

UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO E SISTEMAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

ARQUITETURA DE INFORMAÇÃO PARA O
GERENCIAMENTO DE PERDAS COMERCIAIS DE
ENERGIA ELÉTRICA

NERY ARTUR ELLER

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina como requisito parcial para obtenção do título de Doutor em Engenharia de Produção.

Florianópolis
2003

NERY ARTUR ELLER

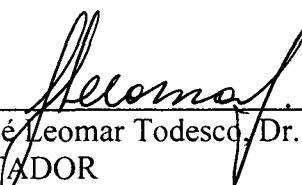
**ARQUITETURA DE INFORMAÇÃO PARA O
GERENCIAMENTO DE PERDAS COMERCIAIS DE
ENERGIA ELÉTRICA**

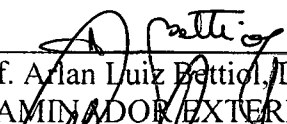
Esta tese foi julgada adequada para a obtenção do título de DOUTOR EM ENGENHARIA DE PRODUÇÃO e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina

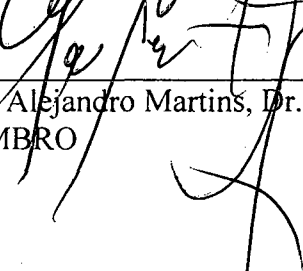
Florianópolis, Janeiro de 2003

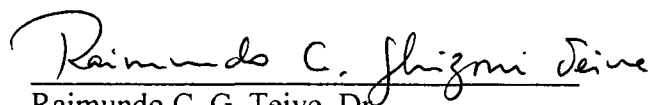
Prof. Edson Pacheco Paladini, Dr.
Coordenador do Curso


BANCA EXAMINADORA

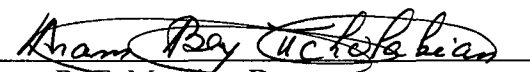

Prof. José Leomar Todesco, Dr.
ORIENTADOR


Prof. Arlan Luiz Bettiol, Dr./Ing.
EXAMINADOR EXTERNO


Prof. Alejandro Martins, Dr.
MEMBRO


Raimundo C. G. Teive, Dr.
MODERADOR


Maria Madalena Dias, Dra.
EXAMINADOR EXTERNO


Aran B. T. Morales, Dr.
MEMBRO

Este trabalho é dedicado aos meus pais, Artur e Nilza, aos meus irmãos, João Batista, Augustinho e Bernadete, à minha esposa, Eliane, e às minhas filhas, Julia e Elisa.

Agradecimentos

Gostaria de registrar os meus sinceros agradecimentos a todas as pessoas que direta ou indiretamente contribuíram para a realização deste trabalho.

Ao meu orientador e amigo, José Leomar Todesco, pela humildade, dedicação e competência na condução desta parceria.

À Universidade Federal de Santa Catarina, pela oportunidade de realizar este trabalho.

Ao professor Alejandro, por ter confiado em mim e por aceitar-me como aluno regular do Programa de Pós-Graduação em Engenharia de Produção.

Ao meu amigo Garbelotto, pelas importantes contribuições no desenvolvimento da arquitetura de informação proposta.

Agradeço também, de forma muito especial, às Centrais Elétricas de Santa Catarina S.A. (Celesc), pelos recursos e pelo tempo a mim destinados enquanto realizava a pós-graduação.

Pelo apoio recebido, externo também meus agradecimentos a Sylvio Luiz Reibnitz e Antônio Carlos Kleinubing, chefes, respectivamente, da Divisão de Suporte Técnico e do Departamento de Informática da Celesc.

Aos meus amigos Manoel Arisoli Pereira e João Airto de Bettio, pelos trabalhos de campo realizados na Agência Regional de Rio do Sul e na Agência Regional de Florianópolis.

À Prefeitura Municipal de Florianópolis, pelo fornecimento gratuito da base de dados do IPTU dos bairros de Coqueiros e Abraão.

À IBM, por ter cedido e instalado a plataforma de desenvolvimento do *Intelligent Mining*.

A todos, empresas e pessoas, que deram alguma contribuição, mas que não estão aqui relacionados, meus sinceros agradecimentos.

SUMÁRIO

LISTA DE FIGURAS	VII
LISTA DE QUADROS	VIII
LISTA DE TABELAS.....	IX
LISTA DE REDUÇÕES	X
RESUMO	XII
1 INTRODUÇÃO	1
1.1 APRESENTAÇÃO	1
1.2 OBJETIVO	2
1.3 OBJETIVOS ESPECÍFICOS.....	2
1.4 JUSTIFICATIVA	3
1.5 ORGANIZAÇÃO DO TRABALHO	6
2 CARACTERIZAÇÃO DO PROBLEMA	7
2.1 INTRODUÇÃO.....	7
2.2 HISTÓRICO DO SETOR ELÉTRICO BRASILEIRO	7
2.3 HISTÓRICO DA CELESC.....	10
2.4 PERDAS DE ENERGIA ELÉTRICA	13
2.5 FRAUDE NO CONSUMO DE ENERGIA ELÉTRICA.....	14
2.6 GERENCIAMENTO DAS PERDAS COMERCIAIS NA CELESC	15
2.6.1 <i>Conceitos</i>	15
2.6.2 <i>Aspectos legais</i>	16
2.6.3 <i>Tipos de inspeções utilizadas na fiscalização</i>	16
2.6.4 <i>Periodicidade das inspeções</i>	17
2.7 GERÊNCIA DE PERDAS COMERCIAIS EM OUTRAS EMPRESAS DO SETOR.....	18
2.8 PERDAS DE ENERGIA NA CELESC E EMPRESAS DO SETOR.....	21
2.9 SOLUÇÕES SENDO PESQUISADAS POR EMPRESAS DO SETOR ELÉTRICO.....	25
2.9.1 <i>Medições às claras</i>	25
2.9.2 <i>Sistema integrado de medição e gerenciamento de perdas</i>	26
2.9.3 <i>Energia pré-paga</i>	27
2.9.4 <i>Sistema de medição centralizada</i>	27
2.10 REESTRUTURAÇÃO DO SETOR ELÉTRICO BRASILEIRO	28
2.11 CONSIDERAÇÕES FINAIS.....	30
3 MINERAÇÃO DE DADOS.....	31
3.1 INTRODUÇÃO.....	31
3.2 TAREFAS REALIZADAS NO USO DAS TÉCNICAS DE MINERAÇÃO DE DADOS.....	33
3.3 TÉCNICAS DE MINERAÇÃO DE DADOS	34
3.4 REDES NEURAIS ARTIFICIAIS	37
3.4.1 <i>Breve histórico</i>	38
3.4.2 <i>Principais arquiteturas</i>	39
3.4.3 <i>Rede Perceptron de Múltiplas Camadas (MLP)</i>	40
3.4.4 <i>Rede Função de Base Radial (RBF)</i>	42
3.4.5 <i>Mapa auto-organizável de Kohonen</i>	45
3.4.6 <i>Segmentação demográfica</i>	46
3.5 TRABALHOS QUE UTILIZAM MD PARA DETECTAR FRAUDES.....	47
3.6 CONSIDERAÇÕES FINAIS.....	51
4 ARQUITETURA DE INFORMAÇÃO PARA GERENCIAMENTO DE PERDAS COMERCIAIS DE ENERGIA ELÉTRICA	53
4.1 INTRODUÇÃO.....	53

4.2	CONSTITUIÇÃO DA ARQUITETURA	53
4.3	MÓDULO 1: PREPARAÇÃO DOS DADOS	55
4.3.1	<i>Estrutura de dados</i>	56
4.3.2	<i>Processamento dos dados</i>	60
4.4	MÓDULO 1: PERDAS NOS ALIMENTADORES	60
4.4.1	<i>Natureza</i>	61
4.4.2	<i>Localização</i>	61
4.4.3	<i>Componentes</i>	61
4.4.4	<i>Fator de carga</i>	62
4.4.5	<i>Cálculo das perdas técnicas</i>	62
4.5	MÓDULO 2: KERNEL IA	74
4.5.1	<i>Introdução</i>	74
4.5.2	<i>Dados de entrada da rede</i>	75
4.5.3	<i>Classificação de consumidores secundários</i>	75
4.5.4	<i>Segmentação de consumidores primários</i>	77
4.5.5	<i>Dados de saída da rede</i>	78
4.6	MÓDULO 3: APLICATIVOS	79
4.6.1	<i>Introdução</i>	79
4.6.2	<i>Aplicativos para a fiscalização</i>	79
4.6.3	<i>Aplicativos para a gerência regional</i>	82
4.6.4	<i>Aplicativos para a gerência central</i>	85
4.7	CONSIDERAÇÕES FINAIS	87
5	EXPERIMENTOS REALIZADOS	89
5.1	INTRODUÇÃO.....	89
5.2	CLASSIFICAÇÃO DE CONSUMIDORES RESIDENCIAIS	89
5.3	CLASSIFICAÇÃO DE CONSUMIDORES COMERCIAIS	93
5.4	SEGMENTAÇÃO DE CONSUMIDORES INDUSTRIAIS.....	97
5.5	VERIFICAÇÃO DE CAMPO	103
5.6	CONSIDERAÇÕES FINAIS	106
6	CONCLUSÕES E TRABALHOS FUTUROS.....	107
6.1	CONCLUSÕES	107
6.2	CONTRIBUIÇÕES.....	108
6.3	TRABALHOS FUTUROS	110
	REFERÊNCIAS BIBLIOGRÁFICAS	112

LISTA DE FIGURAS

Figura 3.1 - Rede MLP de 3 camadas	40
Figura 3.2 - Rede RBF de 3 camadas	43
Figura 4.1 - Arquitetura de informação proposta	55
Figura 4.2 - Modelo E-R da base de dados do Módulo 1	57
Figura 4.3 - Sistema Elétrico de Distribuição	63
Figura 4.4 - Balanço energético	64
Figura 4.5 - Representação de demanda por patamares	65
Figura 4.6 - Representação de um trecho de rede	66
Figura 4.7 - Representação de circuito primário	67
Figura 4.8 - Representação da alimentação de uma carga	68
Figura 4.9 - Trecho de uma rede secundária	70
Figura 4.10 - Arquitetura básica para suportar aplicativos da área de fiscalização	80
Figura 4.11 - Relação de possíveis fraudadores localizados geograficamente	81
Figura 4.12 - Zoom de um possível fraudador	81
Figura 4.13 - Tela do aplicativo em um PDA	82
Figura 4.14 - Tela inicial do aplicativo de gerência regional	83
Figura 4.15 - Seleção de possíveis fraudadores para uma determinada taxa de confiança	83
Figura 4.16 - Gráfico do consumo de um possível fraudador	84
Figura 4.17 - Listagem de possível fraudador - aplicativo de gerência regional	84
Figura 4.18 - Tela de registro de ocorrência na fiscalização	85
Figura 4.19 - Seleção de fraudadores por região geográfica	86
Figura 4.20 - Fraudes agrupadas por agências regionais	87
Figura 5.1 - Gráfico de formação dos grupos a partir de uma segmentação demográfica	99
Figura 5.2 - Valores estatísticos das variáveis numéricas do grupo de fraudadores	99
Figura 5.3 - Valores estatísticos das variáveis de todos os tipos do grupo de fraudadores	100
Figura 5.4 - Valores estatísticos do grupo de fraudadores	101
Figura 5.5 - Valores estatísticos de alguns grupos	101
Figura 5.6 - Características e similaridades dos grupos	102
Figura 5.7 - Gráfico do consumo de energia do fraudador do caso 19	104
Figura 5.8 - Gráfico do consumo de energia do fraudador do caso 19	104
Figura 5.9 - Gráfico do consumo de energia do caso 16 – imóvel em reformas	105
Figura 5.10 - Gráfico do consumo de energia do caso 15 – imóvel em reformas	105

LISTA DE QUADROS

Quadro 4.1 - Descrição e tipo de dados da tabela CNS_Residencial	57
Quadro 4.2 - Descrição e tipo de dados da tabela CNS_Comercial	58
Quadro 4.3 - Descrição e tipo de dados da tabela CNS_Industrial	59
Quadro 4.4 - Descrição e tipo de dados da tabela CNS_Classificacao	78
Quadro 4.5 - Descrição e tipo de dados da tabela CNS_Segmentacao	78

LISTA DE TABELAS

Tabela 2.1 - Perdas globais de energia elétrica na Celesc	21
Tabela 2.2 - Índices de perdas nos sistemas MT e BT	21
Tabela 2.3 - Resultado da fiscalização de consumidores do grupo B em 2001 - Monfásico e Bifásico.....	22
Tabela 2.4 - Resultado da fiscalização de consumidores do grupo B em 2001 - Trifásicos.....	23
Tabela 2.5 - Resultado da fiscalização de consumidores do grupo A em 2001.....	24
Tabela 2.6 - Resultados financeiros das perdas comerciais na Celesc	24
Tabela 4.1 - Perdas no ferro e no cobre de transformadores	68
Tabela 5.1 - Experimentos realizados com consumidores residenciais.....	91
Tabela 5.2 - Experimentos realizados com consumidores comerciais.....	94

LISTA DE REDUÇÕES

Abreviaturas

ASMAE	Agente Negociador do Mercado Aberto de Energia
BNU	Blumenau
CCO	Chapécó
CDA	Concórdia
CUA	Criciúma
DIC	Duração da Interrupção do Consumidor
FIC	Frequência da Interrupção do Consumidor
FNS	Florianópolis
IA	Inteligência Artificial
IAI	Itajaí
JAM	Java Agents for Meta-Learning
JCA	Joaçaba
JGS	Jaraguá do Sul
JVE	Joinville
KDD	Knowledge Discovery in Database
LGS	Lages
MBR	Memory-Based Reasoning
MD	Mineração de dados
MFA	Mafra
MLP	Multi Layer Perceptron
PDA	Personal Digital Assistant
RBF	Rede Função de Base Radial
RNA	Redes neurais artificiais
RSL	Rio do Sul
RTD	Recomendação Técnica de Distribuição
SBS	São Bento do Sul
SD	Segmentação demográfica
SGE	São Miguel do Oeste
TC	Transformador de corrente
TRO	Tubarão
VII	Videira

Siglas

ABNT	Associação Brasileira de Normas Técnicas
ABRADEE	Associação Brasileira de Distribuidoras de Energia Elétrica
ANEEL	Agência Nacional de Energia Elétrica
BNDE	Banco Nacional de Desenvolvimento Econômico
CCON	Comitê Coordenador de Operação do Norte/Nordeste
CEB	Companhia de Eletricidade de Brasília
CELESC	Centrais Elétricas de Santa Catarina
CELPE	Centrais Elétricas de Pernambuco
CEPEL	Centro de Pesquisa da Eletrobrás
CHESF	Companhia Hidroelétrica do São Francisco
CITENEL	Congresso de Inovação Tecnológica em Energia Elétrica
CNAE	Conselho Nacional de Águas e Energia
COD	Centro de Operação da Distribuição
CODI	Comitê de Distribuição da Região Sul/Sudeste
COELBA	Companhia de Eletricidade da Bahia
COPEL	Companhia de Eletricidade do Paraná

CPREDE	Caixa Padrão de Rede
DPSC/DVMD	Departamento de Consumidores/Divisão de Medição
ELETOBRÁS	Centrais Elétricas Brasileiras S/A
ELETROPAULO	Eletricidade de São Paulo S.A.
ELETRONUCLEAR	Eletobrás Termonuclear S.A.
ELETROSUL	Elétrica do Sul do Brasil S.A.
EMPRESUL	Empresa Sul Brasileira de Eletricidade S.A.
ENERAM	Comitê Coordenador de Estudos Energéticos da Amazônia -
ESCELSA	Espírito Santo Centrais Elétricas S.A.
FIESC	Federação da Indústria do Estado de Santa Catarina.
GCOI	Grupo Coordenador para Operação Interligada
MAE	Mercado Atacadista de Energia Elétrica
MME	Ministério das Minas e Energia
ONS	Operador Nacional do Sistema Elétrico
PROCEL	Programa Nacional de Conservação de Energia Elétrica
PND	Programa Nacional de Desestatização
UFRJ	Universidade Federal do Rio de Janeiro

RESUMO

As empresas distribuidoras de energia elétrica possuem como um forte item na queda de receita a perda comercial de energia. A perda global de energia é verificada pela diferença entre a energia requerida pelo sistema elétrico e a energia realmente faturada, e possui duas componentes: a perda técnica e a perda comercial. As perdas técnicas são aquelas associadas à passagem e à interação do fenômeno físico nos meios e materiais utilizados e nos efeitos daí resultantes. As perdas comerciais são motivadas principalmente por problemas apresentados na unidade de medição que influenciam a leitura, por fraudes cometidas no consumo de energia elétrica e por ligações clandestinas, realizadas diretamente no sistema de distribuição.

São vários milhões de reais perdidos com perdas comerciais de energia, principalmente aquelas que envolvem fraudes. Somente em 2001, a Celesc - Centrais Elétricas de Santa Catarina S.A. recuperou R\$ 1.967.716,90 com fraudes que foram descobertas.

Para auxiliar no gerenciamento de perdas comerciais de energia elétrica propõe-se a presente arquitetura, que atuará principalmente na indicação de possíveis fraudadores. Esta arquitetura baseia-se na técnica de mineração de dados, que permite buscar esses consumos fraudados. Dentro da solução de mineração de dados, destaca-se o uso de redes neurais artificiais (RNA) para descobrir comportamentos suspeitos. Duas tarefas da RNA foram exploradas: classificação e segmentação. A classificação foi utilizada para se trabalhar com consumidores residenciais e comerciais, e a segmentação voltou-se a consumidores industriais.

Destaca-se também, nesta arquitetura, a possibilidade de se calcular as perdas global, técnica e comercial por alimentador, de forma que se tenha um controle mais rígido e específico sobre elas. A partir do presente trabalho, abriram-se várias oportunidades para novas pesquisas.

Palavras-chave: Perdas de energia; Mineração de dados; Arquitetura de informações

ABSTRACT

The companies that distribute electricity have suffered strong income losses because of commercial losses in energy itself. The global loss of energy is calculated taking into account the required energy distributed and the energy affectively billed, and can be classified in 2 types: technical losses and commercial losses. Technical losses are those regarding transport and interaction of electrical energy through the wired and other devices. The commercial losses are those derived from bad reading of measurement devices and usage frauds that arise from illegal connections that are not taken into account by the meters.

Several millions of brazilian reais are lost with such problems, specially those involving frauds. In 2001 the company Celesc – Centrais Elétricas de Santa Catarina S.A recalled R\$ 1.967.716, 90 from frauds that were uncovered.

In order to help the management of commercial losses involving electricity distribution we propose the present architecture, which will be used in chasing possible frauds. This architecture is based on data mining to gather information on abnormal energy consumption. This data mining process will use neural nets (AI – Artificial Intelligence) to point any uncommon or unexpected behavior over the gathered data. Two tasks are presented: classification and segmentation. Classification is used to deal with home customers and companies, and segmentation deals with industrial consumers.

Another strong feature of this architecture is the possibility of detailing global, technical and commercial losses from every power supply, in order to increase the control over each one of them. The present work also opens up several opportunities for further research.

Key-words: Losing of energy; Data mining; Information architecture

1 INTRODUÇÃO

1.1 Apresentação

Segundo estimativas das Centrais Elétricas de Santa Catarina S.A. (Celesc), a empresa perde cerca de quatro milhões de reais por mês com ligações clandestinas, desvios e fraudes de energia elétrica. Ocorre uma perda em torno de 3% do consumo, ficando 0,5% nas favelas e 2,5% nos outros setores da sociedade (PAVEI, 2001).

As irregularidades mais comuns ocorrem na adulteração de medidores e na instalação de circuito paralelo durante a construção do imóvel. Somente no ano de 2001, a Celesc precisou deslocar equipes em todas as agências regionais para recuperar o equivalente a R\$ 1.967.716,90 em energia roubada. Foram 11.579.565 KWH de energia fraudada (BEC, 2001).

O professor Jorge Mário Campagnolo, da Universidade Federal de Santa Catarina (UFSC), destaca que a ousadia dos fraudadores evoluiu tanto quanto as maneiras de burlar a fiscalização (PAVEI, 2001).

A partir de uma pesquisa realizada no primeiro semestre de 2001, envolvendo cinco mil residências de seis capitais brasileiras, o Centro de Políticas Sociais da Fundação Getúlio Vargas constatou que cerca de 10% dos consumidores no país são gatunos, não pagando pelo que gastam de energia elétrica. Do total, 6% são de classe média, e 2%, da elite (PAVEI, 2001).

No acompanhamento sistemático das perdas anuais globais da empresa (compra de energia menos o consumo medido) dos últimos quatro anos, verifica-se que o valor médio foi de 7,62%. É um dos melhores índices do setor elétrico brasileiro. No entanto, pode ser melhorado ainda mais, combatendo-se eficazmente o desvio e a fraude de energia, mantendo-se com qualidade o cadastro de consumidores, verificando-se periodicamente defeitos nas unidades de medição e controlando-se a qualidade técnica da rede de distribuição.

Para identificar fraudadores, a Celesc segue a sua meta de unidades a serem fiscalizadas anualmente (informações mais detalhadas no capítulo 2), agindo através de denúncias de terceiros, registros dos leituristas e alertas lançados pelo sistema de consumidores.

No ano de 2001, foram realizadas 15.487 fiscalizações em consumidores alimentados em alta-tensão (indústrias, basicamente) – eram 7.114 consumidores, constatando-se 52 fraudes no uso da energia elétrica; a relação fraudes/fiscalizações dá um valor de 0,34. Nos demais consumidores (1.766.343), foram realizadas 217.954 fiscalizações, resultando em 1.383 constatações de fraudes; a relação fraudes/fiscalizações dá como resultado 0,63.

Apesar do número de fraudes constatadas ser relativamente pequeno em relação ao número de fiscalizações empreendidas, o valor arrecadado atingiu a cifra de R\$ 1.967.716,90 somente no ano de 2001.

Melhorando-se a sistemática de relacionar possíveis fraudadores e dotando-se as áreas de fiscalização com soluções informatizadas que agilizem os trabalhos dos fiscais no escritório e no campo, a detecção de fraudes/irregularidades na medição se tornará mais eficiente, fazendo com que a perda comercial de energia sofra uma sensível redução.

1.2 Objetivo

Este trabalho tem como objetivo propor uma arquitetura de informação para melhorar o processo de identificação e combate às fraudes e irregularidades na medição de energia elétrica, utilizando-se técnicas de mineração de dados.

1.3 Objetivos específicos

Na busca de mecanismos para combater as perdas comerciais de energia elétrica, têm-se os seguintes objetivos específicos:

- identificar padrões de comportamento de consumidores fraudadores;
- identificar variáveis endógenas e exógenas à empresa (Celesc) que tenham algum vínculo com os padrões de comportamento fraudulento;
- utilizar técnicas de inteligência artificial para descobrir padrões de fraudes no consumo de energia elétrica;
- definir possíveis soluções tecnológicas através da informática para o combate e o acompanhamento de fraudes e irregularidades na medição, tendo como público-alvo a administração central da Celesc, as agências regionais da referida empresa e seu contingente de fiscais;

- estudar isoladamente, e de forma mais consistente, consumidores residenciais, comerciais e industriais, aproveitando as características comuns entre eles;
- testar em campo a eficiência dos prognósticos emitidos pelo *engine* IA, módulo 2 da arquitetura de informação proposta;
- propor uma sistematização da Gerência de Perdas Comerciais na Celesc;
- definir um modelo para calcular perda técnica e comercial por alimentador;
- munir a empresa (Celesc) de um método mais eficaz para detectar fraudes e irregularidades na medição, otimizando assim os trabalhos de seu efetivo de fiscalização;
- propor melhorias na qualidade da fiscalização realizada em campo.

1.4 Justificativa

A energia elétrica é um bem de consumo cujo custo é calculado no montante consumido, ou seja, quanto mais se consome, mais se gasta. Quando a energia representa um alto custo no negócio ou na economia do lar, alguns comerciantes, industriais e demais consumidores partem para caminhos alternativos no desvio de energia, fazendo ponte no medidor ou invertendo a ligação na entrada da unidade de medição. Os caminhos da planejada esperteza são vários, todavia, qualquer um deles tem um nome: furto.

Em Santa Catarina, o maior volume de fraudes concentra-se em consumidores residenciais. Em seguida, vêm os consumidores comerciais e, em menor número, as indústrias.

A Celesc possui aproximadamente 7.320 (set/2002) indústrias no grupo “A” (alimentadas em alta-tensão). Anualmente, a equipe de fiscalização (especializada nesse tipo de consumidor) verifica a unidade de medição de cada um deles. Mesmo assim, constataram-se 52 fraudes somente no ano de 2001. Esse número pode ser ainda maior, pois se pode questionar a eficiência da metodologia adotada na fiscalização como também conhecimentos atualizados em relação aos mecanismos de fraudes utilizados pelos fraudadores.

Os desvios de energia a partir dos postes, os defeitos nas unidades de medição e a falta de qualidade do cadastro de consumidores podem acarretar grandes perdas na arrecadação de uma empresa de energia elétrica. Essas variáveis estão totalmente sob controle da área gerencial da empresa.

O mesmo não ocorre quando o consumidor desvia energia em sua residência, comércio ou indústria, ou realiza fraudes sutis nas unidades de medição. O pior ainda acontece quando ele efetua caminhos alternativos para a energia em uma nova construção.

Com a arquitetura de informação proposta, pretende-se atuar no combate à redução de fraudes, tornando ainda menor o atual índice (em torno 3%) de perdas comerciais na empresa. Através da identificação do padrão comportamental dos indivíduos considerados fraudadores, pretende-se indicar possíveis fraudadores com um grau satisfatório de precisão.

Essa nova arquitetura contempla soluções que visam melhorar a eficiência da equipe de fiscalização como também dotar a área gerencial (central e regionais) com instrumentos que propiciem meios de melhor acompanhar e combater as perdas comerciais de energia elétrica na empresa.

Nos anos de 1998 e 1999, houve uma receita adicional com as descobertas de fraudes nos valores de R\$ 468.000,00 e R\$ 1.063.910,00, respectivamente.

No ano 2000, para um contingente de 44.848 unidades de medição do grupo "B" (alimentados em baixa tensão) fiscalizadas, detectaram-se 249 fraudes; já no grupo "A", 12.334 fiscalizações resultaram em 40 fraudes. A receita conseguida foi de R\$ 1.444.871,00.

Em 2001, através de um esforço concentrado, chegou-se à descoberta de 52 fraudes no grupo "A", a partir de 15.487 fiscalizações; no grupo "B", foram 1.383 fraudes para 217.954 fiscalizações. A recuperação de receita com as fraudes descobertas ficou em R\$ 1.967.716,90.

Dados acumulados ao longo do ano de 2002 (até setembro) mostram uma receita adicional de R\$ 2.627.706,98 com descobertas de fraudes. Para um universo de 7.320 consumidores do grupo "A", 10.741 fiscalizações foram realizadas, resultando em 28 fraudes. No grupo "B", de 1.716.026 consumidores monofásicos, 158.844 foram fiscalizados, resultando em 1.563 fraudes. Ainda no grupo "B", para 189.094 consumidores trifásicos, 42.143 unidades foram fiscalizadas, detectando-se 433 fraudes.

Segundo dados da Celesc, as perdas comerciais em 1998, 1999, 2000 e 2001 ficaram em 7,19%, 7,74%, 7,78% e 7,78%, respectivamente. Considerando-se uma perda comercial ("gatos", fraudes e irregularidades na medição) na ordem de 3,2% e tendo-se um faturamento bruto mensal (set/2002) de R\$ 213.467.150,07, pode-se concluir que a perda de receita em setembro foi na ordem de R\$ 6.830.948,00.

Há vários pontos que devem ser atacados para reverter esse quadro, iniciando-se pelo controle de qualidade dos dados do Sistema de Consumidores, combatendo-se os “gatos” e gerenciando-se as irregularidades nas unidades de medição.

Segundo a COELBA (Companhia de Eletricidade da Bahia), no ano 2000 os fraudadores causaram um prejuízo de 30 milhões de reais, sendo 25 mil ligações somente com “gatos” (FRANCISCO, 2001).

A Eletropaulo (São Paulo) calcula que na rede elétrica de São Paulo existam 200 mil ligações clandestinas (JOESP, 2001).

Na CELPE (Pernambuco), de um faturamento de 900 milhões de reais, 180 milhões (19,4%) são de perdas técnicas e comerciais (JDP, 2001).

A Light (Rio de Janeiro) estima que 10% do total de sua energia são roubados por ligações clandestinas, acarretando uma perda anual de 300 milhões de reais (DARIANO, 2001).

Estima-se que no Rio Grande do Sul o furto de energia elétrica chegue a 2% do total do consumo gaúcho (SFREDO, 2001).

Os números mostrados dão uma idéia da magnitude do problema que está sendo tratado. O grau de dificuldade para equacioná-lo é de natureza diversa, variando de região para região do país. Algumas distribuidoras de energia elétrica sofrem mais com os “gatos”, motivados pelo grande número de favelas (Rio de Janeiro, São Paulo, Bahia, Pernambuco); outras são mais visadas pelas fraudes na própria unidade de medição (Paraná, Santa Catarina, Minas Gerais, Rio grande do Sul).

Várias empresas estão se mobilizando para reduzir drasticamente suas perdas comerciais. A Light, por exemplo, está investindo numa parceria com a Universidade Federal do Rio de Janeiro (UFRJ) para reduzir seu alto índice de perdas técnicas e comerciais. Na CELPE, o Grupo Rede está trabalhando com software e hardware para combater as perdas comerciais, inclusive vendendo tecnologias a outras concessionárias.

