ANÁLISE DE COMPORTAMENTO DO SISTEMA RODOVIÁRIO DO NORDESTE DO BRASIL

UMA ABORDAGEM DE MINERAÇÃO DE DADOS

*ANALYSIS OF THE BEHAVIOR OF THE ROAD SYSTEM IN THE NORTHEAST OF BRAZIL*

*A DATA MINING APPROACH*

Othon L. T. de Oliveira

Univerdidade de Pernambuco

Recife, Brasil

olto@ecomp.poli.br

Fernando B. de L. Neto

Universidade de Pernambuco

Recife, Brasil

Email: fbln@ecomp.poli.br

Resumo — As rodovias federais que atravessam a Região Metropolitana de algumas cidades 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. Propomos modelo de classificação para o comportamento das rodovias federais que atravessam o estado de Pernambuco na região Nordeste do Brasil, de modo que seja possível antecipar eventos que poderão causar constrangimentos, como retenção, redução de fluxo de tráfego. A fonte de dados dessa pesquisa provém da base de dados da Polícia Rodoviária Federal de Pernambuco (PRF) a partir de 2007 tendo considerado veículos, traçado da via e trechos da rodovia relacionados a acidentes, dentre outros. Com base nas informações obtidas, foi realizada uma Mineração de Dados utilizando a metodologia CRISP-DM para encontrar padrões comportamentais nas rodovias e em seu entorno. Foram empregados algoritmos de aprendizagem de máquina para classificação e regressão, sendo priorizadas, Árvores de Decisão e Redes Neurais. Os valores da área sob a curva Roc (AUC) obtidos foram acima de 0.7 que reflete um alto grau de confiabilidade. O modelo de predição proposto significa um avanço em termos de mobilidade e gestão do transporte de cargas, uma vez que possibilita antecipar eventos e comportamentos, favorecendo a escolha de rotas alternativas e ampliando o espaço temporal de escolha para determinadas rotas.

Palavras Clave – Modelo de predição; Mineração de dados; Tráfego em rodovias; Árvores de decisão, CRISP-DM.

Abstract — The federal highways that cross metropolitan regions are constantly congested, not be by number of vehicles, but because they are target of manifestations of the most diverse, such as, strike, road traffic accident, holes, natural inclemencies, and other types of downtime. In extreme situations could even paralyze factories and in their environment. We propose a classification model for the behavior of federal highways that cross the state of Pernambuco in Northeast of Brazil

Keywords – Prediction model; Data mining; Road traffic; Decision tree; CRISP-DM.

1. Introdução

O transporte de cargas que atravessa as regiões metropolitanas das grandes cidades brasileiras é realizado principalmente pelas rodovias federais. Essas rodovias frequentemente se encontram congestionadas em determinados dias/horários. Recentemente, tem sido contabilizado aumento expressivo de veículos a cada ano. O estado de Pernambuco, localizado na região Nordeste do Brasil, possuía, em 2015, uma frota de 2.765.521 de veículos, sendo que boa parte dessa frota trafega pelas rodovias que cruzam o estado. A Polícia Rodoviária Federal e outros órgãos de controle público atendem e registram os acontecimentos nessas rodovias. A solução para absorver parte dessas informações requer o estabelecimento de várias etapas, para a definição dos dias, horários e rotas, escolhidos por critérios cientificamente estudados para acompanhar os veículos e conduzindo-os até seu destino de maneira segura e no menor tempo possível.

1. Fundamentação teórica

No processo de extração do conhecimento (KDD), um dos importantes passos a ser considerado e a mineração de dados, que se caracteriza pela aplicação de algoritmos específicos para descoberta de padrões e/ou comportamentos em grandes bases de dados, também (repositórios). A mineração se distingue das técnicas estatísticas pelo fato de que não trabalha com dados hipotéticos, mas se apoia nos próprios dados para extrair os padrões[1].

