

PANAS-t: Uma Escala Psicométrica para Medição de Sentimentos no Twitter

Pollyanna Gonçalves¹, Wellington Dores¹, Fabricio Benevenuto¹

¹Departamento de Computação – Universidade Federal de Ouro Preto (UFOP)
Ouro Preto – MG – Brasil

{pollyannaogoncalves, wellingtonjdores, benevenuto}@gmail.com

Abstract. *Twitter has become an important social communication mean where users post messages about everything. Certain messages express information about their author's emotional state, which may be useful in developing applications that predict emotional trends of a population or simply to better understand the effects of global and local phenomena in the mood of people. In this study, we adapted a psychometric scale known as the PANAS-X, commonly applied as a questionnaire, to measure the feelings of Twitter users in many social, political and sporting events. Our results suggest that the PANAS-t, our adapted version of the PANAS-X, correctly captures feelings for the events analyzed.*

Resumo. *Twitter tem se tornado um meio importante de comunicação social, onde usuários postam mensagens sobre tudo. Algumas dessas mensagens expressam informações sobre o estado emocional do usuário, o que pode ser útil no desenvolvimento de aplicações que prevêm tendências emotivas de uma população ou simplesmente para melhor entender os efeitos de fenômenos mundiais ou locais no humor das pessoas. Nesse trabalho, adaptamos uma escala psicométrica conhecida como PANAS-x, comumente aplicada em forma de questionário, para medir os sentimentos dos usuários do Twitter sobre uma série de eventos de tema social, político e esportivo. Nossos resultados sugerem que o PANAS-t, nossa versão adaptada do PANAS-x, captura corretamente sentimentos para os eventos analisados.*

1. Introdução

Recentemente, a rede social Twitter vem se tornando popular como uma plataforma que permite aos seus usuários a postagem de mensagens curtas sobre fatos do seu cotidiano, compartilhando informações em tempo real [Cha et al. 2012] sobre diversos eventos de escala global como tragédias e revoluções [MacAskill 2009, Sakaki et al. 2010, Gomide et al. 2011], na previsão de ataques terroristas [Tumasjan et al. 2010, Diakopoulos and Shamma 2010, O'Connor et al. 2010] e na compreensão de eleições políticas [Cheong and Lee 2011]. Estimativas recentes sugerem que 200 milhões dos usuários ativos no Twitter postam cerca de 150 milhões de *tweets* diariamente [Kirkpatrick 2011, Rao 2010], ou seja, se o Twitter fosse um país apenas esses usuários o classificariam como o quinto maior do mundo.

Como essas mensagens podem conter o humor de cada usuário, consideramos como hipótese que essas também são capazes de revelar o sentimento do autor em relação

a um determinado evento. Dessa forma, detectar sentimentos do público sobre eventos utilizando uma base de dados extraída do Twitter nos permitiria compreender como, quando e porque os sentimentos das pessoas oscilam em relação a algum evento. Além disso, essa medição permitiria uma melhor compreensão de características intrínsecas da natureza humana. Apesar de todas essas potenciais aplicações, existe um número limitado de esforços nessa direção, dos quais a grande maioria está focada na extração de opiniões e não de sentimentos [Pang and Lee 2008], ou não foram devidamente projetadas para esse contexto [Dodds and Danforth 2010].

Nesse artigo propomos o PANAS-t, uma escala psicométrica de medição de sentimentos adaptada para o contexto do Twitter que se baseia na versão expandida da conhecida escala PANAS-x (*Positive Affect Negative Affect Scale*). O PANAS-t consiste de sentimentos positivos e negativos, sendo adequado para a medição de qualquer evento aleatório no Twitter. Para isso foram utilizados dados empíricos a partir de um conjunto único de dados coletados contendo cerca de 1,8 bilhões de *tweets*. Utilizamos esses dados para computar valores de normalização para cada sentimento, de modo que o crescimento ou decréscimo de cada um pode ser medido na base de dados. Para validar nossa metodologia forneceremos uma análise sobre variações no humor dos usuários do Twitter em relação a eventos que abrangem tragédias, política, saúde, bem como eventos esportivos.

A nossa abordagem faz com que o PANAS-t seja bastante simples e prático de ser utilizado para grandes quantidades de dados e mesmo para análise em tempo real. Nossa avaliação quantitativa oferece evidências de que o PANAS-t captura corretamente os sentimentos esperados para os eventos analisados.

A seguir, fornecemos um breve histórico sobre metodologias existentes para medir sentimentos. E então detalhamos como o PANAS-t funciona e, além disso, fornecemos evidências experimentais que comprovam nossa abordagem. Finalmente, concluímos o artigo e oferecemos direções para trabalhos futuros.