Essa nova proposta de detecção de fraudes não tem como propósito ser um método infalível de descobertas de fraudadores. É, na realidade, um mecanismo que irá auxiliar a gerência da fiscalização e seus fiscais na descoberta de fraudes, como também facilitar o trabalho de campo da equipe, com o intuito de melhorá-lo.

1.5 Organização do trabalho

Este trabalho está dividido em seis capítulos: no primeiro foi feita a apresentação e foram definidos o objetivo e a justificativa.

No segundo capítulo, é descrita a caracterização do problema. Apresenta-se um breve histórico do setor elétrico nacional e estadual como também informações sobre perdas de energia elétrica na Celesc e em outras empresas distribuidoras do Brasil. Discutem-se também algumas soluções que estão sendo pesquisadas para combater as perdas comerciais de energia. Finaliza-se o capítulo apresentando-se rapidamente a reestruturação do setor elétrico brasileiro.

O terceiro capítulo trata de mineração de dados. A mineração de dados é descrita basicamente em função de suas tarefas e técnicas. As técnicas utilizadas para minerar informações são descritas objetivamente para possibilitar um bom entendimento do seu uso na arquitetura de informação trabalhada. Para ilustrar o uso da mineração de dados na detecção de fraudes, apresentam-se alguns trabalhos que abordam esse assunto. Detalha-se um pouco mais o tema *redes neurais artificiais*, uma vez que este ocupa um importante papel na solução do problema em questão.

No quarto capítulo, descreve-se de forma detalhada a arquitetura de informação para o gerenciamento de perdas comerciais de energia elétrica. Todos os módulos constituintes da arquitetura são explorados, mostrando-se de forma clara a função de cada um no processo geral de detecção de fraudes.

A implementação da arquitetura proposta é apresentada no quinto capítulo. Os experimentos realizados e alguns dos resultados obtidos estão aqui descritos. As fases de preparação dos dados e as técnicas de mineração utilizadas como também os aplicativos para o usuário final são adequadamente descritos, seguindo a estrutura de cada um dos módulos apresentados.

O trabalho é finalizado com o sexto capítulo, no qual constam as considerações finais e os trabalhos futuros que poderão ser conduzidos a partir de aprofundamentos ou desmembramentos de idéias e soluções lançadas no desenvolvimento da arquitetura que se propõe.

2 CARACTERIZAÇÃO DO PROBLEMA

2.1 Introdução

Neste capítulo, relaciona-se um conjunto de informações para que o tema “Perdas Comerciais de Energia” fique adequadamente contextualizado no tema maior, que é o setor elétrico.

A história do setor elétrico nacional, como também da empresa que serviu de laboratório (Celesc), é resumidamente destacada para dar uma breve idéia da importância dessa empresa para o desenvolvimento do Estado catarinense. A história do setor elétrico nacional é pertinente, pois nossa matriz energética atual está grandemente alicerçada na energia elétrica, constituindo-se no principal insumo do desenvolvimento econômico do país.

A magnitude do problema tratado é retratada nas descrições das perdas comerciais de energia elétrica, tanto da Celesc como de outras empresas do setor.

A partir do ano de 1997, houve uma reestruturação substancial no setor elétrico brasileiro, ocorrendo uma série de privatizações e também a criação da Agência Nacional de Energia Elétrica (ANEEL). Esse assunto é visto aqui com o nível de detalhe suficiente para se ter uma idéia de suas implicações no setor elétrico em seu conjunto.

Apresentam-se também alguns trabalhos que estão sendo desenvolvidos visando a combater as perdas comerciais de energia nas empresas distribuidoras de energia elétrica.

2.2 Histórico do setor elétrico brasileiro

Os relatos cronológicos da história do setor elétrico brasileiro estão sendo apresentados da mesma forma como se apresentam no site da Associação Brasileira de Distribuidoras de Energia Elétrica (ABRADEE, 2002). A história da energia elétrica no Brasil começa com o interesse de D. Pedro II pelas descobertas de Thomas Edison. O relato do atual acontecimento do setor fica por conta da sua reestruturação, em que o governo sai parcialmente do mercado, atuando a partir de então como agente fiscalizador e regulamentador.

A energia elétrica no Brasil passou por diversas fases que podem ser assim elencadas: primórdios (1879-1899), implantação (1903-1927), regulamentação (1934-1945), expansão (1952-1961), consolidação (1962-1973), estatização (1975-1986), privatização (1988-1999) e os tempos atuais, caracterizados pela desregulamentação do setor, a partir do ano 2000.

Em 1879, D. Pedro II concedeu a Thomas Alva Edison o privilégio de introduzir no país aparelhos e processos de sua invenção destinada à utilização da eletricidade na iluminação pública. Nesse ano, também era inaugurada na Estação Central da Estrada de Ferro D. Pedro II, atual Estrada de Ferro Central do Brasil, a primeira instalação de iluminação elétrica permanente.

A Diretoria Geral dos Telégrafos, em 1881, instalou na cidade do Rio de Janeiro a primeira iluminação externa pública do país em trecho da atual Praça da República.

Em 1883, entrou em operação a primeira usina hidrelétrica no país, localizada no Ribeirão do Inferno, afluente do rio Jequitinhonha, na cidade de Diamantina. Nesse mesmo ano, D. Pedro II inaugurou na cidade de Campos o primeiro serviço público municipal de iluminação elétrica do Brasil e da América do Sul.

A primeira hidrelétrica de maior porte do Brasil, Marmelos-Zero, da Companhia Mineira de Eletricidade, pertencente ao industrial Bernardo Mascarenhas, entrou em operação em 1889.

A fase de implantação da energia elétrica no Brasil inicia-se com o primeiro texto de lei disciplinando o uso de energia elétrica no país, aprovado pelo Congresso Nacional em 1903. Já em 1908, entra em operação a Usina Hidrelétrica Fontes Velha, na época, a maior usina do Brasil e uma das maiores do mundo.

Logo no início da fase de regulamentação, em 1934, foi promulgado pelo Presidente Getúlio Vargas o Código de Águas, assegurando ao poder público a possibilidade de controlar rigorosamente as concessionárias de energia elétrica. Em 1939, o presidente Getúlio Vargas criou o Conselho Nacional de Águas e Energia (CNAE), para sanear os problemas de suprimento, regulamentação e tarifa referentes à indústria de energia elétrica do país.

No ano de 1945, foi criada em Recife a primeira empresa de eletricidade de âmbito federal, a Companhia Hidroelétrica do São Francisco (CHESF).

Entrando na fase expansionista do setor de energia elétrica, em 1952 era criado o Banco Nacional de Desenvolvimento Econômico (BNDE), para atuar nas áreas de energia e transporte.

Em 1954, entrava em operação a primeira grande hidrelétrica construída no rio São Francisco, a Usina Hidrelétrica Paulo Afonso I, pertencente à CHESF. Nesse mesmo ano, entrou em operação a Usina Termelétrica Piratininga, a óleo combustível, primeira termelétrica de grande porte do Brasil.

Em 1957, foi criada a Central Elétrica de Furnas S.A., com o objetivo expresso de aproveitar o potencial hidrelétrico do rio Grande e solucionar a crise de energia na Região Sudeste. Passados três anos, em 1960, o Presidente Juscelino Kubitschek, dentro de seu programa Plano de Metas, criou o Ministério das Minas e Energia (MME). Para coordenar o setor de energia elétrica brasileiro, o presidente Jânio Quadros criou, em 1961, a Eletrobrás.

A fase de consolidação do setor elétrico é marcada com a entrada em operação da Usina Hidrelétrica de Furnas, a maior usina do Brasil na época de sua construção. Em 1968, nascia a empresa Elétrica do Sul do Brasil S.A. (ELETROSUL), na qualidade de empresa subsidiária da Eletrobrás. Também entrava em operação a maior termelétrica do país, a Usina Termelétrica Santa Cruz, de Furnas - Centrais Elétricas S.A. Para supervisionar os estudos hidroenergéticos relacionados à Região Amazônica, foi criado o Comitê Coordenador de Estudos Energéticos da Amazônia (ENERAM).

Concluindo essa fase de consolidação, destaca-se o ano de 1973, como consequência do tratado firmado entre Brasil e Paraguai, regulando a construção e operação de hidrelétricas no rio Paraná, foi criada a Itaipu Binacional (ITAIPU).

No início da fase de estatização, em 1975, foram criados o Comitê de Distribuição da Região Sul/Sudeste (CODI) e o Comitê Coordenador de Operação do Norte/Nordeste (CCON).

Em 1984, entrou em operação a Usina Hidrelétrica Itaipu, maior hidrelétrica do mundo, com 12.600 MW de capacidade instalada.

No ano de 1985, foi constituído o Programa Nacional de Conservação de Energia Elétrica (PROCEL), com o objetivo de incentivar a racionalização do uso da energia elétrica. Entrou em operação, também nesse ano, a Usina Termonuclear Angra I, primeira usina nuclear do Brasil.

Um fato marcante da fase de privatização ocorreu em 1990, com a Lei nº 8.031 sendo sancionada pelo presidente Fernando Collor de Mello. Era então criado o Programa Nacional de Desestatização (PND).

Em 1995, as empresas controladas pela Eletrobrás foram incluídas no Programa Nacional de Desestatização, que orientava a privatização dos segmentos de geração e distribuição. Realizado o leilão de privatização da Escelsa, inaugurando nova fase do setor de energia elétrica brasileiro em consonância com a política de privatização do Governo Federal.

No ano de 1997, a operação da Usina Hidrelétrica Itaipu atingiu 88.560 GWh, estabelecendo novo recorde mundial de produção de energia elétrica em uma só central hidráulica. Nesse ano, também foi criada a Eletrobrás Termonuclear S.A. (ELETRONUCLEAR), empresa que passou a ser a responsável pelos projetos das usinas termonucleares brasileiras. Ainda nesse ano era constituído o novo órgão regulador do setor de energia elétrica sob a denominação de Agência Nacional de Energia Elétrica (ANEEL).

Como destaque na fase atual do setor, cita-se o ano de 1998, com a regulamentação do Mercado Atacadista de Energia Elétrica (MAE), consolidando a distinção entre as atividades de geração, transmissão, distribuição e comercialização de energia elétrica. Foram estabelecidas as regras de organização do Operador Nacional do Sistema Elétrico (ONS), para substituir o Grupo Coordenador para Operação Interligada (GCOI).

Em 1999, a primeira etapa da Interligação Norte-Sul entrou em operação, representando um passo fundamental para a integração elétrica do país.

E, finalmente, no ano 2000, o presidente Fernando Henrique Cardoso lançou o Programa Prioritário de Termelétricas visando à implantação no país de diversas usinas a gás natural. Nesse ano, entrou em operação a Usina Termonuclear Angra II.

2.3 Histórico da Celesc

Este breve histórico das Centrais Elétricas de Santa Catarina S.A (Celesc) está contido no site da empresa (CELESC, 2002).

A Celesc foi criada em dezembro de 1955 pelo Decreto Estadual nº 22, assinado pelo governador Irineu Bornhausen. Até a metade do século, as necessidades energéticas do Estado eram supridas por pequenos e médios sistemas elétricos regionalizados, geralmente mantidos pela iniciativa privada.

Ainda na primeira década do século, Blumenau já dispunha de um rudimentar sistema de iluminação pública. Lá, a usina hidrelétrica Salto Weissbach, datada de 1916, significou uma evolução dos pequenos geradores mantidos pelo espírito empreendedor dos imigrantes desde a virada do século. Essa usina foi definitiva para a extraordinária expressão industrial de todo o Médio Vale do Itajaí.

Em Joinville, a usina hidrelétrica Piraí entrou em funcionamento em 1908 e, em 1913, foi a vez da São Lourenço, em Mafra. Para o suprimento da Capital, o governador Gustavo Richard ordenou a construção da usina hidrelétrica Maroim, em São José. Essa usina está desativada e encontra-se em processo de recuperação arquitetônica.

Esse modelo, no entanto, começou a se mostrar incapaz de responder ao aumento da demanda, pressionada pelo surto desenvolvimentista que tomou conta do país no governo de Juscelino Kubitschek. Preocupado em oferecer condições de infra-estrutura aos investimentos, o governo do Estado decide, então, pela criação da estatal.

Como resultado imediato, o início das operações da Celesc viabilizou a entrada de Santa Catarina no Sistema Elétrico Interligado Sul-Sudeste, medida que garantiu o fornecimento de eletricidade adequado ao parque industrial catarinense.

A princípio, a Celesc funcionou mais como um órgão de planejamento do sistema elétrico estadual. Depois, assumiu o papel de *holding* até começar a incorporar, gradativamente, o patrimônio das velhas empresas regionais. Foi assim que começou seu ciclo de expansão, e a região Sul, já na década de 70, foi a última a ser atendida. Nessa região ainda se concentra o maior número de cooperativas de eletrificação em Santa Catarina.

O primeiro grupo de empresas tomadas subsidiárias foi formado pela Empresa Sul Brasileira de Eletricidade S.A. (Empresul), com base em Joinville, pela Força e Luz Videira S.A., pela Companhia Oeste de Eletricidade, de Concórdia, pela Companhia Serrana de Eletricidade, de Lages, pela Companhia Pery de Eletricidade, com base em Curitiba, e pela Empresa Força e Luz Santa Catarina S.A., com atuação no Vale do Itajaí e sede em Blumenau.

A consolidação técnica da Celesc e o delineamento definitivo do sistema elétrico estadual ocorreram a partir de 1965, com a construção da Linha de Transmissão Tubarão–Lages–Joaçaba–São Miguel D'Oeste.

Gradativamente, a Celesc cobriu todo o Estado e tornou-se a estatal catarinense mais importante.

A Celesc, em dados atuais (set/2002), está assim constituída:

▪ Área de concessão:	96% da área do Estado
▪ Número de clientes:	1.798.034
▪ Número de empregados:	4.229
▪ Consumo de energia:	1.050.978.001 KHW
▪ Geração própria:	3,6%

A Celesc passa atualmente por uma reestruturação significativa, que dará origem a um novo modelo de empresa. Os responsáveis por esse novo modelo (Governo do Estado - 1999/2002, FIESC e empregados, basicamente) almejam fazer dele uma referência nacional em gestão de empresas estatais.

Essa reestruturação possui duas etapas distintas. Na primeira etapa, a distribuidora de energia elétrica firma-se como uma *holding*, tendo abaixo dela as subsidiárias integrais de geração e telecomunicações. Numa segunda etapa, será criada uma *holding* pura, com três ou mais subsidiárias.

Inicialmente a Celesc será desmembrada em três ramos/empresas: Celesc Distribuidora, Celesc Geração e Celesc Telecom.

As bases do novo modelo estão estabelecidas como descrito abaixo.

- Acordo de acionistas.
 - Firmado entre o Estado de Santa Catarina e os acionistas minoritários com direito a voto.
- Regras de governança corporativa.
- Contrato de gestão e resultados.
 - Firmado entre o Conselho de Administração (contratante) e a Diretoria Executiva (contratada).
- Conselho de administração.
 - Composição: 6 representantes do Estado de Santa Catarina; 3 representantes dos acionistas minoritários; 3 representantes dos consumidores; 1 representante dos empregados.

- Diretoria executiva.
 - Formada por quatro membros: Diretor Presidente; Diretor Econômico-Financeiro; Diretor Técnico; Diretor de Gestão e Desenvolvimento Organizacional. A escolha do presidente se processa através de lista tríplice, para um mandato de três anos.

2.4 Perdas de energia elétrica

As perdas de energia elétrica encontradas em um sistema de distribuição são basicamente de duas naturezas: técnica e comercial. Ao conjunto das duas perdas dá-se o nome de Perda Global do Sistema.

O documento técnico do Comitê de Distribuição (CODI, 1997, p. 7) define perdas técnicas como “a energia perdida no transporte, na transformação e nos equipamentos de medição da energia elétrica quando do fornecimento da mesma. As perdas comerciais são aquelas decorrentes da energia efetivamente entregue aos consumidores finais ou a outras concessionárias, mas não computadas na venda”.

Rocha et al. (1997, p. 5) definem perdas globais como “a diferença entre a energia requerida por uma fonte consumidora e a energia faturada nessa área. Desmembrando-se as perdas globais, ter-se-á as perdas técnicas e as perdas comerciais”. As perdas técnicas são aquelas associadas à passagem e interação do fenômeno físico nos meios e materiais utilizados e os efeitos daí resultantes. Compreendem toda a energia dissipada por efeito joule no transporte de corrente elétrica, nos condutores, equipamentos e conexões, acrescentadas a estas, à energia associada às correntes de fuga à massa em um sistema de potência. As perdas comerciais são aquelas oriundas de falhas administrativas, na política de comercialização, ou da ação de terceiros no uso irregular de energia.

Para Nunes Filho (2000, p. 1), as “perdas comerciais são aquelas associadas à comercialização da energia fornecida ao usuário final. Existe uma parcela de energia que é produzida, transportada e fornecida, no entanto, não é faturada, propiciando uma perda direta no faturamento”. Elas são classificadas em administrativas e de intervenção ilícita. As perdas comerciais de natureza administrativa possuem como causas: sistema de medição incompleto, obsoleto ou descalibrado; cadastro de consumidores ineficiente; erro na leitura registrada pelo medidor ou no processo de informação dessa leitura; e erro no processo de faturamento. As intervenções ilícitas correspondem ao montante de energia não faturado, devido às

intervenções que não são autorizadas pela concessionária, realizadas pelos usuários legalizados ou não, com a finalidade de alterar o registro de energia elétrica ou o seu desvio.

2.5 Fraude no consumo de energia elétrica

Conceitua-se fraude no uso de energia elétrica o ato de má-fé praticado contra a empresa fornecedora de energia elétrica, o qual impede a correta medição e/ou faturamento (BETTIO, 2001).

No consumo de energia elétrica a fraude, ou furto de energia, é estabelecida quando ocorre o desvio da energia elétrica em pontos localizados antes da medição; a adulteração dos mecanismos internos do medidor; a alteração dos registros de consumo; a interferência na chave de aferição; a interferência na fiação dos circuitos dos sistemas de medição indireta; ligação à revelia; a religação à revelia; a interferência ou avaria provocada nos transformadores para instrumentos e quaisquer outros atos provocados intencionalmente com o objetivo de diminuir o registro da energia elétrica efetivamente consumida na unidade consumidora (CODI, 1997).

Pode-se também definir fraude como a ocorrência de qualquer procedimento irregular na unidade consumidora cuja responsabilidade não lhe seja atribuível e que tenha provocado faturamento inferior ao correto, ou no caso de não ter havido qualquer faturamento. Constitui-se também em fraude a ligação clandestina ou religação à revelia (ANEEL, 2001).

Os tipos mais comuns de fraudes encontradas em consumidores de baixa tensão da Celesc são (BETTIO, 2001):

- derivação antes da medição;
- abertura da ponte de potencial do medidor de energia;
- bloqueio do disco do medidor através de furo na carcaça, vidro ou com abertura da tampa;
- inversão dos fios de entrada com os de saída da unidade medidora;
- descalibragem proposital do medidor;
- pressão de um mancal contra o outro, visando ao funcionamento irregular;
- desacoplamento do registrador;
- troca de engrenagens do registrador;

- abertura de um pólo do disjuntor nos consumidores rurais ligados em 440/220 V.

Com relação aos consumidores de alta-tensão da Celesc, os procedimentos mais comuns de fraude são (BETTIO, 2001):

- abertura da chave de aferição;
- alargamento do pólo da chave de aferição, impedindo dessa forma o contato elétrico;
- colocação de verniz isolante na chave de aferição, impedindo assim a correta medição do consumo;
- fios do circuito secundário dos transformadores de corrente (TC) de medição descascados, formando pontos por contato entre os condutores, reduzindo assim a corrente no medidor;
- fios do circuito secundário dos transformadores de potencial (TP) interrompidos em determinados horários;
- desacoplamento do registrador;
- troca dos transformadores de corrente por relação de transformação superior mantendo a placa original;
- curto-circuito no circuito primário dos transformadores de corrente.

2.6 Gerenciamento das perdas comerciais na Celesc

O gerenciamento das perdas comerciais na Celesc é realizado através do acompanhamento mensal do total das perdas globais e da atuação sistemática no combate às fraudes e irregularidades na medição. O combate às perdas comerciais dá-se através de inspeções periódicas, informações fornecidas por leituristas e por terceiros, e relatórios de ocorrências do Sistema de Consumidores.

2.6.1 Conceitos

Dá-se o nome de inspeção de medição à vistoria técnica a ser realizada no padrão de entrada da unidade consumidora, visando a:

- a) eliminar problemas de segurança;

- b) verificar equipamentos de medição queimados e/ou danificados;
- c) verificar ligações erradas;
- d) detectar fraudes e/ou desvios de energia;
- e) verificar erros de cadastramento.

2.6.2 Aspectos legais

Para o tratamento de fraude, desvio de energia e irregularidades na medição, a Celesc mune-se dos seguintes aspectos legais:

- Resolução 456, de 29/11/2000, da ANEEL (artigos 71 a 78);
- Recomendação Técnica de Distribuição - RTD nº 9, de 22/09/82, do CODI;
- Normativa I-321.0006, da Celesc;
- Código penal, Artigo 155;
- Código de defesa do consumidor - Lei nº 8.048, de 11/09/1990, Artigo 27.

2.6.3 Tipos de inspeções utilizadas na fiscalização

As inspeções realizadas na empresa, para efeitos de fiscalização, ocorrem da seguinte forma:

Inspeção programada: são as inspeções que possuem uma programação prévia, atendendo a recomendações da ANEEL e à normativa interna (I-321.0011).

Os parâmetros mais comuns que influem na determinação de uma inspeção programada são:

- a) níveis de carga das unidades consumidoras a serem visitadas;
- b) características das unidades consumidoras;
- c) área geográfica abrangida pelas agências regionais.

Inspeção por amostragem: como o próprio nome sugere, neste tipo de inspeção toma-se uma amostra como objeto da ação de fiscalização.

Pode-se adotar os seguintes critérios:

- a) por rotas de leituras ou razões;
- b) critérios estatísticos, com base no ano da instalação da medição, ou da última inspeção;

- c) incidência de fraudes e/ou desvios em determinados tipos de consumidores ou em região específica da área de concessão da Celesc.

Inspeção por análise de dados históricos: este tipo de inspeção leva em consideração o Relatório de Ocorrência emitido pelo Sistema de Consumidores. Tal relatório baseia-se em variações significativas dos seguintes parâmetros:

- a) demandas máximas mensais;
- b) fatores de carga;
- c) fatores de potência médios mensais;
- d) consumos mensais;
- e) outros.

Inspeção por denúncia ou informação de leituristas: neste caso, são consideradas todas as informações, anônimas ou não, principalmente as obtidas de leituras e de consumidores, em cujas unidades consumidoras foram constatadas irregularidades, inclusive com violação ou retirada de lacres.

Inspeção noturna e finais de semana: inspeções especiais e eventuais para gerar o fator surpresa em algumas regiões problemáticas.

2.6.4 Periodicidade das inspeções

Compete ao DPSC/DVMD fiscalizar anualmente nas agências regionais 30% das unidades consumidoras do grupo “A” (fornecimento em alta-tensão), 100% de todas as unidades consumidoras com medições especiais (Tarifação Horosazonal).

Compete às agências regionais efetuarem anualmente a fiscalização de todas as unidades consumidoras do grupo “A” e, a cada 36 meses, todas as unidades consumidoras do grupo “B” (consumidores trifásicos).

A agência regional também é responsável pela inspeção anual de pelo menos 8% (meta) dos consumidores monofásicos e bifásicos.

2.7 Gerência de perdas comerciais em outras empresas do Setor

Em dezembro de 2001, enviou-se um questionário às principais empresas de distribuição de energia elétrica do Brasil, solicitando informações sobre a gerência de perdas comerciais efetuadas em cada uma delas. Entre as que responderam ao questionário, serão relacionadas neste trabalho a Companhia de Eletricidade do Paraná (COPEL) e a Companhia de Eletricidade de Brasília (CEB).

A parte principal do questionário estava assim formada:

1. É medido e acompanhado o nível de perdas comerciais em sua empresa? De que forma?
2. Quais os mecanismos utilizados para verificar fraudes e desvios de energia?
3. Como são feitas as inspeções nas diversas classes de consumidores? Com que frequência?
4. Existe processo, equipamento ou software para detectar fraude e desvio de energia? Caso exista, como se processa/funciona em linhas gerais?
5. Como são utilizados o cadastro de consumidores e o faturamento para gerenciar as perdas comerciais?
6. Existe cooperação entre áreas/processos internos e externos à empresa para combater fraudes?
7. Os sistemas de informação (comercial e engenharia) interagem para gerar conhecimentos que possam auxiliar o gerenciamento das perdas comerciais?

A COPEL assim se manifestou:

1. As perdas comerciais são normalmente medidas, melhor falando, estimadas, pela seguinte equação:

$$\text{Perdas Comerciais} = \text{Perdas Globais} - \text{Perdas Técnicas}$$

Isso porque embora as empresas utilizem mecanismos para detectar e quantificar os vários tipos de perdas comerciais, como descrito nos itens seguintes, desconhecem, por definição, onde se localizam e quais os valores das fraudes, medidores avariados, etc.

2. As perdas globais no sistema de distribuição da COPEL (redes de 34,5 kv e de 13,8 kv, incluindo-se as subestações de 34,5 e 13,8 kv e redes de BT, composta por transformadores e condutores), variaram no período de 1994 a 2000 entre um mínimo de 5,111% e um máximo de 5,938%. Considerando-se as características do mercado e do sistema construído para atendê-lo, conclui-se que as perdas técnicas são da mesma ordem de grandeza. Sendo assim, não faz sentido, no caso

deste trabalho, proceder ao cálculo das perdas técnicas, processo extremamente trabalhoso e de pouca precisão: o erro normal desse cálculo é, provavelmente, maior que as perdas comerciais, e não seria possível tomar qualquer conclusão.

3. Enquanto a situação comentada no item 2 se mantiver, julga-se suficiente limitar as perdas comerciais às medidas permanentes e rotineiras relacionadas a seguir.
4. Quando existe uma diferença apreciável entre as perdas globais e o nível otimizado das perdas técnicas (dado pela curva de otimização definida em trabalhos do CODI, onde são consideradas as características de cada empresa), conclui-se que as perdas comerciais são importantes. Nesse caso, é aconselhável fazer o cálculo das perdas técnicas reais, com a finalidade quase exclusiva de determinar as perdas comerciais: o cálculo das perdas técnicas é pouco eficiente no combate à perda global, mas necessário para, ao longo do tempo, monitorar os resultados obtidos no combate às perdas comerciais.
5. As questões restantes serão tratadas neste item. Os cuidados para detecção de fraudes, furtos e desvios de energia se iniciam já no momento da ligação da unidade consumidora à rede. A vistoria de uma ligação nova deve ser acompanhada de inspeção apurada no local, pelo electricista que fará a ligação, para verificar possíveis artifícios utilizados para desviar a energia. As fraudes mais difíceis de serem detectadas são aquelas estabelecidas anteriormente à ligação da unidade consumidora, pois os históricos de consumo já se iniciam com valores deturpados.

No ato das leituras mensais, o leiturista também é orientado a apontar suspeita de irregularidades na unidade consumidora, tais como interligações clandestinas e desvios e furtos de energia, que mais tarde geram relatórios para que as áreas possam fazer uma inspeção mais apurada no local, com vistas a constatar a irregularidade.

Variações significativas de consumos e demandas são apontadas em relatórios de consistência de faturamento, que podem ser utilizados pelas áreas para verificações na unidade consumidora.

As auto-relições (quando a unidade consumidora é desligada por falta de pagamento e o consumidor religa à revelia da empresa) são apontadas automaticamente pelo sistema (compara a leitura atual com a anterior e informa que houve consumo em unidade desligada)

em relatório de consistência, possibilitando às áreas a atuação no sentido de regularização junto ao consumidor e aplicação das medidas cabíveis.

As inspeções na empresa são feitas a partir de análise de relatórios de irregularidades apontadas por leituristas e análise de histórico de consumo, ou seja, não de forma ostensiva, mas sempre que houver indícios de fraudes.

Eventualmente, dependendo da região (como áreas invadidas e alta rotatividade da população), as inspeções são realizadas de forma mais intensiva, muitas vezes por amostragem.

A CEB retornou o seguinte documento:

1. Não estimamos a perda com fraude, apenas a perda global.
2. Antes da publicação da Resolução 456 da Agência Nacional de Energia Elétrica (ANEEL), o cliente era comunicado formalmente através de um documento (Comunicação de Irregularidade, com as assinaturas do cliente, da CEB e de uma testemunha) que descrevia a fraude e lhe dava um prazo para comparecer à CEB sob risco de suspensão do fornecimento de energia. Na CEB, após os devidos esclarecimentos, era assinado outro documento (ambas as partes) contendo informações sobre legislação e posteriormente efetuada a cobrança, autorizando a CEB à proceder a regularização da medição. Posteriormente, o cliente recebia uma carta com o memorial descritivo do cálculo do valor a ser pago devido à fraude, dando-lhe um prazo para recurso. Após esse prazo, era emitida a fatura. Em alguns casos (consumo alto ou suspeita de formação de quadrilha), a autuação era feita juntamente com o Instituto de Criminalística, com a devida ocorrência policial. Nos dias atuais, em função da nova legislação, esse procedimento foi alterado.
3. Atualmente, devido ao grande volume de solicitações de serviços, para o trabalho em questão foram priorizados apenas os casos com perspectivas de maior retorno.
4. Não.
5. Do Sistema de Consumidores, trabalhamos algumas informações para tomada de decisão.
6. Não.
7. Não.

