

Utilização da biblioteca TerraLib para algoritmos de agrupamento em Sistemas de Informações Geográficas

Maurício P. Guidini¹ e Carlos H. C. Ribeiro¹

¹Instituto Tecnológico de Aeronáutica – Comando-Geral de Tecnologia Espacial
Programa de Pós-Graduação em Aplicações Operacionais
Praça Marechal Eduardo Gomes n. 50 – Jardim das Acácias
São José dos Campos – SP – CEP: 12228-900
Caixa Postal 15.064 – 91.501-970 – São Paulo – SP – Brasil
guidini@ita.br, carlos@comp.ita.br

***Abstract.** The integration of Data Mining algorithms to extract knowledge from Geographic Information Systems that hold large amounts of data and directly affects economics and social sectors, is becoming indispensable. In this way, TerraLib, an important library for geographic applications development in geoprocessing world scenarios, appears as a strong candidate to incorporate these features. This work shows an experiment on the integration of a clustering algorithm with TerraLib to generate a classification on unidentified air traffic, aiming at applying this solution to an analysis of these traffics.*

***Resumo.** A integração de algoritmos de Mineração de Dados para extrair informações de sistemas de informação geográficos, que contêm grandes volumes de dados e que afetam diretamente diversos setores econômicos e sociais, está se tornando imprescindível. Desta maneira, TerraLib, uma biblioteca para desenvolvimento de aplicativos geográficos de importância no cenário do geoprocessamento mundial, surge como forte candidata a incorporar estes recursos. Este trabalho mostra um experimento de integração de um algoritmo de agrupamento na TerraLib, para gerar uma classificação de dados de tráfegos aéreos desconhecidos, visando à aplicação da solução na análise destes tráfegos.*

1. Introdução

Nos dias atuais, devido à facilidade crescente de aquisição de imagens e coleta de dados para aplicações, o volume de dados armazenados cresce muito além da capacidade humana de processá-los para extrair informações relevantes e conhecimentos úteis à tomada de decisão.

Em aplicações geográficas, além dos dados alfanuméricos a serem processados, surge a necessidade de lidar com geometrias e suas relações, determinando áreas de contato, de sobreposição, etc., e com os aspectos temporais, que podem ser contínuos ou discretos. Estas dificuldades levam, naturalmente, à aplicação de técnicas de Mineração de Dados como solução para extrair conhecimento destes sistemas.

Considerando a oportunidade de utilização de uma ferramenta nacional para desenvolvimento de Sistemas de Informação Geográfica (SIG), optou-se pela biblioteca

TerraLib [Câmara 2000], desenvolvida no Instituto Nacional de Pesquisas Espaciais (INPE), em conjunto com o aplicativo TerraView, também desenvolvido no INPE, que permite a criação e controle de bases de dados geográficos.

Uma ferramenta que possibilite a aplicação de algoritmos de Mineração de Dados em um SIG é de grande utilidade para o problema de análise de tráfegos aéreos desconhecidos. Esse problema relaciona-se ao uso ilegal do espaço aéreo brasileiro, e considera aeronaves que sobrevoam o território sem plano de vôo ou contato com os órgãos de controle. Assim, para que o registro dessas ocorrências gere conhecimento e permita operações de planejamento no policiamento do espaço aéreo, é interessante a aplicação de algoritmos de Mineração de Dados nas bases de dados, a fim de levantar informações úteis à tomada de decisão.

Este trabalho visa apresentar um experimento realizado com a aplicação do algoritmo de agrupamento (clusterização) *Kmeans* em uma base de dados de tráfegos aéreos, para evidenciar o traçado das rotas preferenciais dos tráfegos desconhecidos, demonstrando como a biblioteca TerraLib pode ser operada sobre uma base de dados geográficos arbitrária, combinada a um algoritmo de Mineração de Dados.

2. Mineração de Dados

O processo de descoberta de conhecimento em bases de dados, também conhecido como *Knowledge Discovery in Databases* (KDD), é uma área relativamente nova, surgida da necessidade de extrair informações de bases de dados, ao invés de retirar apenas coleções de dados organizados. Segundo Fayyad [Fayyad 1996], a Mineração de Dados é “um processo não trivial de identificação de padrões válidos, novos, úteis e implicitamente presentes em grandes volumes de dados”.

A Mineração de Dados aplica conhecimento multidisciplinar, envolvendo as áreas de gerenciamento de bancos de dados, inteligência artificial e análise estatística, entre outras. Essa junção de áreas aumenta a complexidade dos algoritmos, exigindo capacidade computacional e critério na busca de padrões. Curotto [Curotto 2003] alerta para o fato de que vários autores utilizam o termo Mineração de Dados em um sentido amplo, significando todo o processo de KDD, como pode ser visto na Figura 1.

