ADMINISTRAÇÃO E NEGÓCIOS DE SERGIPE – FANESE
CURSO DE PÓS-GRADUAÇÃO ADMINISTRAÇÃO DE BANCO DE DADOS

MAXWELL AZEVEDO VIANA MORAES

GEOPROCESSAMENTO:

Uma abordagem sobre mineração de dados para descoberta e análise de informações em dados geográficos

Aracaju – SE
2016.1

MAXWELL AZEVEDO VIANA MORAES

GEOPROCESSAMENTO:

Uma abordagem sobre mineração de dados para descoberta e análise de informações em dados geográficos

Trabalho de Conclusão de Curso apresentado ao Núcleo de Pós-Graduação e Extensão – NPGE, da Faculdade de Administração de Negócios de Sergipe – FANESE, como requisito para a obtenção do título de Especialista Administração de Banco de Dados.

Aracaju - SE

2016/01

GEOPROCESSAMENTO:

Uma abordagem sobre mineração de dados para descoberta e análise de informações em dados geográficos

Maxwell Azevedo Viana Moraes[1]

RESUMO

O presente artigo qualitativo e bibliográfico, objetiva apresentar uma visão geral sobre a aplicação de técnicas de Sistemas Inteligentes a Sistemas de Informação Geográfica (SIG) e seus Bancos de Dados Geográficos. Para encontrar diversos tipos de conhecimentos, pode-se utilizar do processo de Descoberta de Conhecimento em Bancos de Dados em uma abordagem alternativa sobre dados e relacionamentos espaciais aplicando o método clássico de data mining por meio de algumas técnicas existentes. Em particular, este trabalho traz uma explanação sobre sistemas baseados em regras de Mineração de dados em Sistemas de Informação Geográfica e idealiza utilizar dessas possíveis alternativas de aplicação no contexto de tomada de decisões e explora algumas abordagens inteligentes no âmbito de sistemas georreferenciados.

Palavras chaves:

Mineração de Dados, Sistema de Informação Geográfica, Banco de Dados Geográficos, Sistemas Inteligentes, KDD**.**

1 Aluno Graduado em Sistemas para Internet - FANESE, Pós-Graduando em Administração de Banco de Dados – FANESE.

SUMÁRIO

RESUMO 4

1 INTRODUÇÃO 3

2 DESENVOLVIMENTO 4

2.1 Fundamentação Teórica 4

2.2 Contextualização do Tema 9

2.3. Técnicas de Mineração de dados 10

2.4. Exemplo de aplicação prática 13

3 CONSIDERAÇÕES FINAIS 13

ABSTRACT 15

REFERÊNCIAS BIBLIOGRÁFICAS 16

LISTA DE FIGURAS

Figura 1. Etapas do processo KDD.....9
Figura 2. Evolução do Valor Estratégico da base de dados.....10

1 INTRODUÇÃO

O sucesso das organizações depende basicamente das decisões tomadas por seus gestores, antes mesmo de apresentar ao mercado seus produtos ou serviços. Tais decisões têm se tornado necessárias em prazos cada vez mais curtos, exigindo dos gestores responsáveis uma atenção redobrada aos ambientes interno e externo da organização. Muitas vezes, más decisões são definidas, não pela inexistência do conhecimento para se escolher melhor, e sim porque o conhecimento não estava disponível para ser utilizado no tempo e lugares certos.

A quantidade informações disponíveis vêm crescendo de forma exacerbada nos últimos anos e vários fatores contribuíram para esse incrível aumento diz Cardoso et. al. (2008). Como consequência, bancos de dados passam a conter verdadeiros tesouros de informação e, devido ao seu volume, ultrapassam a habilidade técnica e a capacidade humana na sua captação e interpretação.

Dessa forma para que o conhecimento seja extraído de forma eficiente, é realizado um processo conhecido como Busca de Conhecimento em Banco de Dados Knowledge Discovery in Database – KDD.

> O processo é definido como um processo não trivial de descoberta de padrões válidos, novos, úteis e acessíveis. A principal vantagem do processo de descoberta é que não são necessárias hipóteses, sendo que o conhecimento é extraído dos dados sem conhecimento prévio. (FAYYAD et al.,1996)

Data mining, ou mineração de dados, é uma técnica que faz parte de uma das etapas da KDD a qual se utiliza de uma metodologia própria de preparação e exploração de dados e interpretação de seus dados, interpretação de seus resultados e assimilação dos conhecimentos minerados. Escreve Neto, et al. (2010) que ela é capaz de revelar, automaticamente, o conhecimento que está implícito em grandes quantidades de informações armazenadas nos bancos de dados de uma organização.