2.8 Perdas de energia na Celesc e empresas do setor

Do total de energia produzida e comprada pela Celesc, em média, 7,62% (média dos últimos quatro anos) do total são perdidos: 4,4% são considerados perdas técnicas (principalmente por efeito joule), e 3,2% são perdas comerciais (fraudes, desvio de energia e irregularidades na medição).

Na Tabela 2.1, pode-se observar as perdas globais de energia elétrica na Celesc ao longo dos últimos seis anos, onde pode-se verificar que a média ficou em torno de 7,72%. Para se chegar aos valores de perdas apresentados, utiliza-se a seguinte fórmula (2.1):

$$PG = ((ES - CT) / ES) * 100 \quad (2.1)$$

onde,

PG = Perda Global (Percentual)

ES = Energia Suprida (MWH)

CT = Consumo Total de Energia (MWH)

Tabela 2.1 - Perdas globais de energia elétrica na Celesc

Mês	1996	1997	1998	1999	2000	2001
TOTAL MWh	851.147	854.360	839.212	953.968	1.031.398	1.065.691
TAXA DE CRESC %	20,69	0,38	(1,77)	13,67	8,12	3,32
TOT. EN. REQ.(MWh)	10.356.551	11.216.678	11.664.574	12.331.535	13.249.765	13.698.992
ÍNDICE PERDAS %	8,22	7,62	7,19	7,74	7,78	7,78

Fonte: DPSC/DVMD/Celesc

A Tabela 2.2 na seqüência relaciona as perdas globais da Celesc com outras empresas do setor, ao longo dos últimos nove anos. Pelos dados apresentados, constata-se que o índice de perdas da Celesc é um dos menores entre as empresas citadas. Todavia, existe espaço para melhorar ainda mais, acompanhando-se e equacionando-se as perdas técnicas, combatendo-se as fraudes e irregularidades nas unidades de medição e mantendo-se o seu cadastro íntegro.

Tabela 2.2 - Índices de perdas nos sistemas MT e BT

ANO	CELESC	COPEL	CPFL	CELPE	CEB	LIGHT	CEMIG
2000	7.78	6.60	6,54	19.40	7,32	10.73	7,81
1999	7.74	6.59	7.70	20.22	8.80	14.90	7.97
1998	7.19	6.22	6.56	20.10	9.61	14.56	8.50
1997	7.62	6.41	6.11	17.72	8.65	16.09	8.47

continua

ANO								conclusão
	CELESC	COPEL	CPFL	CELPE	CEB	LIGHT	CEMIG	
1996	8.22	4.94	6.21	17.07	8.47	16.43	9.04	
1995	7.29	6.07	5.86	15.80	7.93	15.85	8.64	
1994	7.89	6.20	6.87	15.66	10.35	15.46	8.06	
1993	7.55	5.50	6.20	14.80	9.50	14.30	10.00	
1992	7.24	6.60	5.90	13.10	8.90	13.00	9.00	

Fonte: ABRADDEE

A meta de fiscalização dos consumidores do grupo “B” (monofásicos e bifásicos) gira em torno de 8% na Celesc. Porém, todas as agências ultrapassam esse limite. Na Tabela 2.3 apresenta-se um quadro-resumo da fiscalização dos consumidores do grupo “B” (monofásicos e bifásicos), realizada por todas as agências durante o ano de 2001. Pelos resultados, pode-se notar que a agência de Florianópolis e a de Itajaí apresentaram um volume de fraudes superior às demais. Nota-se também que existe um número considerável de irregularidades na medição que não constituem fraude, porém afetam diretamente a arrecadação da empresa.

Tabela 2.3 - Resultado da fiscalização de consumidores do grupo B em 2001 - Monofásico e Bifásico

Agência	Fiscalizado	C/ Influência	Fraudes
FNS	61.085	346	317
BNU	8.557	278	41
JVE	12.379	349	71
LGS	7.724	53	47
VII	1.741	53	5
CDA	6.857	182	0
JGS	2.793	113	27
JCA	3.857	110	2
CUA	3.723	132	97
SGE	5.280	273	4
TRO	4.952	699	63
RSL	5.470	87	12
MFA	4.333	242	4
SBS	3.046	35	13
IAI	26.384	493	219
CCO	5.375	213	51
DVMD	1.012	28	12
Total	164.568	3.686	985

Fonte: DPSC/Celesc

Para os consumidores do grupo “B” trifásicos, a meta de fiscalização é cobrir 33% do total de cada agência, anualmente. Na Tabela 2.4, constata-se que nas agências de Videira e Itajaí, em 2001, ocorreram mais irregularidades na medição, ficando Florianópolis com o maior número de fraudes.

Tabela 2.4 - Resultado da fiscalização de consumidores do grupo B em 2001 - Trifásicos

Agência	Fiscalizado	C/ Influência	Fraudes
FNS	13.434	369	80
BNU	2.328	429	8
JVE	3.354	659	82
LGS	1.080	3	3
VII	2.961	1.727	4
JGS	834	86	11
JCA	5.997	245	0
CUA	1.482	262	71
SGE	3.940	152	0
TRO	1.340	551	4
RSL	2.339	43	2
MFA	1.308	132	0
SBS	1.296	176	10
IAI	7.444	963	73
CCO	2.215	175	8
DVMD	2.309	263	42
Total	53.386	2.180	398

Fonte: DPSC/Celesc

Todos os consumidores do grupo “A” devem ser inspecionados no decorrer de um ano (meta vigente na empresa). O número de fraudes nesse grupo de consumidores é mais reduzido, porém o seu impacto na arrecadação é muitas vezes superior ao do grupo “B”. Pode-se observar pela Tabela 2.5 que a agência de Blumenau teve o maior número de fraudadores no ano de 2001.

Tabela 2.5 - Resultado da fiscalização de consumidores do grupo A em 2001

Agência	Fiscalizado	C/ Influência	Fraudes
FNS	4.575	1	1
BNU	859	70	39
JVE	1.161	0	0
LGS	355	0	0
VII	411	5	0
CDA	379	20	0
JGS	241	0	0
JCA	462	39	0
CUA	757	26	3
SGE	192	6	0
TRO	513	23	3
RSL	599	25	1
MFA	405	19	0
SBS	490	1	1
IAI	926	47	0
CCO	378	15	1
DVMD	2784	83	3
Total	15.487	380	52

Fonte: DPSC/Celesc

O volume de recursos que se deixa de arrecadar por anormalidades na medição devido às fraudes e demais irregularidades é um assunto que merece bastante atenção por parte da empresa. Apresenta-se na Tabela 2.6 o total de energia comprada e não faturada por falhas e fraudes detectadas pela empresa. Com um estudo mais aprofundado sobre perdas nos alimentadores, poder-se-á ter uma idéia mais precisa do montante real de perdas comerciais e técnicas na empresa. Acredita-se em um valor bem mais expressivo.

Tabela 2.6 - Resultados financeiros das perdas comerciais na Celesc

continua

ANO	FRAUDE		IRREGULARIDADE		TOTAL	
	KWH	RS	KWH	RS	KWH	RS
1995	3.285.279	341.529,08	2.432.256	343.420,71	5.717.535	684.931,79
1996	1.950.404	304.627,78	2.032.260	288.209,30	3.982.664	592.837,08
1997	7.336.010	648.011,69	2.483.182	295.984,53	9.819.192	943.996,22

conclusão

ANO	FRAUDE		IRREGULARIDADE		TOTAL	
	KWH	RS	KWH	RS	KWH	RS
1998	3.517.291	468.641,02	4.185.889	511.900,05	7.703.180	980.541,07
1999	8.927.056	1.063.910,58	5.043.955	524.374,76	13.971.011	1.588.285,34
2000	9.340.184	1.444.871,87	8.025.822	1.472.443,11	17.366.006	2.917.314,25
2001	11.579.565	1.967.716,90	10.433.568	1.825.829,32	22.013.133	3.793.546,22

Fonte: DPSC/DVMD/Celesc

2.9 Soluções sendo pesquisadas por empresas do setor elétrico

Como mencionado anteriormente, as empresas de distribuição de energia elétrica estão perseguindo de várias formas a redução de custos, visando a um decorrente aumento de receita. O item “perda comercial de energia” é um dos grandes sumidouros de recursos financeiros. Várias empresas nacionais de distribuição de energia elétrica estão buscando alternativas que possibilitem a redução desse item de custo.

Como exemplos de soluções desenvolvidas para combater a fraude no consumo de energia elétrica serão descritos quatro trabalhos.

O primeiro chama-se *Medições às Claras*, que é uma solução desenvolvida pelo Grupo Rede e implementada na Companhia Energética do Pará, sua afiliada. O segundo trabalho, *Sistema Integrado de Medição e Gerenciamento de Perdas*, está sendo desenvolvido pela Eletrobrás. O terceiro é o *smart card*, um sistema pré-pago de consumo de energia, apresentado no I Congresso de Inovação Tecnológica em Energia Elétrica (I CITENEL), realizado em Brasília, em novembro de 2001. O quarto e último trabalho é o *Sistema Centralizado de Medição*, que está em experiência na Celesc.

2.9.1 Medições às claras

É um sistema criado pelo Grupo Rede, cuja afiliada é a empresa Centrais Elétricas do Pará, para controlar o consumo de energia elétrica. É formado pela Caixa Padrão de Rede (CPREDE), fixada nos postes, e por um conjunto de softwares. Os softwares atuam basicamente no controle entre a energia comprada e a energia faturada.

O *medições às claras* foi criado inicialmente para resolver o problema dos funcionários da CELPA que tinham dificuldades para chegar até os medidores para efetuar a

leitura. Em 80% dos casos, os aparelhos ficavam dentro das residências. Os medidores foram então substituídos por aparelhos fixados nos postes, com o número da residência do consumidor e uma lente de aumento para propiciar a leitura do consumo.

Segundo a empresa, as perdas comerciais, que eram de 29% em agosto de 1998, passaram para 12% em 2000 e para 9,5% em 2001.

Em maio de 2001, a empresa já tinha instalado 410 mil medidores nos postes do Pará e vendido 160 mil para outras concessionárias (LIGHT, COELBA, ESCELSA).

2.9.2 Sistema integrado de medição e gerenciamento de perdas

A partir do conceito de medição centralizada, o Centro de Pesquisa da Eletrobrás (CEPEL) desenvolveu um medidor e uma metodologia para reduzir ao mínimo as perdas comerciais. O sistema de medição é formado por unidade de medição eletrônica, medidor eletrônico, unidade de leitura e medidor globalizador. Esse sistema pode ser simplificado para uma simples caixa fixada no poste. Através de ondas de rádio, pode ser controlado e monitorado à distância.

Alguns benefícios desse sistema são:

- a) medição eletrônica;
- b) leitura automática para faturamento;
- c) desconexão e religamentos remotos;
- d) informações para o cálculo de DIC e FIC;
- e) curva de carga dos consumidores e flexibilidade para futuras expansões;
- f) redução de perdas com fraudadores e irregularidades na medição;
- g) redução de perdas com inadimplentes.

A CEPEL está desenvolvendo também um software para detecção de fraudes, cuja seleção de análise pode ser realizada por:

- a) informações cadastrais;
- b) faixa de consumo;
- c) base de fraudadores;
- d) queda de um percentual do consumo;

- e) redes neurais artificiais;
- f) visualização gráfica do consumo médio anual;
- g) visualização gráfica do consumo por ano;
- h) visualização gráfica do consumo mensal;
- i) visualização da média móvel do consumo mensal;
- j) padrão de consumo.

2.9.3 Energia pré-paga

Um *smart card* (cartão inteligente) que permite a adoção do sistema pré-pago no setor energético foi apresentado no I Congresso de Inovação Tecnológica em Energia Elétrica (I CITENEL), realizado em Brasília, em novembro de 2001.

O invento é proveniente do Instituto de Tecnologia para o Desenvolvimento (Latec) do Paraná e foi viabilizado pela empresa Procomp.

O cartão é abastecido com créditos por uma máquina especial. Quando o crédito estiver acabando, uma mensagem aparecerá no visor, avisando que é hora de carregar o cartão novamente.

O consumidor terá a opção de, minuto a minuto, saber o quanto está gastando.

O maior benefício ficaria por conta da emissão de fatura que atualmente custa para uma distribuidora em torno de R\$ 1,70 por conta.

Com relação às fraudes, haveria pontos positivos e negativos. O desvio de energia, por exemplo, estaria praticamente no mesmo patamar, uma vez que o equipamento estará sob o domínio do consumidor. Poderão ocorrer também fraudes no próprio equipamento, usando-se energia sem consumo de créditos. As fraudes cometidas através do medidor desapareceriam, pois o novo equipamento não é eletromecânico.

2.9.4 Sistema de medição centralizada

Também chamada de Tarifa Amarela, é um sistema que visa a reduzir o consumo de energia na hora de pico, que compreende o período entre 18h30min e 21h30min. O consumo de energia no horário de pico que ficar abaixo de um valor predeterminado sofre uma redução de 50%.

É um projeto experimental que está sendo testado em Joinville, em 450 residências do Conjunto Habitacional Ademar Garcia. Para participar desse experimento, o consumidor deve concordar formalmente com o novo método.

Nesse novo sistema de medição, a unidade de medição eletrônica (que comporta até oito residências) fica fixada no poste. Em um ponto determinado do Conjunto Habitacional, instalou-se um concentrador de leitura, que centraliza as leituras instantâneas de consumo. O consumidor tem acesso ao consumo, na ponta e fora de ponta, a qualquer horário, de uma cabine munida de monitor e teclado.

A leitura é efetuada remotamente. A área de medição realiza essa tarefa através de ligação via modem. É possível efetuar cortes e religar também remotamente, de forma rápida e fácil, diferentemente do modelo manual atual, constrangedor e burocrático.

Aliado ao objetivo de diminuir o consumo no horário de pico, esse sistema também inviabiliza as fraudes que eram realizadas nos medidores de energia como também nos desvios feitos antes da entrada do medidor, em derivações realizadas no interior das residências.

2.10 Reestruturação do setor elétrico brasileiro

A reestruturação do setor de energia elétrica brasileiro tem como principal característica a desverticalização da indústria, isto é, as atividades de geração, transmissão, distribuição e comercialização são agora desempenhadas de forma autônoma e independente (CEPEL, 1999).

O mercado torna-se competitivo, e todos os interessados em participar dele têm livre acesso aos sistemas de transmissão e distribuição.

A característica principal desse novo modelo mercantil é a criação de um mercado chamado Mercado Atacadista de Energia (MAE).

São requisitos básicos nesse novo modelo (CEPEL, 1999):

- a) eficiência econômica da indústria;
- b) auto-sustentação da indústria, de modo a garantir a expansão do sistema;
- c) operação do sistema com elevado grau de confiabilidade em consonância com requisitos de qualidade impostos pela sociedade;

d) a prestação dos serviços não pode ser discriminatória e a universalização destes deve ser perseguida.

A partir desse novo modelo do setor elétrico, surgiram as figuras dos agentes (os principais são ONS e ANEEL), para operar, planejar e regular o sistema elétrico brasileiro. O Operador Nacional do Sistema Elétrico (ONS) é a entidade responsável pelo planejamento operacional, programação e despacho. A Agência Nacional de Energia Elétrica (ANEEL) é o órgão máximo de regulamentação do setor elétrico nacional. Para atender ao Mercado Aberto de Energia (MAE), foi criada a figura do operador de mercado, denominado Agente Negociador do Mercado Aberto de Energia (ASMAE).

A função da produção da indústria ficará a cargo dos agentes de produção (geração), ficando aberta à concorrência.

A função de transporte da energia é desempenhada pelos agentes de transmissão e distribuição, sendo necessário que estes ofereçam o livre acesso a todos os agentes do mercado de forma não discriminatória. A transmissão e distribuição constituem monopólios naturais.

A partir de 2005, qualquer consumidor poderá escolher a companhia comercializadora de sua preferência que fará o seu suprimento de energia, independentemente de esta ser a companhia que lhe presta o serviço de distribuição, desde que pague pelo uso do sistema de transmissão e distribuição.

Os contratos de compra e venda de energia podem ser efetuados das seguintes formas (CEPEL, 1999):

- a) contratos bilaterais entre comercializadores de energia e geradores, com preços acertados livremente;
- b) contratos bilaterais entre comercializadores de energia, com preços acertados livremente;
- c) através do mercado *spot*, com preços que reflitam o custo marginal do sistema.

O agente regulador (ANEEL) tem a atribuição de fixar preços e padrões de qualidade, estimulando a eficiência econômica da indústria e a universalização do serviço, evitando abusos nas estruturas de custo do sistema.

As funções de regulações técnicas da ANEEL são (CEPEL, 1999):

- a) critérios de expansão da geração - atendimento de ponta e energia;

- b) critérios de expansão da transmissão e distribuição;
- c) qualidade do fornecimento - penalidades;
- d) interrupções do fornecimento – penalidades.

As funções de regulações econômicas da ANEEL são (CEPEL, 1999):

- a) controle dos preços praticados nos serviços de transmissão e distribuição;
- b) controle dos preços para os consumidores cativos;
- c) preservação do equilíbrio econômico-financeiro dos agentes.

2.11 Considerações finais

Neste capítulo descreveu-se brevemente a história do setor elétrico brasileiro, iniciando-se com o interesse de D. Pedro II pelos inventos de Thomas Edison, chegando-se até os dias atuais, em que o governo de Fernando Henrique Cardoso inicia o processo de privatização do setor. A história da Celesc também foi relacionada para se conhecer um pouco da empresa que serviu de laboratório para que a arquitetura de informação proposta fosse trabalhada.

Conceituaram-se as perdas existentes na distribuição de energia elétrica e o modo como a Celesc e outras empresas do ramo as gerenciam.

A Celesc possui baixos índices de perdas de energia elétrica, sendo uma das melhores do setor, estando muito próxima da COPEL (a melhor do setor em termos de baixas perdas). Mesmo assim, estima-se que as perdas comerciais anuais cheguem a um valor aproximado de 80 milhões de reais. A CELPE calcula que atualmente suas perdas globais anuais giram em torno de 900 milhões de reais, e a Light, de 300 milhões. São valores expressivos que justificam todo o esforço das empresas no sentido de combater tal sumidouro de recursos.

Atualmente, o Brasil possui 80% das empresas distribuidoras de energia elétrica privatizadas. Mais do que nunca, as perdas de energia passaram a ter o merecido tratamento, pois agora toda despesa desnecessária é combatida, e todo lucro possível é perseguido. Algumas soluções que estão sendo pesquisadas estão relacionadas no item 2.8.

A partir de 2005, as empresas distribuidoras de energia elétrica estarão expostas à concorrência, uma vez que o mercado será livre. A competitividade forçará as empresas a terem um controle de custos mais enxuto e um fornecimento de energia de alta qualidade. Quem tiver um bom programa de combate às perdas de energia e de previsão de cargas levará uma boa vantagem sobre as demais que iniciarem tardiamente esse processo.

3 MINERAÇÃO DE DADOS

3.1 Introdução

Atualmente em muitas organizações o volume de dados é tão grande que manipulá-los por métodos e linguagens de programação tradicionais torna-se uma atividade bastante complexa e laboriosa. Todavia, essas organizações necessitam de informações para atender melhor aos seus clientes, alocar os recursos de forma mais proveitosa e minimizar perdas. Ou seja, necessitam de informações úteis à tomada de decisão.

Para tratar grandes volumes de dados e extrair deles informações e questionamentos, pode-se utilizar o processo conhecido por *Knowledge Discovery in Database* (KDD), processo este que envolve a seleção, o pré-processamento e a transformação dos dados, como também a aplicação de algoritmos, a interpretação dos resultados e a geração de conhecimento (FAYYAD et al., 1996).

Fayyad et al. (1996) destacam que a mineração de dados (MD) é uma das etapas do KDD, em que ocorre a aplicação dos algoritmos com a finalidade de identificar padrões em um conjunto de dados ou a geração de um conjunto de regras que descrevem o comportamento desses dados.

Em linhas gerais pode-se dizer que MD é a exploração e análise, realizada de forma automática ou semi-automática, de grandes quantidades de dados com o intuito de descobrir padrões e regras significativas (BERRY e LINOFF, 1997; HARRISON, 1998).

A MD usa uma proposta alternativa que parte da premissa de que não se sabe quais padrões comportamentais existem em uma base de dados de clientes, por exemplo. A MD providencia respostas, sem que se tenha proposto questões específicas (BARAGOIN, 2001).

Segundo Berry e Linoff (1997), existem basicamente dois modelos de MD: o probabilístico e o descritivo. O modelo probabilístico parte do uso dos dados e resultados conhecidos, para desenvolver modelos que possam prever resultados a partir de diferentes dados. Já o modelo descritivo procura por padrões que descrevem os dados e que sejam interpretáveis, para utilizá-los na tomada de decisão.

Na definição do problema, deve-se levar em consideração as metas desejadas pelo usuário como também os critérios de desempenho e interoperabilidade com o usuário final.

No pré-processamento é feita a limpeza dos dados, com a remoção dos dados inválidos e dos que possuem ruídos; os dados perdidos também são tratados nessa etapa.

Na etapa de representação dos dados, procura-se modelá-los de maneira tal que possam ser trabalhados adequadamente pelos algoritmos de extração de conhecimento.

Na fase de mineração dos dados, buscam-se informações relevantes que não sejam facilmente identificáveis.

Ao se utilizar a MD para a extração de conhecimentos, faz-se uso das seguintes etapas (FAYYAD, 1996):

- **definição do processo:** basicamente classificação, agrupamento ou sumarização;
- **seleção do método:** estatística, redes neurais, árvores de decisão, além de outros;
- **processo de mineração de dados:** dados de entrada, processamento e saída;
- **interpretação dos resultados:** verifica-se a validade do conhecimento extraído da base de dados, apresentando-o na forma que o usuário entenda melhor.

Segundo Chen et al. (1996), um sistema de mineração de dados pode ser classificado levando-se em conta os seguintes aspectos:

- **Tipo da base de dados:** dá-se o nome de MD Relacional àquela que extrai conhecimentos de uma base de dados relacional. Quando a MD trabalha em uma base de dados orientada a objetos, recebe o nome de MD Orientada a Objetos.
- **Tipo de conhecimento:** um primeiro modelo preocupa-se com o conhecimento propriamente dito, incluindo-se entre estes modelos os que adotam regras de associação, classificação, agrupamento, evolução, análise de desvios, entre outros. Um segundo modelo está ligado ao nível de abstração do conhecimento descoberto, podendo ser genérico, específico ou múltiplo.
- **Tipo de técnica:** para minerar os dados na busca de conhecimentos, um conjunto de técnicas pode ser explorado. Tais técnicas estão intimamente ligadas às informações que se deseja obter e aos dados com que se está trabalhando. Considerando-se a abordagem que se vai fazer, esta pode ser baseada na teoria estatística, em padrões e em generalizações.

3.2 Tarefas realizadas no uso das técnicas de mineração de dados

Com relação à busca de conhecimentos em uma base de dados, pode-se agrupar as principais tarefas realizadas pelas técnicas de MD em: classificação, estimativa, associação e agrupamento.

a) Classificação

A classificação tem por objetivo a construção de um modelo que possa ser aplicado a dados não classificados, visando a categorizá-los em classes. O dado assim categorizado pode ser examinado de acordo com as características de sua classe (HARRISON, 1998).

Para Fausett (1994), em um problema envolvendo padrões de classificação, cada dado de entrada irá pertencer ou não a uma classe particular, também chamada de categoria.

A classificação, conforme Baragoin et al. (2001), nada mais é do que o assinalamento de registros de dados para categorias predefinidas. Quando um novo registro é classificado, a técnica utilizada determina a categoria e a probabilidade de o registro pertencer a essa categoria.

Freitas (2000) vincula a tarefa de classificação a uma função não definida e indeterminística.

b) Estimativa

Em um modelo de regressão, ou estimativa, usa-se uma série de valores existentes para que se possa prever outros. Valores difíceis de serem preditos por outras técnicas (previsão de vendas, custos de armazenamentos, taxas de falhas, etc.) podem ser calculados por meio desse modelo.

Segundo Harrison (1998), a estimativa pode ser muito bem explorada para definir um valor para alguma variável contínua desconhecida (receita, saldo, altura, etc.).

A regressão, na prática, é muito utilizada para fazer classificação (BERRY e LINOFF, 1997).

c) Associação

Para Witten e Frank (2000), a associação não difere muito da classificação, com exceção da capacidade de predizer qualquer atributo, e não somente classe. Isso possibilita a independência de predizer combinação de atributos.

Através da tarefa de associação, determinam-se quais itens tendem a ocorrer juntos em uma transação (HARRISON, 1998). É um modelo que pode ser utilizado na área de supermercados, por exemplo, para auxiliar na disposição de produtos nas prateleiras, de acordo com hábitos de compras dos consumidores.

A associação pode ser considerada uma tarefa relativamente simples e determinística (FREITAS, 2000).

d) Agrupamento

Utiliza-se esta tarefa para segmentar um conjunto de dados em grupos diferentes, ou categorias, reunindo-se em cada grupo elementos com características semelhantes.

No agrupamento, ou segmentação, uma população heterogênea é constituída de subgrupos homogêneos, denominados grupos (HARRISON, 1998; BERRY e LINOFF, 1997). Nesta tarefa não existem classes predefinidas ou exemplos, como ocorre na classificação. As ocorrências são reunidas com base na sua similaridade.

Para Baragoïn et al. (2001) e Hair et al., (1998), o termo “segmentação” é utilizado para designar um conjunto de técnicas que possibilita o agrupamento de um conjunto de dados pelo grau de similaridade que eles possuem entre si.

Cada um dos agrupamentos constituídos é definido basicamente em função da maximização da similaridade intraclasses e na minimização da similaridade interclasses (CHEN et al., 1996; JOHNSON e WICHERN, 1998).

3.3 Técnicas de mineração de dados

A mineração de dados utiliza diversas técnicas para detectar padrões de conhecimentos automaticamente, fazendo uma análise inteligente dos dados. Existem diferentes técnicas para atender a diferentes propósitos.

É determinante que se conheça a fundo a natureza do problema que será tratado na mineração de dados, como também que se tenha um bom conhecimento das técnicas existentes, antes de definir qual técnica usar.

a) Árvore de decisão

A técnica *árvore de decisão* é um modelo produzido por um conjunto de técnicas que inclui classificação (CART) e indução automática quiquadrado (CHAID). Esta técnica divide os registros, numa fase de treinamento, em subconjuntos disjuntos de dados. Cada um desses subconjuntos é descrito por uma regra simples em um ou mais atributos (BERRY e LINOFF, 1997).

Em uma árvore de decisão, cada nó não terminal representa um teste ou decisão sobre o item de dado que está sendo considerado (GOEBEL e GRUENWALD, 1999).

A sua estrutura funcional é bastante simples: os dados são divididos em subgrupos, de acordo com os valores das variáveis envolvidas. A constituição do resultado se dá através de uma hierarquia de declarações do tipo “se ... então ...”, utilizada principalmente na classificação de dados.

Existem vários algoritmos que implementam a técnica de árvore de decisão. Alguns deles são os seguintes: CART, CHAID, C5.0, ID-3 (CHEN et al., 1996), SLIQ (METHA et al., 1996) e SPRINT (SHAFER et al., 1996).

b) Regras de associação

Esta técnica consiste basicamente da descoberta de relacionamentos entre variáveis. Ela é formada de dois parâmetros básicos: um, que estabelece o suporte, e o outro, que estabelece a confiança. O suporte pode ser denominado o número mínimo de ocorrência de $X \cup Y$, ou seja, a união de itens no conseqüente e antecedente da regra está presente em um suporte mínimo S% na base (AGRAWAL, R.; IMICLINSKI, T., SWAMI, 2002). A confiança é estabelecida por um valor percentual das transações realizadas em uma base de dados que satisfaz o antecedente da regra (X) e também o conseqüente da regra (Y) (CHEUNG et al., 1996).

Para Goebel e Gruenwald (1999), a regra de associação nada mais é do que uma correlação estatística entre itens de dados encontrada em um conjunto de dados.

O algoritmo *Apriori* (AGRAWAL e SRIKANT, 2002) é um dos principais algoritmos utilizados na descoberta de regras de associação. Existem várias derivações dele, como o algoritmo DHP (CHEN et al., 1996), AprioriTid, Partion, AIS, entre outros.

c) Raciocínio baseado em casos

Esta técnica é também conhecida como MBR (*Memory-Based Reasoning*).

Esta técnica utiliza exemplos conhecidos como um modelo para previsões sobre casos não conhecidos (BERRY; LINOFF, 1997). Uma das grandes vantagens deste modelo é a capacidade de executar virtualmente qualquer fonte de dados, sem qualquer modificação desses dados.

Na busca dos vizinhos mais próximos, através dos exemplos já conhecidos, a MBR combina seus valores para atribuir valores de classificação ou de previsão (HARRISON, 1998). Para cada novo dado processado, todos os dados são utilizados, para que seja encontrado um subconjunto de dados que seja semelhante ao novo dado. Isso a torna não muito eficaz, pois os dados têm que ser lidos em sua totalidade. Para tornar um processo de classificação mais eficaz, esta técnica é utilizada em conjunto com outras técnicas.

CLIQUE (AGRAWAL et al., 2002) E BIRCH (ZHANG et al., 1996) são dois exemplos de algoritmos que utilizam a técnica MBR.

d) Algoritmos genéticos

Baseia-se nos conceitos genéticos de evolução de seres biológicos, mutação e seleção natural, para resolver problemas de otimização e aprendizado.

Para Goebel e Gruenwald (1999), o algoritmo genético pode ser definido como um método iterativo utilizado para fazer evoluir uma população de organismos. É usado em MD para formular hipóteses sobre dependências entre variáveis, de maneira tal que estabeleça um formalismo interno.

