

MODELO DE ANÁLISE ESPACIAL DE MERCADO DE ENERGIA ELÉTRICA UTILIZANDO TÉCNICAS DE CLASSIFICAÇÃO

JOÃO CAETANO NETO*, WAGNER DA SILVA LIMA†

**Escola de Engenharia Elétrica e de Computação
Universidade Federal de Goiás
Goiânia, Goiás, Brasil*

*†Escola de Engenharia Elétrica e de Computação
Universidade Federal de Goiás
Goiânia, Goiás, Brasil*

Emails: joaocaetano@gmail.com, wagnerdasilvalima@gmail.com

Abstract— The electrical distribution companies pass through an antagonistic reality - a growing increase of energetic demand, concerning to the continuous economic and population growth, and a financial resources restriction to invest in the expansion of their electrical network. If so, the activities improvement of energy distribution systems planning is indispensable to the efficient application of available resources. The application of Geographic Information System associated to assortment and grouping methods fit in this context to potentialize the planning process, providing to the users a complete vision about the electrical system. This paper presents a system that uses Geographic Information System associated to assortment and grouping methods K-means, Self-Organizing Maps and Fuzzy C-Means with the goal of analyzing the network load distribution and comparing these methods performance. Each accomplished grouping creates a spatial representation (scene). This scene amongst a grouping performance measurement index (intragroup and intergroup) implemented in this thesis, provide a favorable environment for outcome analysis.

Keywords— Geographic Information System, Electrical Distribution Network, Planning, Grouping, Assortment.

Resumo— As companhias de distribuição de energia elétrica convivem com uma realidade antagônica; um crescente aumento da demanda energética, devido ao constante crescimento econômico e populacional, e uma limitação em seus recursos financeiros para expansão de sua rede. Assim sendo, torna-se imprescindível um aprimoramento na atividade de planejamento do sistema de distribuição de energia procurando aperfeiçoar a aplicação dos recursos disponíveis. Neste contexto se enquadra a aplicação do Sistema de Informação Geográfica aliado a técnicas de agrupamento e classificação com o objetivo de potencializar o processo de planejamento, proporcionando ao planejador uma visão mais completa do sistema elétrico. Este artigo apresenta um sistema que faz uso do Sistema de Informação Geográfica combinado com a aplicação das técnicas de agrupamento e classificação K-Médias, Mapas Auto-Organizáveis e Fuzzy C-Médias, com o intuito de analisar a distribuição de carga da rede e de comparar o desempenho das técnicas utilizadas. Cada agrupamento realizado origina uma representação espacial (cenário). Este, juntamente com um índice de medida da performance do agrupamento (intra-grupo e inter-grupo) implementado neste trabalho, proporciona um ambiente favorável para a análise dos resultados.

Keywords— Sistema de Informação Geográfica, Rede de Distribuição de Energia, Planejamento, Agrupamento, Classificação.

1 Introdução

Com o processo de desverticalização ocorrido no Setor Elétrico Brasileiro na última década, os agentes tradicionais (geração, transmissão, distribuição e comercialização) foram redefinidos, passando a ter estruturas individualizadas (separação entre transporte, produção e comercialização de energia elétrica).

A atividade de planejamento, que é essencial a qualquer sistema, torna-se imprescindível à distribuição de energia elétrica, de forma a atender o crescimento de carga em níveis de qualidade de serviços compatíveis com suas características, procurando aperfeiçoar a aplicação dos recursos disponíveis que são relativamente escassos.

Os sistemas de análise de mercado utilizados hoje normalmente fornecem ao planejador apenas um conjunto de dados organizados em relatórios, planilhas e gráficos, o que não proporciona um

ambiente de análise espacial para crítica dos valores fornecidos pela base de dados da empresa. Portanto, um ambiente que forneça, além dos dados dos clientes, informações sobre sua localização, curvas de carga, criação de cenários, histogramas de carga, agrupamento de classificação das curvas de carga e a análise do crescimento horizontal e vertical da rede de distribuição é parte integrante do processo de previsão de carga futura.

