

Mineração de Métricas de Segurança Pública Baseadas em Dados Geoespaciais

Adelson Dias de Araújo Júnior

Departamento de Engenharia de Computação e Automação
Universidade Federal do Rio Grande do Norte
Natal, Brasil
adelsondias@live.com

Resumo—Muitos das recentes pesquisas em cidades inteligentes envolvem espaço e tempo como variáveis de análise em seus sistemas. Para prover uma exploração na base de dados de segurança pública da região metropolitana de Natal com o objetivo de descobrir conhecimento, este trabalho pretende demonstrar como pode ser feito essa mineração de dados em sistemas de gerenciamento de banco de dados com extensão espacial, propondo métricas estatísticas e apresentando correlações entre essas métricas mineradas.

Palavras-chave: *cidades inteligentes, bancos de dados geoespaciais, mineração de dados.*

I. INTRODUÇÃO

Para integrar técnicas de conectividade e mobilidade nas cidades, em uma época de grande expansão urbana, e com o objetivo de se tornar uma cidade mais humana, o conhecimento é, de fato, uma peça chave. Segundo [1], é sobre uma rede de conhecimento formado por pessoas e informações o que pode fazer uma cidade inteligente, com estruturas de tomada de decisão eficientes. Nessa perspectiva, construir essa interação entre pessoas e conhecimento é fundamental nessa jornada.

Ainda, de acordo com [2], políticas estão se moldando para adaptar soluções que incluam tecnologias de informação e comunicação (em inglês, ICTs) para desenvolver decisões inteligentes no aspecto das cidades. Para tanto, é necessário utilizar as tecnologias existentes para contextualiza-las nesse aspecto. Um exemplo disso são os sistemas de gerenciamento de banco de dados (SGBD) com extensões espaciais, os quais tem a capacidade de armazenar dados geográficos, de grande utilidade para gerenciamento das cidades.

Nesse contexto, esses dados geográficos armazenados nesses SGBD espaciais (como o PostGIS, MySQL Spatial, Oracle Spatial) abrem amplas aplicações na área de análise de dados. Uma dessas aplicações mais discutidas é a mineração de dados, que utiliza muito, inclusive, de informações espaciais, como podemos perceber em [3] e [4] ou até em outras aplicações de geoprocessamento como [5]. Assim, o processo para se descobrir conhecimento em bases de dados espaciais já é bem discutido na literatura, e deve ser abordado pelas aplicações de cidades inteligentes.

Sendo assim, este trabalho apresenta a descoberta de métricas estatísticas nas bases de dados de segurança pública da cidade do Natal e região metropolitana baseadas em informações georreferenciadas sobre ocorrências policiais (vistas como pontos), trajetórias de veículos policiais (vistas

como linhas) e áreas estratégicas de patrulhamento, ditas AISPs (vistas como polígonos), a fim de fazer um recorte sobre o desempenho de variáveis de segurança de uma cidade, como apresenta-se na Figura 1 a representação de uma das métricas a serem expostas, referente ao percentual de ruas patrulhadas.

Dentre tantos os usos dessas métricas, as quais são apresentadas na seção III, um deles pode ser analisar as correlações entre essas métricas apresentadas e verificar quais são dependentes entre si, de modo a verificar se existe uma relação em uma métrica ao agir em outra.

II. PRÉ-PROCESSAMENTO

Um primeiro pré-processamento consiste em adequar essas tuplas georreferenciadas para o modelo de geometrias com o qual o SGBD trabalha. No caso, foi utilizado o PostGIS [6] pela grande variedade de funções de análise e por não possuir licença proprietária. Transformar a latitude e longitude no tipo de geometria ‘POINT’, ‘LINESTRING’ e ‘POLYGON’ foi necessário para trabalhar com os cruzamentos espaciais disponíveis em funções no próprio SQL.

Outro pré-processamento feito foi a partição das trajetórias em sub-trajetórias. Utilizou-se do parâmetro temporal com referência à diferença entre as amostragens, por exemplo, se a próxima amostragem da viatura 13 demorar mais do que 10 minutos, isso pode indicar que podemos particionar em outra trajetória, já que podemos interpretar como uma nova rota que está sendo registrada. Ainda, foram adicionados nos dados das ocorrências, colunas referentes à qual AISP (“bairros” policiais) ela está inserida, bem como o bairro, com um script de *update* em SQL com um cruzamento entre o polígono delimitador e a instância das ocorrências.



Figure 1. Representação gráfica da análise de ruas patrulhadas. À esquerda, dados de um dia; à direita, dados de duas semanas.