2. Abordagens para Medição de Sentimentos

Com o crescimento das redes sociais na Web, a análise de sentimentos e mineração de opiniões se tornaram assunto de estudos para muitas pesquisas. Uma visão geral das abordagens e técnicas de processamento de linguagem natural que são utilizadas por pesquisadores na extração de sentimento foi apresentada no *survey* [Pang and Lee 2008]. Algumas abordagens consistem no uso de técnicas para o aprendizado de máquina. Embora aplicada em vários cenários, técnicas supervisionadas requerem uma intervenção manual para a pré-classificação dos dados de treinamento, o que pode se tornar inviável em alguns casos. A seguir, faremos um breve levantamento das diferentes técnicas já utilizadas para a medição de sentimentos e como elas foram aplicadas.

Uma estudo na área [Dodds and Danforth 2010] se define pela detecção de índices de felicidade de um texto. Os autores mostraram que o índice de felicidade em letras musicais possuem uma tendência decrescente desde os anos de 1960 a meados de 1990 porém o contrário foi observado em blogs, onde o índice de felicidade aumentou entre 2005 e 2009. A desvantagem desse tipo de abordagem está no proporcionamento de apenas uma escala para a medição do índice de felicidade, não fornecendo nenhuma categorização para o sentimento, o que é o foco do nosso trabalho.

Em sua pesquisa [Balog et al. 2006] coletaram dados da rede social LiveJournal,

que disponibiliza aos seus usuários um sistema de criação de blogs ou diário online para o compartilhamento de informações. O interessante desse sistema é que seus usuários também podem definir seu humor atual. Como resultado das análises, os autores concluíram que técnicas simples baseadas em comparação de frequências são eficazes na identificação de variações de humor de uma população em relação a eventos globais.

Dois estudos estão mais estreitamente relacionados aos nossos objetivos. [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 Michael Jackson foi associada com uma escala de três dimensões emocionais, são elas: prazer, excitação e dominância. No entanto, como podemos perceber, essa escala é capaz de capturar apenas sentimentos positivos que um usuário possa expressar.

Já o segundo estudo fez o uso da escala psicométrica *Profile of Mood States* (POMS) [Bollen et al. 2009]. 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.

Propomos aqui um método de detecção de sentimentos utilizando uma escala psicométrica que detecta até 11 escalas de sentimentos, capturando não apenas sentimentos positivos como também sentimentos negativos.

3. Medindo Sentimentos com o PANAS-t

Nossa abordagem de detecção e medição de sentimentos no Twitter consiste na adaptação de uma escala psicométrica bem conhecida, a *Positive Affect Negative Affect Scale* (PANAS). Começaremos descrevendo o PANAS-x, uma versão expandida do PANAS adaptada para o contexto da Web.

3.1. A Escala PANAS-x

A escala PANAS foi desenvolvida em 1998 por Watson [Watson and Clark 1985]. Indivíduos recebem um questionário contendo adjetivos que descrevem cada sentimento. Cada palavra desse questionário é então classificada ao seu sentimento correspondente, em uma escala de 1 a 5, de acordo com o que o entrevistado está sentindo. Mais especificamente, o PANAS-x, versão expandida da escala original, mede 11 afetos específicos: *medo, tristeza, culpa, hostilidade, timidez, fadiga, surpresa, jovialidade, auto-confiança, serenidade e atenção*.

A Tabela 1 sumariza os itens de composição da escala PANAS-x, que são utilizados no nosso trabalho para a análise de *tweets*. Antes de explicar como adaptamos o PANAS-x para o contexto do Twitter precisaremos primeiro descrever a nossa base de dados.

Tabela 1. Termos que compõem a escala PANAS-x.

Estado afetivo	Palavras relacionadas (em inglês)
Medo	afraid, scared, frightened, nervous, jittery, shaky.
Hostilidade	angry, hostile, irritable, scornful, disgusted, loathing.
Culpa	guilty, ashamed, blameworthy, angry at self, disgusted with self, dissatisfied with self.
Tristeza	sad, blue, downhearted, alone, lonely.
Jovialidade	happy, joyful, delighted, cheerful, excited, enthusiastic, lively, energetic.
Auto-confiança	proud, strong, confident, bold, daring, fearless.
Atenção	alert, attentive, concentrating, determined.
Timidez	shy, bashful, sheepish, timid.
Fadiga	sleepy, tired, sluggish, drowsy.
Serenidade	calm, relaxed, at ease.
Surpresa	amazed, surprised, astonished.

