

Técnicas de mineração visual de dados aplicadas aos dados de instrumentação da barragem de Itaipu

Visual data mining techniques applied for the analysis of data collected at Itaipu power plant



Marco Aurélio Silva Neto³
Rosângela Villwock⁴
Sérgio Scheer¹
Maria Teresinha Arns Steiner²
Andréa Sell Dyminski¹

Resumo: A usina de Itaipu, maior geradora de hidroeletricidade do mundo, possui mais de 2.200 instrumentos que monitoram seu comportamento geotécnico e estrutural, os quais possuem leituras armazenadas em um banco de dados há mais de 30 anos. A alta dimensionalidade e a grande quantidade de registros contidos nas bases de dados são problemas não triviais tendo-se em vista a busca pelo “conhecimento” a partir desses dados. Neste artigo é apresentado um estudo no qual foram utilizados algoritmos de Mineração Visual de Dados (MVD), integrando técnicas de Mineração de Dados (MD) com técnicas de Visualização de Informação (VI), para analisar estes dados de instrumentação, tendo como objetivo estabelecer relações existentes entre diferentes instrumentos, que possibilitem detectar falhas indesejáveis no sistema de monitoramento e possíveis danos à segurança e integridade da barragem. Um maior número de informações pôde ser extraído mais facilmente quando diferentes técnicas de VI, juntamente com as de MD, são aplicadas aos dados. A análise visual dos dados mostrou-se eficiente por agilizar a detecção de anomalias nos dados, sendo uma valiosa ferramenta de apoio à tomada de decisões.

Palavras-chave: Processo KDD. Visualização de informações. Mineração de dados. Mineração visual de dados. Monitoramento de barragens.

Abstract: *Itaipu, the greatest hydroelectric power plant in operation in the world, has more than 2,200 monitoring instruments, which have been storing readings in databases for more than 30 years. The large data sets of high dimensionality and the large amount of records inserted into databases are non-trivial problems when conducting a search for “knowledge” through the data. This paper introduces a study using Visual Data Mining (VDM) algorithms integrating Data Mining (DM) techniques with Visualization of Information (VI) techniques to analyze the data collected at Itaipu dam. The main objective was to establish relationships between the variables in order to detect undesirable failures that can compromise the security and integrity of the dam. More information may be more easily extracted when different techniques of Visualization of Information, together with techniques of Data Mining, are applied for data analysis. The visual analysis of the data has proved efficient in detecting patterns of anomalies, and thus it can be considered a valuable tool to support decision making.*

Keywords: *KDD process. Information visualization. Data mining. Visual data mining. Dam monitoring.*

1 Introdução

Uma vez que os potenciais prejuízos e riscos decorrentes de acidentes em barragens podem assumir grandes dimensões, um projeto seguro, uma construção adequada e a correta operação de barragens são preocupações de âmbito mundial. Além disso, um efetivo monitoramento em grandes barragens

¹ Departamento de Construção Civil, Programa de Pós-Graduação em Métodos Numéricos em Engenharia – PPGMNE, Universidade Federal do Paraná – UFPR, CP 19081, CEP 81531-990, Curitiba – PR

² Coordenação da Engenharia de Produção, Programa de Pós-Graduação em Métodos Numéricos em Engenharia – PPGMNE, Universidade Federal do Paraná – UFPR, CP 19081, CEP 81531-990, Curitiba – PR

³ Instituto Tecnológico SIMEPAR, Centro Politécnico da UFPR, Programa de Pós-Graduação em Métodos Numéricos em Engenharia – PPGMNE, Universidade Federal do Paraná – UFPR, Jardim das Américas, CP 19100, CEP 81531-980, Curitiba – PR

⁴ Programa de Pós-Graduação em Métodos Numéricos em Engenharia – PPGMNE, Universidade Federal do Paraná – UFPR, Universidade Estadual do Oeste do Paraná – UNIOESTE, Campus Francisco Beltrão, Rua Maringá, 1200, Vila Nova, CEP 85605-010, Francisco Beltrão – PR

é imprescindível para a segurança da estrutura. Diretrizes internacionais visando a segurança de barragens e muitas discussões produtivas sobre este tema têm sido propostas e conduzidas, tais como a da Comissão Internacional de Grandes Barragens (ICOLD – *International Commission on Large Dams*) (INTERNATIONAL..., 2008). No Brasil, diretrizes visando à segurança de barragens foram publicadas pelo Comitê Brasileiro de Grandes Barragens em 1983 (COMITÊ..., 1983) e um projeto de lei sobre assunto tramita pelo congresso nacional desde 2003 (BRASIL, 2009).

O monitoramento da estrutura de uma barragem, de importância bem conhecida, pode gerar uma enorme massa de dados, definidos em domínios multidimensionais, cuja análise e interpretação nem sempre são triviais. A análise minuciosa dos dados dos instrumentos de auscultação exige a combinação de conhecimentos de engenharia com matemática e estatística, bem como experiência prévia do engenheiro ou técnico responsável pela interpretação destes dados, consumindo muito tempo e muitas vezes inviabilizando o cumprimento rápido desta tarefa. Por esta razão, o uso de técnicas e ferramentas computacionais, que auxiliem o tomador de decisões, é relevante.

O objetivo do presente artigo é analisar a integração de técnicas de Mineração de Dados (MD) com técnicas de Visualização de Informações (VI) utilizando dados reais obtidos de instrumentos de monitoramento da barragem de Itaipu. O resultado da integração destas técnicas é conhecida na literatura por Mineração Visual de Dados (MVD) (WONG, 1999). O uso de técnicas e ferramentas para a MVD permite ao analista extrair informações existentes e “escondidas” nos dados, além de identificar grupos, anomalias e correlações de padrões, de forma visual, rápida e efetiva.

Num grande volume de dados, frequentemente há informação escondida e que não está prontamente evidente. Segundo Tan et al. (2005), analistas humanos podem levar semanas para descobrir este conhecimento e muitos dados nunca são analisados. Enquanto aumenta a quantidade de dados, aumenta a aplicação de técnicas de MVD. Para Witten e Frank (2000), dados inteligentemente analisados são um valioso recurso para tomada de decisões.

Este artigo está organizado da seguinte forma: na seção 2, são apresentadas as principais informações sobre a barragem de Itaipu e sobre os extensômetros, instrumentos cuja base de dados foi aqui analisada. Na seção 3, aborda-se a questão da extração de conhecimento a partir de bases de dados, ou seja, sobre o processo *KDD* (*Knowledge Discovery in Databases* ou Descoberta de Conhecimento em Bases de Dados) e é apresentada uma introdução ao tema de agrupamento de dados (padrões ou instrumentos no problema aqui apresentado), forma utilizada para o tratamento. Na seção 4, encontra-se o detalhamento

da metodologia adotada; na seção 5, são mostrados e analisados os resultados, seguidos das conclusões, na seção 6.

2 A barragem de Itaipu e obtenção dos dados

Nesta seção é feita uma apresentação sucinta da Itaipu Binacional, maior hidrelétrica em produção de energia do mundo, enfatizando a sua instrumentação e, na sequência, são apresentados os dados sobre os instrumentos utilizados no presente estudo.

A Itaipu Binacional teve o início da sua construção em 1973 em um trecho do Rio Paraná conhecido por Itaipu que, em tupi, quer dizer “a pedra que canta”, localizado no coração da América do Sul na divisa entre o Paraguai e o Brasil (ITAIPU, 2008). Em 1982, chegaram ao fim as obras da barragem, sendo a última unidade geradora inaugurada em 2008.

Atualmente, a barragem de Itaipu possui 20 unidades geradoras de 700 MW (megawatts) cada, gerando uma potência total instalada de 14.000 MW. No ano 2000, a Itaipu Binacional bateu seu recorde em geração de energia, cerca de 93,4 bilhões de quilowatts-hora (KWh). É responsável pelo abastecimento de 95% da energia elétrica consumida no Paraguai e 24% de toda a demanda do mercado brasileiro.

A barragem de Itaipu possui 7.919 m de extensão e altura máxima de 196 m, dimensões que transformaram esta obra em referência nos estudos de concreto e na segurança de barragens. É composta por dois trechos de barragens de terra, um trecho de barragem de enrocamento e trechos de concreto, compondo as estruturas mais altas do conjunto. A Figura 1, a seguir, mostra um esboço da estrutura geral da barragem e a Tabela 1 apresenta as principais características dos trechos da barragem apontados na Figura 1.

