

Uma Análise da Polaridade Expressa nas Manchetes de Notícias Brasileiras

Pedro Ramos, Julio Reis, Fabrício Benevenuto

¹Universidade Federal de Minas Gerais (UFMG) – Brasil

{pedroramos, julio.reis, fabricio}@dcc.ufmg.br

Abstract. *In the last years, with the growth of the internet as a means of communication, a large number of people changed the way they consume news, replacing the traditional newspapers and magazines with the virtual and online versions. In order to lure people's attention and attract their clicks, news websites have to create effective strategies. Considering this, we investigate the possible strategies used by Brazilian news websites in the design of their headlines. Specifically, we analyze the number of 59.510 news produced by 8 different and important Brazilian newspapers over a 3-month period in 2015. To find the strategies that could be used to attract clicks, we extracted news text features with the polarity within the headlines. We also present a methodology to evaluate the headline polarity's strength over time. Our results reveal a number of interesting observations on polarity within the news and, in addition, we identified strong evidence that the polarity of the headlines is impacted by the occurrence of specific events.*

Resumo. *Nos últimos anos, com o crescimento da internet como meio de comunicação, um grande número de pessoas mudou a maneira de consumir notícias, substituindo os tradicionais jornais e revistas impressos por suas versões virtuais online e/ou web blogs. Sites de notícias têm que criar estratégias eficazes para chamar a atenção das pessoas e atrair seus cliques. Considerando isso, neste trabalho investigamos as possíveis estratégias utilizadas por jornais de notícias online na concepção de suas manchetes. Especificamente, analisamos o conteúdo de 59.510 notícias produzidas por 8 importantes e diferentes jornais brasileiros em um período de aproximadamente 3 meses do ano de 2015. A fim de descobrir as estratégias que poderiam ser usadas para atrair cliques, foram extraídas características do texto das notícias relacionadas com a polaridade expressa nas manchetes. Apresentamos ainda uma metodologia para avaliar a força e o teor da polaridade contida nas manchetes ao longo do tempo. Nossos resultados revelam várias observações interessantes sobre a polaridade expressa em notícias e, além disso, identificamos fortes evidências de que a polaridade das manchetes é impactada pela ocorrência de eventos específicos.*

1. Introdução

Nos últimos anos, com o crescimento da internet como meio de comunicação, um grande número de pessoas mudou a maneira de consumir notícias, substituindo os tradicionais jornais e revistas impressos por suas versões virtuais online e/ou *web blogs* [Mitchelstein and Boczkowski 2009, Tewksbury 2005].

Atualmente, as notícias online têm sido responsáveis pela captação ou atração de uma quantidade significativa de usuários na Internet [Tatar et al. 2014]. A facilidade e o baixo custo para criação e compartilhamento deste tipo de conteúdo online mudou, inclusive, as regras tradicionais de competição pela atenção do público. É comum que as grandes fontes de notícias concentrem parte de sua atenção em mídias online, onde podem divulgar suas notícias de forma eficaz e de forma mais acessível para uma grande parte da população [Bandari et al. 2012].

Neste contexto, o ambiente de jornais online tem se tornado bastante competitivo. Considerando que as pessoas estão dispostas a disponibilizarem um tempo limitado para o consumo de notícias [Varian 2010], é fundamental que esses jornais criem esquemas eficazes capazes de atrair a atenção das pessoas e conquistarem seus cliques. Entretanto, um aspecto ainda não totalmente compreendido está relacionado às estratégias adotadas por estas empresas para tornarem as notícias, e consequentemente os jornais online, mais atrativos para os usuários.

Uma das características fundamentais relacionadas com esta divulgação de notícias e sua atratividade é a manchete que a descreve. O objetivo da manchete é de forma breve e rápida chamar a atenção do leitor para a história que será apresentada. Este recurso tem sido utilizado há vários anos como mecanismo para atração dos leitores [Evans 1974].

Potencialmente as manchetes podem determinar a forma com que muitas pessoas leem as notícias [Konnikova 2016], e com o grande aumento desta ampla concorrência no mundo online, não é surpreendente nos depararmos com manchetes cada vez mais agressivas, exageradas, e em alguns casos até um pouco enganadoras. Entretanto, as manchetes não são apenas a primeira impressão que adquirimos em relação à determinada notícia. Esforços recentes sugerem que elas podem até mesmo determinar a forma de percepção dos leitores em relação ao restante do conteúdo associado, afetando inclusive a maneira como as pessoas vão se lembrar destas notícias [Ecker et al. 2014]. Apesar da importância absoluta das manchetes, pouco se sabe sobre estratégias envolvendo o seu uso durante a produção e consumo de notícias online. Acreditamos que uma melhor compreensão deste ambiente nos permita realizar observações com inúmeras aplicações para a concepção e adequação de sistemas de apoio ao usuário neste contexto. E é sobre esta lacuna que firma-se o objetivo deste trabalho. Nesta pesquisa damos um primeiro passo nesta direção ao realizarmos uma caracterização de manchetes em notícias online.