3.2. Base de Dados do Twitter

Os dados utilizados neste trabalho incluem um extenso conjunto de dados proveniente de um estudo de medição anterior que incluía um histórico completo de *tweets* postados por todos os usuários desde sua criação, em 2006, até Agosto de 2009 [Cha et al. 2010]. Nosso conjunto 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*. Na nossa análise ignoramos esse usuários.

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 a medição do aumento ou diminuição de um certo sentimento relacionado a *tweets* de um evento específico.

Devido a maior porcentagem de usuários cadastrados até 2009 no Twitter estarem localizados nos Estados Unidos [Kulshrestha et al.], focaremos nossas análises na medição de sentimentos apenas em *tweets* que estão na língua inglesa.

3.3. Limpeza e Parsing dos Dados

Afim de analisar apenas *tweets* que possivelmente expressam sentimentos individuais de um usuário, consideramos apenas *tweets* contendo as expressões em inglês "I'm", "I am", "I", "am", "feeling", "me", e "myself". No total, agrupamos 479.356.536 *tweets* que correspondiam a algum desses padrões, o que corresponde a cerca de 27% da nossa base total de *tweets*.

Após a filtragem dos dados, uma limpeza e parsing dos dados foram feitas seguindo as seguintes etapas: (i) Primeiramente aplicamos abordagens comuns de processamento de linguagem como radicalização, remoção de *stopwords*, vírgulas, traços e outros caracteres não alfa-numéricos, URLs e formas verbais comuns. (ii) Em seguida, separamos termos individuais utilizando espaços em branco como delimitadores. Por exemplo, para o *tweet* "I am so scare about swine flu" teríamos o conjunto de termos normalizados para: [i, am, scare, swine, flu]. No restante desse artigo utilizamos essa base de dados filtrada contendo apenas *tweets* normalizados.

3.4. Calculando o PANAS-t

Tweets que expressam certos sentimentos podem se tornar mais populares que outros. Assim, a fim de dizer se um sentimento cresceu ou decresceu em relação a um evento específico (ex. morte do Michael Jackson), precisaremos de uma base para comparação. Através da quantificação da fração de *tweets* que estão relacionados a cada sentimento, seremos capazes de dizer se um sentimento aumentou para um evento específico em relação a presença desse sentimento na base de dados inteira. Em seguida, apresentamos detalhadamente como calculamos essa base de comparação.

Cada *tweet* do conjunto de *tweets* normalizados foi mapeado a um único sentimento. Para isso, quando um adjetivo da Tabela 1 associado a um sentimento s aparece como um termo de um *tweet* normalizado nós consideraremos esse *tweet* com possuindo o sentimento s . Em casos de empate associaremos o *tweet* como possuindo o primeiro sentimento identificado. No entanto, casos de empate foram muito raros e portanto foram desprezíveis na quantificação. Com base nessa informação, definimos nossos níveis de base para cada sentimento, como segue.

Seja T o conjunto de *tweets* normalizados e T_s o subconjunto destes relacionados a um sentimento s . Assim, o valor base para cada sentimento, α_s , é definido como $\alpha_s = |T_s|/|T|$. Em outras palavras, nós dividimos a ocorrência de *tweets* de cada sentimento pelo número total de *tweets* normalizados na nossa base de dados. A Tabela 2 mostra os valores de base da ocorrência de cada sentimento considerado. Podemos notar que alguns sentimentos ocorrem em maior ordem de grandeza que outros. Por exemplo, *tweets* que expressam sentimento de *fadiga* ocorrem cerca de 32 vezes mais do que os que expressam *timidez*, o que reforça que uma normalização é necessária.

Tabela 2. Fração de *tweets* para cada sentimento na base de dados total.

Sentimento (s)	Linha de Base (α_s)
Medo	0.0063791
Tristeza	0.0086279
Culpa	0.0021756
Hostilidade	0.0018225
Timidez	0.0007608
Fadiga	0.0240757
Surpresa	0.0084612
Jovialidade	0.0182421
Auto-confiança	0.0036012
Atenção	0.0008997
Serenidade	0.0022914

Com base nesses valores, podemos calcular o aumento ou diminuição relativa de um sentimento em uma amostra particular de *tweets* como segue. Seja S um conjunto de *tweets* (por exemplo uma amostra de *tweets* relacionados a morte do Michael Jackson) e S_s o subconjunto desses *tweets* relacionados ao sentimento s . Definimos β_s como $\beta_s = |S_s|/|S|$. E então, definimos a pontuação do PANAS-t como um vetor de dimensão igual a 11, onde a função de pontuação $P(s)$ é calculado como abaixo:

$$P(s) = \begin{cases} (\alpha_s - \beta_s)/\alpha_s & \text{if } \beta_s \leq \alpha_s \\ -1(\beta_s - \alpha_s)/\beta_s & \text{if } \beta_s > \alpha_s \end{cases}$$

Baseados nessa estratégia, cada valor de $P(s)$ poderá variar de -1 a 1 para cada

Tabela 3. Índice de eventos analisados por data e palavras-chave relacionadas.