Em toda sua extensão, para acompanhar o desempenho das estruturas de concreto e fundação, são encontrados 2.218 instrumentos (1.362 no concreto e 856 nas fundações e aterros) sendo que destes, 270 estão automatizados; além disso, existem 5.239 drenos (949 no concreto e 4.290 nas fundações). Todas estas leituras ocorrem em diferentes frequências, podendo ser, por exemplo, diária, semanal, quinzenal, mensal, dependendo do tipo de instrumento.

Embora os nove trechos da barragem de Itaipu sejam instrumentados e monitorados, um trecho da Barragem Principal (trecho F) merece destaque e um estudo mais aprofundado. No trecho F encontram-se as turbinas para a geração da energia elétrica, além de ser o trecho de maior altura em coluna de água e o mais instrumentado. Este trecho é constituído de vários blocos, sendo que cada um deles possui instrumentos que fornecem dados a respeito de seu comportamento físico, tanto na estrutura de concreto como na sua fundação.

Por estes motivos, optou-se por estudar o trecho F de Itaipu, no qual se encontram 30 extensômetros, 74 piezômetros, 11 medidores triortogonais, 4 medidores de nível de água e 9 medidores de vazão (SILVA NETO, 2008). Destes, foram selecionados para análise os extensômetros, que são do tipo múltiplo de hastes, instalados em furos de sondagem. Este tipo de instrumento é considerado um dos mais importantes para o monitoramento da barragem, pois são responsáveis pelas medições de recalques de uma barragem de concreto. Estes recalques, ou deslocamentos verticais, consistem em uma das observações mais importantes na supervisão do comportamento da estrutura durante os períodos de construção, enchimento do reservatório e operação da barragem. Segundo Silveira (2003), nos blocos onde há galerias de acesso transversais ao eixo, a instalação de extensômetro a montante e a jusante permite medir deslocamentos angulares da barragem junto à fundação.

Também se utilizam extensômetros para medir deslocamentos verticais do maciço rochoso basáltico, onde a fundação da barragem tem sua base. São importantes no monitoramento de uma barragem, sendo dada especial atenção a descontinuidades no maciço rochoso, tais como juntas, falhas e contatos entre rochas. Cada extensômetro é instalado em um local específico e pode ser composto por várias hastes de comprimentos diferentes. Assim, é possível controlar separadamente o deslocamento vertical de cada descontinuidade geológica, conforme se pode observar na Figura 2.

Vale salientar que o estudo aqui desenvolvido a respeito dos extensômetros pode ser facilmente estendido de forma similar aos demais instrumentos.

2.1 Coleta de dados

Os dados com os quais o presente trabalho foi desenvolvido referem-se, conforme já mencionado, aos instrumentos de monitoramento da barragem de Itaipu, mais especificamente, aos extensômetros instalados no trecho F da barragem. Neste trecho F, esboçado na

Figura 1, estão localizados 30 extensômetros, cada um com uma, duas ou três hastes totalizando, assim, 72 medidas de deslocamento. Como estas 72 medidas são independentes, serão aqui denominadas de hastes de extensômetros. Estas medidas são aqui identificadas da seguinte forma: equip4_1, que significa haste 1 do extensômetro 4 e assim por diante.

Os registros históricos das hastes de extensômetros, contidos no banco de dados da empresa, contêm dados mensais, datados de janeiro/1995 a dezembro/2004, totalizando 120 leituras. O período foi assim determinado por sugestão da equipe de engenheiros da Itaipu, pois é posterior à construção da hidrelétrica e anterior à implantação do sistema de aquisição automática de dados. Vale salientar que, durante a fase de instalação deste sistema, alguns instrumentos ficaram sem leituras manuais e que, além disso, as 24 hastes de extensômetros já automatizadas sofreram modificações que podem ter influenciado as leituras posteriores. Desta forma, as referidas 120 leituras mensais referentes às 72 hastes de extensômetros ficaram isentas destas irregularidades.

Estes 120 registros históricos foram analisados segundo o processo *KDD*, descrito em linhas gerais na seção 3, a seguir, particularizado para o problema aqui descrito, na seção 4.

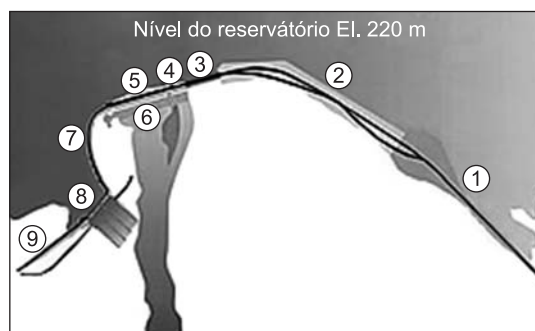


Figura 1. Um esboço da estrutura geral do complexo de Itaipu (ITAIPU, 2008).

Tabela 1. Características dos trechos da Barragem de Itaipu com destaque ao trecho F analisado neste estudo.

	Trecho	Estrutura	Comprimento (m)	Altura máxima (m)
1	L-Barragem Auxiliar	Terra	2294	30
2	K-Barragem Auxiliar	Enrocamento	1984	70
3 e 7	E, I e D-Barragens Laterais	Contraforte	1438	81
4	H-Estrutura de Desvio	Concreto Maciço	170	162
5	F-Barragem Principal	Gravidade Aliviada	612	196
9	Q-Barragem Auxiliar	Terra	872	25
	Outros Trechos		Características	
8	A-Vertedouro		350 m de largura	
6	U-Casa de Força		20 Unidades Geradoras	

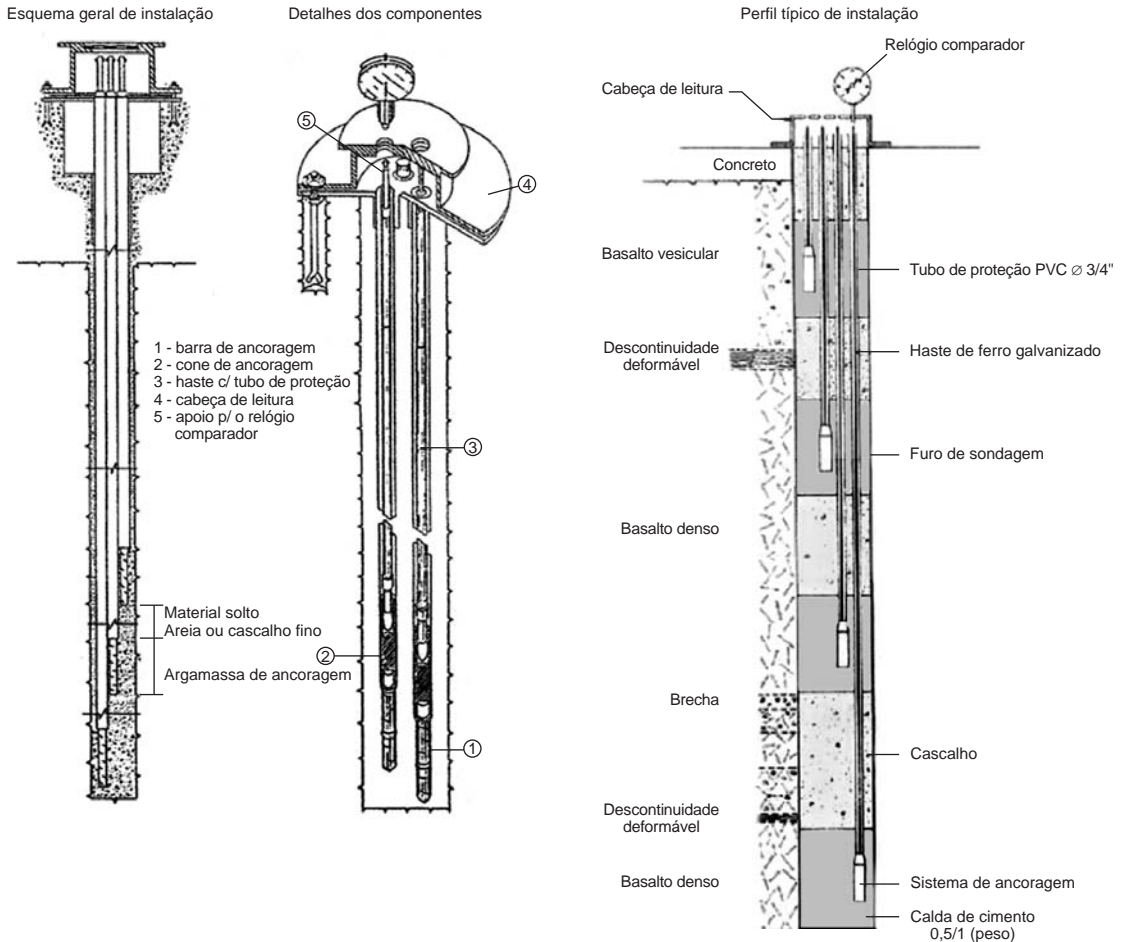


Figura 2. Ilustração de um extensômetro múltiplo de haste (MATOS, 2002), instrumento utilizado para o desenvolvimento do presente estudo.