Para isso, propomos uma metodologia experimental baseada em análise de sentimentos como forma de capturar a força das polaridades expressas nas manchetes de notícias (ou seja, se elas são exageradamente negativas ou positivas, ou se elas são neutras). Mais especificamente, propomos uma análise da polaridade das notícias produzidas pelos jornais online brasileiros para respondermos as seguintes questões: Será que existe uma tendência em relação à quantidade de notícias de polaridade positiva, neutra ou negativa produzidas? Será que a polaridade das notícias publicadas é afetada pela ocorrência de algum evento específico?

Para respondermos a estas questões, coletamos e analisamos o conteúdo de 59.510 notícias produzidas por 8 importantes e diferentes jornais brasileiros durante um período de aproximadamente 3 meses do ano de 2015. Nossos resultados apresentam descobertas interessantes. Mostramos que existe um padrão em relação à polaridade de notícias produzidas pelos jornais brasileiros. Apresentamos ainda que a polaridade das manchetes é

impactada pela ocorrência de alguns eventos específicos. Além disso, apresentamos uma estratégia para avaliar a força e o teor da polaridade contida nas manchetes ao longo do tempo. Conforme discutiremos mais tarde, estas observações têm inúmeras implicações para a concepção de sistemas que apoiem o usuário neste contexto. Em última análise, nossos resultados contribuem para a compreensão da comunicação no ambiente contemporâneo, bem como da sociedade em que vivemos.

O restante desse artigo está organizado da seguinte forma: a seguir, descrevemos os trabalhos relacionados. Depois, descrevemos a nossa metodologia experimental proposta e a base de dados utilizada para o desenvolvimento do trabalho. Em seguida, apresentamos os resultados obtidos. Por fim, concluímos este trabalho e apresentamos direções para trabalhos futuros.

2. Trabalhos Relacionados

Jornais têm estabelecido portais de notícias digitais para fornecerem conteúdo público. Esses portais têm sido bastante utilizados e conseqüentemente, atraído cada vez mais e mais visitantes [Esiyok et al. 2014]. Logo, este ambiente tem sido alvo de estudos em diversos domínios, incluindo ciência da computação.

Recentemente, pesquisadores realizaram um experimento cujo objetivo era examinar o efeito de manchetes enganosas [Ecker et al. 2014]. Para realização do referido estudo, as notícias (factuais ou de opinião) foram apresentadas aos participantes, mas com diferentes manchetes. Os autores concluíram que as manchetes enganosas podem afetar a memória, o raciocínio e intenções comportamentais dos leitores. Eles argumentam que esses efeitos surgem não somente pelo fato das manchetes restringirem mais o processamento das informações e a polarização de leitores no sentido de uma interpretação específica, mas também porque os leitores lutam para atualizar sua memória, com o propósito de que sejam corrigidos equívocos iniciais. Estes esforços destacam a importância das manchetes, que vão além de atrair usuários para a leitura, podendo inclusive até alterar a percepção ou atitudes do indivíduo em relação ao conteúdo, o que motiva a realização de um estudo que se dedique exclusivamente à caracterização de manchetes.

Outro trabalho investigou, por meio de questionários, se a ordem de leitura das notícias, considerando conteúdo negativo ou positivo, apresentava impactos comportamentais no usuário [Nguyen et al. 2011]. Os resultados indicaram que usuários que primeiramente leram más notícias relataram maior intenção para a mudança de comportamento, o que não aconteceu para usuários que primeiramente leram boas notícias. Esse estudo mostra que a polaridade do conteúdo lido apresenta impacto sobre o comportamento das pessoas. Um estudo mais recente mostra que, inclusive, o usuário percebe que exista uma relação entre a polaridade do conteúdo lido e o seu humor [Reis et al. 2014a], o que intensifica a relevância na realização deste estudo a partir da polaridade expressa em manchetes.

Uma análise da polaridade expressa em manchetes de notícias no idioma inglês é apresentada em [Neves et al. 2013]. O trabalho, cujo objetivo era visualizar e detectar eventos positivos e negativos ao longo do tempo a partir da observação de notícias, analisou manchetes de diferentes fontes de jornais publicadas durante os anos de 2008 até 2013, com a utilização da plataforma AlchemyAPI¹. Os autores propuseram uma análise da tendência geral da polaridade das manchetes e desenvolveram uma ferramenta para

¹<http://www.alchemyapi.com/>

visualização do resultado do experimento (disponível em https://github.com/kelmerp/headline_sentiment_rating). O projeto, que foi desenvolvido em Chicago, visou estudar a correlação entre época do ano e sentimento das notícias, em especial, a correlação entre o atual momento político e social e os sentimentos das notícias nos diferentes jornais online.

