Detecção de Casos de Violência Patrimonial a partir do Twitter

João Paulo Clarindo[1], Fábio Coutinho[1], André Lage Freitas[1]

[1]Instituto de Computação – Universidade Federal de Alagoas (UFAL) – Maceió, AL – Brasil

{jpcs,fabio,andre.lage}@ic.ufal.br

Abstract. *In recent years, cases of patrimonial violence have become part of the daily lives of Brazilians. This situation imposes on public security managers a big challenge: look for alternatives to the official statistics in order to reflect a more consistent overview of the reality. The popularity of social networks has resulted in continuous data generation about stories, feelings and opinions of their users. This work presents DETECT, a system which analyzes Twitter data in order to detect messages (tweets) that indicate the occurrence of patrimonial violence crimes, filling the gap presented in official data. We have experienced DETECT and initial results showed to be equivalent to the official data.*

Resumo. *Nos últimos anos, casos de violência patrimonial passaram a fazer parte do cotidiano dos brasileiros. Tal situação impõe aos gestores da segurança pública um grande desafio: buscar alternativas às estatísticas oficiais a fim de refletir uma visão mais condizente com a realidade. A popularização das redes sociais tem resultado na geração contínua de dados sobre relatos, sentimentos e opiniões de seus usuários. Este trabalho apresenta o sistema DETECT, que analisa dados da rede social Twitter com o objetivo de detectar mensagens (tweets) que indiquem a ocorrência de crimes de violência patrimonial, suprindo a lacuna existente em dados oficiais. Um experimento foi realizado e os resultados encontrados mostraram-se equivalentes a dados oficiais.*

1. Introdução

Inicialmente, as redes sociais na Web eram vistas apenas como ambientes que permitiam a seus membros se comunicarem com amigos e familiares a partir das conexões mantidas entre os usuários. Com o decorrer do tempo, tais plataformas transformaram-se em um profícuo mecanismo de comunicação por meio do qual usuários são capazes de expressar sentimentos e manifestar suas opiniões.

O Twitter é uma das redes sociais mais utilizadas no mundo atualmente. O Brasil ocupa a segunda posição entre os países com mais usuários no Twitter [Nisha 2016]. Em janeiro de 2016, o Twitter possuía cerca de 320 milhões de usuários ativos que geravam aproximadamente 1 bilhão de mensagens diariamente [Twitter 2016]. Naturalmente, essa rede social representa uma importante fonte para a troca de opiniões.

Nos últimos anos, a segurança pública tem se tornado um dos maiores desafios da sociedade brasileira. De fato, os dados são alarmantes e evidenciam uma escalada da violência que atinge principalmente os grandes centros urbanos. De acordo com uma pesquisa realizada em São Paulo [CPP/Insper 2013], no período entre 2003 e 2013 registrou-se um aumento de 37% nos casos de furto, quando considerados os dados oficiais do governo estadual. Todavia, o mesmo estudo conclui que o aumento real foi de 81,5%. Tal diferença pode ser explicada pelo fato de que muitas vítimas não registram

a ocorrência desses crimes. Em se tratando de casos de roubo, apenas 35,3% das pessoas entrevistadas fizeram o registro, sendo o "medo de represália" (30%) e a "descrença que faria alguma diferença" (18,9%) os principais motivos alegados por quem não registrou o boletim de ocorrência.

Sendo assim, faz-se necessário buscar alternativas que permitam melhorar as estatísticas utilizadas pelos gestores e, preferencialmente, que reflitam uma visão mais recente dos acontecimentos e mais condizente com a realidade. Uma fonte viável para a obtenção desses dados são as redes sociais. De fato, devido à ocorrência cotidiana de assaltos, roubos e furtos nas principais cidades brasileiras, é natural que os seus cidadãos se utilizem das redes sociais para manifestar indignação e alertar amigos mediante o relato dos crimes aos quais foram submetidos.

