

**Detecção de fraudes em consumidores de energia elétrica
da classe Alta Tensão Horo Sazonal, utilizando Mapas
Auto Organizáveis Self-Organizing Maps (SOM),
baseado em análise de memórias de massa.**

Fernando Antonio Camargo Guimarães

CAMPO GRANDE

2008

UNIVERSIDADE FEDERAL DO MATO GROSSO DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO
EM ENGENHARIA ELÉTRICA

**Detecção de fraudes em consumidores de energia elétrica
da classe Alta Tensão Horo Sazonal, utilizando Mapas
Auto Organizáveis Self-Organizing Maps (SOM),
baseado em análise de memórias de massa.**

Tese submetida à
Universidade Federal de Mato Grosso do Sul
como parte dos requisitos para a
obtenção do grau de Mestre em Engenharia Elétrica.

Fernando Antonio Camargo Guimarães

Campo Grande, Junho de 2008.

**Detecção de fraudes em consumidores de energia elétrica
da classe Alta Tensão Horo Sazonal, utilizando Mapas
Auto Organizáveis Self-Organizing Maps (SOM),
baseado em análise de memórias de massa.**

Fernando Antonio Camargo Guimarães

‘Este trabalho de dissertação foi julgado adequado como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia Elétrica, Área de Concentração em Inteligência Artificial Teoria e Aplicações’.

Banca Examinadora:

Prof. Dr. João Onofre Pereira Pinto
Orientador
DEL/CCET/UFMS

Prof. Dr. Evandro Mazina Martins
DEL/CCET/UFMS

Prof. Dr. Germano Lambert Torres
UNIFEI/Itajubá

*À minha esposa Raquel Andrés Caram Guimarães
À minha filha Catarina Andrés Caram Guimarães
Ao meu filho Victor Andrés Caram Guimarães
Aos meus pais,
Ruy Cardoso Guimarães e Gessy Camargo Guimarães*

Agradecimentos

A Deus, porque me deu a oportunidade da vida.

Ao Professor João Onofre Pereira Pinto, pela paciência e compreensão de minhas limitações.

Aos colegas e amigos de mestrado Anderson Teruya e Faete sem os quais eu não teria conseguido.

Resumo da Dissertação apresentada à UFMS como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia Elétrica.

Detecção de fraudes em consumidores de energia elétrica da classe Alta Tensão Horo Sazonal, utilizando Mapas Auto Organizáveis Self-Organizing Maps (SOM), baseado em análise de memórias de massa.

Fernando Antonio Camargo Guimarães

Junho/2008

Orientador: Dr. João Onofre Pereira Pinto, PhD.

Área de Concentração: Inteligência Artificial, teoria e aplicações.

Palavras-chave: SOM - Self-Organizing Maps, Detecção de Fraudes, Memória de Massa, Consumo de Energia Elétrica, Grupo Tarifário Horo Sazonal, Redes Neurais

Número de Páginas: 78

RESUMO

O presente trabalho tem por objetivo pesquisar e identificar a melhor configuração de rede SOM - Self-Organizing Maps, capaz de apontar a partir das memórias de massa retiradas dos medidores de clientes de alta tensão da concessionária Enersul, potenciais fraudadores. Uma vez determinada a melhor configuração da rede SOM, esta poderá orientar a área de perdas das concessionárias, quais os clientes que apresentem características de fraudadores e que deverão sofrer inspeções.

Abstract of Dissertation presented to UFMS as a partial fulfillment of the requirements for the degree of Master in Electrical Engineering.

Fraud Detection in Hourly fashioned Electrical Energy Consumer through the use of Artificial Neural Networks (SOM) based on Mass Memories Analysis.

Fernando Antonio Camargo Guimarães

Junho/2008

Orientador: Dr. João Onofre Pereira Pinto, PhD.

Área de Concentração: Inteligência Artificial, teoria e aplicações.

Palavras-chave: SOM - Self-Organizing Maps, Detecção de Fraudes, Memória de Massa, Consumo de Energia Elétrica, Grupo Tarifário Horo Sazonal, Redes Neurais

Número de Páginas: 78

ABSTRACT

The actual presentation target is to research and identify the best configuration for the SOM (Self Organizing Maps) network, ready to show from the mass memories taken from the service provider (Enersul) high tension consumers' meters', potential fraudulent consumers. Once established the best SOM network configuration it will be one which will guide the service provider losses department in pointing customers that hint fraudulent characteristics behavior that must go through mandatory inspections.

Abreviaturas / Siglas

ANEEL	Agência Nacional de Energia Elétrica
ABRADE	Associação Brasileira de Distribuidores de Energia Elétrica
Ch. Af.	Chave de aferição
CELESC	Centrais Elétricas de Santa Catarina
CESP	Centrais Elétricas de São Paulo
DEC	Duração Equivalente de Interrupção por Unidade Consumidora
$d(x,y)$	Distância do vetor x para o vetor y
Dem	Demanda registrada na memória de massa
d_{kj}	Distância lateral entre o vizinho (k) e o neurônio vencedor (l)
ELETROBRÁS	Centrais Elétricas do Brasil S/A
ESB	Eletronic Services
FEC	Frequência Equivalente de Interrupção por Unidade Consumidora
$h_{kj(t)}^{(i)}$	função de vizinhança para o vizinho (k) em relação ao vencedor (l) para o dado (t) na iteração (i), determina o nível de operação;
I	Corrente consumida em ampères
KDD	Descoberta de Conhecimento em Base de Dados
kV	quilo volt
kVar	quilo volt – ampère -reativo
kVarh	quilo volt – ampère – reativo - hora
kW	quilo watts
kWh	quilo watts - hora
LVQ	Learning Vector Quantization
MATLAB	Programa de computador
MWh	megawatts-hora

MDA	Minimal decision algorithm
n	Número total de dimensões ou atributos.
P	Potência requerida
PG	Perdas globais
PLAWIN	Programa desenvolvido pela ESB
PNT _r	Perdas não técnicas regulatórias
PT _r	Perdas técnicas regulatórias
RNA	Rede Neural Artificial
SOM	Self Organizing Maps
TCs	Transformadores de Corrente
TMA	Tempo Médio de Atendimento
TPM	Topology Preserving Map
TPs	Transformadores de Potencial
x_i	Elemento da dimensão ou atributo i do vetor x
y_i	Elemento da dimensão ou atributo i do vetor y
$\sigma(i)$	Largura da vizinhança (abrangência) na iteração (i).

Lista de Figuras

Figura 1.1 – Flagrante de furto de energia elétrica.....	01
Figura 1.2 - Distribuidoras de Energia Elétrica no Brasil (51) Fonte ANEEL 2006....	04
Figura 1.3 – Influência Consumo Irregular nas Perdas.....	05
Figura 1.4 – Trajetória de perdas.....	05
Figura 1.5 – Perdas por região – (Concessionária estudada ano 2003).....	06
Figura 1.6 – Número de Irregularidades por tipo (Fonte: Concessionária em estudo 2004)	08
Figura 1.7 – Exemplo de desvio no ramal de ligação.....	09
Figura 1.8 – Exemplo de desvio no ramal de entrada.....	09
Figura 1.9 – Exemplo de fraude interna no medidor.....	10
Figura 1.10 – Exemplo de fraude interna no medidor.....	11
Figura 1.11 – Exemplo de fraude no medidor, sem a retirada da tampa de vidro e sem a violação do lacre da tampa de bornes.....	12
Figura 1.12 – Exemplo de desvio na chave de aferição.....	13
Figura 3.1. Rede Neural do tipo Perceptron.....	24
Figura 3.2 - Rede Auto-Organizável.....	25
Figura 3.3 – Rede feedforward.....	27
Figura 3.4 - Representação do processo competitivo, todos os modelos recebem a mesma informação, mas apenas um é ativado.....	30
Figura 3.5 - Função de vizinhança aplicada em um mapa hexagonal 4x4.....	31
Figura 3.6 – Linha 1 – Função Seno – 63 Registros.....	34
Figura 3.7 – Linha 2 – Função Co-seno – 63 Registros.....	34
Figura 3.8 – Linha 3 – Função Tangente – 63 Registros.....	34

Figura 3.9- Classificação da rede SOM, com uma base de dados gerada a partir das curvas das Funções Seno, Co-seno e Tangente.....35

Figura 4.1 - Tela do Programa PLAWIN.....39

Figura 4.2 – Tela MATLAB.....44

Figura 4.3 - Fluxograma da seqüência de trabalho.....44

Figura 4.4 – Mapa SOM – Quarta Feira Sem Fraudadores.....47

Figura 4.5 – Mapa SOM – Quarta Feira Com Fraudadores.....48

Figura 4.6 – Consumo x Pulsos – Normal.....52

Figura 4.7 – Consumo x Pulsos – 1º semana normal e após com fraude em 1 fase (1/3 de redução).....53

Figura 4.8 –Consumo x Pulsos-1º semana normal e após com fraude em 2 fases(2/3 de redução).....54

Figura 4.9 – Consumo x Pulsos – 1º e 2º semanas normais e após com fraude em 1 fase (1/3 de redução)55

Figura 4.10 – Consumo x Pulsos – 1º e 2º semanas normais e após com fraude em 2 fases (2/3 de redução)55

Figura 5.1 – Mapa SOM – Quarta Feira Com Fraudadores Reais e 4 Novos Fraudadores Construídos.....57

Lista de Tabelas

Tabela 3.1 – Histórico das Redes Neurais.....	23
Tabela 3.2 – Planilha com curvas Seno Co-seno e Tangente.....	33
Tabela 4.1 – Tabela com características de consumo dos dias da semana dos clientes Comerciais, Industriais, Rurais, Livres, Poder Público e Serviço Público para os meses de Janeiro e Fevereiro.....	41
Tabela 4.2 – Tabela com características de consumo dos dias da semana dos clientes Comerciais e Industriais para os meses de Janeiro e Fevereiro.....	42
Tabela 4.3 – Tabela com os arquivos utilizados para os testes de aplicação da SOM..	43
Tabela 4.4 – Análise dos Clientes x Clientes Fraudadores.....	50
Tabela 5.1 – Análise da apresentação dos novos clientes incluindo os novos fraudadores criados.....	58
Tabela 5.2 – Relações apresentação dos novos clientes (e fraudadores criados).....	59

SUMÁRIO

Abreviaturas / Siglas.....	VIII
Lista de Figuras.....	X
Lista de Tabelas.....	XII
1 – Conceito de Fraude em Consumo de Energia Elétrica.....	01
1.1 – Introdução.....	01
1.2 – Conceito de Fraude.....	07
1.3 - Tipos mais comuns de fraudes.....	07
1.3.1 - Desvio no ramal de ligação.....	08
1.3.2 - Desvio no disjuntor.....	09
1.3.3 - Desvio do ramal de entrada.....	09
1.3.4 - Interferência no medidor.....	10
1.3.5 - Interferência na chave de aferição.....	12
1.3.6 - Interferência nos transformadores para instrumentos.....	13
1.3.7 - Interferência na fiação secundária.....	14
1.3.8 - Ligação à revelia.....	14
1.3.9 - Religação à revelia.....	14
1.4 - Ações para combater as Fraudes.....	15
2 – Revisão bibliográfica	16
3 – Ferramenta Proposta para a Solução do Problema SOM.....	21
3.1 – História.....	22
3.2 - O Algoritmo.....	28
3.3 – Formação do Mapa.....	29
3.4 – Exemplo do uso da SOM para classificação de sinais.....	32
4 – Metodologia Proposta para o desenvolvimento do trabalho.....	36
4.1 - Levantamento das memórias de massa dos clientes.....	36
4.2 - Levantamento das memórias de massa dos clientes fraudadores.....	36
4.3 - Escolha do método de análise – especialista.....	37

4.4 - Formatação das memórias de massa.....	38
4.5 - Aplicação da Ferramenta – SOM.....	43
4.6 – Criação de memórias de massa de clientes fraudadores.....	51
5 – Simulações e resultados.....	56
6 – Conclusão.....	60
7 – Bibliografia.....	62

1 - Conceito de Fraude em Consumo de Energia Elétrica

1.1 - Introdução

A legislação do setor elétrico nacional estipula um valor máximo de repasse de perdas, técnicas e comerciais, à tarifa de energia elétrica dos consumidores de uma distribuidora. Deste modo a diminuição das perdas, além de possibilitar a redução da tarifa para todos os seus consumidores, possibilita a diminuição da perda de receita das concessionárias.

O combate e a prevenção à fraude/furto de energia representam um dos fatores de maior importância para a minimização das perdas comerciais das concessionárias de energia elétrica. Chegam a quase 3,5 bilhões de reais as perdas das distribuidoras de energia no Brasil, com furtos e fraudes de eletricidade por parte dos consumidores. A Figura 1.1 abaixo mostra um flagrante de furto de energia em um bairro da periferia de uma das grandes cidades brasileiras. (Fonte: acervo da ABRADE 2000 a 2006).



Figura 1.1 – Flagrante de furto de energia elétrica.

O maior rombo ocorre na Light, que atende a 3,79 milhões de unidades consumidoras em 28 municípios do estado do Rio de Janeiro. A concessionária perde 720 milhões de reais anuais com esses desvios.

Proporcionalmente, a maior distribuidora prejudicada no Brasil é a Ceron, de Rondônia, em que as chamadas perdas não-técnicas alcançam 35% do mercado que ela atende.

Os números fazem parte do levantamento preparado pela Agência Nacional de Energia Elétrica (Aneel) em 2006, com base nas informações enviadas ao órgão regulador pelas próprias distribuidoras no momento em que elas se submetem ao processo de revisão tarifária, a cada quatro anos.

Foram levantados os números de 61 concessionárias, de todas as regiões do país. A conclusão é alarmante: as perdas atingem 15 milhões de Mwh por ano – o suficiente para abastecer ininterruptamente o Distrito Federal por quase quatro anos.

A seguir está detalhado o conceito de “Perdas Comerciais”, sua diferença em relação as perdas técnicas, e os principais tipos de fraude encontradas.

De acordo com os contratos de concessão para distribuição de energia elétrica celebrados com a União, a receita inicial da concessionária de energia elétrica é dividida em duas parcelas: parcelas A e B.

A parcela A envolve os chamados “custos não gerenciáveis”, explicitados no contrato e cujos montantes e variações fogem à vontade ou influência direta da distribuidora, como a compra de energia, os custos de transmissão e os encargos setoriais. A parcela B compreende o valor remanescente de receita, envolvendo, portanto, os ditos “custos gerenciáveis”. Esses constituem custos próprios da atividade de distribuição e de gestão comercial dos clientes, que estão sujeitos ao controle ou influência das práticas gerenciais adotadas pela empresa, ou seja, os custos de operação (pessoal, material e serviços de terceiros). Além desses, a parcela B inclui a remuneração de capital e os tributos. Os contratos de concessão contemplam procedimentos específicos para reajuste dessas parcelas durante cada ano do ciclo tarifário.

É conhecida a existência de perdas de energia ao longo da cadeia produção-transporte-consumo de energia elétrica. Em um enfoque por segmento, as perdas de energia influem na quantidade de energia comprada e que compõe, como já foi esclarecido, a Parcela A da receita de distribuição. Logo, quanto menor o valor das

perdas, maior o benefício auferido pelos consumidores, com reflexos positivos na modicidade tarifária.

As perdas são separadas em:

- a) Perdas Técnicas: constituem a quantidade de energia elétrica, expressa em megawatt-hora por ano (Mwh/ano), dissipada entre os suprimentos de energia da distribuidora e os pontos de entrega nas instalações das unidades consumidoras ou distribuidoras supridas. Essa perda é decorrente das leis da Física relativas aos processos de transporte, transformação de tensão e das perdas inerentes aos equipamentos de medição; e
- b) Perdas Não Técnicas: apuradas pela diferença entre as perdas totais e as perdas técnicas, considerando, portanto, todas as demais perdas associadas à distribuição de energia elétrica, tais como furtos de energia, erros de medição, erros no processo de faturamento, unidades consumidoras sem equipamento de medição, etc.

Conforme mencionado, um nível elevado de perdas será suprido com acréscimo na energia gerada. Dado que o custo marginal de longo prazo de geração tende a ser mais alto que os custos associados à redução de perdas técnicas e não técnicas na atividade de distribuição de energia, torna-se importante a gestão dessas perdas, com o objetivo de reduzi-las. Como mostra a Figura 1.2, para uma amostra de 51 distribuidoras do primeiro ciclo de revisão tarifária, a média de perdas de energia foi de 5,67% (não técnicas) e 8,18% (técnicas). Contudo, têm-se exemplos de todas as composições (elevadas perdas técnicas e não técnicas, ou somente uma das duas componentes), existindo empresas com níveis de perdas acima de 30% de seu mercado de energia.

