



Programa de Pós-Graduação em Engenharia de Sistemas

PROJETO DE DISSERTAÇÃO



Recife, Fevereiro de 2016.



Universidade de Pernambuco (UPE)
Escola Politécnica de Pernambuco (POLI)
Instituto de Ciências Biológicas (ICB)

SISTEMA DE SUGESTÃO DE ROTEAMENTO RODOVIÁRIO DE CARGAS CONSIDERANDO HISTERESE E FATORES SÓCIO-AMBIENTAIS

Mestrando: Eng. Othon Luiz Teixeira de Oliveira
Orientador: Prof. Dr. Fernando Buarque de Lima Neto

Projeto de qualificação apresentado ao
Programa de Pós-Graduação em Engenharia
de Sistemas
Área de concentração: **Cibernética.**

Banca de qualificação:

Prof. Dr. Carmelo José A. Bastos Filho.....Engenharia de Sistemas/POLI/UPE

Prof. Dra. Rita de Cássia M. Nascimento.....Engenharia de Sistemas/POLI/UPE

Recife, Fevereiro de 2016.

*“Quem escolheu a busca
não pode recusar a travessia”
(Guimarães Rosa)*

Resumo

O Transporte de cargas, que atravessa as regiões metropolitanas das grandes cidades brasileiras o fazem principalmente pelas rodovias federais. Essas rodovias estão constantemente congestionadas e têm recebido aumento expressivo de novos veículos a cada ano.

No entorno desses trechos urbanos têm crescido desordenadamente comunidades que demandam por políticas sociais que atendam às suas necessidades. Para reivindicar, dos entes públicos, essas comunidades bloqueiam as rodovias, aumentando a pressão sobre os congestionamentos. Em alguns trechos o traçado das rodovias está próximo a morros e florestas ficando suscetíveis às intempéries climáticas. Essas variáveis impõem constantes paralisações às rodovias, representando atrasos nas entregas, custos adicionais às empresas e prejuízos à competitividade nacional.

Propor novas soluções, que venham minimizar esses constrangimentos é condição 'sine qua non', para mitigar o que vem sendo chamado de "Custo Brasil". Um sistema de suporte a decisão, ao roteamento de cargas, que utilize informações vindas de sítios eletrônicos como google-maps, base de dados históricas da Polícia Rodoviária Federal, das redes sociais e das condições socio-ambientais estão disponíveis em forma de dados na Internet e nos servidores locais dessas instituições, podendo ser processados numa plataforma auto-adaptável.

Assim, esse estudo tem por objetivo desenvolver um sistema de suporte a decisão para a problemática crescente da Logística de Cargas, antecipando prováveis eventos, que possam interferir no fluxo normal das rodovias brasileiras, propondo assim uma alternativa ao traçado de rotas determinísticas e incorporando novas opções de rotas inspiradas na predição de eventos futuros.

Palavras-chave: Mineração de dados, Bases de dados históricas, Redes sociais, Logística de transportes, Roteamento

Lista de Abreviações e Siglas

| | |
|-----|------------------------------------|
| DM | <i>Data Mining</i> |
| BG | <i>Big Data</i> |
| PSO | <i>Particle Swarm Optimization</i> |
| ACO | <i>Ant Colony Optimization</i> |
| FSS | <i>Fish School Search</i> |
| B | <i>Byte</i> |
| KB | <i>kyloByte</i> |
| MB | <i>MegaByte</i> |
| GB | <i>GigaByte</i> |
| TB | <i>TeraByte</i> |
| PB | <i>PetaByte</i> |
| EB | <i>ExaByte</i> |
| ZB | <i>ZettaByte</i> |
| YB | <i>YottaByte</i> |

Lista de Figuras

| | | |
|-----|---------------------------------|----|
| 2.1 | Etapas da metodologia | 13 |
|-----|---------------------------------|----|

Lista de Tabelas

| | | |
|-----|---------------------------------|----|
| 2.1 | Cronograma – 12 meses | 13 |
|-----|---------------------------------|----|

