



## XVIII Seminário Nacional de Distribuição de Energia Elétrica

SENDI 2008 - 06 a 10 de outubro

Olinda - Pernambuco - Brasil

### Desenvolvimento de uma Estratégia de Combate a Perda Comercial a Partir de Aspectos Socioculturais das Comunidades de Baixa Renda

<b>Márcia Coutinho</b>	<b>Marliane de Jesus Mendonça</b>	<b>Landi Hudson da Silva Costa</b>	<b>Henrique de Oliveira Henriques</b>
<b>Light Serviços de Eletricidade S.A</b>	<b>Light Serviços de Eletricidade S.A</b>	<b>UFF</b>	<b>UFF</b>
marcia.coutinho@light.com.br	marliane.mendonca@light.com.br	landihudson@globo.com	rique@vm.uff.br

#### Palavras-chave

Cadastro  
Detecção de Fraude  
Kohonen  
Lógica Fuzzy  
Pesquisa

#### RESUMO

As perdas comerciais em população de baixa renda atingem, no Rio de Janeiro, o valor em torno de 30% das perdas comerciais totais. Essas comunidades, na sua maioria, enfrentam problemas sociais sérios, o que torna difícil o êxito de qualquer ação massiva de combate ao furto e à inadimplência. Esse trabalho apresenta algumas técnicas para auxiliar na estratégia, avaliando os resultados das ações através de simulações antes de investir em soluções de alto custo e retorno duvidoso. Um questionário é feito para uma amostra de clientes das comunidades, dimensionada para um erro de 4%, abordando os principais temas de interesse, como ações de combate mais receptíveis pelos clientes, negociações mais factíveis, etc.. Baseados em dados de cadastro desses clientes amostrados da empresa, associados a outros, que estão disponíveis para todos os clientes da companhia, como é o caso de outros bancos de dados internos ou externos, agrupam-se os clientes em clusters de mesma característica. A técnica usada nesse trabalho é a rede de Kohonen. As respostas aos questionários e os dados cadastrais são conectados através do identificador da instalação. Utilizando-se técnicas de lógica fuzzy, várias simulações podem ser feitas através da escolha de questões chave.

#### 1. INTRODUÇÃO

Vários processos de detecção de fraude têm sido tentados pelas empresas de distribuição de energia elétrica, com resultados ainda modestos face o crescimento das perdas comerciais. Esses métodos, em sua maioria, baseiam-se ou numa correlação entre dados cadastrais de consumidores e seu “status”, normal, com fraude ou inadimplente, ou em variações anormais no consumo desses clientes. Ambos os métodos dependem fortemente do banco de dados de

consumidores das empresas, que em sua formação houve apenas a preocupação natural de se ter informações básicas de identificação do cliente. Portanto, para a solução do problema proposto, dificilmente esses processos tradicionais podem resolver, pois a correlação entre dados cadastrais e o “status” dos clientes é baixa. Este trabalho desenvolve um modelo com o objetivo de nortear as ações de prevenção e combate às perdas e descreve os seguintes processos:

- 1) Criação de clusters de inspeções. Esse módulo tem o objetivo de agrupar inspeções de características próximas, como a descrição do cliente, local e o resultado da inspeção.
- 2) Classificação: Processo de classificação da inspeção candidata, com base na formação dos clusters desenvolvidos no item anterior.
- 3) Determinação do status do cliente: Uma vez selecionado para a inspeção por qualquer método, o objetivo desse módulo é avaliar a possibilidade de se encontrar fraude ou outra irregularidade na inspeção.

## **2. DESENVOLVIMENTO**

### **2.1. Processos Adotados**

Nesse trabalho, são apresentados dois processos de detecção de fraude que foram testados, ambos utilizando como dados de entrada uma estrutura de clusters, gerados por uma Rede de Kohonen, à partir do banco de cadastro. O primeiro processo procurou verificar nas respostas da Pesquisa padrões de clientes normais, fraudulentos e inadimplentes que pudessem ser reconhecidos dentro dos clusters. O segundo processo procurou analisar, utilizando técnicas de lógica fuzzy, características de cada cluster.

