

Como nos Sentimos: Uma Ferramenta de Mineração Visual de Sentimentos no Twitter

Artur O. Rodrigues¹, Raquel C. de Melo-Minardi¹, Wagner Meira Jr.¹

¹Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brazil

{artur, raquelcm, meira}@dcc.ufmg.br

Abstract. *Social network analysis and mining remains a sophisticated task performed by researchers of a few, specific research areas. In this work, we present Como nos Sentimos, a visual data mining tool of sentiments on Twitter, developed by the authors, with an underlying methodology that approximates the social network user to the discovery of patterns and trends in the information there produced. Using the tool, users can answer questions such as, in which Brazilian state people are more stressed; how the country or a region felt when a soccer team won a national championship; what is the predominant feeling on a sunny day.*

Resumo. *A análise e mineração de redes sociais permanece uma tarefa sofisticada realizada por pesquisadores de áreas bem definidas. Neste trabalho, apresentamos o Como nos Sentimos, uma ferramenta de mineração visual de sentimentos no Twitter, desenvolvida pelos autores, cuja metodologia aproxima o usuário da rede social e a descoberta de padrões e tendências nas informações ali produzidas. Através dela, é possível responder, por exemplo, em qual estado brasileiro os habitantes se sentem mais estressados; como o país ou uma região se sentiu quando um time de futebol ganhou o campeonato; qual o sentimento predominante em um dia de sol.*

1. Introdução

O advento do surgimento das redes sociais, especialmente os microblogs como o *Twitter*, que usam o modelo de *stream* de dados, vem apresentando interessantes desafios para a comunidade que analisa e estuda esses meios [Bifet and Frank 2010], como por exemplo os requisitos de se obter informação útil e interessante em tempo real a despeito das limitações de espaço e tempo. Nessas redes, uma aplicação que tem tido grande importância no contexto apresentado é a análise de sentimentos expressos nesses meios. Vários estudos, como em [Silva et al. 2011], têm sido desenvolvidos com o intuito de se propor, aperfeiçoar e avaliar modelos e algoritmos de mineração de dados para análise de sentimentos em ambientes sociais virtuais.

Mecanismos tradicionais de análise de dados não conseguem acompanhar a progressiva taxa com a qual informações de diversas naturezas inundam as mídias e redes sociais. Técnicas de mineração de dados visam a descoberta de conhecimento em volumosas bases e têm sido aplicadas com sucesso na busca de padrões e tendências e na sumarização desses grupos de dados. Parcela considerável das pesquisas na área foca no desenvolvimento de modelos e algoritmos eficientes e escaláveis. Contudo, grande

parte dessas técnicas gera como produto um massivo volume de informações que será analisado, em última instância, por um ser humano [Kosala and Blockeel 2000].

Por esse motivo, a mineração visual de dados é uma área de estudo que tem sido bem sucedida e que deve crescer em simbiose com a mineração de dados tradicional. Enquanto a última se preocupa fundamentalmente com questões relacionadas aos algoritmos e sua escalabilidade, a primeira facilita a percepção e cognição por parte do usuário, envolvendo-o no processo de análise e tornando a descoberta do conhecimento efetiva [Keim 2002]. É necessário, portanto, considerar as habilidades de percepção, assim como ter em vista as limitações da apresentação visual dos dados, de forma a viabilizar ao usuário a geração de *insights* e as interações diretas com os dados e conclusões. A abordagem visual é especialmente útil quando se sabe pouco sobre os dados e as metas de exploração são vagas. As técnicas devem ser simples e intuitivas, não apresentando barreiras adicionais ao usuário na análise do conjunto de informações [Wong 1999].

Nesse contexto, a mineração visual de dados se apresenta como uma alternativa para o estudo e compreensão de redes sociais. A primeira contribuição do presente trabalho é o desenvolvimento de uma metodologia de mineração visual de dados que reúne conceitos e padrões bem consolidados na literatura. Posteriormente, contribui-se ao aplicar essa metodologia através do *Como nos Sentimos*, uma ferramenta para análise de sentimentos expressos em Língua Portuguesa na rede social *Twitter* que são identificados por meio de expressões regulares. Finalmente, analisamos vários contextos nos quais a metodologia se mostrou interessante na identificação de tendências, padrões e exceções nas menções de sentimentos. Da associação entre mineração visual de dados e análise de redes sociais fomos capazes de atender as motivações feitas em [Wong 1999], respondendo, por exemplo, em qual estado brasileiro os habitantes sentem mais saudade; como o país ou uma região se sentiu quando um time de futebol ganhou o campeonato; qual o sentimento predominante em um dia de sol; ele é frequente também em dias chuvosos.

