

Identificação e Prevenção de Perdas Comerciais no Faturamento

Juan G. Lazo Lazo¹, Marley Maria B. R. Vellasco¹, Marco Aurélio C. Pacheco¹, Karla Figueiredo¹, Carlos Hall¹, Ricardo Tanscheit¹, João R. Carrilho Jr.¹, José Eduardo N. da Rocha²

¹ICA: Laboratório de Inteligência Computacional Aplicada

Departamento de Engenharia Elétrica

Pontifícia Universidade Católica do Rio de Janeiro, PUC-Rio

Rua Marquês de S. Vicente 225, Gávea, Rio de Janeiro, CEP 22453-900, RJ Brasil

²Light Serviços de Eletricidade S.A.

Av. Marechal Floriano 168/ 1º Andar Centro Rio de Janeiro, CEP 20080-002, RJ Brasil

juan@ele.puc-rio.br, marley@ele.puc-rio.br, marco@ele.puc-rio.br, karlaf@uerj.br,
hall@ele.puc-rio.br, ricardo@ele.puc-rio.br, joaocarrilho@globo.com,
jose.rocha@light.com.br

Resumo

Este trabalho apresenta o desenvolvimento de uma metodologia capaz de identificar divergências na classificação de clientes da empresa Light S.A., utilizando técnicas de Mineração de Dados (Data Mining) a partir da análise das curvas de consumo, fatores de demanda e de carga das unidades consumidoras (UC) e, ainda, outras informações exógenas presentes na base de dados. A metodologia é baseada em duas etapas: *categorização ou agrupamento*, para agrupar unidades consumidoras com padrões de consumo similares; e *classificação*, para descobrir mudanças no perfil de comportamento dos clientes, configurando a irregularidade na medição de energia elétrica. O agrupamento de registros é uma das tarefas realizadas no processo de Mineração de Dados, onde se deseja identificar registros que possuem características semelhantes, de tal forma a refletir a estrutura de um conjunto de dados em grupos (*clusters*). Dessa forma, foi desenvolvida uma metodologia baseada no algoritmo *Fuzzy C-Means*, (FCM) para agrupamento de dados, nos casos em que se conhece, a priori, o número de *clusters*. O desempenho deste modelo está fortemente correlacionado à escolha de seus parâmetros e aos métodos de interpretação de resultados. Os dados da base de clientes utilizada para teste foi inicialmente pré-processada e normalizada, com intenção de aumentar a precisão do método. Os resultados obtidos com modelagem utilizando técnica de inteligência

computacional foi considerado muito bom por especialistas responsáveis pela área de análise de faturamento de grandes clientes na empresa Light S.A.

1. Introdução

O cenário globalizado da economia mundial exige que as empresas atinjam padrões crescentes de excelência. Estes novos paradigmas envolvem o binômio eficiência (baixo custo de produção) e eficácia (produto com alta confiabilidade). Neste sentido, as concessionárias buscam reduzir as perdas técnicas e comerciais de forma a maximizar a energia disponível para comercialização. Outra questão importante é a influência do índice de perdas na composição da tarifa de negociação para o cliente do mercado livre. Quanto menor for este índice, mais competitivo será o valor final da tarifa.

Para conseguir a redução das perdas comerciais e administrativas nas unidades consumidoras (UC) se faz necessário um sistema que identifique rapidamente quais unidades consumidoras apresentam irregularidade na medição.

A este artigo descreve o desenvolvimento de um sistema inteligente para a identificação das perdas comerciais, visando dar suporte à decisão no direcionamento dos investimentos de recuperação.

O sistema inteligente integra uma metodologia que possui duas etapas: categorização (agrupamento) de unidades consumidoras em grupos de mesmo perfil de

consumo e de demanda de energia elétrica, e classificação para identificar quais unidades consumidoras mudaram de perfil de consumo e de demanda a cada nova medição realizada pela empresa. Esta metodologia considerou as características históricas e sazonais de cada unidade consumidora.

A seção 2 deste artigo apresenta o problema de detecção de irregularidades. A seção 3 descreve a metodologia proposta, a seção 4 descreve os casos de estudo considerados e por último, na seção 5, são apresentadas as conclusões do trabalho.