Nome	Data	Palavras-chaves	# Tweets
H1N1	09 - 16.06.2009	tamiflu, outbreak, antiviral, virus, influenza, pandemia, pandemic, flu, pig, mexico, h1n1, swine, World Health Organization.	35.559
AirFrance	01 - 06.06.2009	victims, passengers, A330, 447, crash, airplane, airfrance.	29.765
Obama	18 - 22.01.2009	barack obama, politics, democrats, election, white house, vote, speech, presidential, republicans, candidate, yes we can.	43.015
US-Elec	02 - 06.11.2008	clinton, biden, palin, voting, vote, elect, candidate, campaign, mccain, democrats, republicans, obama, bush.	185.477
MJ-death	25 - 30.06.2009	rip, mj, michael jackson, death, died, king of pop, overdose, drugs, heart attack, conrad murray, thriller.	56.259
Susan Boyle	11 - 16.04.2009	susan boyle, I dreamed a dream, britain's got talent, les miserables.	7.142
Olympics	06 - 26.08.2008	olympics, medals, china, beijing, sports, peking, sponsor.	12.815

sentimento. Ou seja, um evento com $P(medo) = 0$ significará que o evento não resultou em um aumento ou diminuição do sentimento *medo*, em comparação ao conjunto de dados total de *tweets* publicados até 2009. Em oposição, um valor de P positivo igual a 0,3 significaria um aumento de 30% e assim por diante.

Nossa estratégia para calcular a pontuação do PANAS-t é simples e adequada para permitir a comparação do aumento ou diminuição de cada sentimento de uma base de dados não tendenciosa. Mais importante, a Tabela 2 fornece uma linha de base para comparação de qualquer tipo de amostra de *tweets*. Por exemplo, pode-se facilmente rastrear uma amostra de mensagens postadas por usuários a partir da API do Twitter e utilizar desses valores para normalizar e obter pontuações de sentimentos.

4. Avaliação do PANAS-t

Com o objetivo de avaliarmos se o PANAS-t mede com precisão sentimentos dos usuários do Twitter, precisaríamos de uma verdade absoluta para comparar os resultados com nossos métodos. No entanto, é difícil de obter tal verdade absoluta se tratando de medição de sentimentos, que são subjetivos por natureza. Consideraremos portanto, uma estratégias para executar esta avaliação. Avaliaremos um conjunto de eventos populares para os quais acreditamos que sentimentos associados a eles sejam esperados ou fáceis de serem verificados. Em seguida, descrevemos essa análise.

4.1. Análise de Eventos Populares

Escolhemos sete eventos que foram amplamente noticiados por terem sido populares no Twitter¹. Estes eventos, resumidos na Tabela 3, abrangem tópicos relacionados com tragédias, lançamentos de produtos e filmes, política, saúde e eventos esportivos. Para extrair apenas *tweets* relevantes de cada evento identificamos o conjunto de palavras-chave que os descrevem em consultas a websites, blogs, na Wikipédia e informações individuais. Dada nossa lista de palavras-chave selecionadas, identificamos os eventos pesquisando por essas na nossa base de dados de *tweets*. Limitamos a duração de cada evento no Twitter pois palavras-chave populares são tipicamente utilizados por *spammers* após certo tempo [Benevenuto et al. 2010, Chhabra et al. 2011]. A Tabela 3 exhibe as palavras-chave utilizadas e o número total de *tweets* filtrados para cada evento considerado. A primeira coluna contém o nome que identifica o evento a que iremos nos referir no restante desse artigo.

¹Top Twitter trends <http://tinyurl.com/yb4965e>

Com o intuito de demonstrar que o PANAS-t pode medir variações de sentimentos compararemos os sete eventos descritos na Tabela 3 utilizando a representação de Kiviat. Em cada gráfico de Kiviat, linhas radiais a partir de um ponto central igual a -1 representa um sentimento com valor máximo de 1 [Jain 1991]. Na Figura 1 plotamos os onze sentimentos e cada sub figura dessa representa um evento específico.