O restante deste trabalho é organizado da seguinte forma: Na Seção 2 apresentamos uma revisão de trabalhos relacionados já publicados; na Seção 3, os principais conceitos do modelo de visualização desenvolvido; na 4, mostramos uma ferramenta que representa a implementação do modelo proposto, assim como uma análise dos resultados obtidos com essa ferramenta. A Seção 5 traz a conclusão do trabalho.

2. Trabalhos Relacionados

O aumento do volume de dados opinativos disponível principalmente em redes sociais fomentou o desenvolvimento de técnicas que possibilitem a coleta, a mineração e a análise de sentimentos [Pang and Lee 2008]. Uma estratégia comumente adotada na análise de sentimentos nessas redes envolve a determinação da neutralidade ou polaridade de uma postagem e, no caso de serem polares, se são positivas ou negativas a respeito de um assunto particular [Wilson et al. 2005]. [Nasukawa and Yi 2003] extraem sentimentos associados à polaridades positivas e negativas sobre assuntos específicos e estudam como os sentimentos se relacionam semanticamente com o assunto discutido. [Whitelaw et al. 2005] propõem um método baseado na Teoria da Avaliação Cognitiva que usa *bag-of-words* na definição de grupos de avaliação que consistem em adjetivos e possíveis modificadores que auxiliam na determinação da polaridade das postagens. [Godbole et al. 2007], por sua vez, propõem um índice estatístico que

descreve a positividade ou negatividade de cada sentimento com relação a um assunto. [Bifet and Frank 2010] discutem os desafios de se minerar esses sentimentos em *stream* de dados do *Twitter* como, por exemplo, o desbalanceamento de classes e propõem uma abordagem baseada em janela deslizante com base na métrica Kappa para avaliação dos dados ao longo do tempo. [Silva et al. 2011] também se propõem a considerar a questão da obsolescência de termos passados através de um classificador baseado em regras de associação e uma solução de esquecimento baseada em uma janela de treinamento deslizante ativa. Finalmente, [Goncalves et al. 2012] apresentam uma escala psicométrica de medição de sentimentos adaptada para o contexto do *Twitter*.

Paralelamente, outros estudos têm sido desenvolvidos com o intuito de se propor, aperfeiçoar e avaliar estratégias de visualização de dados que potencializem as capacidades de percepção e cognição do ser humano. Em particular, tem sido descritas técnicas de mineração visual de dados [Keim 2002] e também meios de se visualizar as métricas de afetividade em documentos [Gregory et al. 2006]. Mais especificamente no que diz respeito a redes sociais, [Stefaner et al. 2012] capturam e visualizam a reação global sobre os eventos e atletas dos Jogos Olímpicos de Londres 2012 no *Twitter* em uma ferramenta de visualização *online*. [Kamvar and Harris 2011] apresentam uma ferramenta que une os conceitos de exploração visual de informação com técnicas de análise de sentimentos em *blogs* e páginas pessoais na internet.

Nossa solução difere dos esforços anteriores ao propor uma metodologia de análise visual de redes sociais aplicável a diferentes vocabulários que não somente sentimentos, colocando os *posts* e outras produções dessas redes como centro da análise. Além disso, o trabalho ainda apresenta uma ferramenta *online*, desenvolvida pelos autores, que utiliza o *Twitter* como fonte das mensagens, explorando visualmente, e de maneira inédita, textos expressos na Língua Portuguesa.

3. Metodologia

Neste trabalho, propomos uma metodologia para mineração visual de dados que é instanciada através do desenvolvimento de uma ferramenta para análise de sentimentos mencionados em português no microblog *Twitter*. Tal metodologia é composta de um conjunto de visualizações exploratórias e lúdicas, e outro, de visualizações direcionadas formadas por relacionamentos quantitativos, além de práticas, técnicas e padrões analíticos descritos adiante.

- **Item do vocabulário:** diz respeito à natureza dos itens observados. Nesse estudo, os sentimentos são os itens;
- **Tempo:** a variação no número de menções aos sentimentos ao longo do tempo;
- **Espaço geográfico:** distribuição das menções a diferentes sentimentos no espaço geográfico;
- **Outros:** refere-se neste trabalho à condição climática no momento da composição das mensagens.

3.1. Visualizações Exploratórias

Cada mensagem coletada deve ser individualizada e representada pelo sentimento ao qual se relaciona, formando o átomo das visualizações exploratórias e sendo representada visualmente como uma partícula (em forma de uma circunferência). Essas partículas ajudam a produzir cada uma das representações em questão. Através de sua simplicidade,

movimentos e cores pretende-se despertar o interesse do usuário, tornando mais fácil a exploração e interpretação dos dados que as partículas representam.