São apresentados os dados selecionados, tanto no âmbito do cadastro disponibilizado pela empresa, quanto no âmbito da pesquisa, os quais foram obtidos a partir da mesma. Dos dados disponibilizados apresentam-se também as variáveis que foram utilizadas, bem como o tratamento dado as mesmas.

#### **2.1.1. Aquisição de Dados**

Os dados utilizados para construção do software foram divididos em duas categorias: os dados de entrada, previamente formatados de acordo com as necessidades exigidas no processo adotado, fornecidos pela concessionária LIGHT e os arquivos referentes à Pesquisa, realizada nas comunidades a serem estudadas.

##### **2.1.1.1. Dados de Entrada**

Os dados de entrada, enviados pela LIGHT, contêm inúmeras variáveis das quais as mais relevantes foram selecionadas para serem utilizadas na construção de clusters ou famílias de clientes de mesmo perfil ou características descritivas.

Escolhidas as variáveis a serem utilizadas na “clusterização”, o próximo passo foi o de seccionar as variáveis por faixas, conforme exemplo a seguir:

Campo: Tarifa

Dados:	A4 Água, Esgoto e Saneamento	Faixa 1
	B1 Baixa Renda – código NIS	Faixa 2
	B1 Convencional	Faixa 3
	B3 Água, Esgoto e Saneamento	Faixa 4
	B3 Convencional	Faixa 5
	B3 Convencional – Ligação Direta	Faixa 6

A faixa correspondente ao dado do cliente recebe o valor de “1” e as demais “-1”. Assim cria-se uma seqüência de 1 e -1 que caracteriza o cliente. Considere que no exemplo apresentado o cliente esteja enquadrado na tarifa B3 Convencional, ou seja, faixa 5. Nesse caso, a variável passa a ser:

**-1 -1 -1 -1 1 -1**

O resultado da rede de Kohonen, agrega os clientes de mesma seqüência de 1 e -1 no mesmo cluster. Essa seqüência recebe o nome de DNA e é igual para todos os clientes do mesmo cluster.

### 2.1.1.2. Dados de Pesquisa

Os dados de pesquisa foram obtidos através de Pesquisas realizadas nas áreas de concessão da LIGHT, selecionadas de acordo com as definições adotadas no projeto. Assim como os dados cadastrais, os dados de pesquisa serão formatados através de faixas.

### 2.1.2. Processo de Ligação dos Dados de Entrada com os Dados de Pesquisa

A variável Instalação foi utilizada para que se fizessem as ligações entre os dados de entrada e de pesquisa. Esta é uma das variáveis mais importantes, pois cada unidade consumidora da concessionária possui este número, além de que não permite duplicidade. Isto é importante, pois quando se ligam vários bancos de dados com esta variável não há riscos de se ligarem clientes com dados trocados.

A seguir tem-se um esquema que mostra a relação dos arquivos de entrada e cadastro pelo campo de ligação.