A análise inteligente e automática de dados para descobrir padrões ou regularidades em grandes conjuntos de dados, através de técnicas que

envolvam métodos matemáticos, algoritmos baseados em conceitos biológicos, processos linguísticos e heurísticos, os quais fazem parte do processo KDD responsável pela busca de conhecimentos em banco de dados é proveniente da mineração de dados.

> Padrões como classes, associações, regras e tendência possuem expressões espaciais bem definidas, tendo, assim, aplicações em técnicas para mineração de dados em bancos geográficos. (MILLER, 2001).

Worboys, Duckham (2004) conceituam que bancos de dados geográficos (BDG) são coleções de dados georreferenciados, manipulados por Sistemas de Informação Geográficas (SIG), que são sistemas computacionais capazes de capturar, modelar, armazenar, recuperar, manipular, analisar e apresentar dados geográficos A mineração de dados geográficos é um tópico de estudo da área de KDD que ainda está começando.

A importância da exploração de dados espaciais pode ser notada pelo fato dos dados geográficos, os quais são um tipo específico de dados espaciais, estarem sendo utilizados em vários domínios diferentes de aplicações, como planejamento urbano, meios de transporte, telecomunicações, marketing. Por ser uma área ainda em desenvolvimento e expansão, este projeto de intervenção realiza um apanhado geral sobre o tema e identifica as técnicas utilizadas para a explorar os dados geográficos a fim de gerar resultados como ferramenta de apoio a tomada de decisão.

O objetivo do presente artigo é apresentar os conceitos e técnicas relacionadas a aplicação de Datamining em Geoprocessamento de dados com exemplos de algumas aplicações apresentadas na literatura. A seção 2 conceitua Datamining explicando seus objetivos e especifica algumas técnicas que têm sido utilizadas e algumas aplicações em estudos anteriores e a seção 3 complementa os estudo com as devidas considerações.

2 DESENVOLVIMENTO

2.1 Fundamentação Teórica

A análise de grandes quantidades de dados pelo homem é inviável sem o auxílio de ferramentas computacionais apropriadas. Portanto, torna-se imprescindível o desenvolvimento de ferramentas que auxiliem o homem, de forma automática e inteligente, na tarefa de analisar, interpretar e relacionar esses dados para que se possa desenvolver e selecionar estratégias de ação em cada contexto de aplicação, e em se tratando da utilização da ferramenta de mineração de dados no meio empresarial, segundo Cardoso et. al. (2008), o objetivo do data mining é descobrir, de forma automática ou semiautomática, o conhecimento que está “escondido” nas grandes quantidades de informações armazenadas nos bancos de dados da organização, permitindo agilidade na tomada de decisão.

> Uma organização que emprega o data mining é capaz de: criar parâmetros para entender o comportamento dos dados, que podem ser referentes a pessoas envolvidas com a organização; identificar afinidades entre dados que podem ser, por exemplo, entre pessoas e produtos e ou serviços; prever hábitos ou comportamentos das pessoas e analisar hábitos para se detectar comportamentos fora do padrão entre outros. (CARDOSO et. al., 2008).

De forma salutar Colaço Jr. (2004) conceitua que a mineração de dados utiliza de estatística e de inteligência artificial bem estabelecidas para construir modelos que predizem o comportamento do cliente.

Goldschmidt, Ronaldo (2005) escreve que, em busca de atender a esse novo contexto foi criada uma nova área denominada Descoberta do Conhecimento em Banco de Dados (KDD - Knowledge Discovery from Data), o termo foi formalizado em 1989 em referência ao amplo conceito de procurar conhecimento a partir de bases de dados.

