

Inferindo as Condições de Trânsito através da Análise de Sentimentos no *Twitter*

Bernardo Pereira Lauand¹, Jonice Oliveira^{1,2}

¹Programa de Pós Graduação em Informática (PPGI) – Universidade Federal do Rio de Janeiro (UFRJ)

² Departamento de Ciência da Computação (DCC) – Universidade Federal do Rio de Janeiro (UFRJ)

Caixa Postal 68.530 – 21.945-970 – Rio de Janeiro – RJ – Brazil

bernardo.lauand@ppgi.ufrj.br, jonice@dcc.ufrj.br

Abstract. *In large cities, traffic is becoming increasingly heavy at a fast pace. As a result, many cities, as Rio de Janeiro, face traffic problems. This project aims at showing how Twitter can be used as a resourceful data source for planning daily drive. For this, we developed the TweetTraffic where comments posted by drivers and news from officials agencies are processed and we infer the traffic conditions. We executed two assessments – one qualitative and other quantitative – and both results were satisfactory.*

Resumo. *O trânsito nas grandes metrópoles tem se intensificado num ritmo cada vez mais rápido. Como consequência, muitas cidades, como o Rio de Janeiro, sofrem com o problema de congestionamentos. Este projeto tem como objetivo mostrar que o Twitter pode ser uma importante fonte de dados no planejamento do deslocamento diário. Para isto foi construído o TweeTraffic. A partir dele os relatos dos motoristas, bem como notícias de fontes oficiais, são processados e a partir deles inferimos as condições do trânsito. Duas avaliações foram executadas - uma qualitativa e outra quantitativa - e em ambos os casos os resultados se mostraram satisfatórios.*

1. Introdução

Como em toda grande cidade que teve um crescimento acelerado e sem planejamento, o Rio de Janeiro precisa lidar com o problema do grande número de carros nas ruas. De acordo com o Plano Diretor de Transporte Urbano da Região Metropolitana do Rio de Janeiro - PDTU/RMRJ [2012], hoje em dia, esta cidade apresenta, em média, 94 quilômetros de trânsito lento.

Além dos congestionamentos típicos da hora do *rush* nas principais vias da cidade, os condutores precisam lidar também com situações imprevistas como acidentes, assaltos e outros tipos de violência, problemas mecânicos e condições climáticas adversas.

Devido aos problemas citados anteriormente as condições do trânsito têm piorado ao longo do tempo. Entre os anos de 2003 e 2010, o número de veículos em

circulação nas grandes cidades aumentou 66% e o tempo médio que o brasileiro gasta no trânsito cresceu 20% de acordo com o Portal do Trânsito [2012].

Em paralelo a estes acontecimentos, temos o crescimento do acesso à Internet e às mídias sociais. Lançado pelo governo, o plano de Banda Larga Nacional permitiu a comercialização de banda larga nas camadas mais populares, o que gerou um crescimento do uso da internet banda larga no Brasil. Segundo o IBOPE [2012], houve um aumento de 91% no número de usuários ativos de conexões de mais de 2 Mb. O Brasil teve um aumento de 16% no número de usuários de internet entre os anos de 2011 e 2012 tornando o país o sétimo maior mercado mundial de internet, segundo o comScore[2012].

Neste cenário, temos o crescimento de internet móvel. Segundo o Ministério das Comunicações [2012], o uso de internet móvel 3G no Brasil vem aumentando, sendo que no período de maio de 2010 a maio de 2011 houve um crescimento de 270% no número de usuários, passando de 15,2 milhões para 56,3 milhões. Em relação às mídias sociais, o Brasil se tornou um dos países mais atuantes nas principais redes sociais mundiais, possuindo a maior taxa de usuários de internet que acessam mídias sociais. O uso de redes sociais é consideravelmente maior no Brasil do que nos demais países da América Latina segundo o site Relevância [2012]: Cerca de 69% dos usuários brasileiros tem como uma das atividades participar de redes sociais contra 37% dos mexicanos de 21% dos argentinos.

Esse artigo é uma versão estendida e melhorada de [Lauand and Oliveira, 2013], onde apresentamos a ferramenta TweekTraffic que visa utilizar os dados do *Twitter* para prever as condições de trânsito na cidade do Rio de Janeiro, de maneira que os cidadãos possam planejar melhor as suas rotas. Nesta versão, além de melhorarmos a descrição da ferramenta, apresentamos novos resultados experimentais e uma comparação detalhada com os trabalhos correlatos, ressaltando as contribuições desta abordagem com o estado da arte.

Esse artigo está estruturado da seguinte maneira. Na seção 2 é descrita a arquitetura conceitual do aplicativo. Na seção 3 é apresentada a estrutura do banco de dados utilizado e na seção seguinte (4) é detalhada a abordagem criada para a inferência do estado atual do trânsito, mostrando passo a passo de como o aplicativo funciona. Um exemplo de uso é descrito na seção 5 e os trabalhos correlatos são descritos e comparados na seção 6. Na seção 7, os resultados das avaliações quantitativa e qualitativa são mostrados.