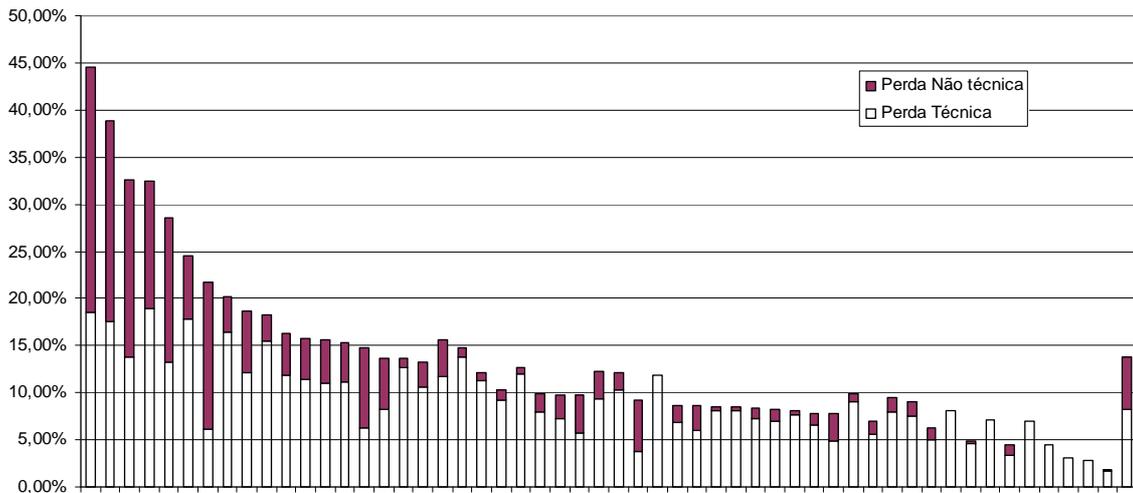


Figura 1.2 - Distribuidoras de Energia Elétrica no Brasil (51)
Fonte ANEEL - 2006

As perdas técnicas não são proporcionais à energia. Variam com o quadrado da potência consumida. Isto é, se tivermos duas cargas com o mesmo consumo, mas com diferentes distribuições de potências, as perdas serão diferentes. Sabe-se ainda que a obrigação da distribuidora está no planejamento de seu sistema para atendimento do mercado faturado. Assim, pode-se prever que as perdas não técnicas devido a furtos e desvios de energia, provocam um adicional de perdas técnicas, pois os alimentadores, transformadores e outros equipamentos podem operar acima do carregamento de projeto, implicando em temperaturas mais elevadas de operação que afetam as perdas técnicas.

A Figura 1.3 representa duas cargas, sendo uma regular (Ir) e outra com fraude (Ii). A Equação 1.1 descreve a influência das fraudes no aumento da potência requerida da distribuidora de energia. A influência das perdas não técnicas nas perdas técnicas não pode ser desprezada. Assim, quantificar e reduzir esse montante são ações imprescindíveis para a sociedade.

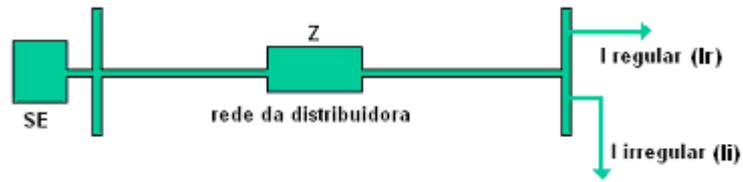


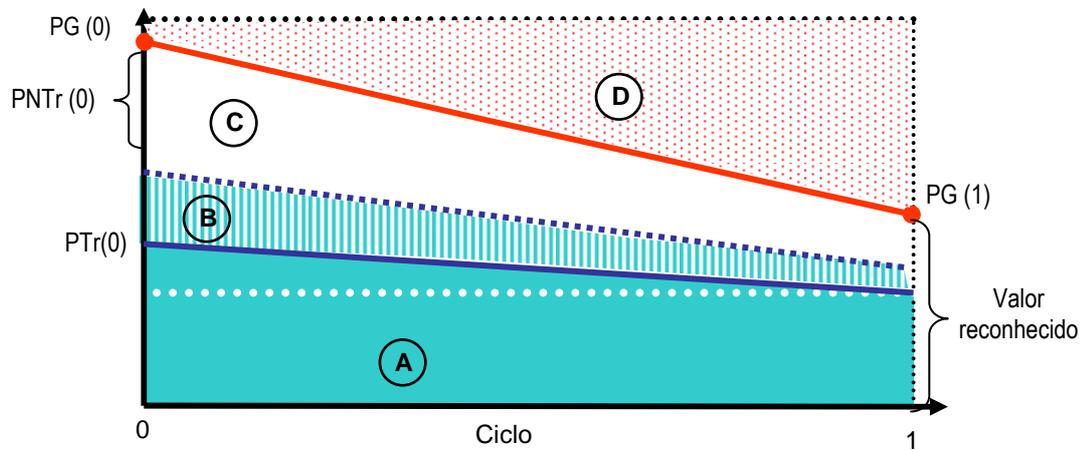
Figura 1.3 – Influência Consumo Irregular nas Perdas

$$P \approx I_r^2 + 2 \cdot I_r \cdot I_i + I_i^2 \quad (1.1)$$

$P =$ Perdas
 $I_r =$ Corrente consumidores regulares
 $I_i =$ Corrente consumidores irregulares – com fraude

}
 Influência do consumo irregular - Fraudes

Graficamente, o impacto e o tratamento da trajetória de perdas estão demonstrados na Figura 1.4, considerando cada componente:



- A - Perdas técnicas devido aos consumidores regulares
- B - Perdas técnicas devido aos consumidores não regulares
- C - Perdas não técnicas não gerenciáveis
- D - Perdas gerenciáveis

PG – perdas globais
 PNTr – Perdas não técnicas regulatória
 Ptr – Perdas técnicas regulatórias

Figura 1.4 – Trajetória de perdas

Na Figura 1.4, são apresentadas as perdas reconhecidas pela ANEEL, cujo custo financeiro é absorvido por todos os consumidores através das tarifas (PG1). A diferença entre as perdas reais (PG0) e as perdas reconhecidas (PG1), implica em prejuízos para as concessionárias de energia elétrica. No caso das concessionárias privatizadas, o prejuízo fica com os acionistas, e nas não privatizadas, com o povo. Combatendo as fraudes, haverá uma diminuição das áreas B (perdas técnicas devido aos consumidores não regulares) e C (perdas não técnicas não gerenciáveis). O maior percentual de perdas não técnicas está associado ao furto de energia e, como foi visto, essas perdas estão associadas a diversos fatores. Uma redução de tais perdas traz conseqüências não só do ponto de vista econômico, mas também do social. No caso da concessionária objeto do estudo, as perdas no ano de 2003, ficaram na ordem de 15%, sendo a contribuição por região e principais cidades, mostradas na Figura 1.5 abaixo:

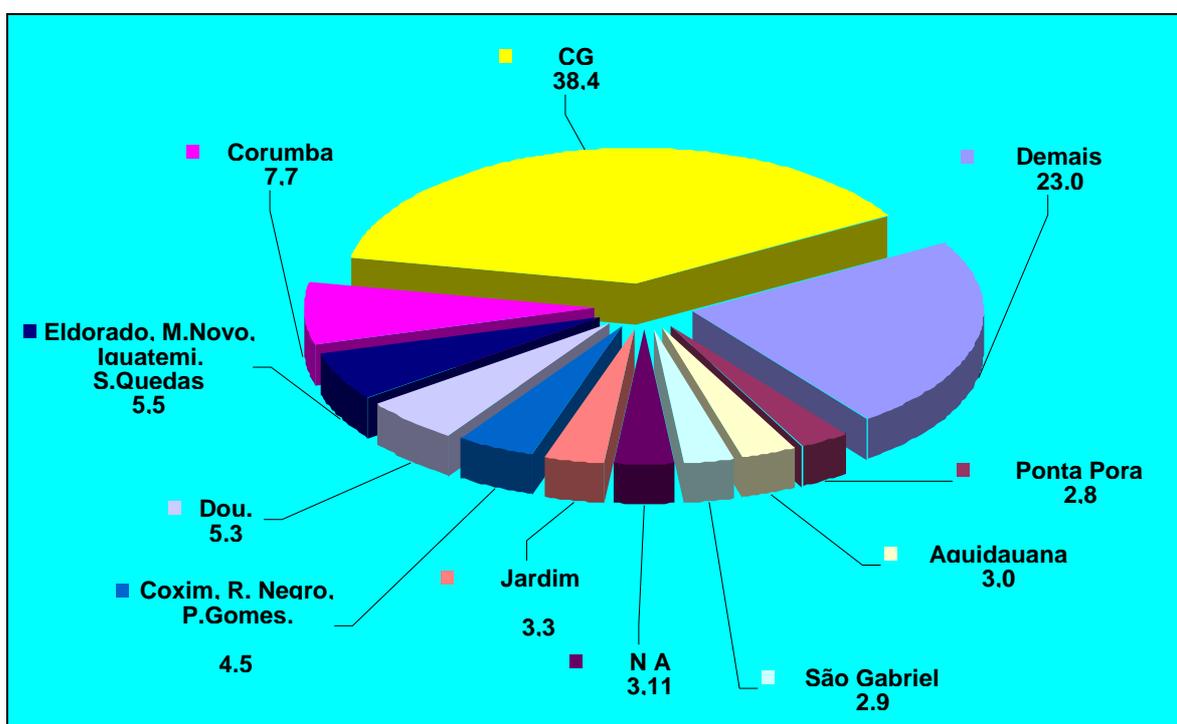


Figura 1.5 – Perdas por região – (Concessionária estudada ano 2003)

Este trabalho tem como objetivo obter uma rede neural artificial tipo Mapas Auto Organizáveis (SOM – *Self Organizing Map*), capaz de apontar potenciais fraudadores a partir das memórias de massa retiradas dos medidores de clientes de alta tensão de uma concessionária de energia elétrica. Sua finalidade é tornar-se uma ferramenta de apoio à área comercial das empresas concessionárias de energia na determinação dos clientes a serem inspecionados em campo, buscando a minimização das perdas comerciais.

1.2 – Conceito de Fraude

O conceito de fraude pressupõe uma ação ilícita e consciente do cliente. Resumidamente pode-se identificar uma fraude como sendo qualquer ato provocado intencionalmente com o objetivo de diminuir o registro da energia efetivamente consumida na unidade consumidora.

1.3 - Tipos mais comuns de fraudes

Considerando as fraudes encontradas em campo, podem-se identificar alguns tipos mais comuns, listados a seguir. A Figura 1.6, abaixo, apresenta o número de irregularidades, por tipo, na Concessionária objeto do estudo acumulado nos meses de Jan/Ago de 2004. Nesta, podemos verificar que as fraudes mais grosseiras, são as que representam o maior volume de ocorrência 74 % (fraudes antes do medidor + auto-religado) e as fraudes mais “trabalhadas” representam 9,1 % (fraude no medidor + fraude nos equipamentos auxiliares de medição TC/TP/Ch. Af.).

O que surpreende, é a crescente criatividade dos ditos “técnicos” que prometem uma redução de consumo aos clientes, mediante pagamento. Normalmente estes técnicos vendem as “soluções de redução de consumo” em troca da economia que o cliente vier a ter nas primeiras 3 ou 4 faturas de energia após a realização do “trabalho”. Alguns clientes acabam embarcando nesta sem a intenção de roubar, mas sendo enganados por estes técnicos que prometem uma redução de consumo após a instalação de algum equipamento por eles inventado.

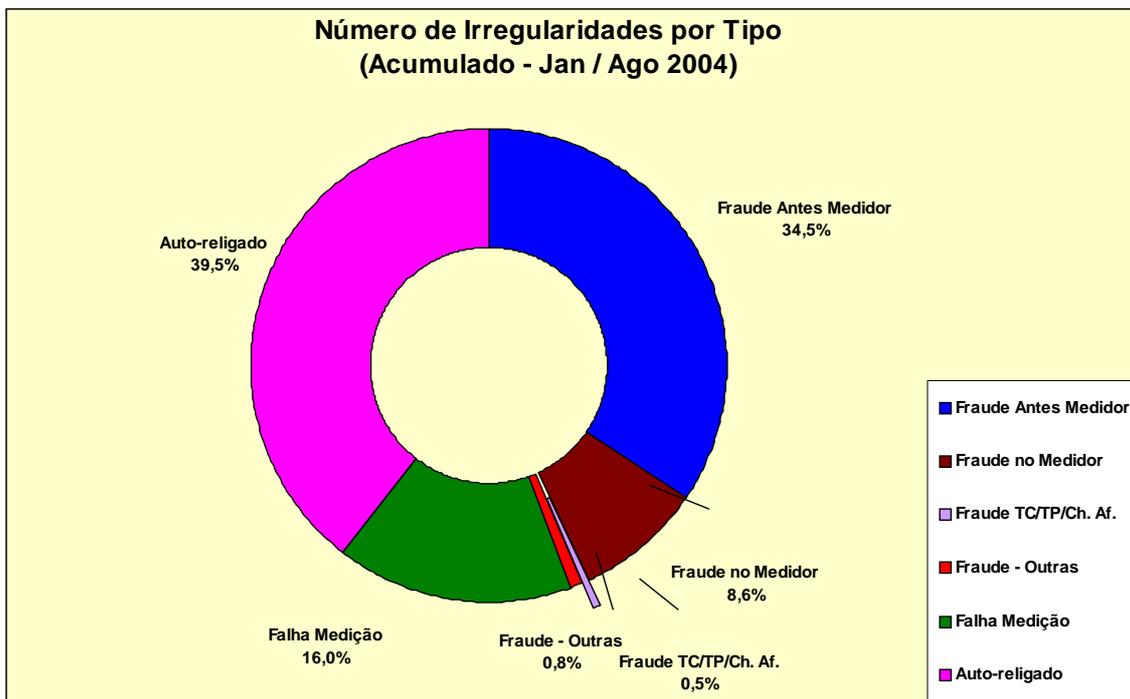


Figura 1.6 – Número de Irregularidades por tipo (Fonte: Concessionária em estudo 2004)

1.3.1 – Desvio no ramal de ligação

Derivação de energia, praticada nos condutores que interligam o poste da concessionária e a entrada de serviço da unidade consumidora. Na Figura 1.7, são mostrados vários fios chegando ao sistema de baixa tensão da concessionária e ligada diretamente no vão dos cabos, isto é, ligações clandestinas efetuadas sem a menor segurança.



Figura 1.7 – Exemplo de desvio no ramal de ligação

1.3.2 - Desvio no disjuntor

Derivação de energia, praticada nos bornes do disjuntor. Pode se dar na entrada ou na saída do disjuntor. Se a derivação se der nos bornes de saída, ao se desligar o disjuntor, a carga que estiver sendo alimentada clandestinamente ficará sem energia. Se a derivação se der nos bornes de entrada do disjuntor, além da prática ilícita do furto de energia, o consumidor estará, ainda, sobrecarregando o ramal de entrada e/ou de ligação que foi dimensionado para a corrente nominal do disjuntor.

1.3.3 - Desvio do ramal de entrada

Derivação efetuada no trecho da fiação compreendida entre o pingadouro e a caixa de medição. Essa forma de desvio normalmente é executada no interior da parede, do muro ou mureta, interceptando-se o eletroduto e a fiação para efetuar a derivação. Na Figura 1.8 tem-se um exemplo de derivação em dois fios dentro da mureta onde está instalado o medidor.



Figura 1.8 – Exemplo de desvio no ramal de entrada

1.3.4 - Interferência no medidor

Pode ser provocada de diversas formas, tais como, a retirada da tampa de vidro, conseqüentemente violando o lacre de aferição, e efetuando-se uma das seguintes ações:

- manipulação dos ponteiros alterando-se o consumo registrado;
- introdução de qualquer objeto que altere o giro do disco do medidor;
- manipulação na fiação interna do medidor;
- provocação de avarias nas bobinas de corrente e/ou de tensão, etc.

Na Figura 1.9, tem-se exemplo de manipulação na fiação interna do medidor através de “jumper” entre a entrada e saída realizada com fios da cor verde. Na Figura 1.10, tem-se outro exemplo de manipulação na fiação interna do medidor através de “jumper” entre a entrada e saída realizada com fios da cor preta.