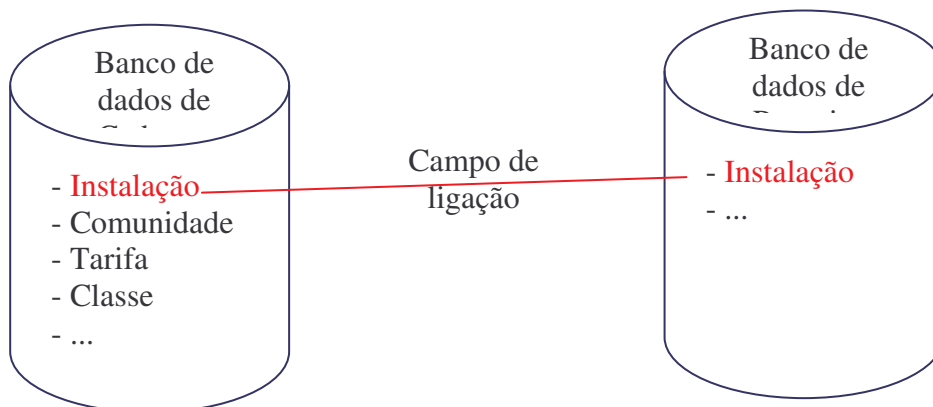


Figura 1 – Elo de Ligação entre os Dados

## 2.2. Processo de Formação dos Clusters

Depois de todo o tratamento realizado com os dados de entrada de cada cliente, a próxima etapa é a classificação destes por rede de Kohonen. Em uma rede desse tipo, os neurônios são colocados em nós de uma grade que é normalmente uni ou bi-dimensional. Os neurônios se tornam seletivamente sintonizados a vários padrões de entrada, estímulos, no decorrer de um processo de aprendizagem. Esse processo é baseado na aprendizagem competitiva, ou seja, os neurônios de saída da grade competem entre si para serem ativados. O neurônio cujo vetor de pesos gerar a menor distância Euclidiana com o vetor de entrada é o vencedor, ou seja, o cliente é classificado nesse neurônio. Todos os clientes classificados no mesmo neurônio pertencem ao mesmo cluster.

A seguir tem-se um esquema da formação dos clusters ou famílias ao se utilizar este tipo de mapa auto-organizável.



Figura 2 – Formação de Clusters

O que se espera é que em um cluster todas as variáveis tenham valores de 1 ou -1 exatamente iguais. Essas famílias terão alguns elementos que fizeram parte da pesquisa de campo, que serão reconhecidos pelo campo de ligação (Instalação). Os clusters que não tiverem dados de pesquisa ou os que possuem em quantidades insuficientes, não poderão participar do processo seguinte de classificação.

### 2.3. Análise dos Dados de Pesquisa

#### 2.3.1. Classificação dos Dados

O processo de classificação dos clientes Normais(N), Inadimplentes(I) e Fraudulentos(F), baseia-se no histórico de pesquisas realizadas anteriormente em outras comunidades com o mesmo objetivo. Através dos campos definidos nos formulários de pesquisa pode-se estabelecer um nível de relevância tal que o cliente analisado possa ser classificado, com base nos dados de históricos de pesquisas.

#### 2.3.2. Preparo dos dados de pesquisa

Para a formação da base de conhecimento usada tanto pelo processo de pesquisa de padrões quanto pelo de análise fuzzy é necessário a realização das seguintes etapas:

- 1) A partir dos arquivos de clientes *N*, *F* e *I* cria-se uma distribuição de frequência para cada uma das variáveis de resposta da pesquisa;
- 2) Em seguida, determina-se o percentual da frequência de cada variável em relação ao total;
- 3) Normaliza-se os dados, já que o número de consumidores *N*, *F* e *I* não são semelhantes;

O resultado desse procedimento é ilustrado na figura 3, tendo como ordenada o número de respostas positivas as questões da pesquisa e como abscissa o número da questão da pesquisa. Em azul estão as respostas dos clientes tipo I, em amarelo as do tipo f e em vermelho as do tipo F.

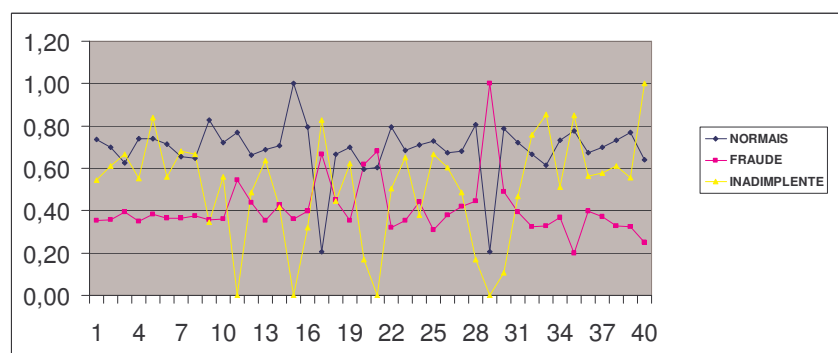


Figura 3 – Frequência de respostas (todos os dados de pesquisa)

Um primeiro passo, que antecede a classificação do cliente em N, F ou I, equivale a fazer um “filtro” nas respostas da pesquisa com o objetivo de diminuir o número de variáveis a serem utilizadas. Ao analisar a frequência relativa das mesmas, percebemos que parte delas não apresenta diferenças nas respostas que possam distinguir clientes Normais, Inadimplentes e Fraudulentos, não interferindo no processo de reconhecimento de padrões.