Ainda neste contexto, em [Reis et al. 2015a] os autores investigaram estratégias utilizadas por organizações de notícias online para projetarem suas manchetes. Eles coletaram e analisaram o conteúdo de 69.907 manchetes produzidas por quatro grandes empresas de mídia conhecidas mundialmente durante um período mínimo de oito meses consecutivos do ano de 2014. Os resultados obtidos mostram que a maioria das manchetes de notícias produzidas possui polaridade negativa, e que esta situação é constante ao longo do tempo. Apresentaram ainda que manchetes extremamente negativas ou positivas geram uma maior popularidade, enquanto manchetes neutras tendem a ser menos atraentes. Uma análise sobre os comentários dos usuários mostrou que os mesmos tendem a ser mais negativos, independentemente da polaridade (negativa, neutra ou positiva) associada à manchete da notícia.

Nossa pesquisa em relação às notícias é complementar aos estudos existentes, entretanto, colocamos em perspectiva a análise do contexto brasileiro, para o qual não identificamos trabalhos que o explorem. Entretanto, este objetivo traz consigo um grande desafio que está relacionado à extração de sentimentos de um texto no idioma português. Este aspecto tem sido alvo de várias pesquisas [dos Santos et al. 2014, Reis et al. 2015b], uma vez que grande parte dos métodos de análise de sentimentos estão disponíveis apenas para o idioma inglês.

Na próxima seção apresentamos a metodologia experimental proposta para este trabalho, que inclui o processo de coleta da base de dados e a nossa estratégia adotada para inferência da polaridade expressa em uma manchete.

3. Metodologia experimental

3.1. Base de Dados

Para determinar as fontes de coleta das notícias no contexto brasileiro, tomamos como base a lista dos maiores jornais em circulação impressa e digital disponibilizada pela *Associação Nacional dos Jornais (ANJ)*². Consideramos os jornais que disponibilizam o conteúdo de notícias via *Feed RSS*. Esse mecanismo funciona como um facilitador no oferecimento do conteúdo web (ou resumos deste conteúdo) juntamente com links para as versões completas e outros metadados, permitindo ao mecanismo utilizador se manter informado sobre as atualizações efetuadas em diversos sites sem precisar visitá-los um a um. A partir disto, coletamos e utilizamos dados oriundos de 8 importantes e distintos jornais online brasileiros: G1 (O Globo) (<http://g1.globo.com/>), R7 (<http://www.r7.com/>), Yahoo Notícias (<https://br.noticias.yahoo.com/>), UOL (<http://www.uol.com.br/>), A Tarde (<http://atarde.uol.com.br/>), O Tempo (<http://www.otempo.com.br/>), Estado de Minas (<http://www.em.com.br/>) e Folha de São Paulo (<http://www.folha.uol.com.br/>).

A coleta foi realizada por meio de um *web crawler* executado durante o período de 06 de Setembro de 2015 até 30 de Novembro de 2015. Para cada uma das páginas web

²<http://www.anj.org.br/>

recuperadas, foi realizado um *parsing* do conteúdo retornado e recuperadas informações relativas à cada uma das notícias. A Tabela 1 sumariza informações relacionadas ao conjunto de dados coletado.

Tabela 1. Sumário de informações do conjunto de dados

Jornal	Total	Média Diária
A Tarde	658	8
Estado de Minas	8.687	101
Folha	7.228	84
G1	13.664	159
O Tempo	3.984	46
R7	8.325	97
UOL	7.707	90
Yahoo	9.257	108
Total	59.510	692

3.2. Inferência da polaridade

Um dos principais desafios deste trabalho consiste em definir o método de análise de sentimentos a ser utilizado para realização da pesquisa. Os métodos para análise de sentimentos existentes estão focados principalmente sobre a análise de texto em inglês, e o aspecto multilíngue tem sido bastante analisado em pesquisas recentes [Araújo et al. 2016]. Logo, ao invés de utilizarmos, por exemplo, um método que seja mais popular, identificamos a necessidade de realização de uma análise específica sobre o método mais adequado para utilização no contexto em questão.

Inicialmente criamos um subconjunto de 100 notícias aleatórias com número proporcional de notícias de cada uma das fontes coletadas. Cada uma das manchetes foi rotulada manualmente por 3 voluntários. Cada registro foi analisado por todos os voluntários com objetivo de que fossem evitados empates e alcançada uma classificação mais consistente e robusta baseada na maioria de votos, com análise do grau de concordância. Para cada manchete, foi solicitado aos voluntários que rotulassem o seu conteúdo como negativo, neutro ou positivo. A percentagem de concordância entre os voluntários foi de 76%. O resultado da classificação foi: 36% de manchetes rotuladas como positivas, 27% neutras e 37% negativas. Com esta base rotulada foi possível analisar a capacidade de predição de cada um dos métodos para o contexto proposto.

O processo de escolha do método de análise de sentimentos a ser utilizado foi baseado em [Gonçalves et al. 2013], que aponta para a não existência de um método de análise de sentimentos que seja melhor em todos os cenários. A partir daí utilizamos este conjunto de dados rotulados e comparamos o desempenho dos métodos de análise de sentimentos disponibilizados considerando 2 situações: (i) execução de um método de análise de sentimentos desenvolvido especificamente para o idioma português; (ii) tradução do texto de entrada para o idioma inglês e execução dos métodos nativos para o respectivo idioma [Araújo et al. 2016].

