



CRICTE 2017

XXVIII Congresso Regional de Iniciação Científica e Tecnológica em Engenharia



BUSCA DE PADRÕES UTILIZANDO ALGORITMO BAYESIANO PARA MINERAÇÃO DE DADOS NAS SUBESTAÇÕES SUBTERRÂNEAS DA CEEE

Luciano Antônio Destefani

Mestrando do Programa de Pós-Graduação Stricto Sensu em Modelagem Matemática.
Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUÍ).
lucianodestefani@gmail.com

Leonardo Bressan Motyczka

Mestrando do Programa de Pós-Graduação Stricto Sensu em Modelagem Matemática.
Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUÍ).
leomoty@gmail.com

Paulo Sérgio Sausen

Professor Doutor, do Departamento de Ciências Exatas em Engenharias.
Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUÍ).
paulosausen@gmail.com

Airam Sausen

Professora Doutora, do Departamento de Ciências Exatas em Engenharias.
Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUÍ).
airamsausen@gmail.com

Resumo. Destaca-se que a partir da instrumentação das redes de distribuição das concessionárias de energia tem sido disponibilizado uma base de dados com um volume considerável, tanto em tamanho, quanto em informações, que podem propiciar além da caracterização dos perfis de consumo, uma nova gama de aplicações relacionadas a descoberta de conhecimento que possam auxiliar na tomada de decisões administrativas vinculadas, por exemplo, a manutenção preventiva de sua estrutura de distribuição. A partir da parceria existente entre o Grupo de Automação Industrial e Controle (GAIC), da Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUÍ), com a Companhia Estadual de Distribuição de Energia Elétrica (CEEE), concessionária de serviços de distribuição de energia elétrica na região sul-sudeste do Estado do Rio Grande do Sul, localizada na cidade de Porto Alegre, possibilitou o acesso

aos dados compostos das aquisições de 30 subestações de energia do mês de janeiro de 2017, sendo analisada somente a primeira semana para rodar o pré-processamento e o processo de Mineração de Dados. Neste contexto este trabalho tem por objetivo a busca de padrões através dos conceitos de Smart Grids, Descoberta de Conhecimento e da Mineração de Dados (MD) utilizando o algoritmo Bayesiano e a ferramenta WEKA nas subestações subterrâneas da CEEE.

Palavras-chave: Mineração de Dados, Algoritmo Bayesiano, Energia Elétrica.

1. INTRODUÇÃO

Nos últimos anos o avanço tecnológico vem proporcionando novas ferramentas para otimizar a utilização da crescente quantidade de dados gerada todos os dias nas mais variadas áreas de atuação, e melhor

aproveitá-las através da Mineração de Dados (MD) eficiente em benefício da sociedade, conforme Zhang *et. Al* [1].

Para trabalhar com esta grande quantidade de informações, Rocha *et. Al* [2] explica que é utilizada a inteligência computacional, sendo necessário o uso de um modelo de MD baseado no processo de descoberta de conhecimento nestas bases de dados, também conhecido como KDD (*Knowledge Discovery in Databases*).

Neste contexto o principal objetivo deste trabalho é utilizar os conceitos das *Smart Grids*, descoberta de conhecimento e mineração de dados para busca de padrões utilizando as regras de classificação e o algoritmo Bayesiano juntamente da ferramenta WEKA visando a melhoria permanente dos serviços disponibilizados pelas concessionárias de energia elétrica.

2. REVISÃO BIBLIOGRAFICA

Com o intuito de facilitar a leitura e compreensão deste trabalho, nesta seção são apresentadas informações elementares sobre a Descoberta de Conhecimento e MD, suas características relevantes e principais propriedades. É realizada uma revisão bibliográfica do estado da arte sobre os métodos/técnicas e ferramentas de MD encontrados na literatura, com ênfase no algoritmo Bayesiano e na ferramenta WEKA (*Waikato Environment for Knowledge Analysis*) utilizada na MD para busca dos padrões em estudo.

2.1 Mineração de Dados

O conceito de Mineração de Dados (*Data Mining*) abordado por Cortes [3] destaca que ela está se tornando cada vez mais popular como uma ferramenta de descoberta de informações com objetivo de revelar estruturas de conhecimento visando a tomada de decisões em condições de certeza limitada.