Propõem-se aqui cinco diferentes visualizações consideradas exploratórias, visto que não partem de uma questão pré-determinada. Tais visualizações devem ser apresentadas em uma tela de dimensões fixas onde as partículas devem estar sempre presentes. Os usuários devem ser capazes de ver a mensagem e o seu contexto na íntegra, através da interação por clique do mouse sobre as partículas. Sua translação, seguida da adição de elementos contextuais, permite perceber a mudança de estado entre as visualizações – coordenada por controles, como se segue:

- **Introdutória:** deve ser capaz de apresentar de forma intuitiva ao usuário os elementos básicos do estudo, ou seja, as mensagens com teor sentimental recuperadas do *Twitter*;
- **Frequência dos sentimentos:** encarregada de representar a frequência de cada um dos itens do vocabulário, no caso, sentimentos, de forma clara e direta;
- **Frequência dos itens do espaço geográfico:** deve representar a distribuição geográfica em termos absolutos para cada uma das unidades geográficas analisadas;
- **Mapa do espaço geográfico:** deve ser capaz de proporcionar a visão geral da dispersão das mensagens no espaço geográfico em questão;
- **Frequência dos itens de outras dimensões:** deve ser capaz de representar a dispersão categórica de itens de dimensões adicionais estudadas.

Cada uma dessas representações visuais procura oferecer um ponto de vista diferente acerca dos dados requisitados, proporcionando análises estatísticas básicas. É importante observar que o objetivo dessas visualizações é de serem lúdicas, simples, intuitivas e atraentes ao público em geral. Nesse sentido, os usuários devem ser capazes de ver a mensagem e o seu contexto na íntegra, através da interação por clique do mouse sobre a partícula. A transição entre as visualizações deve ser feita através de controles adicionais.

3.2. Visualizações Direcionadas

Em complemento às cinco visualizações baseadas em partículas, propomos uma série de representações visuais complementares que objetivam dar ao usuário uma visão estatística mais completa sobre a ocorrência dos sentimentos nas dimensões *tempo*, *espaço geográfico* e *condição climática*. Dessa forma, essas visualizações são consideradas direcionadas. Em linhas gerais, há dois tipos de relacionamentos quantitativos principais a serem visualizados nessa categoria: séries temporais e parte-todo. Uma série temporal consiste em um conjunto de séries de atributos dependentes do tempo como, o número de menções a um sentimento ao longo de um dia ou semana. Já um relacionamento do tipo parte-todo é utilizado quando desejamos explicitar a participação de uma parte em uma quantidade total como, por exemplo, a porcentagem de menções a um dado sentimento em condição climática chuvosa [Tuft 1983]. Para cada um desses relacionamentos há visualizações adequadas. Neste trabalho, usamos gráficos de linha para a representação e comparação entre séries temporais, e gráficos de barras e de barras empilhadas para representar como as partes se unem na formação do todo.

3.2.1. Padrões Analíticos

Essas visualizações, embora bastante tradicionais, têm a capacidade de revelar inúmeros padrões visuais. Alguns exemplos, não exaustivos, são:

- **Co-variação:** duas linhas em um gráfico de linhas podem mostrar se as menções a dois ou mais sentimentos se relacionam no tempo de maneira que mudanças em uma se refletem em mudanças na(s) outra(s).
- **Ciclos:** padrões que se repetem em intervalos regulares de tempo, sejam eles diário, semanal ou mensal. Os gráficos de linha mostram como os sentimentos são mencionados ao longo dos dias.
- **Exceções:** valores fora da norma ou que diferem extraordinariamente do restante podem ser facilmente identificados em gráficos de linhas. Podem também representar importantes eventos que ocorrem em determinada unidade geográfica e que despertam uma relevante comoção social.
- **Espalhamento:** reflete como os valores estão espalhados ao longo do tempo ou mesmo entre diferentes classes de atributos categóricos. Esse padrão pode revelar, por exemplo, se sentimentos são mencionados uniformemente ou se há concentrações de menções a um sentimento.

Ao analisar os padrões de expressões de sentimentos, verificamos que existe um comportamento cíclico bem definido no intervalo de um dia. Nesse caso, gráficos de radar são notadamente úteis para explicitar o padrão cíclico em séries temporais. Para tornar essa análise acessível ao público geral, visto que tais gráficos de radar podem parecer complexos à primeira vista, esses dados são apresentados por um gráfico icônico composto por barras dispostas de forma circular, similar ao gráfico de pétalas discutido em [Tan and Fraser 1998]. A metáfora da luminosidade das horas do dia para coloração das barras proporciona uma melhor interação e entendimento para as pessoas que as visualizam. Dessa maneira, meia-noite deve ser representada pela cor preta, enquanto meio-dia, pelo alaranjado.

