

Modalidade do trabalho: Relatório técnico-científico
Evento: XXIII Seminário de Iniciação Científica

DEFINIÇÃO E APLICAÇÃO DE METODOLOGIAS DE MINERAÇÃO DE DADOS PARA UM SISTEMA DE MONITORAMENTO DE SUBESTAÇÕES DE ENERGIA ELÉTRICA¹

Gabriela Linck De Jesus², Paulo Sausen³.

¹ Projeto de iniciação científica desenvolvida pelo Grupo de Automação Industrial e Controle, para o projeto: “Desenvolvimento e Implantação de um Lote Pioneiro de um Sistema de Monitoramento e Supervisão de Unidades Transformadoras e Subestações”

² Aluna do Curso de Graduação em Ciência da Computação da UNIJUI, bolsista PROBIC/FAPERGS, gabriela.linck@outlook.com

³ Professor Doutor do Departamento de Ciências Exatas e Engenharias da UNIJUI, orientador, sausen@unijui.edu.br;

INTRODUÇÃO

A evolução da tecnologia da informação possibilitou que grandes volumes de dados possam ser armazenados e processados mais rapidamente com uma maior precisão. Estima-se que as estruturas de distribuição de energia elétrica deverão ser providas de inteligência nos próximos anos. No projeto “Desenvolvimento e Implantação de um Lote Pioneiro de um Sistema de Monitoramento e Supervisão de Unidades Transformadoras e Subestações”, desenvolvido pelo Grupo de Automação Industrial e Controle (GAIC) da Unijuí em parceria com a Companhia Estadual de Energia Elétrica (CEEE) foi criada uma infraestrutura que provê um sistema de monitoramento, em tempo real, de todas as grandezas físicas de um conjunto de 160 subestação transformadora de energia elétrica com intervalo de amostragem de 10s. Este sistema é um exemplo de como é usual a criação de bases de dados que, em decorrência da quantidade de grandezas e periodicidade de aquisição, podem facilmente atingir a cifra dos terabytes de informações.

Esta volumosa base de dados ao mesmo tempo que é necessária para o correto monitoramento da subestação passa a ser de difícil interpretação em sua forma bruta, muitas vezes sendo impraticável a extração de conhecimento. Uma das formas de contornar este problema é utilizar técnicas de mineração de dados, que é um processo que visa retirar conhecimento das informações, e traduzir ao usuário final, estas informações, de forma simples e clara. O conceito de mineração de dados é um processo relativamente recente, mas que tem sido amplamente utilizada como forma de melhorar a visualização de grandes bases de dados.

Neste contexto, o principal objetivo deste artigo é analisar as principais metodologias de mineração de dados, e definir parâmetros para derivar grandes quantidades de dados vindos das subestações em alertas simples e claros, que deem ao usuário, em tempo real, uma visão da situação das subestações.

METODOLOGIA

Vários autores afirmam que a quantidade de dados existente hoje é exorbitante, mas que o conhecimento e as informações realmente importantes estão escassas. Atualmente o problema deixou de ser a armazenagem destes dados e passou a ser a forma de como extrair conhecimento

Modalidade do trabalho: Relatório técnico-científico
Evento: XXIII Seminário de Iniciação Científica

destas bases. A mineração de dados surgiu na década de 80, proposta por Jiawei Han e Micheline Kamber para ajudar nestas questões, sendo possível através dela identificar padrões significativos, e informações que provavelmente não seriam vistas de forma analítica.

A mineração de dados dispõe de Processos, Métodos, Modelos, Técnicas e Algoritmos que definem passos a serem seguidos para traduzir grandes quantidades de dados em conhecimento, ou seja, em informações que sejam legíveis ao usuário. Segundo Fayyad (1996) a Descoberta de Conhecimento nas Bases de Dados, ou Knowledge Discovery From Data (KDD), e é o processo base, que visa a extração de padrões significativos e análise destes padrões para a descoberta do conhecimento. É um processo com cinco fases que podem ser repetidas, conforme a necessidade. As fases deste processo são compostas de: limpeza e redução de dados, escolha de métodos e tarefas, escolha de algoritmos de acordo com o método escolhido e por final a análise e documentação dos resultados, como pode ser visto na Figura 1.