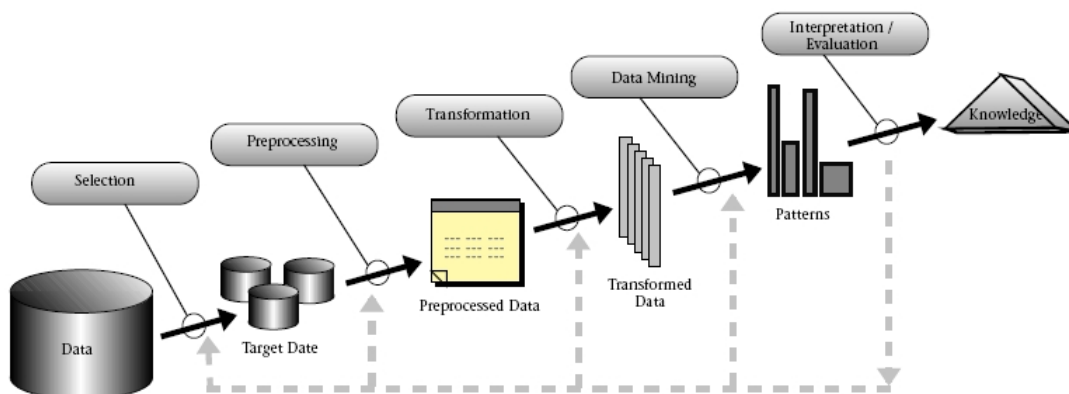


Figura 1. Etapas do processo de KDD [Fayyad 1996].

A aplicação de algoritmos de Mineração de Dados na análise de tráfegos aéreos desconhecidos envolve componentes espaciais (área de ocorrência, rotas), os relacionamentos entre essas entidades, bem como as relações de distância, direção e tempo [Neves 2001]. Assim, os fatores complicadores são maiores, pois não se tratam apenas de dados alfanuméricos, mas também de elementos geométricos e suas relações. Chelghoum e Zeitouni [Chelghoum 2004] propuseram uma solução para resolver este problema através da conjunção dos dados alfanuméricos e das relações espaciais em tabelas auxiliares, para permitir, a partir destas, a extração do conhecimento procurado.

Sistemas de Mineração de Dados que implementem uma boa parte dos algoritmos desenvolvidos não são facilmente encontrados com código-fonte aberto e licença livre, porque fazem parte de aplicações comerciais. Como exemplos de sistemas livres encontramos o WEKA (*Waikato Environment for Knowledge Analysis*), desenvolvido na Universidade Waikato de Hamilton, Nova Zelândia, implementado em linguagem JAVA [WEKA] e o MLC++ (*A Machine Learning Library in C++*) [MLC].

3. Algoritmos de Agrupamento

Para o ser humano é natural realizar a tarefa de agrupar objetos por suas similaridades, como, por exemplo, aves, peixes, plantas, etc. É dessa maneira que aprende a distinguir o que há a sua volta. De forma similar, o processo artificial de agrupamento visa separar os dados em grupos que contenham atributos similares, facilitando a compreensão das informações que guardam.

Um bom algoritmo de agrupamento caracteriza-se pela produção de classes de alta qualidade, nas quais a similaridade intra-classe é alta e a inter-classes é baixa. A qualidade do resultado depende do método utilizado para medir esta similaridade, normalmente baseado em análise de distância entre pares de objetos, e da habilidade em descobrir algum ou todos os padrões escondidos [Aurélio 1999].

Os algoritmos de agrupamento podem ser organizados dentro de cinco categorias principais [Han 2001]:

- *Métodos de particionamento*: dados n objetos, o método constrói k partições, onde cada partição representa um grupo, e $k \leq n$. O conjunto dos grupos satisfaz os seguintes requisitos: (1) cada grupo tem ao menos um objeto e (2) cada objeto pertence somente a um grupo. Este tipo de algoritmo utiliza normalmente uma dessas duas técnicas: *Kmeans*, onde o grupo é representado por uma medida média de todos seus objetos, e a *Kmedoids*, onde o grupo é representado por um objeto próximo ao centro deste.
- *Métodos Hierárquicos*: este método decompõe os dados do conjunto em uma hierarquia, que pode ser formada de forma aglomerativa ou divisiva. Na aglomerativa, cada objeto inicia com sendo um grupo, e a cada iteração é ligado a um grupo similar. Na divisiva, todos começam em um grande grupo, que vai sendo repartido em cada iteração. Os principais tratamentos para melhorar a qualidade dos métodos hierárquicos são: (1) realizar uma análise cuidadosa das conexões dos objetos em cada partição hierárquica, como é utilizado no CURE (*Clustering Using REpresentatives*) [Guha] e no Chameleon [Karapys 2003], ou (2) utilizar um algoritmo de hierarquização aglomerativo e aplicar em seguida

uma relocação iterativa para refinar os resultados, como é utilizado no BIRCH (*Balanced Iterative Reducing and Clustering using Hierarchies*) [Zhang 1997].