2. Arquitetura Conceitual

O projeto consiste em um cliente e um servidor. O cliente, mostrado na Figura 1, é um aplicativo *Android* onde o usuário pode verificar as condições de trânsito nas ruas da cidade do Rio de Janeiro, avaliar a resposta dada pelo sistema, verificar se há novos *tweets* de seus seguidores ou “tuitar”¹ qualquer mensagem. O servidor é responsável por extrair e analisar informações provenientes do *Twitter*. Ele se conecta periodicamente ao *Twitter* em busca de informações relevantes sobre as principais ruas, seleciona o conteúdo, as salva no banco de dados e realiza a análise de sentimentos. Esta é a etapa

¹ Tuitar refere-se ao ato de postar no *Twitter* comentários, informações, fotos, etc.

chamada “análise estática”, que é descrita na seção 4.3. O servidor, quando consultado pelo cliente, realiza a “análise dinâmica” (seção 4.4), inferindo a condição de trânsito da via consultada. Na Figura 2 temos a arquitetura conceitual do projeto TweepTraffic.

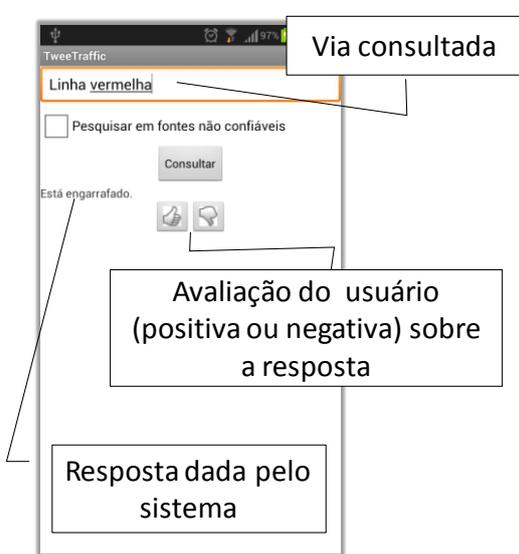


Figura 1 – Interface do Cliente [Lauand and Oliveira, 2013]

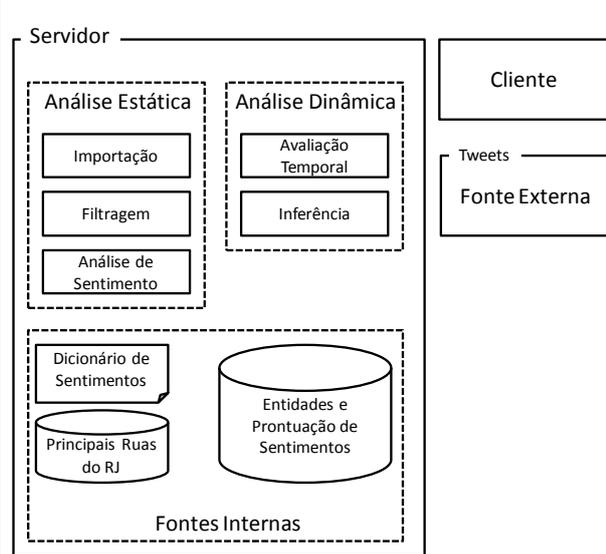


Figura 2 – Arquitetura Conceitual [Lauand and Oliveira, 2013]

3. Estrutura do Banco de Dados

A figura 3 apresenta o esquema do banco relacional utilizado.

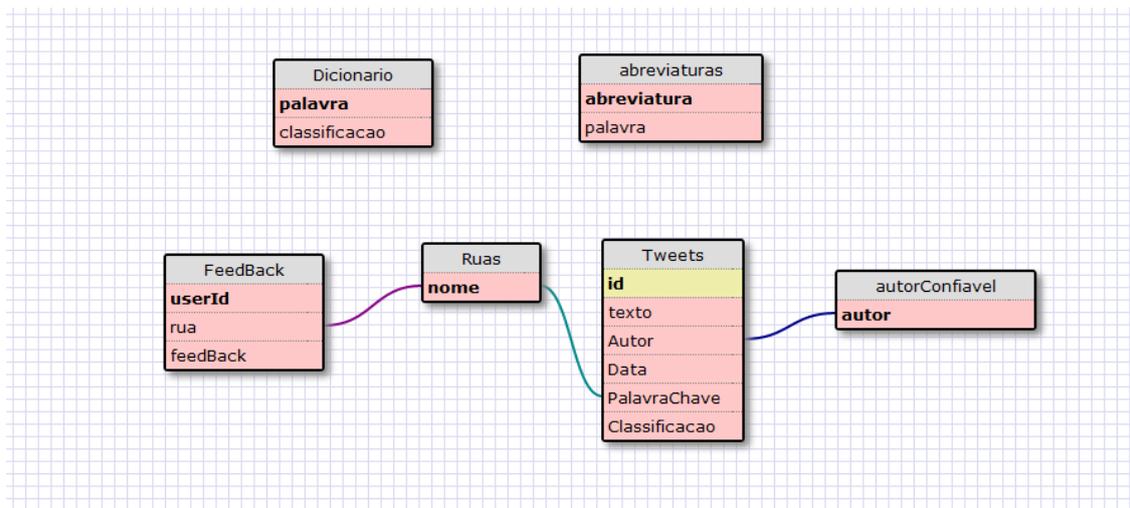


Figura 3 - Estrutura do banco de dados

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, como meios oficiais de comunicação ou órgãos/entidades de controle do 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.

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:

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 distinguem 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).