Este trabalho apresenta o sistema DETECT (**DET**ecção de **E**ventos **C**riminais a partir do **T**witter), que analisa dados da rede social Twitter com o objetivo de detectar mensagens (*tweets*) que relatem a ocorrência de crimes de violência patrimonial. Além de auxiliar aos gestores, tais dados podem servir também para ajudar na orientação dos cidadãos em áreas que não estão familiarizados, possivelmente, alimentando sistemas de navegação a fim de alertar os usuários sobre trechos com maior incidência de crimes.

Este artigo encontra-se organizado da seguinte maneira. Inicialmente, a Seção 2 discute os principais trabalhos relacionados encontrados na literatura. A Seção 3 detalha o funcionamento do sistema enquanto a Seção 4 descreve uma prova de conceito do sistema DETECT. Finalmente, a Seção 5 apresenta as considerações finais.

2. Trabalhos Relacionados

O problema de extrair e analisar *tweets* relacionados à violência pode ser visto como um problema de categorização ou classificação de tópicos, ou seja, decidir se um *tweet* será classificado como violento ou não-violento. Nesta seção, são discutidos os principais trabalhos encontrados na literatura que analisam dados do Twitter no contexto da violência.

[Cano Basave et al. 2013] apresentam um trabalho de detecção de *tweets* sobre violência através de um modelo de detecção de violência (VDM – *Violence Detection Model*). Este modelo propõe uma classificação da violência em nível de documento para domínios gerais em conjunção com detecção de tópicos e análise de tópicos relacionados à violência. O trabalho apresenta um modelo interessante para identificação de tweets violentos, contudo, não possui abordagem para a definição e tratamento da localização onde o *tweet* foi submetido.

[Li et al. 2012] introduzem um sistema de análise e detecção de eventos relativos à criminalidade e a desastres naturais baseados no Twitter, denominado TEDAS. O sistema permite detectar novos eventos, analisar padrões temporais e espaciais e identificar a importância dos eventos através do uso de mapas de calor gerados a partir de dados de geolocalização de *tweets*. TEDAS utiliza uma amostra menor de *tweets* para a análise com enfoque principal voltado para a detecção de desastres naturais.

ReDites [Osborne et al. 2014] constitui um sistema de detecção, rastreamento e monitoramento em tempo real de eventos em redes sociais. No experimento realizado neste trabalho, foi avaliada a possível detecção do ataque terrorista ao shopping Westgate, que ocorreu em setembro de 2013 no Quênia, utilizando apenas *tweets* obtidos através da

Streaming API. ReDites propõe a análise de *tweets* geolocalizados em tempo real, identificando as localidades das ocorrências de imediato.

Similarmente aos trabalhos acima relacionados, DETECT extrai e analisa os *tweets* em busca de padrões. Um dos diferenciais desta proposta é a possibilidade de integração dos dados encontrados com dados governamentais relacionados à segurança pública, possibilitando a validação dos mesmos.

3. O Sistema DETECT

Esta seção descreve as etapas seguidas neste trabalho para o processamento dos dados da plataforma Twitter: extração dos *tweets*, análise dos dados e visualização.

A Figura 1 apresenta uma visão geral do sistema, cuja execução é dividida em quatro passos: *extração, análise, identificação de localização* e *resultados*. O Passo 1 representa a fase de extração, que dispõe do módulo de extração, responsável por obter os *tweets* através da *Streaming API* e gerar arquivos de saída para serem utilizados pelo módulo de seleção executado no Passo 2. O módulo de seleção irá recuperar os *tweets* que contêm termos sobre violência patrimonial. Durante a fase de análise, a cada *tweet* é atribuído um índice de relevância a fim de detectar aqueles que de fato relatam um crime de violência patrimonial. Em seguida, no Passo 3, DETECT tentará identificar o local da ocorrência do crime através de seu próprio conteúdo ou de tags de geolocalização a partir do módulo de localização do *tweet*. Finalmente, no Passo 4, os resultados são visualizados e armazenados pelo módulo de visualização e armazenamento.