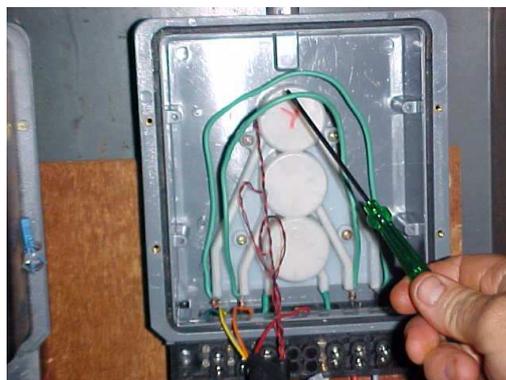


Figura 1.9 – Exemplo de fraude interna no medidor

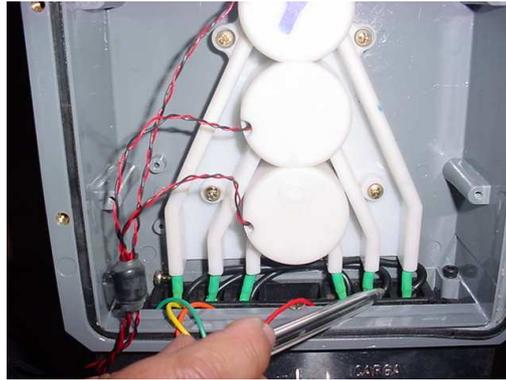


Figura 1.10 – Exemplo de fraude interna no medidor

De outro modo, sem a retirada da tampa de vidro é possível efetuar as seguintes ações:

- perfuração da carcaça ou da tampa e introdução de objetos ou manipulação dos mecanismos internos do medidor;
- manipulação da fiação de ligação do medidor invertendo o sentido da corrente nas bobinas do medidor - sem a violação do lacre da tampa de bornes;
- manipulação da fiação de ligação do medidor ou das conexões internas dos bornes invertendo o sentido da corrente nas bobinas do medidor - com a violação do lacre da tampa de bornes;

Na Figura 1.11, tem-se exemplo de manipulação da fiação de ligação do medidor invertendo o sentido da corrente nas bobinas do medidor - sem a violação do lacre da tampa de bornes.

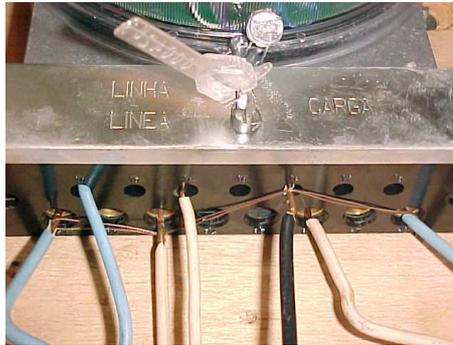


Figura 1.11 – Exemplo de fraude no medidor, sem a retirada da tampa de vidro e sem a violação do lacre da tampa de bornes

1.3.5 - Interferência na chave de aferição

A interferência na chave de aferição normalmente é precedida das violações do lacre da caixa de medição e do lacre da tampa da chave de aferição. São interferências mais simples e mais fáceis de serem identificadas, indicando que foram realizadas por uma pessoa com poucos conhecimentos na área. Exemplos de interferências normalmente encontradas nas chaves de aferição:

- manipulação das lâminas da chave de aferição provocando a interrupção do fluxo normal da corrente secundária que deve circular pelo medidor;
- provocação de alterações sobre as conexões da fiação com a chave;
- operação eventual de um ou mais pólos da chave de aferição, para que o medidor registre apenas o que se deseja.

Na Figura 1.12, tem-se exemplo de manipulação da fiação das lâminas da chave de aferição.



Figura 1.12– Exemplo de desvio na chave de aferição

1.3.6 - Interferência nos transformadores para instrumentos

A interferência nos transformadores para instrumentos normalmente é precedida de um desligamento longo na entrada de energia. São interferências mais complexas e mais difíceis de serem identificadas, indicando que foram realizadas por uma pessoa com conhecimentos na área. Exemplos de interferências normalmente encontradas nos transformadores para instrumentos:

- alteração provocada na forma de ligação dos TPs e/ou TCs;
- adulteração da placa de identificação;
- alteração dos taps de ligação;
- inversão de ligações, etc.

1.3.7 - Interferência na fiação secundária

A interferência na fiação secundária normalmente não é precedida de um desligamento na entrada de energia. São interferências mais simples, porém difíceis de serem identificadas, sendo normalmente a de maior ocorrência. Exemplos de interferências normalmente encontradas nas fiações secundárias:

- inversão das ligações da fiação, na saída dos transformadores, nas régua de bornes das caixas de passagem ou nas conexões com a chave de aferição;
- seccionamento do condutor internamente à isolação dificultando que se perceba em uma inspeção visual;
- aplicação de material isolante nas conexões da fiação com a chave de aferição ou com os TPs e TCs, etc.

1.3.8 - Ligação à revelia

Ligação de unidade consumidora, derivando-se diretamente da rede secundária ou de qualquer ramal de ligação de unidade consumidora vizinha, clandestinamente.

1.3.9 - Religação à revelia

Religação de unidade consumidora, sem medição, que tenha sido desligada a pedido do consumidor ou por iniciativa da concessionária, clandestinamente.

1.4 – Ações para combater as fraudes

A concessionária tem um conjunto de ferramentas e possibilidades para gerenciar essas situações, em geral muito mais efetivas e sem as consequências negativas da suspensão do fornecimento.

É importante que a concessionária desenvolva soluções técnicas e comerciais adequadas para dificultar a fraude e difundir a “cultura” de uso regular da energia elétrica e da adimplência no pagamento. Entre essas ações, destacam-se:

- Ação de combate permanente e efetivo por parte de toda a empresa, principalmente, das áreas com relacionamento direto com os consumidores;
- Promoção da cultura de eficiência energética e adequação do consumo de energia as possibilidade de pagamento por parte dos consumidores;
- Implementação de novas tecnologias para as redes de distribuição, ramais de serviço, medidores e programas de computação para análise;

Dentro da filosofia de implementação de novas tecnologias é que o presente trabalho se encaixa, buscando ajudar a coibir uma prática tão nociva à sociedade.

2 - Revisão bibliográfica

Nesta seção são apresentados os principais artigos relativos a fraudes em concessionária de energia elétrica pesquisados durante a elaboração deste trabalho. Nestes artigos, encontram-se aplicações de Árvore de Decisão, Rough Sets, SOM - Self Organizing Map, os estudos de Kohonen (KOHONEN, T., 2001) além de referências sobre perdas comerciais (fraudes), sua detecção e métodos de análise. Foram pesquisados também, alguns artigos sobre fraudes em cartões de créditos e em concessionárias de distribuição de água.

Juntamente com as empresas distribuidoras e energia elétrica, as empresas de distribuição de água, são uns dos segmentos empresarias que mais sofrem a ação de fraudes. Passini 2002, utilizando o programa DB2 Intelligent Miner indica que na Sanasa, empresa de distribuição de água de Campinas em um total de 26,6% de perdas na distribuição, 5% pode ser considerada como fraudes.

Outro segmento importante e onde se concentra a maioria dos trabalhos publicados sobre fraude é a área de cartões de crédito. Kou et al., 2004 faz uma revisão dos principais métodos de detecção de fraudes utilizando técnicas de inteligência artificial nesta área. Normalmente estas fraudes em cartões de crédito não vêm a publico e são normalmente abafadas pelas próprias instituições financeiras com a finalidade de preservar a credibilidade junto aos clientes. Deste modo estes tipos de fraudes são pouco conhecidos.

Aleskerov et al., 1997 apresenta o CARDWATCH sendo um programa de mineração de dados voltado para a detecção de fraudes em cartões de crédito, baseando-se em uma rede neural artificial de alimentação direta (feedforward). Tendo conseguido como resultado uma taxa de acerto de 85% na detecção de fraudes.

Reis et al., 2004 apresentaram um sistema baseado em “Árvore de decisão” com intuito de pré-selecionar consumidores de energia elétrica para inspeção. O objetivo principal foi o de detectar fraudes e erros de medição. Foi conseguida uma taxa de acerto de 40% para detecção de fraudadores, utilizando um universo de 40.000 registros e cinco atributos previamente selecionados. O resultado foi bem melhor que os 15% normalmente conseguidos pela concessionária estudada. No trabalho foi utilizada “Árvore de Decisão” como ferramenta de mineração de dados. A idéia foi a de desenvolver um sistema que levaria ao aumento das inspeções bem sucedidas de clientes de baixa tensão de concessionária de energia elétrica, visando diminuir as perdas comerciais. O sistema elaborado identificou consumidores potencialmente suspeitos e elaborou uma lista para inspeção. Apesar de o trabalho ter atingido o objetivo proposto, o autor enumerou vários problemas encontrados durante a análise e aplicação da ferramenta: dificuldade de acesso direto aos dados de origem; dados incompletos ou com inconsistências; utilização de consumo mínimo em determinada faixa; dados nulos e a simplificação na classificação utilizada.

Cabral et al., 2004 identificaram padrões de comportamento fraudulentos em dados históricos utilizando a técnica de inteligência artificial “Rough Sets”. Foi organizado um sistema de informação contendo um conjunto de clientes e seus respectivos atributos. Neste sistema, foram aplicados os conceitos de aproximação inferior, reduto e do algoritmo da decisão mínima, ou *minimal decision algorithm* (MDA), conseguindo a redução do número de atributos a serem utilizados. A partir do Sistema de Informação reduzido, derivou-se um conjunto de regras as quais representaram perfis de comportamento dos clientes. Com os perfis de comportamento fraudulento, consolidou-se um sistema de regras de classificação, alcançando uma taxa de acertos de 20% na indicação das fraudes.

Em (Cristian Mara M. M. Patrício, 2005), foi desenvolvido um sistema baseado em técnicas de inteligência artificial na teoria de Rough Sets que definiu perfis de comportamentos diários de consumidores de energia elétrica. Foram apontados

comportamentos normais e anormais (possíveis fraudadores ou erros na medição) dentro de um conjunto de banco de dados apresentados (consumidores e seus históricos de consumo). Foram localizados 68,7% dos fraudadores utilizados no processo de teste e 22,2% dos clientes normais foram acusados indevidamente.

Os Mapas Auto-Organizáveis (SOM - Self Organizing Map) se tornaram um poderoso método para visualização de dados com grandes dimensões, gerando mapeamentos de um espaço de dimensão elevada em estruturas cuja dimensão topológica é inferior à original. Estes mapeamentos são capazes de preservar as relações de vizinhança dos dados de entrada.

A SOM consegue uma redução (compressão) das informações, sem perder a métrica ou hierarquia da relação entre os elementos além de conseguir manter as características intrínsecas da relação das variáveis envolvidas. Estas características de visualização e manutenção das características intrínsecas são bastante úteis na engenharia em análise de processos, reconhecimento de padrões, controle e comunicações.

Um dos trabalhos utilizando a SOM na área de Engenharia Elétrica, mais precisamente na área de Sistemas de Potência, foi a Classificação de conjuntos de consumidores de energia elétrica via Mapas Auto-Organizáveis e Estatística Multivariada (SPERANDIO et al. 2004). Neste, foi apresentada uma metodologia para formação de agrupamentos de conjuntos consumidores de energia elétrica, a partir dos dados de 260 conjuntos sob concessão das Centrais Elétricas de Santa Catarina (CELESC). Utilizando uma validação cruzada entre a rede neural auto-organizável SOM e o método estatístico das k-médias, os grupos obtidos foram classificados de forma a sinalizar quando o nível de continuidade fosse inferior ao nível de referência. A concessionária de energia, conhecendo melhor a situação de cada conjunto, pode aplicar o controle de qualidade necessário, e rever as tarifas de maneira a remunerar os investimentos indispensáveis à prestação do serviço de acordo com a qualidade regulamentada, ou negociar junto ao órgão regulador a situação do conjunto em

relação à meta determinada. Este trabalho mostrou a SOM como uma excelente ferramenta de agrupamento.

LIMA et al.1999 avaliaram o desempenho de classificadores neurais baseados em Mapas de Kohonen, para análise e classificação dos perfis de curvas de carga com influência climática. A análise foi feita para os dados do Estado de São Paulo, sistema CESP, entre setembro e dezembro de 1997. Apesar da escassez de informações climáticas mais precisas e um histórico mais detalhado com informações contínuas, foi possível avaliar que as condições climáticas têm maior importância no período da ponta, validando a aplicação da ferramenta SOM.

No segundo trabalho de LIMA et al. 1999, a SOM conseguiu mostrar a correlação entre o clima e as ocorrências na rede elétrica. Foram utilizadas duas bases de dados, sendo uma de variáveis climáticas (temperatura), e outra de variáveis técnicas (DEC, FEC, TMA).

Em Merja Oja et al. 2002, foram listados 3343 artigos científicos entre 1981 e 1997 que se beneficiaram da SOM e entre 1998 a 2002 listou mais 2092 novos artigos, totalizando 5435 artigos. Dentre as áreas de pesquisa beneficiadas pela aplicação da SOM, foram selecionadas abaixo, as que mais apresentaram artigos:

- Imagem e vídeo
- Negócios e Administração
- Informação científica e documentação
- Análise de circuitos
- Processamento de sinal
- Controle
- Reconhecimento da fala

- Teoria da informação e codificação
- Engenharia biológica e medicina
- Software
- Inteligência artificial
- Matemática técnica
- Reconhecimento de padrões

3 - Ferramenta Proposta para a Solução do Problema

Neste trabalho, propomos a utilização da teoria dos Agrupamentos, ou *Cluster Analysis*, que são ferramentas para análise exploratória de dados a fim de resolver problemas de classificação.

O objetivo é associar argumentos variáveis (consumo de energia) em clusters, de forma que o grau de similaridade seja grande entre os membros de um mesmo cluster e pequeno entre clusters diferentes. Deste modo a análise de agrupamentos torna-se uma ferramenta de extração de conhecimento, podendo revelar associações e estruturas em uma base de dados que não são visíveis a princípio. Os resultados dessa análise podem contribuir para uma classificação, tal como a de possíveis clientes fraudadores, como é o caso deste trabalho.

A ferramenta específica escolhida foi a SOM, principalmente pela sua capacidade de classificação e agrupamento para reconhecimento de padrões e análise de dados (especialmente de dados com grandes dimensões e sem conhecimento antecipado destas relações).

Primeiro é estabelecido uma medida de proximidade entre os elementos que compõem a amostra em estudo para poder formar os grupos. Normalmente chamada de distância, esta medida de proximidade pode ser determinada utilizando diversos critérios, sendo que neste trabalho é utilizada a norma Euclidiana, geralmente a mais comum. A equação que determina a distância entre dois elementos (vetores) em um espaço euclidiano de n-dimensões é exemplificada na Equação (4.1):

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.1)$$

onde:

$d(x, y)$ = Distância do vetor x para o vetor y ;

x_i, y_i = Elemento da dimensão ou atributo i dos vetores x e y ;

n = Número total de dimensões ou atributos.

3.1 História

A modelagem dos sistemas utilizados neste trabalho vem do campo de conhecimento denominado Redes Neurais Artificiais - RNA; estes por sua vez são sistemas inspirados na abordagem biológica do processamento da informação nos organismos vivos.

Basicamente, as redes neurais artificiais são programas de computador que tentam copiar a maneira como o cérebro resolve problemas.

Historicamente, as RNA, tiveram um período de entusiasmo inicial até os anos 60, um período de desencantamento até os anos 80 e o seu ressurgimento no final dos anos 80 e início dos anos 90.

Na Tabela 3.1, tem-se uma cópia do histórico, apresentado por Adolfo Bauchspiess (2004), das redes neurais, contendo pesquisadores e inovações que tiveram grande repercussão na área de Redes Neurais Artificiais.

Tabela 3.1 – Histórico das Redes Neurais

1943	McCulloch	Neurônio Bool	Entusiasmo Inicial
1949	Hebb	Regra de aprendizado	
1957	Rosenblatt	Perceptron	
1960	Widrow-Hoff	ADALINE/MADALINE LMS	
	Rosenblatt	Perceptron Multicamadas, sem treinamento	
1969	Minsky-Papert	Perceptrons	Desencantamento
1974	Werbos	Algoritmo Error Backpropagation – sem repercussão	
1982	Hopfield	Rede realimentada	
1986	Rumelhart, Hinton & Williams	PDP – MIT Backpropagation p/ Perceptron Multicamadas	Ressurgimento
		Função de ativação contínua sigmóide	
1987	Kosko	BAM	

No início do século XX descobriu-se o funcionamento básico dos neurônios biológicos, como o disparo de impulsos elétricos através de suas densas conexões. Este princípio foi utilizado nos primeiros modelos matemáticos na década de 40, onde vários elementos interagem formando uma soma ponderada de sinais que produz uma resposta quando essa soma excede um valor limite. Esse paradigma foi primeiramente empregado em circuitos analógicos.

Então na década de 70, com o avanço da computação começou-se a explorar algoritmos que utilizavam tal paradigma em redes encadeadas, onde um estímulo passa por camadas de modelos até que a resposta seja conhecida.

