

Uma ferramenta para distribuição e mapeamento de dados Paleogeográficos

Joaquim Assunção, Maria Pivel, Paulo Fernandes, Duncan Ruiz

¹Pontifícia Universidade Católica do Rio Grande do Sul
Faculdade de Informática - FACIN
Porto Alegre, Brazil (+55)51-3320-3558

{joaquim.assuncao,maria.pivel,paulo.fernandes,duncan}@pucrs.br

Resumo. *As bacias sedimentares marginais despertam grande interesse, tanto de cientistas da terra, como da indústria petrolífera. Pesquisas nestas áreas geram grandes volumes de dados provenientes de técnicas de geofísica e perfurações. Esses dados se encontram, muitas vezes, esparsos e armazenados de formas gráficas e esquemáticas, o que dificulta a aplicação de técnicas computacionais para análise e descoberta de conhecimento. Além disso, não há informação suficiente para toda a área de interesse. Pensando nisso, criou-se uma ferramenta para coleta, distribuição e mapeamento destes dados. Com isso, criou-se novas oportunidades, pois com o auxílio de algoritmos de mapeamento, novos dados foram estimados de modo a preencher uma área da qual anteriormente não havia informação geológica. Os algoritmos desenvolvidos compõem a ferramenta que serve para extração, transformação e carga de dados (ETL). Estes dados compõem um banco de dados paleogeográficos de grande volume. Esse banco visa agregar informações existentes com informações estimadas para se obter novas informações via processos de descoberta de conhecimento em banco de dados (Knowledge Discovery in Database, KDD).*

1. Introdução

Ao longo de décadas de pesquisa, cientistas das geociências acumularam quantidades substanciais de dados geológicos. Esses grandes volumes de dados têm potencial para possuírem informações ocultas, úteis para as geociências. Porém, quanto maior a quantidade e a diversidade destes dados, maior a dificuldade de se extrair informação dos mesmos [Miller e Han 2001] [Fayyad et al. 1996].

Dados geológicos que representam longos períodos de tempo, usualmente são gerados e acumulados de maneiras distintas, tanto na representação como na forma e local de armazenamento. De fato, muitos são os meios utilizados para armazenamento destas informações. Dentre eles destacam-se os gráficos gerados por levantamentos sísmicos, onde grande parte das informações obtidas, são unidas e sintetizadas em cartas estratigráficas, que possuem foco na ocorrência de litologias em um local e tempo [Milani et al. 2007].

O trabalho descrito neste artigo tem como principal objetivo obter o máximo proveito dos dados presentes nas cartas estratigráficas, unindo-os com outros dados paleogeográficos. Para isso, foram feitos estudos sobre os geodados em questão, além de estudos

de técnicas para adaptação e mapeamento de dados paleogeográficos. Para atingir este objetivo, este trabalho foi dividido em três etapas: (1) coleta e adaptação dos dados, (2) transformação e estimativa de novos dados, (3) carga e mapeamento dos dados.

O produto final, resultante das três fases, constitui um conjunto de dados no espaço e no tempo. Para isso, na primeira fase, os dados são extraídos das cartas e das demais fontes de interesse, e ambos são adaptados para terem formatos compatíveis entre si. Na segunda fase, os dados são transformados para um padrão numérico, de forma a facilitar sua manipulação e servir como entrada para um algoritmo de estimativa e mapeamento, que na terceira fase realiza o mapeamento dos dados e os insere no banco de dados.

A abordagem descrita acima foi aplicada a dados das bacias sedimentares marginais brasileiras. A seção 2 constitui uma breve descrição do conhecimento necessário para o entendimento da solução. A seção 3 relata a solução criada e os resultados obtidos. Na seção 4 são feitas considerações sobre o que foi criado, a contribuição e os trabalhos futuros.

2. Background e dados alvo

Cartas estratigráficas são ferramentas úteis para o estudo de bacias sedimentares. Elas representam graficamente as mudanças geológicas que ocorreram em uma bacia em função do tempo. Além disso, carregam grandes quantidades de informações relativas as litologias presentes naquela bacia [Milani et al. 2007]. A figura 1 mostra um exemplo de carta estratigráfica; neste caso, a carta da Bacia de Santos.

