar apenas sentimentos negativos.

Existem ainda alguns estudos relacionados ao uso de emoticons em aplicações Web. Jonathon Read [Read 2005] coletou textos contendo emoticons, postados em fóruns, para criar um conjunto de treinamento para classificação de sentimentos utilizando emoticons como "`:-)`"e "`:-(`". Já Amy Ip [Amy 2011] mostrou em seu trabalho que o uso de emoticons em mensagens enviadas por sistemas de mensagens instantâneas não aumentam sua positividade, e sim fazem com que essas se tornem mais extremas, assim como com o uso de pontuações

como "!".

Sarah Rojas e colaboradores [L. et al. 2012] utilizaram em sua pesquisa uma base de dados coletada da conhecida ferramenta de troca de mensagens e vídeo, Skype. Seus resultados mostram que emoticons são bons indicadores da positividade das mensagens trocadas, concluindo que esse tipo de comunicação são utilizados para compensar a falta de informação visual presentes nesse tipo de conversa.

Hogenboom A. e colaboradores [Hogenboom et al. 2013] analisaram como emoticons podem influenciar as mensagens postadas por usuários do Twitter. Seus experimentos, executados em uma base contendo pouco mais de 2 mil tweets no idioma alemão, indica que os sentimentos associados aos emoticons utilizados nestes geralmente dominam os expressos em mensagens textuais. Os autores também concluíram que as pessoas geralmente utilizam emoticons na tentativa de desambiguar o sentimento que se deseja expressar em suas mensagens.

No geral, nenhum dos trabalhos apresentados aqui parecem investigam como, e o que, emoticons podem dizer sobre o sentimento do coletivo no Twitter. Acreditamos que este conhecimento poderia permitir não só uma melhoria nas técnicas de análise de sentimentos existentes, mas também relevar potenciais limitações e tendências no seu uso.

3. Extraindo Emoticons do Twitter

A seguir, descrevemos nossa base de dados do Twitter e a estratégia utilizada para mapear emoticons a um conjunto de sentimentos. A Tabela 1 exibe as palavras-chave utilizadas e o número total de tweets filtrados para cada evento considerado. A primeira coluna contem o nome que identifica o evento ao qual será referido no restante desse trabalho.

Evento	**Data**	**Palavras-chave**	# Tweets
H1N1	03.01 - 07.31.2009	tamiflu, outbreak, antiviral, virus, influenza, pandemia, pandemic, flu, pig, mexico, h1n1, swine, World Health Organization.	335,969
Queda-AirFrance	06.01 - 06.2009	victims, passengers, A330, 447, crash, airplane, airfrance.	29,765
Obama	01.18 - 22.2009	barack obama, politics, democrats, election, white house, vote, speech, presidential, republicans, candidate, yes we can.	43,015
Eleições-EUA	11.02 - 06.2008	clinton, biden, palin, voting, vote, elect, candidate, campaign, mccain, democrats, republicans, obama, bush.	185,477
Morte-MJ	06.25 - 30.2009	rip, mj, michael jackson, death, died, king of pop, overdose, drugs, heart attack, conrad murray, thriller.	56,259
Susan Boyle	04.11 - 16.2009	susan boyle, I dreamed a dream, britain's got talent, les miserables.	7,142
Harry Potter	07.13 - 17.2009	harry potter, half-blood prince, rowling.	194,356
Olimpíadas-Pequim	08.06 - 26.2008	olympics, medals, china, beijing, sports, peking, sponsor.	12,815
Tsunami	09.28 - 10.04.2009	tsunami, samoa islands, tonga, earthquake.	23,881
Terremoto-Haiti	01.11 - 17.2010	haiti, earthquake, richter, port au prince, jacmel, leogane.	236,096

Tabela 1. Lista dos eventos analisados.

3.1. A Base de Dados

Os dados utilizados neste trabalho incluem um extenso conjunto de dados proveniente de um estudo de medição anterior [Cha et al. 2010] que inclui um histórico completo de tweets postados na rede social Twitter por todos os usuários desde sua criação, em 2006, até Agosto de 2009. O conjunto de dados contém **54.981.152 usuários** com **1,9 bilhões de links** entre eles e **1.755.925.520 de tweets** postados. Cerca de 8% das contas desses usuários estavam definidas como privadas, o que implica que apenas seus amigos poderiam visualizar suas ligações e tweets. Para este trabalho, esses usuários foram ignorados.

