

Intuição Geoespacial

Por **MundoGEO** | 0100, 15 de Dezembro de 2002

O "Spatial Data Mining" pode nos ajudar a descobrir padrões e tendências que estão além da nossa percepção e intuição?

Tome uma corda esticada, unindo um ponto em São Paulo a um outro ponto no Rio de Janeiro. Suponha que a distância entre esses pontos seja de exatamente 400 quilômetros. Tome uma outra corda, com exatamente **um metro a mais** que a anterior, ou seja, 400.001 metros, e fixe também suas extremidades nos mesmos pontos anteriores. Ela ficará bamba. Levante esta corda pelo seu ponto médio, formando um triângulo, conforme a figura abaixo. Pergunta: a altura deste triângulo formado será maior ou menor que um metro? Pense um pouco antes de prosseguir. O que nossa percepção sugere?

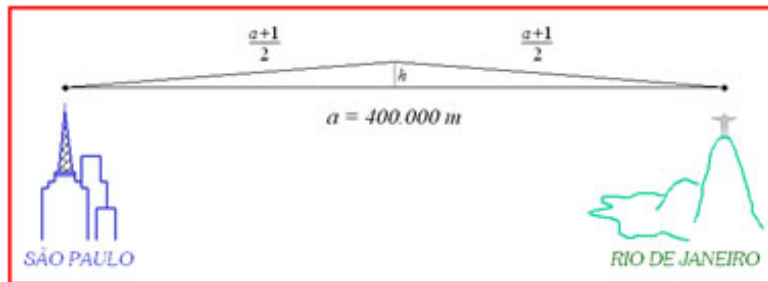


Figura 1: O problema da corda

Por mais absurdo que possa parecer, caberia dentro desse triângulo um prédio, de forma retangular, com cerca de 126 andares de altura e 50 quarteirões de comprimento – o triângulo tem mais de 447 metros de altura! Para se ter uma idéia, as torres gêmeas do World Trade Center tinham, cada uma, 417 metros e 110 andares. (... Pausa para reflexão ...) Perplexos com o resultado, ficamos a imaginar: por que falha a nossa intuição? (veja a resolução matemática no final desse artigo)

É por essas e outras que nem sempre devemos contar com a nossa boa e velha intuição. Nossa percepção espacial prega peças. E é aí que entra o **Spatial Data Mining**.

É público e notório o benefício que a visão geoespacial trouxe para a solução de problemas. Um mapa transmite-nos instantaneamente resultados que quase sempre demandariam horas de trabalho em planilhas e números para serem atingidos. A percepção de concentração, magnitude, segmentação e variabilidade que um mapa nos dá é insubstituível.

Em contrapartida, cada vez mais enormes quantidades de dados têm sido armazenadas em GIS e Data Warehouses por aí. Grandes companhias manipulam da ordem de centenas de terabytes de dados, desde a exploração de recursos naturais a análises mercadológicas. Estima-se que o Sistema de Observação da Terra da NASA irá transmitir cerca de 50 gigabytes de imagem digital por hora! Esse volume de informação torna muitas vezes impraticável a intervenção humana no processo de detecção de relações entre os dados de caracterização espacial.

Mas existem técnicas, que têm sido assunto de recentes pesquisas, que permitem que essa "percepção" seja sugerida pela máquina. Esses trabalhos são úteis para a detecção de potenciais de crédito, de risco, de fraude, de inadimplência, etc. – baseados em características espaciais e não espaciais. Sistemas como esse poderiam, por exemplo, mostrar aos usuários que determinados tipos de solo são propensos a sofrer erosão quando detectados em regiões próximas a aglomerados urbanos, ou, para os marqueteiros de plantão, que toda padaria bem-sucedida em bairros de classe média tem quase sempre uma farmácia/drogaria e um açougue nas vizinhanças.

O *Spatial Data Mining* (ou Mineração de Dados Espaciais) trata da extração de conhecimento implícito, relacionamento espacial e outros padrões não explícitos armazenados em bases de dados espaciais. Geralmente, seus métodos se baseiam em classificação, generalização, proximidade, topologia e concentração. Diferentes técnicas são utilizadas para dados em formatoraster e vetorial, que podem ter a componente espacial dominante ou não no processo de mineração.

A figura 2 apresenta uma arquitetura que modela o Spatial Data Mining. A **Base de Conhecimento** armazena conhecimentos "já estabelecidos", inerentes aos modelamentos espaciais, como hierarquias, topologias e metadados. Os dados são trazidos das bases de armazenamento usando uma **Interface de Bancos de Dados**, que permite a otimização das consultas. Estruturas de indexação espacial podem ser usadas para tornar o processamento eficiente. O componente de Foco decide que partes dos dados serão úteis para o reconhecimento de padrões. É aí que está o núcleo conceitual do processo. Por exemplo, ele pode decidir que somente alguns atributos são relevantes para as tarefas de descoberta de conhecimento, ou pode extrair somente alguns objetos que pareçam mais apropriados para uma determinada estratégia. Regras e padrões são descobertos no módulo de **Extração de Padrões**. Esse componente pode utilizar estatística, geoestatística, *machine learning*, redes neurais, algoritmos genéticos e outras técnicas de Data Mining, em conjunto com algoritmos de geometria computacional e manipulação espacial para tornar performática a tarefa de encontrar regras e relações. A relevância e significância dos padrões encontrados são processadas pelo módulo de **Avaliação**, para eliminar relações óbvias e redundantes. Os componentes podem interagir entre si através do **Controlador**, que também provê *feedbacks* para refinamento das *queries*. As descobertas são, enfim, passadas ao usuário para verificação. O usuário pode, inclusive, controlar cada etapa do processo. Os padrões relevantes podem ser armazenados na própria Base de Conhecimento, configurando o aprendizado.