Nosso primeiro caso de estudo é o dia em que a gripe suína (também denominada H1N1) foi declarada pela Organização Mundial da Saúde (OMS) como a primeira pandemia desde a gripe de Hong Kong, em 1968, causando grande preocupação em toda população mundial. Após o anúncio, a OMS lançou várias advertências e precauções que deveriam ser tomadas pelos governos e pelo público, levando a população mundial a um estado de alerta contra a doença. Como podemos visualizar na Figura 1(a), esse evento modificou o estado de humor dos usuários aumentando os sentimentos de *atenção* ($P(s) = 0,8774$) e *medo* ($P(s) = 0,6768$) nos dias posteriores ao anúncio. Na verdade, esses dois sentimentos correspondem aos sentimentos esperados mais prováveis para esse evento, já que as pessoas estavam atentas às precauções bem como com medo de uma epidemia mundial.

Em seguida, analisamos a queda de um avião em 1º de Julho de 2009, que causou uma grande comoção no Twitter. O voo comercial 447 da empresa aérea AirFrance estava programado para ir do Rio de Janeiro à Paris, mas caiu no oceano matando todos os 216 tripulantes. Após o acidente, podemos esperar que as pessoas se sentissem tristes pelas vítimas e também com medo de que esse tipo de acidente pudessem acontecer a elas. A Figura 1(b) mostra a representação Kiviat para esse evento. Como esperado, os sentimentos de *medo* ($P(s) = 0,7291$) e *tristeza* ($P(s) = 0,6992$) foram os dois mais predominantes nos *tweets* associados esse evento.

Quando se trata de política, as eleições Norte Americanas costumam ser eventos com amplo envolvimento de usuários da Web. Com a eleição, muitos eleitores podem se sentir apreensivos e empolgados com o poder de escolha que lhe são dados. Nossos resultados mostram sentimentos nessa direção. Como podemos ver na Figura 1(c), os usuários tiveram os sentimentos de *auto-confiança* ($P(s) = 0,6741$), *jovialidade* ($P(s) = 0,4277$) e *medo* ($P(s) = 0,3072$) aumentados.

Após as eleições, a posse de Barack Obama como presidente dos Estados Unidos foi amplamente esperada e discutida no Twitter. Como relatado na referência [HCD], a maioria dos norte-americanos estavam mais confiáveis na melhoria do país após o discurso do presidente Obama. Nossa análise de estado de ânimo dos usuários do Twitter, executada no dia do discurso do novo presidente, mostra um grande aumento nos sentimentos de *auto-confiança* ($P(s) = 0,7980$), seguido por *surpresa* ($P(s) = 0,5802$) e *jovialidade* ($P(s) = 0,5227$). Mas apesar de toda a manifestação positiva em relação ao novo governo, também podemos perceber um pequeno aumento no sentimento de *tristeza* ($P(s) = 0,1789$), o que poderia naturalmente representar *tweets* de opositores ao Barack Obama. A Figura 1(d) mostra que os sentimentos medidos com o PANAS-t estão de acordo com os relatos da referência [HCD].

O gráfico de Kiviat seguinte mostra o impacto da morte do cantor Michael Jackson. De acordo com o DailyMail [Bates] sete dos dez tópicos mais populares no Twitter no período foram dedicados a sua morte. Após a morte do cantor, muitas especulações

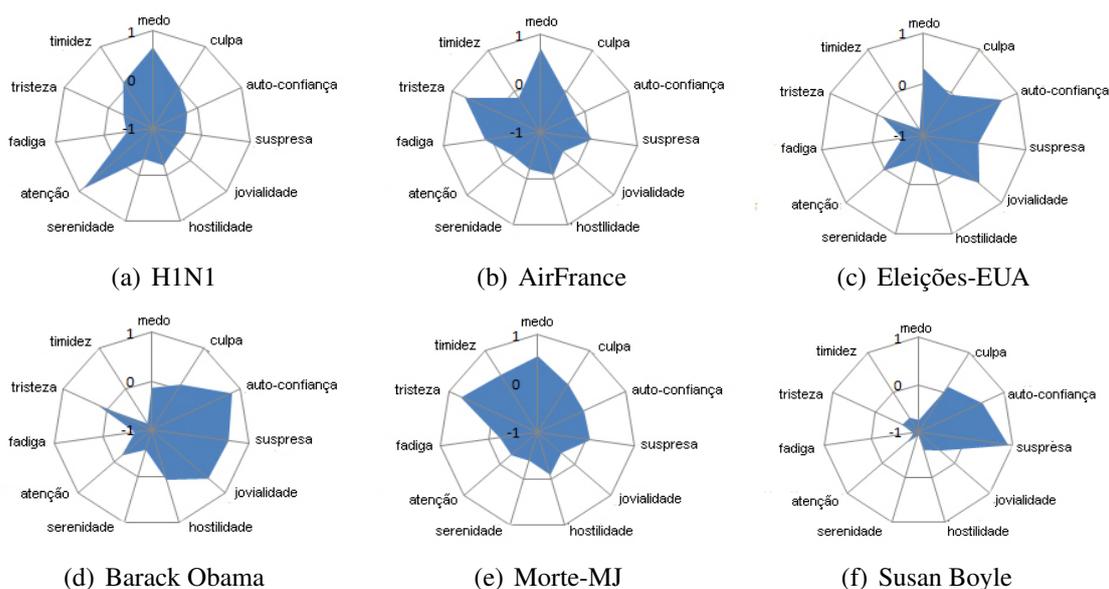


Figura 1. Eventos e sentimentos associados a cada um.