Sumário

| | | |
|----------|---|-----------|
| 1 | Problema, Motivação e Objetivo | 9 |
| 1.1 | Problema | 9 |
| 1.2 | Motivação | 10 |
| 1.3 | Objetivo Geral | 10 |
| 1.3.1 | Objetivos Específicos | 10 |
| 1.4 | Reflexão sobre as tecnologias | 11 |
| 2 | Metodologia | 12 |
| 2.1 | Metodologia Proposta | 12 |
| 2.2 | Cronograma | 13 |
| 3 | Resumo da Proposta, Discussão | 14 |
| 3.1 | Resumo da Proposta | 14 |
| 3.2 | Discussão | 14 |
| | Referências Bibliográficas | 15 |
| | Referências Bibliográficas | 15 |

O Projeto

*“E se o mundo não corresponde
em todos os aspectos a nossos desejos,
é culpa da ciência ou dos que querem
impor seus desejos ao mundo?”
(Carl Sagan)*

1

Problema, Motivação e Objetivo

1.1 Problema

A partir do início do século XXI o mundo digital, especialmente a Internet, conheceu sua primeira grande crise (1). As empresas ligadas a esse mundo, conhecidas como PontoCom, para sobreviverem, adaptaram-se à Internet abrindo suas estruturas, desde então houve um *boom* nesse segmento. As informações geradas e disponibilizadas à Internet, nos mais recentes anos, representam 90% de tudo o que já foi criado nos anos anteriores pela humanidade ou desde que nossa civilização aprendeu a guardar informação.

Para armazená-los, seriam necessários milhões de computadores; se fosse possível dispô-los num único *DataCenter*, esses ocupariam uma área do tamanho do estado de São Paulo.

Os dados produzidos pelo ser humano atualmente dobram a cada 5 anos; esses dados são desde artigos publicados, novas técnicas para os mais diversos problemas da vida humana e outros, ficando impossível serem armazenados pelo cérebro humano (2).

Com a chegada da Internet das Coisas (IoC), acrônimo de *Internet of Things* (IoT), a previsão é de que o número de informações dobre a cada 2 anos.

A Internet da Coisas pode ser entendida como “coisas conectadas às coisas”, que pode ser, o refrigerador comunicar-se com os alimentos ali depositados, que contenham uma etiqueta identificadora por rádiofrequência (Radio Frequency IDentification – RFID), podendo ter autonomia para enviar a um supermercado um rol de compras futuras.

Isso irá fazer com que os dados trafegados pela Internet tenham um crescimento exponencial. A essas informações circulantes dá-se o nome de *Big Data*. Um *Big data* é um conceito, na verdade são as coleções de tudo o que é disponibilizado na Internet, desde os dados dos *Data Centers* aos dados gerados pela Internet das Coisas. Essa enormidade de dados poderia ver a ser um problema de difícil solução se não houvessem abordagens computacionais com habilidade de extrair semântica bem como possibilidade de oferecer suporte a agentes decisores.

Uma instância do problema a ser tratado nessa pesquisa será a integração de bases heterogêneas de dados em um aplicação de suporte à decisão de logística de cargas rodoviárias.

1.2 Motivação

As rodovias federais que atravessam a Região Metropolitana do Recife (RMR) estão constantemente congestionadas, não apenas pela quantidade de veículos, mas por serem alvo de paralisações das mais diversas matizes, como protestos de trabalhadores, acidentes, buracos, intempéries naturais e outros tipos de paralisações. Em situações extremas poderiam paralisar até a produção das fábricas no seu entorno, por exemplo a Fiat - FCA (3). Esta será responsável por aproximadamente 1 000 caminhões cegonheiros nessas rodovias, quando do seu pico de produção (200 000 veículos/ano).

A RMR é a 5ª região mais populosa do Brasil, concentra 3.690.485 habitantes (dados de 2012) (4) em 14 municípios, além da RMR será considerada para a pesquisa a Zona da Mata Norte (ZMN) com 577.191 habitantes e a Zona da Mata Sul (ZMS) com 733.447 habitantes. Nessas regiões (RMR, ZMN e ZMS) a frota (automóveis particulares, ônibus, caminhões, motocicletas, tratores e outros veículos) foi contabilizada, em 2014, com 635.686 veículos (5).

O que acontece na região metropolitana do Recife, é frequentemente visto no entorno das principais cidades brasileiras. Por outro lado, câmeras de monitoramento de trânsito, redes sociais, aplicativos de celular e outros dispositivos, fornecem informações diárias sobre o que acontece nessas rodovias e no entorno delas, atualizando e alimentando bases de dados históricas, em repositórios espalhados pelos centros de monitoramento de trânsito.