- *Métodos baseados em densidade*: este método baseia-se na idéia de densidade, ou seja, continuar aumentando um dado grupo até que a densidade (número de objetos) na vizinhança exceda algum limite estabelecido. Assim, para cada objeto em um grupo, a vizinhança deste, dentro de um raio dado, deve conter ao menos um número mínimo de outros objetos. Os algoritmos típicos que utilizam este método são o DBSCAN (*Density-Based Clustering Method Based on Connected Regions with Sufficiently High Density*) [Ester 1999] e o OPTICS (*Ordering Points To Identify the Clustering Structure*) [Ankerst 1999].
- *Métodos baseados em grade*: estes métodos quantificam os objetos em um número finito de células, que formam uma estrutura de grade. As operações de agrupamento são executadas nesta estrutura de grade. A principal vantagem é a velocidade de processamento, que depende apenas do número de células da grade. Um algoritmo típico deste método é o STING (*Statistical Information Grid*) [Wang 1997].
- *Métodos baseados em modelo*: neste método, um modelo hipotético é criado para cada grupo, e encontra-se o melhor ajuste do objeto ao modelo dado. Podem localizar os grupo através de uma função densidade que reflete a distribuição espacial dos objetos.

Os algoritmos de agrupamento nem sempre são facilmente classificados dentro de apenas uma das categorias citadas, pois utilizam vários métodos na sua construção. A aplicação das técnicas e métodos disponíveis depende do problema a ser resolvido e da quantidade de dados a ser trabalhada.

4. Tráfegos Aéreos Desconhecidos

O Espaço Aéreo Brasileiro está dividido em 4 Regiões de Defesa Aeroespacial (RDA), que formam zonas de responsabilidade atribuídas a Centros de Operações Militares (COPM). Os COPM constituem os Órgãos de Controle de Operações Aéreas Militares (OCOAM), responsáveis pela vigilância do espaço aéreo e por prestar serviços de informação de vôo e alerta a aeronaves engajadas em operações de defesa aérea [Brasil 1997].

As informações sobre tráfegos aéreos que chegam aos OCOAM provêm de diversos sítios-radar, distribuídos por todo o território nacional. Quando um tráfego ingressa na área de cobertura de um radar, um ponto é detectado na tela do controlador, que possui determinado período de tempo para identificá-lo e classificá-lo. Se o alvo é considerado desconhecido, duas linhas de ação podem ser tomadas [Bernabeu 2004]:

1. O tráfego pode ser interceptado, se constituir ameaça para algum ponto sensível do território nacional; ou
2. Ser acompanhado pelo controlador de vôo do OCOAM, gerando um Relatório de Tráfego Aéreo Desconhecido (RELTAD), que é encaminhado à seção de inteligência do Comando de Defesa Aeroespacial Brasileiro (COMDABRA), onde é armazenado e posteriormente analisado.

A análise dos dados retirados desses relatórios, através de técnicas de Mineração de Dados, pode levar à identificação do padrão comportamental dos tráfegos aéreos desconhecidos, relacionando suas principais rotas e áreas de atuação com características de desempenho das aeronaves detectadas, e com elementos que situem no tempo a maior frequência de ocorrência. Deste modo, será possível um melhor planejamento de operações de policiamento, com maior propriedade na escolha do período de execução e na quantidade de meios que devam ser empregados.

Como a localização destas rotas envolve componentes espaço-temporais, a integração de algoritmos de Mineração de Dados e Sistemas de Informação Geográfica proporcionará a ferramenta necessária para realizar a análise desses padrões comportamentais.

5. Integração de Algoritmos de Mineração de Dados com SIG

Dentro do escopo de integração de um algoritmo de Mineração de Dados com um SIG, foram considerados o algoritmo de agrupamento *Kmeans*, a biblioteca TerraLib [TERRALIB] e a ferramenta gráfica TerraView [TERRAVIEW].