2. Problema da detecção de irregularidades

A recuperação das perdas comerciais e administrativas nas unidades consumidoras só é possível a partir da rápida identificação das irregularidades, como a divergência na classificação das atividades dos clientes, na medição e variações excessivas nos consumos e demandas. Outra dificuldade no combate às perdas é o pouco conhecimento do comportamento da carga e demanda das categorias inseridas nas várias classes de atividades econômicas. Este levantamento é, em geral, oneroso e de difícil manutenção, sendo influenciado pelas permanentes modificações nas políticas financeiras e econômicas nacionais, as quais influenciam nos perfis de consumo destes clientes.

Estas irregularidades encontram-se registradas e ao mesmo tempo ocultas no enorme volume de dados históricos dos clientes, depositados na base de dados da empresa Light S.A.. Para extrair da base de dados a identificação das unidades consumidoras que apresentam irregularidades, tem-se utilizado técnicas computacionais clássicas de programação matemática e estatística, comumente aplicadas na análise da variação percentual de consumo, todavia com resultados pouco precisos. As técnicas de mineração de dados da inteligência computacional apresentam-se como uma alternativa viável para realizar uma análise abrangente de todas as classes econômicas da base de dados. Este trabalho propõe a categorização natural do espaço amostral de unidades consumidoras pelo algoritmo FCM [1], em função das características elétricas de demanda e consumo destas unidades nos meses de janeiro a dezembro, independentemente de suas atividades econômicas e considerando o comportamento sazonal do consumo e demanda.

3. Metodologia Proposta

A metodologia proposta foi concebida de forma que o modelo gerado fosse imune à sazonalidade e a

tendência de crescimento no perfil de consumo e demanda de energia elétrica. Este foi um fator fundamental para se alcançar bons resultados neste projeto.

Inicialmente, é feita a clusterização da base de dados históricos de faturamento dos produtos contratados por consumidores de energia elétrica utilizando o método Fuzzy C-Means. Em seguida, a classificação das UC segundo estes grupos a cada mês.

A categorização (*clustering*) é uma das técnicas de mineração de dados (*data mining*) para descobrir padrões ocultos em uma base de dados.

A técnica de categorização (*clustering*) particiona um conjunto de padrões de entrada em grupos (*clusters*) homogêneos [2], de tal forma que os dados de um mesmo *cluster* são mais similares entre si que os dados encontrados em diferentes *clusters* [3].

A similaridade destes pontos é definida de acordo com critérios pré-estabelecidos. Por exemplo, considerando os registros de um banco de dados de uma empresa que contém os itens comprados pelos seus clientes, o procedimento de categorização poderá agrupar clientes de forma que aqueles com padrões similares de compras estejam no mesmo *cluster*. Logo, o principal objetivo do processo de categorização é revelar a organização dos padrões em grupos que nos permitam descobrir similaridades e diferenças, bem como permitir extrair conclusões sobre esta distribuição.

No processo de categorização não há classes pré-definidas, nem exemplos que possam mostrar que tipos de relações desejáveis devem estar válidos entre os dados. Por esta característica o algoritmo utilizado neste processo é dito não supervisionado [4].

A categorização produz categorias iniciais nas quais os valores dos padrões encontrados nos dados são classificados. Este processo pode resultar em diferentes particionamentos dos dados dependendo dos critérios específicos utilizados. Logo, existe a necessidade de pré-processamento antes de se iniciar a tarefa de categorização. A seção 3.1 a seguir descreve o pré-processamento.

3.1. Pré-processamento dos dados

A base de dados fornecida pela Light num arquivo texto, apresenta a organização dos dados como na Tabela 1.

Esta base de dados contém 27 produtos diferentes que podem ser contratados pelas UC segundo suas necessidades. A Tabela 2 mostra a relação dos produtos que uma unidade consumidora pode contratar.