Nós utilizamos o conjunto de dados rotulado como *baseline* para compararmos o desempenho das estratégias de análise de sentimento. No caso (i) utilizamos a versão do método SentiStrength [Thelwall et al. 2010] desenvolvida e disponibilizada pelo autor em <http://sentistrength.wlv.ac.uk/>. Já no caso (ii) utilizamos a *Microsoft Translator API*³ para a tradução das manchetes para o idioma inglês. A escolha da desta

³<https://www.microsoft.com/en-us/translator/translatorapi.aspx>

API deve-se ao fato dela ter sido avaliada, num estudo comparativo, como a *API* mais eficiente de tradução de máquina disponível⁴, além de sua disponibilidade para utilização de forma gratuita. Em seguida, submetemos as manchetes ao iFeel [Araújo et al. 2014, Araújo et al. 2016], uma ferramenta online que disponibiliza a implementação e execução automática de vários métodos de análise de sentimento.

O método SentiStrength [Thelwall et al. 2010] específico para o idioma português (caso (i)) foi o método que apresentou melhor desempenho neste contexto, com uma acurácia de 64%, considerando as 3 classes (negativa, neutra e positiva). Outros métodos apresentaram resultados competitivos com o SentiStrength, entretanto uma propriedade atraente deste método é o formato de saída escalar, que proporciona a medição da força da polaridade em uma escala que varia de -4 (muito negativa) a 4 (muito positiva). Consideramos que isto fornece uma perspectiva interessante da intensidade da polaridade expressa na manchete. A Tabela 2 apresenta exemplos de títulos ou manchetes classificados de acordo com o SentiStrength.

Tabela 2. Exemplo de manchete por pontuação de sentimento

Força da Polaridade	Manchete
-4	Júri absolve policiais “Highlanders” acusados de decapitar deficiente em SP
-3	Acidente com ônibus deixa ao menos 15 mortos em estrada de Paraty (RJ)
-2	Lobista preso na Lava Jato nega propina e diz não conhecer Cunha
-1	ONU denuncia tchecos por prisão sistemática de migrantes e refugiados
0	Comissão bilateral Cuba-EUA se reúne em Havana
1	Ex-camelô vira empresário de sucesso e fatura R\$ 12 mil com palestras
2	Escola realiza “aulão” de redação do Enem de graça, em Porto Velho
3	Aprenda a fazer um delicioso pirulito de brigadeiro para a criançada
4	SP recebe sírios “com muito orgulho e muito prazer”, diz Haddad

Nas próximas seções apresentaremos os resultados das análises efetuadas com base nas estratégias definidas nesta etapa do trabalho, incluindo uma discussão das possíveis implicações dos resultados.

4. Resultados

4.1. Polaridade das Manchetes

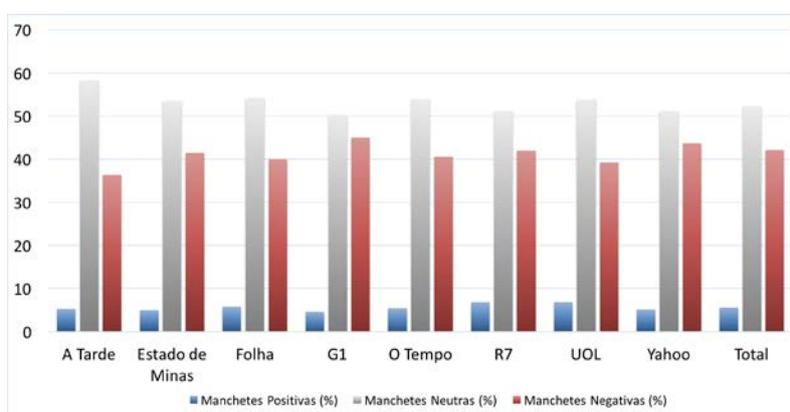
Inicialmente, medimos a força da polaridade de todas as manchetes coletadas. O resultado desta análise inicial é apresentado na Tabela 3. Observamos que o número de notícias no contexto brasileiro classificadas com força de polaridade extremas, ou seja, -4 e +4, é bastante inferior em comparação ao número de notícias classificadas com outras forças de polaridade. Notamos ainda que a maior concentração é de notícias classificadas com polaridade 0 (neutra).

