

Análise de Sentimentos para Previsão das Condições de Trânsito*

Bernardo Lauand¹, Jonice Oliveira¹

¹ Universidade Federal do Rio de Janeiro – DCC/IM, PPGI

Caixa Postal: 68.530, Cep: 21.941-909 – Rio de Janeiro – RJ – Brasil

belauand@ufrj.br, jonice@dcc.ufrj.br

Abstract. *Due to several buildings and routes changes related with preparations for the World Cup and Olympic Games, as well as the growing number of vehicles, the Rio de Janeiro's citizens constantly suffer with traffic jam. This project, called TweeTraffic, aims to use Twitter to predict traffic conditions, by mining the notifications provided by drivers and official sources.*

Resumo. *Devido às inúmeras obras e mudanças de vias por conta dos preparativos da Copa do Mundo e Jogos Olímpicos, além do crescimento do número de veículos em circulação, os moradores do Rio de Janeiro sofrem constantemente com congestionamentos. Este projeto, chamado TweeTraffic, visa utilizar o Twitter para a previsão das condições do trânsito, minerando os relatos dos motoristas, bem como notificações de fontes oficiais.*

1. Introdução

Atualmente, a cidade do Rio de Janeiro apresenta, em média, 94 quilômetros de trânsito lento por dia [PDTU/RMRJ 2012]. Condição que deve piorar pelo aumento constante de sua frota, com previsão que alcance em 2016 a 3.000.000 (três milhões) de veículos [Oliveira 2011]. Fora isso, a cidade enfrenta inúmeras alterações devido aos preparativos de dois grandes eventos mundiais: a Copa do Mundo (2014) e os Jogos Olímpicos (2016). Além dos congestionamentos típicos da hora do *rush*, os condutores precisam lidar também com situações imprevistas (ex: acidentes e enchentes). Através da análise dos sentimentos, o projeto *TweeTraffic* prevê as condições de trânsito na cidade do Rio de Janeiro, de maneira que os cidadãos possam planejar melhor as suas rotas.

2. Trabalhos Relacionados

O Google Maps Brasil [Google 2012] informa as condições de trânsito nas principais vias de algumas cidades brasileiras, além de disponibilizar um histórico do tráfego com base em cada dia da semana. O Google Maps obtém informações do trânsito de autoridades locais e também utiliza dados dos próprios usuários para ajudar a medir a

* Agradecemos ao CNPq e à FAPERJ. Um vídeo de demonstração da ferramenta está disponível em <http://sbbd2013.cin.ufpe.br/screenscasts>. Para executar a ferramenta, execute o arquivo disponível em <https://www.dropbox.com/s/3wffpmoa5ouka2z/TweeTrafficInterface.jar>.

velocidade, coletando de forma anônima os dados como posição e velocidade dos usuários que estejam utilizando o aplicativo [Hardware 2012].

O Endernoto [2011] é uma proposta para divulgar as condições de trânsito das ruas da cidade de Jakarta, na Indonésia. A polícia de Jakarta utiliza o Twitter para divulgar notícias sobre as condições de trânsito locais, que são apresentadas em um mapa através de um aplicativo para plataforma móvel Android.

Através do Waze [2012] o usuário pode compartilhar as informações sobre as condições de trânsito e utilizar as informações disponibilizadas por outros usuários como: localização de paradas policiais (*blitz*), congestionamentos, acidentes e obras.

O trabalho de Ribeiro et al. [2012] descreve um sistema de mineração de textos aplicado ao *Twitter* procurando por padrões de texto relacionados ao trânsito da cidade de Belo Horizonte utilizando apenas fontes oficiais. Para auxiliar o geoprocessamento, utiliza um *gazetteer* que contém os nomes, geometria e geolocalização de logradouros, bairros, cruzamentos e trechos de vias.

