

Uma Aplicação de Data Mining sobre Circuitos Elétricos de Baixa Tensão utilizando Algoritmos Genéticos

Isabela Anciutti, Alexandre L. Gonçalves, Frank A. Siqueira, Paulo S. S. Borges

Universidade Federal de Santa Catarina (UFSC)

Trindade / Caixa Postal: 476 – Florianópolis / SC - 88040-900

isabela@stela.ufsc.br, alexl@stela.ufsc.br, frank@inf.ufsc.br,
pssb@inf.ufsc.br

Abstract. *The Low voltage circuits operate under a variety of factors, including electric, physical, and human. The design and management of those circuits is a challenge for power companies. This paper applies data mining with genetic algorithms to suites of low voltage circuit data. The data were prepared using statistical analysis. Relevant association rules were found, with high fitness. This study aims at contributing to low voltage circuit design and management, as well as to the development of expertise and knowledge on domain-specific optimization problems.*

Keywords: *Data Mining, Genetic Algorithms, Low Voltage Circuits, Association Rules.*

Resumo. *As redes de distribuição de energia elétrica operam sob a complexidade de fatores elétrico, físicos e também humanos. Projetar e controlar estes sistemas de forma efetiva é um desafio constante para as empresas distribuidoras de energia elétrica. Este artigo descreve os processos de data mining aliados à computação evolucionária que, aplicados sobre um conjunto de circuitos elétricos de baixa tensão, objetivou encontrar regras de associação relevantes. Os dados foram preparados por técnicas estatísticas e o algoritmo genético encontrou regras com alto fitness. Este estudo e o conhecimento encontrado através dele deverão ser úteis em futuras pesquisas nesta área e no projeto e manutenção de circuitos elétricos de baixa tensão.*

Palavras-chave: *Mineração de Dados, Algoritmos Genéticos, Redes de Baixa Tensão, Regras de Associação.*

1. Introdução

Data Mining é um conjunto de técnicas e ferramentas aplicado para a descoberta do conhecimento em bases de dados. O conhecimento minerado é utilizado em nível estratégico, para a tomada de decisão.

As aplicações de Data Mining encontram em outras áreas de estudo a construção de abordagens mistas, isto é, soluções multidisciplinares que obtenham melhores resultados,

acrescentando desempenho, confiabilidade e permitindo a otimização do processo de mineração de dados.

Este artigo descreve uma dessas abordagens mistas, que reúne, no contexto de Data Mining, a Computação Evolucionária e técnicas de Estatística, para extrair informações significativas sobre circuitos elétricos de baixa tensão. O objetivo deste estudo vai além do simples trabalho teórico; seu conteúdo visa principalmente à aplicação destas regras no processo estratégico e de tomada de decisão por empresas distribuidoras de energia elétrica.

A próxima seção apresenta os problemas de projeto e manutenção de circuitos de baixa tensão e em seguida discutem-se técnicas alternativas para mineração de dados sobre estes. Na sequência descreve-se o tratamento dos dados para sua adequação a algoritmos genéticos – técnica selecionada. A aplicação do algoritmo genético e os resultados alcançados são discutidos na seção 5. Por fim, a conclusão sumariza os resultados deste trabalho e faz recomendações quanto às direções de pesquisa.

2. Problemas de projeto e manutenção de circuitos de baixa tensão

A situação das redes de distribuição de energia elétrica enfrenta a imprevisibilidade inerente a um sistema que está à mercê de mudanças meteorológicas, variações constantes na demanda de potência (adição no número e no tipo de consumidores), falhas de equipamentos, uso ilegal e impróprio de energia (ligações clandestinas para furto de energia e utilização de eletricidade para fins não discriminados - tipos declarados como "residenciais" utilizando equipamentos comerciais ou industriais), etc.

Acrescenta-se a esses fatores fora do controle humano, outros aspectos passíveis de previsão, gerenciamento e otimização, como é o caso das perdas de energia, avarias em transformadores e alimentadores causadas por sobrecarga de potência ou tensão, quantidade de kVA ociosa, falta de manutenção, entre outros. Entre esses principais aspectos da rede elétrica citados, estão aqueles que são percebidos pelos consumidores e supervisionadas pela Agência Nacional de Energia Elétrica (ANEEL) como indicadores de confiabilidade.