Em seguida analisamos o agrupamento das notícias em 3 classes: negativa, neutra e positiva. O resultado desta análise é apresentado na Figura 1. Notamos uma tendência em todas as oito fontes de notícias analisadas: o número de manchetes neutras é o maior, seguido do número de notícias negativas e positivas. Particularmente, O jornal *GI* apresentou maior quantidade relativa de notícias negativas (45,06%), enquanto o jornal *A Tarde* conteve 36,32% de notícias negativas, que corresponde ao menor valor observado. O mesmo jornal *A Tarde* foi o que deteve a maior proporção de notícias neutras (58,36%). Percebemos ainda que a quantidade de notícias positivas publicadas está abaixo de 7%

⁴<https://www.neno-translate.com/en/blog/48-best-machine-translation-api>

Tabela 3. Percentagem de manchetes por força de polaridade

Jornal Online	Força de Polaridade								
	Negativa				Neutra	Positiva			
	-4	-3	-2	-1	0	1	2	3	4
A Tarde	36.32%				58.36%	5.32%			
	0.00%	1.67%	7.29%	27.36%		3.80%	1.52%	0.00%	0.00%
Estado de Minas	41.44%				53.57%	4.98%			
	0.02%	1.50%	8.02%	31.90%		4.03%	0.90%	0.03%	0.02%
Folha	39.98%				54.21%	5.81%			
	0.03%	1.44%	8.41%	30.11%		4.36%	1.43%	0.03%	0.00%
G1	45.06%				50.31%	4.63%			
	0.00%	1.76%	6.14%	37.16%		3.05%	1.55%	0.03%	0.00%
O Tempo	40.59%				53.99%	5.42%			
	0.00%	1.36%	7.23%	32.00%		4.49%	0.90%	0.03%	0.00%
R7	41.98%				51.20%	6.82%			
	0.01%	1.53%	7.89%	32.55%		5.32%	1.49%	0.01%	0.00%
UOL	39.29%				53.86%	6.85%			
	0.01%	2.13%	7.85%	29.30%		5.59%	1.23%	0.03%	0.00%
Yahoo	43.73%				51.15%	5.12%			
	0.01%	2.87%	9.96%	30.88%		4.25%	0.85%	0.00%	0.02%
Total	42.13%				52.31%	5.56%			
	0.01%	1.84%	7.84%	32.44%		4.29%	1.24%	0.02%	0.01%

**Figura 1. Polaridade das manchetes**

para todos os jornais, indicando que notícias com polaridade positiva são mais raras em comparação a notícias classificadas com outras polaridades.

Para uma melhor compreensão dos resultados obtidos, particionamos as análises numa perspectiva temporal, com o objetivo de verificar se a polaridade das notícias publicadas é afetada por uma determinada época ou devido a ocorrência de algum evento específico. Os resultados são apresentados na seção subsequente.

4.2. Polaridade ao longo do tempo

Nós investigamos a polaridade das manchetes em função do tempo considerando diferentes aspectos. Em primeiro lugar, nós verificamos se a polaridade das manchetes mudou ao longo do período para os dados coletados. Em seguida, analisamos se houve mudanças na polaridade das notícias que poderiam indicar reflexo de temas quentes ou eventos específicos. Depois, verificamos se existem flutuações de polaridade em determinados dias da semana, ou seja, investigamos se os dias da semana apresentam influência sobre a polaridade de notícias publicadas. Por fim propomos uma estratégia para analisar o teor da polaridade durante o período.

Considerando a distribuição de notícias publicadas ao longo do período, observamos na Figura 2 que o número de manchetes neutras e negativas é constantemente

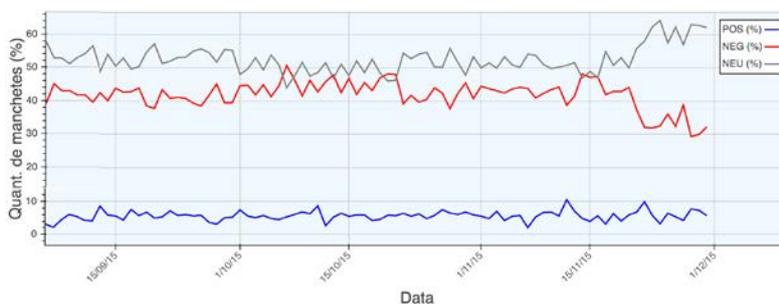


Figura 2. Percentual de notícias negativas, neutras e positivas no decorrer do tempo



(a) 13 de outubro

(b) 14 de novembro

Figura 3. Nuvem de palavras de manchetes - por dia

superior ao número de notícias positivas, com ligeiras flutuações entre as duas primeiras onde, em algumas datas, o percentual de notícias negativas inclusive superou o de notícias neutras. Com base neste aspecto averiguamos e percebemos que, particularmente, para os dias 7, 13, 20 e 21 de Outubro e 14 de Novembro de 2015, é possível notarmos algumas variações significativas (picos ou curvas) que indicam que o número de manchetes negativas publicadas é maior. Assim, para investigarmos a hipótese de que os acontecimentos deste período influenciaram a polaridade das manchetes, selecionamos as notícias veiculadas nestas datas e utilizamos um extrator de entidades⁵ para obtermos uma nuvem de palavras contendo as entidades mais citadas nas respectivas manchetes. O resultado para as datas de 13 de outubro e 14 de novembro é exibido na Figura 3⁶.