Todas as fases são importantes para a obtenção de resultados mais precisos, porém inicialmente será trabalhado apenas com a limpeza dos dados, escolha de métodos e tarefas e a análise e documentação de resultados. As fases de escolha de algoritmos e de redução de dados serão aplicadas após os testes iniciais, que visam a definição de procedimentos corretos com o problema proposto.

Inicialmente é preciso entender os dados, o que eles representam e aonde deseja-se chegar no final do processo, identificando pontos principais da abordagem e as prioridades. Entende-se que a base de dados a ser trabalhada é volumosa, pois contém todas as informações relativas ao controle das subestações, trabalhar com todos esses dados de uma vez seria dispendioso computacionalmente. Por esta razão escolhe-se um conjunto de dados que será trabalhado primeiro, neste caso, o conjunto de dados a ser visto neste artigo é o conjunto referente as temperaturas das subestações, a temperatura ambiente e a temperatura do transformador.

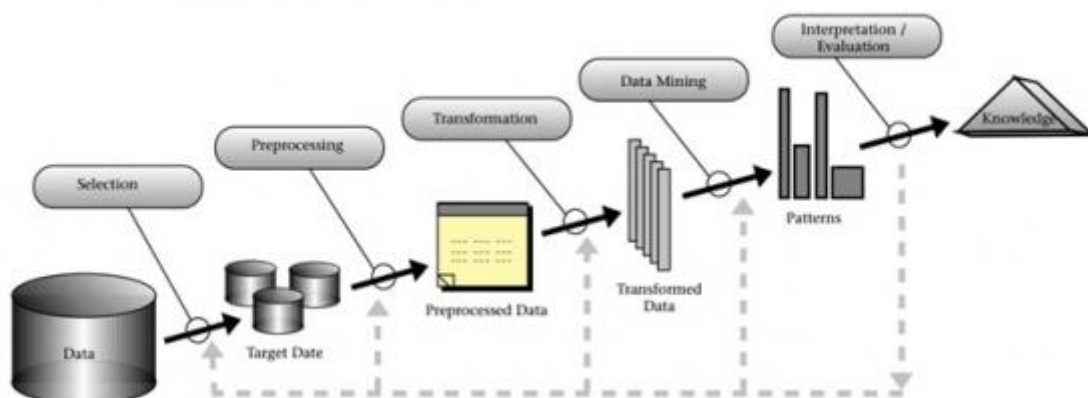


Figura 1. Uma visão geral das etapas que compõem o processo de KDD (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996, p.41)

Modalidade do trabalho: Relatório técnico-científico
Evento: XXIII Seminário de Iniciação Científica

Em sistemas de monitoramento como este é evidente que haverá dados que não correspondem à realidade, seja por falha na transmissão dos dados, seja por problemas na subestação ou na própria máquina que recebe as informações. Estes dados “errados” são chamados de ruídos e são dados que não nos levarão ao conhecimento, então o conjunto de dados escolhido passa por uma “limpeza” aonde os ruídos são tratados ou excluídos.

Após o tratamento de ruídos a base de dados passa a ser mais maleável ao uso de métodos de extração do conhecimento. Os métodos ou tarefas de mineração de dados são a forma a qual os dados serão tratados, e os mais conhecidos segundo Han e Kamber (2006) são os métodos de Classificação, Regressão, Clusterização, Predição e Associação. Existe ainda a possibilidade de combinar estes métodos para obter um melhor resultado.

Os métodos que mais se adequam a situação da base de dados de monitoramento das subestações, que é o objeto de estudo deste artigo, são os métodos de Classificação, Clusterização e Predição.

O método de classificação se identifica com o problema das subestações pois oferece a separação de registros de acordo com alguma classe estipulada pelo desenvolvedor, e desta forma tem-se uma melhor visibilidade do que está acontecendo no momento da coleta dos dados.