Esta base de dados é apropriada para fins do presente trabalho pelas seguintes razões. Primeiro, o conjunto de dados engloba todos os usuários cadastrados desde o surgimento da rede até Agosto de 2009. Assim, ela não é baseada em técnicas de amostragem que podem introduzir tendências para algumas características dos usuários. Em segundo lugar, este conjunto de dados contém todos os tweets desses usuários, o que a torna essencial para o estudo de eventos passados que ocorreram no Twitter.

3.2. Mapeamento de Emoticons

Milhares de símbolos são utilizados por usuários da Web para expressar sentimentos. Baseados em diversas listas de emoticons mais utilizados na Internet [iran.twitter a, iran.twitter b, iran.twitter d], decidimos por filtrar tweets que continham pelo menos um dos 21 emoticons selecionados. Como se sabe, usuários podem escrever um mesmo emoticon de diversas maneiras, por exemplo, o sentimento de felicidade pode ser expresso através do símbolo "`:)`", ou mesmo "`=]`". Portanto, para este trabalho será considerado algumas derivações de um emoticon.

A Tabela 2 sumariza os emoticons utilizados neste trabalho, assim como os símbolos mais utilizados associados a cada um.

4. Padrões de Uso de Emoticons no Twitter

A seguir, analisamos diferentes aspectos relacionados ao uso de emoticons nos tweets e sua capacidade de medir sentimento coletivo em relação a eventos globais.

4.1. Abrangência de Emoticons

Pode ser interessante sabermos qual a porcentagem de tweets postados na base de dados é caracterizada por ter ou não emoticons. Uma análise feita revelou que cerca de 138.071.217 tweets (cerca de 7,8%) possuem apenas 1 emoticon. Alguns tweets podem conter 2 ou mais emoticons, como por exemplo o tweet "Good morning everyone `:) =D` ". O número de casos como esses são relevantes desde que, um número pequeno pode dizer algo sobre o tweet. O número de mensagens com essa característica é 162.549.422 que corresponde a cerca de 9,2% da base de dados. Esse baixo valor, mostra que uma análise de sentimentos baseada somente em emoticons pode ficar limitada ou mesmo tendenciosa a um conjunto pequeno de dados.

Com o objetivo de mostrar como emoticons são postados juntos em um mesmo tweet, calculamos a porcentagem de ocorrência de cada emoticon dado que outro emoticon da lista apareceu. A Tabela 3 apresenta os resultados da ocorrência de cada emoticon.

Podemos perceber nos valores da diagonal apresentado na Tabela 3 que usuários que utilizam mais de 1 emoticon no mesmo tweet parecem preferir repetir o emoticon já utilizado na mensagem. Esse comportamento pode indicar que esses usuários tentam enfatizar um sentimento expressado pelo emoticon, utilizando-o duas ou mais vezes no mesmo post. Mas os resultados também mostram outros comportamentos interessantes, como o uso do emoticon *Confuso* seguindo do emoticon *Surpreso*, no mesmo tweet, sendo que ambos os sentimentos são bastante relacionados. Uma mesma análise pode ser feita quando percebemos que 22,82% dos tweets com essa característica que apresentam o emoticon *Muito Triste* também contém o emoticon *Envergonhado*.

Emoticons	Significado	Símbolos utilizados
	Feliz	:) :] :o) :o] :o} :} :-] :-) :-} =) =] =} =^] =^) =^}
	Muito Feliz	:B :-D :-B :^D :^B =B =^D =^B
	Muito Triste	D: D= D-: D^: D^=
	Triste	:(:[:{ :o(:o[:o{ :^(:^[:^{ :-[:-(:-{ =(=[={ =^[=^(=^{
	Raiva	>=(>=[>={ >=(>:-{ >:-[>:-(>=^[>=^(>=^{
	Pensando	:| =| :-|
	Surpreso	:o :0 =O :@ =@ :- :-@ :^o :^@
	Chorando	:'(:'[:'{ ='{ ='(='[
	Lágrimas de Felicidade	:') :'] :'} =') ='] ='}
	Segredo	:x =X :# =# :-x :-# :^x :^#
	Animado	\o/ \o\ /o/
	Amor	<3
	Alegre	^.^ ^-^ ^_^ ^^
	Suspeito	¬¬ ¬.¬ ¬_¬ -.- -.-' -_- -_-'
	Língua	:P :p :b :-p :-P :-b :^p :^P :^b =P =p =b =^p =^P =^b
	Problema	>.< >< >_<
	Beijo	:* =* :-*
	Confuso	o.O O_o Oo
	Desapontado	=\ :\ =/ :/
	Piscando	;) ;] ;}
	Envergonhado	:$ =$

Tabela 2. Emoticons e suas variações.