	Campo	Descrição do Campo
1	Lote	Lote de Faturação
2	UC	Código da Unidade Consumidora da Light
3	Nome do cliente	
4	Mês da Fatura	Formado "01-MON-YYYY"
5	Tipo de Fornecimento	Energia e Energia Reativa
6	Produto	Código de Especificação
7	Carga Declarada	Carga Declarada em KW
8	Quantidade Medida	Carga Medida em KW
9	Quantidade Faturada	Carga Faturada em KW
10	Código Classe de Consumo	Classe a Qual Pertence a Unidade Consumidora
11	Classe de Consumo	Classe de Consumo a qual o cliente está cadastrado com a Light
12	Código Mensagem de Leitura	Código de erro retornado pelo leitor do medidor

Tabela 1 - Organização da Base de Dados

	Código de Especificação (Produtos)	
1	ADP	DEMANDA ADICIONAL PONTA
2	ADF	DEMANDA ADICIONAL FORA PONTA
3	ADT	DEMANDA ADICIONAL GERAL
4	CEF	CONSUMO EMERGENCIAL FORA PONTA
5	CEP	CONSUMO EMERGENCIAL PONTA
6	CFP	CONSUMO FORA PONTA PERÍODO SECO
7	CFU	CONSUMO FORA PONTA PERÍODO ÚMIDO
8	CNF	CONSUMO FORA PONTA
9	CNP	CONSUMO PONTA
10	CON	CONSUMO GERAL
11	COP	CONSUMO PONTA PERÍODO SECO
12	CPU	CONSUMO PONTA PERÍODO ÚMIDO
13	DEM	DEMANDA GERAL
14	DEP	DEMANDA PONTA
15	DFP	DEMANDA FORA PONTA
16	DMC	DEMANDA REATIVA GERAL
17	DMF	DEMANDA REATIVA FORA PONTA
18	DMP	DEMANDA REATIVA PONTA
19	EPL	ENERGIA PLUS
20	UFE	EXCEDENTE DE CONSUMO REATIVO GERAL
21	UFP	DEMANDA DE ULTRAPASSAGEM FORA PONTA
22	UFS	EXCEDENTE DE CONSUMO REATIVO FORA PONTA PERÍODO SECO
23	UFU	EXCEDENTE DE CONSUMO REATIVO FORA PONTA PERÍODO ÚMIDO
24	ULP	DEMANDA ULTRAPASSAGEM PONTA
25	UPS	EXCEDENTE DE CONSUMO REATIVO PONTA PERÍODO SECO
26	UPU	EXCEDENTE DE CONSUMO REATIVO PONTA PERÍODO ÚMIDO
27	UTT	DEMANDA ULTRAPASSAGEM GERAL

Tabela 2 - Relação de Produtos

Numa primeira etapa foi realizada a transformação da base de dados original em uma base de dados que possua integridade relacional. Esta transformação foi feita a partir da análise da base de dados retirando-se os dados repetidos, ambíguos, *outliers* e variáveis com informações redundantes.

Esta transformação é realizada a partir da definição de que cada registro da base de dados deve conter os dados da fatura mensal de uma unidade consumidora. A nova base de dados foi criada usando unicamente os valores faturados de cada produto contratado pelas UCs.

Como resultado deste processo obteve-se uma base de dados capaz de ser tratada pelo processo de clusterização.

A partir da base de dados limpa foi criada a matriz para a clusterização, contendo unicamente os atributos numéricos. Esta base de dados foi dividida em duas partes: a primeira correspondente aos clientes que apresentam produtos de consumo e demanda geral; e a segunda referente aos clientes que apresentam produtos de consumo e demanda ponta e fora ponta. Foram criados dois tipos de matrizes para cada lote de faturamento de cada mês. Com o objetivo de reduzir o número de atributos e agrupar valores relacionados, foram agrupados (somados) os produtos faturados que representam uma mesma categoria de consumo. Estas matrizes foram normalizadas e foi adicionado um atributo proveniente do cálculo do fator de carga.

As matrizes de saída apresentam os seguintes atributos mostrados na Tabela 3 e Tabela 4.