O processamento genético serve para otimizar a aptidão de indivíduos sobre gerações sucessivas, propagando o material genético através dos indivíduos mais aptos, de uma geração para outra (BERRY e LINOFF, 1997).

Utilizando-se de operadores de seleção, cruzamento e mutação, sucessivas gerações de soluções são formadas. Com a evolução das soluções, somente aquelas com maior poder de

previsão sobrevivem, chegando-se na otimização da solução, ou seja, na convergência de uma solução ideal (HARRISON, 1998).

Não são poucos os exemplos de algoritmos genéticos utilizados em soluções de otimização. Pode-se citar os trabalhos de GENITOR e CHC (WHITLEY, 1993), GA-NUGGETS e GA-PVMINER (ARAÚJO et al., 1999).

e) Redes neurais artificiais

A RNA tem como modelo uma rede neural biológica.

Segundo Goebel e Gruenwld (1999), as redes neurais artificiais (RNA), ou simplesmente redes neurais, constituem um sistema que processa as informações de forma análoga ao funcionamento do cérebro humano, tendo-se em sua arquitetura neurônios artificiais conectados de maneira similar aos neurônios do cérebro humano.

Uma RNA é caracterizada pelo padrão de conexão entre os neurônios (arquitetura), pelo método de determinação dos pesos das conexões (treinamento ou aprendizagem) e pela função de ativação (FAUSETT, 1994).

No item 3.5, podem ser encontradas informações adicionais sobre esta técnica.

3.4 Redes neurais artificiais

As Redes Neurais Artificiais (RNA) são um dos subgrupos de IA (Inteligência Artificial), cujo intuito é buscar através de um modelo artificial de processamento uma aproximação com a inteligência humana.

Pode-se definir RNA como sistemas paralelos distribuídos, compostos por unidades de processamento simples, que calculam determinadas funções matemáticas, normalmente não lineares (BRAGA et al., 2000).

Na visão de Fausett (1994), RNA seria um modelo de processar a informação que possui certas características em comum com a rede neural biológica. Tal processamento é realizado através da aplicação de modelos matemáticos.

Um das principais características de uma RNA é a sua capacidade de aprender a partir de seu ambiente e de aperfeiçoar sua performance através do aprendizado (HAYKIN, 1994; BRAGA et al., 2000).

Um dos problemas na utilização de RNA, conforme Harrison (1998), é que os dados de entrada da rede são difíceis de serem formatados, e os modelos produzidos são de difícil entendimento.

Como características básicas de uma RNA (PANDYA et al., 1995) pode-se citar:

1. *Adaptabilidade*: a utilização de regras auto-organizáveis permite a adaptação em ambientes básicos;
2. *Processamento*: a RNA possui a capacidade de executar tarefas de relacionamento não lineares e tolerantes a ruídos;
3. *Processamento paralelo*: as várias unidades de processamento possibilitam o armazenamento de informações distribuídas.

Uma RNA básica é caracterizada por meio de três componentes (FAUSSET, 1994):

1. *Arquitetura*: definida pelo padrão de conexão entre os neurônios da rede;
2. *Algoritmo, treinamento ou aprendizagem*: é o método que define o conhecimento da rede, ou seja, estabelece os pesos das conexões;
3. *Função de ativação*: responsável pela produção da saída da rede.

Segundo Skapura (1996), ao se utilizar RNA como solução de um determinado problema, três itens merecem especial atenção: (1) o processo de aquisição e modelagem dos dados, (2) a escolha da arquitetura mais adequada para ser aplicada na solução do problema e (3) o treinamento da rede.

Neste item, serão apresentadas as redes MLP e RBF, pela importância que elas assumem no processo de classificação (TODESCO, 1995; MOREAU e VANDEWALLE, 1997), e a rede de Kohonen, pelo seu uso em problemas que envolvem segmentação (NOTARE, 2000). Uma outra técnica descrita neste item é a segmentação demográfica, utilizada na segmentação de consumidores industriais.

3.4.1 Breve histórico

A história de RNA inicia-se em 1943, com Warren McCulloch e Walter Pitts, quando eles definem o primeiro modelo artificial de um neurônio biológico.

No ano de 1949, Donald Hebb apresenta o primeiro trabalho que relaciona aprendizado às redes biológicas artificiais.

Em 1958, Frank Rosenblatt vislumbrou o uso de RNA para classificar certos tipos de padrões, criando o modelo Perceptron.

A rede Perceptron recebeu uma pesada crítica em 1969, através de Minsk e Papert, pois segundo eles, o modelo Perceptron só resolveria problemas linearmente separáveis. O Perceptron não resolveria, por exemplo, problemas de paridade e conectividade, que não são linearmente separáveis.

A crítica à rede Perceptron repercutiu negativamente para o desenvolvimento e uso de RNA. Nos anos 70, ocorreu um certo desaquecimento da abordagem conexionista, motivado em grande parte pela repercussão do trabalho de Minsk e Papert.

O ressurgimento da importância de RNA somente ocorre em 1982, quando se retoma o interesse a partir do artigo publicado por John Hopfield, no qual são descritas as propriedades associativas de uma RNA.

Em meados da década de 80, ocorre uma explosão de interesse por RNA, motivada pelos avanços tecnológicos e pelo pouco avanço da escola simbolista na solução de alguns problemas simples para o ser humano.

3.4.2 Principais arquiteturas

A arquitetura de uma rede neural é determinada pelo número de camadas da rede, número de nodos em cada camada, tipo de conexão entre os nodos e topologia (BRAGA et al., 2000).

Quanto ao número de camadas, a RNA pode ser classificada em:

1. *Camada única*: só existe um nó entre qualquer entrada e qualquer saída da rede.
2. *Múltiplas camadas*: entre a entrada da rede e a saída existe uma ou mais camadas.

Quanto ao número de conexões, a RNA pode ser:

1. *Feedforward*: o sinal da rede é propagado somente para a frente, ou seja, nenhuma saída retorna a uma camada anterior;
2. *Feedback*: possibilita a retropropagação, ou seja, a saída de um neurônio pode servir de entrada na mesma camada ou em uma camada anterior.

Quanto à conectividade, a RNA pode ser classificada em:

1. *Parcialmente conectada*: cada entrada da rede não se liga a todos os neurônios;
2. *Completamente conectada*: cada entrada se liga a cada um dos neurônios da camada seguinte.

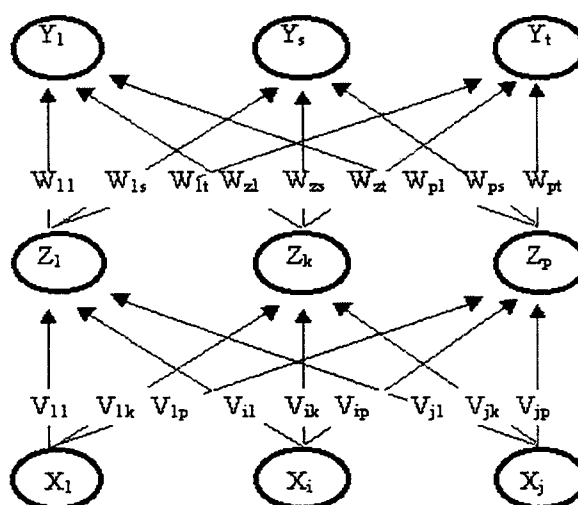
3.4.3 Rede Perceptron de Múltiplas Camadas (MLP)

A RNA MLP é composta de neurônios do tipo Perceptron, proposta por Rosenblatt em 1962. Ela consiste de uma camada de entrada, uma ou mais camadas intermediárias e uma camada de saída (TODESCO, 1995). É uma RNA não linear, de múltiplas camadas e de encaminhamento para frente (*feedforward*), ou seja, os dados seguem da entrada com fluxo para saída sem retrocesso. Segundo Skapura (1996), a rede MLP, treinada com o algoritmo *Backpropagation*, é o tipo de RNA de propósito geral mais utilizado.

Por possuírem camadas intermediárias, as redes MLP conseguem um poder computacional bem maior que aquelas que não as possuem as camadas intermediárias (BRAGA et al., 2000). Já a saída da rede MLP é calculada explicitamente em função dos valores de entrada e dos pesos (BISHOP, 1997).

Uma rede MLP, como pode ser visto na Figura 3.1, é composta de uma camada de entrada, uma ou mais camadas intermediárias e uma camada de saída. Os valores de entrada são inseridos na rede através da camada de entrada; a camada intermediária armazena o conhecimento da rede, e a camada de saída manipula os valores de saída da rede. A matriz de pesos representa a intensidade de conexão dos valores de entrada em cada um dos neurônios da camada intermediária.

Figura 3.1 - Rede MLP de 3 camadas



Diversas funções não lineares e diferenciáveis têm sido propostas como função de ativação. Entre elas, a mais utilizada é a sigmoideal logística (BRAGA et al., 2000), binária e bipolar.

A função binária (3.1) trabalha com valores de saída entre 0 e 1, e a bipolar (3.2), com valores entre -1 e 1.

A função sigmoïdal binária é definida pela fórmula (3.1)

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (3.1)$$

A definição da função sigmoïdal bipolar é dada pela fórmula (3.2)

$$f(x) = \frac{2}{(1 + \exp(-x)) - 1} \quad (3.2)$$

onde “X” representa o nível de ativação do neurônio

Treinamento de uma Rede MLP

Para o treinamento de redes MLP o algoritmo comumente utilizado é o *Backpropagation* (TODESCO, 1995), também conhecido como *Generalized Delta Rule* (FAUSETT, 1994). Ele se baseia no método da gradiente descendente para minimizar o erro quadrático total da saída da rede (FAUSETT, 1994; HAYKIN, 1994).

O treinamento de uma rede MLP ocorre em duas etapas distintas: na primeira, define-se a saída da rede para um dado padrão de entrada e, na segunda fase, utiliza-se a saída desejada e a saída fornecida pela rede para atualizar os pesos de suas conexões, desde a saída até a entrada.

Um algoritmo *Backpropagation* simplificado pode ser definido como (BRAGA et al., 2000):

- 1 - inicializar pesos e parâmetros;
- 2 - repetir até o erro ser mínimo ou até a realização de um dado número de ciclos:
 - 2.1 - para cada padrão de treinamento X:
 - 2.1.1 - definir saída da rede através da fase *forward*;
 - 2.1.2 - comparar saídas produzidas com as saídas desejadas;
 - 2.1.3 - atualizar pesos dos nodos através da fase *backward*.

A fase *forward* é composta dos seguintes passos:

1 - a entrada é apresentada à primeira camada da rede (camada C^0);

2 - para cada camada C^i a partir da camada de entrada:

2.1 - após os nodos da camada C^i ($i > 0$) calcularem seus sinais de saída, estes servem como entrada para a definição das saídas produzidas pelos nodos da camada C^{i+1} .

3 - as saídas produzidas pelos nodos da última camada são comparadas às saídas desejadas.

A fase *backward* é composta dos seguintes passos:

1 - a partir da última camada, até chegar na camada de entrada:

1.1 - os nodos da camada atual ajustam seus pesos de forma a reduzir seus erros.

1.2 - o erro de um nodo das camadas intermediárias é calculado utilizando-se os erros dos nodos da camada seguinte conectados a ele, ponderados pelos pesos das conexões entre eles.

Dois problemas são encontrados no uso do algoritmo *Backpropagation*: lentidão em superfícies mais complexas de erro e convergência para mínimos locais, em vez do global. Para amenizar esses problemas, seguem-se os procedimentos abaixo (BRAGA et al., 2000).

- Utiliza-se taxa de aprendizagem decrescente;
- adicionam-se nós intermediários;
- utiliza-se o termo *momentum*;
- adicionam-se ruídos aos dados.

O termo *momentum* auxilia na rapidez do treinamento em regiões muito planas da superfície do erro e também suprime as oscilações de pesos em vales e ravinas. Dessa forma, procura-se aumentar as chances de a rede estabilizar num ponto de erro mínimo global, evitando-se estabilização em um ponto mínimo local.

3.4.4 Rede Função de Base Radial (RBF)

Assim como a rede MLP, a rede RBF também possui múltiplas camadas. A propagação do sinal de entrada é no sentido entrada/saída (*feedforward*).

A base de funcionamento de uma rede RBF está montada na procura do ajuste de uma curva em um espaço de grande dimensão (HAYKIN, 1994).

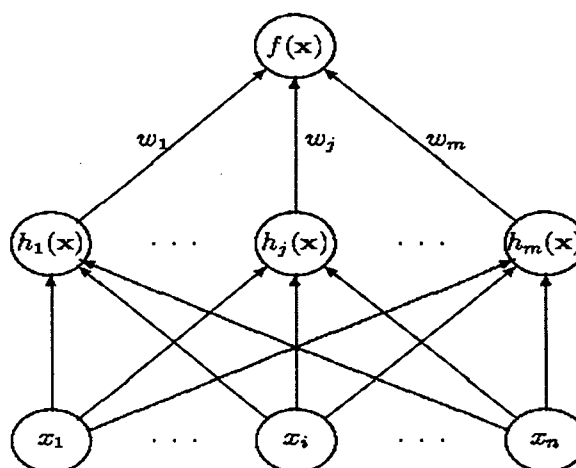
Como atesta Power (1987 apud BISHOP, 1997), o método função de base radial tem suas origens nas técnicas utilizadas em interpolação exata de um conjunto de dados em um espaço multidimensional.

Uma rede RBF típica é formada de uma camada de entrada, uma camada intermediária e uma saída. A camada de entrada é não linear, e a camada de saída fornece respostas lineares (TODESCO, 1995).

A RBF utiliza a distância euclidiana entre o vetor de entrada e o centro de cada unidade na camada intermediária como função de ativação (TODESCO, 1995; BRAGA et al., 2000).

Na Figura 3.2 é representada uma estrutura básica de uma rede RBF, com as camadas de entrada, intermediária e com a saída.

Figura 3.2 - Rede RBF de 3 camadas



Na primeira camada, os sinais de entrada são introduzidos nos respectivos neurônios de entrada da rede. Na segunda camada, os dados são agrupados em *clusters*, utilizando-se uma função de base radial. Esta camada transforma um conjunto de padrões de entrada não linearmente separáveis em saídas linearmente separáveis.

A camada de saída classifica os valores fornecidos pela camada intermediária, usando função de separação linear (regra delta, por exemplo).

As funções de bases radiais mais utilizadas são (BRAGA, 2000):

1. *Gaussiana*: $f(x) = \exp(-v^2 / 2\sigma^2)$
2. *Multiquadrática*: $f(x) = (v^2 + \sigma^2)^{-1/2}$
3. *Thin-Plate-Spline*: $f(x) = v^2 \cdot \log(v)$

onde,

$v = \|x - \mu\|$ é a distância Euclidiana

x = Vetor de entrada

μ = Largura da função radial

σ = Centro da função radial

A saída de cada nó de uma camada intermediária é definida por (3.3) (HAYKIN, 1994):

$$F(x) = \sum_{i=1}^M w_i h(\|x - x_i\|) \quad (3.3)$$

Onde,

$\{h(\|x - x_i\|) \mid i = 1, 2, \dots, M\}$ -

Conjunto das “M” funções (agrupamento) da camada oculta, conhecidas como funções de base radial;

x - Padrão de entrada;

$\|x - x_i\|$ -

Norma da distância (usualmente é a distância Euclidiana);

$x_i \in \mathbb{R}^p, i = 1, 2, \dots, M$ -

Centros das funções de base radial (agrupamentos);

W - Coeficientes desconhecidos (pesos) que minimizam a diferença da função $F(x)$ em relação ao alvo.

As principais características da arquitetura RBF, segundo Haykin (1994), são de duas naturezas, a saber:

- capacidade de aprendizado rápido, em problemas bem definidos;
- baixa sensibilidade à ordem de apresentação dos dados treinados.

Treinamento de uma Rede RBF

A maioria dos métodos de treinamento utiliza duas fases (BRAGA, 2000). Na primeira fase, o número de funções radiais e seus parâmetros são determinados por métodos não supervisionados. A segunda fase ajusta os pesos dos nodos da camada de saída.

Para a clusterização dos valores de entrada (primeira fase), processados na camada intermediária, o método mais utilizado é o algoritmo *K-Means-Clustering*, também conhecido como *K-Means* (HAYKIN, 1994).

Uma alternativa para o algoritmo *K-Means* é o mapa de Kohonen, conhecido como SOM, que também é utilizado para calcular centros de funções (BISHOP, 1997; TODESCO, 1995).

A segunda fase do treinamento é responsável pelo ajuste do peso da camada de saída (problema linear). Como se trata de uma solução de separação linear, uma alternativa seria usar o algoritmo *regra delta*.

Segundo Todesco (1995), a RBF é muito vantajosa com relação ao tempo de processamento, uma vez que esse tempo é bastante inferior quando comparada com outros tipos de RNA.

3.4.5 Mapa auto-organizável de Kohonen

O Mapa Auto-Organizável de Kohonen, ou rede de Kohonen, baseia-se na construção de um mapa topológico onde os nodos que estão topologicamente próximos respondem de forma semelhante a padrões de entrada semelhantes (BRAGA, 2000). Ele utiliza um algoritmo que se caracteriza pela competitividade, em que os nodos de saída competem entre si para se tornarem ativos, isto é, o nodo que produz a maior saída é consagrado vencedor, sofrendo então a atualização de seus pesos como também de sua vizinhança.

O algoritmo de Kohonen é composto dos seguintes passos (FAUSETT, 1994):

1 Inicialização dos pesos W_{ij} .

Definição da topologia da vizinhança

Atribuição da taxa de aprendizagem

2 Fazer enquanto a condição de parada é falsa

2.1 Para cada vetor X

2.1.1 Para cada j , fazer:

$$D(j) = \sum (W_{ij} - X_i)^2$$

2.1.2 Encontrar o índice J tal que $D(j)$ é mínimo

2.1.3 Para todas as unidades de J com uma determinada vizinhança de j , e para todos i , fazer:

$$W_{ij}(\text{novo}) = W_{ij}(\text{antigo}) + \alpha[X_i - W_{ij}(\text{antigo})]$$

2.2 Atualizar taxa de aprendizagem

2.3 Reduzir o raio da topologia da vizinhança

2.4 Testar condição de parada

A área de influência do nodo vencedor pode ter formato circular, quadrático ou hexagonal. A redução da área de influência geralmente é uma função linear do número de passos (BRAGA, 2000). A taxa de aprendizagem (α) é reduzida em função do número de passos da fase de treinamento (FAUSETT, 1994).

Uma vez definido o número de grupos que será utilizado, a sua inicialização pode ser feita de forma randômica, ou refletir o conhecimento prévio que se tenha a respeito da distribuição dos dados (FAUSETT, 1994).

3.4.6 Segmentação demográfica

A técnica de segmentação demográfica (SD) – *demographic clustering* – está intimamente ligada ao princípio chamado *condorset*, para medir as distâncias dos registros de entrada e assinalá-los para um grupo específico. O princípio *condorset* estabelece que o valor da similaridade entre os registros é dado pela soma das similaridades de pares de todas as entradas no mesmo grupo menos a soma das similaridades de pares de todos os registros em diferentes grupos.

Na execução do algoritmo, grupo e centros são atualizados continuamente em cada passo. A condição de parada se dá pelo número máximo de passos fornecido, ou quando o número máximo de grupos encontrado e os centros desses grupos não se alteram significativamente, de acordo com uma margem fornecida.

O algoritmo que esta técnica implementa é incremental, sendo os registros de entrada considerados um a um. Cada novo registro é inserido para um grupo existente, ou um novo grupo é criado com um novo ponto apenas. Para decidir a qual grupo um registro pertence, um critério global é minimizado. Para minimizar o risco de se inserir um ponto em um grupo errado, múltiplos passos são realizados com os dados (CHOU et al., 2002).

Se todos os dados do registro (entrada) são numéricos, então cada registro pode ser visto como um ponto em um espaço numérico multidimensional, e a distância entre dois pontos é calculada pela distância Euclidiana. Para atributos categóricos, ocorre uma transformação para mapear do espaço original para o espaço numérico. Um registro no novo

espaço terá valor 1 na dimensão que corresponde ao valor do atributo categórico no espaço original e 0 nos outros. Dois pontos no novo espaço têm distância 0 se, e somente se, os pontos no espaço original têm o mesmo valor categórico.

3.5 Trabalhos que utilizam MD para detectar fraudes

Com o advento da Internet e do comércio eletrônico, a segurança das informações/transações nunca antes teve um tratamento tão especial.

A crescente demanda de cartões de crédito também acarretou um aumento significativo do número de uso indevido desse recurso.

A telefonia celular também sentiu o impacto do crescimento do uso de aparelhos e, paralelamente, verificou que as tecnologias de fraudá-los também cresceram na mesma proporção (através da clonagem, principalmente).

São vários os ramos das atividades em que as fraudes acontecem, e a MD está se constituindo em um grande aliado para detectá-las.

A seguir, são relacionados alguns trabalhos que tratam do uso de MD para detectar fraudes.

1) Usando técnicas de mineração de dados na detecção de fraude na área fiscal

O trabalho de Bonchi et al. (1999) mostra como técnicas baseadas em classificação podem ser usadas para comportar a tarefa de planejamento de estratégias de auditoria fiscal.

O comportamento fraudulento pode ser detectado *a priori* ou *a posteriori*. No primeiro caso, encontram-se as fraudes com cartão de crédito, telefone celular, etc. No segundo caso, encontram-se as fraudes baseadas em análises de dados históricos auditados, típicos da área de auditoria fiscal.

Neste trabalho os autores exploraram todo o ciclo KDD e utilizaram o modelo de classificação, através da técnica *árvore de decisão*. Como algoritmo de árvore de decisão, eles utilizaram C5.0.

2) Experiências com fraude na telefonia móvel

O artigo de Hynninen (2000) apresenta algumas técnicas usadas na telefonia celular para detectar fraudes e relaciona métodos para combatê-las.

- *Ligações duplicadas*: a rede consegue observar o mesmo aparelho sendo usado em vários locais ao mesmo tempo.
- *Velocidade das ligações*: a rede detecta, através do *log* de utilização, hora e local da ligação, o uso indevido de celulares.
- *RF (Rádio Freqüência) Fingerprint*: é uma tecnologia militar. A cada ligação o software de rede grava e compara marcas *fingerprint* para todos os telefones gerenciados por esta tecnologia.
- *Usage Profiling*: cada um dos assinantes possui em uma base de dados o seu comportamento registrado. A cada ligação, ou conjunto de ligações, é verificado esse comportamento. Havendo discrepâncias, o cliente é avisado.
- *Contador de Chamadas*: o serviço é desabilitado se as ligações ultrapassarem valores estabelecidos.
- *PIN Codes*: o cliente bloqueia o telefone através de *PIN Codes*, e só o desbloqueia quando efetua ligações.

O autor salienta que com uma variedade cada vez maior de serviços oferecidos pelas companhias telefônicas as possibilidades de fraudes só tendem a aumentar.

Ele alega também que um dos grandes fatores para a ocorrência de fraudes em celulares é o alto custo das ligações internacionais, envolvendo GSM e *roaming*.

Como algoritmo de árvore de decisão, os autores utilizaram C5.0.

3) De acordo com o aprendizado escalável com classe não uniforme e distribuição de custo: um estudo de caso para detecção de fraude com cartão de crédito

Este trabalho se propõe a detectar fraude no uso de cartão de crédito (CHAN e STOLFO, 1998). Baseia-se na criação de subconjuntos de dados com uma apropriada distribuição de classe, aplicando-se algoritmos de aprendizagem de forma independente e paralela nos subconjuntos, e integrando-os para otimizar a performance dos classificadores. Esse método utiliza todos os exemplos de treinamento disponíveis, sem alterar o algoritmo de aprendizagem.

Em seus experimentos, os autores utilizaram quatro algoritmos de aprendizagem: C4.5, CART, RIPPER e BAYES.

4) Detecção de fraude na telefonia móvel usando RNA supervisionada: um primeiro protótipo

Neste trabalho, Moreau e Vandewalle (1997) apresentam uma solução para a detecção de fraude na telefonia móvel usando rede neural supervisionada.

Cada usuário de telefone celular (pós-pago) residente na base de dados possui parâmetros relevantes gravados, denominados *user profile*.

Quando uma ligação é produzida, acontece o seguinte: a *user profile* é recuperada; na seqüência, é atualizada com os dados contidos no bilhete da ligação; após esse procedimento, é passada para o *engine* de detecção de fraudes para que seja verificada a atividade do usuário; se houver suspeita de fraude, um alarme é emitido a quem de direito.

A análise é realizada pela divisão da *user profile* em *short-term (current user profile)* e *long-term (user profile history)*, que nada mais é do que a representação comportamental do usuário. A análise leva em conta a comparação dos dois resultados.

A principal entrada deste modelo é a frequência e duração das ligações.

Uma lista de usuários que produzem alarmes é gravada, juntamente com os bilhetes das ligações suspeitas. Essas informações formam uma base de dados para que possam ser auditadas mais tarde por outros procedimentos.

A arquitetura utilizada nos experimentos foi a *Multilayer Perceptron*.

5) Detecção de fraude e intrusão baseada em agente para sistemas de informação financeira

Com o crescente uso da Internet pelo comércio eletrônico, as fraudes nesse tipo de transação cresceram bastante, principalmente no tocante ao uso de cartões de créditos.

Para tratar esse tema, Stolfo et al. (1997) apresentam uma arquitetura de MD distribuído, desenvolvida na Columbia University, New York, chamada JAM (*Java Agents for Meta-Learning*).

Para preservar o sistema financeiro, eles propõem uma proteção baseada em sistemas de inferência dirigidos ao padrão de uso, utilizando modelos comportamentais de transações anormais ou ambulantes.

Essa proposta requer análises de grandes bases de dados distribuídas, que contenham informações sobre o comportamento das transações, para que sejam produzidos modelos de prováveis transações fraudulentas.

É uma técnica emergente, usada de forma centralizada por AT&T, NYNEX e GTE, para detectar fraudes em telefones celulares.

O projeto JAM possui os seguintes propósitos:

- reduzir o custo das fraudes através de um sistema conveniente de detecção e prevenção de fraudes;
- minimizar as perdas com fraudes pela detecção com maior rapidez e minimizar os custos associados a falsos alarmes;
- melhorar o nível de proteção através da coordenação de serviços.

A tecnologia possui dois principais componentes:

- *Agentes locais para detectar fraudes*: tem como objetivo aprender a detectar fraudes e providenciar serviços de detecção destas, agindo como um componente de instituição corporativa;
- *Sistema seguro e integrado de metadecção*: reúne os conhecimentos adquiridos pelos agentes locais para constituir uma visão das transações da rede, usando essa estratégia para melhor detectar fraudes, capacitando dessa forma o sistema para detecção de ataques distribuídos.

6) Concepção, Desenvolvimento e Análise de um Sistema de Gerência de Segurança para Redes de Telecomunicações

O trabalho de Notare (2000) tem como objetivo detectar fraudes na telefonia móvel (clonagem e inadimplência), praticamente no momento em que elas ocorrem. Quando o número de ligações chega a dez, o sistema inicia o processo de verificação da legitimidade da ligação.

Em seu trabalho, a autora utilizou a rede neural Kohonen para construir agrupamentos.

Os fraudadores tratados são de três categorias, a saber:

- aqueles que alteram o padrão de uso de telefones celulares;
- aqueles que compram um telefone móvel somente por um mês, com a intenção expressa de não pagar pelo serviço;
- aqueles que compram telefones móveis em nome de uma terceira pessoa, também com a intenção de não pagar pelo serviço.

Para treinar e testar a rede, a autora utilizou as seguintes variáveis:

- *Caller_number*: número do telefone do usuário (só para testes);
- *Called_number*: número do telefone chamado. Utilizado para identificar o tipo de chamada (local, interurbana, etc.);
- *Time*: horário da ligação. Utilizado para classificar as diferentes tarifas/horário durante o dia;
- *Duration*: duração da ligação. Utilizado para identificar longas chamadas fora do padrão do usuário;
- *Date*: data da ligação. Utilizado para identificar os diferentes padrões durante os dias da semana, finais de semana e feriados.

A solução proposta por Notare (2000) pode ser assim resumida:

Etapa 1

Nesta primeira etapa, através do uso de redes neurais, ocorre um agrupamento de todos os usuários cadastrados. Foram geradas quatro classes, a classe em que o usuário ficou posicionado é gravada no cadastro de clientes da empresa.

Etapa 2

Para cada conjunto de ligações realizadas para um determinado cliente, é chamado o software (rede neural) que verifica a legitimidade delas. A classe de saída apontada pela rede é confrontada com a classe que o usuário possui, gerada na primeira etapa. Caso tenha ocorrido uma mudança de classe, isso se constitui num indício de fraude. A partir dessa constatação, é disparado um alarme ao usuário (e-mail, *short message*, etc.), informando-o sobre um possível uso indevido do aparelho.

3.6 Considerações finais

Atualmente, uma parte significativa da rentabilidade de muitas empresas é perdida através de práticas fraudulentas. Em empresas de distribuição de energia elétrica tais práticas também são comuns.

Só na Celesc, no ano de 2001, foram quase dois milhões de reais recuperados com as descobertas de fraudes.

Uma das técnicas que está sendo bastante explorada para detecção de fraudes é a mineração de dados. Alguns dos trabalhos que exploram a mineração de dados para a detecção de fraudes são: telecomunicações (HYNNINEN, 2000; MOREAU e VANDEWALLE, 1997; NOTARE, 2000); cartão de crédito (CHAN, 1999; CHAN e STOLFO, 2002); setor fiscal (BONCHI et al., 2002); setor financeiro (STOLFO et al., 1997), além de outros.