3 O processo Descoberta de Conhecimento em Bases de Dados – *KDD*

O processo *KDD* é definido por Fayyad et al. (1996) 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.

KDD refere-se ao amplo processo de descoberta de informações em banco de dados, no qual se enfatiza a aplicação de alto nível dos métodos de MD. Enquanto a etapa de MD se destaca pela extração de padrões escondidos nos dados, o processo completo *KDD* é mais amplo e abrange várias etapas necessárias para que isso ocorra, tornando possível, após técnicas de MD, avaliar e interpretar os resultados obtidos.

O processo *KDD* é um conjunto de atividades contínuas composto, basicamente, por cinco etapas: Seleção dos Dados, Pré-Processamento e Limpeza,

Formatação, Mineração de Dados e Interpretação, conforme ilustrado na Figura 3.

O processo inicia-se com o entendimento do domínio da aplicação e dos objetivos a serem atingidos. Então uma seleção poderá ser realizada nestes dados a fim de se trabalhar com os dados de interesse. A etapa de pré-processamento é responsável pela análise e tratamento de dados faltantes ou inconsistentes. Na etapa de formatação, há uma preparação dos dados visando a aplicação da MD como, por exemplo, transformando dados categóricos em valores numéricos ou, então, usando métodos de redução de dimensionalidade nos dados. As etapas de pré-processamento e formatação, segundo Silver (1996), podem levar até 80% do tempo necessário para todo o processo.

Prosseguindo no processo, chega-se à etapa de MD, principal etapa do processo *KDD*, na qual diversos métodos podem ser usados para extração de informações, as quais, por sua vez, são apresentadas na

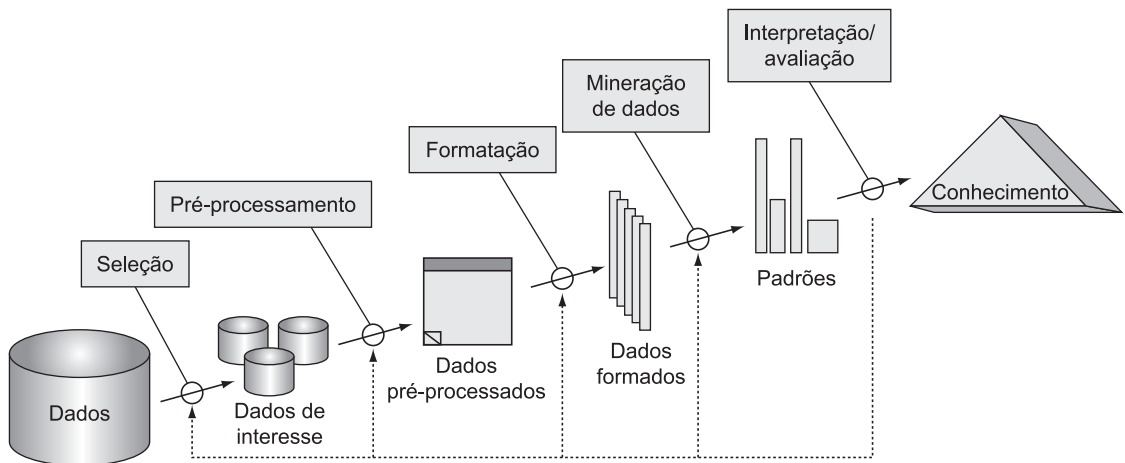


Figura 3. Etapas do processo KDD (FAYYAD et al., 1996).

última etapa, a interpretação, em que o conhecimento é adquirido. Se o resultado final não for satisfatório, todo o processo pode ser realimentado alterando algumas informações as quais podem ser reprocessadas nas etapas anteriores.

O principal objetivo do processo KDD é obter o conhecimento de informações escondidas nos dados que sejam úteis nas tomadas de decisões, utilizando métodos, algoritmos e técnicas de diferentes áreas científicas que, segundo Tan et al. (2005) e Gimenes (2000), incluem aprendizagem de máquina, reconhecimento de padrões, estatística e matemática, aquisição de conhecimento para sistemas especialistas e visualização de dados.

Já a VI é um processo indispensável na etapa de MD, segundo Rezende (2003). Para Keim (1979), a VI permite ao usuário adquirir percepções sobre os dados, podendo provocar o surgimento de novas hipóteses. Este autor acrescenta que quando comparada às técnicas automáticas de MD, como Estatística e Aprendizagem de Máquina, a exploração visual dos dados apresenta vantagens excedentes, tais como: lida mais facilmente com dados altamente heterogêneos e ruidosos, é intuitiva, e não requer maior entendimento de complexos algoritmos ou parâmetros da matemática ou estatística. Ankerst e Keim (2001) definem MVD, como sendo um passo no processo KDD, utilizando a VI como um canal de comunicação entre o computador e o usuário.

A essência deste artigo é apresentar algumas das técnicas de MD para o agrupamento de padrões (extensômetros) e de VI (melhor visualização dos resultados), verificando o relacionamento entre os padrões de um mesmo grupo, conforme detalhamento apresentado nas seções a seguir.

3.1 Agrupamento de padrões

Fayyad et al. (1996) mostram que as tarefas de MD podem ser, de uma forma geral, preditivas ou descritivas. As preditivas usam algumas variáveis para prever valores desconhecidos ou futuros de outras variáveis, enquanto que as descritivas encontram padrões para descrever os dados. As principais tarefas de MD estão relacionadas à Classificação, Agrupamento e Associação de padrões. Conforme já comentado, neste artigo, a tarefa da MD é o agrupamento de padrões.

O Agrupamento ou Segmentação (*Clustering*) procura grupos de padrões tais como padrões pertencentes a um mesmo grupo são mais similares uns aos outros e dissimilares a padrões em outros grupos. Segundo Hair Jr. et al. (2005), a análise de agrupamentos é uma técnica analítica para desenvolver subgrupos significativos de objetos. Seu objetivo é classificar os objetos em um pequeno número de grupos mutuamente excludentes. Para Freitas (2002), na análise de agrupamento é importante favorecer um número pequeno de grupos.

Os algoritmos de agrupamentos podem ser divididos em categorias de diversas formas de acordo com algumas características. As duas principais classes de algoritmos de agrupamentos são os métodos hierárquicos, dentre os quais se destaca o Método de Ward (da Análise Estatística Multivariada), utilizado neste estudo e métodos de particionamento como, por exemplo, o método *k*-médias.

Métodos hierárquicos englobam técnicas que buscam hierarquicamente os grupos e, por isso, admitem obter vários níveis de agrupamento. Os métodos hierárquicos podem ser subdivididos em divisivos ou aglomerativos. O método hierárquico aglomerativo considera, a princípio, cada padrão como um grupo e, iterativamente, agrupa o par de

grupos com maior similaridade em um novo grupo até formar um único grupo contendo todos os padrões. O método hierárquico divisivo, encontrado em Diniz e Louzada-Neto (2000), ao contrário, inicia com um único grupo e executa um processo de sucessivas subdivisões.

Outra forma de analisar a formação de grupos, fazendo o relacionamento entre padrões, é por meio de técnicas de VI, em que parâmetros visuais, como cores, forma, tamanho, posição são usados para representar características e propriedades dos dados (ARTERO, 2005; KEIM, 1996). Outros parâmetros estudados em Computação Gráfica, como tipo de material, luminosidade e transparência, e técnicas interativas, como rotações e passeios por meio dos dados, convidam os usuários a explorar e manipular sistemas de informações grandes e complexos, também pode ser usado para facilitar a busca do “conhecimento” (SILVA NETO, 2008).