3.2.2. Técnicas e Práticas Analíticas

- **Pequenos múltiplos:** para facilitar a comparação e o contraste entre os sentimentos, os gráficos de barras circulares devem ser apresentados segundo o padrão de projeto de pequenos múltiplos [Tufté 1990], sendo agrupados e ordenados pela similaridade entre sua dispersão. O padrão de pequenos múltiplos consiste em apresentar vários gráficos similares em tamanho reduzido, na mesma tela e próximos uns dos outros de forma que os valores e padrões possam ser comparados de maneira facilitada.
- **Detalhes sob demanda:** devem ser incorporados eventos de interação sob demanda [Shneiderman 1996] com os detalhes visuais, significando que, em todas as visualizações, o usuário seja capaz de obter detalhes adicionais ao passar o cursor do *mouse* sobre os objetos visuais. O usuário, deve ainda, ser capaz de clicar sobre as séries dos gráficos de linhas e barras e deletá-las da análise.
- **Linhas de referência:** trata-se de representar uma linha com o valor esperado ou medida de média, permitindo a comparação entre os valores atuais com os

esperados para uma dada métrica. Essa medida de referência deve ser apresentada como um linha de referência e posta sob a linha com os valores do dia corrente.

4. Implementação da Ferramenta *Como nos Sentimos*

Essa seção apresenta *Como nos Sentimos*¹, uma aplicação *online*, desenvolvida pelos autores, que implementa os conceitos discutidos na Seção 3. São introduzidos todos os componentes que permitem a produção dos elementos dessa ferramenta de mineração visual de dados.

4.1. Coleta

O passo inicial é a coleta ininterrupta de mensagens que contenham as expressões “eu tô”, “me sentindo”, “me sinto”, “estou”, “fiquei”, “fico” e “fiquei” no *Twitter*. A escolha dessas expressões foi feita empiricamente, ao se comprovar que esses fragmentos se mostraram presentes em grande parte das sentenças e textos com teor sentimental. Também são coletadas informações dos autores das mensagens como nome, apelido e localização.

4.2. Identificação do Sentimentos

Sentimento	RE	RGB
alegre	a+l+e+gr	255,164,001
culpado	(klc)u+l*pa+d	000,078,111
feliz	fe+l+i+(slz)	255,231,013
nervoso	ne+(rll)vo+(slz+)	255,176,003
triste	tri+st	004,063,105

Tabela 1. Excerto das expressões regulares dos sentimentos

Para cada uma das mensagens, o processador de identificação dos sentimentos irá tentar fazer a correspondência por uma das expressões regulares que denotam os sentimentos cadastrados. Essa relação de expressões regulares foi compilada manualmente e foi feita com a intenção de contemplar gírias, erros ortográficos e construções típicas do ambiente social virtual, que também são relevantes para o estudo realizado. Destarte, a expressão regular para o sentimento “feliz” identificará tanto “feliz”, quanto “felis” e até mesmo “feelliis”. Além disso, foram desconsiderados, quando possível, os sufixos e marcadores de gênero, permitindo que tanto “morta”, “morto” e “mortinha” fossem mapeados para o sentimento “morto”, por exemplo. Quando os termos de coleta são precedidos ou sucedidos por vocábulos que denotam negação, como “não” e “nada”, a mensagem é descartada. Juntamente com as expressões regulares e os sentimentos, é relacionado um valor de cor RGB que será utilizado para discriminar as entidades na visualização. É importante ressaltar que a identificação de sentimentos pode falhar em textos onde um sentimento é empregado como advérbio, ou simplesmente empregado de maneira irônica. Um excerto da relação das expressões regulares com os sentimentos e cores a que elas correspondem é encontrada na Tabela 1.

¹<http://www.sentimos.com.br/>

4.3. Tratamento da Localização e Condição Climática

As mensagens recuperadas do *Twitter* podem conter informações geográficas e os autores das mensagens podem inserir manualmente sua localidade. Essas informações são utilizadas para categorizar cada uma das mensagens no espaço geográfico analisado. Nessa ferramenta são consideradas as unidades federativas do Brasil.

Para que os dados coletados sejam armazenados de forma estruturada é realizada a geocodificação reversa da localização disponível. Isso torna o processo de mineração visual dessas informações mais direto quando a unidade federativa não está identificável de maneira estruturada na mensagem coletada. Quando é possível identificar a cidade dos autores das mensagens, é recuperada a condição climática dessa localidade, sendo essa a dimensão adicional da ferramenta no processo de mineração visual. Essas informações são obtidas através de caches locais e serviços públicos.