> KDD é um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados. (FAYYAD et al., 1996)

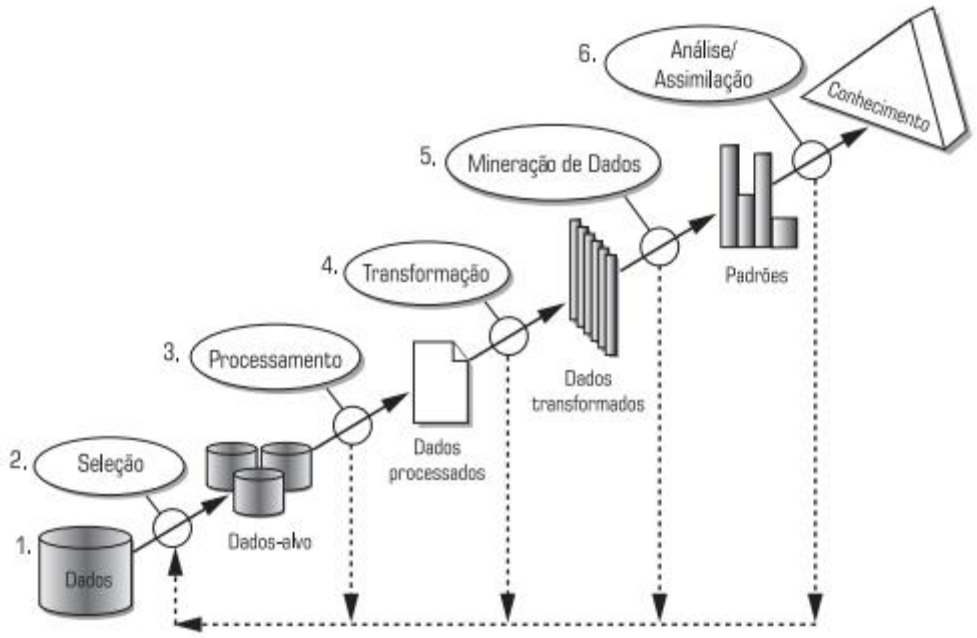
Segundo Han et al. (2011), Data Mining é uma das etapas do processo KDD, e consiste em um processo essencial, no qual técnicas

estatísticas e computacionais são aplicadas para identificar padrões em grandes volumes de dados. As tarefas de Data Mining podem ser classificadas em duas categorias com diferentes aplicações: descritivas (sumarização, análise de associações, segmentação) e preditivas (previsões e classificação).

O processo da descoberta de conhecimento é exemplificado na figura 1 pelo autor Colaço Jr. (2004), o objetivo foi dividi-la em seis partes, as três primeiras etapas iniciais: seleção, pré-processamento e transformação, que compõe o que denominou de preparação de dados. As demais etapas precedem a análise do conhecimento que será gerado a partir da etapa final de análise e assimilação de resultados.

Figura 1.

ETAPAS DO PROCESSO KDD



Fonte: Colaço Jr.

Data: 2004

Diniz, Neto (2000) dizem, para que esses dados sejam devidamente manipulados e, se tenha extração de informações importantes, faz-se necessário a utilização de técnicas que propiciem a automação desse processo a partir de estruturas artificialmente inteligentes, as quais envolvam técnicas necessárias para a compreensão da linguagem, percepção, raciocínio, aprendizagem e resolução de problemas, buscando a criação de teorias e modelos com

capacidade cognitiva e a implementação de sistemas computacionais baseados nestes modelos, objetivando a descoberta dos conhecimentos engendrados no banco de dados.

A maioria das técnicas de data mining derivam de três tipos básicos de modelos: físicos, característicos e cognitivos;

- O primeiro, com base em objetos pré-estabelecidos, replica objetos discriminatórios que podem ser observados de forma precisa numa base de dados;
- O segundo modelo classifica e/ou identifica eventos por meio de dados geográficos numa base de dados;
- O terceiro extrai informações de uma base de dados por meio de simulação e automatização do processo de tomada de decisão humano.

Dentre as técnicas de data mining, é possível destacar as seguintes: lógica fuzzy, redes neurais, algoritmos genéticos, análise de grupamento, espectrograma e procura relacional.

As ferramentas Data Mining identificam todas as possibilidades de correlações existentes nas fontes de dados. Através das técnicas para exploração de dados, pode-se desenvolver aplicações que venham extrair, dos bancos de dados, informações críticas, com o objetivo de subsidiar plenamente o processo decisório de uma organização.

A figura 2 demosntra o posicionamento lógico de diferentes fases da tomada de decisão com seu valor potencial para as dimensões tática e estratégica. Geralmente o valor da informação para apoiar a decisão aumenta a partir da base da pirâmide.