2 Técnicas de Agrupamento e Classificação

Para a análise de carga espacial da rede de distribuição, faz-se primordial realizar o agrupamento de suas áreas. Para esta tarefa pode ser empregada a clusterização (formação de agrupamentos), processo que pelo qual procuram-se classificar objetos em categorias.

A clusterização tem por finalidade reunir os

padrões em vários grupos, sendo que os dados a serem analisados terão um grau de similaridade alto com o grupo de sua classificação e tão diferentes quanto possível, dos elementos dos outros grupos.

Há diferentes abordagens para o tratamento do problema de clusterização, dentre elas, a abordagem convencional, onde cada elemento a ser classificado pertencerá totalmente a uma determinada categoria, e a abordagem nebulosa, onde um elemento pode ser classificado em várias categorias, possuindo diferentes graus de associação a cada uma delas (Santos, 2005).

Na clusterização são identificadas as classes de cada um dos elementos através de suas descrições, como um vetor de pares (atributo, valor do atributo). Baseando-se em certas suposições e/ou critérios, os algoritmos de clusterização tentam particionar o conjunto de objetos de entrada, o que conseqüentemente pode produzir ou não saídas significativas e úteis destes dados de entrada.

Neste trabalho foram utilizadas a abordagem convencional, através das técnicas de agrupamento e classificação K-Médias (KMD) e Mapas Auto-Organizáveis (SOM), e a abordagem nebulosa, mais flexível, através do método Fuzzy C-Médias.

2.1 Redes Mapa-Organizáveis

As redes auto-Organizáveis (redes SOM) são uma classe de redes neurais artificiais (RNA), as quais possuem capacidade de aprender através de exemplos, sem a necessidade de ser explicitamente programada.

Esta classe de RNA é treinada sem a necessidade de um supervisor externo ou punição/recompensa, sendo um conjunto de padrões de entrada a única informação fornecida para o treinamento. Desta forma, estas redes definem seus parâmetros por si próprias, sem auxílio externo.

Nestas redes, os nós estão espacialmente ordenados dentro de áreas, onde os nós topologicamente próximos tendem a responder a padrões ou estímulos semelhantes. Essa ordenação topológica é resultado do uso de feedback lateral entre os nós, geralmente modelado por uma função conhecida como "chapéu mexicano" (Carvalho et al., 1998).

O mapeamento dos padrões de entrada nos neurônios da rede SOM é realizado durante a fase de treinamento, a qual de acordo com (Haykin, 2001) possui três etapas básicas:

- **Competição:** Para cada padrão de entrada, os neurônios do mapa calculam seus respectivos valores de uma função discriminante. Esta função fornece a base para a competição entre os neurônios. O neurônio com o maior valor da função discriminante é declarado vencedor da competição;

- **Cooperação:** O neurônio vencedor determina a localização espacial de uma vizinhança topológica de neurônios excitados que cooperarão entre si;
- **Adaptação sináptica:** Os neurônios excitados aumentam seus valores individuais da função discriminante em relação ao padrão de entrada através de ajustes adequados aplicados a seus pesos sinápticos. Os ajustes feitos são tais que a resposta do neurônio vencedor à aplicação subsequente de um padrão de entrada similar é melhorada.

Neste trabalho foi utilizada a variação da rede SOM, denominada Kohonen em Anel, a qual possui uma camada de saída unidimensional e uma vizinhança em forma de anel, onde o primeiro nó do anel é vizinho do último.

O treinamento desse modelo, baseia-se simplesmente na procura do neurônio cujos pesos são mais próximos de um determinado padrão de entrada (com a menor distância Euclidiana) e no aumento da similaridade entre eles (padrão de entrada e pesos do neurônio vencedor) (Batista, 2004).

2.2 K-Médias

O K-Médias é um tipo de algoritmo de agrupamento, que pode ser utilizado para a classificação não-supervisionada. Nesta técnica os grupos baseiam-se em sua proximidade em relação a um centróide, o qual é obtido das médias das características que descrevem os padrões de entrada.