A proposição de uma solução para resolver esse problemática requer várias etapas, para além da proposição de algumas técnicas de mineração dos dados. Propomos, nesse projeto, uma solução peculiar, ao enviar essa frota de caminhões por diversas rotas, escolhidas por critérios cientificamente estudados. Isso poderá ser de suma importância para solucionar a problemática do transporte de cargas, que poderá advir com a plena produção da FCA, e que não se aplica apenas à região metropolitana do Recife, mas a toda e qualquer fábrica do país. Permitirá fornecer toda informação que se faz necessária para acompanhar os caminhões na transposição dos obstáculos que possam surgir ao transitar por Pernambuco, conduzindo-os até seu destino de maneira segura e no menor tempo possível.

1.3 Objetivo Geral

Esse projeto de pesquisa tem como objetivo desenvolver um sistema de Suporte à Decisão para a problemática das retenções crescentes de logística de cargas rodoviárias. Para isto propomos uma solução multidisciplinar através da integração de diversas tecnologias disponíveis desde a análise dos dados históricos das rodovias até a situação cotidiana.

1.3.1 Objetivos Específicos

- Representar a problemática da logística de cargas em uma plataforma adaptável:
 - Está sendo desenvolvido uma plataforma adaptável que integre dados de redes sociais (Twitter) que analisa o contexto das rodoviárias através da mineração de textos que verifica palavras chaves como: protestos, acidentes e outras.
- Desenvolver um modelo preditivo dos fenômenos que envolvem as rodovias.
 - O modelo preditivo integra várias bases de dados, tais como: Polícia Rodoviária Federal – PRF, Batalhão de Polícia de Trânsito – BPRV e dados históricos do Instituto de Pesquisas Espaciais – INPE.

- Propor uma simulação interativa da estrutura com a dinâmica.
 - A simulação interativa é uma plataforma baseada na API do Google Maps

1.4 Reflexão sobre as tecnologias

Não existe uma técnica de mineração que generalise os mais diversos ambientes preditivos, mas sim um “pool” dessas técnicas onde uma complementa outra. As técnicas preditivas tradicionais que contemplam análise de grandes massas de dados como base heterogêneas são possíveis quando adaptadas para uma forma comparável à que foram inicialmente concebidas. Algumas técnicas de IA são altamente sensíveis a dados ausentes os “missing data” à dados com pouca consistência e outros tipos de dados comuns em bases mantidas sem um bom critério de inserção dos dados. Nesta pesquisa, bases heterogêneas foram integralizadas num única grande base, onde as variáveis independentes foram em sua maioria preservadas e/ou construídas novas, nas bases onde não haviam correspondência, respeitando a lógica do negócio. A variável dependente foi designada como **gargalo** e as variáveis independentes (ou explicativas) são:

| | |
|--------------|--|
| KM | <i>Numeração do Quilômetro</i> |
| BR | <i>Numeração da BR</i> |
| condPista | <i>Condição da Pista: Seca, Molhado, ...</i> |
| restVisibili | <i>Restrição de Visibilidade: Inexistente, Neblina, .., outros</i> |
| tipoAcident | <i>Tipo de Acidente: Atropelamento, Colisão, Paralisação,...</i> |
| tipoDano | <i>Tipo de Dano: Leve, Médio, Grave</i> |
| Município | <i>Localidade onde ocorreu</i> |
| Ano | <i>Ano que o acidente ocorreu</i> |
| Mês | <i>Mês que o acidente ocorreu</i> |
| Dia | <i>Dia que o acidente ocorreu</i> |
| Hora | <i>Hora que o acidente ocorreu</i> |

A base de dados da PRF, relativas a paralisações das vias, por motivos diversos, não haviam variáveis tais como (visibilidade, condições da via, gravidade paralisação e outras). Presumivelmente protestos são realizados com boa visibilidade, em condições de via rezoáveis e a gravidade da paralisação foi considerada leve.

As técnicas como Redes Neurais Artificiais (MLP), Árvores de decisão (CART), Regressão logística (MLR) nos forneçam uma visão generalizada dos fatores preponderantes quando os dados estão tratados.