O método de clusterização oferece ao problema proposto o destaque de registros que não estão em nenhum padrão, podendo este registro ser uma situação de risco, ou que inspire cuidados. Esta solução é viável, porém dispendiosa pois há uma grande quantidade de registros na base de dados.

O método de predição pelo seu próprio nome já soa como uma solução, pois a ideia é prever como os registros irão se portar, desta forma prevenindo danos graves. Porém este método é tão ou mais dispendioso que o método de clusterização, pois analisa registro por registro procurando relações entre eles. Em uma base de dados que irá contar com 160 subestações, com aquisição de dados a cada 10 segundos, sendo 8640 registros inseridos por subestação por dia, realizar uma comparação registro a registro torna-se inviável.

A escolha do método de mineração de dados influencia de muitas formas na escolha do algoritmo de mineração de dados. Mas neste caso o algoritmo utilizado é um algoritmo básico de classificação. A quantidade de dados escolhida é filtrada e os dados desejados recebem marcadores que identificam os dados analisados, que então são inseridos em uma tabela separada dos demais. Com os dados classificados é possível aplicar futuramente estratégias como o cubo de dados, e algoritmos de clusterização, que separam os dados em clusters, e os novos dados são inseridos nos clusters assemelhados.

Após uma análise inicial foi verificado que o método mais coerente com a proposta inicial, e a ser utilizado nos primeiros testes, é o método de classificação. Porém, juntamente com a classificação dos registros, serão aplicados futuramente conceitos de clusterização, pois com os dados classificados, torna-se mais fácil a visualização em cluster de registros “anormais”. Neste caso, haverá a combinação de dois métodos para obter-se melhores resultados.

Cabe ressaltar que na análise dos dados é possível retornar a fases anteriores caso o método escolhido não seja compatível, mesmo para fins de comparação de resultados. Neste caso, passos anteriores serão revistos futuramente com a finalidade de inserir novos passos ao processo e aplicar algoritmos de clusterização. Com os dados classificados e interpretados resta documentá-los e/ou

Modalidade do trabalho: Relatório técnico-científico
Evento: XXIII Seminário de Iniciação Científica

usá-los para outros fins, em programas ou em sistemas mais complexos. Neste caso, para gerar alertas de risco nas subestações, para corrigir erros antes que se tornem graves demais, ou até mesmo prevenir que danos aconteçam.

RESULTADOS

Os passos definidos pelo processo de Descoberta de Conhecimento nas Bases de Dados contam com a limpeza de dados, escolha de métodos, escrita de algoritmos e com a análise de dados para enfim se ter dados legíveis e verídicos. Seguindo este processo, o primeiro passo é analisar a base de dados e retirar um conjunto de dados a ser tratado primeiramente. A base de dados em questão conta com dados que fornecem um monitoramento em tempo real de subestações de energia elétrica. Sabe-se também que a quantidade de dados é extremamente volumosa, e cresce diariamente.

Dentre as 24 grandezas oriundas das subestações de energia estão os dados referentes as temperaturas da subestação: a temperatura ambiente que se refere a temperatura interna da subestação, e a temperatura do transformador que se refere a temperatura interna do transformador de energia. Para o sistema em questão, estas temperaturas nunca atingirão valores inferiores a zero graus uma vez que a subestação é uma subestação subterrânea completamente isolada e com equipamentos que dissipam calor. Neste caso quando for encontrada/registrada uma temperatura inferior a zero graus este registro deve ser descartado.

A temperatura do transformador tende a ser maior que a temperatura ambiente, pois o transformador quando em funcionamento dissipa calor. Estes dados serão inicialmente classificados em três categorias: Os dados considerados Normais, ou seja, funcionamento normal de uma subestação, dados estes que possuem temperatura ambiente entre 1 e 39 graus e temperatura do transformador entre 1 e 49 graus. Os dados considerados Graves, ou seja, a subestação está aquecendo e expira cuidados, dados estes que possuem temperatura ambiente entre 40 e 44 graus e temperatura do transformador entre 50 e 59 graus. E os dados considerados Gravíssimos, que refletem uma subestação superaquecida e que deve ser monitorada imediatamente, dados estes que possuem temperatura ambiente maiores de 45 graus e temperatura do transformador maiores que 60 graus.