Os pesos de ponderação dos modelos são ajustados por retro-propagação, quando são apresentados estímulos (dados) com resposta conhecida à rede, e o erro é retornado para ser compensada, esta fase é conhecida por treinamento.

Após a rede ter sido treinada com uma base de dados, ela passa a ter, por exemplo, a capacidade de interpolação das respostas esperadas.

Esse método ficou conhecido como Perceptrons Multicamada (Multi-Layer Perceptrons - MLP), e é o tipo de RNA mais difundido. Pode-se visualizar esta rede na Figura 3.1 a seguir:

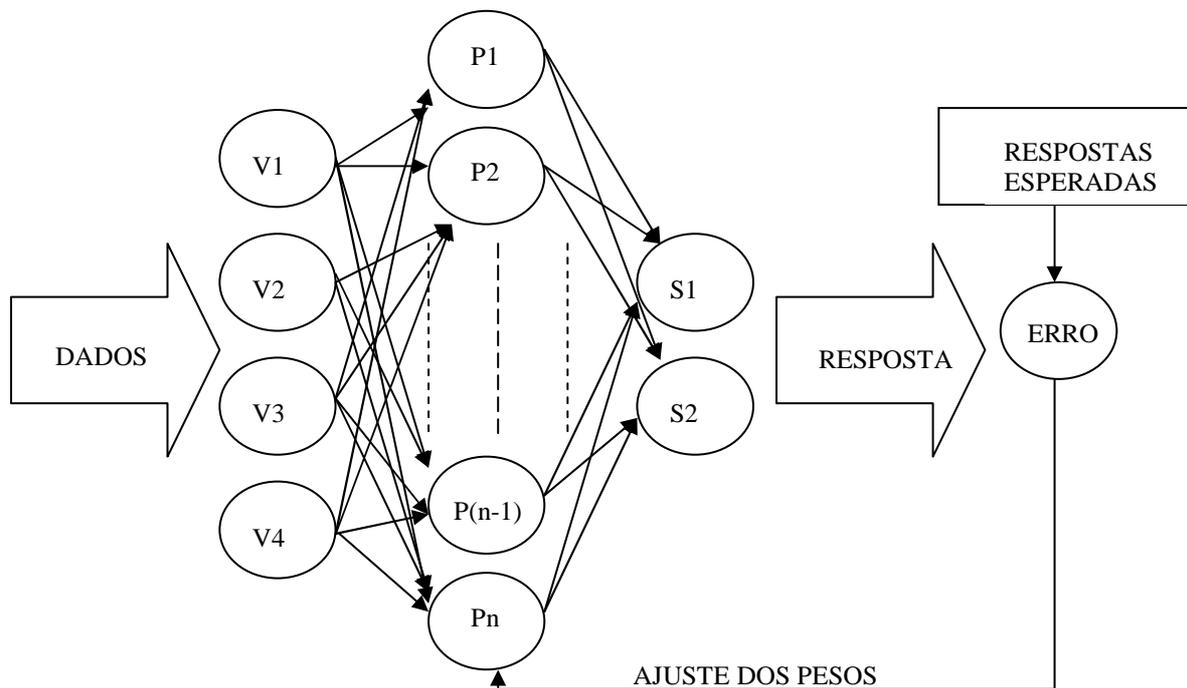


Figura 3.1. Rede Neural do tipo Perceptron.

A partir dos anos 80 alguns estudos voltaram-se para redes de aprendizados competitivos, inspirados na configuração celular do tecido cerebral, onde todos os neurônios disputam o mesmo estímulo e apenas o vencedor dispara a resposta.

Nesse método não há um processo de retro-propagação, o que significa que o treinamento da rede não é supervisionado, sendo que os modelos são atualizados pela informação que cada um recebe.

Esse tipo de RNA geralmente é formada por uma rede bidimensional com vários neurônios interligados (Figura 3.2), e por um processo de organização, dispendo os dados topologicamente, caracterizando um mapa que reconstrói a estrutura hierárquica dos dados.

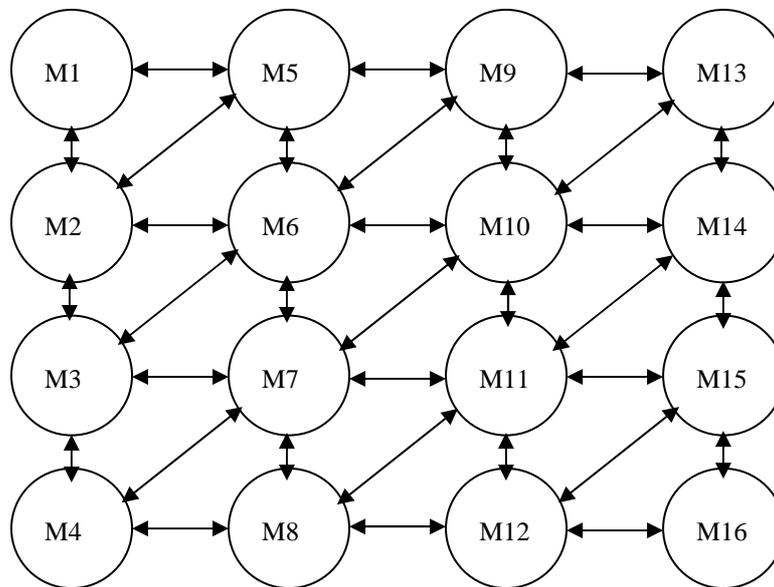


Figura 3.2 - Rede Auto-Organizável.

Vários estudos demonstraram que no cérebro existem neurônios que se organizam em zonas específicas, de modo que as informações originárias dos órgãos sensoriais se apresentam internamente em locais determinados e em forma de camadas bidimensionais. Existem exemplos de mapas detectados do sistema visual (após estímulos visuais) em zonas do córtex (camada externa do cérebro) e também do

sistema auditivo onde se detectou organizações de neurônios segundo a frequência a que foram submetidos. E apesar do pensamento de que estas organizações estão pré-determinadas geneticamente, estudiosos acreditam que ela possa se originar mediante um processo de aprendizagem. Isto sugere que o cérebro tem a capacidade de formar mapas topológicos resultados das informações recebidas do exterior. Esta teoria explica a facilidade do cérebro em trabalhar com elementos semânticos.

Com base nestas idéias, T. Kohonen apresentou, em 1982, um sistema com um comportamento semelhante, isto é, um modelo de rede neural com capacidade de formar mapas de características de maneira similar ao que ocorre no cérebro. O objetivo de Kohonen foi o de demonstrar que um estímulo externo (uma informação de entrada) era suficiente para se conseguir a formação de mapas, isto supondo uma estrutura própria e uma descrição funcional do comportamento da rede.

Os mapas auto-organizáveis pertencem à categoria das redes competitivas com aprendizagem não supervisionada. Possuem uma arquitetura de duas camadas, entrada e saída, funções de ativação lineares e fluxo de informação unidirecional.

As unidades de entrada recebem dados contínuos normalizados onde são definidos pesos aleatórios que definirão a camada de saída. Cada padrão de entrada ativará uma única saída.

O objetivo deste tipo de rede é classificar os padrões de entrada em grupos de características similares, de maneira que cada grupo ativará sempre a(s) mesma(s) saída(s). Cada grupo de entrada é representado pelos pesos das conexões da unidade de saída triunfante.

A arquitetura da versão original de Kohonen, LVQ (Learning Vector Quantization), era baseada em apenas uma só dimensão e um só sentido, isto é do tipo feedforward - Cada um dos N neurônios de entrada se conectava aos M neurônios de saída através de conexões diretas em apenas um sentido, como pode ser visto na Figura 3.3.

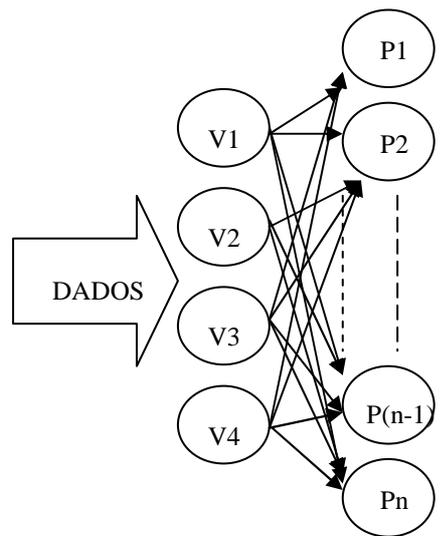


Figura 3.3 – Rede feedforward

Entre os neurônios de entrada e a camada de saída, pode-se dizer que existem conexões “implícitas” de laterais de inibição (com peso negativo), pois mesmo que não estejam conectados, cada um dos neurônios vão ter influência sobre os vencedores. Os valores que os pesos assumem durante o processo de aprendizagem da rede, dependem desta interação lateral. A influência que um neurônio exerce nas demais é função da distancia entre elas, sendo alta influência quando juntos e pequena quando estão afastadas. Normalmente a distribuição desta influência se dá da forma de um chapéu mexicano (gaussiana).

A outra versão apresentada por Kohonen, TPM (Topology Preserving Map) estabeleceu uma correspondência entre os dados de entrada e um espaço bidimensional de saída, criando mapas topológicos de duas dimensões. Assim dados de entrada com características semelhantes ativam neurônios em zonas próximas na camada de saída.

Hoje em dia, o SOM já é um método amplamente utilizado nas mais variadas áreas que exigem classificação de dados.

3.2 O Algoritmo SOM (Self Organizing Map)

O algoritmo determina a forma com que o mapa irá responder à entrada de dados externos. Considerando que o algoritmo pode variar de acordo com cada aplicação, é apresentada a seguir uma forma de representação do algoritmo computacional de Kohonen, (KOHONEN, 1996):

1º: Inicializar:

- definir a taxa de aprendizagem inicial
- definir o raio topológico
- definir pesos: com valores aleatórios entre 0 e 1
- Informar o número máximo de épocas a serem executadas

2º: Critério de parada:

- Comparar o número máximo de épocas e a alteração dos valores da matriz peso.

3º: Treinamento:

Para cada vetor:

Fase Competitiva

- Calcular as distâncias de cada linha da matriz de dados com todas as colunas da matriz peso

- Encontrar o índice tal que a distancia seja o mínimo (é o neurônio vencedor)

Fase Cooperativa

- Para todas as colunas da matriz peso, localizar os vizinhos do neurônio vencedor

pelo raio topológico

Fase Adaptativa

- Atualizar toda a matriz peso:

Retornar para Passo 2: com a entrada de um novo vetor

4º: Atualização:

- Atualizar taxa de aprendizagem: (uma função decrescente linear, exponencial ou geométrica em função das épocas)

- Atualizar raio topológico: (definido como uma função monotonicamente decrescente em função das épocas)

5º: Voltar para o 2º (Critério de parada)

3.3 Formação do Mapa

Para a formação do mapa de Kohonen, existem três processos envolvidos, que são citados abaixo conforme HAYKIN (2001):

1 - Competição: para cada padrão de entrada, os neurônios da rede computam os seus respectivos valores de uma função discriminante. Esta função provê as bases para a competição entre os neurônios. Conforme representado na Figura 3.4, o neurônio com maior valor da função discriminante é declarado vencedor da competição.

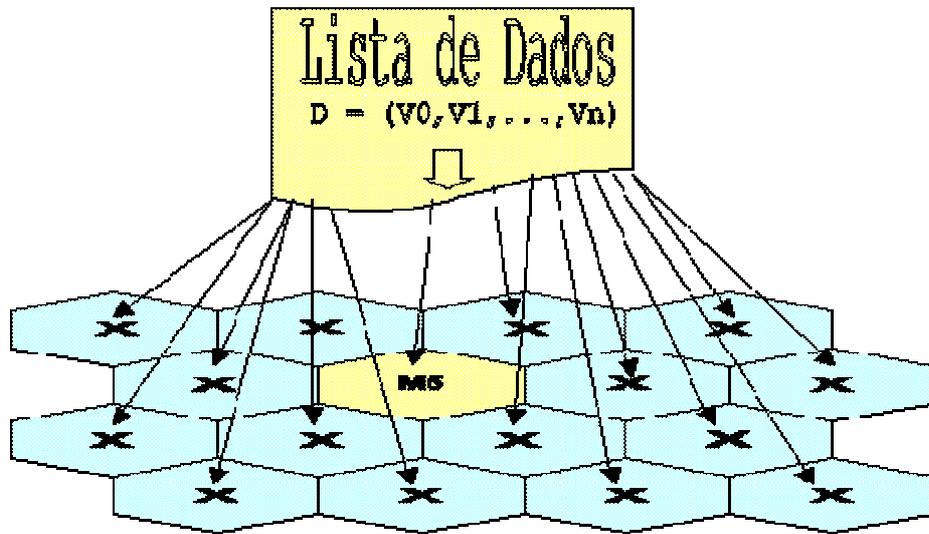


Figura 3.4 - Representação do processo competitivo, todos os modelos recebem a mesma informação, mas apenas um é ativado.

2 - Cooperação: o neurônio vencedor determina a localização espacial de uma vizinhança topológica de neurônios excitados, esta função de vizinhança deve ajustar com maior ênfase os neurônios imediatamente vizinhos ao vencedor e decair com a distância lateral. Esta função deve ser simétrica e possuir amplitude máxima em relação ao centro e decrescer com a distância lateral até zero - condição necessária para a convergência. Conforme exemplificado na Figura 3.5, uma função de vizinhança aplicada em um mapa hexagonal 4x4.

A mais utilizada é a função de vizinhança gaussiana, descrita na equação (4.2), a qual é invariante à translação, isto é, independente da localização do neurônio vencedor (CASTRO; CASTRO, 2001).

$$h_{kj(t)}(i) = \exp\left(-\frac{d_{kj}^2}{2\sigma(i)^2}\right) \quad (3.2)$$

onde :

$h_{kj(t)}(i)$ = função de vizinhança para o vizinho (k) em relação ao vencedor (l)

para o dado (t) na iteração (i), determina o nível de cooperação;

d_{kj} = distância lateral entre o vizinho (k) e o neurônio vencedor (l);

$\sigma(i)$ = largura da vizinhança (abrangência) na iteração (i).

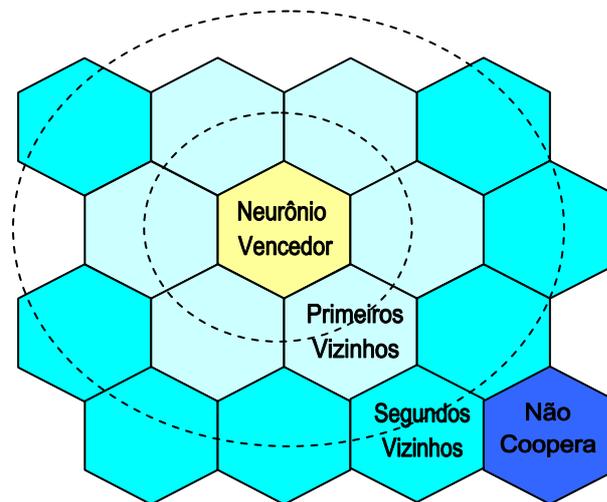


Figura 3.5 - Função de vizinhança aplicada em um mapa hexagonal 4x4.

3 - Adaptação Sináptica: quando o vetor de pesos sináptico é modificado em relação ao vetor de entrada. Os ajustes feitos são tais que, a resposta do neurônio vencedor à subsequente aplicação de um padrão similar de entrada é realçada.

A taxa de aprendizagem também deve diminuir ao longo das iterações, o que corresponde ao caso da aproximação estocástica, começando de um valor inicial e decrescendo gradualmente com o aumento das iterações. Pode ser aplicado o decaimento exponencial.

3.4 Exemplo do uso da SOM para classificação de assinaturas de sinais

A interpretação do resultado final do algoritmo de auto-organização é uma das etapas mais importantes para a correta obtenção dos resultados, e para extrair o máximo de informações que esse método proporciona.

O processo de análise, a princípio, estava voltado para a divisão por cluster, a qual é uma técnica exploratória de dados que tem por objetivo formar agrupamentos de objetos semelhantes a partir de um banco de dados. O conceito de clusterização difere do conceito de classificação no sentido que a análise de cluster é mais “primitiva”, na qual nenhuma suposição é feita a respeito dos grupos, assim como seu número e estrutura. Após o processo de organização, cada elemento da base de dados utilizada para treinar o mapa será representado por um, e apenas um, dos neurônios da rede. Porém, cada neurônio poderá representar qualquer quantidade de elementos (inclusive nenhum). Neste caso, a representação do neurônio vencedor e dos elementos que este representa mostrou-se mais importante que a própria análise dos clusters.