Quando todas essas dimensões são acrescentadas às mensagens, as informações coletadas são armazenadas em um banco de dados de forma a preservar o momento de sua composição e completando a dimensão temporal do processo de mineração visual. Essas informações estão prontas para serem recuperadas pela ferramenta de visualização.

4.4. Visualizações Exploratórias

As implementações de cada uma das representações visuais mostradas na Seção 3.1 são descritas a seguir. Em todas as visualizações, as mesmas mensagens, representadas por partículas coloridas, estão presentes. Assinale-se que, no início da ferramenta, são carregadas as 140 mensagens mais recentemente coletadas.

4.4.1. Introdutória

Nessa visualização, cada partícula se movimenta livremente na tela, introduzindo o usuário à ferramenta através de uma exibição visualmente interessante e lúdica, que representa exatamente a heterogeneidade dos sentimentos, das mensagens e dos autores que inundam o meio virtual com suas histórias e sentimentos.

Caso o usuário passe o *mouse* em alguma das partículas, ela tem seu raio aumentado, e é exibido ao seu lado um texto discreto que identifica o sentimento. Sendo assim, o usuário possui um recurso simples e objetivo de navegar pelas partículas que estão sendo exibidas. Na ocasião do clique, são exibidos na parte superior da tela e reservado em uma área, o texto no qual o sentimento foi identificado, o apelido e nome do autor, bem como sua localização. Essa visualização pode ser vista no sítio da ferramenta. Essa funcionalidade é exibida na Figura 1.

4.4.2. Frequência de Sentimentos

Como discutida anteriormente, essa visualização insere uma noção estatística às representações exploratórias. É apresentada uma espécie de gráfico de barras dos sentimentos, representando a frequência de cada um deles dentre a totalidade das partículas exibidas. Assim sendo, as partículas se orientam no sentido de formar as barras do gráfico.



Figura 1. Visualização introdutória

Cada um dos sentimentos é identificado por um texto, seguido de sua respectiva frequência. Na ocasião de existirem mais partículas para aquele sentimento do que o compartimento de uma sequência, é montada uma nova linha. Da mesma maneira, caso existam mais sentimentos do que comporta somente uma coluna, é montada uma nova coluna com os sentimentos menos frequentes como visto na Figura 2(a).

4.4.3. Mapa

Nesta visualização, Figura 2(b), é apresentado um mapa do Brasil, e as partículas se dirigem para seu estado de origem, explicitando a frequência de cada uma das unidades federativas. Essa representação possibilita a visão geral de como a internet é utilizada pela população brasileira: quais são as regiões do país que mais utilizam o *Twitter*, qual estado desponta em sua região e assim por diante.

4.4.4. Frequência de Estados e de Condições Climáticas

As visualizações Estados e Clima (Figuras 2(c) e 2(d), respectivamente) são semelhantes à visualização de frequência dos sentimentos, uma vez que são novamente apresentados gráficos de barras formados pela orientação das partículas. Partículas cujas mensagens não tiveram sua localização ou condição climática identificadas se direcionam para o canto inferior direito formando um ponto de interrogação. Novamente, as unidades indivisíveis da ferramenta (os textos) se combinam com o intuito de transmitir uma mensagem visual. Essas visualizações podem ser encontradas no sítio da ferramenta.

4.5. Visualizações Direcionadas

Como apresentado na Seção 3.2, as visualizações direcionadas transmitem de forma mais direta e completa os dados estatísticos da base de dados ao não se limitar às 140 mensa-