A habilidade automatizada de reconhecimento de padrões e exceções da mineração de dados faz dessa tecnologia uma poderosa alternativa para que as empresas possam implementar uma boa política de combate às fraudes, tanto *a priori* (no momento em que elas ocorrem) como também *a posteriori* (efetuando-se consulta a dados históricos).

4 ARQUITETURA DE INFORMAÇÃO PARA GERENCIAMENTO DE PERDAS COMERCIAIS DE ENERGIA ELÉTRICA

4.1 Introdução

Nesta etapa do trabalho será detalhada a arquitetura de informação, proposta, que tem como objetivo o levantamento de possíveis fraudes no consumo de energia elétrica. A arquitetura é composta de três módulos: o primeiro é responsável pela organização dos dados; o segundo, concentra a inteligência da arquitetura; e o terceiro, relaciona interfaces para melhor acompanhar e administrar as perdas comerciais de energia elétrica.

Para o desenvolvimento desta arquitetura, utilizou-se o modelo administrativo atual das Centrais Elétricas de Santa Catarina S.A. (Celesc) como também as suas informações de consumidores, consumo e rede elétrica. A indicação de possíveis fraudadores está concentrada principalmente em características comportamentais de consumo ao longo dos quatro períodos do ano, e do ano anterior, registrados na base histórica da rede.

Como pode ser visto na Figura 4.1, o “kernel IA” tem como suporte a utilização de RNA para conceber a indicação de possíveis fraudadores. Dois tipos de RNA foram trabalhados: a MLP, para classificação de consumidores residenciais e comerciais, e a rede de Kohonen, para a segmentação de consumidores industriais. Neste último tipo também se utilizou segmentação geográfica (SD).

Para a classificação de consumidores residenciais e comerciais, informou-se à rede casos de fraudadores e não fraudadores, na fase de treinamento e teste. Os fraudadores foram extraídos de uma base de histórico de fraudes, e o conjunto dos não fraudadores foi formado a partir dos consumidores recentemente vistoriados.

Relaciona-se também neste capítulo um conjunto de aplicativos que manipulam a saída produzida pelo segundo módulo, fornecendo uma série de funcionalidades que possam contribuir eficazmente na melhoria do combate ao consumo irregular de energia.

4.2 Constituição da arquitetura

A Arquitetura de Informação para o Gerenciamento de Perdas Comerciais de Energia Elétrica é composta basicamente por três módulos, como mostra a Figura 4.1.

O **Módulo 1** é subdividido em dois outros módulos: um voltado para o tratamento das informações que alimentarão diretamente o Kernel IA, e outro que permite uma abordagem de perdas por alimentadores, possibilitando dessa forma a concentração de esforços para tratar alimentadores com os piores índices de perdas.

No **Módulo 2** concentra-se o *engine* responsável pela indicação de possíveis fraudadores e, para tanto, utiliza-se RNA para buscar esses indivíduos. Como pode ser observado na Figura 4.1, a entrada deste módulo pode ser feita tanto pelo processo que busca as informações diretamente nas bases de dados como também pela seleção dos piores alimentadores, através do processo que efetua os cálculos de perdas dos alimentadores.

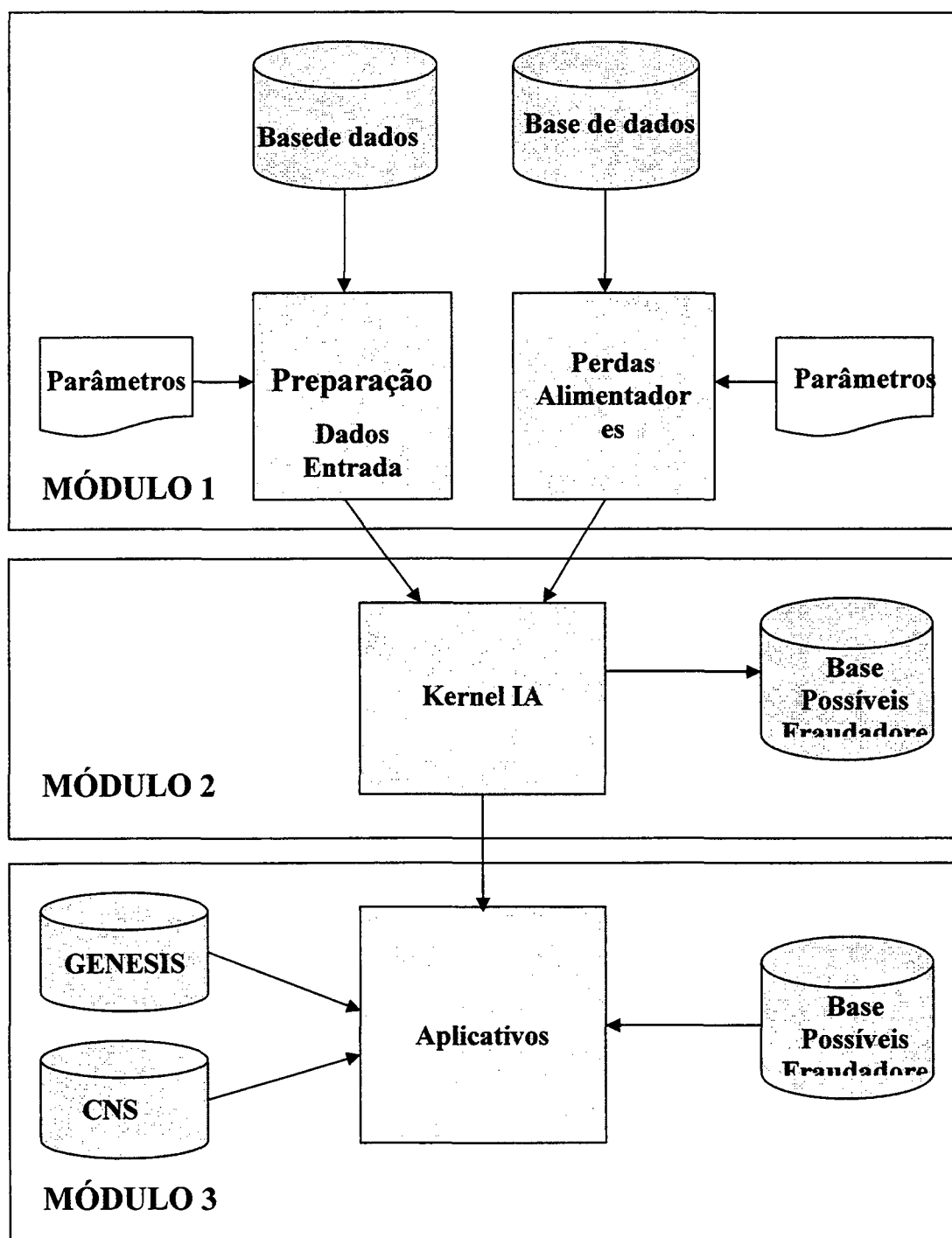
O **Módulo 3** reúne os aplicativos que irão servir à gerência central, gerência regional e área de fiscalização. Em campo, os fiscais poderão fazer uma série de cruzamentos de informações e visualizá-las graficamente para tornar o seu trabalho mais ágil e eficiente. Nas gerências regionais, serão disponibilizados aplicativos que possibilitarão um melhor controle das perdas comerciais e um melhor planejamento e acompanhamento das fiscalizações. Na agência central, poder-se-á acompanhar os resultados obtidos com essa nova sistemática de tratar perdas comerciais como também poderão ser criados aplicativos, à medida que se fizerem necessários.

As bases de dados que estarão envolvidas no processo são oriundas essencialmente do Sistema de Consumidores e do Sistema Genesis. O primeiro sistema é responsável pelo cadastro dos consumidores, arrecadação e faturamento destes. O segundo sistema é o que mantém o cadastro da rede elétrica da empresa através do geoprocessamento.

Atualmente, o Sistema de Consumidores está definido de forma centralizada, sob DB2 (software de banco de dados da IBM) em um *mainframe*. O Sistema Genesis, hoje, funciona de forma descentralizada, sob Oracle (software de banco de dados da Oracle), em máquinas descentralizadas.

Para possibilitar a comunicação entre os bancos Oracle e DB2, existe o *gateway* (*Oracle Transparent Gateway*). Assim, uma vez estando-se conectado ao banco de dados Oracle, é possível acessar, de forma transparente, as informações contidas no banco de dados DB2, no *mainframe*.

Figura 4.1 - Arquitetura de informação proposta



4.3 Módulo 1: Preparação dos dados

Segundo Baragoin et al. (2001), o pré-processamento dos dados é utilizado para identificação, coleta, filtragem e para agregá-los num formato que atenda aos requisitos do modelo de dados e da função de mineração selecionada.

Pyle (1999) destaca que quando os dados são preparados adequadamente, o mesmo acontece com o processo de mineração, pois na utilização dos dados, os melhores e mais rápidos modelos serão produzidos.

Cabe a este módulo preparar adequadamente os dados que serão utilizados pelo Módulo 2 como entrada das redes neurais. Os dados são oriundos do Sistema de Consumidores (Cadastro de Consumidores), do Sistema Genesis (Cadastro da Rede Elétrica) e de uma base de dados de IPTU (prefeitura de Florianópolis).

Neste Módulo também existe a possibilidade de se trabalhar as perdas dos alimentadores. O cálculo das perdas do alimentador produz como saída o valor das perdas global, técnica e comercial. Os alimentadores com os piores índices de perdas poderão ser então selecionados e, a partir daí, todos os consumidores desses alimentadores serão relacionados para que a entrada da rede seja formada de acordo com o item 4.3.2.

Este módulo produz como saída as tabelas CNS_Residencial, CNS_Comercial e CNS_Industrial. Como o próprio nome já sugere, a primeira contém informações de consumidores residenciais, a segunda reúne informações de consumidores comerciais e a terceira contempla consumidores primários, ou industriais.

4.3.1 Estrutura de dados

Existem três tabelas que são utilizadas como entrada das redes neurais: CNS_Residencial, CNS_Comercial e CNS_Industrial. A tabela CNS_Residencial contém informações de consumidores residenciais e seus respectivos consumos; a tabela CNS_Comercial relaciona consumidores comerciais com os respectivos consumos; e a tabela CNS_Industrial reúne um conjunto de consumidores industriais e seus consumos e demandas, basicamente.

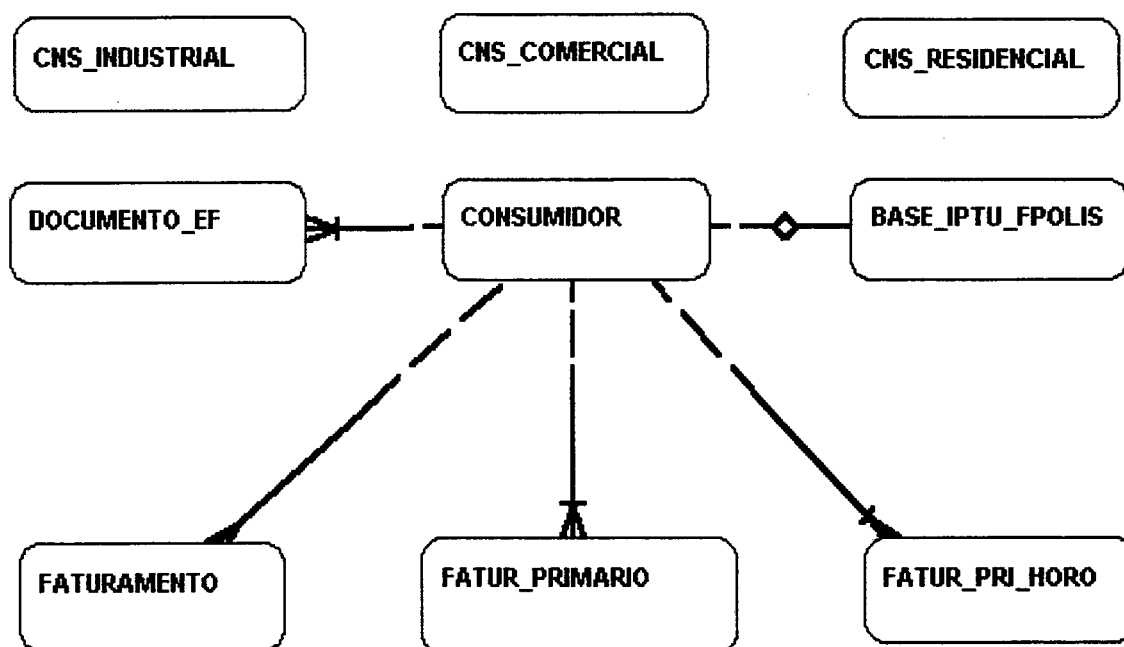
Para formar as tabelas CNS_Residencial e CNS_Comercial, são lidas as tabelas Consumidor e Faturamento, ambas do Sistema de Consumidores. A tabela CNS_Residencial é complementada com dados da tabela Base_Iptu_Fpolis. Esta última contém informações sobre imóveis de Florianópolis. É uma base de dados externa, conseguida junto à prefeitura de Florianópolis.

A tabela CNS_Industrial é formada a partir das leituras das tabelas Consumidor, Faturamento, Fatur_Primary e Fatur_pri_horo. A tabela DOCUMENTO_EF relaciona os fraudadores já conhecidos pela Celesc, servindo, dessa forma, para criar o conjunto de dados

de fraudadores, tanto para a fase de treinamento como também para a fase de testes do Módulo 2.

Na Figura 4.2, é apresentado um modelo E-R da base de dados relacionada ao Módulo 1.

Figura 4.2 - Modelo E-R da base de dados do Módulo 1



Na identificação de possíveis fraudadores residenciais, utilizou-se uma rede *multilayer perceptron*, com o algoritmo *backpropagation*. Os neurônios da camada de entrada dessa rede são formados a partir da tabela *CNS_Residencial*, detalhada no Quadro 4.1 apresentado na seqüência.

Quadro 4.1 - Descrição e tipo de dados da tabela *CNS_Residencial*

continua

NR_CONTA_CNS	NUMBER(10)	Identifica unicamente o consumidor
QT_CONS_PRIMAVERA	NUMBER(10)	Consumo acumulado de outubro, novembro e dezembro
QT_CONS_VERAO	NUMBER(10)	Consumo acumulado de janeiro, fevereiro e março
QT_CONS_OUTONO	NUMBER(10)	Consumo acumulado de abril, maio e junho
QT_CONS_INVERNO	NUMBER(10)	Consumo acumulado de julho, agosto e setembro

conclusão		
QT_CONS_PRIMAVERA_DIF	NUMBER(10)	Diferença entre o consumo (primavera) acumulado de um determinado ano e o ano anterior
QT_CONS_VERAO_DIF	NUMBER(10)	Diferença entre o consumo acumulado (verão) de um determinado ano e o ano anterior
QT_CONS_OUTONO_DIF	NUMBER(10)	Diferença entre o consumo acumulado (outono) de um determinado ano e o ano anterior
QT_CONS_INVERNO_DIF	NUMBER(10)	Diferença entre o consumo acumulado (inverno) de um determinado ano e o ano anterior
ID_FRAUDE	NUMBER(3)	Identificador de fraude (0 = não fraudador; 1 = fraudador)
ID_GPO_TENSAO_CNS	NUMBER(1)	Número de fases no fornecimento de energia
QT_MESES_ULT_INSP	NUMBER(10,2)	Quantidade de meses decorridos desde a última inspeção realizada
ID_MARCA_MEDIDOR	NUMBER(2)	Código da marca do medidor
QT_AREA_CONSTR	NUMBER(5)	Área construída da edificação (em m ²)
TP_EDIFICACAO	NUMBER(2)	Código que identifica o tipo da edificação (madeira, alvenaria, etc.)
PD_CONSTRUCAO	NUMBER(2)	Código que identifica o padrão da construção da edificação (normal, luxo, etc.)

Na identificação de possíveis fraudadores comerciais, utilizou-se uma segunda rede *multilayer perceptron*, com o algoritmo *backpropagation*. Os neurônios da camada de entrada dessa rede são formados a partir da tabela CNS_Comercial, descrita no Quadro 4.2.

Quadro 4.2 - Descrição e tipo de dados da tabela CNS_Comercial

continua		
NR_CONTA_CNS	NUMBER(10)	Identifica unicamente o consumidor na base de dados.
QT_CONS_PRIMAVERA	NUMBER(10)	Consumo acumulado de outubro, novembro e dezembro
QT_CONS_VERAO	NUMBER(10)	Consumo acumulado de janeiro, fevereiro e março
QT_CONS_OUTONO	NUMBER(10)	Consumo acumulado de abril, maio e junho
QT_CONS_INVERNO	NUMBER(10)	Consumo acumulado de julho, agosto e setembro
QT_CONS_PRIMAVERA_DIF	NUMBER(10)	Diferença entre o consumo (primavera) acumulado de um determinado ano e o ano anterior
QT_CONS_VERAO_DIF	NUMBER(10)	Diferença entre o consumo acumulado (verão) de um determinado ano e o ano anterior

conclusão

QT_CONS_OUTONO_DIF	NUMBER(10)	Diferença entre o consumo acumulado (outono) de um determinado ano e o ano anterior
QT_CONS_INVERNO_DIF	NUMBER(10)	Diferença entre o consumo acumulado (inverno) de um determinado ano e o ano anterior
ID_FRAUDE	NUMBER(3)	Identificador de fraude (0 = não fraudador; 1 = fraudador)
ID_GPO_TENSAO_CNS	NUMBER(1)	Número de fases no fornecimento de energia
QT_MESES_ULT_INSP	NUMBER(10,2)	Quantidade de meses decorridos desde a última inspeção realizada
ID_MARCA_MEDIDOR	NUMBER(2)	Código da marca do medidor
ID_RAMO_ATVDD_CNS	NUMBER(3)	Código do ramo de negócio
ID_SUB_RAMO_ATVDD	NUMBER(3)	Código do sub-ramo de negócio
CR_DISJUNTOR_MEDR	NUMBER(5)	Valor da corrente do disjuntor

No levantamento de possíveis fraudadores industriais, trabalhou-se com a tarefa de segmentação, utilizando-se a rede de Kohonen e a segmentação demográfica. Os neurônios da camada de entrada da rede e a entrada para o algoritmo de segmentação foram formados a partir da tabela CNS_Industrial, descrita no Quadro 4.3 a seguir.

Quadro 4.3 - Descrição e tipo de dados da tabela CNS_Industrial

continua

NR_CONTA_CNS	NUMBER(10)	Identifica unicamente o consumidor
ID_FRAUDE	NUMBER(10)	Indicador de fraude
QT_KWH_VER	NUMBER(10)	Consumo acumulado de janeiro, fevereiro e março (verão)
QT_KWH_OUT	NUMBER(10)	Consumo acumulado de abril, maio e junho (outono)
QT_KWH_INV	NUMBER(10)	Consumo acumulado de julho, agosto e setembro (inverno)
QT_KWH_PRI	NUMBER(10)	Consumo acumulado de outubro, novembro e dezembro (primavera)
QT_KW_VER	NUMBER(10)	Demanda acumulada do verão
QT_KW_OUT	NUMBER(10)	Demanda acumulada do outono
QT_KW_INV	NUMBER(10)	Demanda acumulada do inverno
QT_KW_PRI	NUMBER(3)	Demanda acumulada da primavera
QT_FP_VER	NUMBER(1)	Média do fator de potência do verão
QT_FP_OUT	NUMBER(10,2)	Média do fator de potência do outono
QT_FP_INV	NUMBER(2)	Média do fator de potência do inverno
QT_FP_PRI	NUMBER(3)	Média do fator de potência da primavera
QT_FC_VER	NUMBER(3)	Média do fator de carga do verão
QT_FC_OUT	NUMBER(5)	Média do fator de carga do outono
QT_FC_INV	NUMBER(5)	Média do fator de carga do inverno
QT_FC_PRI	NUMBER(5)	Média do fator de carga da primavera
QT_KWH_MAX_VER	NUMBER(5)	Maior consumo do verão
QT_KWH_MAX_OUT	NUMBER(5)	Maior consumo do outono
QT_KWH_MAX_INV	NUMBER(5)	Maior consumo do inverno

conclusão

QT KWH MAX PRI	NUMBER(5)	Maior consumo da primavera
QT KW MAX VER	NUMBER(5)	Maior demanda do verão
QT KW MAX OUT	NUMBER(5)	Maior demanda do outono
QT KW MAX INV	NUMBER(5)	Maior demanda do inverno
QT KW MAX PRI	NUMBER(5)	Maior demanda da primavera

4.3.2 Processamento dos dados

O processamento dos dados para formar a entrada da RNA e SD é bastante simples e prático. Um procedimento conectado a um banco de dados Oracle na central é iniciado, sendo feitas as devidas leituras no banco de dados DB2, através de um *gateway*, e também no banco de dados Oracle.

Uma vez lidos, os dados passam por um processo de verificação de valores nulos e inconsistências (datas não atualizadas, consumos distorcidos, etc.). Outras verificações são realizadas para resolver problemas como, por exemplo, consumidores que não possuem o número de faturamentos esperado.

Já validados, os dados são então processados, dando origem às ocorrências nas tabelas CNS_Residencial, CNS_Comercial e CNS_Industrial.

Neste trabalho realizaram-se experimentos envolvendo consumidores residenciais dos bairros Coqueiros e Abraão, em Florianópolis. A tabela Base_IPTU_Fpolis, como pode ser visto na Figura 4.2, contém informações dos imóveis dos referidos bairros e é utilizada para formar exclusivamente a tabela CNS_Residencial.

O procedimento que trata das perdas nos alimentadores tem processamento próprio, envolvendo acesso às bases de dados do Sistema de Consumidores e Genesis.

4.4 Módulo 1: perdas nos alimentadores

O que será descrito neste trabalho é parte de um estudo que ainda está em andamento para a determinação de perdas nos alimentadores. Trabalhos de campo ainda estão sendo realizados para a definição de uma metodologia de cálculo de perdas. Modelos estão sendo gerados e testados para a verificação da precisão dos resultados; paralelamente, medições estão sendo realizadas em alguns alimentadores de teste, para confrontar o resultado calculado com o resultado medido. Parte da metodologia que será apresentada faz parte do trabalho do CODI (1996).

4.4.1 Natureza

As perdas em sistemas de distribuição de energia elétrica que implicam em extravio de receitas podem ser classificadas em dois tipos:

a) perda de demanda: é a diferença existente entre a demanda de entrada (requerida) e a demanda de saída (vendida), em um dado instante t .

$$P(t) = D(t)_{\text{entrada}} - D(t)_{\text{saída}} \text{ [kW]} \quad (4.1)$$

b) perda de energia: é a diferença existente entre a energia de entrada (requerida) e a energia de saída (vendida) em um intervalo de tempo.

$$e(\Delta t) = E(\Delta t)_{\text{entrada}} - E(\Delta t)_{\text{saída}} \text{ [kWh]} \quad (4.2)$$

Este trabalho tratará especificamente de um método para determinação de perdas de energia.

4.4.2 Localização

As perdas localizam-se nos diversos segmentos de um sistema elétrico, podendo ser classificadas em perdas globais, perdas na transmissão e perdas na distribuição.

As perdas globais são as perdas totais de energia elétrica e demanda existentes, considerando-se o conjunto dos sistemas de geração, transmissão e distribuição; as perdas na transmissão são as perdas de energia elétrica e demanda existentes, considerando-se os sistemas de geração e transmissão; e as perdas na distribuição são as perdas de energia elétrica e demanda existentes, considerando-se apenas o sistema de distribuição.

4.4.3 Componentes

As perdas técnicas na distribuição podem ser classificadas de acordo com o componente elétrico onde ocorrem, tais como condutores da rede primária, transformadores de distribuição, condutores da rede secundária, ramais de ligação, bobina de potencial dos medidores dos consumidores, equipamentos (capacitores, reguladores de tensão, pára-raios, etc.) e diversas (fuga em árvores e isoladores, efeito corona, conexões, etc.).

4.4.4 Fator de carga

Define-se fator de carga (F_C) como a relação entre a demanda média (D_{med}) e a demanda máxima ($D_{máx}$), ao longo de um período de tempo T :

$$F_C = \frac{D_{med}}{D_{máx}} = \frac{\frac{1}{T} \cdot \int_0^T D(t) \cdot dt}{D_{máx}} \quad (4.3)$$

onde:

- F_C - fator de carga
- D_{med} - demanda média [kW]
- $D_{máx}$ - demanda máxima [kW]
- $D(t)$ - demanda no instante t [kW]
- T - período de tempo [h]

Uma vez que a demanda média é inferior à demanda máxima (ou igual, no caso de cargas constantes), o fator de carga varia nos seguintes limites:

$$0 \leq F_C \leq 1 \quad (4.4)$$

A integral de $D(t)$ (demanda instantânea) representa a energia fornecida ao sistema durante o período T :

$$E = \int_0^T D(t) \cdot dt \quad [\text{kWh}] \quad (4.5)$$

A expressão (4.5) pode, portanto, simplificar-se:

$$F_C = \frac{E}{T \cdot D_{máx}} \quad (4.6)$$

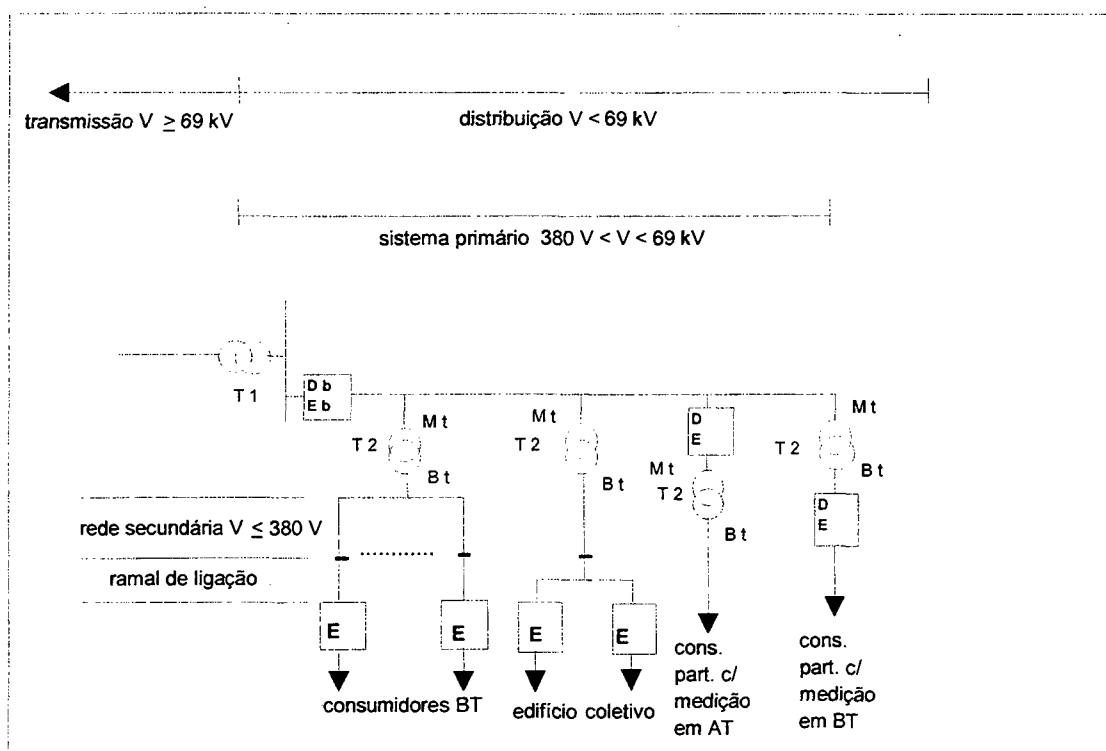
As variáveis E e $D_{máx}$ são grandezas normalmente medidas nas subestações e também nos pontos de fornecimento dos consumidores mais importantes.

4.4.5 Cálculo das perdas técnicas

a) Perda total de energia

Para efeito de estimativa das perdas, o sistema de distribuição é representado de forma simplificada, como na Figura 4.3.

Figura 4.3 - Sistema Elétrico de Distribuição



Fonte: Adaptado de CODI (1996)

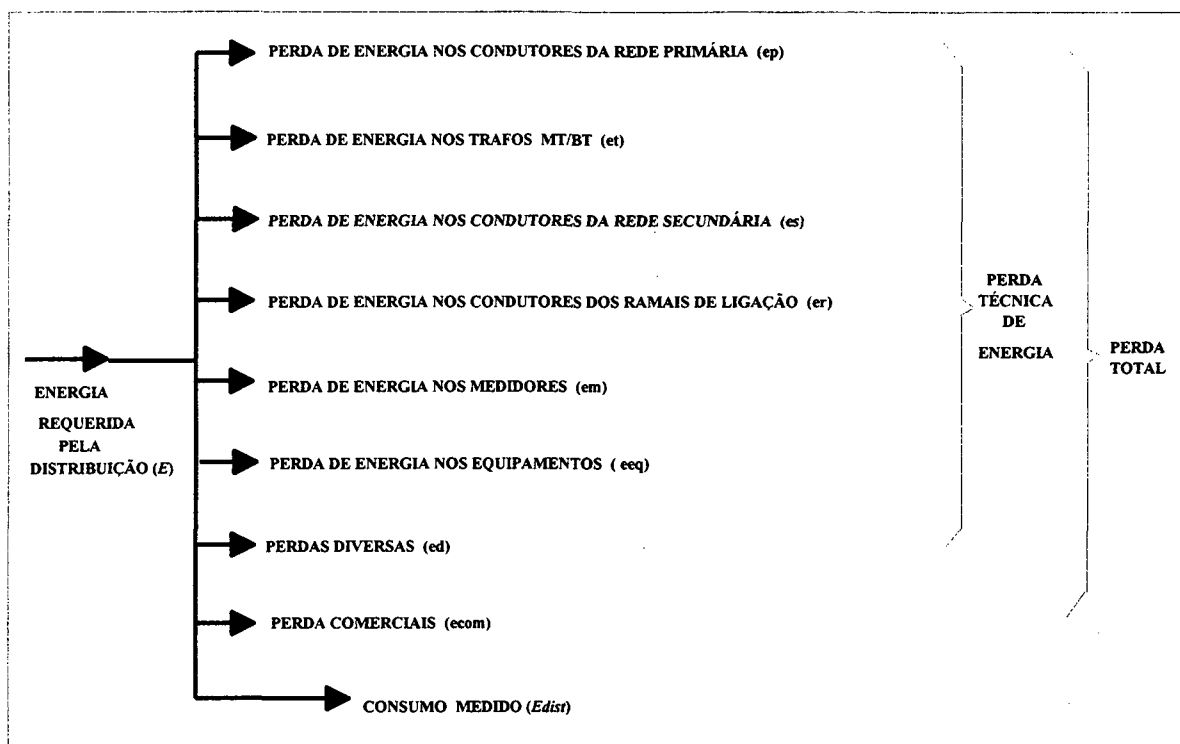
onde:

T_1 - subestação distribuidora (SE de fronteira transmissão/distribuição)

T_2 - transformador de distribuição (MT/BT)

Para esse sistema elétrico, o balanço energético pode ser esquematizado como na Figura 4.4, apresentada na seqüência.