Uma decisão baseada em dados nas camadas mais baixas, onde há tipicamente milhões de registros de dados, não possui muito valor agregado, já aquela apoiada em dados altamente resumidos nas camadas superiores da pirâmide tem possibilidades de alto valor estratégico e indica quem são os usuários das informações em cada camada da pirâmide.

Na última década a utilização de técnicas de geoprocessamento cresceu substancialmente. Este crescimento pode ser observado pelo grande número de ferramentas criadas como Google Earth, Google Maps, ArcGIS que podem ser utilizadas na cartografia, fotogrametria, geoestatística, banco de dados geográficos, *WebMapping*, SIG e auxiliando também em tomadas de

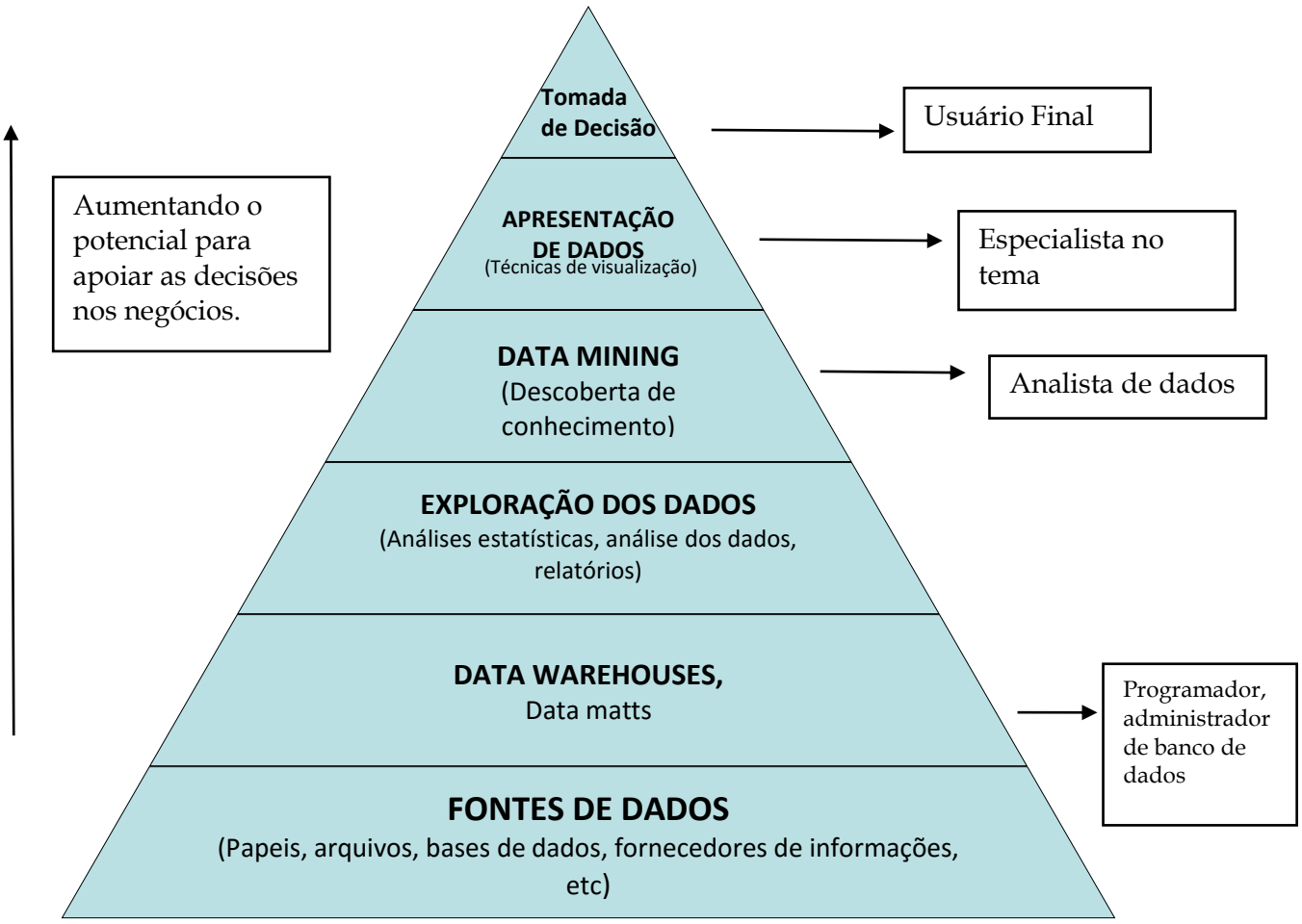
decisão fundamentadas nas informações extraídas através das ferramentas de mineração de dados cita Kusunoki (2013).