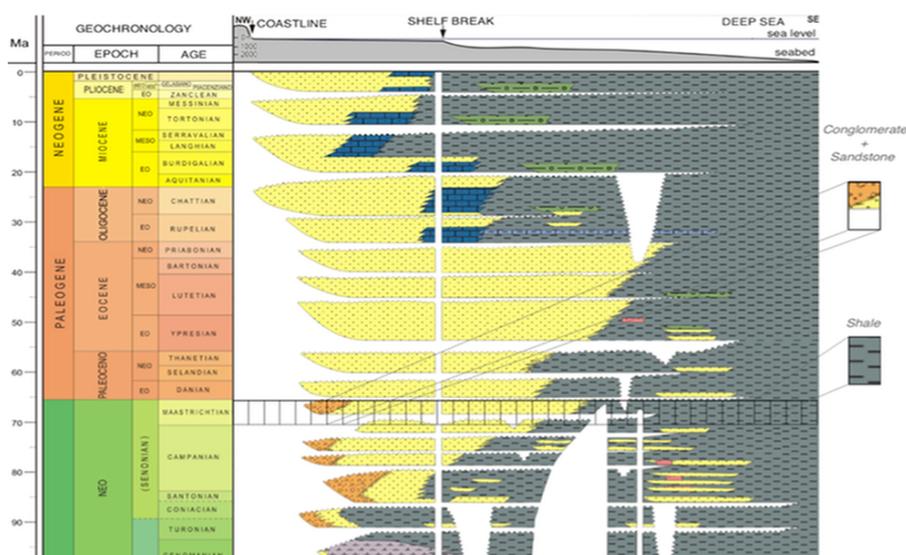


Figura 1. Simplificação da Carta estratigráfica da Bacia de Santos. À esquerda, a barra colorida representa a escala de tempo geológico, do Cretáceo até o Recentes (amarelo). À direita, diferentes padrões e cores são usados para representar as litologias que são distribuídas de acordo com a distância da costa (eixo horizontal) e idade geológica (eixo vertical). Espaços em branco representam ausência de depósitos de litologias para uma determinada distância da costa, em uma determinada idade geológica. O grid mostrado sobre a idade Maastrichtiana ilustra uma das áreas de coleta de dados.

Todos os dados extraídos das cartas estratigráficas são divididos por idades geológicas. Cada carta representa um limite entre a linha da costa brasileira até a isóbata

de 3000m. Isto significa que a largura da Bacia é variável. Todavia, o limite marítimo brasileiro é restrito a 200 milhas náuticas; como grande parte das bacias está contida neste limite, cada Idade da carta foi dividida em 37 partes, onde cada parte representa aproximadamente 10km.

Com o objetivo de caracterizar a configuração presente da costa brasileira, além dos dados estratigráficos, a solução desenvolvida também agrega ao banco, dados de batimetria (*i.e.* profundidade do mar) [Smith e Sandwell 1997] e anomalias gravimétricas (*i.e.* desvios do valor teórico da aceleração da gravidade que auxiliam na compreensão da estrutura interna das bacias) [Sandwell e Smith 2009].

3. Solução

Para cada tipo de litologia foi atribuído uma representação por potência de dois. Deste modo, várias litologias podem ser representadas com um único valor. Esta abordagem também é prática para a decomposição de valores e, por consequência, obtenção das litologias presentes no local.

Para estimar o valor de um ponto x no espaço, foi criado um algoritmo que se baseia na distância entre o ponto em questão e os pontos mais próximos de cada limite da bacia em questão. Deste modo o algoritmo cria uma matriz de valores que são mapeados em pontos no espaço. Para isso, foi utilizada a fórmula de Haversine, juntamente com uma variação da mesma para obtenção da curvatura entre dois pontos.

$$d = R.c \quad (1)$$

Onde :

$$R = \text{Raio da Terra} = 6.371\text{km}$$

$$c = 2.\text{atan2}(\sqrt{a}, \sqrt{(1-a)})$$

$$a = \sin^2(\Delta\text{lat}/2) + \cos(\text{lat1}).\cos(\text{lat2}).\sin^2(\Delta\text{lon}/2)$$

$$\Delta\text{lat} = \text{lat2} - \text{lat1}$$

$$\Delta\text{lon} = \text{lon2} - \text{lon1}$$

$$\Theta = \text{atan2}(\sin(\Delta\text{lon}).\cos(\text{lat2}), \cos(\text{lat1}).\sin(\text{lat2}) - \sin(\text{lat1}).\cos(\text{lat2}).\cos(\Delta\text{lon})) \quad (2)$$