Figura 4.4 - Balanço energético



Do balanço energético, pode-se deduzir as equações abaixo.

$$E = e_{tot} + E_{dist} \quad (4.7)$$

$$e_{tot} = e_{tec} + e_{com} \quad (4.8)$$

$$e_{tec} = e_p + e_t + e_s + e_r + e_m + e_{eq} + e_d \quad (4.9)$$

onde:

E - energia requerida pelo sistema de distribuição [kWh]

e_{tot} - perda total de energia [kWh]

e_{tec} - perda técnica de energia [kWh]

e_{com} - perda comercial de energia [kWh]

E_{dist} - energia requerida (medida) pelos consumidores cadastrados [kWh]

e_p - perda de energia nos condutores da rede primária [kWh]

e_t - perda de energia nos transformadores MT/BT [kWh]

e_s - perda de energia nos condutores da rede secundária [kWh]

e_r - perda de energia nos condutores dos ramais de ligação [kWh]

e_m - perda de energia nos medidores de consumo [kWh]

e_{eq} - perda de energia nos equipamentos instalados no sistema de distribuição [kWh]

e_d - perdas de energia diversas [kWh]

As perdas totais de energia são expressas, portanto, pela equação abaixo (4.10).

$$e_{tot} = e_{tec} + e_{com} = E - E_{dist} \quad (4.10)$$

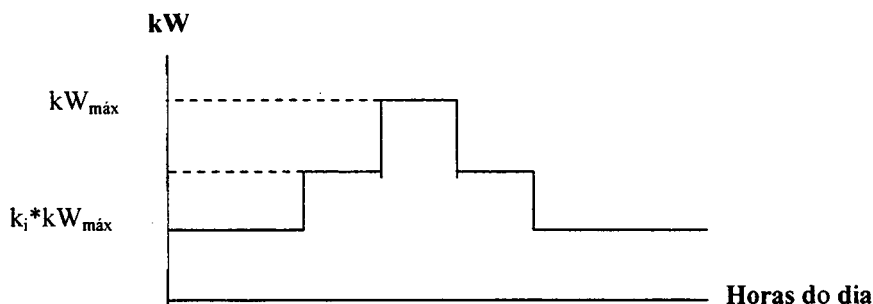
Os disjuntores das linhas ou alimentadores devem ser considerados como ponto de entrega do sistema de transmissão para o sistema de distribuição.

b) Perda de energia nos condutores da rede primária

Para que possam ser determinadas as perdas de energia nos condutores da rede primária necessita-se saber suas especificações, seus comprimentos, a localização dos pontos de derivação, a quantidade de energia por período de tempo circulante em cada trecho e a curva de carga dos consumidores conectados em cada ponto de derivação da rede.

Por questão de simplicidade, as curvas-padrão devem ter patamares, como apresentado na Figura 4.5.

Figura 4.5 - Representação de demanda por patamares



Dado o trecho de circuito primário da Figura 4.6, onde é conhecida a resistência por quilômetro do cabo, seu comprimento, a demanda máxima mensal e a curva de carga característica dos consumidores supridos por essa porção do alimentador, pode-se estabelecer as relações que se seguem.

$$e = R * i^2 * \Delta t \quad (4.11)$$

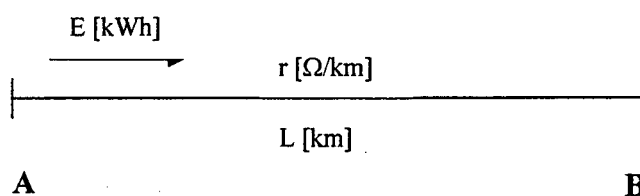
$$i = \frac{P}{\sqrt{3} * V} \quad (4.12)$$

$$i^2 = \frac{P^2}{3 * V^2} \quad (4.13)$$

$$e_{AB} = \sum_{i=1}^n R * \left[\frac{(k_i * D_{m\acute{a}x})^2}{3 * V^2} \right] * \Delta t_i \quad (4.14)$$

$$e_{AB} = \frac{r * L * D_{m\acute{a}x}^2}{3 * V^2} * \sum_{i=1}^n k_i^2 * \Delta t_i \quad (4.15)$$

Figura 4.6 - Representação de um trecho de rede



onde:

e_{AB} - perda de energia no trecho AB [Wh]

r - resistência elétrica por quilômetro dos cabos [Ω /km]

L - comprimento dos cabos no trecho sob análise [km]

$D_{m\acute{a}x}$ - Demanda máxima mensal no trecho [W]

V - tensão primária de linha [V]

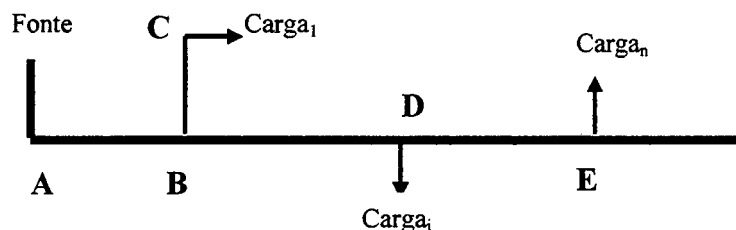
k_i - fator de multiplicação da demanda máxima no intervalo de tempo Δt_i

Δt_i - intervalo de tempo correspondente a um patamar da curva de carga [h]

n - número de patamares da curva de carga característica

Na Figura 4.7, representa-se um circuito primário constituído de m trechos.

Figura 4.7 - Representação de circuito primário



A perda de energia e_p nos condutores primários será então:

$$e_p = \sum_{j=1}^m \left[\frac{r_j * L_j * D_{\max_j}^2}{3 * V^2} * \left(\sum_{i=1}^n k_i^2 * \Delta t_i \right) \right] \quad (4.16)$$

c) Perda de energia nos transformadores de distribuição

A perda de demanda em um transformador é constituída de duas componentes: perda no ferro P_{fe} e perda no cobre (enrolamentos) P_{cu} .

A perda de demanda no ferro é provocada principalmente por correntes de Foucault e histerese. Está relacionada com a corrente de excitação e pode ser considerada constante, uma vez que praticamente depende apenas da tensão de operação do transformador.

A perda de demanda no cobre é provocada pela corrente de carga que atravessa os enrolamentos do transformador.

A Tabela 4.1 apresenta os limites máximos fixados pela Associação Brasileira de Normas Técnicas (ABNT) das perdas no ferro e no cobre (para tensão e corrente nominais), em transformadores de diversas potências.

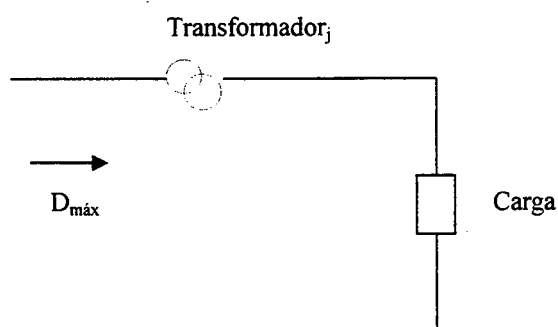
Tabela 4.1 - Perdas no ferro e no cobre de transformadores

POTÊNCIA [kVA]	13,8 KV		34,5 kV	
	P _{fe} [kW]	P _{cu} [kW]	P _{fe} [kW]	P _{cu} [kW]
15	0,120	0,340	0,130	0,390
30	0,200	0,570	0,215	0,645
45	0,260	0,780	0,290	0,870
75	0,390	1,140	0,425	1,275
112,5	0,520	1,550	0,575	1,725
150	0,640	1,910	0,715	2,145
225	0,900	2,700	0,970	2,910
300	1,120	3,360	1,200	3,600

Fonte: Celesc

Na Figura 4.8 representa-se, simplificada, um transformador atendendo a uma carga que lhe exige uma certa demanda máxima $D_{m\acute{a}x}$.

Figura 4.8 - Representação da alimentação de uma carga



As perdas de demanda e energia que ocorrem em cada transformador podem ser determinadas por:

$$p_t = p_{fe} + p_{cu} \quad (4.17)$$

$$e_t = e_{fe} + e_{cu} \quad (4.18)$$

$$e_{fe} = P_{fe} * \Delta t \quad (4.19)$$

$$e_{cu} = p_{cu \text{ nom}} * \frac{D_{m\acute{a}x}^2}{P_{nom}^2} * \left(\sum_{i=1}^n k_i^2 * \Delta t \right) \quad (4.20)$$

onde:

- p_t - perda de demanda no transformador [kW]
 p_{fe} - perda de demanda no ferro do transformador [kW]
 p_{cu} - perda de demanda no cobre do transformador [kW]
 e_t - perda de energia no transformador [kWh]
 Δt - período de análise [h]
 e_{fe} - perda de energia no ferro do transformador [kWh]
 e_{cu} - perda de energia no cobre do transformador [kWh]
 P_{cunom} - perda no cobre do transformador, a potência nominal [kW]
 $D_{máx}$ - Demanda máxima no transformador [kW]
 P_{nom} - Potência nominal do transformador [KVA]

A perda de energia em m transformadores pode ser calculada pela fórmula (4.20):

$$e_{cu} = \sum_{j=1}^m \left[P_{cunomj} * \frac{D_{máxj}^2}{P_{nomj}^2} * \left(\sum_{i=1}^n k_i^2 * \Delta t \right) \right] \quad (4.21)$$

A demanda máxima do transformador é função do tipo predominante de consumidores ligado a ele e do consumo total dos consumidores. Logo,

$$D_{MAX} = f(E_t) \quad (4.22)$$

Por exemplo, para transformadores com predominância de consumidores residenciais, segundo estudos feitos pela Celesc (DPSD/GTGEO), a demanda máxima pode ser calculada pela fórmula (4.23).

$$D_{MAX} = \left(\left(\sum_{i=1}^n E_i \right) * 0,0039 \right) + 3,9 \text{ (KW)} \quad (4.23)$$

onde:

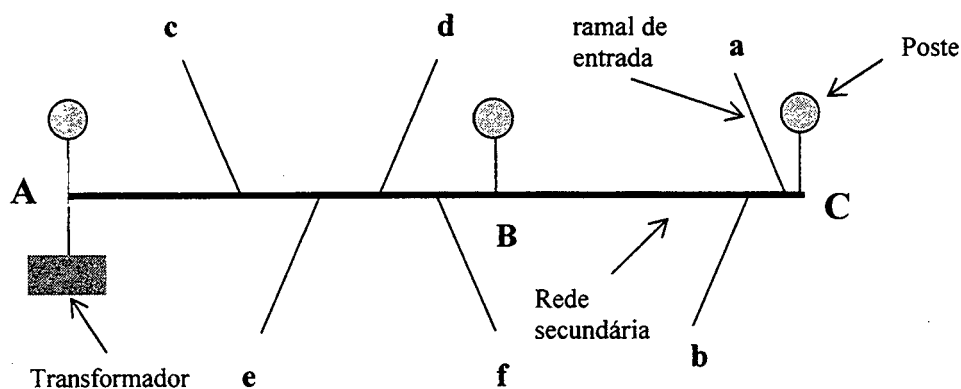
n – número de consumidores

E – energia consumida por cada um dos consumidores do transformador (kWh)

d) Perda de energia nos condutores da rede secundária

No trecho *BC* da rede secundária da Figura 4.9, flui a energia gasta pelos consumidores *a* e *b*. A demanda máxima neste trecho é dada pela equação 4.22. No trecho *AB*, além da energia que serve *a* e *b*, flui a energia consumida por *c*, *d*, *e* e *f*. Novamente, usando-se a equação (4.22), determina-se a demanda máxima neste trecho.

Figura 4.9 - Trecho de uma rede secundária



A perda de energia e_s nos condutores da rede secundária é dada pela equação (4.24):

$$e_s = \sum_{j=1}^m \left(r_j * L_j * \left(\sum_{i=1}^n \left[\frac{(k_i * D_{m\acute{a}x})^2}{3 * V^2} * \Delta t \right] \right) \right) \quad (4.24)$$

onde:

- e_s - perda de energia no circuito secundário associado a determinado transformador [Wh]
- r_j - resistência elétrica por quilômetro do cabo do trecho *j* [Ω/km]
- L_j - comprimento dos cabos no trecho *j* [km]
- $D_{m\acute{a}x}$ - demanda máxima mensal [W]
- V - tensão secundária de linha [V]
- k_i - fator de multiplicação da demanda máxima no intervalo de tempo Δt_i
- Δt_i - intervalo de tempo correspondente a um patamar da curva de carga [h]

- n - número de patamares da curva de carga característica
 m - número de trechos secundários

e) Perda de energia nos condutores dos ramais de ligação

O método, nesse caso, consiste do uso da equação 4.15. Conhecendo-se o comprimento médio do ramal típico e a bitola média dos condutores, pode-se obter a perda de energia nos m ramais de ligação de consumidores secundários de um alimentador da seguinte forma:

$$e_r = m * r * L * \sum_{i=1}^n \left(\frac{k_i^2 * D_{máx}^2 * \Delta t_i}{3 * V^2} \right) \quad (4.25)$$

onde:

- e_r - perda de energia nos ramais de ligação de consumidores secundários de determinado alimentador [Wh]
 r - resistência elétrica por quilômetro dos condutores típicos do ramal de ligação [Ω /km]
 L - comprimento típico do ramal de ligação [km]
 $D_{máx}$ - Demanda máxima mensal média dos consumidores secundários [W]
 V - tensão primária de linha [V]
 k_i - fator de multiplicação da demanda máxima no intervalo de tempo Δt_i
 Δt_i - intervalo de tempo correspondente a um patamar da curva de carga [h]
 n - número de patamares da curva de carga característica
 m - número de consumidores secundários do alimentador

f) Perda de energia nos medidores

A perda de demanda verificada nos medidores de energia ativa pode ser estimada a partir do conhecimento do número de medidores monofásicos, bifásicos e trifásicos instalados no sistema de distribuição.

Os medidores monofásicos, bifásicos e trifásicos apresentam um, dois e três pares de bobinas Bc/Bp (bobina de corrente / bobina de potencial), respectivamente. Enquanto a perda

na bobina de corrente é contabilizada pelo medidor e, conseqüentemente, assumida pelo cliente, a perda que ocorre na bobina de potencial é de responsabilidade da concessionária.

Uma bobina de potencial apresenta uma perda média de 1,2 W e, portanto, as perdas totais de demanda e energia nos medidores de energia ativa são calculadas pelas expressões:

$$p_m = \frac{1,2 \cdot N_m \cdot (j_1 + 2 \cdot j_2 + 3 \cdot j_3)}{1000} \quad (4.26)$$

$$e_m = p_m \cdot \Delta t \quad (4.27)$$

onde

p_m - perda de demanda nos medidores de energia do alimentador [kW]

e_m - perda de energia nos medidores de energia do alimentador [kWh]

N_m - número de medidores instalados no alimentador

j_1, j_2, j_3 - incidência de medidores monofásicos, bifásicos e trifásicos, respectivamente

Δt - período de análise [h]

g) Perda de energia nos equipamentos

Conhecendo-se o número de equipamentos instalados no alimentador (reguladores de tensão, capacitores, medidores de energia reativa, medidores de demanda, etc.), estimam-se as perdas de demanda e energia que ocorrem nesses equipamentos através das equações:

$$p_{eq} = \sum_{i=1}^n N_{eq_i} \cdot P_{eq_i} \quad (4.28)$$

$$e_{eq} = p_{eq} \cdot \Delta t \quad (4.29)$$

onde:

p_{eq} - perda de demanda nos equipamentos do alimentador [kW]

N_{eq_i} - número de equipamentos tipo i (regulador, capacitor, medidor de energia reativa, medidor de demanda, etc.)

- p_{eqi} - perda de demanda unitária no equipamento tipo i
[kW]
- n - número de diferentes tipos de equipamentos (ex., se há apenas medidores de energia reativa e medidores de demanda, $n = 2$)
- e_{eq} - perda de energia nos equipamentos do alimentador
[kWh]
- Δt - período de tempo da análise [h]

h) Perdas diversas

Outras parcelas de perdas técnicas ocorrem nos sistemas de distribuição devido às correntes de fuga em árvores ou isoladores, por efeito corona, em conexões, etc.

Algumas dessas perdas podem apresentar valores desprezíveis, e outras exigem processos de cálculo extremamente elaborados, além de pesquisas de campo e em laboratório.

Dadas a dificuldade de avaliação e a inviabilidade de investigação, estima-se que o montante dessas perdas corresponda a um percentual das perdas técnicas calculadas anteriormente, conforme a expressão:

$$e_d = K \cdot (e_p + e_t + e_s + e_m + e_{eq}) \quad (4.30)$$

onde:

- K - valor a ser definido pela concessionária em função de suas características e particularidades - recomenda-se $K \leq 0,1$
- e_d - perdas de energia diversas [kWh]
- e_p - perda de energia nos condutores da rede primária [kWh]
- e_t - perda de energia nos transformadores [kWh]
- e_s - perda de energia nos condutores da rede secundária [kWh]
- e_m - perda de energia nos medidores de energia [kWh]
- e_{eq} - perda de energia nos equipamentos instalados no sistema de distribuição [kWh]

Vale ainda lembrar que nos sistemas de distribuição ocorre perda nos reatores do sistema de iluminação pública que não deve ser quantificada, uma vez que o seu faturamento deve contemplar, além da potência da lâmpada, a perda que ocorre no seu reator.

i) Cálculo das perdas comerciais

Consideram-se perdas comerciais aquelas que decorrem de erro de medição (ou não medição) de potência e consumo de energia, e são consequência da existência de consumidores clandestinos, consumidores deixados sem medição, desvio de energia, medidores descalibrados, erro de leitura, erro de constante do medidor, erro de relações de TC e TP, entre outros.

Conhecendo-se a perda total de energia e_{tot} e a perda técnica total de energia e_{tec} , estima-se a perda comercial de energia, de acordo com a equação (4.31).

$$e_{com} = e_{tot} - e_{tec} = E - E_{DIST} - (e_p + e_t + e_s + e_r + e_m + e_{eq} + e_d) \quad (4.31)$$

4.5 Módulo 2: Kernel IA

4.5.1 Introdução

O Kernel IA, como o próprio nome já diz, é implementado com base em inteligência artificial, mais precisamente RNA.

Deseja-se com o seu uso classificar consumidores residenciais e comerciais em fraudadores e não fraudadores. Uma outra função é a de segmentar consumidores industriais, procurando-se isolar fraudadores em segmentos próprios (BORAGOIN, 2001).

A utilização de RNA para detecção de fraudes é um mecanismo que está sendo explorado em vários trabalhos. Destaca-se o trabalho de Moreau e Vandennalle (1997), que utilizam uma RNA MLP para detectar fraudes no uso de telefones celulares. Já Notare (2000) utilizou Kohonen para implementar um sistema para a identificação de clonagem de celulares.

Para se trabalhar consumidores residenciais e comerciais, optou-se pela tarefa de classificação. Uma das razões para tal foi o número razoável de casos registrados de fraudes nesses grupos de usuários.

Para consumidores industriais, os estudos e experimentos foram dirigidos para a tarefa de segmentação, pois os casos de fraudes registrados nesse segmento são desprezíveis, inviabilizando o processo de RNA supervisionada.

No Kernel IA, trabalha-se com RNA para classificar consumidores secundários (grupo “B”) e SD para segmentar consumidores primários (grupo “A”). A RNA e a SD possuem

entradas distintas, definidas pelas tabelas *CNS_Residencial*, *CNS_Comercial* e *CNS_Primary*, como pode ser visto na Figura 4.2.

Os experimentos com RNA e SD foram realizados com a ferramenta *Intelligent Mining*, da IBM.

4.5.2 Dados de entrada da rede

Os dados de entrada da RNA e SD estão detalhados nos quadros 4.1, 4.2 e 4.3.

Os tipos de dados trabalhados foram os seguintes: categórico, binário, discreto e contínuo. Esses tipos de dados são definidos quando se mapeiam as tabelas relacionais para a ferramenta *Intelligent Mining*.

No processo de normalização dos dados de entrada da RNA e SD, realizado pela própria ferramenta, as variáveis numéricas contínuas e discretas são inseridas no intervalo de 0.0 a 1.0, através do algoritmo *piece-wise* linear. O valor médio é convertido para 0.5; o dado de entrada com valor até a média é convertido entre 0.0 e 0.5; o dado de entrada com valor acima da média é convertido entre 0.5 e 1.0. Esse algoritmo possibilita a distribuição de variáveis contínuas quando elas possuem uma distribuição não normal.

As variáveis categóricas são convertidas em vetores. Uma variável categórica com um grande número de valores discretos pode causar uma expansão no número de unidades de entrada da rede, podendo afetar o tempo de treinamento da rede. Para resolver esse problema, a ferramenta disponibiliza funções de pré-processamento para reduzir o número de valores discretos: o atributo pode ser convertido para valores discretos numéricos ou pode-se criar um novo atributo na própria ferramenta, com funções específicas, mapeando-se o atributo da tabela.

Todos os dados manipulados pela RNA e SD são residentes em banco de dados Oracle, de forma centralizada. Não existem trocas de arquivos nesse processo, estando tudo integrado no banco de dados.

4.5.3 Classificação de consumidores secundários

A tarefa de classificação é utilizada para buscar possíveis fraudadores entre consumidores residenciais e comerciais. É uma tarefa supervisionada, composta de três fases distintas: treinamento, teste e aplicação.

O treinamento é utilizado para se construir um modelo inicial. Uma vez construído o modelo, a fase de teste é utilizada para que esse modelo seja refinado. Finalmente, a rede passa a ser utilizada na função para a qual foi destinada.

Uma das formas de se separar os dados de treinamento e teste da rede é através da divisão randômica dos dados (BARAGOIN et al., 2001). É importante ter aproximadamente o mesmo número de exemplos (fraudador e não fraudador) em cada uma das entradas.

A RNA utilizada para realizar a classificação foi a *Multilayer Perceptron*, com o algoritmo de treinamento *Backpropagation*.

O processo de classificação é constituído das seguintes etapas:

- a) através das tabelas DOCUMENTO_EF, CONSUMIDOR e FATURAMENTO, todas de um banco de dados DB2, acessam-se os exemplos de fraudadores e as suas informações, de acordo com as tabelas CNS_RESIDENCIAL e CNS_Comercial. Para essa entrada, atribui-se o valor “1” ao atributo ID_FRAUDE;
- b) para selecionar exemplos de não fraudadores, acessam-se as tabelas DOCUMENTO_EF, CONSUMIDOR e FATURAMENTO, relacionando-se os consumidores que foram recentemente fiscalizados e que não constam na tabela DOCUMENTO_EF. A essa entrada, atribui-se o valor “0” para o atributo ID_FRAUDE. Para consumidores residenciais, complementam-se as informações com dados da tabela BASE_IPTU_FPOLIS;
- c) montam-se dois conjuntos de dados, tanto para a tabela CNS_Residencial como também para a tabela CNS_Comercial, para a etapa de treinamento e teste, tendo-se a preocupação de manter em cada conjunto a representatividade das amostras;
- d) vários experimentos são realizados na fase de treinamento, trabalhando-se com dados e parâmetros da rede para se chegar a um nível aceitável de acertos de fraudadores e não fraudadores. A maior preocupação é obter o maior índice possível de acertos de não fraudadores, pois destes se tem certeza de seu comportamento;
- e) feitos os refinamentos adequados da rede para a fase de treinamento, parte-se para a fase de teste. Caso o resultado obtido nesta fase não atenda à expectativa, retorna-se ao item “d”, para tentar melhorar um pouco mais o nível de acertos,

retornando-se em seguida a este item. Fica-se neste ciclo até se chegar a um nível aceitável de acertos na fase de teste;

- f) novas entradas de consumidores podem então ser formatadas para serem avaliadas pela rede. Esta fase é conhecida como fase de aplicação. Para que se possa verificar a autenticidade do prognóstico da rede, é necessário ir a campo.

4.5.4 Segmentação de consumidores primários

Para buscar fraudes entre consumidores industriais, optou-se pela tarefa de segmentação ou agrupamento. Uma das razões é o número pequeno de casos de fraudes registrados.

A segmentação agrupa consumidores que têm características similares e, ao mesmo tempo, maximiza a diferença entre grupos distintos de consumidores.

A idéia aqui é deixar a SD construir o maior número possível de grupos, para tentar encontrar comportamentos anormais em pequenos grupos (BARAGOIN et al., 2001).

Para cada saída, a *Intelligent Mining* fornece os seguintes campos: *cluster id1*, *score1*, *cluster id2*, *score2* e *confidence*. O *cluster id1* identifica o grupo mais ajustado à entrada; *score1* estabelece a qualidade do ajustamento da entrada em relação ao *cluster1*; *cluster id2* identifica o segundo grupo mais apropriado para a entrada; *score2* estabelece a qualidade do ajustamento da entrada em relação ao *cluster2*; *confidence* é a taxa de confiança do assinalamento da entrada, em valor percentual.

Enquanto a classificação vista no item 4.5.3 dá como resposta os valores “0” para não fraudador e “1” para fraudador, com uma determinada taxa de confiança na segmentação a interpretação do resultado é bem mais complexa.

Essa interpretação pode ser feita pela análise gráfica de cada um dos grupos criados, avaliando-se uma série de informações de cada grupo individualmente, do grupo em relação aos outros e também de cada um dos atributos dentro do grupo. Uma outra maneira é a avaliação da taxa de confiança, na qual um consumidor individual tem sido colocado em um particular grupo e o grupo alternativo que lhe foi indicado para estar associado.

Uma alternativa seria a avaliação através da mudança de grupo, considerando-se o mesmo período do ano anterior, podendo-se levar em consideração os campos *cluster id1*, *cluster id2* e *confidence* do ano anterior e do atual, caso ocorra a mudança.

Na segmentação de consumidores industriais, testaram-se as redes de Kohonen e SD. Com Kohonen não houve sucesso na tentativa de isolar consumidores industriais fraudadores em pequenos grupos. Em vários experimentos realizados, o valor absoluto dos grupos permaneceu alto e estável. Com o uso de SD, houve progresso pois já se conseguiu gerar uma quantidade de grupos maior e de pequeno valor absoluto.

As tabelas acessadas para formar a entrada do processo de segmentação foram DOCUMENTO_EF, CONSUMIDOR, FATURAMENTO, FATUR_PRIMARIO e FATUR_PRI_HORO, todas oriundas do banco de dados DB2, no *mainframe*.

4.5.5 Dados de saída da rede

Na tarefa de classificação, a rede MLP produz como saída os valores “0” para não fraudadores e “1” para fraudadores, e um valor de taxa de confiança para essa indicação.

A tarefa de classificação de consumidores residenciais e comerciais produz uma saída que é gravada diretamente no banco de dados, de acordo com a estrutura mostrada no Quadro 4.4.

Quadro 4.4 - Descrição e tipo de dados da tabela CNS_Classificacao

NR_CONTA_CNS	NUMBER(10)	Identifica unicamente o consumidor
ID_FRAUDE	NUMBER(1)	Indicador de fraude
TX_CONFIANCA	NUMBER(5,2)	Valor entre 0 e 1, que mede o grau de confiança do prognóstico

A saída da tarefa de segmentação de consumidores primários também é gravada em uma tabela Oracle, cuja estrutura é apresentada no Quadro 4.5.

Quadro 4.5 - Descrição e tipo de dados da tabela CNS_Segmentacao

NR_CONTA_CNS	NUMBER(10)	Identifica unicamente o consumidor
ID_FRAUDE	NUMBER(1)	Indicador de fraude
ID_CLUSTER1	NUMBER(2)	Identificador do grupo primário
VL_SCORE1	NUMBER(5,4)	Valor que mede a afinidade do registro com o grupo primário
ID_CLUSTER2	NUMBER(2)	Identificador do grupo secundário
VL_SCORE2	NUMBER(5,4)	Valor que mede a afinidade do registro com o grupo secundário
TX_CONFIANCA	NUMBER(5,2)	Valor entre 0 e 1, que mede o grau de confiança do prognóstico

4.6 Módulo 3: Aplicativos

4.6.1 Introdução

Este módulo tem como objetivo prover um conjunto de aplicativos para atender adequadamente às gerências regionais e à gerência central, como também à área de fiscalização das agências, permitindo, dessa forma, um melhor gerenciamento das perdas comerciais de energia elétrica.