Foi confeccionado um banco fictício de movimentos aéreos, baseado em dados reais fornecidos pelo COMDABRA, e que reflete uma amostragem da utilização ilegal do espaço aéreo brasileiro.

Os campos considerados para cada registro foram: o ponto de contato inicial, plotado através de suas coordenadas geográficas (longitude e latitude decimal), o ponto de perda de contato, também em coordenadas geográficas, um código de identificação e o mês da ocorrência. A base de dados completa contém aproximadamente 8.000 registros de tráfegos.

Na análise do comportamento dos tráfegos, o conhecimento do domínio é de grande importância, sendo representado pela atuação do especialista na redução do espaço de busca, na fase de pré-processamento, visando estudar um conjunto com características relevantes, como, por exemplo, aeronaves voando baixo e em velocidade reduzida. Esta pré-seleção é realizada por meio de uma consulta SQL normal. Nesse trabalho foram considerados dois conjuntos distintos: o primeiro, um grupo de controle contendo 36 instâncias de tráfegos, previamente selecionadas e que compõem 3 grupos; e o segundo contendo 321 instâncias, composta por tráfegos voando baixo, na área de vigilância de dois radares específicos.

A implementação utiliza como base o aplicativo Terraview, com um menu integrado que permite o acesso às funcionalidades de importação de dados dos tráfegos e aplicação do algoritmo de agrupamento. A utilização específica da TerraLib, dentro das funções criadas, ocorre na leitura da base de dados fonte (não obrigatoriamente no padrão TerraLib), que contém as informações dos tráfegos, a importação desses para um banco de dados geográfico já existente, com a criação de um plano (*layer*) específico, a criação da tabela com informações de linhas (rotas dos tráfegos), utilizando a classe *TeLines*, e registro das demais informações sobre cada tráfego em uma tabela de atributos.

Como os tráfegos são representados por segmentos de reta, serão referenciados como retas, compostas pelos atributos: latitude e longitude do ponto inicial, distância

(comprimento da reta) e proa (rumo geral do deslocamento). Uma instância de tráfego comporta sua reta representativa e demais atributos alfa-numéricos. Durante a carga de dados, a partir das coordenadas inicial e final das rotas é calculada a distância percorrida, em milhas náuticas, e o rumo geral da rota, em graus, através de trigonometria esférica [Miguens 1993] [Moreira], cujas funções de cálculo foram incluídas em uma classe para controle de rotas. Esses dados são armazenados em uma tabela de atributos.

Após a importação dos dados, pode-se acionar o algoritmo de agrupamento que carrega os dados das instâncias em tabelas na memória, utilizando um portal da classe TeDatabasePortal, e as processa utilizando classes próprias, implementadas para controle dos grupos gerados (classe Cluster) e o andamento do algoritmo (classe Clusterer) que, após a execução, retorna o número de grupos e as retas que representam os centros desses. O resultado gera um novo plano, cuja tabela de geometrias registra os centros dos grupos, e os demais valores, como nome do grupo e quantidade de rotas que o compõe, são armazenados em uma tabela de atributos. O algoritmo segue o pseudo-código apresentado na Figura 2.

```

Pré-condição: valores de desvio máximo para coordenadas, distância e proa ajustados
Início: K=0
Enquanto não atingir critério de término (desvio nos grupos)
    Aumenta K
    Eleja K centros aleatórios (dentro do rol de rotas)
    Enquanto houver trocas (kmeans)
        Aloque as rotas nos grupos através do critério peso
        Recalcule o valor dos centros
    Fim trocas
    Verifica critério de término
Salva Resultados
Fim

```

Figura 2. Pseudo-código do algoritmo de agrupamento.

O algoritmo inicia com K igual a 1, e aplica o $Kmeans$ com este valor. K instâncias são escolhidas aleatoriamente, e os valores das retas dessas instâncias são inicialmente ajustados para serem rotas médias dos grupos. Para cada nova instância, calcula-se a distância entre sua reta e a reta média dos grupos existentes, e aloca-se a instância no grupo cuja distância Euclidiana seja a menor. A fórmula 5.1 é utilizada para este cálculo.

$$\sqrt{\textit{peso} * |\textit{lat}_i - \textit{lat}_m|^2 + \textit{peso} * |\textit{long}_i - \textit{long}_m|^2 + |\textit{dist}_i - \textit{dist}_m|^2 + |\textit{pr}_i - \textit{pr}_m|^2} \quad (5.1)$$