As contribuições do *TweeTraffic*, em comparação ao estado da arte, podem ser resumidas da seguinte forma:

- Esta proposta se baseia exclusivamente no conteúdo textual provido por livre participação da população (*crowdsourcing*) ou por fontes oficiais, sem utilizar qualquer outro recurso. Embora tenha sido construído e testado com as ruas do Rio de Janeiro, esta abordagem facilitaria a ampliação do seu uso em outras cidades brasileiras sem grandes mudanças arquiteturais ou de conteúdo.
- Utiliza informações providas por fontes oficiais e por fontes não-oficiais (população). Este é um ponto a se ressaltar, pois a população é um importante mecanismo na detecção e divulgação rápida de informação, muitas vezes mais rápida que as fontes oficiais. Fora isso, para localidades poucos conhecidas, a população torna-se o único meio de sensoriamento da região. Comparando-o especificamente com o Waze, que também trabalha com o conceito de *crowdsourcing*, seus mapas do Rio de Janeiro não estão bem construídos e não há muitas informações sobre o trânsito já que a comunidade brasileira no Waze é pequena. Isso torna o programa difícil de ser usado.
- As regras utilizadas na inferência das condições de trânsito foram construídas após observação de uma massa de dados.

3. TweeTraffic: A Ferramenta

O TweeTraffic é uma ferramenta que consiste em um cliente e um servidor. O servidor é responsável por extrair e analisar informações provenientes do *Twitter*, onde se conecta periodicamente em busca de informações relevantes sobre as principais ruas. Ele seleciona o conteúdo, o registra no banco de dados e realiza a análise de sentimentos. Esta é a etapa chamada “análise estática”. O servidor, quando consultado pelo cliente, realiza a “análise dinâmica”, inferindo a condição de trânsito da via consultada. O cliente (Figura 1) é um aplicativo *Android* onde o usuário pode verificar as condições de trânsito nas ruas da cidade do Rio de Janeiro e avaliar a resposta dada pelo sistema. Na Figura 2 temos a arquitetura conceitual do projeto TweeTraffic.

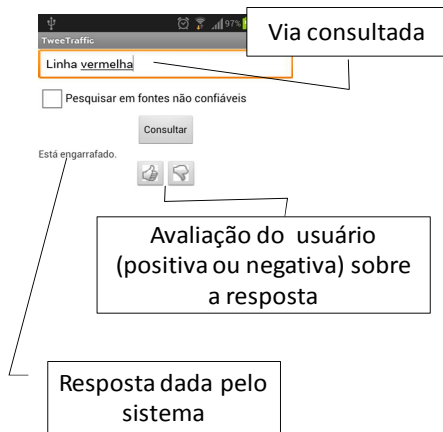


Figura 1 – Interface do Cliente [Lauand 2013]

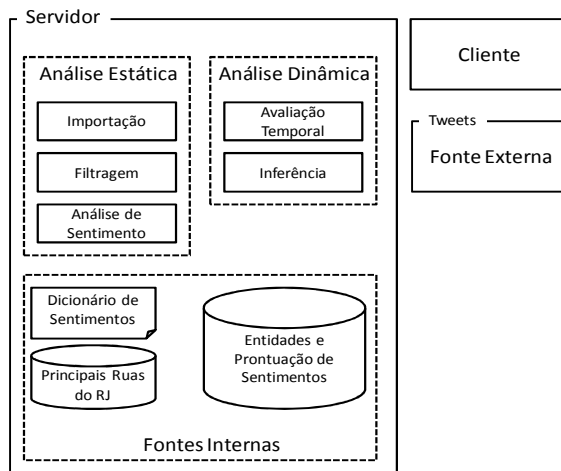


Figura 2 – Arquitetura Conceitual [Lauand 2013]

4. Verificação do estado atual do trânsito

Para realizar uma consulta, o cliente se conecta ao servidor e envia a rua a ser consultada (Figura 1). A seguir são realizadas as seguintes ações:

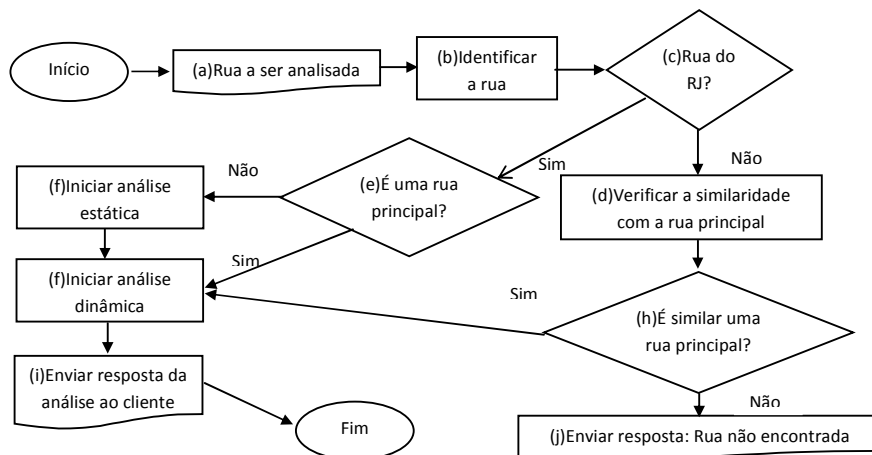


Figura 3. Fluxograma da identificação do estado [Lauand 2013]

4.1. Identificação da Rua

Ao receber o nome da rua a ser consultada, o servidor procura por abreviaturas comuns e as substitui pela sua palavra correspondente (figura 3.b), já que diferentes logradouros possuem nomes iguais e o que os distingue são as denominações de via. Por exemplo: Avenida Rui Barbosa (bairro Flamengo, zona sul do Rio de Janeiro), Rua Rui Barbosa (bairro Guaratiba, zona oeste) e Travessa Rui Barbosa (zona norte).

4.2. Verificação da Região

Após a fase de Identificação da Rua (figura 3.b), o servidor verifica se o nome recebido é de uma rua da cidade do Rio de Janeiro (figura 3.c). Caso a rua pesquisada não seja uma rua principal, será feita uma Análise Estática (seção 4.3), caso contrário, a Análise Dinâmica será feita (seção 4.4). Consideramos como 'ruas principais': "linha amarela",

"linha vermelha", "avenida presidente vargas", "avenida rio branco", "rua humaitá", "rua são clemente", "avenida nossa senhora de copacabana", "praça da bandeira", "ponte rio niteroi", "avenida brasil", "avenida das americas", "perimetral", "av nossa senhora de copacabana" e "rebouças".

4.3. Análise Estática

A análise estática é realizada continuamente sobre os dados captados periodicamente, processando-os e armazenando o resultado da análise de sentimento sobre cada entidade. Caso a via a ser consultada não seja uma via principal, a análise estática é acionada após o usuário informar a rua procurada. Esta análise é composta de três etapas: importação, filtragem e análise de sentimento.

- **Importação:** O servidor importa do Twitter todas as mensagens que contenham o nome da rua, postadas nos últimos 60 minutos e que não estejam armazenados no banco de dados.

- **Filtragem:** Caso o usuário opte por fazer a pesquisa considerando apenas os autores confiáveis, os *tweets* dos autores que não constam na relação de autores confiáveis são eliminados nessa etapa. Consideramos um "autor confiável" os perfis do Twitter de importantes veículos de comunicação como perfis oficiais de revistas, jornais, estações de rádio e televisão, além de perfis de entidades da prefeitura, como CET-RIO, Rio Trânsito e o Centro de Operações da Prefeitura do Rio de Janeiro.