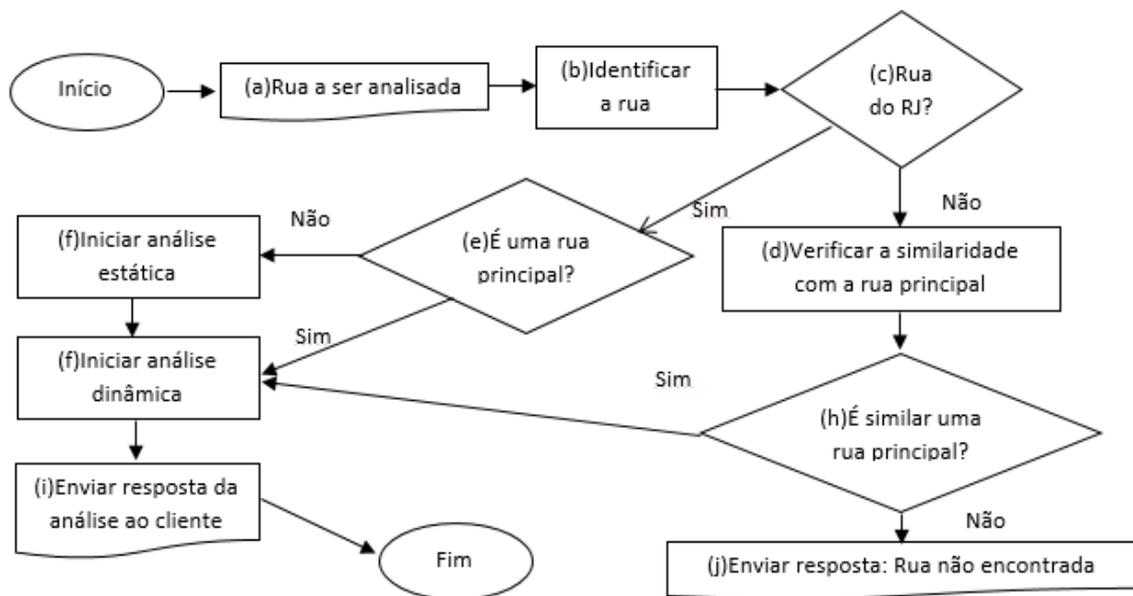


Figura 3. Fluxograma do procedimento para a identificação do estado [Lauand and Oliveira, 2013]

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). Por exemplo: Av. Brasil, como essa é uma rua principal, ela iria direto para a fase de Análise Dinâmica (figura 3.f), caso o usuário estivesse escrito errado o nome da rua principal, o aplicativo tentaria verificar se uma similaridade entre o escrito e uma rua principal (figura 3.d), para caso existisse iniciar a fase de Análise Dinâmica (figura 3.f).

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. Essa etapa acontece em períodos de 5 minutos para as ruas principais, visando dar uma resposta mais rápida quando uma dessas vias é pesquisada.

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. Nessa etapa também são excluídos os *tweets* de autores confiáveis caso o usuário opte por fazer uma pesquisa somente por autores não confiáveis.

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) a palavra modificadora. O sistema classifica um *tweet* de acordo com a quantidade de palavras negativas e positivas. Por exemplo, no *tweet*: “Avenida Brasil encontra-se sem lentidão.”, apesar da palavra lentidão ter uma conotação negativa, a palavra modificadora “sem” muda a conotação para positiva e a contagem para esse *tweet* é positiva, o tornando um *tweet* positivo. Caso o *tweet* fosse: “Sem carros na rua, trânsito livre.” como a palavra modificadora “sem” encontra-se a uma distância maior do que duas palavras (nesse caso são 4) da palavra positiva “livre”, a contagem é positiva e

o *tweet* é considerado positivo.

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.”. Exemplo: vamos supor que o usuário realize uma consulta sobre a Linha Vermelha e no momento da consulta havia 2 *tweets* sobre essa via conforme a tabela a seguir (tabela 1).

Tabela 1. Exemplo do Caso 2

Análise Estática	Hora
Positivo	10:30
Negativo	10:10

Neste caso há um conflito na classificação dos *tweets*. Como o intervalo de tempo entre os *tweets* conflitantes é maior do que 15 minutos o *tweet* negativo é desconsiderado e o sistema retorna a seguinte mensagem: “Não 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.”. Exemplo: vamos supor que o usuário realize uma consulta sobre a Linha Vermelha e no momento da consulta havia 3 *tweets* sobre essa via conforme a tabela abaixo (tabela 2).

Tabela 2. Exemplo do Caso 3

Análise Estática	Hora
Positivo	10:20
Positivo	10:15
Negativo	10:10

Neste caso há um conflito na classificação dos *tweets*. Como o intervalo de tempo entre os *tweets* conflitantes é menor do que 15 minutos e há mais *tweets* positivos (2) do que negativos (1) o sistema retorna a seguinte mensagem: “Provavelmente não 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. Caso este seja positivo, o sistema retorna a seguinte mensagem: “Provavelmente não está engarrafado.”. Caso contrário a seguinte mensagem é retornada: “Provavelmente está engarrafado.”. Exemplo: vamos supor que o usuário realize uma consulta sobre a Linha Vermelha e no momento da consulta havia 4 *tweets* sobre essa via conforme a tabela abaixo (tabela 3).

Tabela 3. Exemplo do Caso 3

Análise Estática	Hora
Positivo	10:20
Positivo	10:18
Negativo	10:16
Negativo	10:14

Neste caso há um conflito na classificação dos *tweets*. Como o intervalo de tempo entre os *tweets* conflitantes é menor do que 15 minutos a quantidade de *tweets* positivos e negativos é a mesma (2) o sistema considera a classificação do *tweet* mais recente (10:20) como certa e retorna seguinte mensagem: “Provavelmente não 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. Exemplo de Uso

Para facilitar a avaliação da ferramenta, criamos uma interface *desktop*. 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 10/07/2013 às 22h58min sobre as condições da Avenida Presidente Vargas utilizando apenas as fontes oficiais, ou seja, autores confiáveis. (Figura 4), todos os *tweets* (Figura 5), somente fontes não oficiais (Figura 6) e o estado real da via (Figura 7).