- Redes Neurais Artificiais do tipo Multi Layer Perceptron – (MLP) têm capacidade de receber várias entradas ao mesmo tempo e distribuí-las de maneira organizada, além são simples de implementar e trazem resultados satisfatórios em grandes bases de dados.
- Árvores de decisão para classificar acidentes do tipo Classification and Regression Tree – (CART) foi empregue por Pakgohar et al no artigo “The role of human factor in incident and severity of road crashes based on the CART and LR regression a data mining approach” com nível de acurácia próximo aos 80%
- Regressão logística tipo Multinomial Logistic Regression – (MLR) fornece a possibilidade de aprofundamento em vários níveis de busca sendo a mais apropriada, já que Regressão logística não permite aprofundamento no espaço de busca.

2

Metodologia

2.1 Metodologia Proposta

O objetivo deste projeto é propor um sistema preditivo de Suporte à Decisão que dê apoio a um gestor quando decidir enviar uma frota de caminhões por determinada rodovia que apresente retenções crescentes de logística de cargas.

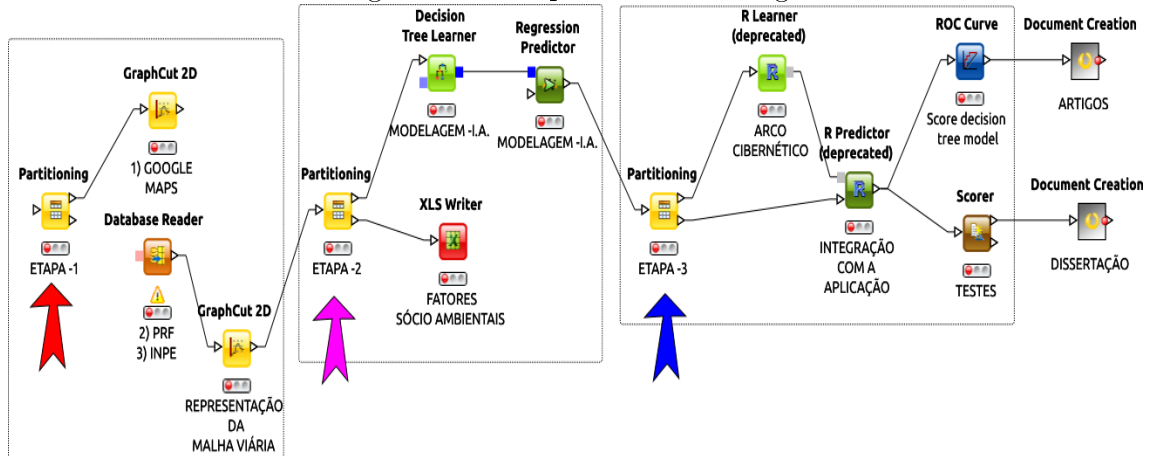
A malha viária está representada na Internet em mapas de bases vetoriais que pretendemos integrar ao nosso sistema de informações e também vamos incorporar as informações que a Polícia Rodoviária Federal (PRF) dispõe, que controla as rodovias BRs. Estas informações estão disponíveis em base de dados históricas nos servidores da PRF além de outras informações para complementar o sistema estão disponíveis na Internet sendo atualizadas pela PRF através de uma API aberta, esta pode ser configurável para se ligar facilmente ao nosso sistema.

Para representação da malha logística serão utilizadas bases vetoriais, por exemplo, o Google Maps. Para o desenvolvimento do modelo preditivo pretende-se utilizar bases de dados históricas, por exemplo: da Polícia Rodoviária Federal (PRF), do Centro de Previsões de Tempo e Estudos Climáticos (CPTEC - INPE) e possivelmente do Instituto Brasileiro de Geografia e Estatística (IBGE).

Para simulação interativa da estrutura preditiva com a dinâmica real serão capturados “feeds” de redes sociais, por exemplo pelo Twitter. Essa técnica fará um arco cibernético mantendo o sistema preditivo atualizado.

Propomos um plano que contemple 3 etapas, cada uma dividida em fases atinentes. A figura 3.1 ilustra essa metodologia descrita graficamente, onde as três etapas são representadas por retângulos.

Figura 2.1: Etapas da metodologia



O retângulo da esquerda é a etapa 1, onde há a representação da malha viária. A etapa 1 contempla as fases da coleta das bases vetoriais, das bases de dados históricas e proposição de um modelo de representação único.

O retângulo central é a etapa 2, que consiste nas fases de Identificação dos fatores sócio ambientais na bases históricas, a modelagem do sistema preditivo e aplicação das técnicas de IA. Propomos inicialmente a Regressão Logística e Árvores de decisão, testes iniciais e depuração do modelo.