4.2. Frequência de cada Emoticon

A seguir, a Tabela 4 quantifica o quanto cada o quanto cada tipo de emoticon aparece nos tweets. Essa análise serve como uma forma de identificar os emoticons mais utilizados pelos usuários. Como podemos perceber, os 3 emoticons mais utilizados pelos usuários da rede social são *Feliz*, com cerca de 41%, *Muito Feliz*, com quase 11%, e *Triste*, também com quase 11% de uso.

4.3. Palavras e *Hashtags* Associadas à Emoticons

Emoticons podem ser utilizados associados a palavras ou *hashtags* que talvez estejam relacionadas aos sentimentos expressos por eles. No Twitter, o símbolo # (*hashtag*) é utilizado para marcar palavras-chaves ou tópicos em um tweet. Normalmente, usuários utilizam *hashtags* antes de uma palavra-chave para categorizar o tweet e tornar mais fácil a busca por aquele tweet

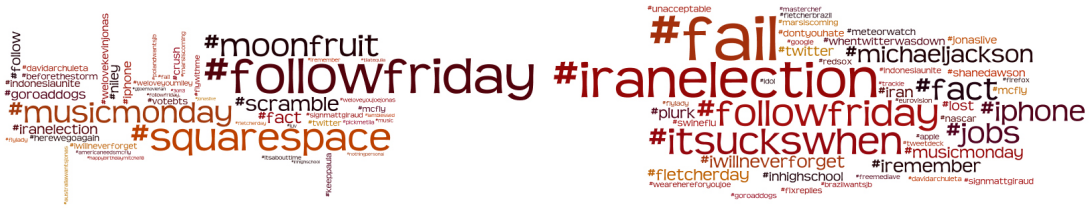
	88.02	0.92	0.07	0.66	0.01	0.06	1.44	0.02	0.01	0.02	0.02	0.99	0.14	0.01	0.52	0.02	0.05	0.09	0.01	0.59	0.07
	3.14	**87.46**	1.15	0.75	0.01	0.08	0.62	0.04	0.02	0.08	0.12	1.39	0.40	0.00	1.20	0.07	0.16	0.22	2.06	0.76	0.26
	1.14	5.50	**62.24**	1.72	1.72	0.09	0.92	0.06	0.02	0.11	0.04	0.48	0.10	0.19	0.52	0.13	0.16	0.42	1.00	0.63	22.82
	2.49	0.83	0.40	**90.90**	1.20	0.14	1.44	0.17	0.01	0.04	0.02	0.60	0.09	0.00	0.50	0.07	0.05	0.15	0.23	0.42	0.26
	0.73	0.45	0.18	36.64	**50.06**	7.56	0.65	0.11	0.01	0.02	0.01	0.24	0.05	1.13	0.26	0.08	0.02	0.11	1.48	0.12	0.11
	3.72	1.42	0.34	2.16	3.88	**82.79**	1.34	0.15	0.08	0.07	0.02	0.48	0.09	0.07	0.77	0.10	0.08	0.25	1.40	0.45	0.33
	8.72	1.10	0.34	2.31	0.03	0.14	**81.07**	0.01	0.09	0.05	0.25	0.25	0.10	0.05	1.18	0.05	0.05	2.84	0.03	1.04	0.28
	1.33	0.95	0.26	3.38	0.07	0.19	0.54	**82.30**	0.06	0.08	0.04	0.83	0.12	0.46	0.50	0.15	0.09	0.19	8.02	0.31	0.13
	2.27	1.88	0.36	0.72	0.01	0.38	0.46	0.25	**77.61**	0.05	0.06	2.09	0.22	0.37	0.77	0.08	0.12	0.18	11.16	0.67	0.28
	1.64	1.85	0.49	0.80	0.01	0.09	1.15	0.08	0.01	**81.12**	0.18	0.37	0.26	0.05	0.86	0.19	0.51	0.38	8.90	0.65	0.40
	1.64	2.64	0.19	0.36	0.01	0.02	0.58	0.04	0.01	0.18	**87.04**	0.23	0.54	0.07	0.75	0.11	0.40	1.49	2.66	0.70	0.32
	8.05	3.32	0.24	1.30	0.02	0.07	0.33	0.09	0.06	0.04	0.02	**83.17**	0.46	0.10	0.73	0.09	0.13	0.15	0.09	0.90	0.65
	2.63	2.17	0.12	0.44	0.01	0.03	0.30	0.03	0.01	0.06	0.13	1.04	**87.67**	0.23	1.04	0.30	0.14	0.39	0.07	2.73	0.46
	4.05	0.05	0.14	0.61	0.01	0.04	0.39	0.04	0.02	0.04	0.19	1.44	0.31	**83.44**	0.14	0.44	0.19	0.64	3.58	3.78	0.44
	3.13	2.12	0.19	0.81	0.01	0.08	1.18	0.04	0.02	0.07	0.06	0.54	0.34	0.01	**88.19**	0.09	0.07	0.33	1.45	0.96	0.32
	1.27	1.09	0.42	0.95	0.03	0.09	0.46	0.10	0.01	0.13	0.07	0.56	0.83	0.06	0.73	**84.02**	0.11	0.60	2.58	5.54	0.34
	3.14	2.82	0.59	0.88	0.01	0.08	0.46	0.07	0.02	0.42	0.33	0.94	0.46	0.07	0.70	0.13	**85.38**	0.22	1.46	1.34	0.47
	1.19	0.86	0.36	0.53	0.01	0.06	6.41	0.03	0.01	0.07	0.27	0.25	0.29	0.02	0.74	0.16	0.05	**87.36**	0.40	0.57	0.37
	5.40	1.73	0.85	1.17	0.05	0.07	2.44	0.08	0.02	0.10	0.07	0.52	0.33	0.00	0.88	0.08	0.06	1.36	**81.37**	3.10	0.31
	2.14	0.87	0.15	0.43	0.00	0.03	0.67	0.02	0.01	0.03	0.04	0.43	0.57	0.00	0.61	0.41	0.09	0.16	1.93	**91.36**	0.06
	0.67	0.79	14.42	0.71	0.01	0.06	0.48	0.02	0.01	0.06	0.04	0.81	0.26	0.01	0.55	0.07	0.08	0.28	2.46	0.16	**78.08**