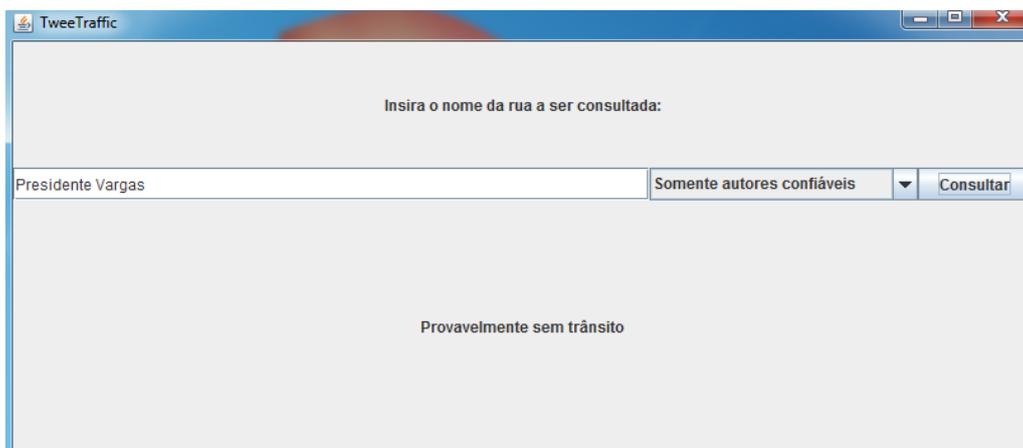


Figura 4. Previsão utilizando apenas fontes confiáveis (oficiais)



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

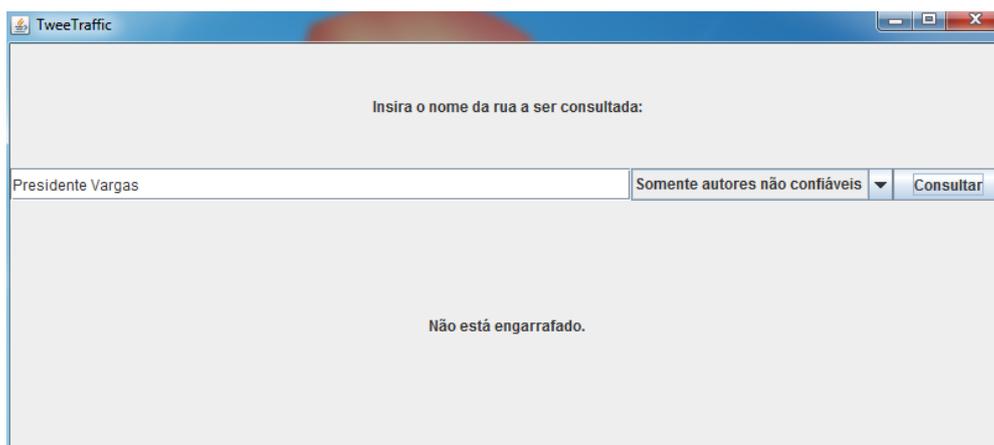


Figura 6. Previsão utilizando apenas fontes não confiáveis (não-oficiais)



Figura 7. Condição real, mostrada pela Câmera da CET-Rio²: Via totalmente livre.

6. Trabalhos Correlatos

Nesta seção será feita uma comparação com os trabalhos correlatos: *Google Maps Brasil* [2012], *Waze* [2012], *Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application* [Endarnoto, 2011], *Observatório do Trânsito: sistema para detecção e localização de eventos de trânsito no Twitter* [Ribeiro et al., 2012], a abordagem de [Ribeiro et al., 2012] e o *Twittraffic: Uma Plataforma de Monitoração, Visualização e Identificação de Ocorrências no Trânsito* [Benevenuto et al., 2013].

² <http://www.rio.rj.gov.br/web/riotransito>

6.1. Sistema de informações geográficas sobre o trânsito a partir da colaboração entre as pessoas pelo *Twitter*

A abordagem de [Wienandts, 2011] apresenta uma proposta para mapeamento do trânsito na cidade de Porto Alegre utilizando o *Twitter*. Essa proposta é bastante semelhante a do nosso projeto, ambas fazem o uso do *Twitter* como base de dados para determinar as condições de trânsito nas ruas de uma cidade.

Apesar dos projetos possuírem o mesmo objetivo e utilizarem a mesma fonte de dados eles apresentam diferentes implementações. O nosso projeto busca periodicamente por *tweets* relacionados ao trânsito dentre um subconjunto de usuários predeterminados. Já em [Wienandts, 2011] há a criação de um perfil no *Twitter* para centralizar as informações do trânsito. Dessa maneira os demais usuários participam de modo colaborativo, enviando mensagens sobre o trânsito através do *Twitter*.

6.2. *Google Maps* função trânsito

O Google Maps Brasil [2012], um dos mapas mais completos e atualizados do mundo, desde 2008, apresenta uma função que mostra, em tempo real, as condições de trânsito nas principais vias de algumas cidades brasileiras. Assim que foi lançado, ele cobria as cidades de São Paulo, Belo Horizonte e Rio de Janeiro, atualmente, ele já está em Recife e Curitiba.

Além de apresentar as condições atuais do trânsito nos corredores capitais dessas cidades, ele mostra também um histórico do tráfego deles com base em cada dia da semana. Essa funcionalidade é o grande diferencial entre o TweeTraffic e a função trânsito do Google Maps Brasil [2012].