Para poder entender o processo, foi elaborado um exemplo da ordenação realizado pelo algoritmo SOM e a sua forma de visualização, visando à extração das características que levaram a tal organização. A idéia principal foi a de entender a aplicação da rede SOM, visualizando os resultados a partir de curvas previamente conhecidas.

Foram escolhidos três modelos de curvas bastante conhecidas e diferentes (seno – co-seno - tangente). Foram separadas de seis a sete curvas de cada (vinte curvas no total) com 63 pontos cada uma (pontos de formação de cada curva), e aplicada a ferramenta SOM.

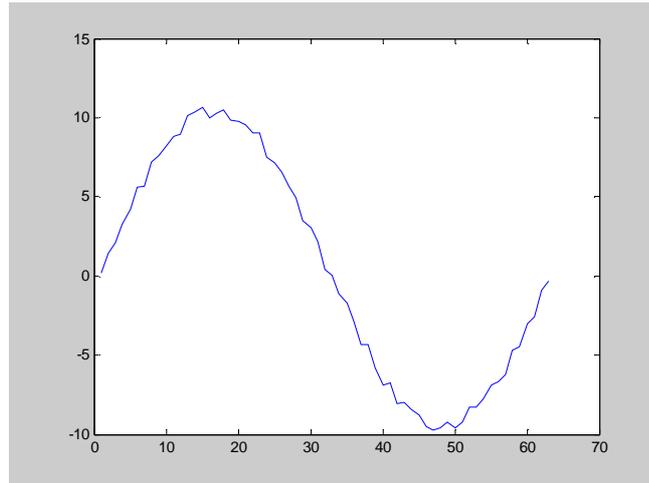


Figura 3.6 – Linha 1 – Função Seno – 63 Registros

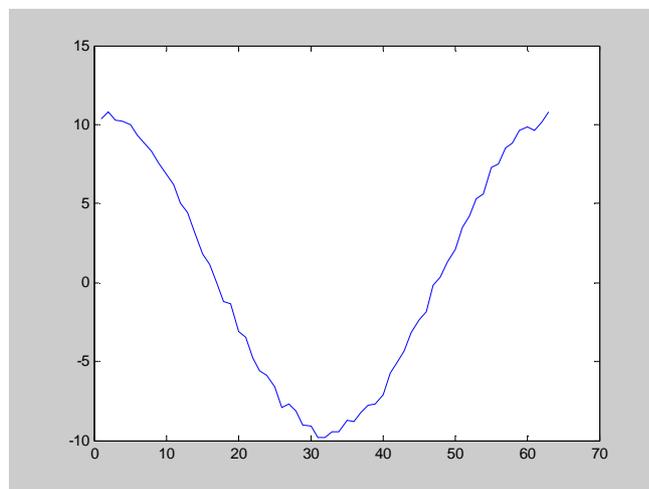


Figura 3.7 – Linha 2 – Função Co-seno – 63 Registros

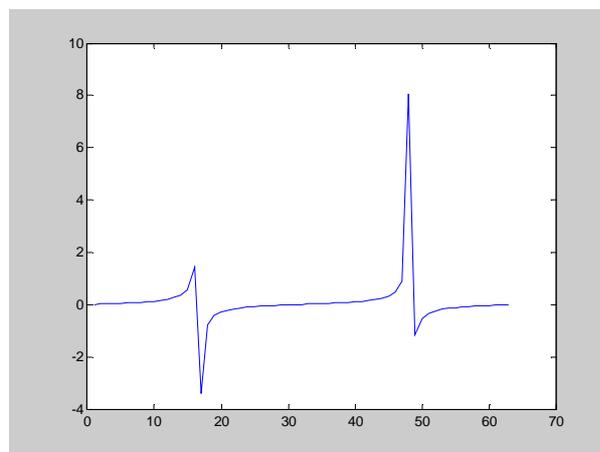


Figura 3.8 – Linha 3 – Função Tangente – 63 Registros

A Figura 3.9 a seguir, mostra o exemplo da SOM para um mapa 16 x 64 usado para classificar uma base de dados gerada a partir das curvas das Funções Seno, Co-seno e Tangente em um espaço tridimensional .

Como se pode observar, a classificação foi totalmente correta a partir de uma base de dados gerada propositalmente com três grupos, validando assim o uso da SOM para agrupamento de dados.

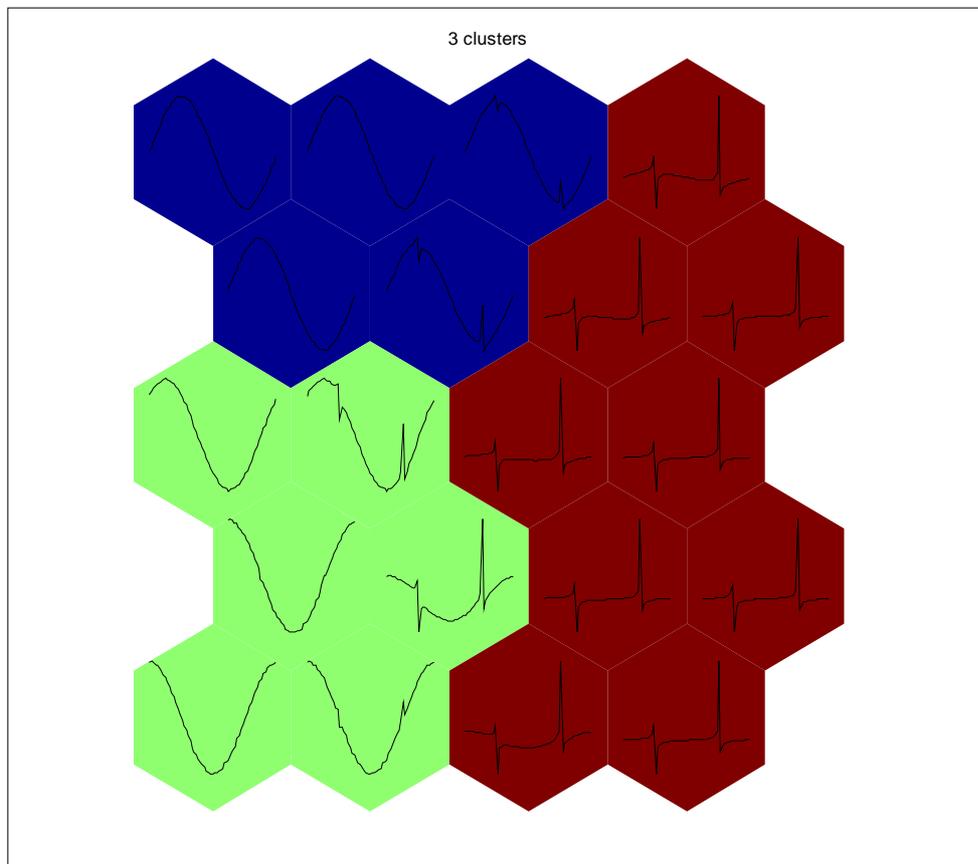


Figura 3.9- Classificação da rede SOM, com uma base de dados gerada a partir das curvas das Funções Seno, Co-seno e Tangente.

4 - Metodologia proposta para o desenvolvimento do trabalho

A metodologia proposta para o desenvolvimento do trabalho seguirá duas (2) etapas principais. A primeira etapa baseia-se no levantamento das memórias de massa dos clientes, o levantamento das memórias de massa dos clientes fraudadores, a formatação das memórias de massa, isto é, o ajuste do banco de dados, a escolha do método de análise pelo especialista da área, finalizando com a aplicação da ferramenta – SOM.

A segunda etapa baseia-se no teste da ferramenta e do método escolhido, com o levantamento de novas memórias de massa de clientes, de clientes fraudadores e a sua aplicação na rede SOM já treinada e salva com uma análise final dos resultados obtidos.

4.1 - Levantamento das memórias de massa dos clientes.

Nesta etapa, foram separadas todas as memórias de massa disponíveis dos clientes de Alta Tensão Horo Sazonais, para os meses de Janeiro e Fevereiro de 2005.

Foram separadas todas as memórias de massa que efetivamente foram utilizadas para o faturamento dos clientes junto à concessionária estudada, fazendo deste modo a utilização de uma base de dados consistente. No total foram separados 726 arquivos de memória de massa.

Na primeira etapa, foram utilizados 580 arquivos de memória de massa, devido à escolha do método especialista, isto é os clientes Industriais e Comerciais, na segunda etapa foram separadas mais 136 memórias de massa de clientes para os testes finais.

4.2 - Levantamento das memórias de massa dos clientes fraudadores.

Nesta etapa, foram separadas todas as memórias de massa disponíveis dos clientes de Alta Tensão Horo Sazonais da concessionária estudada, que efetivamente foram identificados com fraudes das mais diversas modalidades.

Esta etapa foi uma das mais críticas de todo o processo, pois as fraudes nos clientes de Alta Tensão são mais difíceis de acontecerem, de serem identificadas e principalmente de serem comprovadas. Algumas fraudes são identificadas, mas acabam não gerando um arquivo de memória de massa completo que possam ser utilizados na análise, são fraudes mais grosseiras onde literalmente são “zerados” os dados disponibilizados pelo medidor.

Com esta restrição, para a primeira etapa, só foi conseguido separar e trabalhar com 8 (oito) arquivos de memória de massa de clientes comprovadamente fraudadores.

Para a segunda etapa, onde foram testados a rede SOM e o método escolhido para análise, foi necessário a criação de 4 (quatro) arquivos de memórias de massa com características de fraudadores. Foi utilizado neste caso o conhecimento dos especialistas da área, buscando abranger a maior gama possível de efeitos dos principais tipos de fraude já encontrados em uma memória de massa de clientes atendidos em Alta Tensão.

4.3 - Escolha do método de análise – especialista.

Inicialmente, foram separadas as amostras em cinco categorias: Industrial, Comercial, Rural, Poder Público (incluídas nesta: Serviço Público e concessionárias de serviço público) e Clientes Livres. Cada uma destas categorias possui características próprias de consumo, demanda e mesmo de relacionamento contratual com a concessionária estudada.

Esta divisão se baseou na Resolução 456 de 29/11/2000 da AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL, onde no Artigo nº 18, é

apresentada a Classificação e Cadastro das unidades consumidoras e na expertise dos especialistas na área. A principal motivação desta separação deve-se ao fato das diferenças marcantes nos históricos de consumo, ao longo dos meses, destas diversas categorias.

Como as chances de fraude são praticamente nulas nos clientes classificados como Poder Público e Clientes Livres, além do fato de que os clientes Rurais possuem uma sazonalidade de funcionamento, em decorrência da própria atividade, a análise, neste trabalho, terá como alvo as memórias de massa dos clientes Industriais e Comerciais.

4.4 - Formatação das memórias de massa.

Após a separação em categorias, as memórias de massa foram abertas para análise por um programa específico denominado PLAWIN Versão 3,13/2002 desenvolvido pela *ESB Electronic Services* e amplamente utilizado pelas concessionárias de energia elétrica. Na Figura 4.1 é apresentada a tela do programa PLAWIN durante a abertura de uma memória de massa.

Após abertas, as tabelas foram salvas como planilhas do Microsoft Excell, sendo que cada uma foi identificada com um nome que facilitasse o reconhecimento de sua categoria e a que cliente pertencia. Neste caso foi adotado o seguinte código:

- Letra inicial = Categoria (I para Industrial, C para Comercial, R para Rural, P para Poder Público e L para Livre)

- Seis caracteres numéricos = Identificando o medidor associado 000000
- Espaço
- Mês correspondente ao período de medição da memória de massa = jan / fev

Modelo: I000241 fev.xls - Memória de massa do cliente com medidor nº 241, pertencente à classe Industrial e ao período de fevereiro/2005.

Registro	Data	Hora	Canal 1	W	Canal 2	varIND	Canal 3	varCAP	SH	SR	Fat.Pot
3	23/2/2007	09:15:00	334	24	135	10	0	0	F	L	93 L
6	23/2/2007	09:30:00	346	25	137	10	0	0	F	L	93 L
9	23/2/2007	09:45:00	377	27	138	10	0	0	F	L	94 L
12	23/2/2007	10:00:00	371	27	143	10	0	0	F	L	93 L
15	23/2/2007	10:15:00	424	31	163	12	0	0	F	L	93 L
18	23/2/2007	10:30:00	414	30	157	11	0	0	F	L	94 L
21	23/2/2007	10:45:00	424	31	157	11	0	0	F	L	94 L
24	23/2/2007	11:00:00	416	30	156	11	0	0	F	L	94 L
27	23/2/2007	11:15:00	417	30	156	11	0	0	F	L	94 L
30	23/2/2007	11:30:00	393	28	148	11	0	0	F	L	94 L
33	23/2/2007	11:45:00	404	29	149	11	0	0	F	L	94 L
36	23/2/2007	12:00:00	403	29	144	10	0	0	F	L	94 L
39	23/2/2007	12:15:00	395	28	142	10	0	0	F	L	94 L
42	23/2/2007	12:30:00	319	23	120	9	0	0	F	L	94 L
45	23/2/2007	12:45:00	243	17	95	7	0	0	F	L	93 L
48	23/2/2007	13:00:00	176	13	48	3	0	0	F	L	96 L
51	23/2/2007	13:15:00	167	12	42	3	0	0	F	L	97 L
54	23/2/2007	13:30:00	147	11	41	3	0	0	F	L	96 L
57	23/2/2007	13:45:00	125	9	36	3	0	0	F	L	96 L
60	23/2/2007	14:00:00	112	8	37	3	0	0	F	L	95 L
63	23/2/2007	14:15:00	97	7	33	2	0	0	F	L	95 L
66	23/2/2007	14:30:00	94	7	29	2	0	0	F	L	96 L
69	23/2/2007	14:45:00	97	7	33	2	0	0	F	L	95 L

Figura 4.1 - Tela do Programa PLAWIN

Após esta etapa, as memórias de massa foram abertas, uma a uma, e colocadas na mesma base escolhida – kW (Kilowatts), isto porque algumas das memórias de massa apresentaram informações em W (Watts), o que poderia distorcer a análise.

A segunda etapa envolveu a normalização do banco de dados. Considerando que os medidores iniciam o armazenamento dos dados na memória de massa assim

que o medidor é lido em campo, e que esta leitura pode ser realizada em qualquer dia da semana e horário, uma análise que leve em conta as mesmas características deve considerar estas diferenças de consumo de dias da semana e horário.

Por exemplo, em uma indústria, a característica de consumo das segundas feiras às 14:00 h são diferentes dos consumos das 2:00 horas da madrugada dos domingos.

Assim, foi desenvolvida uma ferramenta em Excel para normalizar os dias e horários de consumo. Nesta ferramenta, ao entrar com a memória de massa são aplicados algumas macros, separando os consumos por dia da semana e por horários.

Ao ser aberta uma memória de massa, encontramos entre 3 a 5 registros de 15 minutos (por exemplo, intervalos de 14:00 hs até 14:15 hs) relacionadas a 3 ou 5 segundas feiras compreendidas no mês. Ao rodar o programa, foi separado por média aritmética, o perfil de consumo de cada memória de massa para cada dia na semana. Deste modo, foi levantado um consumo médio para cada dia da semana e para cada cliente separadamente, característica esta definida e esperada pelo especialista. Deste modo, foi elaborado para cada categoria de consumo, 7 (sete) tabelas, uma para cada dia da semana, com suas características de consumo médio diário, pode-se ver um resumo na Tabela 4.1. Para identificar todas as memórias de massa salvas, foi utilizado o seguinte método para guardar as tabelas resultantes:

IJF SEG.xls – Tabela com características de consumo das segundas feiras dos clientes Industriais para os meses de Janeiro e Fevereiro. Resultando no caso em uma tabela de 313 linhas por 96 colunas. Isto é, 313 clientes com 96 registros médios da Segunda Feira. O número de registros é devido a: 96 registros – Registros de 15 minutos em 24 horas ($24 \times 4 = 96$).

Tabela 4.1 – Tabela com características de consumo dos dias da semana dos clientes Comerciais, Industriais, Rurais, Livres, Poder Público e Serviço Público para os meses de Janeiro e Fevereiro.