Todos esses problemas geram altos custos à empresa distribuidora de energia; para citar um exemplo, segundo a agência central da Companhia Elétrica do Estado de Santa Catarina (CELESC), são gastos por ano cerca de R\$111.000,00 com aproximadamente 224.000 mil consumidores em problemas de queima de transformadores por sobrecarga (CELESC, 2001).

Para fornecer suporte às atividades estratégicas operacionais envolvidas no problema de projeto e manutenção de redes elétricas, as empresas distribuidoras de energia têm investido na pesquisa de novas tecnologias e técnicas. Tais investimentos possuem o intuito de melhorar a qualidade dos serviços de distribuição elétrica, reduzindo assim as multas impostas pelas agências controladoras, as reclamações de danos a partir dos consumidores, os gastos com manutenção corretiva e os elevados custos na compra e operação de equipamentos; cita-se como exemplo, neste mesmo contexto, (Dabbaghchi et al, 1997), (Fuller, Fuchs e Roesler, 1988), (Laurent et al., 2003) e (Henriques et al, 2003).

Uma das técnicas passível de uso neste contexto é a mineração de dados, que será empregada neste trabalho para encontrar regras de associação¹ em circuitos elétricos de baixa tensão.

3. Mineração de dados em circuitos de baixa tensão

A seguir é discutido como foram selecionadas as técnicas de mineração de dados aplicadas neste trabalho, conceitualizando-as brevemente e expondo de forma resumida os problemas encontrados em outras abordagens dentro do contexto deste mesmo problema.

3.1. Mineração de Dados

A descoberta do conhecimento em bases de dados (KDD) é um processo que envolve seleção, pré-processamento dos dados e transformação dos mesmos; utiliza-se de algoritmos, interpretação dos resultados e a transformação do conhecimento. É dentro desse processo que Data Mining reside, sendo o estágio no processo KDD responsável pela aplicação de algoritmos com o propósito de identificar padrões em uma base de dados ou até mesmo para gerar um conjunto de regras que descreva o comportamento de um banco de dados (Fayyad, 1996).

Mineração de Dados também pode ser descrita como uma área multidisciplinar compartilhada por muitos campos como sistemas de bancos de dados, aprendizado de máquina, sistemas de informação inteligente, estatística, data warehouse e aquisição do conhecimento em sistemas especialistas (Fayyad, 1996).

3.2 Seleção de técnicas para mineração de dados

Ao analisar as redes de distribuição de energia em baixa tensão, tanto em seus aspectos físicos quanto em seus aspectos elétricos, encontramos muitas variáveis que as caracterizam física e eletricamente e que influenciam no comportamento da rede para o consumidor final. Nota-se que os problemas encontrados no nível de rede são resultados do somatório desse mesmo problema em vários circuitos².

Embora individualmente as consequências de tais problemas estejam restritas ao próprio circuito - pelo pequeno valor que as mesmas ocupam na rede -, de uma maneira coletiva, ou seja, no somatório dos circuitos com o mesmo problema, as consequências se estendem a um nível macro. Portanto, o estudo feito aborda o escopo de cada circuito, apoiando-se no fato de que se caso cada um deles esteja operando de forma adequada, a rede de baixa tensão composta pelos mesmos também funcionará de forma correta.

Desse modo, para estudar o comportamento de um circuito elétrico, testou-se algumas abordagens para mineração de dados que não obtiveram resultados satisfatórios:

- Regressão Linear e Não Linear: as funções encontradas alcançaram um alto erro padrão da estimação, além de grandes concentrações não aleatórias de resíduos.

¹ Regras que relacionam a ocorrência simultânea de diferentes características dos dados de um conjunto.

² Chamou-se aqui de circuito o conjunto composto de um aparelho transformador e das ligações elétricas – cabos aéreos, subterrâneos, etc. – entre este e os consumidores da energia por ele distribuída.

- Análise Multivariada: o erro quadrático ajustado (R^2) – que representa o quanto à variância da variável dependente pode ser explicada pelas variáveis independentes – ficou limitado a 50%, isto é, não foi possível explicar, de forma relevante, a variância dos principais atributos do circuito através das demais características nele presentes.

