

# Uma proposta de Solução de Mineração de Dados aplicada à Segurança Pública

Fernando José Braz<sup>1</sup>, Walter Silvestre Coan<sup>2</sup>, Andrei Rosseti<sup>2</sup>

<sup>1</sup>Instituto Federal Catarinense (IFC)  
Rodovia SC 303, Campus Experimental - Videira – SC – Brasil

<sup>2</sup>Departamento de Informática – Universidade da Região de Joinville - UNIVILLE  
Joinville – SC – Brasil.

fernando.braz@ifc-videira.edu.br, walter.s@univille.br, mdxico@gmail.com

**Abstract.** *The Public Security is a major task in the planning of any city. The data analysis of police occurrences is a mandatory step in policies to combat crime. This paper presents an information system based on the database and data mining technologies in order to allow access to resumes of georeferenced data related to the services rendered by a military police battalion. It also allows access to pre-processed data mining models on the database. Considering the location issuer's geographic query, both the data summaries as the related mining models are presented to user, revealing the knowledge related to the area space-time.*

**Resumo.** *A segurança pública tem importância fundamental no planejamento de qualquer cidade. A análise dos dados das ocorrências policiais é fundamental nas políticas de combate ao crime. Este artigo apresenta um sistema de informação desenvolvido sobre as tecnologias de bancos de dados e mineração de dados com o objetivo de permitir o acesso a resumos de dados georeferenciados relacionados com os atendimentos prestados por um batalhão da Polícia Militar. Também permite o acesso a modelos de mineração de dados previamente processados sobre a base de dados. Considerando a localização do emissor da consulta, resumos de dados e modelos de mineração são apresentados ao usuário, revelando o conhecimento relativo àquela área espaço-temporal.*

## 1. Introdução

O desenvolvimento das tecnologias de banco de dados, mineração de dados e georeferenciamento, tem resultado em novas soluções de sistemas de informação. A área da segurança pública pode ser especialmente beneficiada por este desenvolvimento. As centrais de atendimento de ocorrências policiais produzem grandes volumes de dados que representam o histórico daqueles atendimentos. A análise deste histórico permite revelar o conhecimento residente dentro daquele volume de dados, apresentando características específicas do comportamento da violência dentro de uma região espaço-temporal. O domínio deste conhecimento permite o desenvolvimento de políticas e ações de segurança pública mais efetivas e eficientes, empregando os recursos de acordo com a necessidade do comportamento revelado para aquele ambiente.

Os registros de atendimentos prestados pelas forças policiais geralmente envolvem os seguintes atributos: *data e hora da ocorrência, endereço e descrição da ocorrência*

(*código da ocorrência*). Entretanto, apenas o armazenamento destes registros não oferece muito poder de análise. É necessário que se efetuem processos de transformação de dados para que se possa revelar o conhecimento residente. Basicamente, o processo de transformação envolve a tradução dos endereços das ocorrências em coordenadas de latitude e longitude, com o objetivo de oferecer a possibilidade da emissão de consultas espaciais em toda a sua amplitude. A simples tradução do endereço em coordenadas permite que se emitam consultas espaciais, o que, sem dúvida alguma, aumenta a capacidade de análise dos dados. As consultas espaciais permitem, por exemplo, a recuperação dos resumos de dados (somatório, média etc) relacionados às ocorrências registradas dentro de um raio a partir de uma coordenada de latitude e longitude. Além disso, permite a emissão de consultas recuperando as ocorrências dentro dos limites de um polígono definido pelo usuário.

A maior contribuição do sistema apresentado neste artigo reside na possibilidade de recuperação das informações relacionadas com modelos de mineração de dados previamente processados. Estes modelos permitem identificar, dentro de uma determinada região espaço-temporal, a probabilidade de co-ocorrência de eventos. Tomando como base a localização do usuário emissor da consulta (uma viatura policial, por exemplo), e considerando uma determinada área ao redor daquela posição, o sistema permite recuperar e apresentar a probabilidade de uma determinada ocorrência (ex: assalto a mão armada) e, além disso, a probabilidade desta ocorrência estar associada a outras diversas (ex: assalto a mão armada → tentativa de homicídio, 85%). Este conhecimento é de fundamental importância para o prévio entendimento do contexto da região em que uma viatura policial está em deslocamento. O comportamento e recursos utilizados pela força policial serão dimensionados de acordo com o modelo que se apresenta naquele momento para aquela área.

O artigo segue com a Seção 2 comentando alguns trabalhos relacionados, a Seção 3 apresentando o sistema e, finalmente a Seção 4 trazendo as primeiras conclusões sobre a proposta em discussão.