6.3. *Waze*

O Waze [2012] é um aplicativo para dispositivos móveis que tem como proposta ser um GPS comunitário. Ele funciona como um GPS normal: oferece informações de rotas e ensina como chegar ao local desejado através do mapa.

No entanto, ele possui outra característica: ao longo do caminho, você pode compartilhar as informações sobre as condições de trânsito e utilizar as informações disponibilizadas por outros usuários. Através dele é possível consultar ou publicar informações diversas sobre as vias como: localização de blitz policiais, congestionamentos, acidentes e obras.

Um diferencial é que ele apenas utiliza os dados postados pelos usuários do Waze, não incluindo fontes oficiais.

6.4. *Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application*

A abordagem de [Endarnoto, 2011] apresenta uma proposta para divulgar as condições de trânsito das ruas da cidade de Jakarta, Indonésia. Segundo o artigo a polícia de Jakarta utiliza o *Twitter* para divulgar notícias sobre as condições de trânsito locais. O objetivo desse artigo é extrair essas informações divulgadas pela polícia e apresentá-las em um mapa através de um aplicativo para plataforma móvel Android.

Como podemos observar essa proposta possui muitas semelhanças com a nossa, entres elas o uso do *Twitter*, como base de dados para extrair as condições de trânsito e

o uso de um aplicativo móvel para exibi-las. Entretanto também há algumas diferenças, apesar do *Twitter* ter sido utilizado como base de dados, apenas as informações publicadas pela polícia local são usadas como fontes para determinar as condições de trânsito das vias. Além disso, as informações sobre as condições de trânsito são apresentadas de forma gráfica através de um mapa. As ruas com congestionamento são exibidas na cor vermelha, ruas com tráfego intenso são exibidas na cor amarela e ruas sem trânsito são exibidas na cor verde.

6.5. Observatório do Trânsito: sistema para detecção e localização de eventos de trânsito no *Twitter*

Esse artigo de [Ribeiro et al., 2012] descreve um sistema de mineração de textos aplicado ao *Twitter* procurando por padrões de texto relacionados ao trânsito. O sistema utiliza uma interface Web para apresentar as informações coletadas para os usuários. O sistema permite definir o local de onde se deseja obter as condições de trânsito, no entanto, neste artigo foi utilizada a cidade de Belo Horizonte.

O processo para a extração das informações de trânsito é dividida em quatro etapas: pré-processamento do texto, identificações do evento, localização por casamento exato e enriquecimento da localização por casamento aproximado.

Na primeira etapa são removidos acentos, links, citações a perfis de *Twitter* e informações de outras cidades. Na segunda etapa é feita a filtragem para eliminar os *tweets* que não possuam ligação com as condições de trânsito. Na terceira etapa é responsável por identificar a região sobre a qual o *tweet* se refere buscando por referências a ruas e bairros. Na última etapa é utilizada a localização da etapa anterior para buscar informações de locais próximos. A motivação para essa estratégia vem do fato de o *tweet* ser limitado a 140 caracteres e os textos sobre trânsito serem, em geral, bastante concisos e citarem locais próximos.

Para auxílio da terceira etapa é usado um gazetteer para a geolocalização dos *tweets*. Desse modo a informação coletada é disponibilizada na forma de um mapa para os usuários. Esse é o principal diferencial para o TweeTraffic, que retorna apenas a informação sobre a via pesquisada de maneira rápida e direta.

6.6. Twittraffic: Uma Plataforma de Monitoração, Visualização e Identificação de Ocorrências no Trânsito

O trabalho desenvolvido por [Benevenuto et al., 2013], é uma plataforma que monitora, serve de visualização e de identificação de ocorrências no trânsito, assim como esse trabalho, ela só usa o *Twitter* como fonte de dados. Porém a fase de busca ocorre de forma diferente. Em vez de pesquisar o nome da rua no *Twitter*, [Benevenuto et al., 2013] busca por termos relacionados a problemas no trânsito como “agarrado”, “garrado”, “parado” e “congestionado” e depois se usa do processamento de linguagem natural para identificação de entidades de possíveis localizações que são validadas através da API do Google Maps.

Há uma diferença também na forma de como a informação coletada e processada é mostrada para o usuário, no caso de [Benevenuto et al., 2013], ela é colocada em um mapa, onde apresenta-se geograficamente as ocorrências de trânsito e suas intensidades em forma de alerta, enquanto no TweeTraffic a resposta é em forma de texto e a informação fornecida é somente da via consultada.

Ao pesquisarmos trabalhos correlatos podemos observar que o uso do *Twitter* como fonte de dados para auxiliar pessoas a contornar o problema crescente do congestionamento nas grandes cidades é um campo que vem sendo estudado.

Ao comparar nosso projeto com os demais identificamos diferenças na forma de como as condições de trânsito nas vias das cidades são exibidas. Ao contrário do nosso, que exibe um texto informando a condição atual da rua consultada, os demais disponibilizam essa informação de uma forma gráfica utilizando mapas. Isso se deve, em grande parte, ao fato de o foco do nosso trabalho ser desenvolver um método de extração e análise de informações sobre o trânsito a partir do *Twitter*.

Tabela 3. Tabela comparativa entre trabalhos correlatos.