Como nas redes de distribuição de energia podem ocorrer as mais diversas combinações de variáveis em cada circuito que compõe a rede, é necessário empregar um método que possa rapidamente analisar todas as possibilidades existentes. A busca de uma solução ótima dentre um grande número de possíveis soluções pode se beneficiar do uso efetivo do paralelismo, no qual muitas possibilidades são exploradas simultaneamente de um modo eficiente. Desta forma, algoritmos que exploram o paralelismo massivo, como os algoritmos genéticos, que serão descritos em seguida, apresentam-se como alternativas promissoras para serem empregadas neste contexto.

3.3. Algoritmos Genéticos

Em meados dos anos 1960, John Holland acreditou que as peculiaridades da Evolução Natural poderiam ser implementadas de forma algorítmica a fim de alcançar uma versão computacional dos processos de evolução (Mitchell, 1996). Assim, o objetivo de Holland ao inventar os Algoritmos Genéticos era abstrair e rigorosamente explicar o processo adaptativo de sistemas naturais, projetando sistemas artificiais de software que contenham os importantes mecanismos de sistemas naturais (Goldberg, 1989).

Algoritmos Genéticos são algoritmos baseados nos mecanismos de seleção natural e da genética. Tais algoritmos empregam uma estratégia de busca paralela massiva e estruturada, porém aleatória, trabalhando não apenas com uma espécie (ou seja, uma possível solução para o problema em questão) de cada vez, mas testando e modificando milhões de espécies em paralelo. Durante sua execução, os algoritmos genéticos reforçam a busca por pontos de alto *fitness*, isto é, pontos nos quais a função a ser minimizada (ou maximizada) tem valores relativamente baixos (ou altos).

Apesar de aleatórios, os algoritmos genéticos efetuam caminhadas aleatórias direcionadas, que exploram informações anteriormente aprendidas de modo a encontrar novos pontos de busca onde são esperados melhores resultados. Isto é feito através de processos iterativos, onde cada iteração é chamada de geração. Ou seja, o processamento genético serve para otimizar a aptidão de indivíduos sobre gerações sucessivas, propagando o material genético através dos indivíduos mais aptos, de uma geração para outra (Berry e Linoff, 1997).

4. Tratamento dos dados para a aplicação do algoritmo genético

Para alcançar o melhor desempenho do algoritmo genético utilizado, foi preciso tratar os dados visando o uso desta técnica em particular, discretizando-os, dando suporte a dados ausentes, etc.

4.1. Seleção dos Dados

Esta primeira tarefa foi necessária para definir, num primeiro estágio, quais características dos circuitos de baixa tensão eram pertinentes à análise pretendida e possuíam alguma possível

contribuição para a obtenção das regras. Esta atividade de seleção de dados foi acompanhada por um engenheiro eletricista especialista em redes de distribuição de energia, o qual orientou e ajudou a definir o escopo do problema.

Os dados foram extraídos de um sistema de software corporativo de uma empresa de distribuição de energia elétrica³, num total de 5210 registros, correspondendo a um igual número de circuitos de baixa tensão existentes em uma região geográfica do estado e compostos de 120 atributos. Dentre esses campos, apenas 30 foram selecionados para análise.

4.2. Pré-processamento dos dados

Como é usual em processos de Data Mining, neste estudo também foi preciso executar a preparação dos dados, de forma a evitar *outliers*, dados ausentes e perdidos, etc. Seguindo esta orientação, apenas 3072 indivíduos adequados foram encontrados a partir da amostra original.

4.3. Transformação dos dados

Após certificar-se de que o conjunto era composto apenas por valores válidos, foi necessário prepará-los para oferecerem o máximo de significância e robustez à execução do algoritmo genético e a solução por ele gerada.

Constatou-se que alguns atributos, como o desequilíbrio do circuito, por exemplo, não estavam consolidados de forma que se pudesse realizar uma análise direta. Para se obter o atributo desequilíbrio nas saídas do trafo, calculou-se a média entre as correntes passantes de cada fase para cada trecho de saída:

$$\bar{x} = \frac{(i_a + i_b + i_c)}{3}$$

Onde: $\bar{x}$: Média das correntes passantes de cada fase do trecho, i_a : Corrente passante⁴ na Fase A, i_b : Corrente passante na Fase B, i_c : Corrente passante na Fase C.