## **2. Trabalhos Relacionados**

Em [Furtado 2002] os autores discutem a utilização das técnicas de mineração de dados na análise das ocorrências policiais. Naquele estudo aborda-se a possibilidade de estabelecimento de perfis de criminosos e/ou vítimas, fazendo uso da tarefa de agrupamento em mineração de dados. Em [Júnior and de Lima Dantas 2012] é apresentada a discussão da utilização das técnicas de mineração de dados para a identificação dos vínculos entre os agentes criminais, com o objetivo de identificar relacionamentos entre criminosos. Os autores do trabalho apresentado em [Alberto Filho 2008] argumentam que a mineração de dados pode trazer resultados mais expressivos no combate ao crime do que a apresentação de resumos de dados de ocorrências criminais. Mais do que isto, a mineração de dados pode revelar o conhecimento residente naqueles grandes volumes de dados, para tanto sugerem a utilização das técnicas de mineração de texto para revelar este conhecimento. No trabalho de [Braz et al. 2009] é possível visualizar os resultados da utilização da mineração de dados sobre uma base de dados de ocorrências policiais. Naquele estudo a proposta era identificar agrupamentos de ocorrências, considerando o tipo de ocorrência, perfil de vítima, data da ocorrência e local do evento. Apesar de envolver a localização do evento, ainda naquela proposta não acontecia o processamento

de modelos, e sua posterior visualização, fazendo uso do georeferenciamento. Na proposta apresentada em [SILVA and ROVER 2011] os autores fazem uma abordagem da mineração de dados aplicada à segurança pública. A utilização das técnicas de mineração de dados e texto tinha o objetivo de revelar a associação existente entre os diversos perfis de vítimas e ocorrências criminais. Nenhuma das propostas anteriores considerava a utilização da mineração de dados buscando a co-ocorrência de eventos considerando dados espaço-temporais na análise e apresentação dos resultados.

### 3. O Sistema Desenvolvido

O aplicativo foi desenvolvido sobre a plataforma do *Google Maps*, permitindo, desta forma, que todas as consultas relativas às ocorrências possam ser acessadas sobre a visualização do mapa de qualquer cidade. Esta característica facilita tanto a visualização das informações quanto a análise dos dados. As ferramentas que possibilitaram o desenvolvimento do projeto foram: *Microsoft SQL Server 2008 R2* (Banco de dados), *Microsoft Visual Studio 2010* (Ambiente WEB de Desenvolvimento Integrado), *X Code.app* (Ambiente Móvel de Desenvolvimento) e *iPhone Simulator.app* (Simulador do sistema operacional do iPhone).

#### 3.1. A Fonte de Dados

O volume de dados utilizado é o produzido por um batalhão da Polícia Militar de Santa Catarina, partir do registro das ocorrências atendidas através das chamadas do telefone 190. Basicamente, as informações residentes nestes volumes de dados apresentam a data, o logradouro e o tipo de chamada atendida. Estes dados, na sua origem, não estão mantidos sob o formato relacional. Este fato dificulta que o processo de análise aconteça. Portanto, a primeira etapa do projeto foi o tratamento dos dados, com o objetivo de prepará-los para utilização no formato pretendido. Cada tupla do volume de dados foi definida pelo seguinte esquema: *identificador da célula, identificador temporal, código da ocorrência*. O espaço foi dividido em um grid regular, composto por um conjunto de células que representam a menor granularidade espacial possível, definida previamente. Desta forma, dentro das fronteiras de uma célula, é possível que existam registradas diversas ocorrências.

#### 3.2. O Funcionamento do Sistema

O sistema permite que as consultas emitidas pelo usuário sejam definidas através do desenho de polígonos sobre o mapa da cidade, e também pela seleção de bairros e ruas da mesma cidade. Além disso, o intervalo temporal também é definido e utilizado como parâmetro da consulta. O mapa da região é obtido fazendo uso da *API* do *Google Maps*. Desta forma é possível fazer uso da ferramenta de zoom, possibilitando ao usuário definir áreas de pesquisa nos mais diversos níveis de *granularidade*, desde o envolvimento de uma cidade inteira até a uma residência específica. As Figuras 1 e 2 apresentam, respectivamente a visualização do ambiente da maior para a menor granularidade.