III. MÉTRICAS

Através da busca por medidas de distância, densidade de similaridades, pode-se reparar que é possível extrair das bases de dados informações que tragam uma semântica àquele contexto. No caso da segurança pública, com as informações disponíveis foi se foi possível extrair estatísticas em temos de:

- A. Densidade de ocorrências por unidade de área (OPA) em metros/km²;
- B. Percentual de ocorrência sobre o total (PO);
- C. Percentual de ruas patrulhadas (PRP);
- D. Distância média entre ocorrências (DMO), em metros;
- E. Dissimilaridade entre trajetórias (DT);
- F. Tempo médio de estadia em AISP (TMEA), em minutos.

Todas essas métricas estão estruturadas em um esquema de instância (ocorrências, trajetórias) e delimitação (AISP ou bairro). Pode-se organizar essas métricas em torno de uma AISP específica, e a evolução dessas métricas em tempos específicos ou por um ranking por AISP, visualizando quais AISPs estão mais críticas para uma determinada métrica, como é o caso da Tabela I, ordenada pelo valor de ODA em um dia de análise, considerando somente as 5 primeiras.

TABLE I. MÉTRICAS REFERENTES A UMA ÚNICA AISP

AISP	Métricas					
	ODA	PO	PRP	DMO	DT	TMEA
4	4.93	2.86	14.08	805.1	0.0129	34.551
2	3.02	5.22	30.37	914.7	0.0137	41.995
8	2.48	7.73	12.62	1143.4	0.0223	25.127
3	2.25	4.04	23.27	1050.3	0.0109	36.765
1	2.13	3.15	25.48	1776.2	0.0123	50.200

Por fim, podemos ter uma ideia sobre a correlação entre essas métricas, através da Tabela II, em que foi utilizado o coeficiente de correlação de Pearson, e o nível de significância do teste foi de 5%. Nela podemos perceber que algumas das métricas estão bastante correlacionadas, positivamente ou negativamente. Negativamente relacionadas estão DMO-OPA, significando que a distância média entre as ocorrências cresce inversamente com a densidade de ocorrências por área, DT-OPA, significando que quanto mais as trajetórias das viaturas se espalham, menor é a densidade de ocorrências por área, TMEA-DMO, significando que quanto mais tempo as viaturas passam nas regiões menos as ocorrências se

espalham, e TMET-DT, significando que quanto mais tempo passam as viaturas nas regiões, mais repetitivas são as trajetórias.

TABLE II. ANÁLISE DE CORRELAÇÃO DE PEARSON

	A	B	C	D	E	F
A	1	-0.015 (0.953)	0.665 (0.001)	-0.597 (0.011)	-0.732 (0.001)	0.363 (0.151)
B		1	-0.195 (0.565)	0.110 (0.674)	0.146 (0.573)	-0.039 (0.879)
C			1	-0.252 (0.329)	-0.428 (0.086)	-0.076 (0.770)
D				1	0.813 (0.001)	-0.496 (0.042)
E					1	-0.518 (0.032)
F						1

Continuando na análise de correlação, tem-se que positivamente relacionadas estão as métricas DMO-DT, isto é, quanto mais as trajetórias se diferem, maior é a distância entre as ocorrências, e por fim, uma conclusão bastante intrigante é a correlação positiva entre OPA-PRP, que quer dizer que quanto maior a densidade de ocorrências na região, maior é o percentual de ruas patrulhadas. Entretanto, a razão lógica para essa correlação deve ser o fato de que áreas perigosas demandam maior densidade de policiamento.

Sendo assim, a avaliação de métricas baseadas em dados geográficos pode constituir-se de rica informação para tomada de decisão em patrulhamento e distribuição de recursos na segurança pública de uma cidade inteligente.

REFERENCES

- [1] Negre, E., Rosenthal-Sabroux, C., Gascó, M. (2015) "A Knowledge-Based Conceptual Vision of the Smart City". 48th Hawaii International Conference on System Sciences.
- [2] Caragliu, A., Del Bo, C., Nijkamp, P. (2009), "Smart Cities in Europe" In: 3rd Central European Conference in Regional Science – CERS.
- [3] Bogorny, V., Palma, A. T., Engel, P. M., Alvares, L. O. (2006) "Weka-GDPM: Integrating Classical Data Mining Toolkit to Geographic Information Systems". In: SBBD Workshop on Data Mining Algorithms and Applications(WAAMD'06), Florianopolis, Brazil, October 16-20, pp.9-16.
- [4] Braz, F. J. (2012) "Uma proposta de Solução de Mineração de Dados aplicada à Segurança Pública". In: Simpósio Brasileiro de Sistemas de Informação, São Paulo. SBSI 2012 - Trilhas Técnicas (Technical Tracks).
- [5] Bordin, M., Silva, J. P., Manske, K. V., Costa, A., Caneparo, S. C. (2013) "O uso do geoprocessamento na segurança pública do Estado do Paraná". Anais XVI Simpósio Brasileiro de Sensoriamento Remoto – SBSR, INPE.
- [6] PostGIS, Spatial and Geographic objects for PostgreSQL. <http://postgis.net>