> A tecnologia de Sistemas de Informação Geográfica (SIG) provê ferramentas de software bastante poderosas para realizar o tratamento de dados geográficos. A principal diferença de um SIG para um sistema de informação convencional é sua capacidade de armazenar tanto os atributos descritivos como as geometrias dos diferentes tipos de dados geográficos (CASANOVA et al., 2005).

Dados geográficos são aqueles que possuem uma dimensão espacial, ou uma localização, diretamente ligada ao mundo geográfico real como

Figura 2.

EVOLUÇÃO DO VALOR ESTRATÉGICO DA BASE DE DADOS



Fonte: Baseado em Alvares et, al.

Data: 2001

as imagens de satélites de sensoriamento remoto, os dados de inventários cadastrais, os dados ambientais coletados em campo e os modelos numéricos de terreno, já os bancos de dados geográficos (BDG) são coleções de dados georreferenciados, manipulados por Sistemas de Informação Geográficas (SIG). Os SIG são sistemas computacionais capazes de capturar, modelar, armazenar, recuperar, manipular, analisar e apresentar dados geográficos. Worboys; Duckham (2004).

2.2 Contextualização do Tema

Num país como o Brasil, com uma grande carência de informações adequadas para a tomada de decisões sobre os problemas administrativos, urbanos, rurais e ambientais, o Geoprocessamento apresenta um grande potencial, de coleta, tratamento, manipulação e apresentação de informações espaciais para tal fim.

O termo Geoprocessamento traduz a disciplina do conhecimento que utiliza técnicas matemáticas e computacionais para o tratamento de informação geográfica e que vem influenciando de maneira crescente as áreas de Cartografia, Análise de Recursos Naturais, Transportes, Comunicações, Energia, Planejamento Urbano e Regional e acima de tudo à Saúde Pública, fala Baluz (2010).

Na última década a utilização de técnicas de geoprocessamento cresceu substancialmente. Este crescimento pode ser observado pelo grande número de ferramentas criadas como Google Earth, Google Maps, ArcGIS que podem ser utilizadas na cartografia, fotogrametria, geoestatística, banco de dados geográficos, *WebMapping*, SIG e auxiliando também em tomadas de decisão fundamentadas nas informações extraídas através das ferramentas de mineração de dados cita Kusunoki (2013).

> A tecnologia de Sistemas de Informação Geográfica (SIG) provê ferramentas de software bastante poderosas para realizar o tratamento de dados geográficos. A principal diferença de um SIG para um sistema de informação convencional é sua capacidade de armazenar tanto os atributos descritivos como as geometrias dos diferentes tipos de dados geográficos (CASANOVA et al., 2005).

Dados geográficos possuem uma dimensão espacial ou uma localização, diretamente ligada ao mundo geográfico real como as imagens de satélites de sensoriamento remoto, os dados de inventários cadastrais, os dados ambientais coletados em campo, os modelos numéricos de terreno e os bancos de dados geográficos (BDG) são coleções de dados georreferenciados, manipulados pelos Sistemas de Informação Geográficas (SIG). (Worboys e Duckham, 2004).

A implementação de todas as funcionalidades da análise espacial requer uma plataforma computacional capaz de reproduzir e reconhecer os relacionamentos espaciais entre as entidades modeladas. Conforme Câmara (2000), os SIG permitem a operacionalização de múltiplas ferramentas de análise espacial, tendo em vista a sua capacidade de modelar as relações topológicas entre entidades do mundo real.

Um SIG com um Data Mining implementado, por exemplo, pode facilitar a interpretação de padrões encontrados. Em outras palavras, os algoritmos encontram padrões, e o SIG os apresenta geograficamente, em mapas e gráficos, tornando mais fácil a interpretação, colaborando, assim, para uma melhor tomada de decisões.

2.3. Técnicas de Mineração de dados

As técnicas utilizadas para mineração de dados são várias, mas segundo Colaço Jr. (2004) as duas que mais se destacam são as Regras de Associação e Classificação, a primeira sendo geralmente uma técnica utilizada para descobrir vendas associadas de produtos nos mais diversos ramos, já a segunda procura formar classes a partir dos dados minerados.