- **Análise de Sentimento:** O servidor possui um Dicionário de Sentimentos, que é composto por um conjunto de palavras e suas respectivas classificações. Uma palavra pode ser classificada em: positiva, negativa ou modificadora. Uma palavra positiva é aquela que sugere uma boa condição de trânsito (ex: livre, bom, rápido). Já uma negativa denota uma má condição de trânsito (ex: ruim, parado, batida). Uma palavra modificadora por sua vez é aquela que altera a classificação de uma palavra de positiva para negativa ou vice versa (ex: sem, não, nenhum). A primeira fase da análise de sentimento separa o texto dos *tweets* em frases e as analisam individualmente, descartando as frases interrogativas. Para cada frase, contabiliza-se o número de palavras positivas e negativas e verifica-se a existência de palavras modificadoras. Caso haja, altera-se o valor das palavras positivas ou negativas próximas (distância de até 2 palavras) à palavra modificadora. O sistema classifica um *tweet* de acordo com a quantidade de palavras negativas e positivas.

4.4. Análise Dinâmica

Ela consiste em avaliar os *tweets* da última hora levando em consideração o resultado da análise estática e o horário em que o *tweet* foi publicado. A partir daí é feita uma inferência, baseando-se nas quatro situações distintas:

- **Caso 1:** Caso não haja nenhum *tweet* sobre a via consultada nos últimos 60 minutos, o sistema retorna: "Não há informações suficientes sobre essa via."

- **Caso 2:** Caso haja *tweets* na última hora com classificações divergentes de acordo com a análise estática, calcula-se o intervalo de tempo entre os *tweets* mais recentes e conflitantes. Se esse intervalo for superior a 15 minutos, o conflito é ignorado e considera-se a classificação do *tweet* mais recente como certa. Caso este seja positivo, o sistema retorna: "Não está engarrafado.". Caso contrário: "Está engarrafado."

•**Caso 3:** Caso haja *tweets* na última hora com classificações divergentes de acordo com a análise estática e o intervalo de tempo entre os *tweets* conflitantes seja menor do que 15 minutos, é contabilizado o total de *tweets* positivos e negativos. Caso haja mais *tweets* positivos, o sistema retorna: “Provavelmente não está engarrafado.”. Caso haja mais *tweets* negativos, o sistema retorna: “Provavelmente está engarrafado.”. Caso o número de *tweets* positivos seja igual ao número de *tweets* negativos o sistema considera a classificação do *tweet* mais recente como certa. Se positivo, retorna: “Provavelmente não está engarrafado.”. Caso contrário: “Provavelmente está engarrafado.”

•**Caso 4:** Caso haja *tweets* na última hora e não haja divergência de classificação de acordo com a classificação estática, o sistema retorna: i) “Não está engarrafado” caso haja ao menos um *tweet* positivo na última hora, ii) “Está engarrafado.” caso haja ao menos um *tweet* negativo na última hora e iii) “Provavelmente não está engarrafado” caso todos os *tweets* da última hora sejam neutros.

5. Estrutura do banco de dados

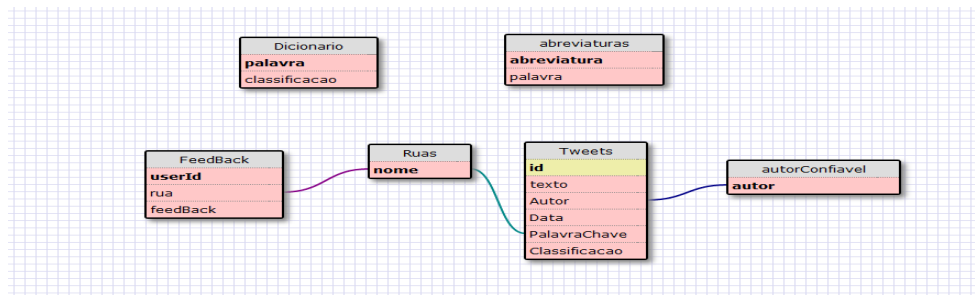


Figura 3 - Estrutura do banco de dados

A tabela dicionário armazena o conjunto de palavras predeterminadas e suas respectivas classificações, a tabela abreviaturas contém abreviaturas normalmente utilizadas em nomes de ruas e os nomes associados a elas.