O que foi inicialmente projetado e implementado foram algumas funcionalidades. Várias outras poderão ser exploradas, criando-se assim um ambiente mais consistente e flexível de Gerência de Perdas Comerciais.

Uma das preocupações com relação a esses aplicativos foi a de desenvolvê-los em ambiente Web, evitando-se assim a dependência de hardware e sistema operacional.

Para a área de fiscalização, houve a preocupação de mobilidade, optando-se então pela arquitetura *wireless*.

A base de dados da arquitetura proposta residirá em um banco de dados Oracle, na administração central, acessível de qualquer agência regional.

A solução *wireless* será implementada inicialmente de forma *off-line*, através de processo de sincronização com a base de dados do sistema Genesis. Quando houver o serviço de conexão *on-line*, através de equipamentos *wireless*, os aplicativos já estarão preparados para trabalhar nessa modalidade.

4.6.2 Aplicativos para a fiscalização

Uma vez gerada a base de possíveis fraudadores, a equipe de fiscalização poderá ir a campo e ter acesso a uma gama bastante grande de informações através de um *palmtop*, por exemplo.

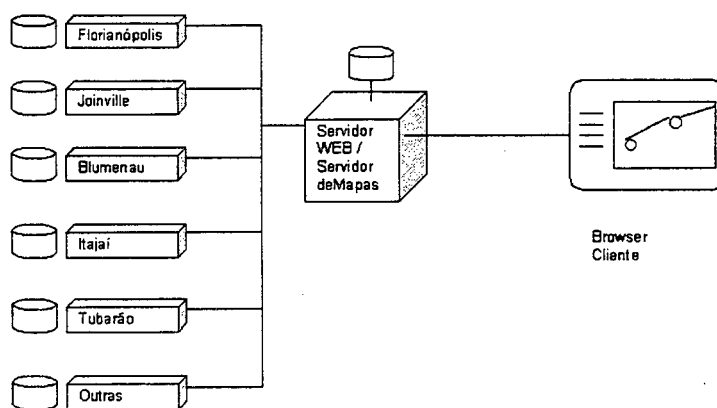
Como os aplicativos fazem uso de geoprocessamento, as alternativas de uso são bastante amplas. Assim, as funcionalidades para a fiscalização seriam:

- a) relacionar consumidores por categoria, num determinado raio a partir do ponto onde está localizado o fiscal;

- b) indicar a localização de consumidores de um determinado ramo de negócio, informando a distância do ponto de onde se encontra, para uma determinada taxa de confiança;
- c) informar dados cadastrais e de consumo, para um determinado consumidor;
- d) informar graficamente a curva de consumo para um determinado consumidor;
- e) informar dados da rede, para a unidade de medição que está sendo fiscalizada;
- f) selecionar uma área (bairro, rua, município) e solicitar os consumidores de uma determinada categoria, para valores de taxa de confiança;
- g) outras.

A configuração básica para prover esses serviços está apresentada na Figura 4.10. É uma arquitetura de três camadas, formada pelo banco de dados Oracle, pelo servidor de aplicação (IIS/MapGuide) e pelo *browser*, de onde as aplicações Web são invocadas.

Figura 4.10 - Arquitetura básica para suportar aplicativos da área de fiscalização



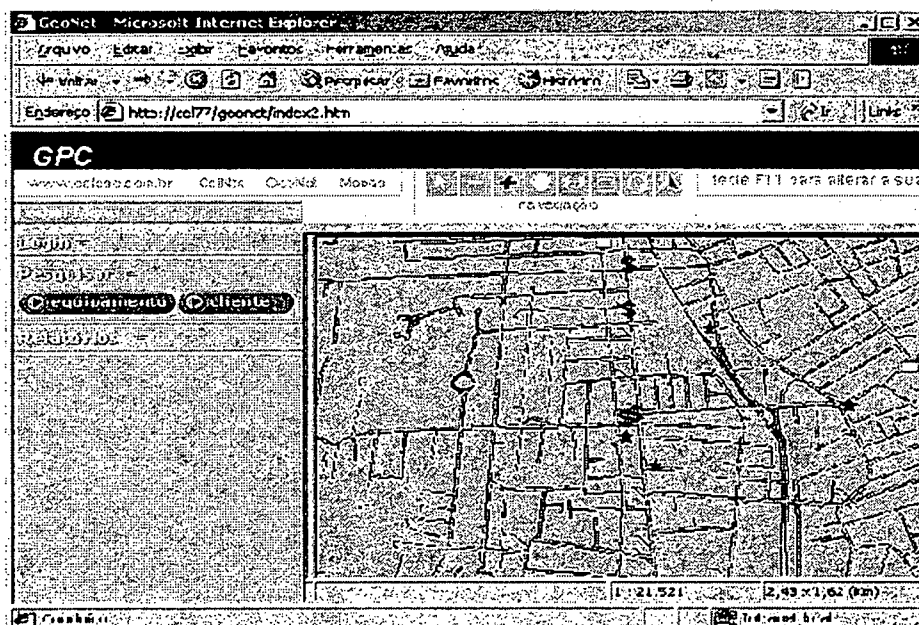
Cada uma das camadas relacionadas na Figura 4.10 é composta ou exige os seguintes componentes (imaginando-se um PC ou Laptop):

- no lado cliente, usa-se o Internet Explorer versão 5.0 ou superior, com ActiveX do MapGuide, da Autodesk;
- na camada intermediária reside a lógica da aplicação e o servidor de mapas vetoriais, Autodesk MapGuide Server, funcionando em Windows 2000 Server com IIS (Internet Information Server);

- na camada servidora são acessados os diversos bancos de dados Oracle em cada uma das agências. Para esse acesso, utiliza-se ASP (*Active Server Pages*).

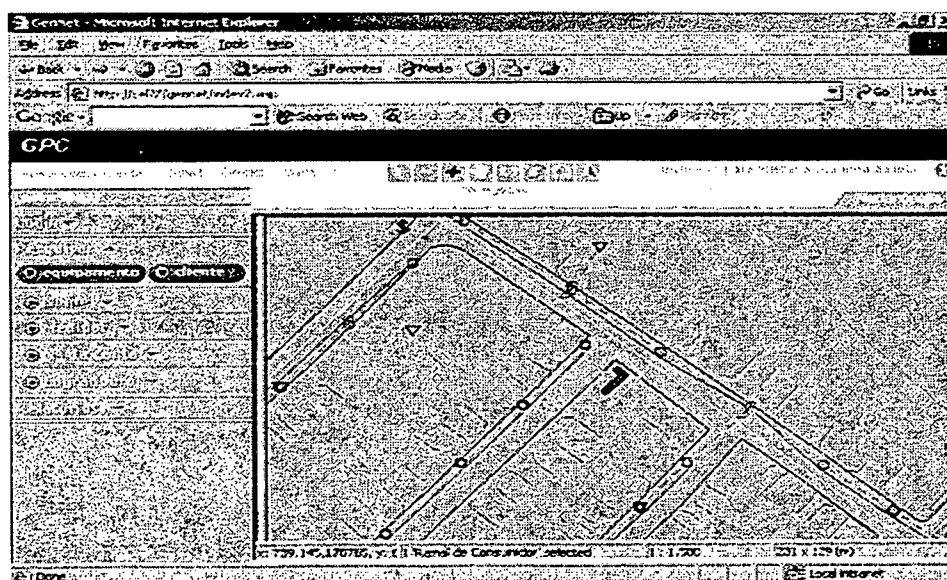
Na Figura 4.11, é mostrada uma das funcionalidades do aplicativo, sendo relacionados graficamente alguns possíveis fraudadores, com as suas localizações geográficas.

Figura 4.11 - Relação de possíveis fraudadores localizados geograficamente



A Figura 4.12 representa um *zoom* de um possível fraudador, apresentado na Figura 4.11. Dessa forma, tem-se a localização do mesmo em termos de rua, bairro, município, transformador, poste, fase, etc.

Figura 4.12 - Zoom de um possível fraudador

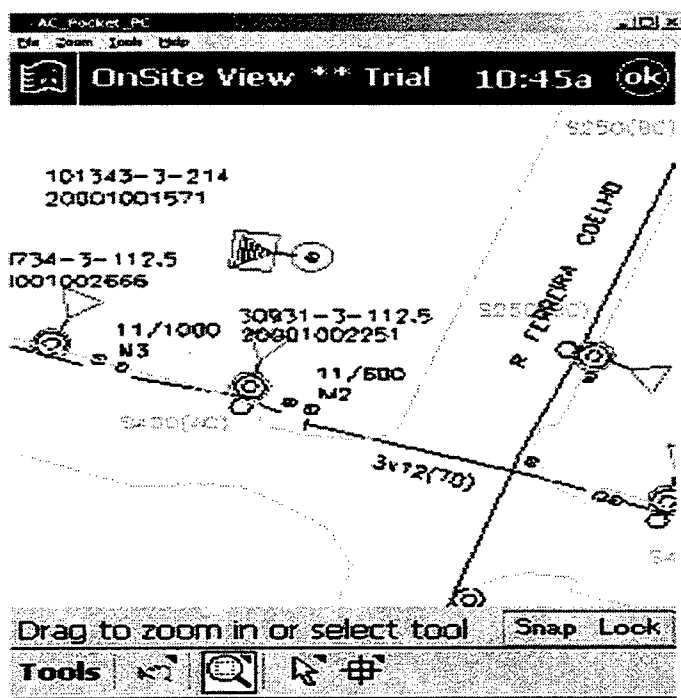


Para os aplicativos destinados a equipamentos PDA (*Personal Digital Assistant*), como iPAQ da Compaq, Jornada da HP ou Casiopea da Casio, rodando Windows CE, outros recursos são necessários:

- a) ActiveSync, da Microsoft, para sincronização de arquivos;
- b) Autodesk Onsite Enterprise, instalado na camada de aplicação, junto ao servidor MapGuide;
- c) Autodesk Onsite View, instalado no PDA.

Na Figura 4.13 tem-se um exemplo da visualização de um trecho da rede elétrica, exibida em um PDA.

Figura 4.13 - Tela do aplicativo em um PDA



4.6.3 Aplicativos para a gerência regional

Para as gerências regionais montou-se um conjunto inicial de soluções, para atendê-las em algumas de suas necessidades.

Para esses aplicativos, a arquitetura de três camadas também foi usada:

- no lado do cliente, o *browser* Internet Explorer 5.5 ou superior;
- na camada intermediária, o servidor de aplicação da Oracle, o IAS9i, rodando Developer6i Server, também da Oracle.
- como banco de dados, o Oracle e o DB2.

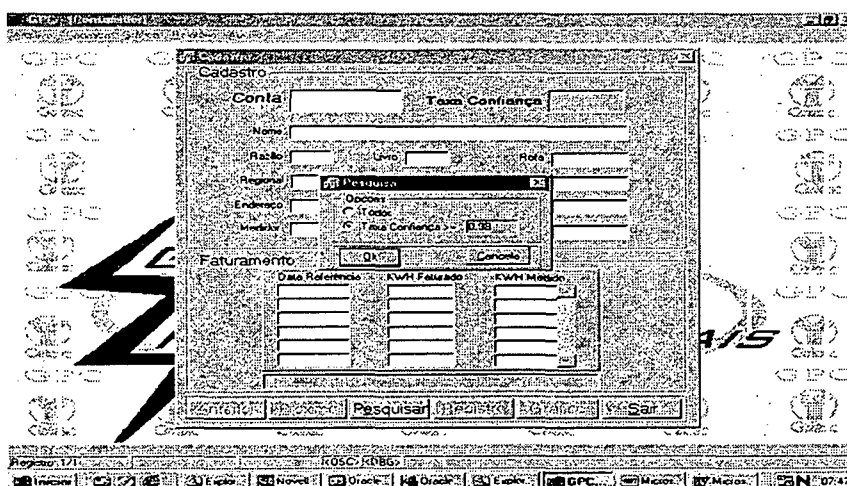
Os aplicativos foram construídos utilizando-se as ferramentas *Oracle Forms6i*, *Oracle Reports6i* e *Oracle Graphics6i*. A execução se dá através do *browser*. Na Figura 4.14, é mostrada a tela inicial do aplicativo.

Uma das funcionalidades desse aplicativo é relacionar um conjunto de possíveis fraudadores de uma determinada categoria de consumidores, para uma taxa de confiança desejada, como pode ser observado na Figura 4.15.

Figura 4.14 - Tela inicial do aplicativo de gerência regional

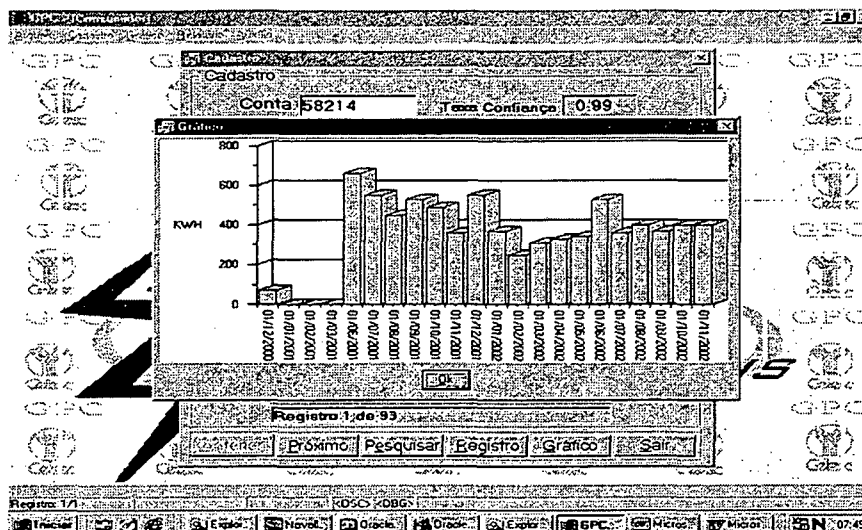


Figura 4.15 - Seleção de possíveis fraudadores para uma determinada taxa de confiança



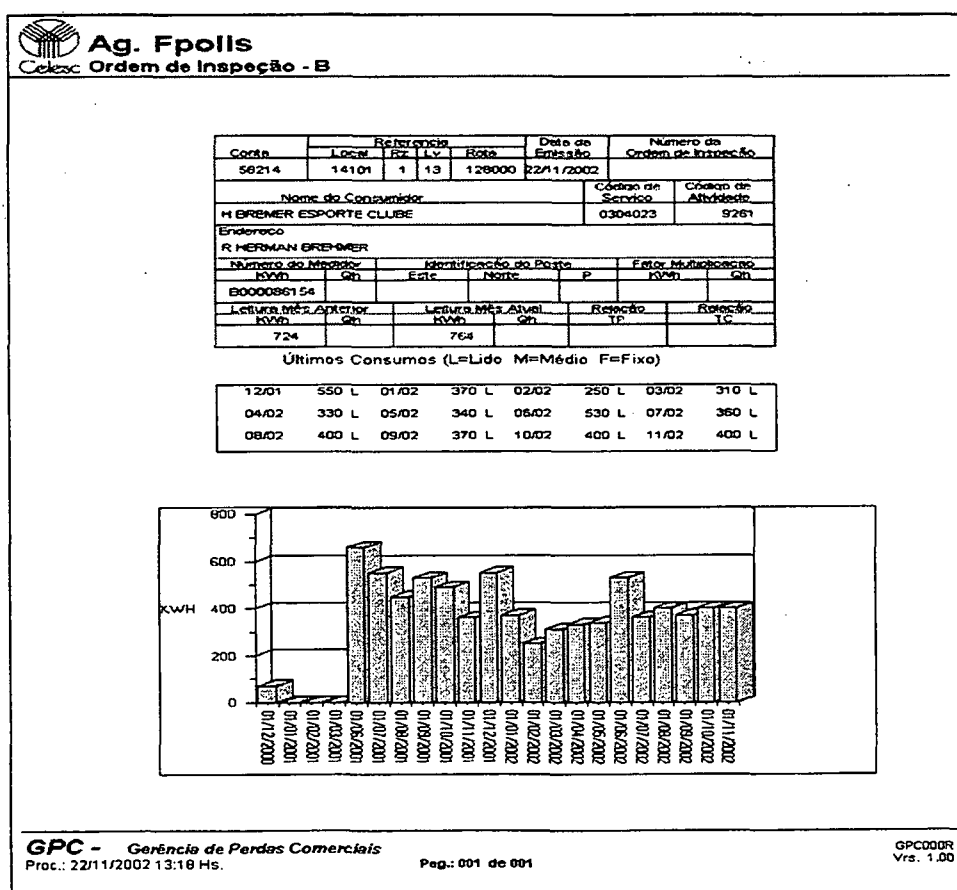
Uma vez selecionado um possível consumidor fraudador de uma determinada taxa de confiança, pode-se visualizar graficamente seus últimos 24 consumos, para uma análise alternativa da possível fraude, como é exemplificado na Figura 4.16.

Figura 4.16 - Gráfico do consumo de um possível fraudador



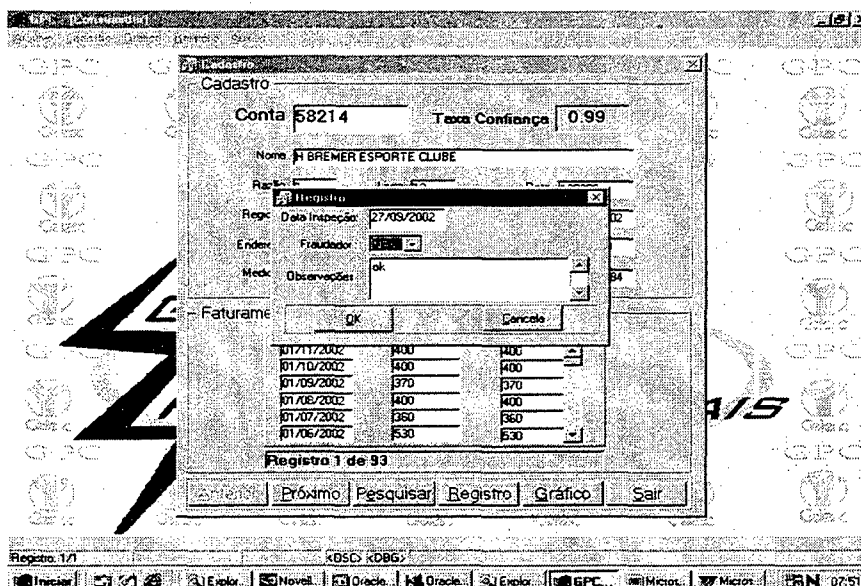
De posse de uma relação de possíveis fraudadores, os fiscais poderão ir a campo para verificar as unidades consumidoras relacionadas. Isso pode ser feito utilizando-se listagens emitidas pelo aplicativo, como é mostrado na Figura 4.17.

Figura 4.17 - Listagem de possível fraudador - aplicativo de gerência regional



Após fazer a vistoria na unidade consumidora relacionada como possível fraudadora, o responsável pela fiscalização deverá fazer o registro dessa ação, ou seja, deverá informar se existia ou não a fraude como também fornecer uma descrição de irregularidades na medição, técnica de fraude utilizada, etc., através da interface mostrada na Figura 4.18. Posteriormente, essa descrição será codificada para que se possa efetuar pesquisas mais elaboradas.

Figura 4.18 - Tela de registro de ocorrência na fiscalização



4.6.4 Aplicativos para a gerência central

Para a área da administração central intimamente ligada ao gerenciamento de perdas comerciais de energia – DPSC/DVMD – concebeu-se um modelo inicial de aplicativos que será, aos poucos, acrescido de novas funcionalidades.

Os aplicativos foram desenvolvidos em arquitetura de três camadas, como foi descrito no item 4.6.2. Eles exploram as facilidades do componente de geoprocessamento, e a interface com o usuário se dá via *browser*.

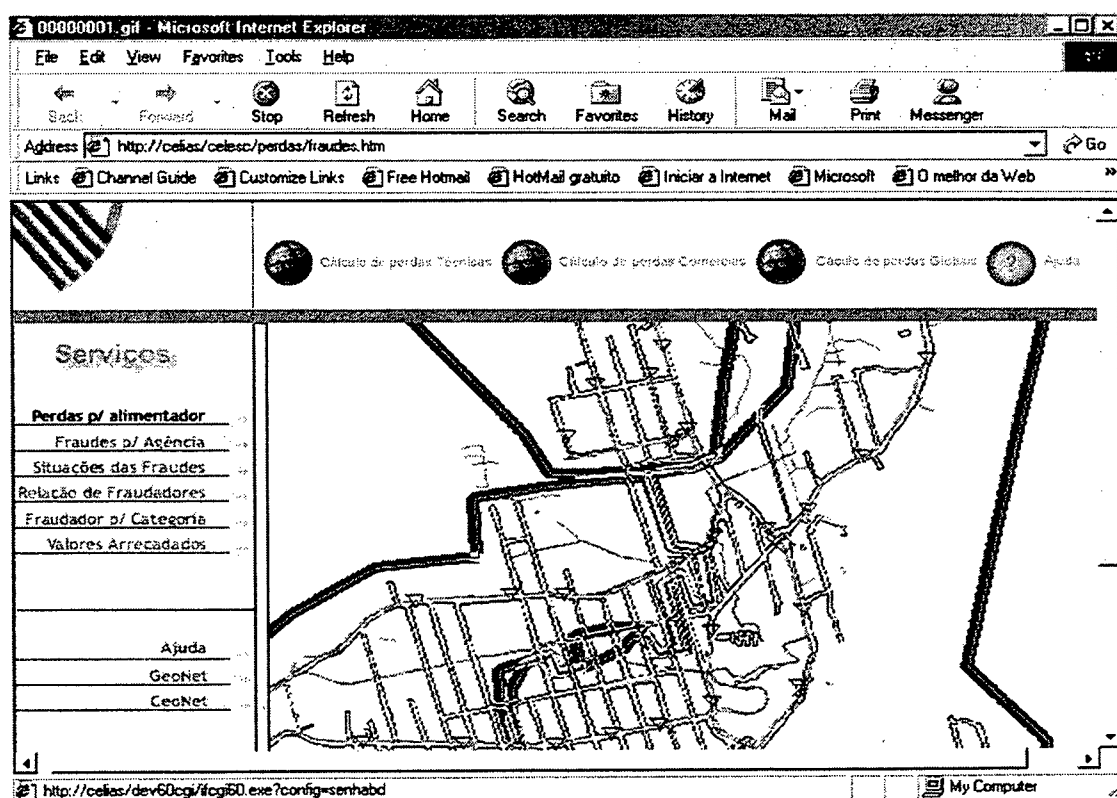
Através dos aplicativos, a gerência central poderá ter acesso a algumas informações, tais como:

- a) acompanhamento das perdas por alimentador, montando, assim, suas estratégias de recuperação financeira naqueles alimentadores com os piores índices de perda comercial de energia;

- b) acompanhamento do nível de fraudes por agência, tanto em nível de ocorrências como também em relação ao volume financeiro envolvido;
- c) localização geográfica das regiões mais problemáticas de ocorrências de fraudes, por classe de consumidor;
- d) relação dos tipos mais comuns de fraudes por região;
- e) relação de fraudadores, com dados históricos de consumo e alguns índices estatísticos;
- f) outros.

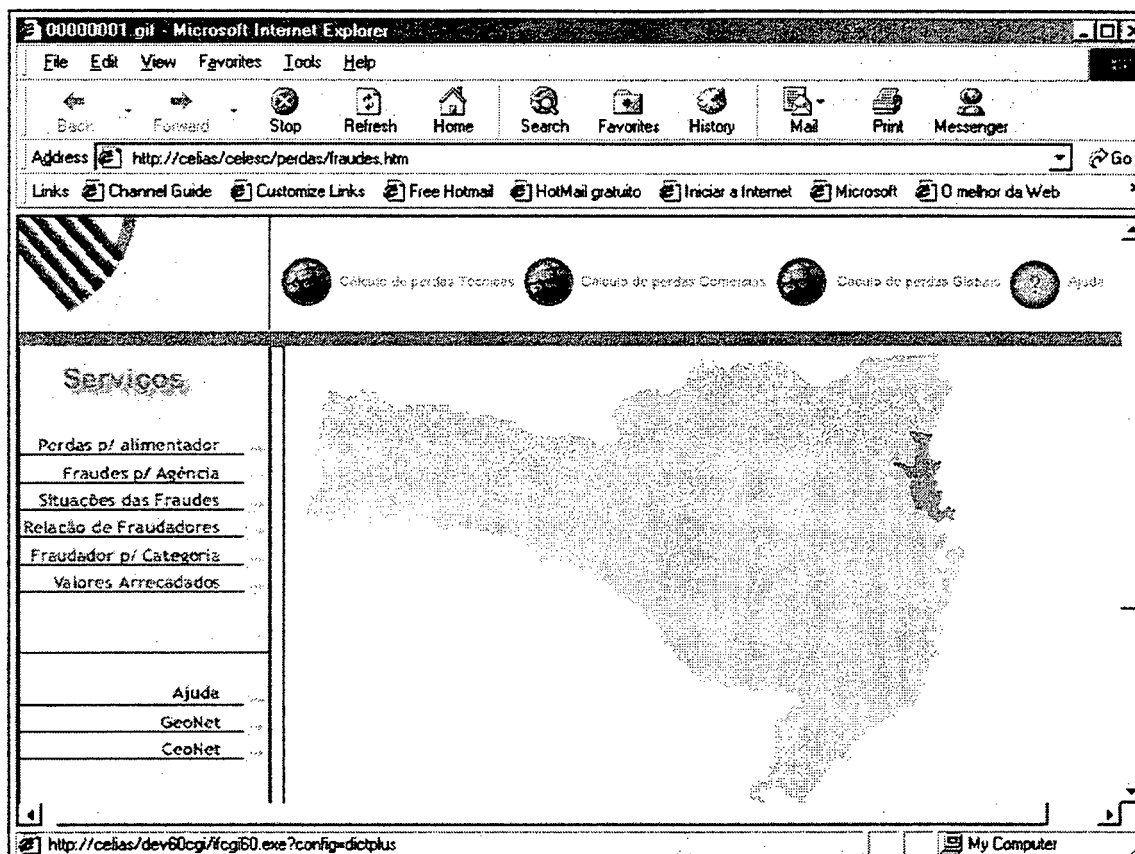
Na Figura 4.19, por exemplo, pode-se ver a seleção de uma região geográfica com o intuito de se observarem ocorrências de fraudes nessa região em destaque, tendo-se inicialmente selecionado o alimentador desejado.

Figura 4.19 - Seleção de fraudadores por região geográfica



Já na Figura 4.20, representa-se um mapa que relaciona as agências regionais, podendo-se, a partir daí, selecionar informações sobre fraudes em cada uma delas.

Figura 4.20 - Fraudes agrupadas por agências regionais



4.7 Considerações finais

Neste capítulo apresentou-se a Arquitetura de Informação para Gerenciamento de Perdas Comerciais de Energia Elétrica, detalhando-se cada um dos seus módulos.

Como foi comentado na introdução deste capítulo, trabalhou-se na implementação desse modelo com dados das Centrais Elétricas de Santa Catarina S.A. (Celesc).

No Módulo 1, foram vistos o processo de extração e tratamento dos dados que formam a entrada do Módulo 2 e também os sistemas e ambientes de origem desses dados. Trabalhou-se basicamente com dois sistemas de informação: Sistema de Consumidores e Sistema Genesis. O primeiro mantém o cadastro dos consumidores bem como o faturamento e a arrecadação deles. O segundo é responsável pelo cadastro da rede elétrica. Uma terceira base de dados foi acessada, externa à empresa, relacionando área construída, padrão de construção e tipo de construção.

Foi mostrado também que os dados de entrada para o segundo módulo podem ser precedidos de um levantamento de perdas nos alimentadores. Uma vez verificados os alimentadores com os piores índices de perdas, poder-se-á então selecioná-los para que a

busca de possíveis fraudadores fique limitada a esses alimentadores, otimizando dessa forma o esforço da equipe de fiscalização.

No segundo módulo detalhou-se a constituição do Kernel IA, tendo sido utilizadas as tecnologias RNA e SD para buscar possíveis fraudadores. Para tratar consumidores residenciais e comerciais, a tarefa selecionada foi a classificação, através de uma RNA MLP. Já para consumidores industriais, a tarefa foi a segmentação, ou agrupamento, utilizando-se para tal Kohonen e SD.

Concluindo-se este capítulo, foram abordados aspectos de aplicativos para os usuários finais. As principais preocupações foram: munir a equipe de fiscalização de recursos que possam melhorar e facilitar seus trabalhos em campo; prover recursos para a gerência regional, para que o planejamento e a efetivação das fiscalizações ocorram de forma organizada e com critérios técnicos de indicações; possibilitar à gerência central maneiras de acompanhar melhor as perdas comerciais na empresa, de forma rápida e fácil, buscando-se com isso a redução das fraudes e a conseqüente redução neste item de custo.

5 EXPERIMENTOS REALIZADOS

5.1 Introdução

Neste capítulo relacionam-se alguns dos experimentos realizados no desenvolvimento da arquitetura de informação proposta.

Em todos os experimentos realizados, utilizou-se a plataforma de desenvolvimento da *Intelligent Mining* da IBM.

Na classificação de consumidores residenciais são relacionados seis experimentos desenvolvidos com uma rede neural MLP (algoritmo *backpropagation*). Uma outra rede neural MLP foi construída para realizar vários experimentos envolvendo consumidores comerciais, sendo apresentados neste trabalho seis casos.