Segundo Miller e Shaw (2001), padrões como classes, associações, regras e tendência possuem expressões espaciais bem definidas, tendo, assim, aplicações em técnicas para mineração de dados em bancos geográficos. A seguir, são descritas algumas técnicas que utilizam esses padrões:

• Classificação espacial – Essa técnica mapeia objetos geográficos em categorias com significados bem definidos, considerando relacionamentos de distância, direção ou conectividade e morfologia desses objetos..

• Associação espacial – Essa técnica define regras de associação espacial que definem contém predicados espaciais na parte precedente ou antecedente de uma cláusula.

• Predição espacial – Utiliza algoritmos de aprendizagem indutiva para extrair informações de mapas topográficos de propósito geral, como os produzidos por organizações nacionais ligadas à cartografia. Nesse método, uma heurística de busca constrói predicados locais baseados em objetos espaciais, conhecimento implícito, conceitos de alto-nível previamente definidos e critérios de performance.

Existem várias ferramentas para realizar mineração de dados em bancos de dados relacionais; por exemplo, o Intelligent Data Miner do DB2/IBM, MineSet da SiliconGraphics e o Clementine da SPSS. No entanto, os modelos de análise usados em mineração de dados convencionais precisam ser modificados ou adaptados para levar em consideração os diferentes estilos e necessidades de análise e modelagem relevantes para o mundo dos SIG.

Existem várias características que devem ser levados em consideração ao analisar dados geográficos (MILLER e HAN, 2001 apud FURTADO), tais como:

• Incerteza dos dados aliada ao fato de os erros serem frequentemente estruturados de forma espacial;

• Relacionamentos frequentemente estão relacionados a uma posição geográfica específica, em vez de serem globais;

• O tempo geralmente interage com o espaço;

• Alto-nível de multi-variedade nos dados, associada à redundância;

• O elemento localização é importante.

> Na maioria das técnicas de exploração de dados, os resultados apresentados são coisas já conhecidas. Por exemplo, um relatório mostrando a diminuição das vendas de certos produtos em uma região pode ser direto para o usuário porque ele intuitivamente sabe que esse tipo de informação já existe no banco de dados. Técnica de Data Mining, por outro lado,

> extraem informação útil, valiosa e previamente desconhecida. (Colaço Jr. 2004).

A implementação destas funcionalidades tem sido viabilizada por vários modelos que buscam representar o espaço e o tempo, em ambiente SIG, de forma integrada e complementar Zhao et al. (1997) e também os autores Miller e Shaw (2001) descrevem seis modelos usados para incorporar a dimensão temporal ao SIG: modelo snapshot, modelo de atualização, modelo espaço-tempo, modelo multidimensional, modelo integrado e modelo de versões.

Os padrões espaciais têm propriedades únicas que são interessantes para o *data mining*: são frequentemente ao-correlacionados; Possuem domínios complexos e descrevem objetos contínuos diz Andrade et. al (2010). como o data mining é um campo em desenvolvimento, então além de haver diversas famílias para os padrões espaciais, ainda podem estar surgindo mais. Entretanto três delas são usualmente utilizadas:

- Predição de localização: a técnica adotada é a classificação, que tem como principal característica a resposta ao questionamento: Onde um certo fenômeno irá acontecer?
- Interação espacial: As técnicas adotadas são correlação, associação e colocations, que tem como principal característica a resposta ao questionamento: Quais subconjuntos de objetos ou fenômenos espaciais interagem?
- Hot spots: As técnicas adotadas são clusterings e detecção de desvios, que tem como principais características as respostas ao questionamento: Quais localizações são não usuais?

A visualização dos resultados de operações de análise espacial sobre dados dinâmicos é fundamental para o processo de tomada de decisão. Miller e Shaw (2001) destacam dois tipos básicos de visualização espacial: a cartográfica convencional e a visualização realista.

A primeira inclui mapas bi e tridimensionais, além de modelos de representação 3D, relativos a um ou mais temas.

A segunda usa técnicas de animação, simulação e realidade virtual para representar objetos e fenômenos do mundo real, explorando as percepções sensoriais humanas e reduzindo a necessidade da abstração.