As variáveis escolhidas nesse processo foram aquelas que teriam a maior representatividade tanto quando cliente normal, fraudulento ou inadimplente. A figura a seguir mostra o nível de corte de 0,8, realçado em cor vermelha, as variáveis mais representativas.

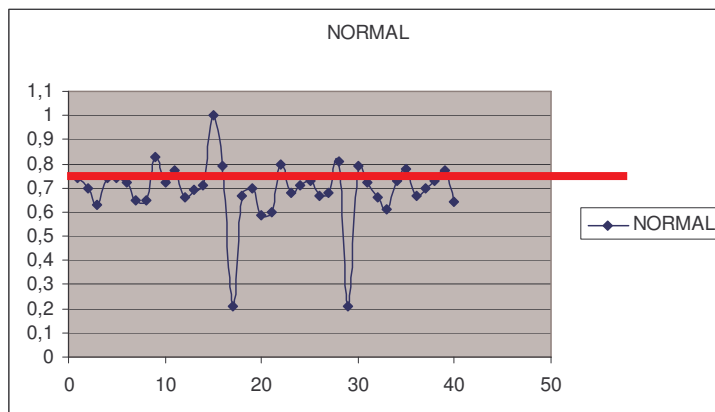


Figura 4 - Seleção de variáveis representativas para clientes normais

O numero de variáveis acima da linha de corte, nível 0,75, selecionadas foram 15. O processo se repete para os clientes fraudulentos e inadimplentes. Para cada DNA, de acordo com o processo, foram levantadas as freqüências de respostas às perguntas selecionadas conforme as faixas.

## 2.4. Metodologia de Detecção de Fraudes e Inadimplência

### 2.4.1 Formação dos Padrões

Através dos dados apresentados anteriormente podemos ver que o resultado de cada cluster, agrupa clientes com a mesma seqüência de 1 e -1 nos dados de entrada, a qual denominamos de DNA. Cada cliente, quando possuir pesquisa, tem o seu DNA ligado as suas respostas correspondentes as perguntas escolhidas, constantes nos dados de pesquisa de clientes Normais, Fraudulentos e Inadimplentes. Cada cluster, dessa maneira, possui um grupo de respostas que pode ou não responder ao padrão da figura 3, apenas com as questões de interesse. A figura abaixo ilustra o processo.

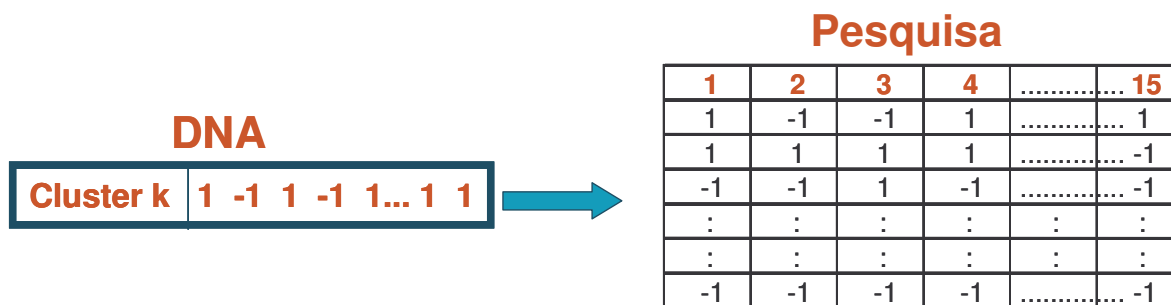


Figura 5 - Respostas dos clientes pertencentes ao cluster k as questões relevantes para clientes N

Nas respostas da pesquisa da figura 4 são computadas, para cada uma das perguntas, a incidência de 1 (um positivo), para cada grupo de N, F e I. Dessa forma, calculando os percentuais e normalizando-se os resultados chega-se a um padrão de resposta do cluster que será comparado com o padrão da figura 3, reduzido às perguntas relevantes. A idéia é a de que o processo funcione como um detector de mentira, ou seja, uma rede neural tipo FF aprende o padrão de respostas da figura 3, com variações aleatórias de mais ou menos um percentual que pode ser, por exemplo, 20%. Cada cluster k tem as suas respostas às perguntas da pesquisa testadas pela rede FF que a classificará em N, F ou I.