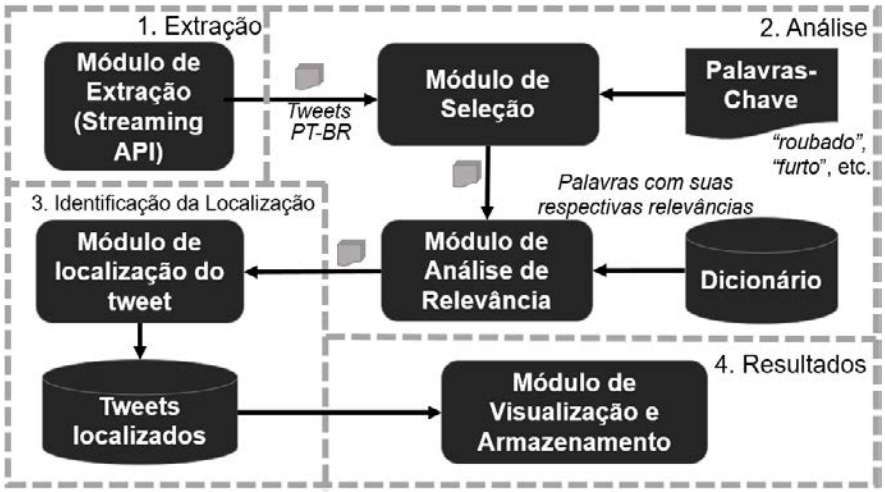

Figura 1. Visão Geral do Sistema DETECT

As seções a seguir, descrevem em detalhes os processos de extração, análise e visualização destes dados.

3.1 Extração de *tweets*

Para o módulo de extração de *tweets*, foi escolhido o mecanismo de extração *Streaming API,* que permite a extração em modo *stream*, ou seja, no momento da geração dos *tweets*.

O módulo de extração implementado em DETECT foi desenvolvido em Python. A abrangência da extração são os *tweets* gerados em tempo real e escritos em língua portuguesa, sem nenhuma filtragem de termos específica.

Além da funcionalidade básica de extração, o módulo exclui os *retweets* (replicação de *tweets*) no momento da extração para evitar repetições.

3.2 Análise

O processo de escolha para as palavras-chave utilizado na seleção de *tweets* teve origem a partir de uma pesquisa acerca de termos comuns presentes em relatos de crimes de violência patrimonial. As derivações destas palavras também foram utilizadas como palavras-chave. A base de termos que serviu como referência para a seleção neste trabalho foi construída levando em conta dados de boletins de ocorrência e comentários postados na Web a partir do site "Onde fui roubado"[1]. Dentre as palavras utilizadas na seleção de *tweets* incluem-se *roubo, assalto, furto, arrombamento*, *mão armada*, e *arrastão*.

Para implementar o módulo de seleção de *tweets*, foi utilizado o Pentaho Data Integration (Kettle), que disponibiliza um ambiente ETL[2]. A transformação de seleção feita pelo Kettle é representada pela Figura 2, onde os *tweets* são filtrados de acordo com as palavras-chave e enviados para o módulo de análise de relevância.

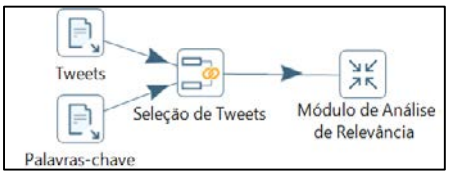

Figura 2: Transformação para seleção de tweets

Após os *tweets* terem sido selecionados de acordo com as palavras-chave, um dicionário com índices de relevância foi criado a partir de uma solução desenvolvida em Python, utilizando a biblioteca de análise léxica NLTK[3]. A criação do dicionário teve como base alguns *tweets* selecionados para estudo. Para cada *tweet*, as *stopwords* (artigos, pronomes, etc.) são removidas e às palavras restantes são atribuídas relevâncias de acordo com a frequência das mesmas. Para aquelas que possuem relação com o tema, atribui-se uma relevância positiva. Às palavras mais frequentes que não possuem relação com o tema e que remetem um novo sentido ao *tweet,* designa-se uma relevância negativa. Por fim, para palavras que não dizem respeito ao contexto da violência patrimonial e não interferem no sentido da mensagem define-se uma relevância neutra.