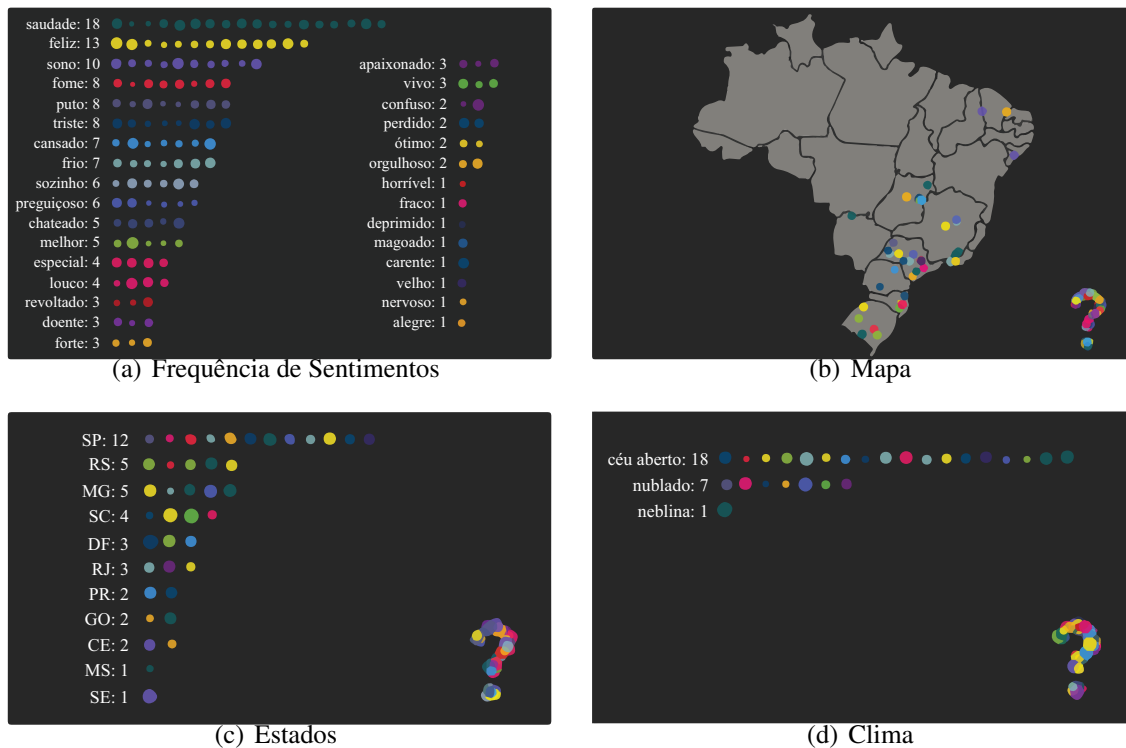


Figura 2. Visualizações exploratórias

gens mais recentes e por representarem dados que fogem à metáfora atômica da cada uma das mensagens.

A Figura 3 apresenta, através de um gráfico de linha, a participação dos cinco sentimentos mais frequentes nas 24 horas anteriores ao carregamento da ferramenta, nesse mesmo intervalo de tempo. A Figura 4, por sua vez, é obtida quando sentimentos específicos são selecionados através de um filtro disponibilizado que evidencia individualmente cada sentimento, acompanhado pela média histórica para o intervalo em questão.

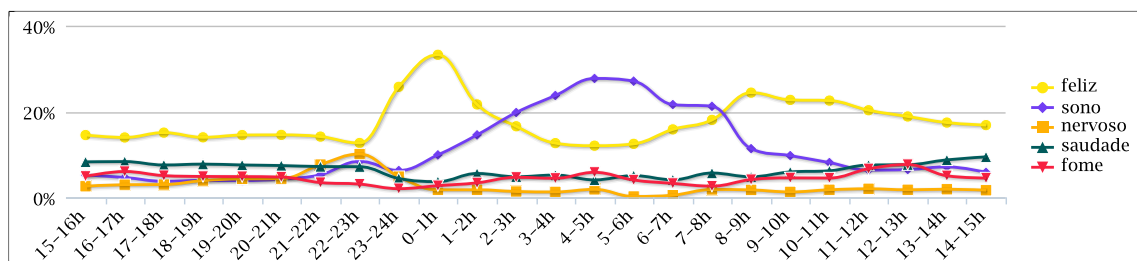


Figura 3. Cinco sentimentos mais frequentes - Gráfico recuperado às 15 horas do dia 5 de julho de 2012, um dia após a conquista da Taça Libertadores de Futebol (ocorrida entre 22 e 24 horas) pelo Sport Club Corinthians Paulista. Interessante observar os picos dos sentimentos “nervoso”, momentos antes da partida, e “feliz”, após o termino desta. Possível também observar a predominância do sentimento “sono” nas primeiras horas do dia.