surgiram sobre quem ou o que matou Michael Jackson, fãs e críticos culpavam o estresse causado por paparazzis e pela mídia. Na Figura 1(e), podemos ver um aumento nos sentimentos de *tristeza* ($P(s) = 0,4055$), *medo* ($P(s) = 0,5676$), *timidez* ($P(s) = 0,4055$), *culpa* ($P(s) = 0,1616$) e *surpresa* ($P(s) = 0,0810$). É curioso perceber que, além dos sentimentos esperados associados a sua morte súbita como tristeza e medo, podemos notar a presença do sentimento de culpa. De fato, após especulações sobre a causa da sua morte é natural que alguns usuários do Twitter se sentissem com uma parte da culpa e assim expressaram em seus *tweets*.

Em seguida, analisamos o evento denotado por Susan Boyle. A aparição da Susan Boyle como uma concorrente de um programa de TV chamado *Britain's Got Talent* teve uma incrível repercussão na mídia. O interesse mundial foi desencadeado pelo contraste entre sua voz poderosa cantando "*I Dreamed a Dream*", da banda "*Les Misérables*", e sua aparência simples no palco. Seu contraste da primeira impressão dada aos telespectadores seguida de uma ovação durante e após sua performance, levou a um imediato viral que se espalhou através das redes sociais e obteve uma atenção enorme dos meios de comunicação globais. A Figura 1(f) mostra que os primeiros sentimentos expressos no Twitter associada a Susan Boyle foram de *surpresa* ($P(s) = 0,9066$) seguidos por *auto-confiança* ($P(s) = 0,4751$) e *culpa* ($P(s) = 0,1367$). Surpresa é definitivamente a sensação que fez com que sua aparição na TV se tornasse um hit na Internet. Esperávamos também que algumas pessoas se sentissem auto-confiantes ao serem encorajados ao ver uma mulher enfrentar com sucesso uma audiência que primeiramente a ironizou. Finalmente, o sentimento de culpa também era esperado levando em consideração que o evento se baseou em um pré-julgamento de valores com base na aparência.

O último evento analisado foi relativo as Olimpíadas de 2008, que ocorreram em Pequim, China. Durante os jogos Olímpicos, vários sentimentos poderiam ter sido expressos de acordo com diferentes jogos que estavam ocorrendo. Assim, ao invés de utilizarmos gráficos de Kiviat, fizemos a medição para cada sentimentos diariamente durante

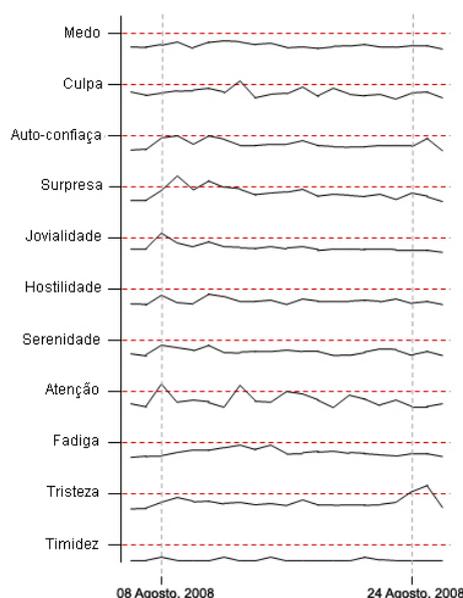


Figura 2. Gráfico de análise temporal do humor do público durante as Olimpíadas de Pequim, 2008.

todo o período do evento. A Figura 2 apresenta um gráfico de *sparklines* que mostra as variações de cada uma dos 11 sentimentos definidos pelo PANAS-t durante os Jogos Olímpicos de Pequim. Como podemos perceber, no dia em que os jogos começaram, em 08 de Agosto de 2008, tivemos predominância dos sentimentos de *surpresa* ($P(s) = 0,7024$), *atenção* ($P(s) = 0,4621$) e *jovialidade* ($P(s) = 0,3298$). No entanto, ao final do evento, no dia 24 do mesmo mês, podemos ver que esses sentimentos tiveram um decréscimo, enquanto o sentimento de *tristeza* aumentou de $P(s) = 0,1222$ para $P(s) = 0,5245$ no dia seguinte.