2.4. Exemplo de aplicação prática

Guidini e Ribeiro (2006) 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, utilizando o software WEKA, onde comprovou-se a possibilidade e a aplicabilidade da fusão de tecnologias, gerando aplicações úteis à tomada de decisão baseada em geoprocessamento de dados.

Francisco e Reina (2007) criaram modelos de influência espacial (GIS) e busca de padrões e tendências com Data Mining visando uma melhor caracterização sócio-econômico-demográfica dos clientes e do Mercado de Concessionária, o que propiciou melhor eficácia em estudos sobre Fraude, Análise de Inadimplência, Contratação de Demanda por região, etc., otimizando e conduzindo ao reposicionamento estratégico dos processos de negócios junto ao mercado nos diversos cenários que se apresentem.

O desenvolvimento do projeto proporcionou o direcionamento do novo sistema de gestão de consumidores da empresa e as gerências de gestão de recebíveis, identificação de perdas comerciais e regularização de clandestinos estão utilizando os resultados em cada um dos seus processos para o desenvolvimento de estratégias diferenciadas a partir do modelo de segmentação comportamental.

3 CONSIDERAÇÕES FINAIS

Devido ao fato de existirem muitos sistemas legados com dados disponíveis em vários formatos, com novos tipos de dados constantemente surgindo, é bastante difícil se chegar a um padrão comum, apesar de existirem muitos esforços nesse sentido. Mineração de dados geográficos é um tópico de estudo da área de Descoberta do Conhecimento (Knowledge Discovery) que ainda está começando.

Um Data Mining pode oferecer novas formas de análise a bases de dados com informações espaciais, além de os resultados da mineração poderem ser visualizados geograficamente. A pesquisa nessa área tende a crescer, visto que cada vez mais cientistas estão dando especial atenção aos dados

geográficos e, com isso, novas técnicas para explorar esses dados vem sendo desenvolvidas. Além disso, o crescimento do poder das plataformas computacionais contribui bastante para viabilizar a aplicação de novos métodos.

ABSTRACT

This article presents an overview of the application of techniques of Intelligent Systems in Geographic Information Systems (GIS) and its Geographic Databases. Where to find different types of knowledge , you can use the Knowledge Discovery process in databases on an alternative approach to data and spatial relationships by applying the classic method of data mining through some existing techniques. In particular, this work provides an explanation of systems based on data mining rules in Geographic Information Systems and idealizes use these possible application of alternatives in the context of decision-making and explores some clever approaches in georeferenced systems.

Key words:
Data Mining, Geographic Information System, Geographical Database, Intelligent Systems, KDD.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, Gabriel Arthur Gerber; GONZAGA, Jorge Júnior. **Data Mining em Bancos de Dados Espaciais**. Departamento de Informática e Estatística – INE, Universidade Federal de Santa Catarina, 2010. Disponível em < http://www.inf.ufsc.br/~alvares/INE5644/G10_DM_DadosEspaciais.pdf > Acesso em: 15 de agosto de 2015.

ALVARES, Lillian; ARAUJO, Rogério Henrique Jr.; TARAPANOFF, Kira; QUONIAM, Luc. **Inteligência obtida pela aplicação de datamining em bases de teste francesas no Brasil.** Brasília, 2001. Disponível em < https://www.researchgate.net/publication/26349917_Inteligencia_obtida_pela_aplicacao_de_data_mining_em_base_de_teses_francesas_sobre_o_Brasil> Acesso em: 15 de agosto de 2015.

CARDOSO**,** Olinda Nogueira Paes; MACHADO, Rosa Teresa Moreira. **Gestão do conhecimento usando *data mining*: estudo de caso na Universidade Federal de Lavras.** Fgv: revista de administração: Rio de janeiro, 2008. Disponível em < http://repositorio.ufla.br/bitstream/1/2524/1/DISSERTA%C3%87%C3%83O_Gest%C3%A3o%20do%20conhecimento%20usando%20Data%20Mining.pdf > Acesso em: 14 de agosto de 2015.

COLAÇO JR, Methanias. **Projetando sistemas de apoio a tomada de decisão baseados em Dataware House.** Axcel Books. Rio de Janeiro, 2004.

CORADINE, Luis Cláudius; LOPES, Roberta Vilhena Vieira; MACIEL, Andrilene Ferreira. **Mineração de Dados: Uma Introdução.** Journal of the Brazilian Neural Network Society, Vol. 9. Sociedade Brasileira de Redes Neurais (SBRN) 168 Universidade Federal de Alagoas – UFAL Instituto de Computação. Disponível em: <http://www.lnlm-sbic.org/papers/vol9-no3-art3.pdf > Acesso em: 18 de agosto 2015.