Depois de calculada a média, divide-se a corrente de cada trecho pela média, subtrai-se 1 e retorna-se o módulo. O máximo do módulo entre os valores calculados é o desequilíbrio. Ou seja, o maior valor entre os resultados das três equações seguintes:

$$\left(\left| \frac{i_a}{\bar{x}} - 1 \right| \right) \left(\left| \frac{i_b}{\bar{x}} - 1 \right| \right) \left(\left| \frac{i_c}{\bar{x}} - 1 \right| \right)$$

Foi determinada também a quantidade de trechos em cada circuito nas seguintes faixas de queda de tensão:

- 0 a 3% [220 a 213.4)

³ A utilização dos dados para estudo foi autorizado pela empresa distribuidora de energia elétrica.

⁴ Corrente Passante: intensidade da corrente elétrica (medida em amperes) em uma dada fase de tensão.

- 3% a 5% [213.4 a 209)
- 5% a 7.5% [209 a 203.5)
- Maior que 7.5% (Abaixo de 203.5)

Para determinados atributos foram elaborados critérios através do cálculo de uma média matemática ponderada, chamada de centróide, para realizar a classificação do circuito. O cálculo do centróide é o seguinte:

$$\text{Centróide} = Pf_1 * \text{Peso}_1 + Pf_2 * \text{Peso}_2 + Pf_3 * \text{Peso}_3 + Pf_4 * \text{Peso}_4 \dots Pf_n * \text{Peso}_n$$

$$\text{Centróide} = \sum_{i=1}^n (pf_i * \text{peso}_i)$$

Onde:

pf_i representa o percentual da *inésima* faixa em relação ao total e o peso_i representa o peso aplicado ao percentual da *inésima* faixa.

A partir dos valores calculados para os centróides, estipularam-se faixas de valores para se realizar uma classificação. Então, para cada atributo, o cálculo do centróide e a classificação do resultado obtido foram realizados conforme definido pelas seguintes regras:

- Cálculo do centróide de tensão:

$$\text{Centróide_Tensão} = (\text{Percentual } 3.5\% \text{ Queda} * 1.5) + (\text{Percentual } 5\% \text{ Queda} * 3) + (\text{Percentual } 7.5\% \text{ Queda} * 6.25) + (\text{Percentual } > 7.5\% \text{ Queda} * 12)$$

- Classificação da queda de tensão:

- Circuito Adequado: $\text{Centróide_Tensão} = 150$
- Tensão Baixa: $150 < \text{Centróide_Tensão} \leq 168$
- Tensão Baixa Precária: $168 < \text{Centróide_Tensão} \leq 230$
- Tensão Baixa Crítica: $\text{Centróide_Tensão} > 230$

- Cálculo do centróide de fases:

$$\text{Centróide_Fase} = (\text{Percentual Trifásico} * 1) + (\text{Percentual Bifásico} * 3) + (\text{Percentual Monofásico} * 9)$$

- Determinação da quantidade de fases:

- Circuito Monofásico: $\text{Centróide_Fase} > 1.8$
- Circuito Trifásico: $\text{Centróide_Fase} \leq 1.8$

Em sua maioria, os dados possuem valores contínuos, o que não é de forma alguma recomendável para o uso de operadores genéticos, por exemplo, ao uso do *crossover*, já que pode não haver intersecção no próprio intervalo de valores entre um atributo e outro. Portanto, antes que os dados pudessem ser utilizados pelo algoritmo genético, foi necessário discretizá-los. Em primeiro lugar encontrou-se o número de classes mais adequado para o tamanho do conjunto, seguindo a fórmula de Sturges (Peso-Aguiar, 1995): $1 + 3.3 \log_{10} N$, onde N é o tamanho da amostra.