Comercial	Tamanho	Livres	Tamanho	Poder Público	Tamanho
CFJ SEG	267 x 96	LFJ SEG	8 x 96	PFJ SEG	66 x 96
CFJ TER		LFJ TER		PFJ TER	
CFJ QUA		LFJ QUA		PFJ QUA	
CFJ QUI		LFJ QUI		PFJ QUI	
CFJ SEX		LFJ SEX		PFJ SEX	
CFJ SAB		LFJ SAB		PFJ SAB	
CFJ DOM		LFJ DOM		PFJ DOM	
Industrial	Tamanho	Rural	Tamanho	Serviço Público	Tamanho
IFJ SEG	313 x 96	RFJ SEG	81 x 96	SFJ SEG	56 x 96
IFJ TER		RFJ TER		SFJ TER	
IFJ QUA		RFJ QUA		SFJ QUA	
IFJ QUI		RFJ QUI		SFJ QUI	
IFJ SEX		RFJ SEX		SFJ SEX	
IFJ SAB		RFJ SAB		SFJ SAB	
IFJ DOM		RFJ DOM		SFJ DOM	

Considerando a análise deste trabalho tendo como alvo os clientes Industriais, Comerciais, montamos uma tabela onde colocamos estas duas categorias juntas, observando a divisão por dias da semana. Ficando resumido à 7 (sete) tabelas conforme mostrado na Tabela 4.2:

Tabela 4.2 – Tabela com características de consumo dos dias da semana dos clientes Comerciais e Industriais para os meses de Janeiro e Fevereiro.

Categoria	Nome		Tamanho	Nome
Comercial	CFJ SEG		580 x 96	CIFJ SEG
Industrial	IFJ SEG			CIFJ TER
Comercial	CFJ TER			CIFJ QUA
Industrial	IFJ TER			CIFJ QUI
Comercial	CFJ QUA			CIFJ SEX
Industrial	IFJ QUA			CIFJ SAB
Comercial	CFJ QUI			CIFJ DOM
Industrial	IFJ QUI			
Comercial	CFJ SEX			
Industrial	IFJ SEX			
Comercial	CFJ SAB			
Industrial	IFJ SAB			
Comercial	CFJ DOM			
Industrial	IFJ DOM			

Deste modo, foram reduzidas todas as tabelas em apenas 7 memórias de massa, cada uma contendo, em seqüência, os clientes comerciais e industriais devidamente identificados. Com estes arquivos, foi construída e treinada a rede SOM para cada dia da semana.

Na segunda etapa deste trabalho, foram separados mais 136 arquivos de clientes Industriais e Comerciais para aplicar o teste da ferramenta SOM. Foi utilizado o mesmo processo de abertura, normalização e armazenagem descrito na primeira etapa. Neste arquivo foram incluídas quatro memórias de massa dos clientes fraudadores,

especialmente criadas para o teste. O resultado pode ser visto na tabela 4.3 abaixo com a formatação utilizada.

Tabela 4.3 – Tabela com os arquivos utilizados para os testes de aplicação da SOM.

Categoria	Nome		Tamanho	Nome	Tamanho com Fraudadores	Nome
Comercial	CFJT SEG	⇒	136 x 96	CIFJT SEG	140 x 96	CIFJTF SEG
Industrial	IFJT SEG			CIFJT TER		CIFJTF TER
Comercial	CFJT TER			CIFJT QUA		CIFJTF QUA
Industrial	IFJT TER			CIFJT QUI		CIFJTF QUI
Comercial	CFJT QUA			CIFJT SEX		CIFJTF SEX
Industrial	IFJT QUA			CIFJT SAB		CIFJTF SAB
Comercial	CFJT QUI			CIFJT DOM		CIFJTF DOM
Industrial	IFJT QUI					
Comercial	CFJT SEX					
Industrial	IFJT SEX					
Comercial	CFJT SAB					
Industrial	IFJT SAB					
Comercial	CFJT DOM					
Industrial	IFJT DOM					

4.5 - Aplicação da Ferramenta – SOM

O mapa auto-organizável foi gerado e treinado a partir de um pacote de programas para MATLAB [SOM toolbox]. Esse pacote permite ajustar todos os parâmetros do SOM, mas a princípio foi utilizada uma função chamada *som_make*, que gera o mapa, inicia e faz o treinamento.

Foi utilizada a configuração padrão desse algoritmo, isto é uma inicialização linear, que geralmente leva a uma mesma topologia ao final do treinamento se não houver alteração da base de dados, vizinhança hexagonal, ou seja, os neurônios centrais estão ligados a seis vizinhos, e uma função de vizinhança gaussiana. A base de dados, para entrada no algoritmo foi composta por uma matriz de dimensão 580 x 96, onde as linhas representam as memórias de massa dos clientes e as colunas o valor médio do intervalo de 15 minutos de integralização da medição por dia.

Uma vez definido o universo de tabelas a estudar, iniciou-se a aplicação da ferramenta SOM. Pode ser observado na Figura 4.2, a tela do programa MATLAB rodando a rede SOM.

```

1 %% SOM
2 %
3 - importfile('CJFSEG.xls')
4 - data=Plan1;
5 % Padronização com média zero e variância unitária
6 - data=zeroscore(data)';
7 - % Dados usados para treino
8 - smap = som_make(data(randperm(size(data(1:267,:),1)),:), 'mapsize', 'normal', 'j');
9 %som_plotplane (smap, smap.codebook)
10
11 %%
12 % Clusterização da SOM
13 - [c,p,err, ind] = kmeans_clusters(smap);
14 - [dummy, i] = min(ind);

```

```

smap =
    type: 'som_map'
  codebook: [84x96 double]
    topol: [1x1 struct]
   labels: (84x1 cell)
   neigh: 'gaussian'
    mask: [96x1 double]
  trainhist: [1x3 struct]
    name: 'SOM 06-Aug-2007'
 comp_names: (96x1 cell)
  comp_norm: (96x1 cell)

```

Figura 4.2 – Tela MATLAB

Nesta primeira etapa, foi utilizado o MATLAB com a seguinte seqüência de trabalho: 1º Importação da Tabela a ser rodada, 2º Criação das Estruturas da SOM, 3º Clusterização da SOM, 4º Plotagem do mapa SOM, 5º Apresentação e classificação dos fraudadores, 6º Plotagem do mapa SOM com fraudadores, 7º Análise dos resultados, 8º Identificação dos consumidores nos clusters indicados pelos fraudadores e 9º Armazenamento dos dados, de acordo com o fluxograma da Figura 4.3.

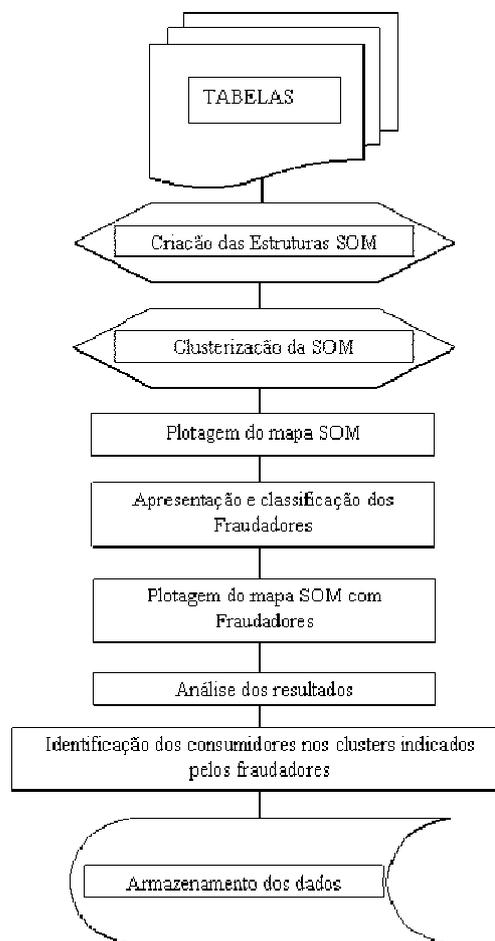


Figura 4.3 - Fluxograma da seqüência de trabalho

Foi rodado o algoritmo SOM para a classe Industrial e Comercial, para cada dia da semana. Foram encontrados e salvos os sete mapas SOM característicos correspondentes aos dias de semana. Como exemplo, pode-se visualizar a Figura 4.4, o mapa SOM para a quarta feira sem fraudadores.

Após obter o mapa SOM característico do dia da semana, foi rodado novamente o mapa com as memórias de massa dos clientes fraudadores (8) oito, formatados nos mesmos moldes das outras memórias de massa.

Resultando em sete mapas da SOM, um para cada dia da semana, com as localizações dos clientes fraudadores em seus respectivos clusters. Como exemplo, pode-se visualizar na Figura 4.5, o mapa SOM para a quarta feira com fraudadores. Foram salvos, também planilhas com os pontos de localizações dos clientes, seus clusters correspondentes e ainda os pontos de localização e clusters dos fraudadores.

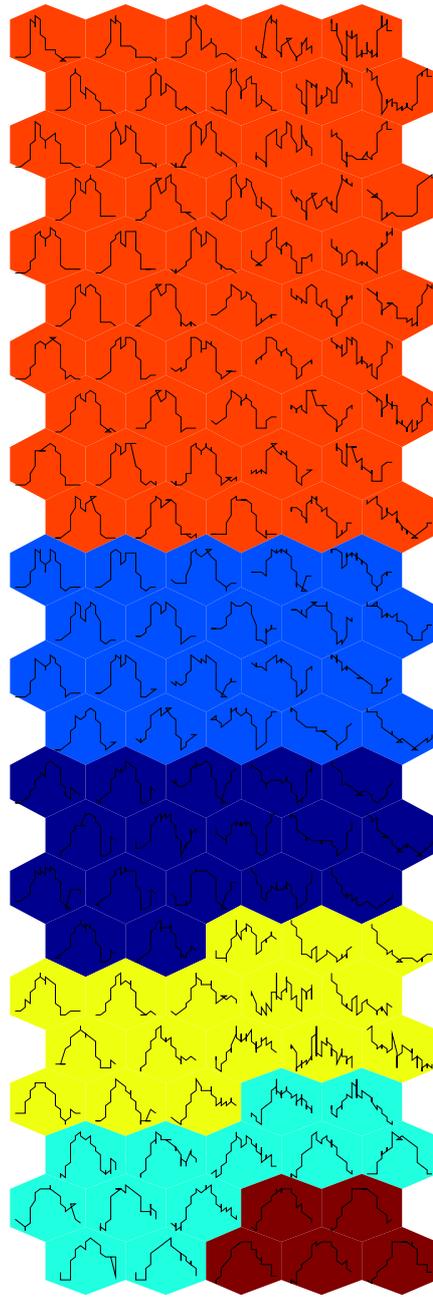


Figura 4.4 – Mapa SOM – Quarta Feira Sem Fraudadores

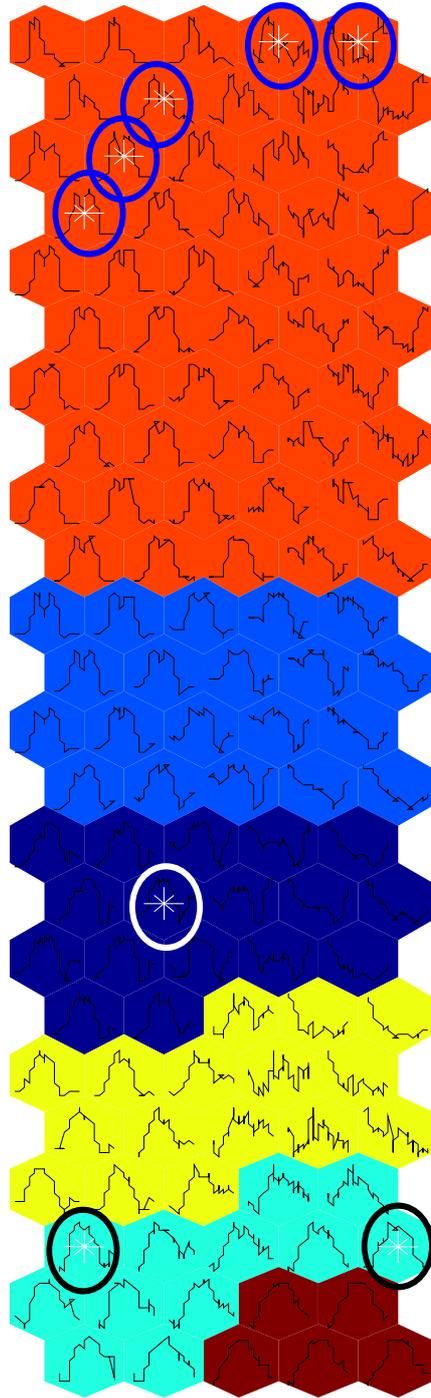


Figura 4.5 – Mapa SOM – Quarta Feira Com Fraudadores

Com estas informações, foram detectados os clusters e os pontos de localização onde apareceram os clientes e os fraudadores. Estes clusters, a princípio indicariam um universo de clientes a serem inspecionados (possíveis fraudadores).

Como o universo de memórias de massa (clientes) indicadas nos clusters para a inspeção ficou muito alto, algo em torno de 82,96 %, em relação ao número de memórias de massa de clientes Industriais e Comerciais apresentados para rodar a SOM (580 no total), ficou evidente a necessidade de uma nova investigação com os clientes que foram alocados nas mesmas posições dos fraudadores. Foi verificado que além de dividir os clientes por clusters (entre 5 e 7 dependendo do dia da semana), a SOM alocou os 580 clientes em torno de 120 posições específicas.

Ao rodar os clientes fraudadores, estes indicaram além dos clusters, as posições coincidentes com os outros clientes, deste modo proporcionando a possibilidade de diminuir o universo de pesquisa dos possíveis clientes com características de fraudadores para 13,13 % (médios) para cada dia da semana.

Continuando nesta linha, os estudos foram aprofundados em relação à alocação dos fraudadores junto aos clientes para todos os dias da semana, isto é uma possível relação entre todas as redes SOM.

Foram encontradas então algumas relações importantes, resumidas na tabela 4.4, que possibilitaram diminuir ainda mais o universo de clientes que apresentam as mesmas características dos fraudadores.

Por exemplo, para clientes que foram alocados na mesma posição dos fraudadores e que apareceram em pelo menos uma vez nos sete dias da semana, foi encontrado um percentual de 40,17%.

Abrindo este percentual e analisando o número de vezes coincidentes na semana em que os clientes aparecem, junto aos fraudadores nas redes SOM, foram encontradas as seguintes relações percentuais que podem ser visualizadas na Tabela 4.4:

- em apenas um dia na semana - percentual de 18,79 %.
- em dois dias na semana - percentual de 9,31 %.
- em três dias na semana - percentual de 4,14 %.
- em quatro dias na semana - percentual de 3,45 %.
- em cinco dias na semana - percentual de 1,21 %.
- em seis dias na semana - percentual de 0,69 %.
- em sete dias na semana - percentual de 2,59 %.

Tabela 4.4 – Análise dos Clientes x Clientes Fraudadores

Clientes								
Total	Nº dos que apareceram pelo menos 1 vez	Nº de vezes que apareceram na semana						
		1	2	3	4	5	6	7
580	233	109	54	24	20	7	4	15
100 %	40,17%	18,79%	9,31%	4,14%	3,45%	1,21%	0,69%	2,59%

Os resultados preliminares indicaram os “clusters” prováveis de clientes com características de fraudadores, conforme esperado. Como alguns “clusters” resultantes foram os que apresentaram um número maior de clientes, a análise do universo de possíveis fraudadores se mostrou maior do que o esperado.

Deste modo, a análise voltou-se para os locais coincidentes de clientes e fraudadores reduzindo o universo de busca a níveis mais aceitáveis para o trabalho proposto.

O próximo passo será o de apresentar novos clientes (novas memórias de massa) à rede SOM já treinada, contendo alguns clientes fraudadores previamente conhecidos (criados). A análise dos pontos de localização destes clientes dentro dos clusters irá permitir avaliar se as análises são convergentes ou não.

4.6 – Criação de memória de massa de clientes fraudadores

Dada a inexistência de memória de massa de clientes fraudadores para avaliar a rede SOM construída e verificar sua aplicação na prática, foi necessário construir memórias de massa com o mesmo perfil das memórias de massa dos clientes fraudadores conhecidos.