Com base nestas formas de interação e navegação, diversas técnicas de VI estão sendo desenvolvidas com o objetivo de facilitar a interpretação de dados. Silva Neto (2008) utiliza estas técnicas baseando-se nas características e natureza dos dados. Um levantamento das técnicas existentes mostrando suas vantagens e desvantagens podem ser encontradas no trabalho deste autor.

Neste artigo, a técnica de MD utilizada para o agrupamento de padrões foi o Método de Ward. Já as técnicas de VI aplicadas aos dados foram Coordenadas Paralelas, Matrizes de Dispersão, Orientadas a Pixels e baseadas em Ícones. Estas técnicas utilizadas, enquadradas na metodologia aplicada ao problema, são apresentadas na seção 4 a seguir.

4 Metodologia adotada

O processo KDD, apresentado de forma genérica na seção 3 e Figura 3, pode ser particularizado para o problema aqui abordado, por meio da Figura 4 que compõe a metodologia aqui proposta. Tal metodologia foi aplicada aos dados de instrumentação da barragem de Itaipu, mais especificamente às 72 hastes de extensômetros que, por sua vez, geraram 120 registros históricos, conforme já comentado na seção 2.1 (1ª etapa da Figura 4).

Para a maioria das 72 hastes de extensômetros analisadas, tem-se uma leitura mensal, porém algumas delas apresentaram mais de uma leitura por mês sendo que, nestes casos, foi obtida a média mensal. Por outro lado, algumas hastes de extensômetros apresentaram leituras faltantes e, nestas situações, foram realizadas interpolações por séries temporais garantindo que todos as 72 hastes tivessem exatamente 120 leituras (2ª. etapa da Figura 4). Para um melhor detalhamento a respeito de técnicas de interpolação envolvendo Séries Temporais, sugere-se o livro de Box e Jenkins, (1976).

Já para a 3ª. etapa, o agrupamento dos padrões (ou hastes de extensômetros), o método utilizado foi o Método de Ward, um método hierárquico aglomerativo. Segundo Johnson e Wichern (1998), o Método de Ward faz a junção de dois grupos baseando-se na “perda de informação”. Considera-se como critério de “perda de informação” a soma do quadrado do erro (SQE ou $SQE - Square Quadratic Error$). Para cada grupo i , calcula-se a média (ou centroide) do grupo e a soma do quadrado do erro do grupo i (SQE_i), que é a soma do quadrado do erro de cada padrão do grupo em relação à média. Assim, para k grupos, tem-se $SQE_1, SQE_2, \dots, SQE_k$, em que SQE é definida pela Equação 1.

$$SQE = SQE_1 + SQE_2 + \dots + SQE_k \quad (1)$$

Para cada par de grupos m e n , primeiramente, calcula-se a média (ou centroide) do grupo formado (grupo mn). Em seguida, calcula-se a soma do quadrado do erro do grupo mn (SQE_{mn}), segundo a Equação 2.

$$SQE = SQE_1 + SQE_2 + \dots + SQE_k - SQE_m - SQE_n + SQE_{mn} \quad (2)$$

Os grupos m e n que apresentarem o menor aumento na SQE , ou seja, tiverem menor “perda de informação”, serão unidos. Segundo Hair Jr. et al. (2005), este método tende a obter grupos de mesmo tamanho devido a sua minimização de variação interna.

Uma forma de representar um agrupamento hierárquico é por meio de um dendrograma. Um

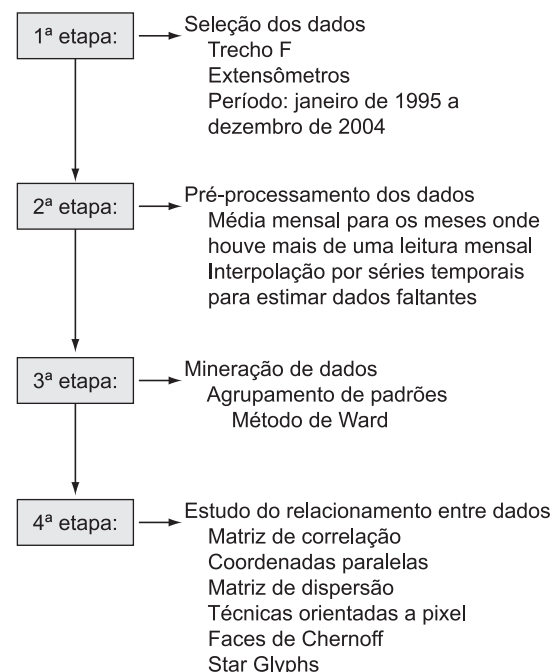


Figura 4. Etapas da metodologia adotada.

dendrograma representa o agrupamento dos padrões e os níveis de similaridade em que os grupos se formam. Pode ser “quebrado” em diferentes níveis, que para Jain et al. (1999), faz a subdivisão em diferentes grupos.

A similaridade entre variáveis pode ser medida a partir do coeficiente de correlação. Desta forma, este coeficiente r , que é calculado pela Equação (3), a seguir, é uma medida pertencente ao intervalo $[-1, 1]$ que identifica o nível de correlação entre pares de variáveis. Quanto mais próximo dos extremos, “-1” ou “1”, maior é a relação entre os dados. Quando os dados são próximos de “-1”, as variáveis são inversamente correlacionáveis, ou seja, os valores crescentes de uma das variáveis estarão associados aos valores decrescentes da outra, e quando este coeficiente é próximo de “1”, estas variáveis possuem comportamentos próximos. Coeficientes próximos de zero sugerem que as variáveis não possuem relação (KACHIGAN, 1986). Assim, tem-se:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_x S_y}, \quad 1 \leq i \leq n \quad (3)$$

em que \bar{x} representa a média do conjunto de valores de x e é definido pela Equação (4); \bar{y} representa a média do conjunto de valores de y e é definido pela Equação (5); $(x_i - \bar{x})$ representa o desvio entre x_i e a média do conjunto \bar{x} ; $(y_i - \bar{y})$ representa o desvio entre y_i e a média do conjunto \bar{y} ; S_x representa o desvio padrão do conjunto x e é definido pela Equação (6) e S_y representa o desvio padrão do conjunto y e é definido pela Equação (7).

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j \quad (4)$$

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j \quad (5)$$

$$S_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{(n-1)}} \quad (6)$$

$$S_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{(n-1)}} \quad (7)$$

Para definir o relacionamento entre os padrões, podem ser aplicadas, também, aos grupos anteriormente definidos, técnicas de Visualização de Informações como, por exemplo: Coordenadas Paralelas, Matrizes de Dispersão, Faces de Chernoff, *Star Glyphs* e Orientada a Pixel (4ª etapa da Figura 4).

Artero (2005) define Coordenadas Paralelas como um espaço de dimensão n mapeado para um espaço bidimensional, usando n eixos equidistantes e paralelos a um dos eixos principais. Cada eixo representa um atributo (neste caso, as 72 hastes dos

extensômetros) e, normalmente, o intervalo de valores de cada atributo é mapeado linearmente sobre o eixo correspondente. Cada item de um dado é exibido como uma linha poligonal que intercepta cada eixo no ponto correspondente ao valor do atributo associado.

Esta técnica permite observar o relacionamento entre as 72 hastes de extensômetros, analisando as linhas entre os eixos. Se acontecerem poucos cruzamentos, então há um bom relacionamento; muitos cruzamentos indicam um relacionamento ruim.

Por outro lado, Matrizes de Dispersão (*ScatterPlots Matrices*) buscam comparar diversos atributos simultaneamente (dois a dois) mapeando os dados para um espaço bidimensional. Informações como correlações e dispersões dos dados podem ser extraídas neste tipo de visualização, contribuindo para um maior entendimento dos dados, além de apoiar no procedimento de redução de dimensionalidade (WARD, 1994).

Outras técnicas como as Orientadas a Pixel, encontradas em Keim e Kriegel (1996); Faces de Chernoff, introduzida por Chernoff (1973); ou *Star Glyphs*, em Pickett e Grinstein (1998), geram imagens para cada uma das 72 hastes de extensômetros. O relacionamento entre estas variáveis pode ser observado visualmente, pela semelhança entre elas, ou seja, imagens parecidas indicam um bom relacionamento entre as variáveis.