Nessa fórmula, o fator lat representa a latitude do ponto inicial da reta, $long$ representa a longitude do ponto inicial da reta, $dist$ representa a distância da reta e pr representa a proa da reta, em relação ao norte. Os fatores de índice i referem-se aos valores da reta da instância, e os valores de índice m aos da reta média dos grupos. O fator $peso$ representa uma constante de multiplicação. Os grupos formados não precisam obrigatoriamente ser densos, mas desvios excessivos nos valores das coordenadas não

podem ser tolerados, considerando que apenas 1° de variação representa em média 60 milhas náuticas de diferença. Assim, foram estabelecidos os parâmetros para o desvio máximo em graus para as coordenadas e proa, e milhas para distância. Os parâmetros para distância e proa podem ser mais amplos que o das coordenadas, sem acarretar diferenças significativas.

O critério de parada do algoritmo baseia-se nesses parâmetros, servindo como ponto de corte para os desvios médios dos grupos. Quando todos os grupos satisfazem esses limites, o algoritmo é encerrado.

Após realizar testes com o grupo de controle, verificou-se a necessidade de incluir um fator de peso para a latitude e longitude, a fim de compensar a amplitude desses valores quando comparados aos de distância e proa. Outra anomalia surgiu ao processar o grupo de controle, onde dois valores de coordenadas distantes do centro do grupo se compensaram, terminando com rotas fora dos limites desejáveis, provocando perda de qualidade dos grupos. Para evitar este problema, verifica-se:

$$|lat_i - lat_m| > param_{coord} \quad (5.2)$$

e

$$|long_i - long_m| > param_{coord} \quad (5.3)$$

O fator $param_{coord}$ refere-se ao parâmetro de desvio máximo para as coordenadas. Em ambos os casos, se o fato ocorre o resultado do cálculo é multiplicado por um fator fixo. Esse sobrepeso mostrou-se eficiente na prevenção da criação de grupos com baixa qualidade.

Após a alocação das instâncias, as retas médias que representam os centros dos grupos são recalculadas, através da média aritmética dos valores das coordenadas, das distâncias e das proas. Se não houver mudanças nos valores médios, o algoritmo *Kmeans* para este valor de K é encerrado, e verifica-se se os parâmetros de desvio para os grupos foram satisfeitos. Caso positivo, o resultado é salvo e termina-se o processo. Senão, K é aumentado de 1 unidade e reexecuta-se o algoritmo *Kmeans*.

As Figura 3 e 4 apresentam, respectivamente, o total de rotas importadas para o banco de dados georeferenciado, e as rotas preferenciais resultantes da aplicação do algoritmo.