Encontrado o número de classes mais adequado (doze faixas), foi preciso separar cada valor de determinado atributo de acordo com a classe a que pertencia, ditada segundo a frequência de registros com aquele mesmo valor. Para os atributos já discretizados em um número de faixas menor do que o encontrado pela fórmula de Sturges manteve-se a classificação. Nestes atributos foi mantido o número original de classes, para que não houvesse prejuízo devido à utilização de valores reais na descrição das faixas - a discretização feita objetivava evitar o uso de valores contínuos e não seria lógico agora permiti-los por conta de exceções.

Tendo em mãos uma amostra válida, descrita através de classes, cujos valores obtidos se mostraram coerentes ao que descrevem (um circuito) e completos em seu conteúdo, foi necessário definir os atributos que comporiam o cromossomo, ou seja, as características que comprovadamente contribuiriam para uma regra útil.

Para selecionar os atributos que mais teriam influência em uma regra cujo conseqüente⁵ é conhecido, foi feita uma análise estatística descritiva, examinando-se a correlação de todos os atributos com o atributo conseqüente da regra. Foi gerada uma análise de correlação de cada uma das variáveis dependentes em relação às demais, ou seja, em relação aos atributos: situação do circuito, potência demandada no circuito secundário e índice de carregamento de potência do circuito secundário. Aqueles campos que atingissem mais de 50% de correlação com as variáveis independentes eram selecionados para compor o cromossomo que serviria para a regra. A partir dos resultados obtidos foi possível diminuir o tamanho do cromossomo, descartando atributos do conjunto de dados que não influenciavam no objetivo da regra.

O procedimento descrito foi repetido para todas as regras que se desejava encontrar e a programação no software, para que o algoritmo buscasse a regra desejada, foi feita conforme estes campos selecionados. Desse modo, basta que o usuário escolha qual a regra que busca descobrir e o algoritmo processará os resultados baseando-se apenas nos atributos comprovadamente pertinentes.

Em relação às regras pretendidas, os campos descritos na tabela alcançaram mais de 50% de correlação com o atributo conseqüente da regra (cabecalho da tabela 1).

Tabela 1. Atributos selecionados para as regras de associação.

Situação	Potência demandada no circuito secundário	Índice de carregamento de potência do circuito secundário
Potência perdida no circuito. Índice de Queda de Tensão entre Fase A e o neutro Índice de Queda de	Módulo da demanda total de potência do posto transformador Índice de carregamento do trafo Potência perdida no circuito Índice de Queda de Tensão entre Fase A e o neutro	Demanda de potência de consumidores secundários Potência perdida no circuito Índice de Queda de

⁵ Uma regra é formada pela premissa e pelo conseqüente, onde a premissa é um conjunto de atributos valorados e o conseqüente é o atributo valorado resultante e dependente da ocorrência desse conjunto. Por exemplo: um determinado conjunto de aspectos climáticos reunidos gera o fenômeno natural furacão.

Tensão entre Fase B e o neutro	Índice de Queda de Tensão entre Fase B e o neutro	Tensão entre Fase B e o neutro
Índice de Queda de Tensão entre Fase C e o neutro	Índice de Queda de Tensão entre Fase C e o neutro	Índice de Queda de Tensão entre Fase C e o neutro
Distância do trafo até o trecho onde ocorreu a maior queda de tensão	Energia total (kWh) medida para a classe de consumidores comerciais	Energia total (kWh) medida para a fase elétrica A
	Energia total (kWh) medida para a classe de consumidores residenciais	Energia total (kWh) medida para a fase elétrica B
	Quantidade de kWh consumido para a fase elétrica A	Energia total (kWh) medida para a fase elétrica C
	Quantidade de kWh consumido para a fase elétrica B	Energia total (kWh) medida para a fase elétrica A
	Quantidade de kWh consumido para a fase elétrica C	Energia total (kWh) medida para a fase elétrica B
	Quantidade total de consumidores na rede de baixa tensão	Energia total (kWh) medida para a fase elétrica C
	Quantidade de consumidores para a fase elétrica A	Energia total (kWh) medida para a classe de consumidores residenciais
	Quantidade de consumidores para a fase elétrica B	
	Quantidade de consumidores para a fase elétrica C	
	Quantidade de consumidores para a classe residencial	

5. Aplicação do Algoritmo Genético

Após todos os atributos do circuito terem sido definidos e terem passado pelos processos de limpeza e validação, os mesmos foram inseridos em um banco de dados. Outras tabelas foram criadas no mesmo banco de dados para apoiar a execução do algoritmo visando otimizar o processo e armazenar os resultados obtidos para posterior análise.