4.1 Mineração Visual de Dados – MVD: integração entre Mineração de Dados – MD e Visualização de Informações – VI

Técnicas de mineração podem apoiar a exploração visual de grandes conjuntos de dados, e também serem apoiadas por recursos visuais. Percebe-se que os dois problemas críticos em MD também são críticos na VI. O primeiro é a alta dimensionalidade e o segundo é a grande quantidade de registros contidos nas bases de dados atuais, demandando o uso de algoritmos cada vez mais eficientes.

A integração de técnicas de MD com técnicas de VI facilita a análise dos dados tornando-se um valioso recurso na tomada de decisões, de forma que estas podem ser usadas para auxiliar ou serem auxiliadas pelas técnicas de MD. Neste estudo, em particular, os resultados da técnica de MD (Método de Ward) que agrupou os padrões (hastes de extensômetros) permitiu a aplicação das técnicas de VI (Coordenadas Polares; Matiz de Dispersão; Orientada a Pixel; Faces de Chernoff e *Star Glyphs*) que foram aplicadas, então, a cada um dos grupos separadamente. Vale salientar que, em geral, as técnicas de VI só podem ser aplicadas a um número pequeno de registros.

5 Obtenção dos resultados

Fazendo a implementação das técnicas descritas na seção 4 ao problema em pauta, tem-se que na interpolação por Séries Temporais, por meio do uso do *software Statgraphics Centurium XV*, o modelo foi escolhido automaticamente, segundo o critério de Akaike (AIC), observando-se, também, a raiz do erro quadrático médio (REQM ou *RMSE – Root Mean Squared Error*). Foi observado o periodograma acumulado dos resíduos e, em alguns casos, após análise dos valores-*p* nos testes *t* dos parâmetros, o modelo foi substituído por outro considerado mais adequado.

Realizadas as interpolações por Séries Temporais, foi então aplicada a Análise de Agrupamento (Método de Ward), utilizando-se o mesmo *software*. O dendrograma da Figura 5 mostra a formação dos grupos para os dados de Itaipu (72 hastes de extensômetros).

Observando-se o 1º. corte da Figura 5, resultam dois grupos. O primeiro grupo, denominado “grupo 1”, é um grupo formado por hastes de extensômetros extremamente importantes para o monitoramento da barragem; são hastes instaladas no eixo do bloco, a montante da barragem e inclinados 60º a montante. O segundo grupo, “grupo 2”, é formado pelas demais hastes.

Observando-se o 2º. corte, tem-se a formação de três grupos. O primeiro grupo coincide com o “grupo 1”, definido pelo primeiro corte. O segundo grupo, denominado “grupo 2”, possui a maioria das hastes de extensômetros instaladas nos derrames basálticos B, C e D (A e B são camadas de rochas mais profundas; C e D são camadas de rochas superficiais) e nos contatos litológicos B/C e C/D. O terceiro grupo, denominado “grupo 3”, possui a maioria das hastes de extensômetros instaladas nas juntas (entre camadas rochosas) A e B e no contato litológico A/B. Nota-se aqui que os instrumentos foram agrupados segundo as características geológicas relevantes do maciço de fundação, mesmo não tendo sido explicitamente apresentadas aos métodos de MD.

Observando-se o 3º. corte, o “grupo 1” continua o mesmo dos demais cortes; o “grupo 2” foi dividido em dois grupos denominados “grupo 2_1” e “grupo 2_2”. Destaca-se o “grupo 2_2”, que é formado, em sua maioria, por hastes de extensômetros instalados no derrame B. O “grupo 3” foi dividido em dois grupos denominados “grupo 3_1” e “grupo 3_2”. Dyminski et al. (2008) apresentam uma metodologia para identificar as hastes de extensômetros mais importantes nestes cinco grupos, utilizando a Análise Fatorial aplicada dentro de cada grupo.

Para um melhor entendimento, considerando os três grupos formados por meio, especificamente, do 2º corte, apresenta-se a Tabela 2, a seguir. Nesta Tabela 2, é apresentada cada uma das 72 hastes dos extensômetros separadas por grupo (conforme o

corte 2 da Figura 5), a sua inclinação a Montante (M) ou a Jusante (J), o afastamento em relação ao eixo da barragem em metros e a feição onde a haste está instalada.

A Figura 6, mais adiante, mostra o gráfico das leituras das 72 hastes de extensômetros durante o período analisado. As linhas foram coloridas conforme o grupo a que as hastes pertencem (preto, azul e amarelo para grupos 1, 2 e 3, respectivamente), podendo assim, identificar visualmente a distinção entre os grupos.

A Tabela 3, por sua vez, mostra os coeficientes de correlação multivariada aplicada às hastes do grupo 1, deste 2º corte. Deve-se salientar que a análise aqui mostrada, foi feita somente para o “grupo 1”, do 2º corte, por ser o grupo cujas hastes de extensômetros são de extrema importância no monitoramento da barragem. Deve-se salientar que uma análise semelhante também poderia ser feita aos demais grupos ou então para quaisquer subconjuntos de dados disponibilizados pela Itaipu.

Valores maiores que 0,5 foram coloridos em vermelho (correlação alta) e valores menores que 0,5 foram coloridos em azul (correlação baixa) na referida Tabela 3. Pode-se observar que, para a maioria das hastes de extensômetros, a correlação é alta. A haste de extensômetro equip26_1 é a única cuja correlação com as outras hastes é menor.

Já com as técnicas de VI, foi possível a comparação e o surgimento de novas interpretações para o referido corte 2, grupo 1. A Figura 7, apresentada a seguir, foi gerada pelo *software ParVis* (disponível em: <http://home.subnet.at/flo/mv/parvis/>) e utiliza a técnica Coordenadas Paralelas para visualizar os dados das 72 hastes de extensômetros do trecho F da barragem de Itaipu. Observe-se que os eixos foram propositalmente ordenados conforme ordem gerada pelo Método de Ward para este grupo (ver dendrograma da Figura 5).

Como era de se esperar, a Figura 7 induz a concluir que, em geral, estas hastes realmente estão bem relacionadas, comprovando a semelhança existente entre elas, já observada pela técnica de Ward. Nesta Figura 7, observa-se que algumas hastes são mais semelhantes entre si; destacam-se os grupos {equip1_1 e equip1_2} e {equip4_1 e equip4_2} pela existência de poucos cruzamentos entre as linhas que ligam os eixos correspondentes a estas variáveis. Da mesma forma, como já observado pela Tabela 3 (coeficientes de correlação), é possível verificar aqui, também, a semelhança entre alguns grupos de hastes, é o caso de {equip1_1 e equip1_2; coeficiente=0,97}, {equip4_1 e equip4_2; coeficiente=0,99} e {equip6_1 e equip6_2; coeficiente=0,99}.

Esta técnica (Coordenadas Paralelas) apresenta a grande vantagem da visualização imediata dos resultados, porém, como desvantagem, não permite

analisar o relacionamento entre variáveis que não estejam em eixos vizinhos. Desta forma não se pode, por exemplo, saber o relacionamento entre as variáveis {equip6_1 e equip6_2}. Porém, de forma interativa, é possível alterar a posição dos eixos permitindo analisar o relacionamento entre quaisquer pares de variáveis.

Outra forma de analisar o relacionamento entre pares de variáveis, mostrada na Figura 8, a seguir, é pela aplicação da técnica matriz de dispersão. Pode-se observar nesta matriz, gerada pelo *software*

XmdvTool (WARD et al., 2007), que as hastes são altamente correlacionáveis, ou seja, o aumento de uma implica num aumento da outra. Esta conclusão pode ser extraída da matriz analisando-se a distribuição dos pontos como, por exemplo, na linha da haste “equip1_1” e coluna da haste “equip1_2”, os dados estão distribuídos se aproximando de uma reta crescente o que induz a um bom relacionamento entre estas variáveis.