O algoritmo K-Médias é um processo iterativo que tenta minimizar o erro quadrático médio atribuindo cada padrão de entrada ao grupo de centróide mais próximo.

Este algoritmo realiza uma busca de um ponto de máximo, o qual é atingido quando não há mais mudança na escolha do centróide para cada entrada, dado um critério de parada a partir da solução inicial (Drummond, 2003). Não existem garantias de que o algoritmo encontre o máximo global, sendo possível que o algoritmo apresente diferentes soluções quando executado mais de uma vez.

2.3 Fuzzy C-Médias

A lógica nebulosa (*fuzzy*) possibilita que seja abordado de forma mais adequada a representação e manipulação de conhecimentos na área de Inteligência Artificial (IA), onde expressões que usualmente não podem ser tratadas pelos sistemas da lógica clássica, tais como "quase", "muito" e "pouco" são atendidas nos sistemas nebulosos.

No agrupamento nebuloso um elemento pode pertencer a vários grupos ou classes, com diferentes valores de pertinência, já no agrupamento com

a abordagem convencional, cada elemento pertence totalmente a um único grupo, como os dois modelos de agrupamento citados anteriormente (Ross, 1995).

O algoritmo de agrupamento de dados nebuloso Fuzzy C-Médias baseia-se em um modelo não linear de otimização que agrupa dados com características similares de acordo com um procedimento iterativo de minimização de uma função objetivo (distância Euclidiana) que representa um critério de partição, ponderado pelos graus de pertinência dos dados aos respectivos grupos (Bezdek, 1981).

2.4 Índice de Comparação de Métodos de Agrupamento e Classificação

Para análise dos métodos de agrupamento e classificação utilizados neste trabalho, foram utilizados dois índices (Zanon and Lima, 2003):

- Intra-grupo (IAG);
- Inter-grupo (IEG).

O índice Intra-grupo (equação 1) é uma função de custo baseada no erro quadrático médio e indica o quanto os padrões de entrada que estão em um mesmo grupo são similares.

$$IAG = \sum_{i=1}^D \sum_{x \in D_i} \|x - m_i\|^2 \quad (1)$$

Onde x é o número de amostras de D_i e m_i é a média dessas amostras. Assim, para um dado *cluster* D_i , o vetor m_i é o melhor representante das amostras em D_i de forma a minimizar a soma dos comprimentos quadrados de cada x em relação a m_i .

Assim o IAG mede o erro quadrático total em representar as amostras x_1, x_2, \dots, x_n por D agrupamentos com o centro em m_1, m_2, \dots, m_n . Quanto menor for o IAG, melhor é o agrupamento ou classificação dos padrões, ou seja, os padrões que estão em um mesmo grupo possuem uma semelhança maior.

O índice inter-grupo (IEG) determina a relação entre os grupos, ou seja, o quão distantes eles estão. Seu cálculo é feito da seguinte forma:

$$IEG = \sum_{j=1}^D \|X_j - X\|^2 \quad (2)$$

Onde X_j é a curva do centróide do grupo j e X é a média global das curvas de carga do conjunto de treinamento.

Diferente do IAG este índice não permite afirmar que quanto menor o valor, melhor é o resultado, mas apenas que se ele tiver um valor alto os grupos estão mais separados ou se ele tiver um valor baixo significa que há uma semelhança maior entre os padrões de grupos diferentes.

3 Sistema Proposto

Este trabalho deu origem a um programa, denominado GeoCom, que utilizando das funções do Sistema de Informação Geográfica (SIG) e das técnicas de classificação Kohonen em anel, K-Médias e Fuzzy C-Médias, cria um ambiente favorável à análise comercial de empresas de distribuição de energia elétrica que possuam um levantamento georeferenciado de sua rede de distribuição.