	Fonte de dados	Exibição da informação	Tipo de fontes utilizadas
Sistema de informações geográficas sobre o trânsito a partir da colaboração entre as pessoas pelo <i>Twitter</i>	<i>Twitter</i>	Mapa – interface web	Qualquer pessoa com perfil no <i>Twitter</i> sem separação entre fontes confiáveis e não confiáveis.
<i>Google Maps</i> função trânsito	Usuários do aplicativo	Mapa - interface web e móvel	Dados do GPS de usuários do aplicativo de forma anônima.
<i>Waze</i>	Usuários do aplicativo	Mapa - interface web e móvel	Qualquer pessoa com perfil no <i>Waze</i> .
<i>Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application</i>	<i>Twitter</i>	Mapa - interface móvel	Polícia de Jakarta, não se utiliza de outros usuários do <i>Twitter</i> .
Observatório do Trânsito: sistema para detecção e localização de eventos de trânsito no <i>Twitter</i>	<i>Twitter</i>	Mapa - interface web	Qualquer pessoa com perfil no <i>Twitter</i> sem separação entre fontes confiáveis e não confiáveis.
Twittraffic: Uma Plataforma de Monitoração, Visualização e Identificação de Ocorrências no Trânsito	<i>Twitter</i>	Mapa – interface web	Qualquer pessoa com perfil no <i>Twitter</i> sem separação entre fontes confiáveis e não confiáveis.

7. Avaliação

A avaliação foi dividida em duas etapas. A primeira composta por uma avaliação qualitativa, através de um estudo de caso para verificar o grau de utilidade e usabilidade da aplicação (seção 7.1). A segunda, através da monitoração do trânsito em diferentes horários, a verificação do resultado dado pelo TweepTraffic e o cálculo da precisão, acurácia (seção 7.2) e de um estudo mais focado nas ruas principais, com o cálculo da precisão, acurácia e revocação (seção 7.3).

7.1 Estudo de Caso

Os participantes utilizaram o aplicativo durante a semana de 19 e 26 de agosto de 2012 e a amostra deste estudo era composta por 7 pessoas. Os testes foram realizados por indivíduos de ambos os gêneros que possuíam, em sua maioria, idade entre 22 e 25 anos e níveis de escolaridade diversos e todos eles tinham alguma experiência com o *Twitter*. Todos os usuários acessavam o *Twitter* pelo menos uma vez por semana. Todos os usuários acharam entre fácil e muito fácil a checagem de trânsito, 80% deles o usariam durante o cotidiano e todos eles dirigiam pelo menos uma vez na semana, sendo que a maioria (60%) dirigia três vezes durante a semana. Todos os usuários informaram que pegam trânsito intenso pelo menos uma vez durante a semana e que o sistema no máximo de duas vezes não foi capaz de informar as condições de trânsito da via pesquisada.

7.2 Acurácia e Precisão

Durante o período 20 a 24 de agosto o trânsito de algumas ruas foi monitorado em tempo real através das imagens de câmeras de trânsito da [CET-Rio 2012], nos horários: 8h, 12h e 18h. As vias monitoradas foram Linha Vermelha, Avenida Brasil, Avenida Nossa Senhora de Copacabana, Avenida Rio Branco e Praça da Bandeira. Para cada observação, realizamos as consultas no TweepTraffic: uma considerando apenas as fontes confiáveis e a outra considerando todas as fontes, como mostra a Figura 8.

Dia	Horário	Imagem CET-Rio	TweepTraffic: Uso de todas as fontes	TweepTraffic: Uso apenas das fontes confiáveis
22/08/12	12:00	 <p>Condição real, extraída a partir da análise da imagem: “Não está engarrafado”</p>	Provavelmente não está engarrafado	Não está engarrafado

Figura 8. – Exemplo de uma análise realizada, para a via com nome Linha Vermelha

Para verificar a eficiência do sistema foram utilizadas duas métricas: precisão (fórmula 1) e acurácia (fórmula 2).

$$\text{Precisão} = \frac{\text{Positivo verdadeiro}}{\text{Positivo verdadeiro} + \text{positivo falso}}$$

Fórmula 1. Precisão

$$\text{Acurácia} = \frac{\text{Positivo verdadeiro} + \text{negativo verdadeiro}}{\text{Positivo verdadeiro} + \text{positivo falso} + \text{negativo verdadeiro} + \text{negativo falso}}$$

Fórmula 2. Acurácia

Consideramos como **positivo verdadeiro** a situação na qual a imagem coletada indicava trânsito na via e a consulta feita no TweeTraffic retornava a mensagem “Está engarrafado”. **Positivo falso** é a situação na qual a imagem coletada indicava trânsito na via e a consulta feita no TweeTraffic retornava a mensagem “Não está engarrafado”. Já **negativo verdadeiro** é a situação na qual a imagem coletada indicava que não havia trânsito na via e a consulta feita no TweeTraffic retornava a mensagem “Não está engarrafado”. **Negativo falso** a situação na qual a imagem coletada indicava que não havia trânsito na via e a consulta feita no TweeTraffic retornava a mensagem “Está engarrafado”.

Para facilitar o entendimento, esses quatro conceitos foram resumidos na tabela 4.

Tabela 4. Resumo de conceitos

		Condições determinadas pela captura de tela	
		Positivo	Negativo
Resultado do teste	Positivo	Positivo verdadeiro	Positivo falso
	Negativo	Negativo falso	Negativo verdadeiro

Os resultados estão na Tabela 5.