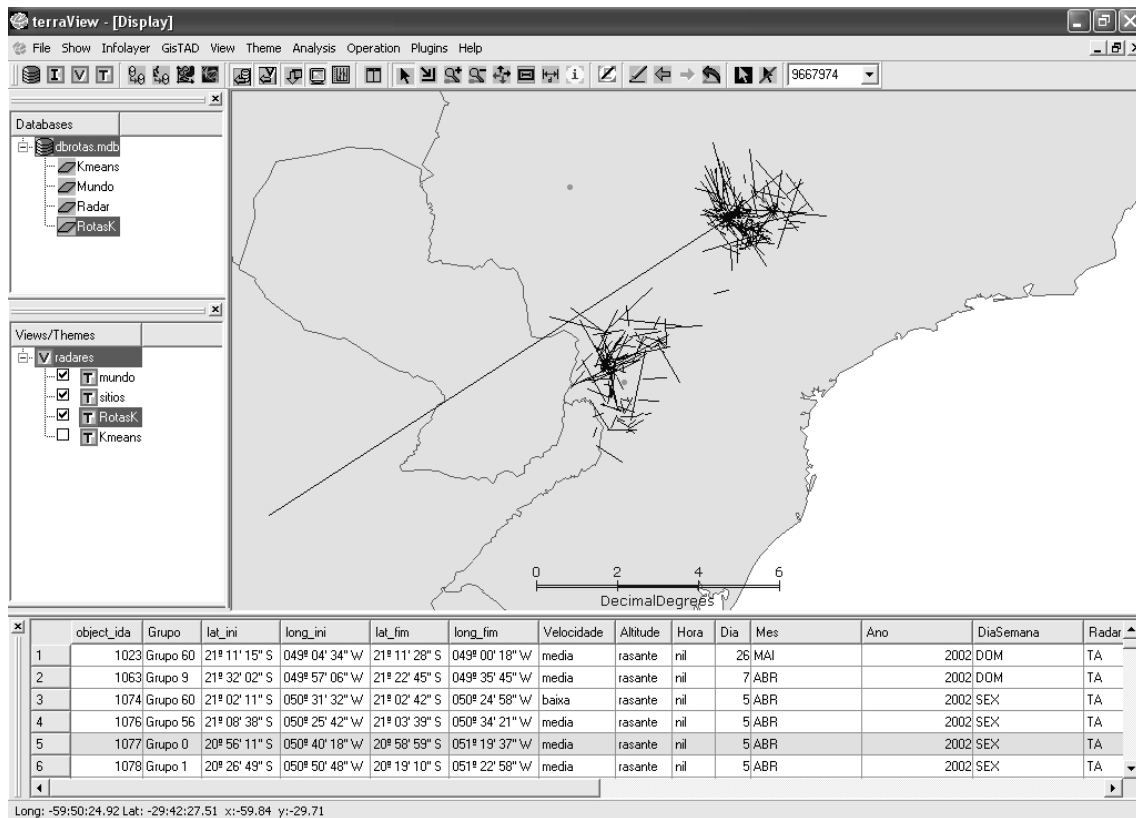


Figura 3. Rotas importadas.

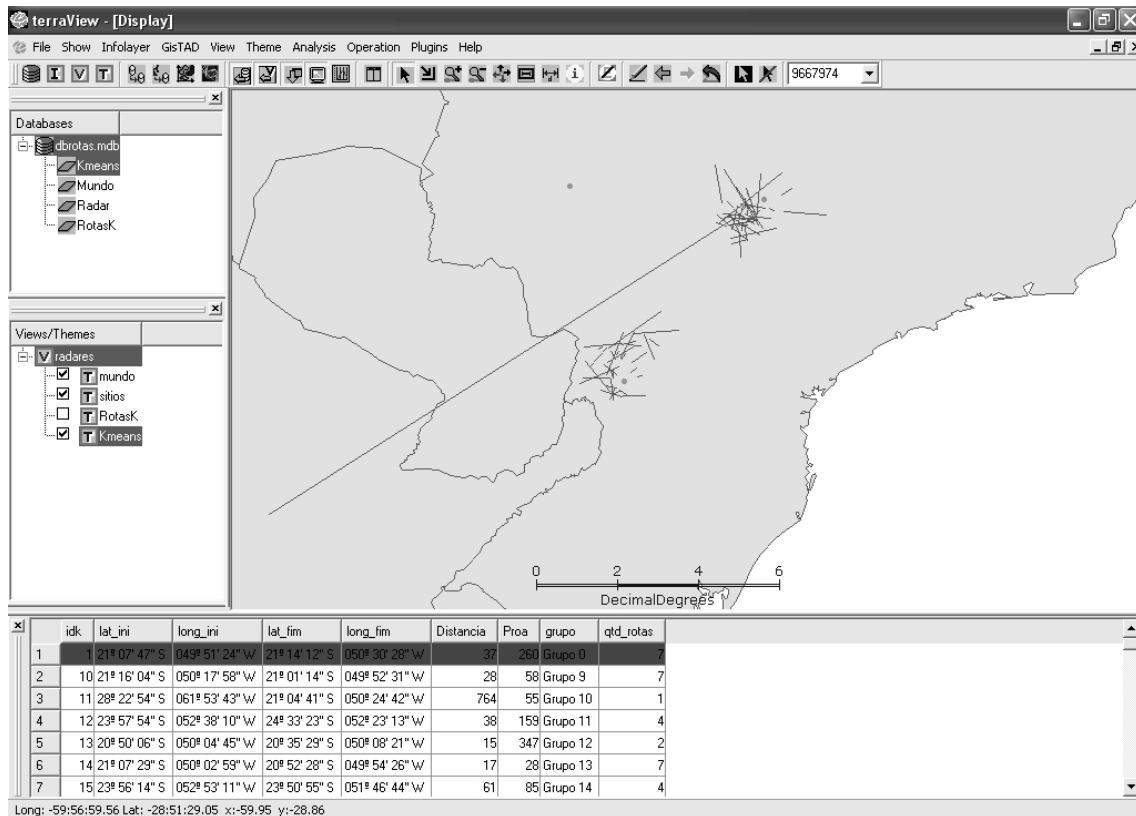


Figura 4. Rotas preferenciais, resultado do algoritmo.

As 319 instâncias do conjunto de teste foram repartidas em 73 grupos. A Figura 5 apresenta o detalhe de um grupo (Grupo 0), com as rotas que o compõe em tom mais claro e a rota preferencial (média) em tom mais escuro.