Na tabela ruas há o nome de todas as ruas da cidade do Rio de Janeiro e a tabela autor confiável contém uma lista de usuários do *Twitter* que possuem uma relação direta com o trânsito. Já na tabela Tweets há os *tweets* dos últimos 60 minutos que o sistema identifica como relevante na análise da condição de trânsito do Rio de Janeiro.

6. Exemplo de Uso

Para facilitar a avaliação da ferramenta, criamos uma interface *desktop*¹. Assim, qualquer leitor poderá testar a solução sem precisar instalá-la em um dispositivo móvel. Neste exemplo, mostramos o resultado da consulta realizada no dia 10/07 às 22h58min sobre as condições da Avenida Presidente Vargas utilizando apenas as fontes oficiais (Figura 4), todos os *tweets* (Figura 5) e o estado real da via (Figura 6).

7. Conclusão

O *TweeTraffic* mostrou-se uma ferramenta prática, trazendo bons resultados. Pelo tratamento ser exclusivamente textual, enfrentamos algumas dificuldades como a distinção de ruas de outras localidades, mas com o mesmo nome de ruas cariocas. Em vias extensas e de mão dupla, ocorre a dificuldade em se descobrir em que região e

sentido da rua se encontra congestionado. Outra dificuldade encontrada foi o fato de os usuários do *Twitter* não georeferenciarem seus *tweets*, tornando impossível descobrir a real localização dos usuários na hora de filtra-los. Outro problema encontrado foi a recente mudança na API do *Twitter* que só permite 180 requisições a cada 15 minutos. Alterações no código foram necessárias e limitações do número de requisições que o programa pode atender apareceram.

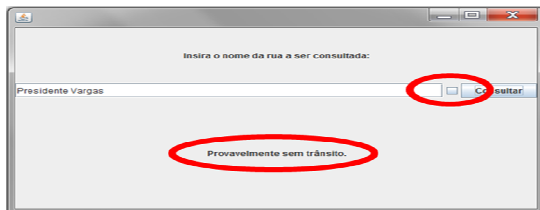


Figura 4 – Previsão com fontes confiáveis (oficiais)

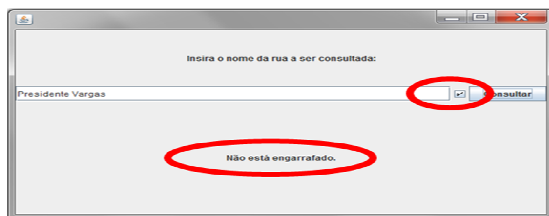


Figura 5 - Previsão com todas as fontes (oficiais e não-oficiais)



Figura 6 – Condição real, mostrada pela Câmera da CET-Rio¹: Via totalmente livre.

Referências

- Endarnoto, S. K. (2011). Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application. In *Electrical Engineering and Informatics (ICEEI)*, pages 1 – 4.
- Google (2012), “Google Maps”, <http://maps.google.com.br>, Setembro
- Hardware (2012) "Dados de trânsito em tempo real no Google Maps para mais cidades brasileiras", <http://www.hardware.com.br/noticias/2012-08/maps-transito-brasil.html>, Setembro
- Lauand, B., Oliveira, J. (2013). TweeTraffic: ferramenta de análise das condições de trânsito baseado nas informações do Twitter. In *II Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2013)*. SBC
- Oliveira, A. (2011). O Árduo Desafio de Ir e Vir na Cidade do Rio de Janeiro. In *Revista Eletrônica Novo Enfoque*, v. 13, n. 13, p. 170 – 187. UCB
- PDTU (2012) “Plano Diretor de Transporte Urbano da Região Metropolitana do Rio de Janeiro”, <http://www.pdtu.rj.gov.br/indexini.php>, Julho.
- Ribeiro, S. et al. (2012). Observatório do Trânsito: sistema para detecção e localização de eventos de trânsito no Twitter. In *Simpósio Brasileiro de Bancos de Dados*. SBC

¹ <http://www.rio.rj.gov.br/web/riotransito>