### 2.4.2. Processo de Análise Fuzzy

Nesse processo são levantadas as frequências das respostas das variáveis de corte na pesquisa e são construídas as curvas de pertinência para cada “status” dos clientes, através do percentual da resposta positiva em relação ao total de pesquisas disponíveis. O quadro a seguir exemplifica o processo. A curva de pertinência, por cluster da Fraude para uma determinada pergunta:

Pergunta	Cluster 1	Cluster 4	Cluster 8	Cluster 10	Cluster 11	Cluster 15	Cluster 16
8	0	2	6	0	0	4	4
Total de Pesquisa	12	8	67	6	40	8	22
%	0,00	0,25	0,09	0,00	0,00	0,50	0,18

Quadro 1 - Frequência de respostas positivas cada pergunta

A curva de pertinência é dada pela figura a seguir.

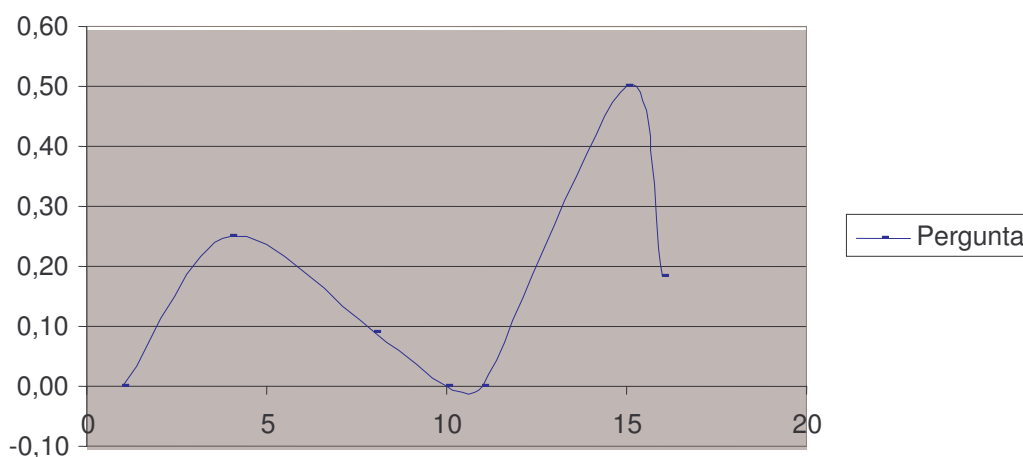


Figura 6 - Curva de pertinência da pergunta exemplo de clientes com fraude

O operador escolhido para definir o perfil dos clientes normais, fraudadores e inadimplentes devido ao nível de corte das perguntas ser alto (0.75), foi o MAX (máximo). Ou seja, o máximo de todas as curvas de pertinência de cada pergunta selecionada. A análise final é feita sobrepondo as curvas Normais, Fraudulentos e Inadimplentes que ressalta o percentual de cada “status” dos clientes nos clusters, possibilitando realizar uma estratégia de se combater o problema.

## 3. RESULTADOS

### 3.1. Processo de Reconhecimento de Padrões

O processo baseia-se na premissa de que o padrão de frequência de respostas as 15 perguntas selecionadas no processo de corte, agregada, pode servir de resposta padrão para todos os clusters de clientes. A figura 7 apresenta esse padrão onde, como exemplo, a pergunta 1 apresentou resultados 0.74 da frequência máxima de respostas positivas que ocorreram na pergunta 6 do padrão. Um comportamento próximo a esse é esperado para os clusters se as respostas dos clientes forem próximas do da Empresa. Para que a curva agregada da Empresa represente o padrão verdade, basta que a diferença entre a resposta padrão e a do cluster seja ligeiramente positiva ou negativa.