O atendimento das requisições emitidas pelo usuário envolve duas etapas: *Definição da Consulta* e *Busca e Registro dos Resultados*. Na etapa de ***Definição da Consulta*** acontece o recebimento e a interpretação das coordenadas marcadas em tela pelo usuário (polígono), delimitando as fronteiras da consulta (ver Figura 1). Através de

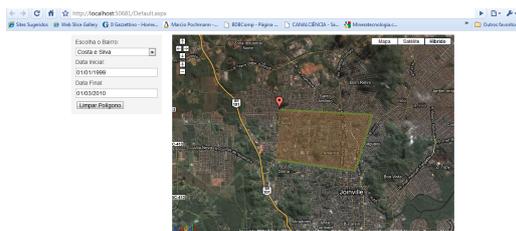


Figura 1. Sistema - Geral

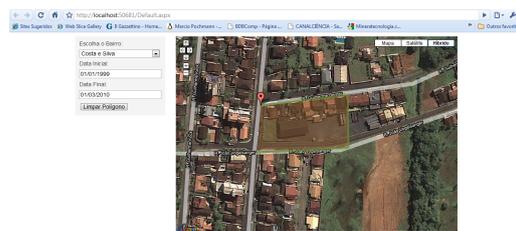


Figura 2. Sistema - Detalhe

módulo desenvolvido especificamente para esta tarefa, a definição do polígono é armazenado em atributo geográfico. A data da criação e o usuário que emitiu a consulta também farão parte do registro da consulta. Na etapa de **Busca e Registro dos Resultados**, uma vez que os detalhes (polígono, data e usuário) da consulta foram registrados no banco de dados, através da utilização de funções espaciais, são recuperadas e resumidas todas as ocorrências que fazem parte daquele conjunto de resultados. Os modelos de mineração que pertencem àquela determinada região espaço-temporal, também são recuperados e disponibilizados para apresentação ao usuário.

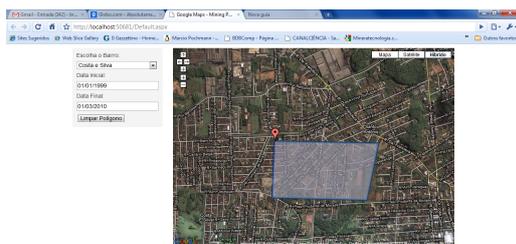


Figura 3. Delimitação da área espaço-temporal da consulta

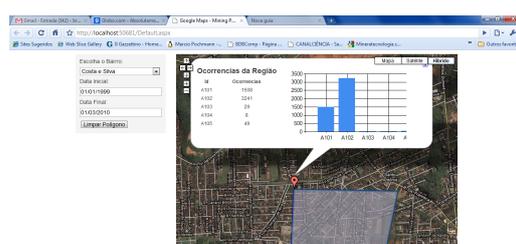


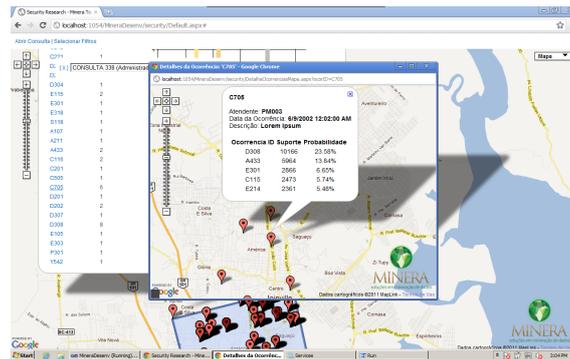
Figura 4. Apresentação dos resultados

As Figuras 3 e 4 apresentam, respectivamente, as telas do sistema mostrando a área espaço-temporal definida pelo usuário e o consequente resultado da consulta emitida à base de dados. Os resultados podem ser visualizados tanto em forma de tabela quanto no formato de gráfico. Para cada código de ocorrência é apresentado o resumo de chamados relacionados dentro da respectiva área espaço-temporal.

### 3.3. Módulo de Mineração de Dados

Os conjuntos de modelos de mineração de dados relacionados com a área espaço-temporal da consulta também são apresentados em tela. Cada ocorrência, cujo resumo de casos é apresentado em tela, oferece um link que dá acesso ao módulo de mineração de dados.

A Figura 5 apresenta a visualização do módulo de mineração de dados. No exemplo, a ocorrência *C705* é selecionada pelo usuário para a requisição do módulo de mineração de dados. Na execução do módulo, os limites da área espaço-temporal, bem como o código da ocorrência (no exemplo *C705*) são utilizadas como variáveis de entrada (input). A execução do módulo de mineração de dados acontece em duas etapas: (1) seleção dos modelos pertencentes à área espaço-temporal e (2) seleção dos modelos que possuam a ocorrência de entrada (*C705* no exemplo) como parâmetro. O sistema possui um conjunto de modelos de mineração de dados previamente processados e baseados



**Figura 5. Módulo de Mineração de Dados**

na tarefa de regra de associação proposta por [Agrawal and Srikant 1994]. A preparação dos dados (ver Seção 3.1) foi executada com a proposta de permitir que tarefas de regra de associação pudessem ser utilizadas no projeto. Os modelos estão relacionados com a área espaço-temporal. Por conta disto, na primeira etapa de execução do módulo de mineração, somente os modelos que estejam envolvidos pela área espaço-temporal serão recuperados. O resultado da primeira etapa servirá como base para a execução da fase em que acontece o filtro dos modelos que possuem a ocorrência selecionada pelo usuário. Todos os modelos do sistema proposto são do tipo regra de associação  $A \rightarrow B, X\%$  (a ocorrência do evento  $A$  está relacionada com a consequente ocorrência do evento  $B$  com um suporte de  $X\%$ ).