4.2. Comparando o PANAS-x com o PANAS-t

Embora tenhamos adaptado o PANAS-t a partir do PANAS-x, há duas diferenças fundamentais entre as duas escalas que poderiam afetar como os sentimentos são medidos. Primeiro, o PANAS-x fornece uma série de adjetivos que um entrevistado deve quantificar, já o PANAS-t colhe sentimentos espontâneos expressos no Twitter. Em segundo lugar, o PANAS-x necessita que o entrevistado exponha o que e o quanto ele está sentindo, enquanto o PANAS-t filtra sentimentos individuais sobre eventos específicos. Além disso, medir sentimentos no Twitter é particularmente desafiador pois usuários podem expressar sarcasmo ou até mesmo utilizar expressões que não são abrangidos pelo PANAS-t. Assim, com o intuito de quantificar possíveis incompatibilidades entre o PANAS-x original e o PANAS-t nós conduzimos o seguinte conjunto de experimentos.

Projetamos um serviço de *feedback* baseado na Web (disponível em <http://200.131.216.78/index.php>)², onde é mostrado a um avaliador uma versão orientada a eventos do tradicional PANAS-x de teste. Em vez de perguntar como o entrevistado está se sentindo, perguntamos aos voluntários como eles se sentiram sobre certos eventos. O mesmo conjunto de adjetivos utilizados no PANAS-x e no PANAS-t foram adicionados

²Convidamos o leitor a fazer o teste e nos ajudar a melhorar nossa avaliação.

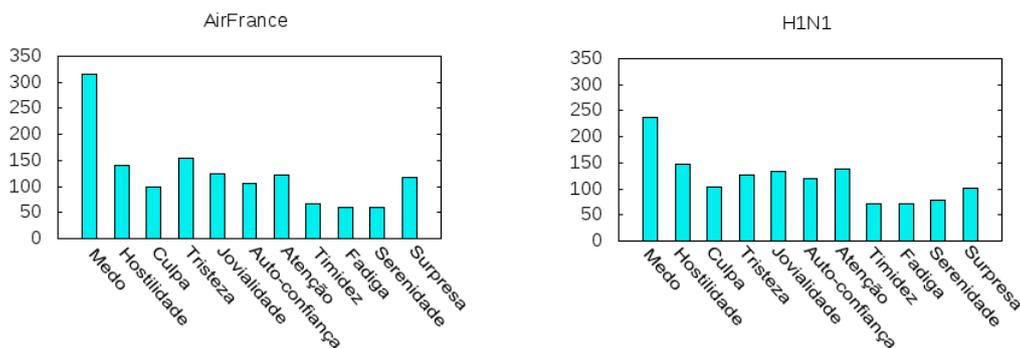


Figura 3. Resultados dos testes feitos com voluntários do PANAS-t para os eventos AirFrance e H1N1.

a ferramenta adicionando a eles uma variação de sentimentos em uma escala de 1 a 5, correspondente a força daquele sentimento.

Como a avaliação requer uma grande quantidade de trabalho manual, permitimos que um voluntário escolhesse entre apenas dois eventos: o acidente com o voo da AirFrance e o anúncio da gripe H1N1 como pandemia. Embora a maior parte dos voluntários avaliaram ambos os eventos, alguns escolheram apenas um deles para a avaliação. No total, foram obtidos 29 avaliações, dos quais 14 estão relacionados ao acidente da AirFrance e os outros 15 à gripe H1N1.

A Figura 3 mostra as pontuações obtidas para cada sentimento para o evento da queda do avião AirFrance. Podemos notar que os sentimentos de *medo* e *tristeza* foram os predominantes. Estes resultados estão em conformidade com o que esperávamos que as pessoas pudessem sentir. Similarmente, quando observamos os resultados obtidos para o evento do anúncio da gripe H1N1 como pandemia (Figura 3), podemos notar a prevalência do sentimento de *medo*, seguido de *hostilidade* e *atenção*. Assim, os sentimentos mais proeminentes expressos no Twitter para esses eventos coincidem com os sentimentos calculados com o método normal de aplicação do questionário do PANAS-x. Isto sugere que, embora o Twitter possa trazer alguns potenciais de predisposição à tendência quando utilizamos o PANAS-t, esse não é suficiente para afetar o sentimento geral dos usuários sobre os eventos analisados.