Segue exemplo:

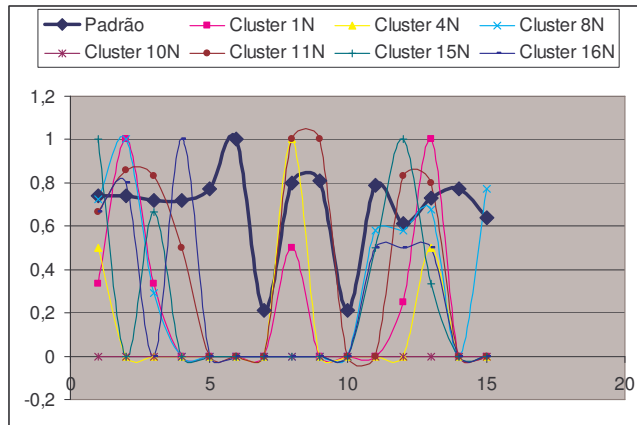


Figura 7. Curvas de frequência de respostas Normais por cluster.

A figura 7 mostra uma variedade muito grande de formas da curva de frequência de respostas normalizadas pela máxima e existe uma diferença grande entre a agregação, considerada como padrão, e os clusters. Esse resultado deve ser analisado com cuidado, pois não demonstra que todos os clusters estão mentindo quanto a sua normalidade. O padrão da empresa, com todas as respostas dos clientes agregadas, não pode servir como referência, pois quanto mais agregadas forem as respostas, mais próximas ficam da curva da Empresa e quanto menor for o cluster mais a forma da curva pode se afastar. Esse resultado é ilustrado claramente na figura a seguir.

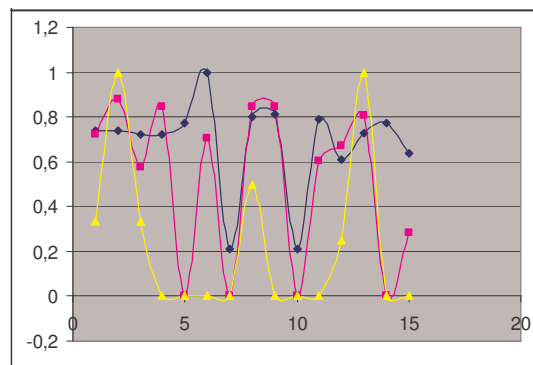


Figura 8 - Diferenças entre padrões Normais

A curva de respostas de uma determinada comunidade, em vermelho, se aproxima mais da curva de resposta da Empresa do que a de um cluster exemplo, em amarelo, sabidamente normal. Se o cluster 10 que contém apenas clientes conhecidamente inadimplentes for comparado com a curva padrão de inadimplência, conforme as curvas de frequências de respostas da figura 9, nta-se que a diferença entre as curvas, faria com que uma rede neural do tipo FeedForward classificasse o cluster 10 erradamente como não inadimplente.

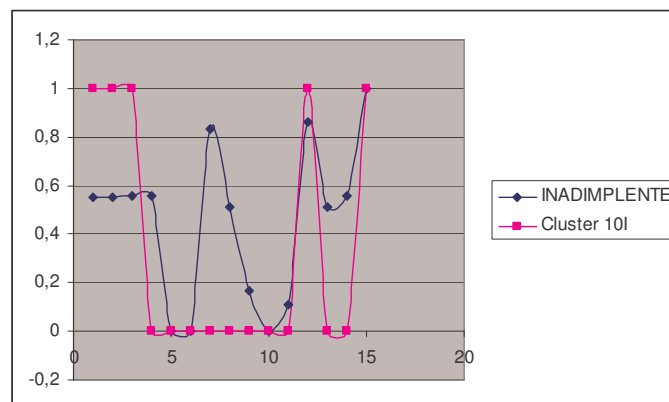


Figura 9 - Curva de respostas do cluster 10 e a curva padrão da Empresa.