Diniz, C.A.R. e Louzada-Neto, F. (2000), Hair Jr., J.F., Anderson, R.E., Tatham, R.L. e Black, W.C. (2005). **Análise Multivariada de Dados**, Tradução, 5ª ed., Bookman, Porto Alegre.

DINIZ, Carlos Alberto R., NETO, Francisco Louzada; **Data Mining: Uma Introdução**, ABE - Associação Brasileira de Estatística, São Carlos – SP, 2000. Disponível em: < http://www.lnlm-sbic.org/papers/vol9-no3-art3.pdf> Acesso em: 17 de agosto 2015.

FAYYAD, U.M. et al. Advances in knowledge discovery and data mining. **Massachusetts: AAAI Press,** 1996**.**

FRANCISCO, Eduardo de Rezende; REINA, Claudia Sciortino de. **Definição de Metodologia para Segmentação de Clientes, Identificação de Grupos e conhecimento do Mercado "Clusters BT". –** 3º Congresso Anual de Tecnologia da Informação da Fundação Getulio Vargas: São Paulo, 2006.

FURTADO, Antônio; GAZOLA, L. Alexandre **Bancos de Dados Geográficos Inteligentes** Monografia em Ciência da Computação.

GUIDINI, M. P.; **RIBEIRO**, C. H. C. Utilização da biblioteca TerraLib para algoritmos de agrupamento em Sistemas de Informações Geográficas. In: SIMPÓSIO BRASILEIRO DE GEOINFORMÁTICA. Campos do Jordão, 2006. Disponível em: <http://urlib.net/83LX3pFwXQZ3V9uMbiY/Nibz8>. Acesso em: 05 ago. 2016.

HAN, J.; KAMBER, M.; PEI, J. **Data mining**: concepts and techniques. San Francisco: Morgan Kaufmann Publishers, 2011.

KUSUNOKI, Breno Naodi. **Identificação de Padrões de Itinerários e de Pontos Intermediários em Dados Espaço-Temporais Utilizando Busca Por Similaridade** Universidade Estadual de Londrina. LONDRINA–PR, 2013**.** Disponível em: <http://www.uel.br/cce/dc/wp-content/uploads/VersaoPreliminarTCC-BrenoKusunoki.pdf> Acesso em: 18 de agosto 2015.

LOH, Stanley**. BI na era do big data para cientistas de dados - indo além de cubos e dashboards na busca pelos porquês, explicações e padrões.** Porto Alegre, 2014. Disponível em: <http://www.intext.com.br/livro-BI-stanley-loh.pdf> Acesso em: 18 de agosto 2015.

NETO, Marco Aurélio Silva; VILLWOCK, Rosangela; SCHEER, Sérgio; STEINER Maria Teresinha Arns; DYMINSKI, Andréa Sell. **Técnicas de mineração visual de dados aplicadas aos dados de instrumentação da barragem de Itaipu.** *Visual data mining Gest. Prod.*, São Carlos, v. 17, 2010**.** Disponível em: < http://www.scielo.br/scielo.php?pid=S0104-530X2010000400007&script=sci_arttext > Acesso em: 18 de agosto 2015

WORBOYS, M.; DUCKHAM, M. GIS A Computing Perspective Second Edition. Boca Raton, Florida, USA: CRC Press, 2004.

VINHAS, Lúbia. **Um Subsistema Extensível para o Armazenamento de Geo-Campos em Bancos de Dados Geográficos**. 2006. 114 f. Tese (Doutorado) –

Instituo Nacional de Pesquisas Espaciais, São José dos Campos - SP, 2006. Disponível em: <http://mtc-m17.sid.inpe.br/col/sid.inpe.br/MTC-m13@80/2006/06.12.17.17/doc/publicacao.pdf > Acesso em: 15 de agosto 2015.

ZUCHINI, M.H. **Aplicações de mapas auto-organizáveis em mineração de dados e recuperação de informações**, dissertação de mestrado, Unicamp, S.P, 2003. Disponível em: < ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/theses/zuchini_mest/zuchini_mest.pdf> Acesso em: 17 de agosto 2015.