Como funcionalidades básicas de todo sistema de informação geográfica, as ferramentas de navegação, controle de camada, escala e símbolos também estão presentes no GeoCom (Figura 1).

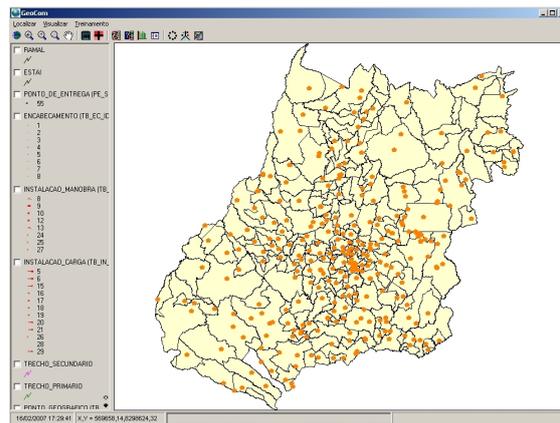


Figura 1: Tela principal do sistema GeoCom

3.1 Metodologia Geral proposta para o Agrupamento

Para possibilitar o agrupamento e classificação das curvas de carga com o apoio do SIG foi proposta uma metodologia que possui quatro etapas, cada uma com suas respectivas atividades.

3.1.1 Coleta de informações - etapa 1

Nesta etapa está prevista a obtenção das informações necessárias para o agrupamento, ocorrendo o ajuste destas informações na etapa seguinte. São obtidos do setor comercial os dados de consumo de energia dos consumidores em kWh, referente aos 12 meses que compõem o cenário. A partir destes dados é avaliado o consumo de energia das entidades da rede elétrica georeferenciadas (pontos de entrega e estações transformadoras).

3.1.2 Pré-carga de dados - etapa 2

A base do agrupamento e classificação das áreas da rede de distribuição é a geração das curvas de carga. Estas curvas são geradas por quadrícula, onde cada quadrícula corresponde a uma grade uniforme de 500 x 500 m. O processo de geração da curva de carga por quadrícula ocorre da seguinte forma:

- Primeiramente é selecionada a área de abrangência do estudo (município).
- Para cada quadrícula são somados os consumos médios (kWh) dos clientes presentes na área da mesma.
- O consumo obtido na quadrícula é dividido por classe de clientes.
- As curvas de carga são normalizadas.

Como os clientes não são entidades espaciais, estes são identificados em uma quadrícula através dos transformadores e dos pontos de entrega (pontos de medição) presentes na área da quadrícula.

A curva de carga da quadrícula obtida é referente ao total dos consumos dos últimos doze meses e a composição das curvas das quadrículas de um município gera o cenário de consumo do mesmo. Assim, por exemplo, para gerar o cenário de 2005 basta carregarmos o histórico de consumo daquele ano (etapa 1) e o sistema automaticamente identificará que os dados são referentes ao cenário de consumo de 2005.

O consumo total obtido em uma quadrícula é dividido em classes, gerando uma curva de classe por consumo. Devido a variabilidade de magnitude encontrada nas curvas de carga, o processo de normalização se faz necessário, sendo o intervalo [-1,1] definido neste trabalho. Foram determinadas 10 classes possíveis de clientes:

- Residencial de 0 a 50 kWh
- Residencial de 50 a 200 kWh
- Residencial de 200 a 400 kWh
- Residencial acima de 400 kWh
- Comercial de 0 a 200 kWh
- Comercial acima de 200 kWh
- Industrial abaixo de 1000 kWh
- Industrial acima de 1000 kWh
- Rural
- Outros

Na geração das curvas por município o sistema calcula tanto pelos transformadores presentes na quadrícula quanto pelos pontos de entrega. Esta análise é possível através de consultas espaciais, onde é realizada a sobreposição de camadas de dados (Casanova et al., 2005). Isso é feito para proporcionar ao planejador a opção de agrupar as curvas por transformador ou pontos de entrega, já que a quantidade de clientes na quadrícula pode ser diferente.