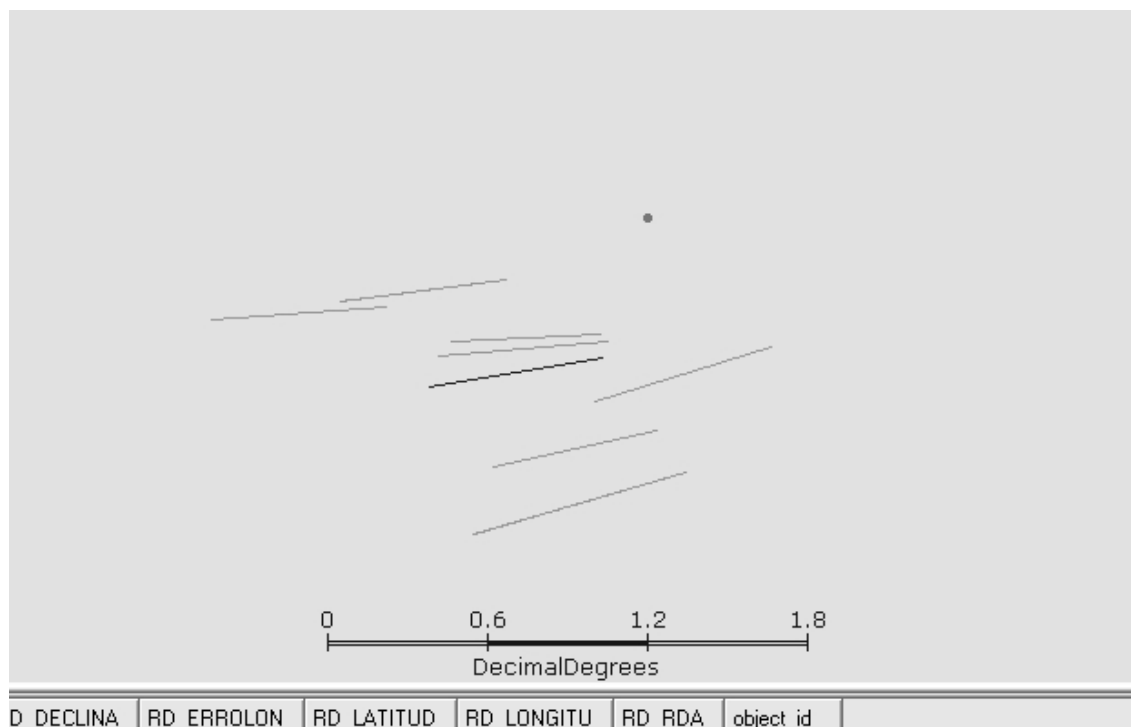


Figura 5. Detalhe de um grupo.

A partir desse resultado fica evidente a possibilidade de integração de algoritmos de Mineração de Dados à biblioteca Terralib, podendo este conjunto ser utilizado para extrair informações da base de dados de tráfegos desconhecidos, levantando rotas preferenciais, padrões de comportamento dos tráfegos, etc., e permitindo a visualização destas informações de maneira gráfica georeferenciada, com plotagem sobre mapas aeronáuticos ou fotos satélite.

6. Conclusão

A necessidade de trabalhar com grandes bases de dados é uma realidade nos dias atuais. A capacidade de extrair dados relevantes, mesmo com objetivos muito específicos, fica cada vez mais complexa. Torna-se necessário, então, ferramentas de Mineração de Dados para lidar com tais bases.

Considerando a integração de algoritmos de Mineração de Dados com ferramentas de apoio à aplicações geográficas, a TerraLib mostra-se uma ótima opção, por ser *software* de código aberto e de distribuição livre, além de ser nacional. Acredita-se que este trabalho comprova a possibilidade e a aplicabilidade desta fusão de tecnologias, gerando aplicações úteis à tomada de decisão baseada em SIG.

Como trabalho futuro, deverão ser incorporados à TerraLib os componentes necessários a outros algoritmos de agrupamento para Mineração de Dados, para utilização em uma aplicação de classificação de tráfegos aéreos desconhecidos.