Da mesma forma, é possível analisar o relacionamento entre outras variáveis, é o caso de {equip1_1

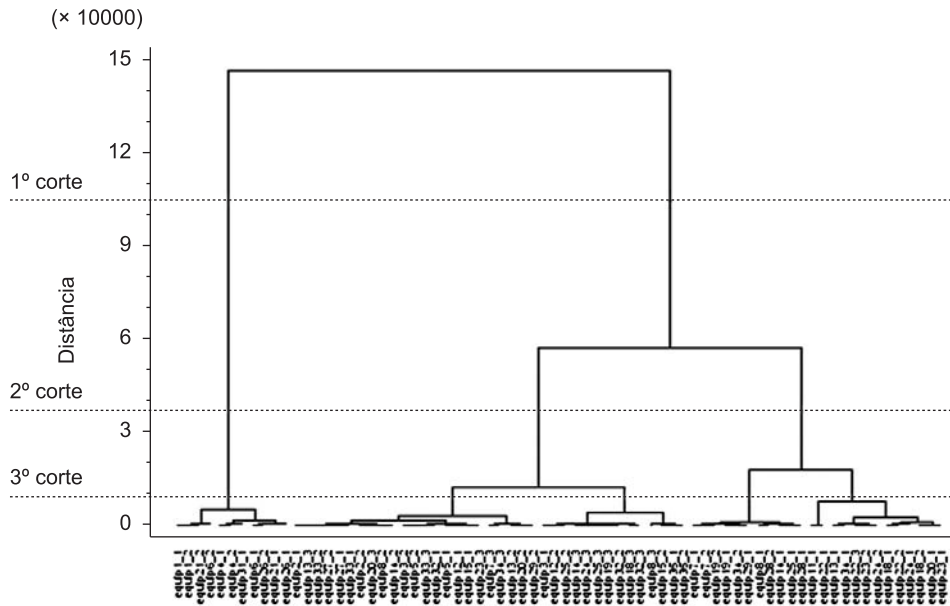


Figura 5. Dendrograma mostrando a formação dos grupos por meio do Método Ward.

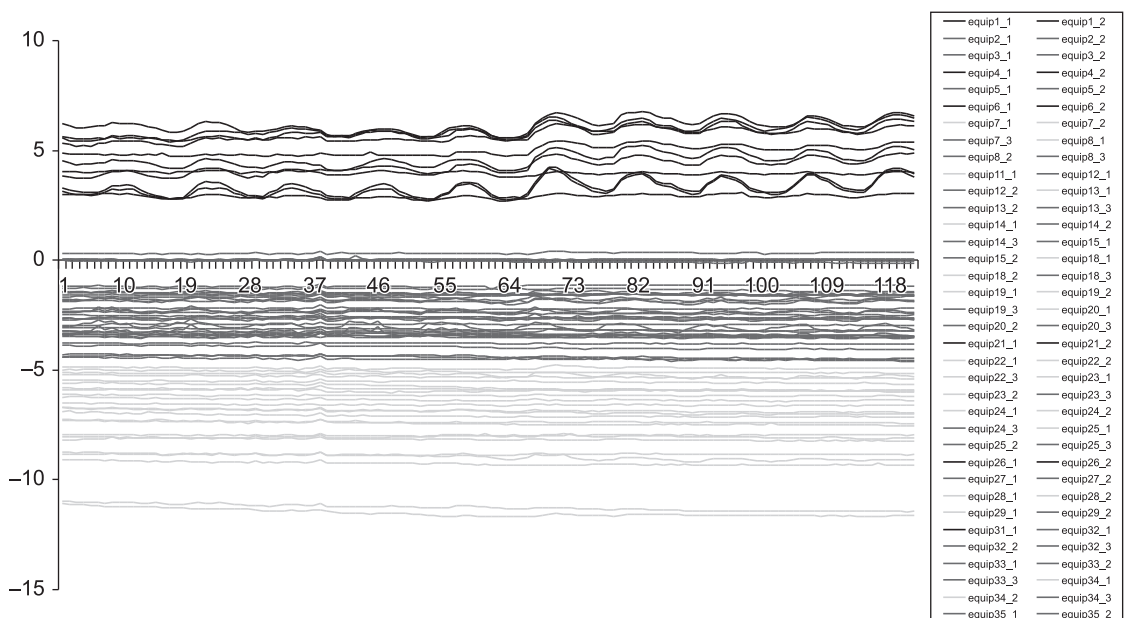


Figura 6. Gráfico das 72 hastes de extensômetros durante o período analisado.

Tabela 2. As 72 hastes de extensômetros enquadradas em cada um dos três grupos da Figura 5 (2º. corte).

Grupo	Haste	Inclinação	Afastamento do eixo da barragem	Feição
1	equip 1_1	60° a M	125,5 m a M	Junta B
1	equip 1_2	60° a M	105,4 m a M	Contato B_C
1	equip 4_1	60° a M	65,3 m a M	Contato C_D
1	equip 4_2	60° a M	60,4 m a M	Rocha Fraturada
1	equip 6_1	60° a M	150,8 m a M	Junta A
1	equip 6_2	60° a M	110,5 m a M	Derrame B
1	equip 21_1	60° a M	159,8 m a M	Junta A
1	equip 21_2	60° a M	135,1 m a M	Derrame B
1	equip 26_1	60° a M	139,2 m a M	Junta B
1	equip 26_2	60° a M	115,6 m a M	Contato B_C
1	equip 31_1	60° a M	64,7 m a M	Contato C_D
2	equip 2_1	0	32,0m a M	Contato C_D
2	equip 2_2	0	32,0 m a M	Rocha Fraturada
2	equip 3_1	0	32,0 m a M	Derrame C
2	equip 3_2	0	32,0 m a M	Derrame D
2	equip 5_1	0	13,0 m a J	Contato C_D
2	equip 5_2	0	13,0 m a J	Derrame D
2	equip 7_3	0	13,0 m a J	Derrame B
2	equip 8_2	0	84,0 m a M	Rocha Fraturada
2	equip 8_3	0	84,0 m a M	Derrame B
2	equip 12_1	60° a J	47,2 m a J	Rocha Fraturada
2	equip 12_2	60° a J	42,5 m a J	Basalto Denso
2	equip 13_2	0	44,0 m a J	Rocha Fraturada
2	equip 13_3	0	44,0 m a J	Derrame B
2	equip 14_2	0	54,0 m a J	Rocha Fraturada
2	equip 14_3	0	54,0 m a J	Derrame B
2	equip 15_1	0	80,0 m a M	Rocha Fraturada
2	equip 15_2	0	80,0 m a M	Derrame B
2	equip 18_3	0	33,0 m a J	Derrame B
2	equip 19_3	0	55,0 m a J	Derrame B
2	equip 20_2	0	82,0 m a M	Rocha Fraturada
2	equip 20_3	0	82,0 m a M	Derrame B
2	equip 23_3	0	36,0 m a J	Rocha Fraturada
2	equip 24_3	0	62,0 m a J	Derrame B
2	equip 25_2	0	75,0 m a M	Derrame B
2	equip 25_3	0	75,0 m a M	Derrame B
2	equip 27_1	30° a M	16,6 m a J	Junta B
2	equip 27_2	30° a M	22,6 m a J	Contato B_C
2	equip 29_2	30° a J	55,7 m a J	Contato B_C
2	equip 32_1	30° a M	36,5 m a M	Junta B
2	equip 32_2	30° a M	14,6 m a M	Derrame C
2	equip 32_3	30° a M	7,5 m a M	Contato C_D
2	equip 33_1	0	0,0	Junta B
2	equip 33_2	0	0,0	Derrame C
2	equip 33_3	0	0,0	Contato C_D
2	equip 34_3	30° a J	7,5 m a J	Contato C_D
2	equip 35_1	90° a M	0,0	Concreto
2	equip 35_2	90° a M	0,0	Concreto
3	equip 7_1	0	13,0 m a J	Junta A
3	equip 7_2	0	13,0 m a J	Contato A_B
3	equip 8_1	0	84,0 m a M	Contato A_B
3	equip 11_1	0	81,0 m a M	Junta A
3	equip 13_1	0	44,0 m a J	Contato A_B
3	equip 14_1	0	54,0 m a J	Contato A_B
3	equip 18_1	0	33,0 m a J	Junta A
3	equip 18_2	0	33,0 m a J	Rocha Fraturada
3	equip 19_1	0	55,0 m a J	Junta A
3	equip 19_2	0	55,0 m a J	Rocha Fraturada
3	equip 20_1	0	82,0 m a M	Rocha Fraturada
3	equip 22_1	0	68,0 m a M	Junta A
3	equip 22_2	0	68,0 m a M	Rocha Fraturada
3	equip 22_3	0	68,0 m a M	Derrame B
3	equip 23_1	0	36,0 m a J	Junta A
3	equip 23_2	0	36,0 m a J	Rocha Fraturada
3	equip 24_1	0	62,0 m a J	Junta A
3	equip 24_2	0	62,0 m a J	Rocha Fraturada
3	equip 25_1	0	75,0 m a M	Junta A
3	equip 28_1	0	40,0 m a J	Junta B
3	equip 28_2	0	40,0 m a J	Contato B_C
3	equip 29_1	30° a J	63,5 m a J	Junta B
3	equip 34_1	30° a J	36,6 m a J	Junta B
3	equip 34_2	30° a J	21,0 m a J	Derrame C