5.1. Representação do cromossomo

Definiu-se um indivíduo como um circuito de baixa tensão, onde cada característica do circuito (índice de carregamento, quantidade de unidades consumidoras por classe, fator de potência, etc.) é um gene do cromossomo.

Apesar do uso de um alfabeto binário para a representação cromossômica ser muito indicado e defendido (Goldberg, 1989), não foi possível criar uma estrutura cujos alelos fossem binários, pois se perderia a potencialidade dos dados quanto ao universo de soluções. Seria necessário limitar os aspectos do circuito quanto a seus valores em apenas duas classes, gerando resultados não tão significativos quanto poderiam se divididos em mais faixas.

5.2. Função Objetivo

A função *payoff* implementada para este problema é razoavelmente simples no sentido de que sua lógica pode ser considerada aplicável para qualquer problema; no entanto, requisita modificação quanto ao seu código já que faz menção em sua descrição aos aspectos do problema em que se aplica.

Para calcular o *fitness* de um determinado indivíduo, inicialmente executa-se uma consulta SQL ao banco de dados tendo como restrição (em sua cláusula "where") o conseqüente da regra; ou seja, se desejamos saber o que caracteriza um circuito com situação adequada de tensão, selecionamos todos os registros da base que contenham o campo

referente à situação de tensão do circuito (campo SITUACAO) correspondente ao valor da "situação adequada".

A partir deste conjunto de dados, cada aspecto do circuito no registro do banco de dados é comparado com o correspondente aspecto no indivíduo da geração. Para cada gene com valor igual ao valor do respectivo atributo em cada um dos registros é acrescentado um ponto ao seu *fitness*. Se o indivíduo como um todo for encontrado no conjunto de registros trazido do banco de dados, vinte pontos são acrescentados ao seu valor total de *fitness*, premiando assim indivíduos cujos valores alcancem dados reais.

Não foi implementada uma lógica de penalidade porque isso envolveria mais processamento computacional - medindo o quanto o valor gerado através dos operadores genéticos distancia-se do valor real presente na base de dados - e complexidade - por exigir definição de quanto e sob que circunstâncias penalizar aquele indivíduo. Entretanto, mais do que por essas razões, o objetivo desse estudo tinha entre suas prioridades a simplicidade na implementação, pois se pretendia basicamente testar e comprovar se o uso de algoritmos genéticos podia alcançar ótimos resultados na geração de regras de associação sobre dados de circuitos de baixa tensão.

5.3. Tamanho da população

Quanto ao tamanho da população, decidiu-se utilizar cerca de um terço dos registros do conjunto de dados para compor a população processada pelo algoritmo. Essa decisão explica-se pelo fato de que determinados aspectos referentes aos circuitos nem sempre estavam presentes em um número adequado na base de dados. Feito deste modo, 1000 registros são utilizados como população, enriquecendo em muito a diversidade possível na evolução dos indivíduos. Os benefícios alcançados por essa abordagem refletem-se no *fitness* máximo obtido quando do uso de apenas 100 indivíduos na população em comparação com a base completa (3072 registros).

5.4. Taxa de Cruzamento e de Mutação

Quanto ao valor das probabilidades, escolheu-se, após vários testes, os valores de 60% para *crossover* e 8% para mutação. À medida que esses valores foram elevados, percebeu-se que os indivíduos com mais alto *fitness* eram encontrados menos vezes pelo algoritmo. Esse aspecto, de modo inverso, também se refletiu pelo número de gerações selecionado para o processo; com mais de seis gerações, nenhum novo indivíduo surgia, apenas encontravam-se mais vezes o mesmo cromossomo, ou seja, o algoritmo convergia para àquele cromossomo, passando a gerar populações constituídas do mesmo material genético.