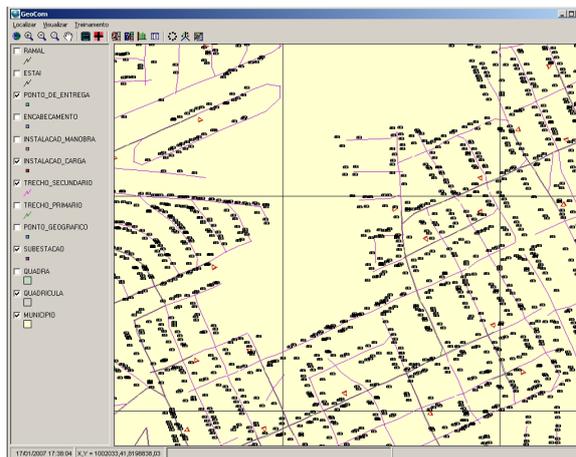


Figura 2: Exemplo de distribuição dos clientes nas quadrículas.

Note na Figura 2 que clientes alimentados por um transformador podem estar localizados em outra quadrícula, já que estes transformadores atendem seus clientes através das redes de baixa tensão (linhas na cor magenta), as quais normalmente possuem extensões que ultrapassam a área de uma quadrícula (500 x 500 m).

3.1.3 Agrupamento e classificação - etapa 3

Após a geração das curvas de carga aplica-se o processo de agrupamento das quadrículas gerando-se grupos que posteriormente serão rotulados. O processo de agrupamento pode ser realizado através dos métodos: Kohonen em Anel, K-Médias e Fuzzy C-Médias.

Para iniciar o agrupamento deve-se selecionar o município, o ano base das curvas geradas (cenário), o número de grupos e preencher os dados solicitados pelo método escolhido. No método Kohonen em Anel deve ser informado o número de iterações, a taxa de aprendizado e o raio de vizinhança. No K-Médias informa-se o critério de parada e no Fuzzy C-Médias além do critério de parada informa-se também o índice de fuzzificação.

Para todos os métodos, é dada a possibilidade de mudança dos dados de entrada pela desativação das classes de consumo, ou seja, o agrupamento pode ser feito levando-se em consideração apenas as classes escolhidas pelo planejador. Isso acarreta uma mudança nos grupos gerados e aumenta a flexibilidade do sistema.

Os valores aplicados para o treinamento não são padrões, sendo necessário a sensibilidade do planejador para a obtenção de um bom conjunto de valores por município. Esta sensibilidade é obtida ao se analisarem as classes geradas por vários treinamentos diferentes.

3.1.4 Visualização das classes

Terminado o treinamento e agrupamento das curvas de carga, podem-se visualizar as classes geradas. Primeiramente, o planejador deve escolher o ano-base do treinamento e o tipo de obtenção dos clientes, se por transformador ou por ponto de entrega.

Após a escolha dos parâmetros a visualização é montada e disponibilizada ao planejador (Figura 3). As quadrículas de mesma classe são coloridas da mesma cor, proporcionando uma visão clara das áreas com curvas de carga semelhante.

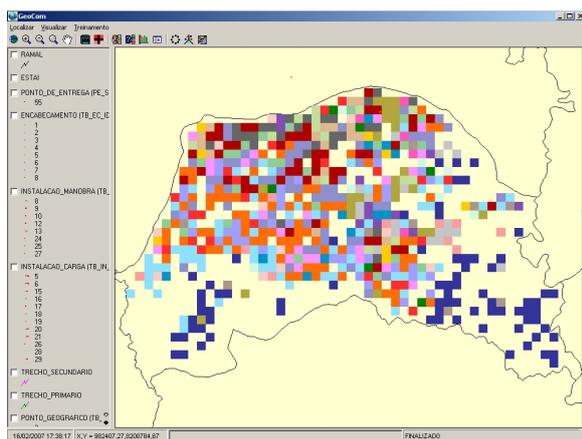


Figura 3: Visualização das classes geradas - Aparecida de Goiânia

O sistema desenvolvido fornece ainda gráficos detalhados das quadrículas como forma de apoio à análise da classificação efetuada.