Um conceito muito difundido e equivocado sobre mineração de dados, na opinião de Kamber [4] define os sistemas de MD como fontes para extração automática de

todos os conceitos mais valiosos que estão escondidos em um grande banco de dados sem intervenção ou direcionamento humano.

Para confrontar esta ideia, Schuch [5] destaca que os principais conceitos aceitos por muitos pesquisadores se referem a descoberta de conhecimento em bases de dados como um processo global de descoberta que envolve seleção, pré-processamento, mineração dos dados, interpretação dos resultados e a transformação do conhecimento. Sendo assim, a MD é uma das etapas deste processo onde são aplicados algoritmos específicos para extração de padrões a partir dos dados ou até mesmo revelar comportamento de um banco de dados.

2.2 Métodos ou Técnicas

Ainda segundo Ref. [5] a escolha das técnicas de MD dependerá da tarefa específica a ser executada e dos dados disponíveis para análise, devendo-se levar em conta a natureza dos dados disponíveis em termos de conteúdo, os tipos de campos de dados e a estrutura das relações entre os registros.

Durante o processo de mineração, Camilo e Silva [6] e Mccue [7], explicam que diversas técnicas devem ser testadas e combinadas a fim de que comparações possam ser feitas e então a melhor técnica (ou combinação de técnicas) seja utilizada. A Figura 1 traz a combinação destas técnicas para solução de um problema de KDD.



Fig. 1: Combinação de Técnicas de MD.

2.3 Ferramentas para MD

Dentre as ferramentas pesquisadas para a descoberta de conhecimento e MD, existem opções gratuitas e outras pagas. Os autores Witten *et. Al* [8] e Ref. [6], destacam que dentre as ferramentas gratuitas o WEKA é

amplamente referenciado como um software reconhecido pela comunidade científica em aprendizado de máquina e MD.

O WEKA (*Waikato Environment for Knowledge Analysis*) desenvolvido na Universidade de Waikato, Nova Zelândia, trata-se de um pacote implementado na linguagem Java, seguindo o paradigma de orientação a objetos, composto de uma série de algoritmos para solucionar problemas de MD.

Os algoritmos podem ser aplicados diretamente a uma série de dados pela ferramenta ou utilizados por programas Java. Por ser uma ferramenta de domínio público (software livre) encontra-se disponível em <http://www.cs.waikato.ac.nz/ml/weka/>.

2.4 Algoritmos Bayesianos

Nas palavras de Hruschka e Teixeira [9], as Redes Bayesianas têm como característica representar uma classe ou atributo no nó pai e não permitem que os nós filhos possuam arco ou ligação entre si. Seu processo de classificação é eficiente, pois geralmente seus atributos ou classes são independentes entre si. Esta rede vem sendo utilizada normalmente em processos de classificação e de predição.

Já Cheng e Greiner [10] e Amaral [11] destacam que as redes Bayesianas estão disponíveis na maioria dos softwares de mineração, além da facilidade de implementação de seus algoritmos, também possuem eficiência como classificadores pela boa previsão e precisão.

3. METODOLOGIA E RESULTADOS

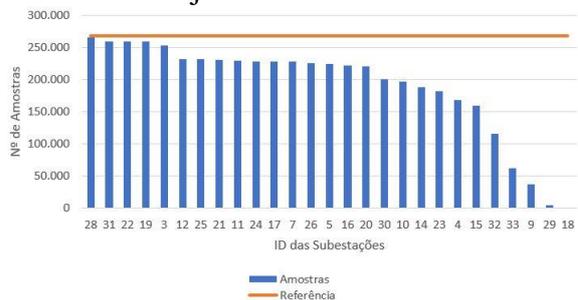
Pretende-se utilizar modelos computacionais e técnicas/algoritmos de descoberta de conhecimento que em conjunto com a análise de um especialista da área possa gerar informações úteis para a tomada de decisões relativa a manutenção da rede de distribuição de energia da CEEE-D.