Tabela 3. Distribuição de emoticons em um mesmo tweet com 2 ou mais emoticons.

no Twitter [iran.twitter c].

Com o objetivo de identificar quais os termos representados por *hashtags* estão mais associados aos emoticons encontrados considerando-se a base completa de tweets, plotamos uma nuvens de *hashtags* para os emoticons *Feliz* e *Triste*, emoticons positivos e negativos, respectivamente, conforme a Tabela 2. Escolhemos apenas esses dois emoticons para fazer a análise de *hashtags* e palavras relacionadas por questões de escassez de espaço na máquina que utilizamos para esse trabalho.

A Figura 1(a) apresenta o ranking das 50 *hashtags* que mais apareceram em tweets que continham o emoticon *Feliz* ou suas variações. Já a Figura 1(b) apresenta o mesmo ranking, porém para as *hashtags* que mais apareceram associadas ao emoticon *Triste*.



(a) Top-Hashtags-Feliz

(b) Top-Hashtags-Triste

Figura 1. Top 50-Hashtag associados a emoticons felizes e tristes, respectivamente.

Também pode ser interessante sabermos quais palavras postadas pelos usuários na rede social estão mais associados a alguns emoticons. A Figura 2(a) apresenta uma nuvem de frequência de palavras para as 50 que mais apareceram junto ao emoticon *Feliz*, já a Figura 2(b) apresenta as top 50 palavras mais associadas aos emoticon *Triste*.

A partir dos resultados obtidos, podemos então concluir que palavras positivas e negativas são utilizadas associadas a emoticons felizes e tristes, respectivamente. Por exemplo, as palavras "Love"("Amor", em Português), "Good"("Bom", em Português) e "Thanks"("Obrigado", em Português), que possuem peso positivo, apareceram em tweets postados por usuários que também utilizaram o emoticon *Feliz* e suas variações. Ao verificarmos a nuvem de palavras referente aos emoticons tristes, a palavra "Sad"("Triste", em Português) foi mais utilizada associada ao emoticon *Triste*, ou suas variações.