Tabela 5. Resultados

	Precisão	Acurácia
Todas as fontes	0,36	0,53
Fonte confiável	0,71	0,79

Com base nos resultados obtidos na Tabela 5, podemos observar um melhor desempenho de resultados ao se utilizar apenas fontes confiáveis. O uso de fontes não-confiáveis não pode ser ignorado, por ser referente ao que o sensoriamento humano (feito através dos cidadãos) nos indica. Muitas vezes, o cidadão comum pode emitir uma notícia mais rapidamente do que um meio oficial, ou se referenciar a uma rua ou via onde não há monitoramento constante. A acurácia de mais de 50% não é desprezível neste caso.

7.3 Avaliação de Vias Principais

Devido aos resultados anteriores, resolvemos estudar mais detalhadamente a comunicação usando o *Twitter* - bem como a resposta do ambiente na inferência da situação de trânsito - nas principais vias da cidade, vias onde há grande circulação de veículos, usualmente ocorrem congestionamentos e ligam pontos chave da cidade. Durante o período de 25 de agosto a 25 de setembro de 2013 foram coletados 1.423.685 *tweets* das ruas Av. Presidente Vargas, Av. Rio Branco, Av. Brasil e Av. das Américas, dos quais 198.243 foram provenientes de fontes confiáveis. Essas ruas foram escolhidas por fazerem parte do grupo de ruas principais, logo há mais *tweets* sobre elas e por possuírem câmeras de trânsito com acesso via internet para checarmos suas condições reais de trânsito no momento da consulta.

A tabela 6 mostra a média de *tweets* por dia, o que evidencia a maior quantidade em todas as ruas de *tweets* originados de fontes não confiáveis.

Tabela 6. Média de tweets por dia

	Fontes confiáveis	Fontes não confiáveis	Todas as fontes
Av. Brasil	32,4	738,6	771
Av. das Américas	91,6	213	304,6
Av. Presidente Vargas	427,2	733,3	1160,5
Av. Rio Branco	96	185,5	281,5

Para avaliar os resultados obtidos pela ferramenta, foram escolhidas três métricas:

→ Precisão: proximidade entre os valores obtidos pela repetição do processo de mensuração. Para calcularmos os valores nos utilizamos da fórmula 1.

→ Revocação: fração dos *tweets* recuperados que foram relevantes para a determinação da condição do trânsito de determinada rua. A fórmula 3 mostra como calcular a revocação.

$$\text{Revocação} = \frac{\text{positivo verdadeiro}}{\text{positivo verdadeiro} + \text{negativo falso}}$$

Fórmula 3. Revocação

→ Acurácia: proximidade da medida ao verdadeiro valor da variável. Para calcularmos os valores nos utilizamos da fórmula 2.

A tabela 7 mostra a média por dia de cada uma das métricas durante o período da avaliação.

Tabela 7. Médias por dia

	Fontes confiáveis	Fontes não confiáveis	Todas as fontes
Precisão	0,4175	0,25	0,2925
Revocação	0,75	0,375	0,625
Acurácia	0,542	0,225	0,275

Como pode ser observado, as fontes confiáveis sempre possuem os melhores índices, porém não podemos só usar *tweets* provenientes dessas fontes pois muitas vezes não há *tweets* delas. O alto índice de revocação mostra que quase a maioria dos *tweets* coletadas são relevantes para a determinação da condição de trânsito das ruas.

$$\text{Taxa de acerto} = \frac{\text{quantidade de vezes certas de determinada resposta}}{\text{quantidade de vezes total de determinada resposta}}$$

Fórmula 4. Taxa de acerto

A tabela 8 mostra a taxa de acerto do aplicativo utilizando-se somente de fontes confiáveis, somente de fontes não confiáveis e todas as. A taxa de acerto (fórmula 4) é a porcentagem de vezes que o programa foi capaz de responder corretamente sobre as condições de trânsito da via e foi calculada de acordo com a resposta do aplicativo que poderiam ser:

- E: engarrafado.
- PE: provavelmente engarrafado.
- L: livre.

→ PL: provavelmente livre.

Nesse caso foi utilizado um processo similar ao do cálculo da acurácia e precisão da seção 7.2, onde o resultado do programa era comparado com as imagens obtidas pelas câmeras de trânsito da [CET-Rio 2012].

Tabela 8. Taxa de Acerto

Vias	Fontes confiáveis				Fontes não confiáveis				Todas as fontes			
	E	PE	L	PL	E	PE	L	PL	E	PE	L	PL
Av Brasil	0,5	0	0	0,62	0,5	0	0	0,08	0,5	0	0	0,67
Av das Américas	0	0	0	0,03	0	0	0	0	0	0	0	0,03
Av Presidente Vargas	0,1	0	0,11	0,33	0,1	0	0,11	0,05	0,1	0	0,05	0,28
Av Rio Branco	0,5	0	0,04	0,3	0,5	0	0,04	0	0,5	0	0,04	0,25
TOTAL	0,15	0	0,02	0,25	0,15	0	0,02	0,04	0,2	0	0,01	0,28

A taxa de acerto menor nas fontes não confiáveis era um resultado esperado, uma vez que é mais difícil conseguir interpretar um *tweet* postado por essas fontes e pela falta de comprometimento com a real situação que as fontes confiáveis (oficiais) possuem.

A baixa taxa de acerto no geral pode ser explicada pela longa extensão das vias pesquisadas e pelo fato de grande parte delas possuírem mão dupla, ou seja, alguns *tweets* informavam que o trânsito estava livre em determinados quilômetro da via e em determinada direção, mas a imagem que possuíamos da câmera mostrava trânsito intenso em outra região.

O valor zero encontrado nas colunas PE (provavelmente engarrafado) ocorreu pois o aplicativo não respondeu em nenhuma das ocasiões que as vias estariam provavelmente engarrafadas.