Tabela 3. Coeficientes de Correlação Multivariada aplicada às hastes do grupo 1, 2º corte.

	equip1_1	equip1_2	equip4_1	equip4_2	equip6_1	equip6_2	equip21_1	equip21_2	equip26_1	equip26_2	equip31_1
equip1_1	1	0,973789	0,8894156	0,910016	0,8256753	0,767965	0,8401519	0,8630023	0,5143424	0,9108726	0,57704335
equip1_2	0,97	1	0,8195501	0,847656	0,7524794	0,676221	0,8361274	0,8297473	0,4709566	0,938734	0,46222959
equip4_1	0,89	0,82	1	0,990258	0,8845925	0,882776	0,7863462	0,8403169	0,593878	0,7894659	0,79846531
equip4_2	0,91	0,85	0,99	1	0,8930264	0,882633	0,8221118	0,8644494	0,5160937	0,8282038	0,77197724
equip6_1	0,83	0,75	0,88	0,89	1	0,989003	0,9105697	0,956394	0,3662914	0,7319502	0,82526618
equip6_2	0,77	0,68	0,88	0,88	0,99	1	0,8576867	0,9165347	0,3804936	0,6555618	0,87195133
equip21_1	0,84	0,84	0,79	0,82	0,91	0,86	1	0,9809751	0,1967116	0,8470128	0,60943684
equip21_2	0,86	0,83	0,84	0,86	0,96	0,92	0,98	1	0,2823973	0,8159251	0,69100332
equip26_1	0,51	0,47	0,59	0,52	0,37	0,38	0,2	0,28	1	0,3582111	0,44396828
equip26_2	0,91	0,94	0,79	0,83	0,73	0,66	0,85	0,82	0,36	1	0,48
equip31_1	0,58	0,46	0,8	0,77	0,83	0,87	0,61	0,69	0,44	0,48	1

e equip1_2}, {equip4_1 e equip4_2}, {equip6_1 e equip6_2} e {equip21_1 e equip21_2}.

A Figura 9, a seguir, gerada pelo software MatLab, mostra o uso de três técnicas: a) técnica Orientada a

Pixel, em que os pixels são mapeados conforme valores da matriz de dados; b) técnica Faces de Chernoff, que mapeia as informações seguindo características baseadas em ícones, características faciais (tamanho

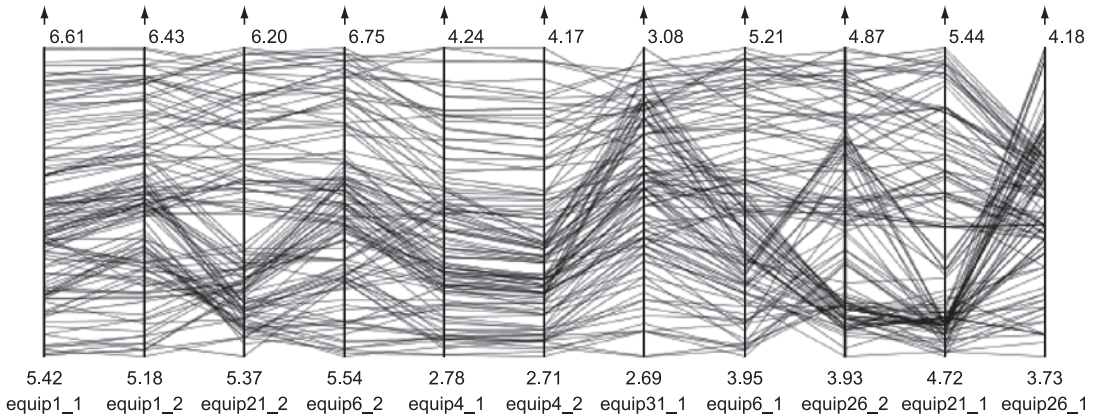


Figura 7. Uso da técnica Coordenadas Paralelas para a visualização dos dados das hastes de extensômetros (grupo 1, 2º. corte).



Figura 8. Representação da técnica matriz de dispersão para as hastes de extensômetros (grupo 1, 2º. corte).

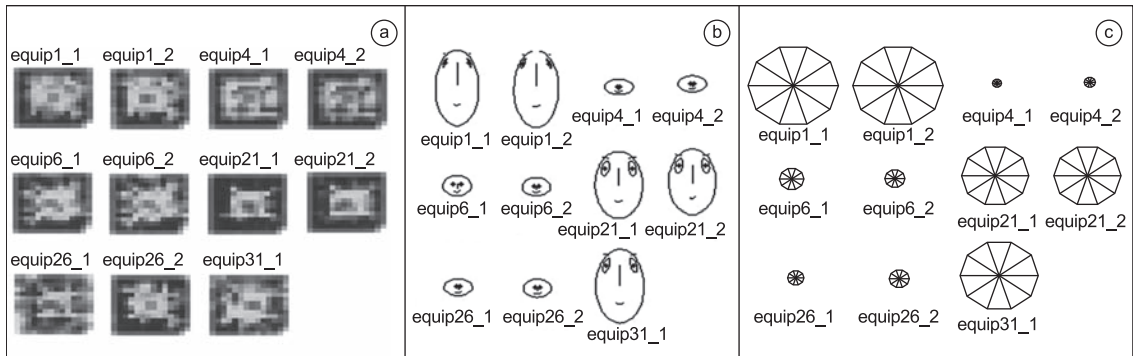


Figura 9. Uso das técnicas a) Orientadas a Pixel, b) Faces de Chernoff e c) Star Glyphs na análise de relacionamentos entre hastes de extensômetros (grupo 1, 2º. corte).

do rosto, tamanho e forma das sobranceiras, etc.); e c) técnica *Star Glyphs*, que representa os atributos pelas pontas das estrelas com tamanho proporcional ao valor representado.

Observando-se a Figura 9a, pode ser visto que algumas hastes são mais semelhantes entre si. É o caso de {equip1_1, equip1_2, equip21_1, equip21_2 e equip26_2} e {equip4_1, equip4_2, equip6_1, equip6_2, equip26_1 e equip31_1}. Já nas Figuras 9b e 9c, esta semelhança ocorre em {equip1_1, equip1_2, equip21_1, equip21_2 e equip31_1} e {equip4_1, equip4_2, equip6_1, equip6_2, equip26_1 e equip26_2}.

Considerando-se todas as técnicas abordadas, pode-se dizer que, de uma forma geral, no grupo 1 (2º corte) analisado, todas as hastes possuem bons relacionamentos, porém, como visto pelas diversas técnicas, algumas hastes são mais semelhantes entre si.

Algumas técnicas são mais refinadas do que outras (por exemplo, o dendrograma apresentado quando aplicado à Análise de Agrupamento pelo Método Ward, mostra a similaridade entre as hastes em todos os níveis), enquanto que outras são mais facilmente interpretadas (por exemplo, as Faces de Chernoff).

Os resultados atingiram as expectativas esperadas e as técnicas de VI integradas às de MD mostraram ser eficientes do ponto de vista computacional e facilidade de uso. No entanto, devido à grande quantidade de técnicas, a escolha daquelas que melhor se ajustam aos dados em análise não é uma tarefa fácil e devem ser escolhidas de acordo com o que se está buscando.

6 Conclusões

Neste artigo é apresentada a abordagem de um importante problema de engenharia, análise de dados de instrumentação de grandes obras, usando técnicas de MVD. Como aplicação destas técnicas, procurou-se conhecer o relacionamento entre os dados de instrumentação da barragem de Itaipu, o qual foi tratado por meio do processo *KDD*. Para tanto, foram utilizadas técnicas para o pré-processamento e agrupamento de dados e de relacionamento entre os dados dentro de um mesmo grupo.

Para o agrupamento de dados, que é uma das possíveis tarefas desempenhadas pela MD, foi utilizado o Método de Ward. Já para verificar o relacionamento entre os dados, em cada um dos grupos obtidos por este método, foram utilizadas algumas técnicas de VI: coordenadas paralelas; matriz de dispersão; orientada a pixel; faces de Chernoff e *Star Glyphs*.