A partir desta análise será possível encontrar padrões de erros ou mesmo ocorrências de falhas/faltas que irão auxiliar

na manutenção preventiva da rede. Para os testes preliminares deste trabalho serão exploradas as regras de classificação utilizando o algoritmo Bayesiano e como ferramenta para MD será utilizado o software WEKA descrito na subseção 2.3.

A partir da parceria existente entre o GAIC e a CEEE foi realizado o pré-processamento para MD da base de dados do mês de janeiro de 2017 contendo 30 subestações subterrâneas de energia elétrica, sendo que a periodicidade na coleta das grandezas que compõem a base de dados em estudo é de uma aquisição a cada 10 segundos o que mesmo em um curto período de tempo pode se transformar em uma rica e grande base de conhecimento para a concessionária de energia elétrica. Na Tabela 1 é possível visualizar a distribuição dos dados no período para cada uma das subestações da CEEE.

Tabela 1. Aquisição de Dados das Subestações Subterrâneas da CEEE em janeiro de 2017



Fonte: Próprio Autor

Para melhor entendimento dos resultados apresentados na subseção 3.1 os atributos utilizados na Mineração de Dados são respectivamente (*substation_id*, *water_barrier1* e *water_barrier2*), que representam respectivamente o número das subestações, o nível de alagamento inferior e o nível de alagamento superior. Sendo que para os testes iniciais foram analisados os dados somente da primeira semana de janeiro de 2017.

3.1 Resultados

Os resultados obtidos a partir das regras de classificação do algoritmo Bayesiano

estão representados na Tabela 2 e foram minerados a partir de 265.474 aquisições referente aos dados extraídos no período informado na secção 3.

Tabela 2. Resultado da MD utilizando Algoritmo Bayesiano na Primeira Semana do Mês de Janeiro de 2017 (*water_barrier2*)

A	B	classified as
27173	216	A = T
5542	1002702	B = F

Fonte: Próprio Autor

4. CONSIDERAÇÕES FINAIS

Nas regras de classificação obtidas através do algoritmo bayesiano é possível observar que, conforme Tabela 2, o classificador gerado a partir do nível de alagamento 1 (*water_barrier1*) possibilita a classificação correta de 99,44% das amostras do nível de alagamento 2 (*water_barrier2*).

Neste sentido o classificador bayesiano foi capaz de fazer uma classificação satisfatórias dos níveis de água das subestações de energia elétrica. A proposta para trabalhos futuros é utilizar a *water_barrier2* para confirmar se é possível fazer a classificação correta da *water_barrier1*.

5. REFERÊNCIAS

- [1] S. Zhang, et al. The research on smart power consumption technology based on big data. In Smart Grid and Clean Energy Technologies (ICSGCE), 2016 International Conference on, pages 12–18. IEEE, 2016.
- [2] D. V. Rocha, et al. Mineração de dados aplicada a detecção de fraudes nas redes de distribuição de energia elétrica. VI Congresso Brasileiro de Engenharia da Produção. Paraná, 2016.
- [3] S. C. Côrtes et al. Mineração de dados: funcionalidades, técnicas e abordagens. PUC, 2002.
- [4] M. Kamber et al. Data mining: Concepts and techniques. Elsevier, 2012.
- [5] R. Schuch et al. Mineração de dados em uma subestação de energia elétrica. Dincon, 2010.
- [6] C. O. Camilo e J. C. da Silva. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. Universidade Federal de Goiás (UFG), p. 1-29, 2009.
- [7] C. Mccue. Data mining and predictive analysis: Intelligence gathering and crime analysis. Butterworth-Heinemann, 2014.
- [8] I. H. Witten et al. Data Mining: Practical machine learning tools and techniques with java implementations. Morgan Kaufmann, 2005.
- [9] E. R. Hruschka, W. Teixeira. Propagação de Crenças em Redes Bayesianas. Brasília: UnB, 1997 (Relatório de Pesquisa CIC/UNB - 02/97).
- [10] J. Cheng, R. Greiner. Learning Bayesian Belief Network Classifiers: Algorithms and System. Edmonton. Canada: Department of Computing Science, University Alberta, 1999.
- [11] F. Amaral, Aprenda Mineração de dados: teoria e prática, Rio de Janeiro, RJ: Alta Books, 2016, p. 42, 165.