O “CRoss Industry Standard Process for Data Mining” CRISP-DM é um processo de mineração de dados que descreve como especialistas nesse campo aplicam as técnicas de mineração para obter os melhores resultados[2]. O CRISP-DM é um processo recursivo, em que cada etapa deve ser revista até quando o modelo apresentar os resultados satisfatórios, preliminarmente definidos. O Analista de Dados ou o Cientista de Dados é o profissional que acompanha e executa o processo.

O contexto da aplicação do CRISP-DM[2] é guiado desde o nível mais genérico até o nível mais especializado, sendo normalmente explicado em quatro dimensões:

* O domínio da aplicação - a área específica que o projeto de mineração de dados acontece;
* O tipo de problema - descreve as classes específicas do objetivo do projeto de mineração de dados;
* Os aspectos técnicos - cobrem as questões específicas como os desafios usualmente encontrados durante o processo de mineração de dados;
* As ferramentas e técnicas - dimensão específica que cada ferramenta/técnica de mineração de dados é aplicada durante o projeto.

A tabela a seguir resume e exemplifica essas dimensões no contexto de aplicação do CRISP-DM.

Tabela 1- Mineração de dados -- contexto de aplicação[2]

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Dimensão** | | |
| Domínio da aplicação | Tipo de problema | Ferramenta e técnica |
| **Exemplo** | Modelo de resposta | Descrição e sumarização | Clementine |
| " | Predição | Segmentação | MineSet |
| " | --- | Descrição do conhecimento | Arvore de decisão |
| " | --- | Classificação | --- |
| " | --- | Analise de dependências | --- |

A aplicação das técnicas de mineração de dados identifica padrões ocultos nos dados, inacessíveis pelas técnicas tradicionais, como por exemplo, consultas a banco de dados, técnicas estatísticas, dentre outras. Além disso, possibilita analisar um grande numero de variáveis simultaneamente, o que não acontece com o cérebro humano[3]. Fayyad[4] destaca a natureza interdisciplinar do KDD que contempla a intersecção de campos de pesquisa tais como Aprendizado de Maquina (Machine Learning), Reconhecimento de Padrões, Inteligência Artificial, estatística, computação de alto desempenho e outros. Propõe, ainda que o objetivo principal seja extrair um conhecimento de alto nível a partir de dados de baixo nível.

O modelo CRISP-DM contempla seis fases para um projeto de mineração de dados, sendo assim determina-se um ciclo de vida compreendido para cada uma dessas fases: A primeira fase, conhecida como Entendimento dos dados, ou “fase de entendimento dos objetivos e dos requerimentos sob a perspectiva do negocio” segundo CHAMPAN [5] e uma fase crucial da mineração. A fase Entendimento dos dados caracteriza-se pelo exame acurado dos dados, procurando identificar a sua qualidade. Dados ausentes “missing data” – são comuns em base de dados não estruturados, configurando-se como um problema a ser considerado,

* 1. Abreviaturas y Siglas

Defina las abreviaturas y los acrónimos cuando se utilizan por primera vez en el texto, incluso después de que se han definido en el resumen. No use abreviaturas en el título o títulos de las secciones a menos que sea totalmente inevitable.

* 1. Unidades

En cuanto al uso de las unidades:

* Utilizar las unidades del Sistema Internacional (SI);
* Evite combinar las unidades del SI con otros sistemas (por ejemplo, la corriente en amperios y el campo magnético en oersteds) porque crea confusión;
* Utilice cero antes de la coma decimal: "0,25", no ",25."
  1. Ecuaciones

Las ecuaciones son una excepción a las especificaciones de este modelo. Debe utilizar la ecuación en fuentes Times New Roman y /o Symbol (no utilice ninguna otra fuente). Numere las ecuaciones consecutivamente con paréntesis y hacia la derecha. Ejemplo[1]:

 

Tenga en cuenta que las ecuaciones deben estar centradas. Asegúrese de que los símbolos en su ecuación se definen antes o después de la presentación de la ecuación. Una excelente guía para la escritura científica es

1. Usando la Plantilla

Duplicar el archivo con la plantilla usando el comando "Guardar como" ("Save As") del sistema, darle un nombre apropiado para su artículo y utilizar el nuevo archivo para la escritura y el formato de su artículo.