Dessa forma, salvo estudos mais aprofundados, desaconselhamos o uso desse processo para detecção de clientes em fraude ou inadimplentes, pois um estudo de padrões de resposta se mostrou complexo e de resultados ainda não aplicáveis.

### 3.2. Processo de Análise Fuzzy

Foi o processo que apresentou os melhores resultados. O primeiro passo cria as curvas de pertinência para as respostas dos clientes N, F e I pelo método descrito anteriormente e determina-se a curva de pertinência máxima, conforme mostrado na figura abaixo.

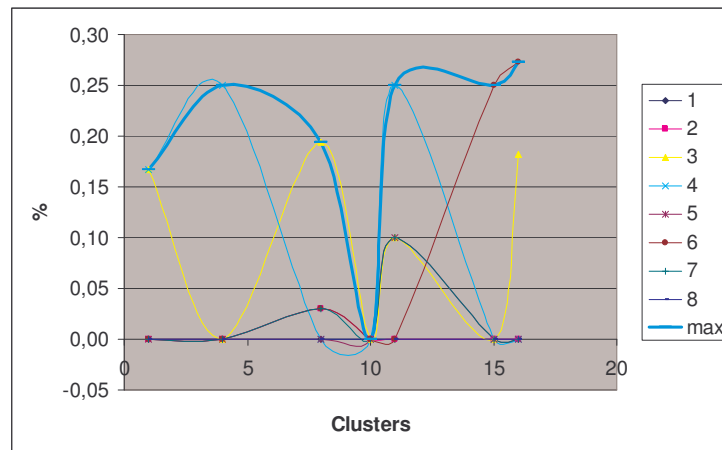


Figura 11 - Curvas de pertinência das respostas de clientes normais as 8 perguntas selecionadas.

O mesmo procedimento é feito para os clientes F e I. Normalizando-se as curvas em relação ao total de respostas positivas entre clientes Normais, Fraudulentos e Inadimplentes tem-se o resultado final na figura abaixo.

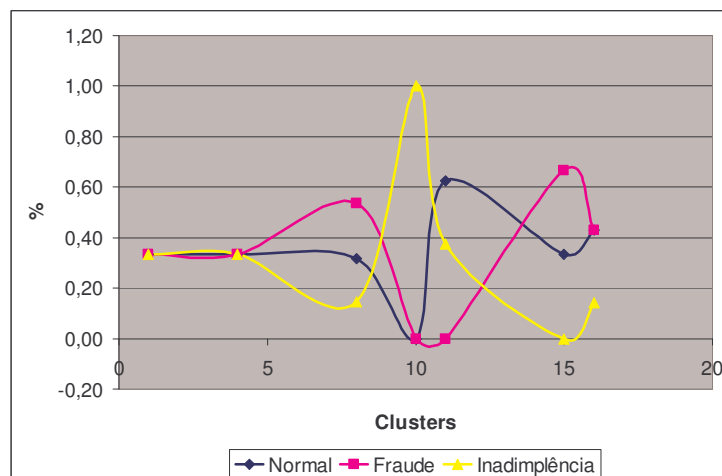


Figura 12 - Resultado final da classificação fuzzy dos clusters

Analisando os resultados conclui-se que os clusters 10 e 11, no exemplo não têm problema de fraude. O cluster 10 apresenta um problema crônico de inadimplência, cujos clientes já são, ou serão brevemente inadimplentes. Para a Light é importante vigiá-los ou se antecipar ao problema chamando-os para negociar. O cluster 11 apresenta uma tendência menor do que o cluster 10 quanto à inadimplência, entretanto apresenta boa quantidade de clientes normais. O cluster 15 apresenta sério



problema de fraude que pode estar camuflado com alguns clientes normais. Para esse grupo não deve haver inadimplência a se considerar. O mesmo acontece com o cluster 16, porém esse grupo apresenta já um pequeno grupo de inadimplentes. Os clusters 1 e 4 apresentam mesma proporção entre clientes normais, fraudulentos e inadimplentes. Para esse grupo o importante primeiro é combater a inadimplência, pois a tendência é que esses clientes sejam cortados e voltem a se religar, aumentando o número de fraudulentos. Por outro lado, aqueles que hoje estão normais tendem a ficarem inadimplentes. O cluster 8 apresenta o 2º maior índice de fraudulentos, numa região onde há também clientes normais e poucos inadimplentes.