	ATRIBUTOS
1	UC (Unidade Consumidora)
2	ADT
3	CON
4	DEM
5	DMC
6	EPL
7	UFE
8	UTT
9	Fator de Carga Geral

Tabela 3 - Matriz com Somatório Geral

	ATRIBUTOS
1	UC (Unidade Consumidora)
2	ADP + ADF
3	CEP + CEF
4	COP + CPU + CFP + CFU
5	Max(DEP, DFP)
6	Max(DMP, DMF)
7	EPL
8	UPS + UPU + UFS + UFU
9	Max(ULP, UFP)
10	Fator de Carga na Ponta
11	Fator de Carga Fora da Ponta

Tabela 4 - Matriz com Somatório Ponta – Fora Ponta

3.2. Clusterização por Fuzzy C Means

A análise de *cluster* é uma técnica exploratória de dados que tem por objetivo formar agrupamentos de objetos semelhantes em um banco de dados. O

conceito de *clusterização* difere do conceito de classificação no sentido de que a análise de *cluster* é uma técnica mais “primitiva”, na qual nenhuma suposição é feita a respeito dos grupos, assim como o seu número e estrutura (Jonhson, 1998). Os *clusters* são obtidos por intermédio da aplicação dos conceitos de similaridade e de distância.

O algoritmo Fuzzy C-means (FCM), originalmente introduzido por J. Bezdek em 1981, é a versão fuzzy do método rígido *k-means*, sendo empregado para classificar um universo de amostras em *categorias fuzzy* de acordo com a sua disposição no *Espaço Euclidiano*.

Em agrupamento não nebuloso, o limite de diferentes *clusters* é *crisp*, ou seja, cada padrão pertence somente a um único grupo. Em agrupamento nebuloso, cada padrão não pertence somente a um único grupo, mas sim a vários grupos com diferentes graus de pertinência para cada grupo.

O algoritmo *Fuzzy c-Means* (FCM) é uma variante do algoritmo *Fuzzy Clustering* iniciado por Bezdek no final dos anos 70 [6]. O algoritmo tenta agrupar medidas através da procura do mínimo local da generalização da soma dos quadrados dos erros das funções.

O problema de agrupamento nebuloso pode ser definido da seguinte forma [6], [7], e [8]:

Seja $X = \{x_1, x_2, \dots, x_n\}$ um conjunto de dados finito composto de elementos do espaço Euclidiano dimensional R^p tal que, $x_j \in R^p, j=1, 2, \dots, n$. Deseja-se particionar os elementos em c conjuntos nebulosos de forma a otimizar uma certa função objetivo, onde c é o número de *clusters*.

O resultado do agrupamento nebuloso pode ser expressado através da seguinte matriz U:

$$U = [u_{ij}], \quad i=1, \dots, c; \quad j=1, \dots, n. \quad (1)$$

onde u_{ij} é um valor entre [0,1] que indica o grau de pertinência de cada elemento x_j para um determinado i -ésimo *cluster*.

Trivedi e Bezdek desenvolveram a seguinte função objetivo para o algoritmo FCM:

$$J(u_{ij}, v_k) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2, \quad 1 < m < \infty. \quad (2)$$

onde m é o coeficiente nebuloso responsável pelo grau de “fuzzificação” dos elementos da matriz U e v_k é o centróide do k -ésimo *cluster*.

Quanto maior é o coeficiente nebuloso m , mais nebuloso se torna a matriz U. Quando $m=1$ a função objetivo $J(u_{ij}, v_k)$ é reduzida ao caso *crisp* do algoritmo de agrupamento *k-means*.

O objetivo é minimizar $J(u_{ij}, v_k)$. Para tanto, basta diferenciar a função objetivo dada pela equação (2) em relação a v_i (fixando $u_{ij}; i=1, \dots, c; j=1, \dots, n$) e em relação a u_{ij} (fixando $v_i; i=1, \dots, c$), de onde obtém-se:

$$v_i = \frac{1}{\sum_{j=1}^n (u_{ij})^m} \sum_{j=1}^n (u_{ij})^m x_j, \quad i=1, 2, \dots, c. \quad (3)$$

$$u_{ij} = \frac{\left(\frac{1}{\|x_j - v_i\|^2} \right)^{1/m-1}}{\sum_{k=1}^c \left(\frac{1}{\|x_j - v_k\|^2} \right)^{1/m-1}} \quad (4)$$

$$i=1, 2, \dots, c;$$

$$j=1, 2, \dots, n..$$

Algoritmo FCM:

1. Escolher o número de *clusters* c ($2 \leq c \leq n$), o coeficiente *fuzzy* m ($1 < m < \infty$), o critério de parada ϵ , e a matriz inicial $U^{(0)}$. Colocar o índice $l=0$.
2. Calcular o centro dos *clusters* $\{v_i^{(l)} \mid i=1, 2, \dots, c\}$
3. Calcular a nova matriz $U^{(l+1)}$
4. Calcular $\Delta = \| U^{(l+1)} - U^{(l)} \| = \max_{i,j} |u_{ij}^{(l+1)} - u_{ij}^{(l)}|$ se $\Delta > \epsilon$, então $l=l+1$ e volta ao passo 2. Se $\Delta \leq \epsilon$ fim.

3.3. Modelo de Classificação

Após a realização da clusterização, tem-se os grupos de UC para cada mês e lote de faturação para o ano anterior ao de referência (neste caso é 2003, este ano será usado para futuras referências). A classificação dos clientes do ano 2004 (ano de referência, ou seja, ano em que se deseja verificar a possibilidade da existência de irregularidades) para detecção de irregularidades é realizada a partir da

seleção dos clientes de 2004 definidos pelos grupos formados pela clusterização para o ano 2003. Considerou-se como UC válidas no grupo, aquelas que constam nos dois anos (2003 e 2004). Dessa forma, são arranjados os grupos com as mesmas UC no 2003 e 2004. As UC do 2004 que não pertencem a nenhum grupo são consideradas clientes novos.

A partir destes grupos calcula-se a matriz de distâncias das UC para cada grupo em cada ano. Esta matriz de distâncias é simétrica, tem dimensão $N \times N$ e possui diagonal igual a zero, onde N é o número de UC.

A seguir descreve-se o processo de cálculo das matrizes do algoritmo de classificação.

Seja a matriz $C03_g$ correspondente a um grupo de um determinado mês e lote do ano 2003, onde g é o número de grupo com N clientes (UC) e M atributos, assim $C03_{gi}$ é o vetor de atributos de uma UC i , onde $i = 1, 2, \dots, N$ e $C03_{gij}$ é um atributo, onde $j = 1, 2, \dots, M$.

Define-se a distancia d_{ik} como a distância Euclidiana de dois UC $C03_{gi}$ e $C03_{gk}$ com $k = 1, 2, \dots, N$ como:

$$d_{ik} = \sqrt{(C03_{gi1} - C03_{gk1})^2 + \dots + (C03_{gij} - C03_{gkj})^2} \quad (5)$$

onde se cumpre que: $d_{ii} = 0$ e $d_{ij} = d_{ji}$.

Então é possível construir uma matriz de distâncias $A03_g$ como:

$$A03_g = \begin{bmatrix} d_{11} & \dots & d_{1N} \\ \vdots & \ddots & \vdots \\ d_{N1} & \dots & d_{NN} \end{bmatrix}$$

Este mesmo processo é realizado para os grupo formados no ano 2004. Assim, define-se a matriz de distâncias para os grupos do ano 2004 como a matriz $A04_g$.

A seguir, calcula-se o vetor das médias das distâncias das UC denotado por $d03_{gk}$ e $d04_{gk}$ para os anos 2003 e 2004 respectivamente, onde:

$$d03_{gk} = \frac{\sum_{i=1}^N A03_{gik}}{N} \text{ e } d04_{gk} = \frac{\sum_{i=1}^N A04_{gik}}{N} \quad (6)$$

onde $i = 1, 2, \dots, N$ e $k = 1, 2, \dots, N$

O cálculo da média do vetor de médias das distâncias das UC para ambos anos, $D03_g$ e $D04_g$ é definido por:

$$D03_g = \frac{\sum_{k=1}^N d03_{gk}}{N} \quad (7)$$

$$D04_g = \frac{\sum_{k=1}^N d04_{gk}}{N} \quad (8)$$

Com eles é calculada a média de referência de ambos anos, M_{0304} :