Considerando, por exemplo, o *tweet* "*Ontem à noite uma mulher foi assaltada no ponto de ônibus do meu lado*", a análise o destacaria como resultado devido à ocorrência de palavras com alta relevância ("ponto de ônibus"). Por outro lado, o *tweet* "*Está na hora de assaltar a geladeira*" não seria retornado pelo DETECT devido ao termo "geladeira" estar muitas vezes associado à expressão popular *assaltar a geladeira* e, consequentemente, possuir baixa relevância.

Sendo assim, através de uma transformação no Kettle, é atribuída a relevância de cada tweet a partir do dicionário. A relevância total de um tweet é calculada a partir da soma das relevâncias de cada palavra. Caso o resultado seja positivo, o tweet é considerado relevante.

3.3 Localização de Tweets

Os *tweets* sem informações de geolocalização correspondem a maior parte de *tweets* gerados [Gonzalez et al. 2012]. Para identificar a origem dessas mensagens, foi implementada uma abordagem que se baseia em: (i) verificar se algum logradouro foi

[1] www.ondefuiroubado.com.br
[2] http://community.pentaho.com/projects/data-integration/
[3] http://www.nltk.org/

incluído no conteúdo do *tweet*; e (ii) identificar a localização do usuário que gerou o *tweet* através da obtenção de dados do perfil de usuário. Em (i), os *tweets* passam por uma análise léxica a partir de um conjunto de palavras-chave (ex. rua, avenida, estrada, etc). Mediante a obtenção do logradouro e da cidade do usuário, tal informação é analisada pelo Google Maps API para atribuir a geocodificação correspondente. Finalmente, os dados de latitude e longitude retornados pela Google Maps API são enviados para o módulo de visualização, que gera um mapa de calor relacionado com as áreas de maior incidência de *tweets* retornados pelo DETECT. Os mapas de calor foram implementados a partir de um script Python que utiliza a Google Maps API. Os resultados são gerados em um arquivo .html que inclui a exibição do mapa.

4. Prova de Conceito

A fim de avaliar as funcionalidades do DETECT, foi realizada uma extração de *tweets* entre os meses de setembro de 2015 e janeiro de 2016. Destes *tweets,* 210.881.326 não possuíam informações de geolocalização e 1.679.008 possuíam informações de geolocalização. O total de *tweets* extraídos no período mencionado foi de 212.560.334.

O sistema DETECT foi executado a partir de um ambiente construído sobre a plataforma de nuvem Microsoft Azure. Tal fato garantiu a estabilidade necessária durante a fase de coleta dos *tweets*. Para o protótipo utilizado neste experimento, foram implementados os seguintes módulos: módulo de extração, módulo de seleção, criação de dicionário e módulo de análise de relevância para *tweets* geolocalizados.

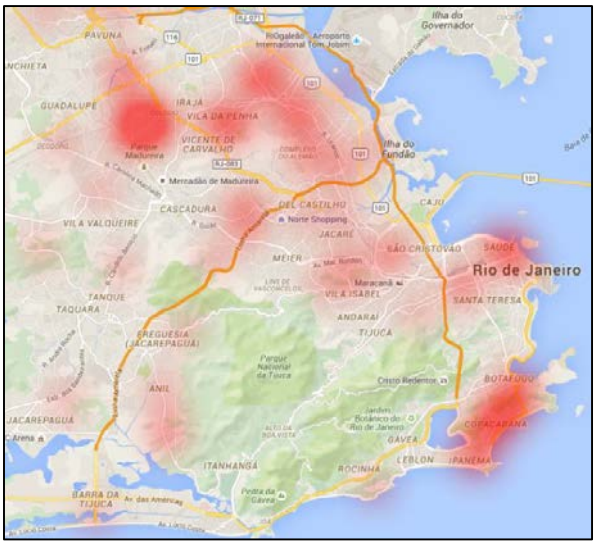

Figura 3: Mapa de calor refletindo tweets submetidos em áreas da cidade do Rio de Janeiro