Emoticon	Significado	#Tweets	%Tweets com emoticons
😀	Feliz	66,977,387	41.20%
😂	Muito Feliz	17,840,029	10.97%
😟	Triste	17,249,171	10.61%
😉	Piscando	13,911,464	8.55%
😜	Língua	9,766,862	6.00%
😕	Desapontado	9,373,591	5.76%
😮	Surpreso	8,244,469	5.07%
😠	Raiva	7,179,136	4.41%
😵	Confuso	3,859,419	2.37%
😁	Alegre	2,933,829	1.80%
😩	Muito Triste	2,173,062	1.33%
🤔	Pensando	517,541	0.03%
💋	Beijo	487,180	0.02%
😂	Lágrimas de Felicidade	390,959	0.02%
👐	Animado	364,656	0.02%
🤐	Segredo	309,592	0.01%
😭	Chorando	303,992	0.01%
❤️	Amor	250,936	0.01%
🤨	Suspeito	209,311	0.01%
😖	Problema	206,835	0.01%
😳	Vergonha	205,194	0.01%
Total	**162,549,422**		

Tabela 4. Popularidade dos Emoticons nos Tweets com emoticons.

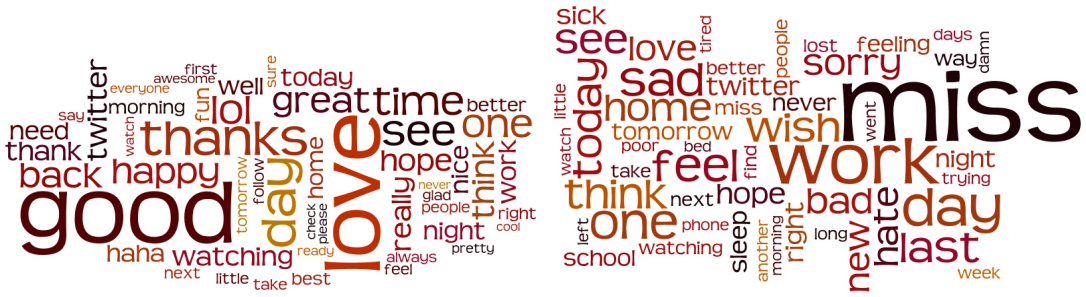
Por outro lado, quando olhamos para as nuvens de *hashtags* mais associadas aos emoticons podemos perceber que esse tipo de termo não está diretamente associado com o sentimento que o usuário estava tentando expressar no tweet com aquele emoticon, muitas vezes porque uma *hashtag* tenta sumarizar o conteúdo ou o assunto do tweet como um todo.

5. Emoticons Associados a Eventos Globais

Finalmente, para investigar se emoticons podem ser utilizados como uma forma de medir alterações de humor dos usuários do Twitter, foram realizadas diversas análises utilizando os eventos descritos na Tabela 1, que são eventos cujo sentimento coletivo é intuitivo ou simples de serem verificados. Primeiramente, construímos gráficos de colunas para mostrar se o uso de emoticons pelos usuários do Twitter podem ser utilizados como uma forma para detecção de sentimentos. Em seguida, foi feito um estudo geográfico que analisa como usuários de diferentes localidades utilizam emoticons para expressar sentimentos em relação a um mesmo evento. Finalmente, apresentamos uma análise capaz de dizer se eventos similares, como por exemplo, a morte do cantor Michael Jackson e a queda do avião da Airfrance, ambos eventos negativos, afeta os usuários de forma parecida ou não.

5.1. Emoticons *vs.* Eventos

Com o objetivo de mostrar que emoticons podem mensurar flutuações de humor comparamos alguns dos eventos descritos na Tabela 1 utilizando gráficos de colunas. A Tabela 5 apresenta os top 5 emoticons mais utilizados pelos usuários em quatro grandes eventos de escala global:



(a) Top-Palavras-Feliz

(b) Top-Palavras-Triste

Figura 2. Top 50-Palavras associados a emoticons felizes e tristes, respectivamente.

o discurso de posse do Presidente Barack Obama, em 2009, a abertura das Olimpíadas de Pequim, em 2008, o lançamento do filme Harry Potter e o Enigma do Príncipe, em 2009 e as eleições Norte-americanas em 2008, respectivamente.