$$M_{0304} = \frac{D03_g + D04_g}{2} \quad (9)$$

A seguir calcula-se a matriz de variação das UC, MV_g de cada grupo, que é simétrica de $N \times N$ elementos, onde cada elemento é calculado da seguinte forma:

$$MV_{gik} = |A03_{gik} - A04_{gik}| \quad (10)$$

A seguir é calculada a média da variação de cada UC e normalizada com a média de referência de ambos anos, que será denominada como média normalizada de variação, μ_{gk} :

$$\mu_{gk} = \left[\frac{\sum_{i=1}^N MV_{gik}}{N} \right] * \frac{1}{M_{0304}} \quad (11)$$

Este valor indica a variação de cada UC com respeito ao grupo, sendo que a UC que tenha maior μ_{gk} apresenta maior mudança no seu perfil de consumo e demanda.

Este processo é repetido para todos os grupos, sendo a resposta final a união de todos os μ_{gk} ordenados de forma decrescente; ficando assim no topo da lista as UC que apresentaram maior mudança no perfil.

4. Estudo de Caso

Esta seção apresenta os teste e resultados obtidos com a metodologia proposta.

Com o histórico dos dados fornecidos pela Light, foi treinado o sistema definindo os grupos de unidades consumidoras com características similares de consumo e demanda, para cada mês, considerando o lote de faturação.

Com base nestes grupos, foram classificados os dados correspondentes ao ano 2004, mês a mês, considerando os lotes de faturamento.

Para avaliar a eficiência do modelo, o resultado da classificação foi comparado com uma lista de clientes com irregularidade comprovada fornecido pela Light.

Neste artigo são apresentados os resultados da classificação para um mês (Julho) escolhido aleatoriamente.

A descrição dos atributos da Tabela 6 de resultados é apresentada na Tabela 5.

UC	Unidade Consumidora
ADT	Demanda Adicional Geral
CON	Consumo Geral
DEM	Demanda Geral
DMC	Demanda Reativa Geral
EPL	Energia Plus
UFE	Excedente de Consumo Reativo Geral
UTT	Demanda Ultrapassagem Geral
FC	Fator de Carga Geral

Tabela 5 - Códigos de descrição das colunas da tabela de resultado

As linhas da tabela de resultados (Tabela 6) estão separadas em pares que indicam as características de esse cliente para o mês em questão, para o ano anterior e para o ano em análise (anos 2003 e 2004).

	U	AD	CO	DE	DM	EP	UF	UT	F
1	296	0	13446	54.9	0	0	324	0	0.3355
	296	0	15044	0	0	0	216	78.2	0.2635
2	8443	0	10519	100	0	0	4903	0	0.1441
	8443	0	15142	100	18.4	0	6728	11.5	0.1860
3	618	0	8360	150	0	0	2320	0	0.0763
	618	0	43656	150	9.3	0	3672	53.2	0.2943
4	7703	0	89435	183.2	3.9	0	5800	0	0.6687
	7703	0	63256	178.8	0	0	875	0	0.4846
5	642	0	57832	150	0	0	64	0	0.5281
	642	0	144712	230	0	0	0	0	0.8619
6	8660	0	9418	65	0	0	1555	0	0.1985
	8660	0	799	2.6	5.3	0	3218	0	0.4210
7	7682	0	4464	80	0	0	1037	0	0.0764
	7682	0	22982	80	5.2	0	3082	26	0.2970
8	1702	0	37336	120	0	0	0	0	0.4262
	1702	0	0	120	0	0	0	0	0.0000
9	237	0	11588	165	0	0	2052	0	0.0962
	237	0	13662	115	0	0	5972	0	0.1627
10	7660	0	27216	110	0	0	1296	0	0.3389
	7660	0	43416	110	0	0	461	45.5	0.3825
11	387	0	105008	207	0	0	3812	55.2	0.5486
	387	0	95105	290	0	0	2722	0	0.4492
12	1767	0	1631	66.5	0	0	13446	0	0.0336
	1767	0	35294	160	0	0	4007	0	0.3022
13	995	0	9947	0	12	0	3856	41.5	0.3283
	995	0	8748	0	5.1	0	2549	38	0.3154
14	3668	0	5244	0	3.5	0	2388	13.9	0.5168
	3668	0	20436	0	12.4	0	9744	118.6	0.2360
15	5078	0	19991	120	0	0	1955	0	0.2282
	5078	0	1350	9.1	0.7	0	734	0	0.2032
16	253	0	12125	75	0	0	893	0	0.2215
	253	0	8755	26.5	0	0	0	0	0.4526
17	2946	0	32227	185.8	3.6	0	5789	0	0.2376
	2946	0	4104	180	0	0	4687	0	0.0312
18	210	0	39506	240	0	0	1210	0	0.2255
	210	0	63558	252.3	0	0	2938	0	0.3451
19	3949	0	13608	0	0	0	497	44.9	0.4152
	3949	0	15304	105	0	0	108	0	0.1997
20	1898	0	334	0	0	0	0	1.2	0.3813
	1898	0	23	0	0	0	0	0.5	0.0630
21	1703	0	9288	50	0	0	3100	0	0.2545
	1703	0	8186	50	0	0	1426	0	0.2243
22	1201	0	23069	110	10.1	0	2225	23.1	0.2374
	1201	0	14623	110	28.5	0	6307	0	0.1821