Conhecer o relacionamento entre os instrumentos de uma barragem é de suma importância, pois, se sabendo destas relações, ao ocorrer uma anomalia num determinado instrumento em leituras futuras, os outros instrumentos relacionados deverão ser analisados para confirmar a anomalia, fazendo com que as medidas cabíveis sejam tomadas pela empresa.

Para o desenvolvimento do trabalho, foram obtidos os dados relativos a 72 hastes de extensômetros localizados no bloco F da barragem no decorrer de 120 meses. Na análise de agrupamento, foi mostrada a formação de possíveis 2, 3 ou 5 grupos (Figura 5). Pode-se observar justificativas técnicas para tais formações. No gráfico das leituras das hastes de extensômetros (Figura 6), por exemplo, foi possível identificar visualmente a formação de 3 grupos.

A sequência da análise deu-se sobre a formação dos 3 grupos (2º. corte, Figura 5), do qual selecionou-se o grupo 1, por conter instrumentos bastante importantes para o monitoramento da barragem. Para este grupo 1 de instrumentos, mostrou-se que, para a maioria das hastes de extensômetros, a correlação é alta (Tabela 3), justificando mais uma vez a formação deste grupo.

Com as técnicas de VI aplicadas ao grupo 1 (2º corte), obtido pelo Método de Ward, foi possível a comparação e o surgimento de novas interpretações, sendo que a grande vantagem destas técnicas é a visualização imediata dos resultados. Com a técnica de Coordenadas Paralelas (Figura 7), por exemplo, foi possível concluir que, em geral, as hastes do grupo 1 estão realmente bem relacionadas. Estas técnicas permitem, sem a necessidade de conhecimentos em áreas como estatística ou matemática, analisar visualmente a imagem formada pelas variáveis e nela, intuitivamente, formar os grupos.

Este tipo de abordagem mostrou-se bastante útil no agrupamento de instrumentos segundo sua relevância em relação ao comportamento da barragem, bem como os agrupou segundo critérios não explicitados no banco de dados, como a localização em diferentes feições geológicas. Anomalias em leituras de instrumentos também podem ser facilmente detectadas, pois as correlações com os demais instrumentos deverão mudar abruptamente caso aconteçam. Desta forma, a introdução de técnicas de MVD em sistemas de tomada de decisão é bastante recomendável.

Agradecimentos

Os autores agradecem à equipe de Engenharia Civil da Itaipu Binacional o apoio técnico, à FINEP o apoio financeiro ao projeto CT-HIDRO – AIEVC – “Análise de Incertezas e Estimação de Valores de Controle para o Sistema de Monitoração Geotécnico-Estrutural na Barragem de Itaipu” e à CAPES a bolsa concedida à segunda autora.

Referências

ANKERST, M.; KEIM, D. Visual data mining and exploration of large databases. In: EUROPEAN CONFERENCE ON PRINCIPLES AND PRACTICE OF KNOWLEDGE DISCOVERY IN DATABASES, 5., 2001, Freiburg, **Proceedings...**

- ARTERO, A. O. **Estratégias para apoiar a detecção de estruturas em visualizações multidimensionais percentualmente sobrecarregadas**. 2005. Tese (Doutorado) – Universidade de São Paulo, São Carlos, 2005.
- BOX, G. E. P.; JENKINS, G. M. **Time series analysis, forecasting and control**. San Francisco: Holden Day, 1976.
- BRASIL. **Projeto de Lei nº 1.181**, de 2003. Estabelece diretrizes para verificação da segurança de barragens de cursos de água para quaisquer fins e para aterros de contenção de resíduos líquidos industriais. Disponível em: <http://www.emtermos.com.br/ABMS/PL_1181.pdf>. Acesso em: 19 jun. 2009.
- CHERNOFF, H. The use of faces to represent points in K-dimensional space graphically. **Journal of American Statistical Association**, v. 68, p. 361-368, 1973.
- COMITÊ BRASILEIRO DE GRANDES BARRAGENS – CBGB. **Diretrizes para a inspeção e avaliação de segurança de barragens em operação**. Rio de Janeiro, 1983.
- DINIZ, C. A. R.; LOUZADA NETO, F. **Data mining: uma introdução**. São Paulo: ABE, 2000.
- DYMINSKI, A. S.; STEINER, M. T. A.; VILLWOCK, R. Hierarchical ordering of extensometers readings from Itaipu dam. In: INTERNATIONAL SYMPOSIUM ON LIFE-CYCLE CIVIL ENGINEERING, Varenna, 2008. **Proceedings...**
- FAYYAD, U.M. et al. **Advances in knowledge discovery and data mining**. Massachusetts: AAAI Press, 1996.
- FREITAS, A. A. **Data mining and knowledge discovery with evolutionary algorithms**. New York: Springer, 2002.
- GIMENES, E. **Data Mining – Data Warehouse: a importância da mineração de dados em tomadas de decisões**. 2000. Monografia de conclusão de curso (Tecnólogo em Processamento de Dados)–Centro Estadual de Educação Tecnológica Paula Souza, Faculdade de Tecnologia de Taquaritinga, Taquaritinga, 2000.
- HAIR JUNIOR, J. F. et al. **Análise multivariada de dados**. São Paulo: Bookman, 2005.
- INTERNATIONAL COMMISSION ON LARGE DAMS – ICOLD. Título do texto/diretriz que foi consultada. Disponível em: <<http://www.icold-cigb.org>>. Acesso em: dia mês 2008.
- ITAIPU. **ITAIPU Binacional**. Disponível em: <<http://www.itaipu.gov.br>>. Acesso em: dia mês 2008.
- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM Computing Surveys**, v. 31, n. 3, 1999.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4. ed. Upper Saddle River: Prentice Hall, 1998.
- KACHIGAN, S. K. **Statistical analysis: an interdisciplinary introduction to univariate and multivariate methods**. New York: Radius Press, 1986.
- KEIM, D. A. Visual exploration of large data sets. **Communications of the ACM**, v. 44, n. 8, p. 38-44, 1979.
- KEIM, D. A.; KRIEGEL, H. P. Visualization techniques for mining large databases: a comparison. **IEEE Trans. Knowledge & Data Engineering**, v. 8, n. 6, p. 923-936, 1996.
- MATOS, S. F. **Avaliação de instrumentos para auscultação de barragem de concreto**. Estudo de Caso: deformímetros e tensômetros para concreto na barragem de Itaipu. 2002. Dissertação (Mestrado em Engenharia Civil)–Universidade Federal do Paraná, Curitiba, 2002.
- PICKETT, R. M.; GRINSTEIN, G. G. Iconographic displays for visualizing multidimensional data. In: IEEE CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS, 1988, Piscataway. **Proceedings...**
- REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. Barueri: Manole, 2003.
- SILVA NETO, M. A. **Mineração visual de dados: extração do conhecimento a partir de técnicas de visualização da informação e mineração de dados**. Experimentos: ITAIPU e SIMEPAR. 2008. Dissertação (Mestrado em Métodos Numéricos em Engenharia)–Universidade Federal do Paraná, Curitiba, 2008.
- SILVEIRA, J. F. A. **Instrumentação e comportamento de fundações de barragens de concreto**. São Paulo: Oficina de Textos, 2003.
- SILVER, D. L. Knowledge discovery and data mining. **Technical Report MBA6522**. CogNova Technologies London Health Science Center, 1996.
- TAN, P. N.; STEINBACH, M.; KUMAR, V. **Introduction to data mining**. Boston: Addison-Wesley Longman, 2005.
- WARD, M. O. et al. **XmdvTool release: the multivariate data visualization tool**. Disponível em: <<http://davis.wpi.edu/~xmdv/>>. Acesso em: 26 jan. 2007.
- WITTEN, I. H.; FRANK, E. **Data mining: practical machine learning tools and techniques with Java implementations**. San Francisco: Morgan Kaufmann, 2000.
- WARD, M. O. XmdvTool: integrating multiple methods for visualizing multivariate data. In: IEEE VISUALIZATION, 1994, Washington, DC. **Proceedings...**
- WONG, P. C. Visual data mining. **IEEE Computer Graphics and Applications**, v. 19, n. 5, p. 20-21, 1999.