8. Conclusão

O trânsito na cidade do Rio de Janeiro se torna cada vez mais caótico com o passar dos anos, através da popularização da internet e do crescimento das redes sociais no Brasil, principalmente do *Twitter*, o TweeTraffic visa melhorar a vida do motorista carioca, informando através da coleta de informações provenientes dessa rede social as reais condições de trânsito das principais vias dessa cidade.

De acordo com os resultados obtidos nas três avaliações, podemos observar um interesse das pessoas em um aplicativo como esse e um grau de acerto satisfatório por parte dele. Vale ressaltar também que os resultados obtidos ao se fazer pesquisas usando todas as fontes não foram tão bons. Uma das explicações é a dificuldade em se distinguir ruas com o mesmo nome de ruas cariocas, mas que sejam de outra cidade. O pouco uso do *Twitter* para informar o trânsito em ruas menores torna difícil a verificação das condições de trânsito nessas vias. O mesmo problema ocorre com vias extensas e de mão dupla, nas quais é um obstáculo descobrir em que região e sentido da rua se encontram os problemas. Outra dificuldade encontrada foi o fato de os usuários do *Twitter* geralmente não georeferenciarem seus *tweets*, tornando impossível descobrir a real localização dos usuários no momento de filtrá-los.

Outro obstáculo encontrado foi que pessoas não costumam postar no *Twitter* *tweets* de trânsito bom, em grande maioria são *tweets* negativos que aparecem em nossas buscas.

Como trabalho futuro está o aprimoramento da filtragem para a melhor identificação da região do problema, possibilitando informar qual o sentido e quilômetro das grandes vias em que você se encontra para que a pesquisa retorne uma condição mais próxima da realidade, o aumento do dicionário sobre o trânsito para que mais acidentes sejam identificados e dos autores confiáveis para ocorra uma melhora na taxa de acerto e um aprimoramento na análise estática através do uso de um dicionário de análise de sentimentos. Outro trabalho futuro é fazer uma coleta de dados de uma via principal ao longo de um dia inteiro e comparar os resultados inferidos pelo programa com as já conhecidas horas do *rush*.

Existe ainda a possibilidade desse estudo ser estendido para outros cenários como sentimentos da população a respeito de um político, a implementação de uma nova lei, uma empresa ou um produto, fazendo apenas algumas pequenas modificações mas mantendo a ideia principal.

Agradecimentos. Gostaríamos de agradecer à FAPERJ através do Programa Jovem Cientista do Nosso Estado (processo: E_23/2013), ao CNPq pelo Edital Universal (processo: 487239/2012-1) e ao CENPES/Petrobras por bolsa de estudo.

Referências

Brazil's Social Networking Activity Accelerates in the Past Year Fueled by Facebook Growth - comScore, Inc. Disponível em: <http://www.comscore.com/Press_Events/Press_Releases/2012/3/Brazil_s_Social_Networking_Activity_Accelerates_in_the_Past_Year>. Acesso em: 7 set. 2012.

Cresce uso de banda larga no Brasil. Disponível em: <<http://www.ibope.com.br/pt-br/relacionamento/imprensa/releases/paginas/cresce-uso-de-banda-larga-no-brasil.aspx>>. Acesso em: 7 set. 2012.

Endarnoto, S. K. (2011) "Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application" in Electrical Engineering and Informatics (ICEEI), 2011

Google Maps (2012), <http://maps.google.com.br>, Setembro

Lauand, B. P., 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), Anais do XXXIII Congresso da Sociedade Brasileira de Computação. pp:1507-1512. SBC.

Em 2 anos de PNBL, número de conexões 3G cresce 270%, afirma secretário-executivo - Ministério das Comunicações. Disponível em: <<http://www.mc.gov.br/dia-a-dia/36-noticias/25354-em-2-anos-de-pnbl-numero-de-conexoes-3g-cresce-270-afirma-secretario-executivo>>. Acesso em: 7 set. 2012.

PDTU - Plano Diretor de Transporte Urbano da Região Metropolitana do Rio de Janeiro. Disponível em: <<http://www.pdtu.rj.gov.br/indexini.php>>. Acesso em: 8 jul. 2012.

Portal do Trânsito » Tempo médio gasto no trânsito no Brasil cresce 20% em 7 anos. Disponível em: <<http://www.portaldotransito.com.br/noticias/tempo-medio-gasto-no-transito-no-brasil-cresce-20-em-7-anos.html>>. Acesso em: 7 set. 2012.

Relevância » Blog Archive » Quase 70% dos usuários de Internet do Brasil usam Redes Sociais. Disponível em: <<http://relevancia.com.br/quase-70-dos-usuarios-de-internet-do-brasil-usam-redes-sociais/>>. Acesso em: 7 set. 2012.

Ribeiro, S. S., Rennó, D., Gonçalves, T. S., Davis, C. A., Meira W. (2012) " Observatório do Trânsito: sistema para detecção e localização de eventos de trânsito no Twitter " in Simpósio Brasileiro de Bancos de Dados - SBBDD, 2012

Schmidt, L., Ferreira, J., Peixoto, V., Oliveira, R., R., Benevenuto, F. (2013) "Twittraffic: Uma Plataforma de Monitoração, Visualização e Identificação de Ocorrências no Trânsito" in Proceedings of the XXXIX Seminário Integrado de Software e Hardware - Semish, 2013.

Waze (2012) "Free GPS Navigation with Turn by Turn - Waze", <http://www.waze.com>