Tabela 6 – Resultado da Classificação para o Lote 1 do mês de Julho

Na Tabela 6 estão ressaltadas em amarelo e sublinhadas as variações mais significativas para cada produto de consumo ou demanda. Os clientes que apresentam o código de unidade consumidora (UC) em fundo cinza e com os números em negrito, correspondem aos clientes com irregularidade comprovada fornecidos pela Light. Na tabela as unidades consumidoras encontram-se listadas em ordem da maior para a menor probabilidade de irregularidade.

Observa-se como a metodologia proposta conseguiu detectar tanto as irregularidades comprovadas, como também outras irregularidades, devidamente reconhecidas por especialistas da Light, que dificilmente seriam detectadas por simples observação.

5. Conclusões

Este artigo apresentou o desenvolvimento de uma metodologia para detecção de irregularidades no consumo e demanda de energia elétrica.

Os resultados preliminares obtidos com o sistema evidenciaram a capacidade do sistema em detectar clientes com diversos tipos de irregularidades, o que foi comprovado e confirmado pelos especialistas da Light.

É importante ressaltar que com o uso desta metodologia pode-se incorporar ao modelo não apenas a tendência de crescimento dos padrões de consumo e demanda, como também a sazonalidade inerente a alguns tipos de atividades econômicas de determinados clientes.

Pretende-se no futuro próximo estender a metodologia para a baixa tensão que apresenta o desafio da grande quantidade de clientes.

Referências Bibliográficas

[1] Kohonen, T.: Self-Organizing Maps. 2nd Edition. Springer-Verlag Berlin Heidelberg New York (1997)

[2] Bezdec, J.; Pal, S. K. Fuzzy Models for Pattern Recognition. New Jersey: IEEE Press, 1991.

[3] Guha, S.; Rastogi, R.; Shim K. Cure: An Efficient Clustering Algorithm for Large Databases. In: ACM SIGMOD International Conference on Management of Data, Seattle. 1998.

[4] Michael, J. A.; Berry, G. L. Data Mining Techniques For Marketing, Sales and Customer Support. New York: John Willey & Sons, 1996.

[5] Johnson, R.A.; Wichern, D.W. (1998). *Applied Multivariate Statistical Analysis*, Prentice Hall.

[6] Hall, O.L., Ozyurt, B., Bezdek, J.C.: The Case for Genetic Algorithms in Fuzzy Clustering. The Seventh Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. (1998).

[7] Flores-Sintas, A., Cadenas, José M., Martín, F.: Membership Functions in the Fuzzy C-Means Algorithm. Elsevier Science B.V. Fuzzy Sets and Systems 101 (1999) 49-58.

[8] Lin, C.-T., Lee, G.: Neural Fuzzy System: A Neuro-Fuzzy Synergism to Intelligent Systems. Prentice-Hall PTR (1996).