Referências

- [Ankerst 1999] Ankerst, M. et al. OPTICS: ordering points to identify the clustering structure. International Conference on Management of Data. Pennsylvania, Philadelphia, 1999.
- [Aurélio 1999] Aurélio, M.; Velasco, M.; Lopes, C.H. Descoberta de Conhecimento e Mineração de Dados. 1999. Apostila – Laboratório de Inteligência Computacional Aplicada (ICA) – Departamento de Engenharia Elétrica PUC, Rio de Janeiro.
- [Bernabeu 2004] Bernabeu, F.G. Mineração de dados aplicada ao estudo do perfil de tráfegos aéreos desconhecidos. 2004. Dissertação (Mestrado) - Instituto Tecnológico de Aeronáutica, São Paulo.
- [Brasil 1997] Comando da Aeronáutica. ORG 1/1 Sistema de defesa aeroespacial brasileiro. Brasília: COMDABRA, 30 maio 1997.
- [Câmara 2000] CÂMARA, G. et al. Terralib: Technology in Support of GIS Innovation. Instituto Nacional de Pesquisas Espaciais, 2000. Disponível em <http://www.dpi.inpe.br/terralib/docs/papers/TerraLib_Paper_GeoInfo2000.pdf>. Acesso em 28 out. 2005.
- [Curotto 2003] Curotto, C.L. Integração de recursos de Data Mining com gerenciadores de bancos de dados relacionais. 2003. Tese (Doutorado) – Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- [Chelghoum 2004] Chelghoum, N.; Zeitouni, K. Spatial Data Mining implementation - Alternatives and performances. University of Versailles, 2004.
- [Ester 1999] Ester, M. et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. University of Munich, 1999.
- [Fayyad 1996] Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. From data mining to knowledge discovery: an overview. Cambridge: MIT Press, 1996.
- [Guha 1998] Guha, S.; Rastogi, R.; Shim, K. CURE: an efficient clustering algorithm for large databases. International Conference on Management of Data. Seattle, Washington, 1998.
- [Han 2001] Han, J.; Kamber, M. Data Mining – Concepts and Techniques. San Francisco, CA: Morgan Kaufmann Publishers, 2001.
- [Karapys 2003] Karapys, G.; Han, E.; Kumar, V. Chameleon: A Hierarchical Clustering Algorithm Using Dynamic Modeling. University of Minnesota, 2003.
- [Miguens 1993] Miguens, A. P. Navegação: A ciência e a arte - Volume II. Diretoria de Hidrografia e Navegação - Marinha do Brasil, 1993. Disponível em <http://www.mar.mil.br/dhn/bhm/publicacao/catalogo/htm/publicacoes_download.htm>. Acesso em 09 ago. 2006.
- [Moreira] Moreira, J.L.K. Qual é a distância entre dois pontos na superfície da Terra? Página da Internet. Disponível em: <<http://obsn3.on.br/~jlk/geopath/>>. Acesso em 09 ago. 2006.

- [Neves 2001] Neves, M.C.; Freitas, C.C.; Câmara, G. Mineração de Dados em Grandes Bancos de Dados Geográficos - Relatório Técnico. 2001 – Instituto Nacional de Pesquisas Espaciais, São Paulo.
- [Rezende 2003] Rezende, S. O. et al. Sistemas Inteligentes. Fundamentos e Aplicações. São Paulo: Manole, 2003.
- [Shannon 1975] SHANNON, C.; WEAVER, W. A. Teoria matemática da comunicação. São Paulo: Difel, 1975.
- [Silva 2005] Silva, M. P. S., et al.: Mining Patterns of Change in Remote Sensing Image Databases. 2005 – Universidade Estadual do Rio Grande do Norte, Rio Grande do Norte. Instituto Nacional de Pesquisas Espaciais, São Paulo.
- [Witten 2000] Witten, I.H.; Frank, E. Data Mining - Practical Machine Learning Tools and Techniques with Java. San Francisco, CA: Morgan Kaufmann Publishes, 2000.
- [TERRALIB] INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. Divisão de Processamento de Imagens. TerraLib. Disponível em: <<http://www.terralib.org/>>. Acessado em: 10 ago. 2005.
- [TERRAVIEW] INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. Divisão de Processamento de Imagens. TerraView. Disponível em: <<http://www.dpi.inpe.br/terraview/index.php>>. Acessado em: 10 ago. 2005.
- [Wang 1997] Wang, W.; Yang, J.; Muntz, R. STING: A Statistical Information Grid Approach to Spatial Data Mining. International Conference on Very Large Databases. Los Angeles, California, 1997.
- [WEKA] UNIVERSIDADE DE WAIKATO. WEKA 3 – Data Mining Software in Java. Disponível em <<http://www.cs.waikato.ac.nz/~ml/weka/index.html>>. Acessado em: 20 ago. 2005.
- [Zhang 1997] Zhang, T; Ramakrishnan, R.; Livny, M. BIRCH: A New Data Clustering Algorithm and Its Applications. Netherlands: Kluwer Academic Publishers, 1997.