* 1. Autores y Afiliaciones

La plantilla fue diseñada para que las afiliaciones del autor no se repitan para los autores de la misma afiliación. Por favor, mantenga sus afiliaciones lo más sucintas posible (por ejemplo, no haga diferencia entre departamentos de la misma organización). Este modelo fue diseñado para dos afiliaciones. Afiliación individual modificar la sección correspondiente para una sola columna. Para tres o más afiliaciones también utilizar una sola columna.

* 1. Los títulos de las Secciones

Hay dos tipos de títulos de Secciones: Títulos de Componentes e Títulos de Texto.

Los Títulos de Componentes identifican los diferentes capítulos tales como "Agradecimientos" ("Agradecimientos") y Referencias ("Referencias") y deben tener el formato con el estilo "Título 5". Utilice “figure caption” para figuras, y "table head” para los títulos de las tablas. Normalmente, el más sencillo es copiar y pegar ("copy-paste") de un título o subtítulo y cambiar el contenido existente convenientemente (donde la numeración se realiza automáticamente por la plantilla).

Títulos de Texto organizan los temas de artículo de moda relacional y jerárquica. Los estilos “Heading 1”, “Heading 2”, “Heading 3”, e “Heading 4” deben ser utilizados.

* 1. Figuras y Tablas
     1. Posicionamiento de Figuras y Tablas: Lugar pies de figura por debajo de las figuras y los títulos de las tablas sobre las tablas. Inserte figuras y tablas después de que se citan en el texto. Use la abreviación "Fig. 1 " para se refieren a las figuras, incluso al principio de una frase.

1. Ejemplo de un pie de figura (figure caption)

Leyendas de Figuras: Use 8 puntos Times New Roman. Utilice palabras en lugar de símbolos o abreviaturas. Buscar figuras legibles incluso si no se imprimen en color (por ejemplo, el uso de diferentes texturas y tipos de línea en lugar de los colores).

1. Título Ejemplo de una Tabla

| Tabla | Tabla | | |
| --- | --- | --- | --- |
| Columna 1 | Columna 2 | Columna 3 |
| Línea 1 | Contenido |  |  |
| Línea 2 |  |  |  |



1. Ejemplo de un pie de figura (figure caption)
2. Conclusiones

El modelo IEEE y otros modelos disponibles, facilitan el trabajo de los resúmenes de congresos de edición y la producción de material científico. Utilizar correctamente al escribir su artículo y deseos para un gran trabajo y contribuir al éxito de nuestra conferencia.

Agradecimientos (Heading 5)

Introduzca agradecimiento a las organizaciones y personas que contribuyeron a la obra. Si la presentación de la conferencia es Double-Blind, introduzca sólo texto DFBR (Deleted for Blind Review) en señal de agradecimiento.

Referências Bibliográfica

Los números de las citaciones son automáticos e entre corchetes [1] [2] [3]. A menos que haya seis autores o más se incluyen todos los autores y no utilizan "et al.". Los artículos que no hayan sido publicados deben ser citados como "unpublished" [4]. Los artículos aceptados para publicación deben ser citados como "in press" [5]. Para los artículos publicados en un idioma que no sea inglés, por favor poner la cita en inglés primero seguido de la cita en el idioma original [6].

1. L. Castanheira, “Aplicação de técnicas de Mineração de Dados em Problemas de Classificação de Padrões”, 2008, pp553.
2. R. Wirth, “CRISP-DM 1.0 – Step-by-step data mining guide”, 2000, pp. 7– 10.
3. B. Possas, M. Carvalho, R. Rezende, and W. Meira jr., “Data mining: técnicas para exploração de dados”, Universidade Federal de Minas Gerais, 1998.
4. P. Fayyad, U. Piatetsky-Suapiro, and G. Smyth, “From data mining to knowledge discovery in databases”, Advances in Knowledge Discovery and Data Mining, 3rd ed, vol.17, 1996, pp.1– 36.
5. P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Thomas, C. Shearer, and R. Wirth.
6. K. Elissa, “Title of paper if known,” unpublished.
7. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
8. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740c741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
9. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.