Para as datas analisadas, percebemos que o dia 13 de outubro (ver Figura 3(a)) foi marcado por notícias do contexto político nacional que afetaram de forma negativa a polaridade associada às manchetes publicadas durante este período. Além disso, podemos observar que o dia 14 de novembro foi marcado por um atentado terrorista à cidade de Paris (ver Figura 3(b)). Esse acontecimento social impactou fortemente a polaridade das manchetes veiculadas nesta data. Desta forma, acreditamos que existe forte influência de determinados acontecimentos na polaridade das notícias publicadas durante o período analisado.

Por fim, em um contexto temporal mais específico, analisamos a evolução da polaridade das manchetes no decorrer dos dias da semana. O resultado apresentado na Figura 4 nos permite notar um aumento na quantidade de notícias negativas publicadas nas segundas feiras, e uma queda na quantidade de notícias neutras publicadas, também no

⁵Tag Crowd: <http://www.tagcrowd.com/>

⁶A nuvem de palavras para as demais datas não é apresentada por limitação de espaço.

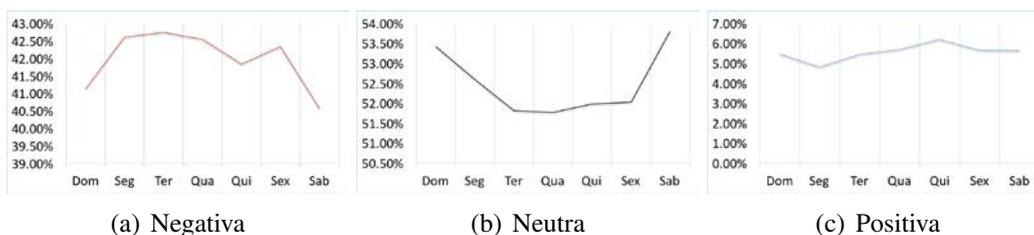


Figura 4. Percentual de manchetes publicadas - dias da semana

começo da semana. Observamos ainda que, em finais de semana, as notícias negativas diminuem. Em relação às notícias positivas, enxergamos apenas leves flutuações, mas em geral, esse percentual se mantém.

Finalmente, utilizamos nossa análise em relação à força da polaridade (ver Seção 4.1 e Tabela 3) para propormos uma forma de mensurar e analisar o teor de negatividade ou positividade expressa nas manchetes durante o período analisado. Uma vez que estamos interessados apenas no teor de negatividade ou positividade diários, nesta análise as manchetes classificadas com polaridade neutra não foram consideradas para o cálculo. Inicialmente, calculamos a média de polaridade diária, ou seja, se tivermos uma notícia negativa classificada como -4 e uma positiva classificada como 2, o teor de polaridade do dia será -1 (ou seja, $((-4) + (+2)) / 2 = -1$, teor negativo, neste caso). O resultado é apresentado na Figura 5 (linha cinza).

Observamos que o teor de polaridade é negativo durante todo o período analisado. Entretanto a média considera o número absoluto de notícias classificadas com determinada polaridade, dificultando, de certa forma a percepção da força da polaridade de cada uma das classes. Para analisarmos este aspecto, propomos uma análise que considera a diferença do módulo das médias positiva e negativa separadamente, em outras palavras, qual média é maior em módulo. Suponhamos por exemplo uma situação em que tenhamos 10 notícias negativas classificadas com força de polaridade -1, a média de polaridade negativa para este caso seria -1. Consideremos também que para esta mesma data tenham sido publicadas somente 2 notícias positivas classificadas com força de polaridade +4. Neste caso a média de polaridade positiva seria +4. A média de polaridade geral para este exemplo seria -0,2, entretanto a diferença do módulo das médias positiva e negativa separadamente seria +3 ($+4-1$). Acreditamos que isso representa de forma mais adequada a força e a expressividade das notícias nesta data. Ou seja, apesar de haver mais notícias negativas as notícias positivas são mais expressivas em termos de força de polaridade. O resultado desta análise também é apresentado na Figura 5 (linha laranja).

Percebemos que para algumas datas a força da polaridade das notícias positivas foi maior, apesar de estarem em menor número. Entretanto, não foi possível identificarmos um padrão claro, uma vez que os valores variam muito. Consideramos que ainda assim esta análise é útil uma vez que nos permite identificar em quais datas as notícias de cada uma das classes foram mais impactantes.