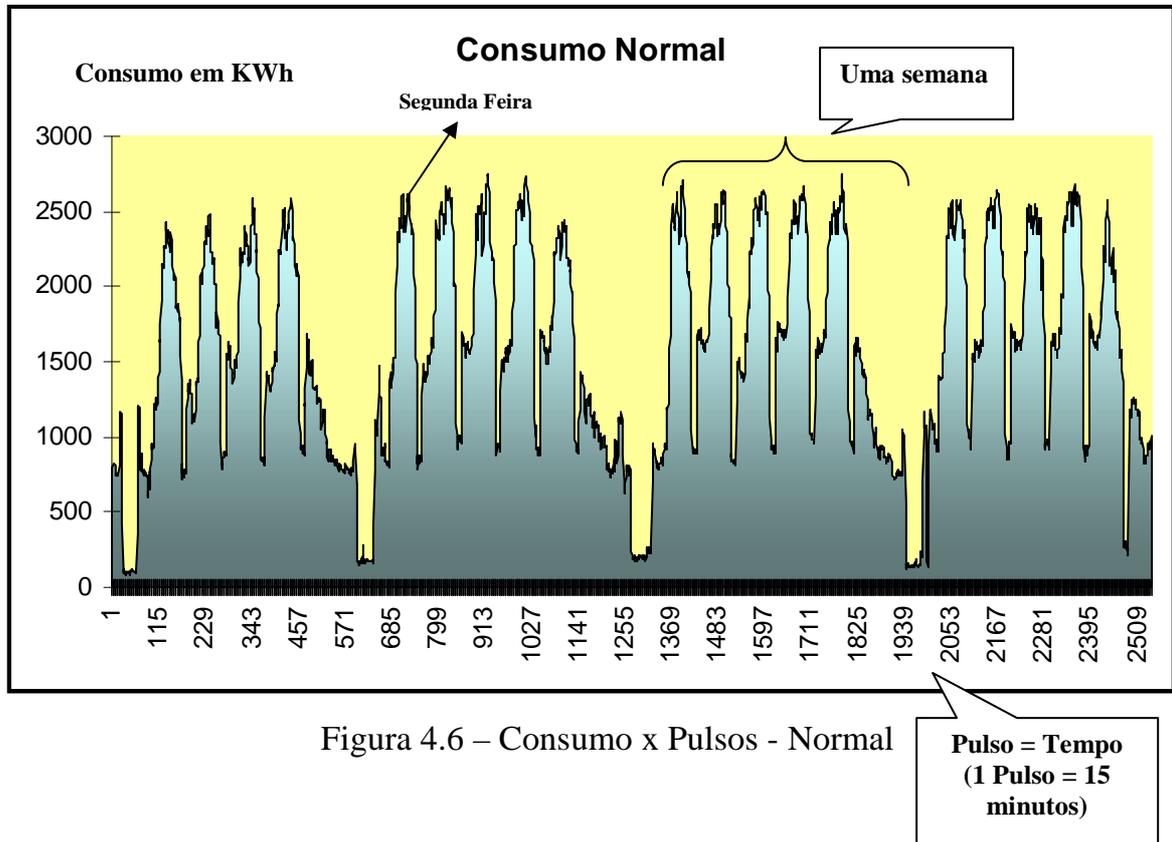
A dificuldade de encontrar memórias de massa de clientes fraudadores se deve a dois fatos principais:

- O número de clientes fraudadores atendidos em alta tensão detectados é mínimo no universo total de clientes atendidos por uma concessionária.
- As memórias de massa dos clientes pegos com uma fraude, nem sempre estão intactas, isto é, muitas vezes a fraude corrompeu o arquivo ou “zerou” as medições, fazendo com que estas memórias de massa não indiquem uma característica de fraude, mas sim de defeito na medição.

Foram utilizadas as seguintes premissas para a construção das quatro memórias de massa com características de fraudadores:

1º – Utilização de quatro arquivos de memórias de massa distintas e com características de consumo normal. Sendo dois arquivos de clientes comerciais e dois arquivos de clientes industriais. A Figura 4.6 apresenta um gráfico de consumo em

função dos pulsos (tempo) de uma memória de massa com características de consumo normal.



2º – Manteve-se a primeira semana com consumo normal, isto é, sem alterar qualquer dado em dois arquivos e nos outros dois manteve-se a primeira e segunda semanas com consumos normais.

A escolha de como seriam “manipuladas” as memórias de massa com características de fraudadores, foi baseada nos principais resultados encontrados em campo, quando da descoberta e comprovação de uma fraude.

Em um dos arquivos, foram reduzidos os valores de consumo em 1/3 após uma semana de consumo normal, caracterizando o desligamento/rompimento da comunicação do sinal de uma fase para o medidor. Característica de fraude externa ao

medidor, sendo provavelmente uma intervenção na fiação ou nos equipamentos auxiliares de medição. A Figura 4.7 abaixo representa este arquivo.

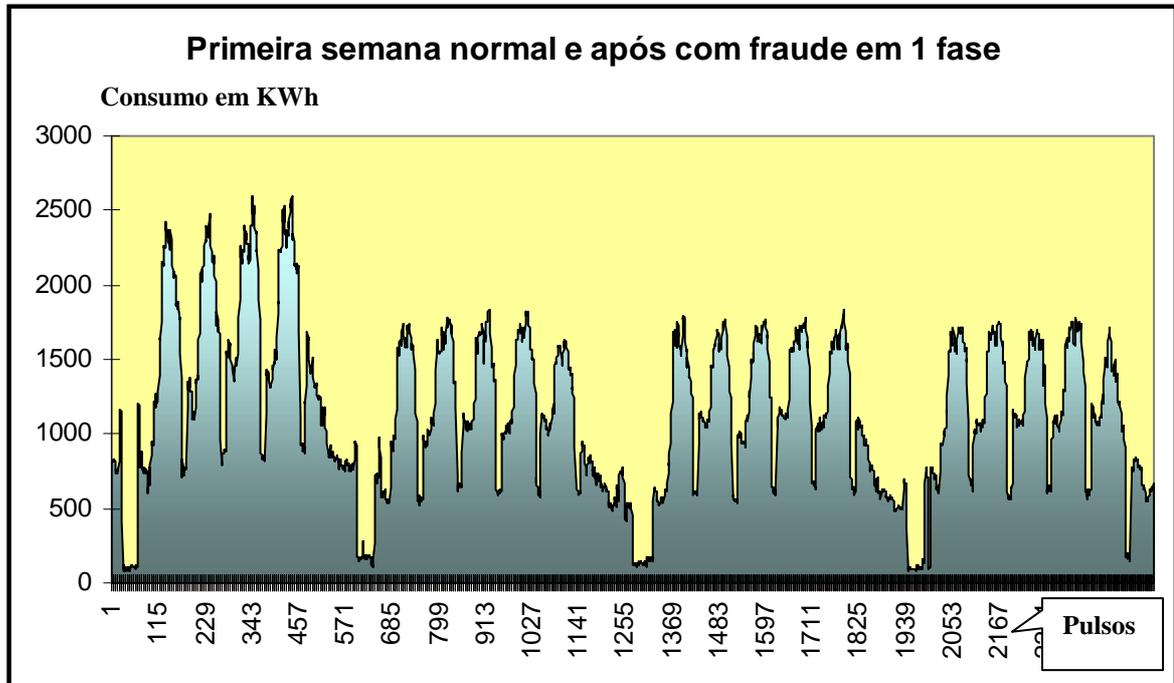


Figura 4.7 – Consumo x Pulsos – 1º semana normal e após com fraude em 1 fase (1/3 de redução)

Em outro arquivo, foram reduzidos os valores de consumo em 2/3 após uma semana de consumo normal, caracterizando o desligamento/rompimento da comunicação do sinal de duas fases para o medidor. Característica esta de intervenção parecida com a anterior, mas com resultados mais drásticos em termos de redução de consumo. A Figura 4.8 abaixo representa este arquivo.

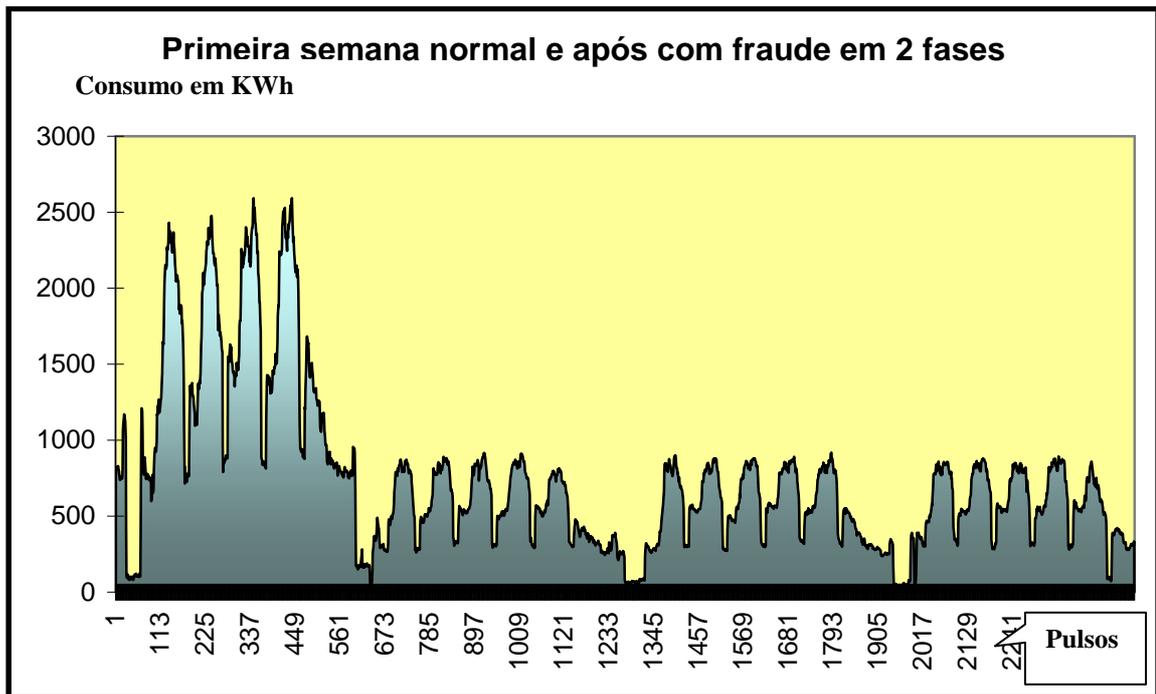


Figura 4.8 – Consumo x Pulsos – 1º semana normal e após com fraude em 2 fases (2/3 de redução)

Os outros dois arquivos utilizaram os mesmos critérios anteriores, mas com uma semana a mais de consumo normal. As Figuras 4.9 e 4.10 representam estes arquivos.

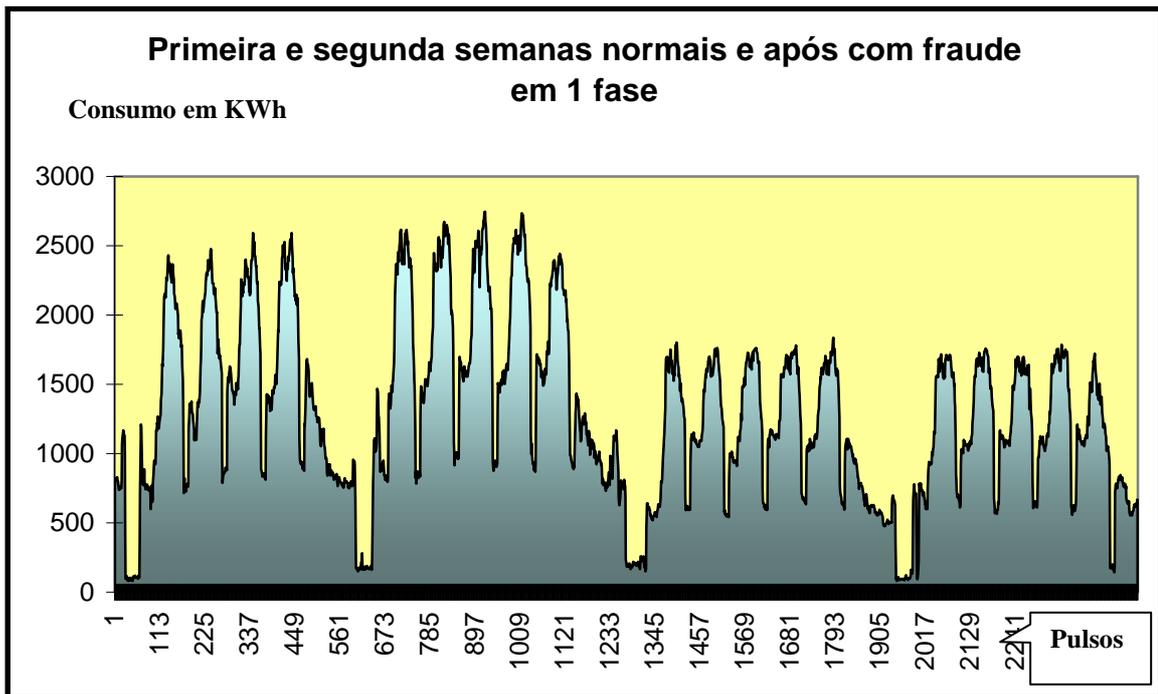


Figura 4.9 – Consumo x Pulsos – 1º e 2º semanas normais e após com fraude em 1 fase (1/3 de redução)

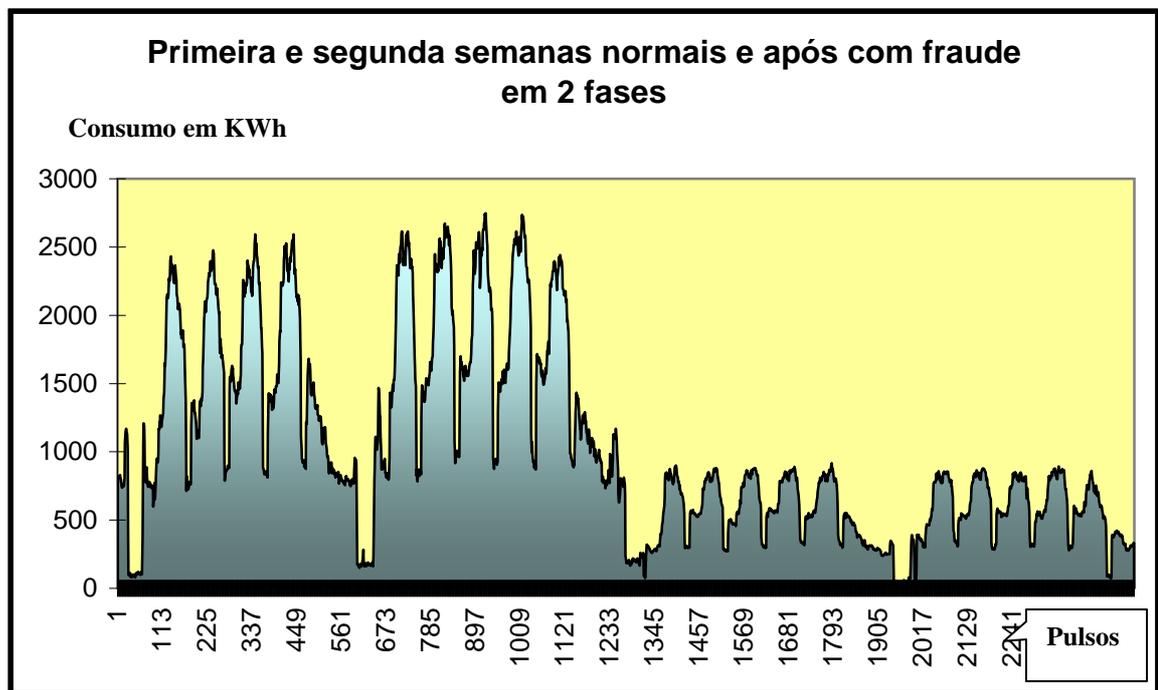


Figura 4.10 – Consumo x Pulsos – 1º e 2º semanas normais e após com fraude em 2 fases (2/3 de redução)

5 – Simulações e Resultados

De posse dos mapas SOM característicos de cada dia da semana, incluso as memórias de massa dos clientes fraudadores trabalhado na primeira etapa deste projeto, foi apresentado, para cada dia da semana, uma nova planilha contendo as memórias de massa de 132 clientes não fraudadores e 4 novos clientes fraudadores (arquivos construídos).

Resultando em sete novos mapas da SOM, um para cada dia da semana, com as localizações dos clientes fraudadores conhecidos na primeira etapa do trabalho, em seus respectivos clusters e a localização dos novos clientes fraudadores (arquivos construídos) e seus respectivos clusters.

Como exemplo, pode-se visualizar na Figura 5.1, o mapa SOM para a quarta feira com fraudadores. Foram salvos, também planilhas com os pontos de localizações dos clientes, seus clusters correspondentes e ainda os pontos de localização e clusters dos fraudadores.