Na Figura 5, tem-se o gráfico de colunas empilhadas para a representação parte-todo das condições climáticas para cada um dos sentimentos. As dispersões são agrupadas pelo algoritmo *K-Means* [Hartigan 1975] para que sentimentos com participações climá-

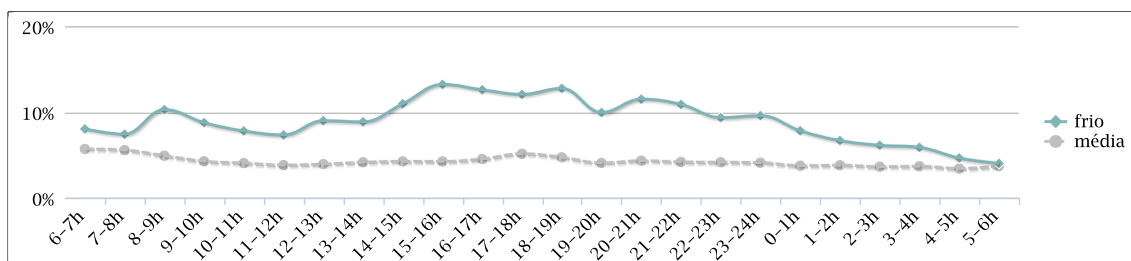


Figura 4. Gráfico de linha com média temporal - Sentimento frio entre 6 horas do dia 16 até às 6 horas do dia 17 de julho de 2012. Observar que nesse intervalo de 24 horas foi possível identificar que a população do Brasil sentiu mais frio que o normal, refletindo a baixa temperatura observada pelos sensores meteorológicos em todo o país nesse período.

ticas similares sejam apresentados próximos uns dos outros.

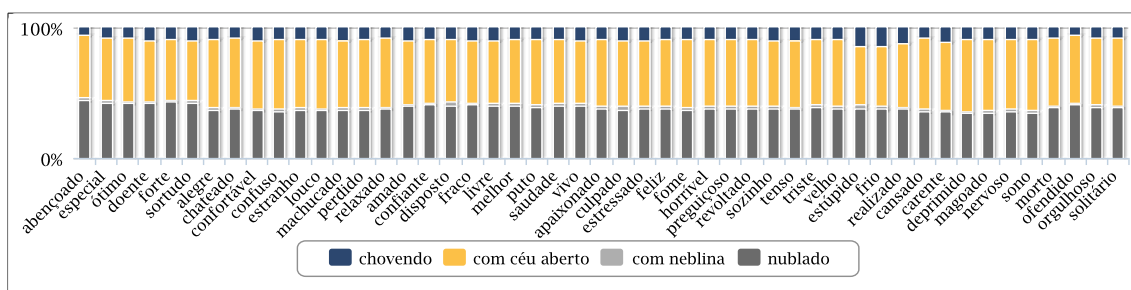


Figura 5. Gráfico de barras empilhadas das condições climáticas para cada um dos sentimentos - Vale notar que as condições climáticas parecem não influenciar significativamente na maneira como a população brasileira expressa sentimentos. Alguns sentimentos como “abençoado” e “estúpido” desviam ligeiramente da normalidade, além de “frio” que, por razões naturais, é mais expresso em condições chuvosas.

A Figura 6 apresenta alguns exemplos do gráfico de barras dispostas de maneira radial, o qual foi discutido na Seção 3.2.1. O resultado é bastante atraente, com barras que se assemelham a pétalas. Estas são maiores durante o período do dia quando o sentimento é mais citado, permitindo uma interessante interpretação e identificação diretas de tendências.

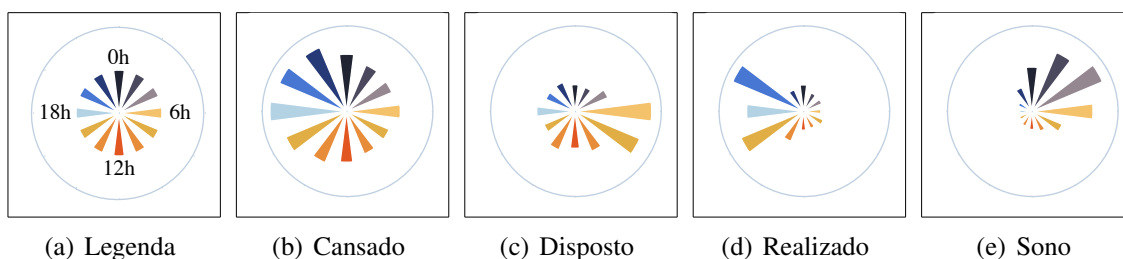


Figura 6. Dispersão dos sentimento “cansado”, “disposto”, “realizado” e “sono” durante as horas do dia - Interessante notar que, como esperado, a população se sente mais disposta nas primeiras horas da manhã, ao passo que o cansaço aumenta gradativamente com o fim do dia. O comportamento do sentimento “realizado”, muito frequente entre 16 e 22 horas, é um dos padrões interessantes que podem ser descobertos através desse tipo de visualização. Novamente, é possível comprovar a predominância do sono durante a madrugada.