5. Considerações Finais e Trabalhos Futuros

O número de leitores de notícias na web tem crescido drasticamente e com isso os jornais online têm investido cada vez mais em estratégias para atrair os cliques dos usuários. Entretanto, pouco se sabe sobre estes aspectos. Propomos neste trabalho um primeiro

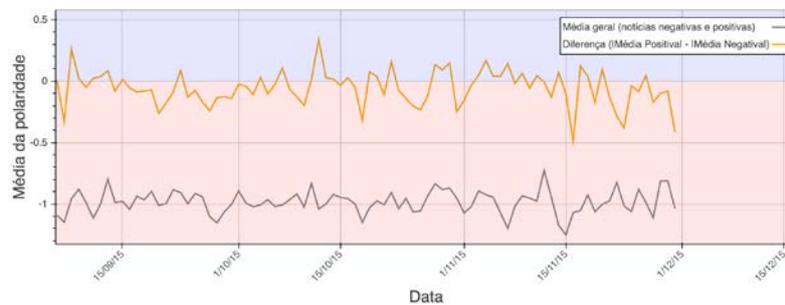


Figura 5. Teor da polaridade expressa nas manchetes

passo no entendimento deste universo com foco no contexto brasileiro. Exibimos uma caracterização de notícias neste ambiente online, que tem se tornado cada vez mais competitivo.

Para isso apresentamos uma metodologia experimental projetada para realização de nossas análises onde colocamos as manchetes como artefato central, a partir da hipótese de que elas são um importante elemento para atração da atenção dos usuários de notícias online. Nossa metodologia propõe a utilização de métodos de análise de sentimentos como meio para o entendimento de aspectos intrínsecos às manchetes.

Inicialmente, caracterizamos os dados de um conjunto de 59.510 notícias produzidas por oito grandes empresas de mídia brasileiras, reunidos durante um período de aproximadamente 3 meses do ano de 2015. Nossos resultados revelam várias observações interessantes sobre a polaridade expressa em notícias e o seu uso neste contexto, e fornecem características que podem ser utilizadas para a concepção de novos sistemas, e finalmente, para compreendermos a sociedade em que vivemos.

Revelamos uma tendência em todos os jornais analisados: o número de notícias neutras é maior, seguido do número de notícias negativas (que é competitivo em relação ao número de notícias neutras) e por fim, as notícias positivas (cuja quantidade é bastante baixa). Diferente de [Reis et al. 2015a], mostramos ainda que, para o cenário brasileiro existe grande concentração de notícias classificadas como neutras, e poucas classificadas no extremo negativo ou positivo. Sabemos que as notícias lidas podem impactar o comportamento dos usuários [Nguyen et al. 2011], e indicadores que apontam nessa direção também surgiram no estudo qualitativo realizado em [Reis et al. 2014a]. Outro aspecto está relacionado ao fato de que notícias neutras não são muito atrativas para os usuários [Reis et al. 2015a]. Logo, esses resultados contribuem para que outros pesquisadores investiguem características ou estratégias que possam apoiar o usuário na seleção de conteúdo neste cenário. Por exemplo, existem esforços recentes para a proposição de ferramentas que proporcionem ao usuário a possibilidade de selecionar um conteúdo com base na polaridade expressa pelo mesmo [Reis et al. 2014b]. Como trabalho futuro, pretendemos explorar este aspecto para o cenário brasileiro.

Já nossa análise temporal revelou que o número de manchetes neutras e negativas é constantemente superior ao número de notícias positivas. Além disso, mostramos evidências de que existe forte influência de determinados acontecimentos na polaridade das notícias publicadas durante o período analisado. Acreditamos que este tipo de descoberta possa contribuir para trabalhos que explorem esquemas de classificação com base

na previsão da popularidade de notícias. Como trabalho futuro pretendemos analisar a correlação entre a popularidade e a polaridade de uma notícia, sob a perspectiva, por exemplo, de um período ou determinado acontecimento.

Mostramos que é perceptível um aumento na quantidade de notícias negativas publicadas em segundas feiras além de uma queda na quantidade de notícias neutras publicadas no começo da semana. Apresentamos ainda que, em finais de semana, as notícias negativas diminuem. Acreditamos que uma possível explicação para esta variação esteja relacionada ao *Efeito Dia Da Semana*. Esta é uma teoria amplamente estudada em áreas do conhecimento como psicologia e economia que aponta para uma possível correlação entre dias úteis com momentos de estresse e mau-humor, e fins-de-semana com momentos de felicidade e prazer [M. Ryan et al. 2010]. Como trabalho futuro, pretendemos explorar a validação desta hipótese por meio de uma análise mais aprofundada focada especificamente neste aspecto.

Propomos também uma estratégia para análise da força da polaridade das notícias veiculadas numa mesma data. Mostramos que para algumas datas a força de polaridade das notícias positivas é maior, apesar de estarem em menor número, entretanto esse número é bastante variável. Diferente da média, esse enfoque sobre a força da polaridade explicita a expressividade da polaridade de cada uma das classes, e acreditamos que este aspecto possa ser utilizado por outros trabalhos que façam uso da análise de sentimentos como mecanismo para captura do sentimento de um texto qualquer.

Ainda como trabalho futuro, pretendemos investigar de forma mais aprofundada os resultados da execução do método de análise de sentimentos para o idioma português com objetivo de propormos estratégias que melhorem o desempenho destas abordagens no contexto de notícias online.