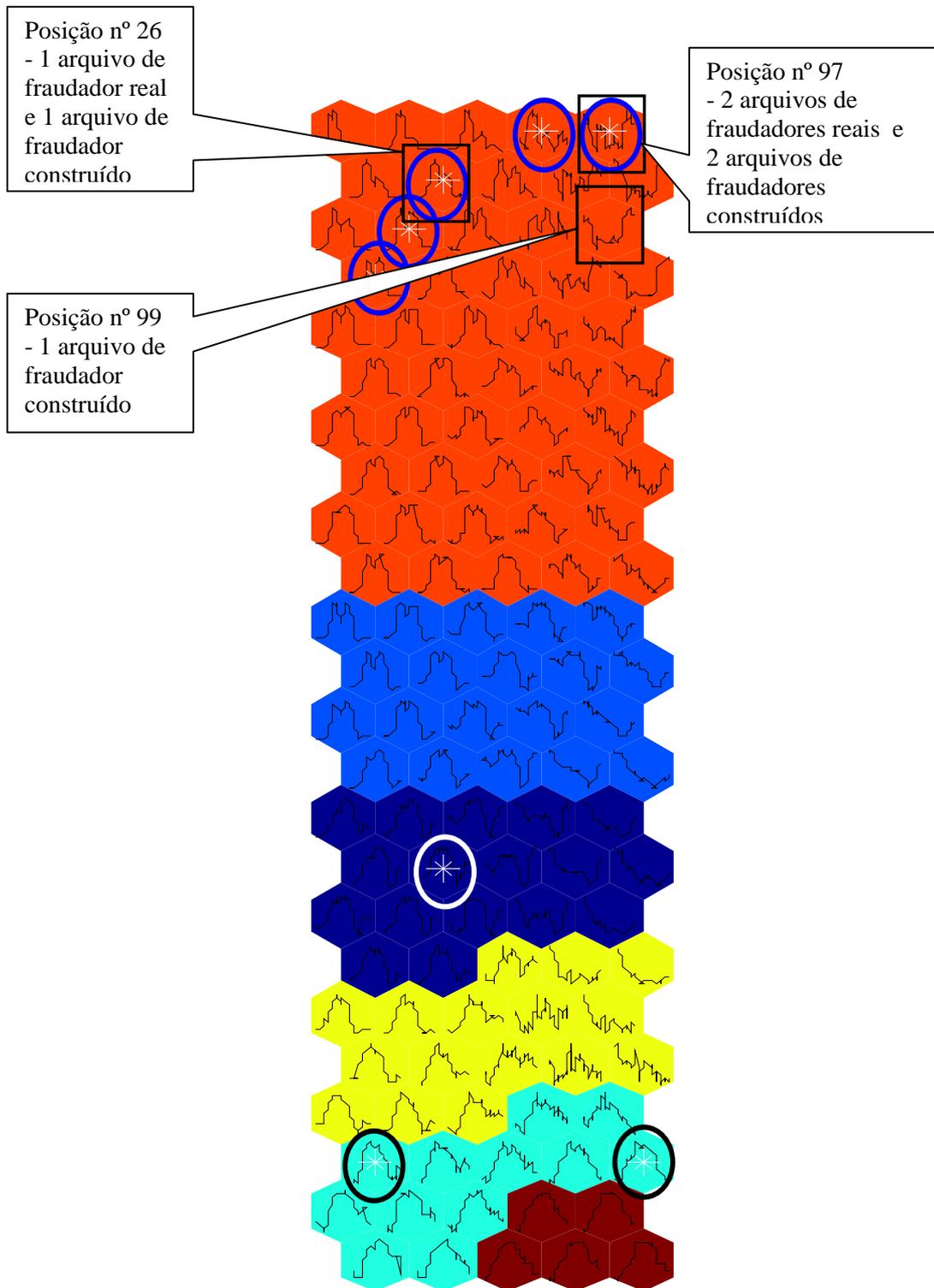


Figura 5.1 – Mapa SOM – Quarta Feira Com Fraudadores Reais e 4 Novos Fraudadores Construídos

Como já era esperado, em todos os dias da semana, os novos fraudadores criados foram alocados nos clusters de maior incidência, isto é, os clusters indicados anteriormente como possuindo as características de possíveis fraudadores. Com relação à análise mais fina, das posições, foi encontrado para os 136 clientes apresentados na quarta feira, por exemplo, 23 clientes coincidentes das posições dos fraudadores conhecidos sendo 3 destes os novos clientes fraudadores. A tabela 5.1 apresenta os resultados obtidos para os sete dias da semana.

Tabela 5.1 – Análise da apresentação dos novos clientes incluindo os novos fraudadores criados

Seg	SOM treinanda (580 + 8 fraudadores)								Soma	%	Nº de Acertos	
	Posição Fraudadores Reais **	120	72	46	23	4	3	60				42
	Quantidade Clientes nas posições	59	16	6	11	1	0	4	97	16,72%		
	Apresentação de novos clientes (132 + 4 fraudadores)											
	Quantidade na Posição **	9	7	1	3	0	0	0	20	14,71%		
	Quantid. Fraud. Criados na Posição	2	1	0	0	0	0	0	4		3	
Ter	SOM treinanda (580 + 8 fraudadores)											
	Posição Fraudadores Reais **	102	54	99	51	120	47	18	81			
	Quantidade Clientes nas posições	47	9	7	6	0	0	5	74	12,76%		
	Apresentação de novos clientes (132 + 4 fraudadores)											
	Quantidade na Posição **	10	2	0	2	0	0	1	15	11,03%		
	Quantid. Fraud. Criados na Posição	2	1	0	0	0	0	0	4		3	
Qua	SOM treinanda (580 + 8 fraudadores)											
	Posição Fraudadores Reais **	97	26	27	3	20	72	63	99			
	Quantidade Clientes nas posições	72	3	2	10	1	0	2	90	15,52%		
	Apresentação de novos clientes (132 + 4 fraudadores)											
	Quantidade na Posição **	18	1	0	4	0	0	0	23	16,91%		
	Quantid. Fraud. Criados na Posição	2	1	0	0	0	0	0	4		3	
Qui	SOM treinanda (580 + 8 fraudadores)											
	Posição Fraudadores Reais **	120	20	71	46	98	73	110	70			
	Quantidade Clientes nas posições	63	13	6	11	0	1	8	102	17,59%		
	Apresentação de novos clientes (132 + 4 fraudadores)											
	Quantidade na Posição **	15	3	3	2	0	1	0	24	17,65%		
	Quantid. Fraud. Criados na Posição	2	0	1	0	0	0	0	4		3	
Sex	SOM treinanda (580 + 8 fraudadores)											
	Posição Fraudadores Reais **	97	101	99	76	5	87	46	75			
	Quantidade Clientes nas posições	65	14	12	5	12	2	0	110	18,97%		
	Apresentação de novos clientes (132 + 4 fraudadores)											
	Quantidade na Posição **	10	4	5	1	0	0	0	20	14,71%		
	Quantid. Fraud. Criados na Posição	2	1	0	0	0	0	0	4		3	
Sáb	SOM treinanda (580 + 8 fraudadores)											
	Posição Fraudadores Reais **	3	35	33	1	65	105	25	116	30		
	Quantidade Clientes nas posições	12	5	8	45	10	3	3	2	76	13,10%	
	Apresentação de novos clientes (132 + 4 fraudadores)											
	Quantidade na Posição **	4	2	2	9	3	1	0	21	15,44%		
	Quantid. Fraud. Criados na Posição	1	0	0	2	0	0	0	4		3	
Dom	SOM treinanda (580 + 8 fraudadores)											
	Posição Fraudadores Reais **	28	25	49	97	20	46	62	29			
	Quantidade Clientes nas posições	14	9	12	46	4	0	2	87	15,00%		
	Apresentação de novos clientes (132 + 4 fraudadores)											
	Quantidade na Posição **	4	7	2	10	1	0	0	24	17,65%		
	Quantid. Fraud. Criados na Posição	0	1	0	2	0	0	0	4		3	

A exemplo das análises preliminares da primeira etapa, quando foi rodada a rede SOM juntamente com os fraudadores reais, foram encontradas algumas relações importantes após a apresentação dos novos clientes com fraudadores construídos, resumidas na tabela 5.2, que possibilitaram uma análise mais crítica dos resultados.

Por exemplo, dos 136 clientes apresentados, 60 foram alocados junto às posições dos clientes fraudadores reais, isto é, 44% apareceram pelo menos em uma vez em um dia da semana. Destes, 23 apareceram apenas em 1 dia da semana.

Tabela 5.2 – Relações apresentação dos novos clientes (e fraudadores criados)

Clientes novos apresentados com fraudadores criados (136)								
Total	Nº dos que apareceram pelo menos 1 vez	Quantidade de vezes que apareceram na semana						
		1	2	3	4	5	6	7
136	60	23	15	10	4	3	2	3
100%	44%	16,9%	11,0%	7,4%	2,9%	2,2%	1,5%	2,2%

Obteve-se uma taxa de acerto superior a 75%, isto é, a cada 4 fraudadores apresentados a rede SOM, esta indicou a inspeção em pelo menos 3 casos. Esse bom desempenho pode ser questionado devido à construção das memórias de massa dos novos clientes fraudadores, mas considerando que estas memórias de massa representam uma cópia bem real das fraudes encontradas, podemos concluir que o trabalho convergiu para os resultados esperados.

6 - Conclusão

O objetivo principal deste trabalho foi obter uma rede neural artificial tipo Mapas Auto Organizáveis (SOM – *Self Organizing Map*), capaz de apontar potenciais fraudadores a partir das memórias de massa retiradas dos medidores de clientes de alta tensão de uma concessionária de energia elétrica.

Para os clientes de alta tensão de energia elétrica, enquadrados nas tarifas horo-sazonais, as concessionárias utilizam medidores eletrônicos, com capacidade de armazenamento, em memória de massa, de informações de consumo ativo e reativo, a cada 15 minutos, de acordo com a Resolução 456 de 29/11/2000 da AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA – ANEEL.

Considerando a concessionária estudada, temos um banco de dados alimentado mensalmente com 1.400 memórias de massa (média) sendo cada uma com algo em torno de 3.000 registros, dependendo do período de medição que pode ser de 27 a 30 dias. Isso gera, após um curto espaço de tempo, um banco de dados extremamente grande e com uma quantidade de informações muitas das vezes ocultas, principalmente quanto ao possível relacionamento entre todas as memórias dos diversos clientes e suas características.

Deste modo, a idéia de se procurar características ocultas em um banco de dados de medição, principalmente características de erro ou fraude na medição utilizando a SOM, se mostrou viável.

O sistema desenvolvido, empregando a metodologia proposta, mostrou-se capaz de ajudar a solucionar problemas das perdas comerciais relacionadas à fraude e erro de medição das concessionárias de energia elétrica. Os resultados obtidos foram

considerados satisfatórios, uma vez que foram indicados 3 dos 4 clientes fraudadores misturados junto com 132 clientes com características de consumo normal

Deste modo, 75% dos fraudadores utilizados no processo de teste foram localizados e dos 132 clientes normais apresentados, 15% foram indicados para inspeção sendo 12,5% apontados indevidamente e 3% apontados corretamente.

Como o universo de fraudadores detectados na prática, bem como os utilizados no trabalho, é pequeno, a eficácia da aplicação do método fica restritos aos poucos casos existentes e exige tempo para uma melhor análise em cima de uma possível fraude detectada ou não pela ferramenta.

Como trabalhos futuros, pretende-se continuar a análise de outras memórias de massa de clientes fraudadores além da realização de inspeções em campo, de clientes que sejam alocados nas mesmas posições dos fraudadores após terem sido submetidos a rede SOM treinada.

Pretende-se ainda, cruzar os resultados dos clientes listados para inspeções com outros programas já em uso da concessionária local buscando um refinamento nos dados e uma possível convergência destas listas, possibilitando uma avaliação do uso da rede SOM treinada, mesmo sem conseguir a determinação positiva de um cliente como fraudador.

7 – Bibliografia

ASSOCIAÇÃO BRASILEIRA DOS DISTRIBUIDORES DE ENERGIA ELÉTRICA

– ABRADE, *Referências*

Bibliográficas, Brasília, acervo de 2000 à 2006

ADOLFO BAUCHSPIESS – Apostila Curso - Introdução aos Sistemas Inteligente

UNB Brasília

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *Referências*

Bibliográficas, NBR 6023. Rio de Janeiro, 1989.

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL Nota Técnica no

026/2006–SRD/SRC/SRE/ANEEL

Aleskerov, E., Freisleben, B. and Rao, B., "CARDWATCH: A Neural Network Based
Data

CABRAL, J. E.; Pinto, J. O. P.; Gontijo, E. M.; Reis, J. Rough Sets Based Fraud

Detection in Electrical Energy Consumers. WSEAS International Conference
on MATHEMATICS AND COMPUTERS IN PHYSICS, Cancun, Mexico,
Apr. 2004, 2 ed.,v.3, p. 413-416.

CABRAL, J. E.; Pinto, J. O. P.; Gontijo, E. M.; Reis, J. Fraud detection in electrical
energy consumers using rough sets. In: 2004 IEEE International Conference on
Systems, Man, and Cybernetics., p. 3625–3629, 2004.

CASTRO, F.C.C.; CASTRO, M.C.F. Redes Neurais Artificiais. Porto Alegre: PUCRS,
2001. Paginação irregular. Apostila para fins didáticos.

- CRISTIAN MARA; Detecção de Fraude e Erro de Medição em Grandes Consumidores de Energia Elétrica Utilizando Rough Sets Baseado em Dados Históricos e em Dados em Tempo, 2005
- CHON, T.-S.; PARK, Y.S.; MOON, K.H.; CHA, E.Y. Patternizing communities by using an artificial neural network. *Ecological Modelling*, v.90, p.69-78, 1996.
- DEBOECK, J.G., 1998: Financial Applications of Self-Organizing Maps, *American Heuristics Electronic Newsletter*.
- HAYKIN, S., 2001, *Redes Neurais : Princípios e Práticas*, Bookman, 2a ed.
- KOHONEN, T., 2001, *Self-Organizing Maps*; terceira edição, editora Springer.
- LIMA, W.S., de CASTRO, L.N., O LIM HISHI, T., 1999: Mapas Auto-Organizáveis Não-Paramétricos para Análise da Influência Climática em Curvas de Carga, XV Seminário Nacional de Produção e Transmissão de Energia Elétrica (SNPTEE), Foz do Iguaçu, Brasil.
- MERJA OJA, SAMUEL KSKI, KOHONEN TEUVO - Helsinki University of Technology, Neural Networks Research Centre, P.O. Box 5400, FIN-02015 HUT, FINLAND - Bibliography of Self-Organizing Map (SOM) Papers 1998-2001 Addendum
- RESOLUÇÃO NORMATIVA Nº 456 da ANEEL – Agência Nacional de Energia Elétrica, de 29 de novembro de 2000.
- REIS, J.; Gontijo, E. M.; Mazina, E.; Cabral, J. E.; Pinto, J. O. P. Fraud identification in electricity company customers using decision tree. In: 2004 IEEE International Conference on Systems, Man, and Cybernetics, p. 3730–3734, 2004.
- SPERANDIO, MAURICIO- Classificação de Conjuntos de Consumidores de Energia

Elétrica via Mapas Auto-Organizáveis e Estatística Multivariada -2004

VERSANTO, J. Data Mining Techniques Based on the Self-Organizing ap.Dissertação
– Helsinki University of Technology, May 1997.

Y. KOU, C.T. Lu, S. Sirwongwattana, Y.P. Huang, .Survey of Fraud Detection
Techniques,. Proceedings of the 2004 International Conference on Networking,
Sensing, and Control, pp. 749-754, Taipei, Taiwan, March 21-23, 2004.

Bibliografia Adicional

MINING SYSTEM FOR CREDIT CARD FRAUD DETECTION, Proceedings of the
IEEE/IAFE,1997.

ANDERSON, D., MCNEIL, G., Artificial Neural Networks Technology, 1992.

ABRADEE Associação Brasileira de Distribuidores de Energia, CODI 19-34 Metodologia
para determinação, Análise e Otimização de Perdas Técnicas em Sistemas de
Distribuição, 1994.

BRAGA, A. P., CARVALHO, A. C. P. L. F., LUDERMIR, T. B., Fundamentos de redes
neurais artificiais, 1998.

BOLTON, R. J., HAND, D. J., Unsupervised Profiling Methods for Fraud Detection, 2001.

BREIMAN, L., FRIEDMAN,R.A., OLSHEN, J.H. e STONE, C.J., Classification and
Regression Trees. Wadsworth, Belmont, CA, 1984.

ELLER, N. A., Arquitetura de informação para o gerenciamento de perdas comerciais de
energia elétrica, Programa de Pós Graduação, Engenharia da Produção, UFSC, 2003.

ENGELS, R. e THEUSINGER, C. Using a Data Metric for Preprocessing Advice for Data Mining Applications, European Conference on Artificial Intelligence, ECAI 1998.

FAYYAD, U., PIATETSKY-SHAPIRO, G. e SMYTH, P. The KDD Process for Extracting Useful Knowledge from Volumes of Data, ACM, 1996

HAN, J., KAMBER M., Data Mining Concepts and Techniques, Morgan-Kaufmann Publishers, 2001.

KPMG TRANSITION AND FORENSIC SERVICES Ltda. A fraude no Brasil Relatório de Pesquisa, 2004.