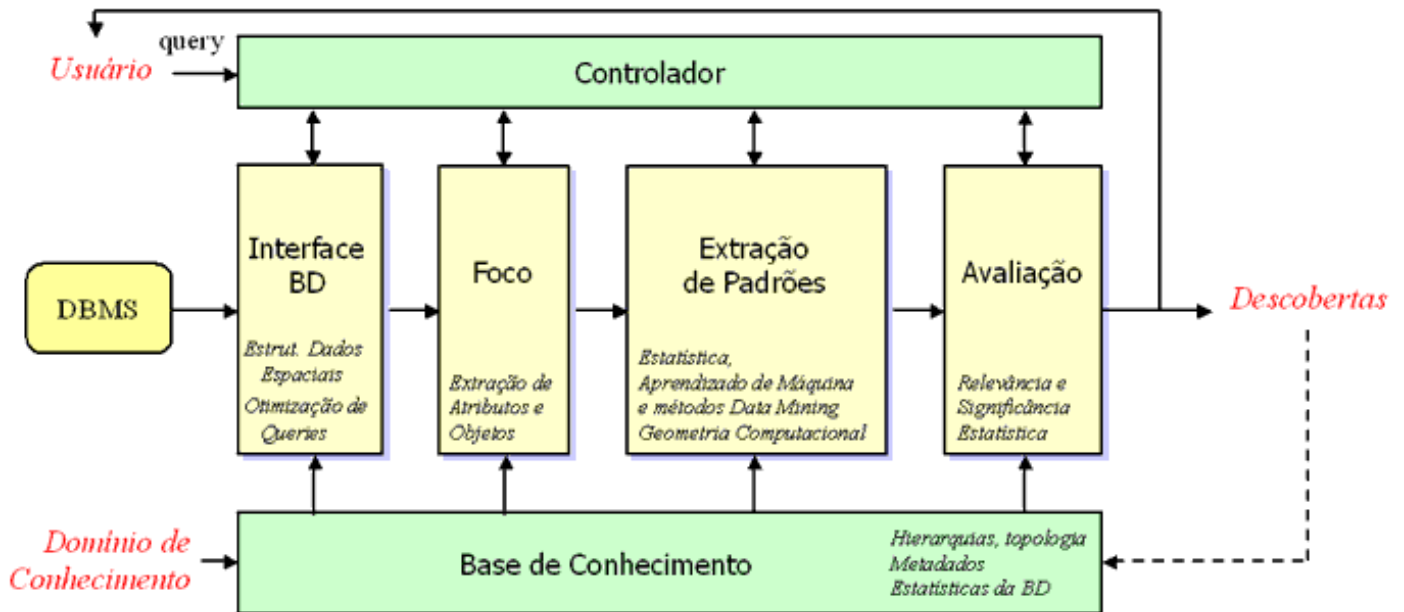


Figura 2: O processo de um Spatial Data Mining

As ferramentas GIS cada vez mais procuram incorporar essas técnicas como módulos e extensões de sua funcionalidade básica, e, da mesma forma, os grandes pacotes de Data Mining tradicional têm procurado expandir-se para entender e utilizar as técnicas e métodos espaciais. Mas isso ainda é incipiente. O uso dessas técnicas em Sensoriamento Remoto, Imageamento Médico, Meteorologia, Geomarketing, e outras muitas áreas, deve ser maximizado. Mas vale ressaltar que tais sistemas são falíveis, às vezes respondem com descobertas sem importância, e dependem exclusivamente da boa avaliação do usuário e da qualidade e estrutura das informações fornecidas. E é aí que a boa e velha intervenção humana não deve falhar.

Em tempo. Eis a solução do problema da corda:

Sejam a o comprimento da corda menor e h a altura do triângulo. Pelo Teorema de Pitágoras temos:

$$h^2 = \left[\frac{a+1}{2} \right]^2 - \left[\frac{a}{2} \right]^2 = \frac{1}{4}(2a+1)$$

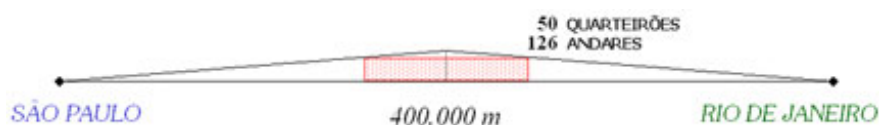
Logo,

$$h = \frac{1}{2} \sqrt{2a+1}.$$

Sendo

$a = 400.000\text{m}$, temos:

$$h = \frac{1}{2} \sqrt{800.001} \approx 447 \text{ m}.$$



Referências Bibliográficas:

Duarte Jr., G. G. De São Paulo ao Rio de Janeiro com uma Corda "Ideal", Revista do Professor de

Matemática, edição 22, 1992.

Koperski, K., Han, J., Jiawei & Adhikary, J. Mining Knowledge in Geographical Data, In Communications of ACM, Simon Frasey University, Burnaby, BC, Canadá, 1998.

Eduardo de Rezende Francisco, Bacharel em Ciência da Computação pelo IME-USP, atua em GIS e Análise de Mercado na AES Eletropaulo e é sócio-fundador da GITA Brasil. erfrancisco@hotmail.com e eduardo.francisco@eletropaulo.com.br