Ranking/Eventos	Barack Obama	Olimpiadas	Harry Potter	Eleições-EUA
1º	(7,8%)	(8,09%)	(17,2%)	(6,9%)
2º	(1,6%)	(3,6%)	(8,7%)	(1,5%)
3º	(1,4%)	(1,9%)	(4,2%)	(1,4%)
4º	(1,2%)	(1,8%)	(3,4%)	(1,1%)
5	(1,0%)	(1,5%)	(2,1%)	(1,0%)

Tabela 5. Top 5 emoticons para quatro eventos globais.

5.2. Emoticons *vs.* Localidade

Esta seção tem como objetivo saber como usuários localizados em diferentes pontos do mundo utilizam emoticons para expressar seus sentimentos no Twitter. O primeiro passo para essa análise se deu com a filtragem de tweets por regiões. Neste trabalho, o foco será analisar e comparar 3 regiões mais populares na rede até 2009, ano da nossa base de dados [SySomos]: América do Norte, que consiste aqui do Canadá e Estados Unidos, Europa, com Reino Unido e Alemanha, e Brasil.

Resultados obtidos na análise dos top emoticons de cada região revelaram que apesar das diferenças geográficas usuários tendem a utilizar emoticons similares. Algumas análises revelaram que usuários norte americanos e brasileiros utilizam mais emoticons associados a felicidade, já usuários europeus geralmente utilizam mais emoticons relacionados a tristeza. Como forma de exemplificar tais diferenças regionais, a seguir vamos analisar a repercussão da morte do cantor Michael Jackson no Twitter.

A Tabela 6 apresenta os *top* 5 emoticons associados a esse evento para as 3 diferentes regiões consideradas.

Como podemos ver, o emoticon *Triste* apareceu em 1º no *ranking* para as 3 localidades, mas outros emoticons como o de *Surpresa*, *Confusão* e *Desapontamento*, presentes no gráfico para o Brasil, não foram muito utilizados por usuários das outras regiões.

Com essas análises geográficas podemos concluir que ao serem expostos a eventos de escalas globais, usuários tendem a expressar seus sentimentos associados a eles de forma si-

Ranking/Região	EUA+Canadá	Reino Unido	Brasil
1º	😢 (49%)	😢 (45%)	😢(24%)
2º	😊 (20%)	😊 (21%)	🤔 (17%)
3º	❤ (8%)	❤ (7%)	😱 (12%)
4º	😃 (5%)	😒 (6%)	🙄 (12%)
5	😒 (4%)	😃 (5%)	😊 (10%)

Tabela 6. Top 5 emoticons para a morte do cantor Michael Jackson em 3 diferentes regiões.

milar. No entanto, quando analisamos a porcentagem de uso de emoticons de cada região na nossa base total percebemos que esse comportamento não se mantém, visto que cada região pode ser alvo de eventos locais que alteram o humor de seus usuários de forma diferente.

6. Conclusões e Trabalhos Futuros

Neste artigo damos um passo na direção de responder a seguinte questão: "'Emoticons podem revelar sentimentos do público em relação a eventos importantes?"'. Emoticons, uma representação pictorial de uma expressão facial com o uso de letras, números e símbolos do teclado, vem sendo utilizados por usuários como uma forma de expressar sentimentos na Web. Utilizando dados empíricos de uma base de dados únicos da rede social Twitter contendo quase 1,8 bilhões de tweets, fomos capazes de calcular pontuações de emoticons para um índic de 21 emoticons listados na Tabela 2.

Mostramos como emoticons são utilizados pelo usuários do Twitter analisando o uso deste tipo de comunicação em vários eventos de escala global, como a morte de celebridades, acidentes trágicos e eventos esportivos. Apresentamos também algumas análises geográficas, apresentando comparações sobre uso desses emoticons por usuários de diferentes regiões quando expostos aos mesmos eventos. Nossos resultados sugerem que usuários de diferentes países expressam seus sentimentos em relação a um mesmo acontecimento utilizando emoticons distintos, mas sempre preservando a positividade ou negatividade do sentimento.