A última etapa da nossa proposta metodológica é a etapa 3, esta conterá um arco cibernético construído com os dados de redes sociais, como por exemplo a API do Twitter, este “per si” fará com que o modelo preditivo seja retro-alimentado, mantendo-o, ao longo do tempo atualizado na perspectiva do usuário gestor. Isso pois, modelos preditivos com o passar do tempo tendem a desfazer-se. Uma proposta algorítmica para substituir a API das redes sociais poderá ser desenvolvida e testada numa fase complementar, o algoritmo Ant-Miner poderá vir a ser um candidato de adaptação.

Concluída as três etapas e com as informações geradas pelo modelo serão escritos artigos científicos pertinentes à pesquisa em lide e a escrita da dissertação.

2.2 Cronograma

Tabela 2.1: Cronograma – 12 meses

| Etapas/2016 | Fev | Mar | Abr | Mai | Jun | Jul | Ago | Set | Out | Nov | Dez | Jan/17 |
|------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------|
| Rev. da literatura. | x | x | x | x | x | x | x | — | — | — | — | — |
| Etapa – 1 | x | x | x | — | — | — | — | — | — | — | — | — |
| Etapa – 2 | — | — | — | x | x | x | — | — | — | — | — | — |
| Etapa – 3 | — | — | — | — | — | — | x | x | x | — | — | — |
| Escrita de artigos | — | — | — | — | — | — | — | — | x | x | x | — |
| Escrita da dissertação | — | — | — | — | — | — | — | — | x | x | x | — |
| Defesa | — | — | — | — | — | — | — | — | — | — | — | x |

3

Resumo da Proposta, Discussão

3.1 Resumo da Proposta

As rodovias federais brasileiras que cruzam as regiões metropolitanas das grandes cidades se apresentarão sempre como um gargalo no fluxo do transporte de cargas devido ao crescimento do número de veículos que nelas trafegam.

A solução proposta visa reduzir e dirimir os atuais gargalos burocráticos e tecnológicos para obtenção das bases de dados históricas de entidades públicas, os direitos autorais para utilização de APIs e de tecnologias específicas necessárias para apropriação via Internet. Entendemos ser normal esse cuidado por se tratar de informações de órgão que devem primar pelas informações de seus usuários. Mas tentaremos suprir demandas reais e importantes sem que isso represente algum risco à privacidade dos geradores de dados. Em suma, nossa proposta pretende mitigar o gargalo da logística de transporte de cargas, oferecendo uma solução possível à gestão de frotas de veículos que trafegam em rodovias, notadamente no caso do entrono metropolitano do Recife.

3.2 Discussão

Algumas propostas vêm sendo amplamente difundidas pelas mídias, tais como o arco metropolitano. Arcos metropolitanos, para além dos transtornos de se contruir um, são muito caros, requerem constantes manutenções e com o passar dos anos, com o crescimento populacional no seu entorno, tornam-se novamente um novo gargalo para o transporte de cargas.

Gerir como as rodovias são utilizadas é a maneira mais racional, elas estão aí para auxiliar no transporte de pessoas, mercadorias e para serviços, portanto é de todos e todos têm o dever de contribuir preservando-as e respeitando o direito dos outros.

Referências Bibliográficas

- 1 QUADROS, C. I. D. Dez Anos Depois do. *Intercom*, v. 1995, p. 65–69, 2005.
- 2 DILSIZIAN, L. E. S. E. Artificial intelligence in medicine and cardiac imaging: Harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. p. 1.
- 3 BNDES. Perspectivas do investimento, n. 2, out. 2013. *Perspectivas do Investimento 2014-2017*, p. 2, 2013.
- 4 BITOUN, J. et al. Região Metropolitana do Recife no Contexto de Pernambuco no Censo 2010. p. 25, 2012. Disponível em: <http://www.observatoriodasmetropoles.net/download/Texto/_BOLETIM/_RECIFE/_FINAL.pdf>.
- 5 IBGE, I. B. de Geografia e E. Região metropolitana do recife no contexto de pernambuco no censo 2010. 2014. Disponível em: <http://www.cidades.ibge.gov.br/painel/frota.php?codmun=261160&search=pernambuco%7Crecife%7Cinfogr%E1ficos:-frota-municipal-de-ve%C3%94culos%27&lang=_ES>.