5.5. Algoritmo utilizado

O algoritmo genético utilizado para este estudo é uma adaptação do algoritmo descrito por Goldberg (1989). Para tal foi implementado um programa com as devidas alterações na função objetivo que satisfazem os propósitos desse estudo. Para cada uma das regras desejadas foram extraídos os registros da base de dados, bem como os atributos pertinentes à regra em questão. O programa principal, que apresenta a arquitetura hierárquica e de controle do software, está descrito abaixo:

```

begin {programa principal}
  gen := 0;
  initialize;
  statistics (popsize, max, avg, min, sumfitness, oldpop);
  repeat_
    gen := gen + 1;
    generation;
    statistics (popsize, max, avg, min, sumfitness, newpop);
    oldpop := newpop;
  until (gen >= maxgen);
end.

```

Onde:

- gen: geração atual
- maxgen: número pré-definido de gerações
- popsize: tamanho pré-definido da população
- max: fitness máximo da população
- avg: média do fitness da população
- min: fitness mínimo da população
- sumfitness: somatório do fitness
- oldpop: antiga geração da população
- newpop: nova geração da população

O primeiro procedimento, *initialize*, executado no algoritmo é a inicialização da população de indivíduos segundo o tamanho que foi determinado para a população (*popsize*). Uma consulta SQL é feita à base de dados, trazendo todos os registros da tabela. A população inicial é então selecionada desse conjunto de modo aleatório e sem repetição.

Em seguida o procedimento *statistics*, que representa a função de avaliação, analisa a população inicial extraída, calculando os parâmetros que irão medir a qualidade dos indivíduos na escolha das próximas gerações. Entre esses parâmetros estão o *fitness* máximo e o somatório do *fitness* daquela geração.

Tendo uma população inicial, com suas características individuais (pertencentes a cada indivíduo) calculadas, o algoritmo inicia o processo repetitivo para otimizar as possíveis soluções no ambiente do problema. Isto é feito produzindo-se novas gerações da população de acordo com o valor do *fitness* alcançado por elas. Este ciclo termina quando o número de gerações é alcançado - preferiu-se deixar ao usuário e não ao algoritmo a análise quanto à significância da variação obtida na população ou sobre a eficácia dos operadores genéticos.

Por fim, no procedimento de geração chamado *generation*, ocorre a aplicação do *crossover* e da mutação. O critério para a seleção dos indivíduos que darão origem a descendentes baseia-se em um cálculo aleatório envolvendo o somatório do *fitness* alcançado pela população anterior, sem restrição quanto a serem escolhidos mais de uma vez. Após selecionar dois indivíduos seguindo estas regras, os mesmo sofrem *crossover* e mutação de acordo, respectivamente, com as probabilidades de *crossover* (recombinação) e mutação (transformação) definidas pelo usuário.

Com este método, os melhores indivíduos deram origem à nova população na medida de dois cromossomos antigos para dois novos cromossomos, modificados geneticamente na tentativa de atingir um maior espaço no escopo de soluções possíveis.

Por este algoritmo utilizar a estratégia "geracional" - onde a população antiga ajuda a criar uma nova população para substituir a anterior ao invés de participar da anterior

(estratégia "estado estável") - acreditou-se que taxas de probabilidade não muito altas ajudariam a manter um equilíbrio, isto é, indivíduos com alto *fitness* poderiam passar sem transformação para a população seguinte ao invés de se perderem por causa de um valor muito alto para os operadores genéticos.

5.6. Resultados obtidos

Resultam da aplicação do algoritmo genético três regras de associação:

- Quais são as características dos circuitos que possuem tensão adequada (intervalo de [220 a 213.4])?
- Quais são as características dos circuitos com potência demandada inferior a 14,45 kVA?
- Quais são as características dos circuitos com índice de carregamento de potência acima de 86,5%?

A tabela 2 ilustra os intervalos de valores para os principais atributos de cada regra de associação detectada pelo algoritmo genético.

Tabela 2. Regras de associação (fragmento).