4 Estudo de Caso - Anápolis

Anápolis é considerada a capital industrial de Goiás com a criação do Distrito Agroindustrial de Anápolis - DAIA. É o terceiro maior município do estado em população e o segundo no ranking de competitividade e desenvolvimento, e compõe a região mais desenvolvida do Centro-Oeste.

4.1 Agrupamento e Classificação

Para a obtenção de resultados satisfatórios, é importante a determinação de alguns parâmetros tais como, tipo de normalização, conjunto de entrada e número de saídas, pois estes influenciam diretamente no resultado final.

Por exemplo, alguns métodos trabalham melhor com valores negativos do que com zeros, sendo assim, optar-se pela normalização no intervalo $[-1,1]$ ao invés de $[0,1]$ é uma escolha melhor. Já o conjunto de entrada pode ser filtrado excluindo-se curvas com amplitudes suspeitas, oriundas de problemas cadastrais. Por fim, a definição do número de saídas influencia diretamente na distribuição das curvas pelos grupos dada a ne-

cessidade de generalização com a diminuição do número de saídas.

Para o processo de agrupamento efetuado neste trabalho foi utilizado um intervalo de normalização de $[-1,1]$, um número de saídas igual a 20 grupos e o conjunto de entrada abrangeu todas as curvas do município, não sendo ignoradas as curvas com amplitudes demasiadamente elevadas.

Entretanto, a obtenção dos parâmetros adequados é um processo empírico, pois depende do nível de experiência do planejador e do número de testes realizados.

A Figura 4, apresenta o resultado do processo de classificação proveniente do método Fuzzy C-Médias no município de Anápolis para o cenário de 2006.

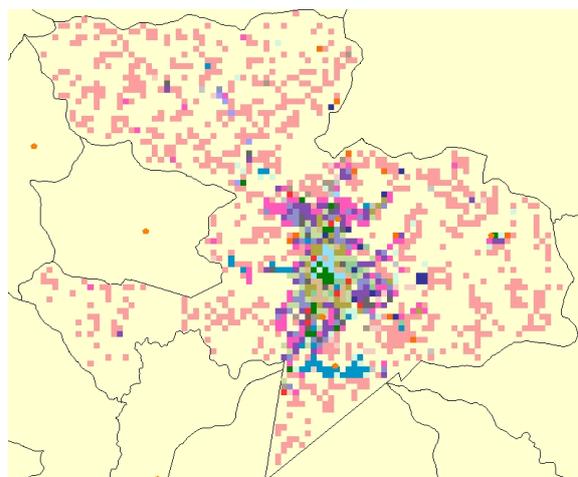


Figura 4: Tela do resultado da classificação pelo método Fuzzy C-Médias

As quadrículas de cor laranja, por exemplo, representam áreas de perfil de consumo predominantemente rural. Todas as quadrículas e grupos gerados podem ser analisados através de gráficos do sistema GeoCom.

4.2 Medidas do agrupamento

Nesta seção, serão apresentados os resultados dos índices de medição do agrupamento e classificação para o município de Anápolis.

Primeiramente, na Tabela 1 é mostrada a relação do tempo de processamento das classificações efetuadas no município de Anápolis.

O método K-Médias obteve um tempo computacional menor na execução da classificação do que o método Fuzzy C-Médias, pois seu processo de atualização dos centróides dos grupos utiliza a média aritmética simples, enquanto o Fuzzy C-Médias utiliza a média aritmética ponderada.

Estes tempos foram inferiores ao tempo gasto pelo método Kohonen em Anel, o qual possui um número de iterações pré-definido pelo planejador. O método Kohonen executou 6000 iterações, sendo destas 1000 para o ajuste fino da rede.