Finalmente, apresentamos nuvens de tags construídas com as 50 palavras e *hashtags* mais utilizadas em tweets que continham emoticons de felicidade e tristeza. Como resultados, concluímos que as palavras encontradas estavam de fato associadas ao sentimento expresso pelos emoticons.

Apesar de nossos resultados concordarem com o esperado para os eventos analisados, se faz necessário um processo de validação mais aprofundado. Assim saberemos o quão eficiente é o método proposto com as técnicas de detecção do estado da arte. Como trabalhos futuros pretendemos comparar a técnica para detecção de sentimentos proposta com outros métodos que vêm sendo utilizadas no contexto de redes sociais. Assim como combiná-la a outros métodos com o objetivo de aumentar sua abrangência e eficiência.

7. Agradecimentos

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

Referências

[Adamic et al. 2011] Adamic, L. A., Baeza-Yates, R. A., and Counts, S., editors (2011). *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*. The AAAI Press.

[Airoldi et al. 2006] Airoldi, E., Bai, X., and Padman, R. (2006). Markov blankets and meta-heuristics search: Sentiment extraction from unstructured texts.

[Amy 2011] Amy, I. (2011). The impact of emoticons on affect interpretation in instant messaging. `http://amysmile.com/doc/emoticon_paper.pdf`.

[Aue and Gamon 2005] Aue, A. and Gamon, M. (2005). Customizing sentiment classifiers to new domains: A case study. In *Recent Advances in Natural Language Processing (RANLP)*.

[Cha et al. 2010] Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. P. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. In *AAAI Conference on Weblogs and Social Media (ICWSM 2010)*.

[Dodds and Danforth 2010] Dodds, P. and Danforth, C. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies*, 11(4):441–456.

[Gomide et al. 2011] Gomide, J., Veloso, A., Jr., W. M., Almeida, V., Benevenuto, F., Ferraz, F., and Teixeira, M. (2011). Dengue surveillance based on a computational model of spatio-temporal locality of twitter. In *ACM Web Science Conference (WebSci)*.

[Hogenboom et al. 2013] Hogenboom, A., Bal, D., Frasincar, F., Bal, M., Jong, F., and Kaymak, U. (2013). Exploiting emoticons in sentiment analysis. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing (SAC)*.

[iran.twitter a] iran.twitter. List of text emoticons: The ultimate resource. `www.cool-smileys.com/text-emoticons`.

[iran.twitter b] iran.twitter. Msn messenger emoticons. `http://messenger.msn.com/Resource/Emoticons.aspx`.

[iran.twitter c] iran.twitter. What are hashtags symbols? `https://support.twitter.com/articles/49309-what-are-hashtags-symbols`.

[iran.twitter d] iran.twitter. Yahoo messenger emoticons. `http://messenger.yahoo.com/features/emoticons`.

[iran.twitter 2011] iran.twitter (2011). Omg! oxford english dictionary grows a heart: Graphic symbol for love (and that exclamation) are added as words. `tinyurl.com/klv36p`.

[Kim and Gilbert 2009] Kim, B. E. and Gilbert, S. (2009). Detecting sadness in 140 characters: Sentiment analysis and mourning michael jackson on twitter. *Web Ecology*, 03(August):–.

[L. et al. 2012] L., R. S., U., K., and M., W. (2012). We have no feelings, we have emoticons ;-). *Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference on*, -(2):642 – 646.

[MacAskill 2009] MacAskill, E. (2009). US confirms it asked Twitter to stay open to help Iran protesters. `tinyurl.com/klv36p`.

[Pak and Paroubek 2010] Pak, A. and Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In Chair), N. C. C., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

[Pang and Lee 2008] Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

[Read 2005] Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *ACL Student Research Workshop*, ACLstudent '05, pages 43–48.

[Sakaki et al. 2010] Sakaki, T., Okazaki, M., and Matsuo, Y. (2010). Earthquake shakes twitter users: real-time event detection by social sensors. In *International conference on World wide web*, WWW '10, pages 851–860.

[SySomos] SySomos. Exploring the use of twitter around the world. `www.sysomos.com/insidetwitter/`. Accessed January 14, 2012.

[Wickre] Wickre, K. Celebrating twitter7. `http://blog.twitter.com/2013/03/celebrating-twitter7.html`. Accessed March 25, 2013.