5. Conclusões e Trabalhos Futuros

Nesse artigo, apresentamos o PANAS-t, uma escala psicométrica que mede 11 sentimentos adaptada ao contexto do Twitter, e que se baseia na versão estendida da conhecida escala *Positive Affect Negative Affect Scale* (PANAS). Utilizando dados empíricos de um conjunto único de dados do Twitter contendo cerca de 1,8 bilhões de *tweets*, fomos capazes de calcular pontuações de normalização para cada sentimento.

Realizamos uma avaliação em 2 etapas, onde se compara o teste PANAS-t com o tradicional PANAS-x, e também aplicamos nossa escala adaptada em sete eventos que foram amplamente discutidos no Twitter afim de analisar qualitativamente os resultados obtidos. Acreditamos que as pontuações de normalização apresentadas na Tabela 2 permitem facilmente que qualquer um utilize o PANAS-t, tornando-o simples e prático para

ser utilizado em grandes quantidades de dados e mesmo para análises em tempo real.

Especulamos que esta escala de medição possa ser utilizada por quaisquer pesquisas com o objetivo de criar ferramentas que podem ser usadas por agências governamentais ou empresas interessadas em melhorar seus produtos utilizando redes sociais. De uma perspectiva da pesquisa, nosso método permite compreender como, quando e porquê os indivíduos sentem e como seus sentimentos variam de acordo com eventos sociais, políticos e econômicos.

Como trabalho futuro pretendemos combinar outras técnicas de aprendizagem de máquina para então dinamicamente incorporar sentimentos ao PANAS-t de acordo com o contexto.

5.1. Agradecimentos

Esse trabalho é apoiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e pela Fundação de Amparo à Pesquisa do estado de Minas Gerais (FAPEMIG). Também agradecemos a todos os voluntários que avaliaram nosso questionário de teste para o PANAS-t.

Referências

- Balog, K., Mishne, G., and Rijke, M. (2006). Rijke. why are they excited? identifying and explaining spikes in blog mood levels. In *In Proceedings 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL)*.
- Bates, C. How michael jackson's death shut down twitter, brought chaos to google...and 'killed off' jeff goldblum. <http://bit.ly/16e6eM>. Accessed January, 2012.
- Benevenuto, F., Magno, G., Rodrigues, T., and Almeida, V. (2010). Detecting spammers on twitter. In *Proceedings of the 7th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS)*.
- Bollen, J., Pepe, A., and Mao, H. (2009). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *Computer*, abs/0911.1:450–453.
- Cha, M., Benevenuto, F., Haddadi, H., and Gummadi, K. (2012). The world of connections and information flow in twitter. In *IEEE Transactions on Systems, Man and Cybernetics - Part A*.
- Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. P. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2010)*.
- Cheong, M. and Lee, V. C. (2011). A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via twitter. *Information Systems Frontiers*, 13:45–59.
- Chhabra, S., Aggarwal, A., Benevenuto, F., and Kumaraguru, P. (2011). Phi.sh/\$ocial: The phishing landscape through short urls. In *Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS)*.
- Diakopoulos, N. A. and Shamma, D. A. (2010). Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the 28th international conference on Human factors in computing systems, CHI '10*, pages 1195–1198.

- 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, 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)*.
- HCD. Confidence levels increase among democrats and independents, decrease among republicans after viewing obama’s press conference. <http://www.hcdi.net/news/MediacurvesRelease.cfm?M=276>. Accessed January 15, 2012.
- Jain, R. (1991). *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. John Wiley and Sons, INC, 1st edition.
- Kim, B. E. and Gilbert, S. (2009). Detecting sadness in 140 characters: Sentiment analysis and mourning michael jackson on twitter. *Web Ecology*, 03(August).
- Kirkpatrick, M. (2011). There Are Now 155m Tweets Posted Per Day, Triple the Number a Year Ago. <http://rww.to/gv4VqA>.
- Kulshrestha, J., Kooti, F., Nikraves, A., and Gummadi, K. P. Geographic Dissection of the Twitter Network. In *In Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- MacAskill, E. (2009). US confirms it asked Twitter to stay open to help Iran protesters. <http://tinyurl.com/klv36p>.
- O’Connor, B., Balasubramanian, R., Routledge, B. R., and Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.
- Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.
- Rao, L. (2010). Twitter Seeing 90 Million Tweets Per Day, 25 Percent Contain Links. <http://tinyurl.com/27x5cay>.
- Sakaki, T., Okazaki, M., and Matsuo, Y. (2010). Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web, WWW ’10*, pages 851–860.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *Word Journal Of The International Linguistic Association*, pages 178–185.
- Watson, D. and Clark, L. (1985). Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of Personality and Social Psychology*, 54(1):1063–1070.