Por fim, acreditamos que nossa metodologia proposta, de forma geral, possa facilitar ou apoiar diferentes esforços de investigação neste domínio, impulsionando assim a produção de conhecimento.

Agradecimentos

Este trabalho foi parcialmente financiado pelo projeto FAPEMIG - PRONEX - MASWeb, Modelos, algoritmos e sistemas para a Web (número do processo APQ - 01400-14), e por projetos individuais do CNPq, CAPES e Fapemig.

Referências

- Araújo, M., Diniz, J. P., Bastos, L., Soares, E., Júnior, M., Ferreira, M., Ribeiro, F., and Benevenuto, F. (2016). ifeel 2.0: A multilingual benchmarking system for sentence-level sentiment analysis. In *Proceedings of the International AAAI Conference on Web-Blogs and Social Media*, Cologne, Germany.
- Araújo, M., Gonçalves, P., Benevenuto, F., and Cha, M. (2014). ifeel: A system that compares and combines sentiment analysis methods. In *WWW (Companion Volume)*. International World Wide Web Conference (WWW'14).
- Araújo, M., Reis, J., Pereira, A., and Benevenuto, F. (2016). An evaluation of machine translation for multilingual sentence-level sentiment analysis. In *Proceedings of the ACM Symposium on Applied Computing (SAC'16)*. ACM.
- Bandari, R., Asur, S., and Huberman, B. A. (2012). The pulse of news in social media: Forecasting popularity. In *ICWSM*, pages 26–33.

- dos Santos, A. G. L., Becker, K., and Moreira, V. (2014). Um estudo de caso de mineração de emoções em textos multilíngues. In *Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM'14)*.
- Ecker, U. K., Lewandowsky, S., Chang, E. P., and Pillai, R. (2014). The effects of subtle misinformation in news headlines. *Journal of experimental psychology: applied*, 20(4):323.
- Esiyok, C., Kille, B., Jain, B.-J., Hopfgartner, F., and Albayrak, S. (2014). Users' reading habits in online news portals. In *Proceedings of the 5th Information Interaction in Context Symposium*, pages 263–266. ACM.
- Evans, H. (1974). *News headlines*. Holt, Rinehart and Winston.
- Gonçalves, P., Araújo, M., Benevenuto, F., and Cha, M. (2013). Comparing and combining sentiment analysis methods. In *1st ACM Conference on Online Social Networks (COSN'13)*, pages 27–38.
- Konnikova, M. (2016). How headlines change the way we think. http://www.newyorker.com/science/maria-konnikova/headlines-change-way-think?mbid=social_twitter. Acessado em Março, 2016.
- M. Ryan, R., H. Bernstein, J., and Warren Brown, K. (2010). Weekends, work, and well-being: Psychological need satisfactions and day of the week effects on mood, vitality, and physical symptoms. In *Journal of Social and Clinical Psychology*, volume 29, pages 95–122.
- Mitchelstein, E. and Boczkowski, P. J. (2009). Between tradition and change: A review of recent research on online news production. *Journalism*, 10(5):562–586.
- Neves, L., Perez, K., Speisman, C., and Westerfield, C. (2013). Headlines sentiment project. <https://headlines-and-data.herokuapp.com/>. Acessado em Março, 2016.
- Nguyen, A., Legg, A., and Sweeny, K. (2011). Do you want the good news or bad news first? news order influences recipients' mood, perceptions, and behaviors. *UCR Undergraduate Research Journal*, 5:31–36.
- Reis, J., Benevenuto, F., Vaz de Melo, P., Prates, R., Kwak, H., and An, J. (2015a). Breaking the news: First impressions matter on online news. *Proceedings of the 9th International AAAI Conference on Web-Blogs and Social Media*.
- Reis, J., Gonçalves, P., Miranda, L., Prates, R., and Benevenuto, F. (2014a). A percepção dos usuários sobre a utilização de análise de sentimentos como apoio a seleção de notícias. In *Proceedings of the XIII Brazilian Symposium on Human Factors in Computing Systems*. Sociedade Brasileira de Computação.
- Reis, J., Gonçalves, P., Vaz de Melo, P. O., Prates, R., and Benevenuto, F. (2014b). Magnet news: You choose the polarity of what you read. *Proceedings of the International Conference on Web and Social Media (ICWSM)*.
- Reis, J., Oliveira, P., Araújo, M., Pereira, A., and Benevenuto, F. (2015b). Uma abordagem multilíngue para análise de sentimentos. In *Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM'15)*.
- Tatar, A., Antoniadis, P., De Amorim, M. D., and Fdida, S. (2014). From popularity prediction to ranking online news. *Social Network Analysis and Mining*, 4(1):1–12.
- Tewksbury, D. (2005). The Seeds of Audience Fragmentation: Specialization in the Use of Online News Sites. *Journal of Broadcasting & Electronic Media*, 49(3):332–348.
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558.
- Varian, H. (2010). Newspaper economics: Online and offline. In *presentation at FTC Workshop on How Will Journalism Survive the Internet Age Washington, March*, volume 9.