Questão 1	Questão 2	Questão 3
Potência perdida no circuito $\leq$ 0,085	Quantidade de kWh consumido para a fase elétrica A $\leq$ 30,5	Demanda de potência de consumidores secundários $\geq$ 75,05
Índice de Queda de Tensão entre Fase A e o neutro $\leq$ 0,675	Índice de carregamento do trafo $\leq$ 42,5	Energia total (kWh) medida para a classe de consumidores residenciais $\geq$ 10,5
	Quantidade de consumidores para a fase elétrica A $\leq$ 0,5	

Com base nos intervalos de valores dos atributos de cada regra de associação é possível determinar as razões pelas quais indicadores vitais para a qualidade da rede de transmissão de energia – como a tensão, a potência e a carga, por exemplo – podem vir a se alterar. Estas informações podem ser utilizadas tanto no projeto quanto na operação das redes de distribuição de energia elétrica de modo a aprimorar a qualidade do serviço prestado pelas distribuidoras de energia.

6. Conclusão

Este artigo apresentou uma abordagem de mineração de dados com algoritmo genético aplicados a circuitos elétricos de baixa tensão. Durante o processo de mineração de dados, diversos testes foram feitos antes de alcançar a adequação dos dados considerando qualidade, formato e pertinência para o uso efetivo do algoritmo genético selecionado.

O algoritmo foi executado com várias combinações de parâmetros para certificar-se que as regras de associação encontradas tinham feito uso do potencial máximo do algoritmo, assim como se certificar e que a função de *payoff* era capaz de manter a robustez dentro de um ambiente com características aleatórias.

Os resultados obtidos corroboram a recomendação de Goldberg (1989) quanto à utilização de algoritmos genéticos em problemas de otimização com busca paralela e característica de explosão combinatória, pois o objetivo principal deste estudo foi alcançado mesmo na utilização de um modelo simples de algoritmo genético.

Outros algoritmos na área de computação evolucionária estão sendo atualmente investigados para mineração de dados na área de circuitos elétricos. Futuros estudos poderão especializar o algoritmo, implementando elitismo e penalidade na função objetivo.

7. Agradecimentos

Ao Prof. Dr. Vinícius Medina Kern pela revisão crítica e orientação na preparação do artigo.

Referências Bibliográficas

CELESC, documento de circulação interna à empresa, 2001.

David E. Goldberg, "Genetic Algorithms in Search, Optimization & Machine Learning, "Addison Wesley Longman, 1989.

Gregory Piatetsky-Shapiro, "Knowledge Discovery in Real Databases", 1991, A Report on the IJCAI-89 Workshop. AI Magazine 11(5): 68-70.

H.O. Henriques, D.M. Falcão, C. L. T. Borges, G.N. Taranto, A. Manzoni, W.S. Andrade, E.C. Viveros, "Aplicações de Sistemas Inteligentes e Processamento Distribuído na Previsão, Localização e Minimização de Perdas em Redes de Distribuição", trabalho submetido ao II CITENEL, 2003.

Henri Laurent, Luiz A. T. Machado, Daniel Vila, Carlos, Morales, Juan Ceballos, Fernando Mirancos e Shiguematsu Nosaki, "Sistema de Previsão Imediata de Ocorrência de Tempestades para Apoio a Tomada de Decisão na Distribuição e Manutenção da Rede Elétrica", trabalho submetido ao II CITENEL, 2003.

I. Dabbaghchi; R. D. Christie, G. W. Rosenwald; Chen-Ching Liu, "AI Application Areas in Power Systems", IEEE Expert. Vol. 12, No. 1 pp. 58-66. 1997.

J. F. Fuller, E. F. Fuchs, and K. J. Roesler, "Influence of harmonics on power distribution system protection", IEEE Trans. Power Delivery, vol. 3, pp. 549-557, Apr. 1988.

Michel J. A. Berry, G. Linoff, "Data Mining Techniques - for Marketing, Sales, and Customer Support", John Wiley & Sons, New York, 1997.

Melanie Mitchell, "An Introduction to Genetic Algorithms", A Bradford Book, 1996.

Peso-Aguiar, M. C., *Macoma contricta* (Bruguière, 1792) (BIVALVIA– TELLINIDAE) como biomonitor da presença crônica de petróleo na Baía de Todos os Santos (BA). São Carlos: UFCar. Tese de Doutorado, 160p, 1995.

U. M. Fayyad, "Data mining and knowledge discovery: making sense out of data", IEEE Expert, 1996.