O sistema detectou 569 tweets geolocalizados com relevância ao tema de violência patrimonial de acordo com o dicionário gerado. Dentre aqueles mais relevantes, a maioria foi gerada na cidade do Rio de Janeiro. Um mapa de calor com as principais áreas de geração destas mensagens no Rio de Janeiro é apresentado na Figura 3.

5. Considerações finais

Conforme descrito na seção anterior, DETECT retornou apenas *tweets* geolocalizados, o que explica o número diminuto do resultado (apenas 0,78% dos *tweets* extraídos possuíam informações de geolocalização). Tal fato ocorreu porque o módulo de identificação da

localização ainda não foi finalizado, logo, os trabalhos futuros incluirão também os *tweets* com localização a partir do conteúdo.

Além do desafio de prover a localização do tweet, outra questão importante é a redução dos "falsos positivos", que são os *tweets* que não são relevantes ao tema, embora o sistema tenha detectado como relevante. *Tweets* que envolvem ironias, piadas ou erros de digitação podem induzir o sistema à geração de falsos positivos. Para resolver este problema, em trabalhos futuros, está prevista a implementação de um módulo de análise dos *tweets* utilizando aprendizado de máquina para identificar os padrões de *tweets* relacionados ao tema. Mesmo com uma faixa limitada de *tweets* para análise, foi possível perceber um índice maior de *tweets* em áreas nobres da cidade do Rio de Janeiro (como Ipanema e Copacabana) e também em comunidades na periferia da cidade, apresentando correspondência com os dados oficiais [ISP-RJ 2016].

Nos trabalhos futuros, os resultados serão contrapostos frente aos dados populacionais do IBGE. Além disso, será finalizada a implementação dos módulos restantes e desenvolvida uma versão para processamento distribuído do sistema DETECT para garantir um melhor desempenho diante do grande volume de dados.

Referências

Cano Basave, A. E., He, Y., Liu, K. and Zhao, J. (out 2013). A weakly supervised Bayesian model for violence detection in social media. In 6th International Joint Conference on Natural Language Processing (IJCNLP)

CPP/Insper (2013). Relatório da Pesquisa de Vitimazação em São Paulo - 2003-2013. São Paulo.

Gonzalez, R., Figueroa, G. and Chen, Y.-S. (2012). TweoLocator: A Non-intrusive Geographical Locator System for Twitter. In Proceedings of the 5th ACM SIGSPATIAL International Workshop on Location-Based Social Networks. , LBSN '12. ACM.

ISP-RJ (2016). Incidências Criminais e Administrativas do Estado do Rio de Janeiro do ano de 2015. http://www.isp.rj.gov.br/Conteudo.asp?ident=108, [acessado em Abr 22].

Li, R., Lei, K. H., Khadiwala, R. and Chang, K. C.-C. (apr 2012). TEDAS: A Twitter-based Event Detection and Analysis System. In *2012 IEEE 28th International Conference on Data Engineering (ICDE)*.

Nisha (13 jul 2015). Top Ten Countries With Most Twitter Users. http://www.perfectinsider.com/top-ten-countries-with-most-twitter-users-in-the-world/, [acessado em Fev 10].

Osborne, M., Moran, S., McCreadie, R., et al. (2014). Real-time detection, tracking, and monitoring of automatically discovered events in social media. Association for Computational Linguistics

Twitter (2016). About Twitter. https://about.twitter.com/company, [acessado em Jan 26].