**Tabela 1. Associações**

Padrão	Suporte
$C705 \rightarrow D308$	23,58%
$C705 \rightarrow A433$	13,84%
$C705 \rightarrow E301$	6,65%
$C705 \rightarrow C115$	5,74%
$C705 \rightarrow E214$	5,48%

No exemplo, através da Figura 5 é possível identificar as possibilidades de relacionamentos com o antecedente  $C705$  apresentadas na Tabela 1. Esta informação pode ser de extrema importância na medida em que apresenta, em ordem de probabilidade decrescente, os relacionamentos possíveis de acontecerem naquela região espaço-temporal.

Outra possibilidade na utilização do sistema é a definição da região espacial fazendo uso das coordenadas de latitude/longitude do usuário emitente da consulta. Neste formato de utilização, a área de delimitação da consulta acontece através da formação da região no entorno da localização do usuário e o raio de cobertura definido pelo emitente da consulta. Neste caso não existe a necessidade de definição do polígono por parte do usuário, tudo acontece de forma automática levando em conta o valor do raio definido para a consulta.

## 4. Conclusões

A etapa de preparação dos dados trouxe uma complexidade extra ao desenvolvimento do sistema. Tendo em vista a utilização de consultas espaciais, houve a necessidade do desenvolvimento de um módulo de tradução dos endereços em coordenadas de latitude e longitude para cada ocorrência.

Outra contribuição deste sistema é a possibilidade de oferecer a visualização dos modelos de mineração de dados no formato *antecedente* → *consequente* fazendo uso da localização geográfica do emitente da consulta. Isto permite que uma viatura policial, por exemplo, durante a sua movimentação por diversos pontos da cidade, tenha conhecimento preciso das características e do contexto da região em que se movimenta. Este é um comportamento diverso da grande maioria dos sistemas que estão restritos a apresentação de resumos de dados relacionados às ocorrências. Vale lembrar que a tarefa de *regra de associação* foi utilizada neste protótipo. Entretanto, a discussão da utilização de outras técnicas para a descoberta dos relacionamentos pode, e deve, ser estudada na continuidade do projeto.

A primeira versão do sistema encontra-se concluída e instalada em dispositivo móvel do tipo *Tablet* equipado com *GPS (Global Position System)* e acesso à internet. Isto possibilita que as requisições do sistema emitidas pelo dispositivo considerem a localização geográfica do emitente da consulta, buscando os modelos de mineração de dados (bem como os resumos de dados) relacionados com a região espaço-temporal definida pelo usuário. Entretanto, a utilização efetiva do protótipo por parte do quadro da Polícia Militar ainda não aconteceu. Espera-se que esta utilização aconteça em breve.

## Referências

- Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *VLDB*, pages 487–499.
- Alberto Filho, L. (2008). Aplicação de técnicas de mineração de dados e textos no apoio à tomada de decisão em segurança pública. Master's thesis, Universidade Federal do Pará, Belem, Pará.
- Braz, L., Ferreira, R., Dermeval, D., Veras, D., Lima, M., and Tiengo, W. (2009). Aplicando mineração de dados para apoiar a tomada de decisão na segurança pública do estado de alagoas. In *Workshop de Computação Aplicada em Governo Eletrônico*, Bento Gonçalves, Rio Grande do Sul, Brazil. SBC Brazilian Computer Society.
- Furtado, V. (2002). *Tecnologia e gestão da informação na segurança pública*. Garamond.
- Júnior, C. and de Lima Dantas, G. (2012). A descoberta e a análise de vínculos na complexidade da investigação criminal moderna. Disponível em <<http://www.egov.ufsc.br/portal/sites/default/files/anexos/13124-13125-1-PB.pdf>>. Acesso em março de 2012.
- SILVA, E. R. G. and ROVER, A. J. (2011). O processo de descoberta do conhecimento como suporte à análise criminal: minerando dados da segurança pública de santa carolina. In *International Conference on Information Systems and Technology Management*, São Paulo, São Paulo, Brazil. Sociedade Brasileira de Pesquisa Operacional.