Tabela 1: Tempo de processamento dos métodos

Método	Tempo
Kohonen Anel	1 h 56 min 04 seg
K-Médias	1 min 38 seg
Fuzzy C-Médias	4 min 43 seg

A Tabela 2 mostra o resultado de medição dos Índices Intra-Grupo (IAG), o qual indica o quanto os padrões de entrada que estão em um mesmo grupo são similares.

O método que obteve o melhor resultado foi o método Fuzzy C-Médias seguido de perto pelo método K-Médias. Isso está relacionado à boa distribuição das quadrículas pelos grupos. O método Kohonen em Anel obteve o menor resultado, pois sua distribuição das curvas no agrupamento não seguiu os demais métodos. Isso ocorreu devido a ação do vetor de pesos, o qual gerou grupos que ficaram sem curvas, aumentando assim a generalização dos grupos formados.

Tabela 2: Resultados dos Índices Intra-Grupo (IAG)

Método	IAG
Kohonen Anel	565
K-Médias	465
Fuzzy C-Médias	426

Por fim, a Tabela 3 apresenta os resultados do Índice Inter-Grupo (IEG), o qual mede o quão distantes os grupos estão de uma curva referencial. Neste índice o método Fuzzy C-Médias também obteve o menor valor.

Tabela 3: Resultados dos Índices Inter-Grupo (IEG)

Método	IEG
Kohonen Anel	819
K-Médias	795
Fuzzy	790

5 Conclusões

O método Fuzzy C-Médias obteve um melhor resultado no estudo de caso, formando grupos com um número de quadrículas mais distribuídos que os demais métodos. Além disso, este método, por usar um grau de pertinência para indicar o quanto um centróide representa uma determinada entrada, permite extrair conhecimento além do agrupamento produzido.

O método K-Médias obteve valores próximos ao Fuzzy C-Médias com um tempo de processamento muito inferior e com o mesmo critério de parada. Isso se deve ao seu simples processo de atualização dos centróides, o qual utiliza da média aritmética simples.

O método Kohonen em Anel apresentou os resultados menos satisfatórios, entretanto a aplicação de novos parâmetros de inicialização pode vir a apresentar resultados melhores, dado o caráter empírico do processo de agrupamento. Além disso, este método, devido à ação de seu vetor de atualização de pesos, gerou grupos com centróides próximos, possibilitando identificar clientes com consumo ligeiramente diferente.

Em todos os métodos foi utilizada a distância Euclidiana para medir a distância entre os padrões do conjunto de treinamento e os centróides dos grupos. Assim o que diferencia cada método é a forma como os centróides são atualizados, produzindo com isso informações diferentes como produto da classificação.

Agradecimentos

À Companhia Energética de Goiás (CELG) por fornecer as informações e dados necessários para a realização deste trabalho.

Referências

- Batista, L. B. (2004). *Aprendizagem e recuperação de imagens utilizando mapas auto-organizáveis e representação log-polar*, Mestrado, Universidade Federal de Campina Grande, Campina Grande, Brasil.
- Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms.*, Plenum Press.
- Carvalho, A. C. P. L. F., Braga, A. P. and Ludermir, T. B. (1998). *Fundamentos de Redes Neurais Artificiais*, RIO.
- Casanova, M. A., Câmara, G., Davis, C. A. J., Vinhas, L. and Queiroz, G. R. (2005). *Banco de dados Geográficos*, MundoGEO.
- Drummond, I. N. (2003). Implementação do método de classificação contínua fuzzy k-médias no ambiente terralib, *Technical report*, INPE.
- Haykin, S. (2001). *Redes neurais: princípios e práticas*, Bookman.
- Ross, T. J. (1995). *Fuzzy Logic with Engineering Applications*, McGraw-Hill.
- Santos, R. (2005). Princípios e aplicações de mineração de dados, *Technical report*, INPE.
- Zanon, D. M. and Lima, W. S. (2003). *Mineração de dados aplicada à análise de curvas de carga*, Mestrado, Universidade Federal de Goiás, Goiânia, Brasil.