Após a criação do algoritmo, foi desenvolvida uma ferramenta para executá-lo. Mais do que isso, a ferramenta realiza alterações de formatos, importa planilhas Excel, realiza cálculos e aplica os algoritmos criados para estimativas e carrega os dados para o banco. À esquerda na figura 2, é mostrada a interface da ferramenta, juntamente com um breve resumo das suas funções. À direita é mostrado o resultado do mapeamento de dados.

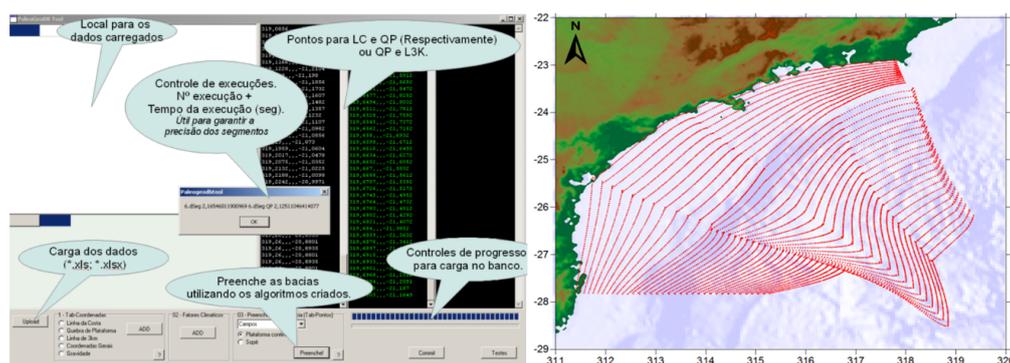


Figura 2. Ferramenta de ETL, juntamente com o resultado do mapeamento para a bacia de Santos. Cada ponto representa a localização de diferentes dados inseridos no banco e mapeados ao longo da margem continental.

4. Conclusão

Neste artigo foi relatada a criação de uma ferramenta que incorpora novos algoritmos para distribuição e mapeamento de dados paleogeográficos. Com esta ferramenta é possível atribuir valores para diferentes proporções de depósitos de sedimentos para toda a área de interesse da costa brasileira.

A ferramenta também possibilita a criação de diversos dados que são parte de um grande banco de dados paleogeográficos. Além disso, a ferramenta também pode ser adaptada para mapear outros dados semelhantes em qualquer superfície esférica.

Nossos testes quanto à posição geográfica se mostraram satisfatórios (ver: parte à direita da figura 2), além disso, não foi identificado nenhum conjunto de técnicas na literatura com propósito semelhante.

As informações que os dados gerados representam estão sendo comparadas com dados de levantamentos sísmicos. Para isso, são usadas técnicas de mineração de dados, que não apenas servem para verificação dos dados mapeados, mas principalmente como parte do processo de KDD. Assim, conclui-se que os algoritmos criados, juntamente com a ferramenta, geram e possibilitam a utilização de novos dados e técnicas computacionais para descobrir informações e gerar conhecimento.

Referências

- Fayyad, U., Piatetsky-Shapiro, G., e Smyth, P. (1996). Knowledge discovery and data mining: Towards a unifying framework. pages 82–88.
- Milani, E., Rangel, D., Bueno, G., Stica, J., Winter, W., Caixeta, J. e Neto, O. (2007). *Boletim de Geociências da Petrobras 3ª Ed.*, volume 2. Sól Gráfica.
- Miller, H. e Han, J. (2001). *Geographic Data Mining and Knowledge Discovery*. Research Monographs in Geographic Information Systems Series. Taylor.
- Sandwell, D. e Smith, W. (2009). Global marine gravity from retracked Geosat and ERS-1 altimetry: Ridge Segmentation versus spreading rate. *Journal of Geophysical Research*, 114:18.
- Smith, W. e Sandwell, D. (1997). Global seafloor topography from satellite altimetry and ship depth soundings. *Science Magazine*, 277.