Para categorizar os dados e ter uma dimensão melhor do comportamento das subestações foram coletados dados da subestação Mercado, subestação localizada perto do Mercado Público de Porto Alegre, do dia 30/03/2014 ao dia 07/04/2014. Foi coletado um total de 18.145 dados, dados estes que passaram pelo filtro e não possuem ruídos. Após categorizar estes dados obteve-se um resultado de 9234 dados classificados como “Normais”, pois possuíam temperatura ambiente abaixo de 40 graus e temperatura do transformador abaixo de 50 graus, 50,90% dos dados. Dados considerados “Graves” são 28,50% dos dados, ou seja, 5172 dados estavam tinham uma temperatura ambiente entre 40 e 45 graus e temperatura do transformador entre 50 e 60 graus. Deste total de 18.145 dados, aqueles classificados como “Gravíssimos” são 20,60% dos dados, ou seja, 3739 dados possuíam temperatura ambiente acima de 45 graus, e temperatura do transformador acima de 60 graus. Na Figura 2 é possível ver o gráfico relativo as temperaturas e o dia em que esta temperatura foi registrada. No dia 02 de Abril de 2014 a temperatura do transformador atingiu uma temperatura

Modalidade do trabalho: Relatório técnico-científico
Evento: XXIII Seminário de Iniciação Científica

superior a 60 graus, e a temperatura ambiente passou os 45 graus, o que para ambas as temperaturas é considerado gravíssimo.

CONCLUSÃO

O objetivo deste processo é reduzir ao máximo a complexidade na visualização dos dados de monitoramento das subestações, com a finalidade de transparecer ao usuário final apenas informações relevantes de formas claras e simplificadas. Este processo chamado de Descoberta do Conhecimento é um processo complexo, que exige estudo profundo de métodos e algoritmos adequados a cada situação. Neste caso, há uma base de dados extremamente volumosa, que tende a aumentar a cada dia, e estes dados na forma como são armazenados não são inteligíveis ao usuário final, o que pode prejudicar o monitoramento.

Os testes realizados mostraram que em dias de pico de calor a subestação tende a alcançar níveis perigosos de temperatura, e se estes dados forem visíveis apenas ao desenvolvedor da base de dados, esta subestação pode, em decorrência de altas temperaturas, sofrer danos que poderiam ter sido evitados. Este primeiro processo categorizou uma faixa de dados de temperaturas, entre temperaturas Normais, Graves e Gravíssimas, alertas que serão dados em forma de gráfico com o intuito de realizar manutenções preventivas nas subestações antes que danos sejam causados.

Este resultados são resultados parciais onde foram utilizados apenas as temperaturas, pretende-se estender os testes a outras grandezas, como Tensão e Corrente, visando monitorar a subestação de diferentes formas sempre objetivando sua manutenção preventiva.

Referências Bibliográficas

- SOUZA, F. B.; FERREIRA, R. S. A.; SOUZA, V. S. M. Desenvolvimento Socioeconômico e Energia Elétrica - Uma Análise na Comunidade Rural do Município Careiro da Várzea no Estado do Amazonas. ISSN 1984-9354, Agosto, 2011.
- HAN, J.; KAMBER, M. Data Mining, Concepts and Techniques. Segunda Edição. São Francisco: Morgan Kaufmann, 2006.
- WITTEN, I. H.; FRANK, E. Data Mining, Practical Machine Learning Tools and Techniques. Segunda Edição. São Francisco: Morgan Kaufmann, 2005.
- FAYYAD, U; PIATETSKY-SHAPIRO, G; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence, Julho, 1996.