4.5.1. Recursos de filtragem

Foi ainda criado um recurso de filtragem que permite que o usuário escolha quais sentimentos e estados exibir na ferramenta, realizando inúmeras combinações como, por exemplo, somente sentimentos “fome” e “relaxado” dos Estados “Rondônia” e “Santa Catarina”. Além disso, ele pode optar por carregar somente mensagens que foi capaz de identificar a localidade dos autores ou somente aquelas com condição climática associada.

5. Conclusão e Trabalhos Futuros

Apresentamos o *Como nos Sentimos*, uma ferramenta de mineração visual de sentimentos no *Twitter*, cuja metodologia é composta por dois conjuntos de visualizações, um exploratório e outro direcionado, considerando os sentimentos como dimensão, além do tempo, espaço geográfico e outros atributos contextuais. Essa ferramenta, disponibilizada publicamente, se mostrou capaz de identificar tendências e comportamentos típicos dos usuários da rede social, ao refletir, por exemplo, o nervosismo e felicidade usuais de torcedores de um time que sai vitorioso de uma final de campeonato. Seu caráter lúdico e acessível viabiliza a análise e mineração das redes sociais pelos seus próprios usuários.

Além da perspectiva de outros vocabulários para a metodologia, a ferramenta pode ser adaptada para considerar somente mensagens que façam menção a uma marca ou produto, bastando substituir os elementos buscados no *Twitter*, como “eu tô” e “me sentindo” por nomes e perfis dessas entidades comerciais. É importante notar que conclusões tiradas através de analisadores de sentimento tradicionais do tipo positivo/negativo também podem ser obtidas pela ferramenta criada, bastando classificar cada um dos sentimentos dessa maneira.

Como trabalhos futuros, pretendemos investigar as possibilidades supracitadas, além de inserir novas metáforas visuais e gráficos direcionais. Finalmente, o caráter visual da ferramenta torna a validação do estudo apresentado um desafio, tornando a avaliação com usuários um importante instrumento.

Referências

- Bifet, A. and Frank, E. (2010). Sentiment knowledge discovery in twitter streaming data. In *Proceedings of the 13th international conference on Discovery science, DS'10*, pages 1–15, Berlin, Heidelberg. Springer-Verlag.
- Godbole, N., Srinivasaiah, M., and Skiena, S. (2007). Large-scale sentiment analysis for news and blogs. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.
- Goncalves, P., Dores, W., and Benevenuto, F. (2012). Panas-t: Uma escala psicométrica para medição de sentimentos no twitter. *Brazilian Workshop on Social Network Analysis and Mining*.
- Gregory, M. L., Chinchor, N., Whitney, P., Carter, R., Hetzler, E., and Turner, A. (2006). User-directed sentiment analysis: visualizing the affective content of documents. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text, SST '06*, pages 23–30, Stroudsburg, PA, USA. Association for Computational Linguistics.

- Hartigan, J. A. (1975). *Clustering Algorithms*. John Wiley & Sons, Inc., New York, NY, USA, 99th edition.
- Kamvar, S. D. and Harris, J. (2011). We feel fine and searching the emotional web. In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, pages 117–126, New York, NY, USA. ACM.
- Keim, D. (2002). Information visualization and visual data mining. *Visualization and Computer Graphics, IEEE Transactions on*, 8(1):1–8.
- Kosala, R. and Blockeel, H. (2000). Web mining research: a survey. *SIGKDD Explor. Newsl.*, 2(1):1–15.
- Nasukawa, T. and Yi, J. (2003). Sentiment analysis: capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, pages 70–77.
- Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135.
- Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, VL '96, pages 336–, Washington, DC, USA. IEEE Computer Society.
- Silva, I., Barbosa, G., Veloso, A., Ferreira, R., and Meira, J. W. (2011). Analise adaptativa de fluxo de sentimento baseada em janela deslizante ativa. In *Proceedings of the Brazilian Symposium on Databases*.
- Stefaner, M., Hemment, D., and NAND, S. (2012). emoto project | 2012 @ONLINE. <http://www.emoto2012.org/>.
- Tan, Y. S. and Fraser, N. M. (1998). The modified star graph and the petal diagram: two new visual aids for discrete alternative multicriteria decision making. *Journal of Multi-Criteria Decision Analysis*, 7(1):20–33.
- Tufte, E. (1983). *The visual display of quantitative information*. Number v. 914 in The Visual Display of Quantitative Information. Graphics Press.
- Tufte, E. (1990). *Envisioning information*. Graphics Press.
- Whitelaw, C., Garg, N., and Argamon, S. (2005). Using appraisal groups for sentiment analysis. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 625–631.
- Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354.
- Wong, P. C. (1999). Guest editor's introduction: Visual data mining. *IEEE Computer Graphics and Applications*, 19:20–21.