### **3. CONCLUSÕES**

Foram apresentados dois métodos candidatos para detecção de fraudes e inadimplência em consumidores da LIGHT. O método de reconhecimento de padrões tem muito potencial, mas foi descartado pela dificuldade em se definir os padrões “verdade” de clientes normais, fraudulentos e inadimplentes, baseados nas respostas dos clientes. Essa dificuldade, não deve descartar totalmente a possibilidade de aplicação no futuro, pois podem ser superadas com mais tempo de pesquisa e com a inclusão, no cadastro de consumidores, de dados qualitativos que possam vir a definir melhor esses padrões. Essa inclusão, também pode melhorar a criação de clusters e separar melhor as respostas dos clientes facilitando a identificação dos tipos de clientes no método de análise fuzzy. O estudo mostrou que, embora os clientes tenham o mesmo DNA eles podem ter comportamentos diferentes dentro do próprio cluster, dificultando qualquer estudo estocástico, baseados em históricos de fraude. O método de análise Fuzzy proporciona um ótimo sentimento visual do problema, sendo uma tendência a ser usada nas empresas de Energia Elétrica.

### **4. REFERÊNCIAS BIBLIOGRÁFICAS E/OU BIBLIOGRAFIA**

Henriques, Henrique de Oliveira et al. "Novas Ações e Alternativas para Redução de Inadimplência e Perdas Comerciais em Regiões Socialmente Desfavorecidas", V CIERTEC, Seminário Internacional sobre Gestão de Perdas, Eficiência Energética e Proteção da Receita no Setor Elétrico, Maceió, Brasil, Agosto 2005.

Calili R. F., Souza R. C., Henriques, H. O., "Desenvolvimento de Sistema para Detecção de Perdas Comerciais em Redes de Distribuição", V CIERTEC, Seminário Internacional sobre Gestão de Perdas, Eficiência Energética e Proteção da Receita no Setor Elétrico, Maceió, Brasil, Agosto 2005.

Henriques, Henrique de Oliveira; TARANTO, Glauco N; FALCÃO, Djalma Mosqueira. Tracking of Frauds in Distribution Feeders. In: IEEE TRANSMISSION AND DISTRIBUTION CONFERENCE, 2002, São Paulo. 2002.

BABBIE, Earl Métodos de pesquisa de survey. Editora UFMG, 1999.

BABBIE, Earl Observing ourselves: essays in social research. Waveland Press, 1986.

FIELDING, Nigel G. LEE, Raymond M. Computer analysis and qualitativa research. Sage Publications, 1998.

GREENBAUM, Thomas L. Focus group research. Sage Publications, 2nd ed., 1998.

KRUEGER, Richard A. El grupo de discussion: guía práctica para la investigaciona aplicada. Ediciones Pirámides, 1991.

ROBERTS, Carl W. Text analysis for the social sciences: methods for drawing statistical inferences from texts and transcripts. Lawrence Erlbaum Associates Publishers, 1997.

ROBINSON, John R. SHAVER, Philip R. WRIGHTSMAN, Lawrence S. Measures of personality and social psychological attitudes. Academic Press, 1991.

SCHUMAN, Howard. Questions and answers in attitude surveys: experiments on question form, wording and context. Sage Publications, 1996

SINGLETON, Royce A. STRAITS, Bruce C. Approaches to social research. Oxford University Press, 3rd ed., 1999.

SUDMAN, Seymour BRADBURN, Norman M. Asking questions: a practical guide to questionnaire design. Jossey-Bass Publishers, 1982.

WEISS Robert S. Learning from strangers: the art and method of qualitative interview studies. The Free Press, 1994.