Os consumidores industriais receberam uma outra abordagem, ou seja, eles foram trabalhados sob a ótica da segmentação. Nos experimentos com esse segmento, utilizou-se a rede de Kohonen e a segmentação demográfica, obtendo-se um pouco mais de sucesso com esta última.

Alguns levantamentos e verificações foram realizados em campo para testar, na prática, os resultados alcançados com o uso das redes neurais desenvolvidas. O resultado de uma das idas a campo está relacionado no item 5.5, onde são retratados 24 casos de indicações de fraudes levantadas na classificação de consumidores comerciais.

5.2 Classificação de consumidores residenciais

No processo de classificação de consumidores residenciais, vários experimentos foram realizados para se obter o máximo de acertos em relação à massa de dados informada à rede neural. Em todos os experimentos realizados, o conjunto de dados para a fase de treinamento da rede continha mil ocorrências de não fraudadores e mil ocorrências de fraudadores. Para a fase de teste, foram selecionadas também mil ocorrências de fraudadores e de não fraudadores.

O conjunto de consumidores residenciais trabalhados foi da agência de Florianópolis, dos bairros Coqueiros e Abraão. É dessa região que se conseguiu a base de IPTU junto à prefeitura de Florianópolis. Um levantamento de campo foi realizado por uma equipe de

funcionários da Agência Regional de Florianópolis para associar o número da conta do consumidor da Celesc com o número de inscrição do imóvel do cadastro de imóveis da prefeitura.

Nos vários experimentos realizados, trabalhou-se basicamente com variações do número de neurônios de entrada da rede e com os seus parâmetros. Os parâmetros da rede MLP trabalhados foram a taxa de aprendizagem, o *momentum* e a época.

O número de camadas intermediárias, conhecidas também como camadas escondidas, é definido por um otimizador da própria ferramenta utilizada (*Intelligent Mining*).

A taxa de precisão, que define o percentual de classificações corretas, foi fixada em 85% para todos os experimentos. A taxa de erro, que estabelece o percentual máximo de classificações incorretas, foi fixada em 5% para os experimentos realizados.

O *momentum* é um parâmetro com valor entre 0 e 1 que modifica o padrão de treinamento da rede. Em cada passo (época), os ajustes dos pesos são baseados em uma combinação do atual ajuste do peso e o peso alterado do passo anterior (FAUSETT, 1994).

A *taxa de aprendizagem* também varia entre 0 e 1. É um parâmetro que controla o total do valor de alteração do peso durante o treinamento. No algoritmo *backpropagation* esse valor é constante, mas, em outros, pode acontecer uma redução progressiva em cada época para manter a estabilidade da rede, como ocorre com Kohonen (FAUSETT, 1994).

A *época*, ou passo, representa cada ciclo de execução do algoritmo de treinamento. Para Fausett (1994), é uma representação de cada padrão de treinamento.

Entre os vários experimentos realizados, destacam-se seis cujos percentuais de acertos (fase de treinamento e teste) estão representados na Tabela 5.1. Em todos esses experimentos, trabalhou-se com um núcleo básico de neurônios de entrada (consumo, marca do medidor, meses decorridos da última inspeção, número de fases de alimentação), conforme é mostrado no Quadro 4.1. Algumas variações foram testadas, manipulando-se diferenças de consumo, área construída, tipo de construção e padrão de construção, além dos parâmetros da rede. A Tabela 5.1 reúne todos os experimentos realizados.

Tabela 5.1 - Experimentos realizados com consumidores residenciais

Experimentos	% Acertos de Fraudadores		% de Acertos de Não Fraudadores	
	Treinamento	Teste	Treinamento	Teste
1	55.0	52.5	91.0	92.7
2	59.6	58.8	85.5	84.5
3	68.0	63.0	86.5	83.8
4	64.2	62.6	82.5	83.0
5	60.0	56.4	92.0	91.2
6	63.5	62.9	82.9	88.7

EXPERIMENTO 1

No experimento 1, informou-se à rede somente o núcleo básico de dados. Vários ajustes da rede foram realizados através dos parâmetros *momentum*, taxa de aprendizagem e época. Os resultados obtidos foram: taxa de aprendizagem = 0.1; *momentum* = 0.94; época = 500.

No experimento 1, como pode ser verificado na Tabela 5.1, os acertos conseguidos com exemplos de consumidores fraudadores ficaram num nível intermediário. Ocorreu alguma melhora com os outros experimentos, porém o resultado ficou um pouco aquém do que se pretendia.

São muitas as variáveis que determinam o comportamento de consumidores residenciais. No entanto, o conjunto de variáveis que foi possível reunir para efetuar essa tarefa foi muito limitado, o que tornou a busca de padrões comportamentais típicos de consumidores fraudadores um pouco prejudicada.

EXPERIMENTO 2

No experimento 2, informou-se à rede, além do núcleo básico de dados, a diferença de consumo em cada um dos períodos do ano, em relação ao ano anterior.

Neste experimento, utilizaram-se os seguintes valores para os parâmetros da rede: taxa de aprendizagem = 0.1; *momentum* = 0.8; época = 500.

Como pode ser constatado na Tabela 5.1, houve uma pequena melhora no acerto de consumidores fraudadores e uma pequena diminuição de acertos de consumidores não fraudadores. Como já foi comentado anteriormente, neste trabalho, tem-se certeza absoluta dos exemplos de consumidores fraudadores informados na fase de treinamento e teste da rede, o que não acontece com exemplos de não fraudadores, pois mesmo com a vistoria de rotina não se tem garantia absoluta de que realmente eles não sejam fraudadores.

EXPERIMENTO 3

Para o experimento 3, acrescentou-se aos dados básicos o total da área construída do imóvel do consumidor.

Os valores dos parâmetros da rede ficaram assim constituídos: taxa de aprendizagem = 0.17; *momentum* = 0.93; época = 500.

Como pode ser observado na Tabela 5.1, o fornecimento da área construída já contribui um pouco para melhorar o acerto da rede.

Tanto na fase de treinamento como na fase de teste, ocorreu uma pequena melhora nos índices de acertos de consumidores fraudadores.

A área construída de consumidores não fraudadores foi informada a partir da base de IPTU da prefeitura de Florianópolis. Para os consumidores fraudadores, como eles estão pulverizados entre várias agências, houve um esforço de algumas dessas agências para levantar a área construída de alguns deles, para que fosse possível usá-los nos experimentos.

EXPERIMENTO 4

O experimento 4 é o experimento 3 acrescido das diferenças de consumo de energia (períodos do ano atual menos o mesmo período do ano anterior).

Os valores dos parâmetros da rede foram os seguintes: taxa de aprendizagem = 0.21; *momentum* = 0.93; época = 500.

Como se pode notar na Tabela 5.1, a entrada da diferença de consumo não abrandou o nível de acertos em relação ao experimento 3. A rede não conseguiu melhorar a identificação de casos fraudulentos informados.

Os acertos entre a fase de treinamento e a fase de teste continuam praticamente os mesmos.

EXPERIMENTO 5

Acrescentando-se o padrão e o tipo de construção do imóvel ao experimento 3, obtém-se o experimento 5.

Os melhores valores para os parâmetros da rede foram os seguintes: taxa de aprendizagem = 0.1; *momentum* = 0.85; época = 500.

Analisando-se o resultado dos acertos de fraudadores e não fraudadores, observa-se que houve um bom índice de acertos de não fraudadores (92%) e um nível satisfatório de acertos de não fraudadores (60%).

Para essas novas variáveis, foram feitos testes com vários tipos de dados (binário, discreto, categórico e contínuo). No entanto, o nível de acertos teve pouca oscilação.

EXPERIMENTO 6

O experimento 6 constitui o experimento 5 acrescido das diferenças de consumo.

Os melhores valores para os parâmetros da rede foram os seguintes: taxa de aprendizagem = 0.2; *momentum* = 0.9; época = 500.

Neste experimento, houve uma ligeira melhora do nível de acertos de consumidores não fraudadores e uma ligeira queda nos acertos dos não fraudadores em relação ao experimento 5.

Diferentemente do experimento 4, no experimento 6 a entrada da diferença de consumo contribuiu para melhorar um pouco o nível de acertos dos exemplos de consumidores fraudadores. Já o índice de acertos de consumidores não fraudadores ficou num patamar bom.

5.3 Classificação de consumidores comerciais

Para o processo de classificação de consumidores comerciais, utilizou-se um conjunto de 300 exemplos de consumidores fraudadores e 300 exemplos de consumidores não fraudadores, tanto na fase de treinamento como também na fase de teste.

Todos os exemplos de consumidores utilizados foram originados a partir da área continental de Florianópolis, mais precisamente da região abrangendo Estreito e Barreiros.

Para classificar consumidores comerciais, uma segunda rede neural MLP foi construída através da ferramenta *Intelligent Mining*.

Nos diversos experimentos realizados, trabalhou-se basicamente com variações do número de neurônios de entrada da rede e também com os parâmetros dela. Os parâmetros da rede MLP trabalhados foram a taxa de aprendizagem, o *momentum* e a época.

Assim como ocorreu na classificação de consumidores residenciais, o número de camadas intermediárias foi definido automaticamente pelo otimizador da ferramenta *Intelligent Mining*. A taxa de precisão foi fixada em 85%, e a taxa de erro, em 5%.

Entre os vários experimentos realizados com consumidores comerciais, destacam-se seis que serão apresentados na Tabela 5.2. Em todos esses experimentos, trabalhou-se com um núcleo básico de neurônios de entrada (consumo, marca medidor, meses decorridos da última inspeção, número de fases de alimentação, ramo e sub-ramo), conforme é mostrado no Quadro 4.2. Algumas variações foram testadas, manipulando-se diferenças de consumo, número da razão de leitura, número do livro e corrente do disjuntor.

Para se chegar ao valor da corrente do disjuntor instalado no imóvel do negócio, contou-se com o apoio da equipe de fiscalização da Agência Regional de Florianópolis. Esse trabalho também foi realizado na Agência Regional de Rio do Sul, envolvendo todos os consumidores comerciais dessa agência. Os que serão relacionados a seguir dizem respeito aos consumidores comerciais de Florianópolis. Na Tabela 5.2, encontram-se os principais experimentos realizados na classificação de consumidores comerciais.

Tabela 5.2 - Experimentos realizados com consumidores comerciais

Experimentos	% Acertos de Fraudadores		% de Acertos de Não Fraudadores	
	Treinamento	Teste	Treinamento	Teste
1	61.0	74.0	78.7	69.5
2	58.0	83.0	69.0	80.4
3	68.0	85.0	78.7	85.8
4	71.0	86.0	79.6	87.4
5	70.0	77.8	86.0	88.0
6	73.0	76.9	89.0	89.1

EXPERIMENTO 1

No experimento 1, informou-se à rede somente o núcleo básico de dados. Vários ajustes da rede foram realizados através dos parâmetros *momentum*, taxa de aprendizagem e época. O que deu o melhor resultado foi a taxa de aprendizagem, com valor igual a 0.1; *momentum* = 0.7 e época = 500.

Como pode ser visto na Tabela 5.2, os acertos conseguidos com exemplos de consumidores fraudadores ficaram num nível satisfatório, o mesmo ocorrendo com os acertos de não fraudadores.

Ocorreu uma boa melhora no nível de acertos da fase de teste em relação aos consumidores não fraudadores, havendo também uma pequena redução dos não fraudadores.

Cabe novamente lembrar que o que se almeja é aumentar o nível de acertos de fraudadores, pois deles se tem certeza absoluta de sua indicação.

EXPERIMENTO 2

Para o experimento 2, foram acrescentadas aos dados básicos as diferenças de consumo dos quatro períodos do ano.

Os parâmetros da rede informados foram: taxa de aprendizagem = 0.1; *momentum* = 0.7; época = 500.

Comparando-se o experimento 2 com o experimento 1, observa-se uma melhora no nível de acertos de consumidores fraudadores e não fraudadores, na fase de teste. Já na fase de treinamento, houve uma pequena queda nos acertos.

EXPERIMENTO 3

O experimento 3 é uma variante do experimento 1, com o acréscimo das variáveis *razão e livro*.

Os melhores valores para os parâmetros foram: taxa de aprendizagem = 0.18; *momentum* = 0.8; época = 500.

O que se observa com este experimento é uma melhora no nível de acertos dos consumidores fraudadores e não fraudadores. Na fase de teste também houve uma sensível melhora, como pode ser notado na Tabela 5.2. Nota-se também uma melhora de aproximadamente 10% para os dados de teste de fraudadores em relação ao experimento 1.

O número da razão está ligado à idéia da região, e o número do livro vincula-se mais à rua (ou às ruas). Essa idéia de localização contribuiu para melhorar a precisão da rede.

EXPERIMENTO 4

No experimento 4, acrescentou-se ao experimento 3 o valor da corrente do disjuntor do estabelecimento comercial.

Os valores dos parâmetros da rede ficaram assim definidos: taxa de aprendizagem = 0.2; *momentum* = 0.85; época = 500.

Novamente, ocorreu uma pequena melhora nos acertos da rede, em relação ao experimento 3, tanto na fase de treinamento como na fase de teste.

Quando se dimensiona o disjuntor, o que se leva em consideração é o valor máximo de demanda esperado, para que possa prover a segurança necessária em caso de sobrecarga. Caso o valor da corrente ultrapasse um percentual do valor nominal, ele desliga o fornecimento de energia, protegendo assim a rede interna de energia e os equipamentos ligados nela.

EXPERIMENTO 5

No experimento 5, acrescentou-se ao experimento 1 o valor da corrente do disjuntor. Os valores dos parâmetros da rede para este experimento foram: taxa de aprendizagem = 0.17; *momentum* = 0.9; época = 500.

Com este experimento, observa-se que a idéia de carga instalada, conseguida com o valor da corrente do disjuntor, fez com que a rede obtivesse uma melhora no número de acertos com relação ao experimento 1, tanto na fase de treinamento como também na fase de teste.

Atualmente, só se realiza leitura de consumo dos consumidores comerciais. Seria interessante que houvesse também a medição da demanda, para que outros comportamentos pudessem ser observados, como, por exemplo, o fator de carga. Com novas unidades de medições eletrônicas, isso seria bastante tranquilo, mas o estudo do custo x benefício teria que ser feito.

EXPERIMENTO 6

Concluindo-se os principais experimentos com consumidores comerciais, no experimento 6 foram acrescentadas ao experimento 2 as variáveis *razão* e *livro*. Os valores dos parâmetros da rede para este experimento foram: taxa de aprendizagem = 0.1; *momentum* = 0.7; época = 500.

Com este experimento observa-se novamente que a idéia de localização melhora o nível de acerto da rede bem como o valor da corrente do disjuntor, o qual presume um determinado valor da carga instalada.

Na fase de teste também foi observado que o número de acertos permaneceu muito próximo, como pode ser constatado através dos valores da Tabela 5.2.

Está em fase de conclusão na Celesc o trabalho de codificação de localização (logradouro, bairro, rua, etc.). Uma vez concluído tal trabalho, poder-se-á utilizar essa variável para tentar melhorar o nível de acertos da rede.

Uma técnica de fraude poderá ser repassada para um ramo específico de negócio (panificadoras, por exemplo), em um determinado bairro de um município. A localização, nesse caso, será muito importante.

5.4 Segmentação de consumidores industriais

No processo de segmentação de consumidores industriais utilizou-se também a ferramenta *Intelligent Mining*. Nesta etapa, foi utilizado o mapa de Kohonen, conhecido como *Kohonen Feature Map Neural Network* e *Demographic Clustering* (segmentação demográfica).

O mapa de Kohonen baseia-se em um aprendizado não supervisionado, tendo uma camada de entrada completamente conectada a uma camada de saída competitiva. A saída, que é julgada mais integrada ao padrão de entrada (menor valor do somatório do quadrado da distância Euclidiana), é declarada vencedora. A unidade vencedora é atualizada, e a vizinhança da unidade vencedora (no processo de atualização é levado em consideração o valor decrescente da taxa de aprendizagem).

Também foi utilizado o algoritmo *Demographic Clustering*, que é um modelo que trata variáveis numéricas e categóricas de forma diferente. Neste algoritmo, utiliza-se um valor absoluto de comparação de igualdade (0 ou 1), ao passo que na rede de Kohonen utiliza-se uma métrica de distância. O algoritmo *Demographic Clustering* usa uma medida de qualidade, chamada critério *condorset*, para determinar o grupo e a qualidade do resultado. O valor do *Condorset* mede a similaridade entre os valores dos dados de um atributo; se o *Condorset* é 1, todos os valores do atributo dos registros no grupo são idênticos; caso contrário (valor igual a 0), indica que todos os valores do atributo dos registros são diferentes.

Com o mapa de Kohonen, tentou-se criar agrupamentos, de maneira que fosse possível isolar os fraudadores, ou alguns deles, em grupos específicos. Uma outra abordagem testada foi a mudança de grupo, em diferentes anos (escolhe-se um ano para gerar algumas classes que agrupem um conjunto de consumidores e, em um outro ano, verifica-se a permanência dos consumidores nas classes estabelecidas anteriormente). Não houve sucesso na tentativa de isolamento de consumidores fraudadores utilizando-se Kohonen, apesar de serem feitos vários experimentos com os dados disponíveis. O mesmo ocorreu com o teste de mudança de grupo.

Partiu-se, então, para uma segunda alternativa, explorando-se também a tarefa de segmentação, utilizando-se o algoritmo *demographic clustering*, disponível na ferramenta

Intelligent Mining. Realizaram-se, então, alguns experimentos através da variação dos dados de entrada e dos parâmetros da rede, sendo possível isolar alguns consumidores fraudadores (30%) em um grupo específico.

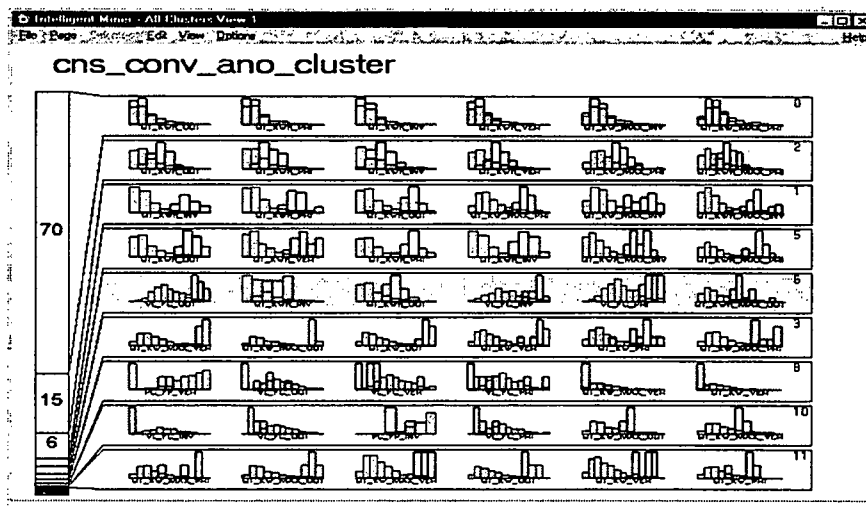
Nos experimentos, foram utilizados dados de 713 consumidores industriais, recentemente vistoriados e constatados como não fraudadores, de todo o Estado catarinense, de vários ramos de atividades, juntamente com dados de 20 consumidores industriais fraudadores.

Quatro parâmetros são informados ao algoritmo no processo de segmentação: (1) número máximo de passos, (2) número máximo de grupos, (3) precisão e (4) limiar de similaridade. O número de passos influencia a precisão da segmentação e estabelece o término da execução, caso a precisão informada não seja encontrada. O número máximo de grupos limita os grupos a serem criados; limitando-se o número de grupos, evita-se criar muitos grupos pequenos, melhorando-se o tempo no processo de segmentação; incrementando-se o número de grupos, aumenta-se a probabilidade de se encontrarem nichos. O valor do parâmetro *precisão* estabelece uma condição de parada no processo de segmentação (caso o valor seja 3, o processo de iteração termina quando a melhoria da qualidade entre dois passos é menor que 3%). O limiar da similaridade limita os valores aceitos como os melhores ajustes para um grupo (caso esse valor seja 0.30, isso quer dizer que registros com 30% de atributos idênticos são assinalados para o mesmo grupo).

No experimento que determinou o melhor resultado, foram utilizados os seguintes valores para os parâmetros: Número Máximo de grupos = 30; Número Máximo de Passos = 2; Precisão = 3; Limiar de Similaridade = 0.5.

Na Figura 5.1, pode-se observar uma amostra parcial da distribuição das variáveis em alguns *clusters*. O mapa apresentado é gerado após ser realizado o processo de segmentação demográfica.

Figura 5.1 - Gráfico de formação dos grupos a partir de uma segmentação demográfica



Na Figura 5.2, são mostrados os valores de algumas variáveis numéricas dos grupos fraudadores, fornecidas durante o processo de segmentação. Para cada variável do grupo, são fornecidos o valor mínimo, o valor máximo, o valor médio, o desvio-padrão e a unidade de distância. Esta última estabelece que se uma variável é 50% similar a uma ou mais variáveis no grupo, a variável difere pelo valor de 1 unidade de distância; quando é menor do que 50%, o valor é maior que 1.

Figura 5.2 - Valores estatísticos das variáveis numéricas do grupo de fraudadores

Cluster Field Characteristics(For Numeric Fields Only):						
Id	Name	Value Minimum	Value Maximum	Mean	Deviation Standard	Unit Distance
1	[ID_RAMO_ATVDD_CNS]	22	92	59.2857	30.2309	-
2	QT_KWH_VER	103	14760	3606.71	5083.35	33007.056
3	QT_KWH_OUT	0	5762	2547.06	2112.77	31348.277
4	QT_KWH_INV	0	3360	1317.06	1283.85	28357.841
5	QT_KWH_PRI	0	42252	7217.86	15487.5	29031.099
6	QT_KW_VER	0	66	21.2857	23.4927	99.199
7	QT_KW_OUT	0	104	26.4286	36.3678	96.836
8	QT_KW_INV	0	31	11.7143	11.6005	91.482
9	QT_KW_PRI	0	121	28.7143	42.2047	90.389
10	QT_KWH_MAX_VER	41	14760	2755.29	5321.86	12299.245
11	QT_KWH_MAX_OUT	0	3834	1201.57	1302.16	11323.018
12	QT_KWH_MAX_INV	0	1440	539.714	533.823	10312.246
13	QT_KWH_MAX_PRI	0	20029	3288.43	7391.67	10844.044
14	QT_KW_MAX_VER	0	66	14.1429	23.2983	35.056
15	QT_KW_MAX_OUT	0	54	12	19.0526	35.618
16	QT_KW_MAX_INV	0	12	4.57143	4.27618	31.728
17	QT_KW_MAX_PRI	0	42	10.7143	14.4535	32.558
18	PC_FP_VER	0	86.06	55.7043	32.3646	4.836
19	PC_FP_OUT	0	98.15	52.1057	44.5592	5.642
20	PC_FP_INV	0	87.01	49.71	36.4762	7.252
21	PC_FP_PRI	0	91.07	58.74	40.9091	5.258
22	VL_FC_VER	0	0.5424	0.163014	0.179045	0.075
23	VL_FC_OUT	0	0.3749	0.1456	0.140964	0.872
24	VL_FC_INV	0	0.5271	0.141671	0.178175	0.070
25	VL_FC_PRI	0	0.6115	0.211729	0.238935	0.073

Na Figura 5.3, são mostrados os valores de algumas variáveis numéricas e não numéricas dos grupos fraudadores, fornecidas durante o processo de segmentação. Para cada variável do grupo, são fornecidos o valor modal, a frequência modal, o valor qui-quadrado, o valor do *condorcet* e a entropia. O valor modal mostra o valor da variável que ocorre mais freqüentemente no resultado. A frequência modal indica o percentual de registros do resultado que têm o valor modal. O qui-quadrado mede o quanto a distribuição dos valores da variável dos registros em um grupo difere da distribuição dos valores da variável para todos os registros da população. O valor da entropia é similar ao valor do qui-quadrado, exceto que valores inválidos ou ausentes no grupo são incluídos no cálculo do valor da entropia.

Figura 5.3 - Valores estatísticos das variáveis de todos os tipos do grupo de fraudadores

Id	Name	Type	Value Modal	Frequency Modal(%)	Chi-Squared	Value Condorcet	Entropy
1	{ID_RAMO_ATVDD_CNS}	DN	92	28.57	0.212	-	2.057
2	QT_KWH_VER	CO	10000	100.00	0.310	0.9572	1.314
3	QT_KWH_OUT	CO	10000	100.00	0.300	0.9397	1.207
4	QT_KWH_INV	CO	10000	100.00	0.238	0.9139	0.962
5	QT_KWH_PRI	CO	10000	85.71	0.179	0.7650	0.804
6	QT_KW_VER	CO	25	85.71	0.954	0.9411	2.223
7	QT_KW_OUT	CO	25	85.71	0.800	0.8657	2.083
8	QT_KW_INV	CO	25	100.00	0.759	0.9711	2.132
9	QT_KW_PRI	CO	25	85.71	0.609	0.8376	1.873
10	QT_KWH_MAX_VER	CO	2500	85.71	0.372	0.8594	1.621
11	QT_KWH_MAX_OUT	CO	2500	100.00	0.503	0.9002	1.932
12	QT_KWH_MAX_INV	CO	2500	100.00	0.302	0.9611	1.444
13	QT_KWH_MAX_PRI	CO	2500	05.71	0.289	0.7598	1.442
14	QT_KW_MAX_VER	CO	10	05.71	0.984	0.7657	2.131
15	QT_KW_MAX_OUT	CO	10	05.71	0.917	0.8158	2.026
16	QT_KW_MAX_INV	CO	10	100.00	0.714	0.9628	1.987
17	QT_KW_MAX_PRI	CO	10	85.71	0.758	0.8284	1.870
18	PC_FP_VER	CO	73.75	28.57	1.999	0.4914	1.725
19	PC_FP_OUT	CO	71.25	14.29	0.413	0.3624	0.735
20	PC_FP_INV	CO	72.5	14.29	0.593	0.5646	0.815
21	PC_FP_PRI	CO	91.25	28.57	0.360	0.4971	1.023
22	VL_FC_VER	CO	0.025	20.57	1.152	0.4321	2.109
23	VL_FC_OUT	CO	0.025	42.86	1.529	0.3506	1.993
24	VL_FC_INV	CO	0.125	42.86	0.667	0.6153	1.662
25	VL_FC_PRI	CO	0.025	28.57	1.100	0.3302	1.903

Na Figura 5.4, são mostradas algumas informações do grupo que reúne alguns dos consumidores fraudadores. Pode-se ver, por exemplo, o tamanho absoluto e relativo do grupo; o valor do *condorcet* global e do grupo; o valor da similaridade com relação aos outros grupos, etc. Se os valores de todas as variáveis dos registros são idênticos, então o *condorcet* é 1 e, no outro extremo, se todos são diferentes, o valor é 0.

As figuras 5.5 e 5.6 dão alguns detalhes do processo de segmentação bem como algumas características dos grupos formados. Uma das informações relacionada na Figura 5.6, por exemplo, é o valor de similaridade de cada um dos grupos com todos os demais grupos.

Figura 5.4 - Valores estatísticos do grupo de fraudadores

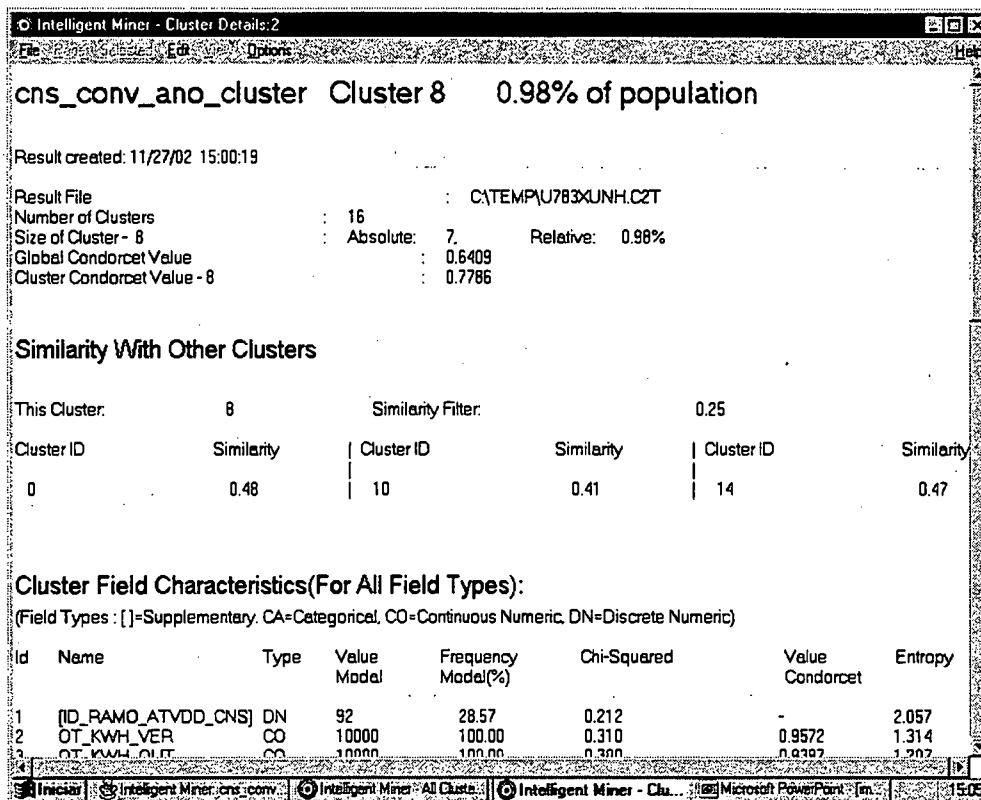


Figura 5.5 - Valores estatísticos de alguns grupos

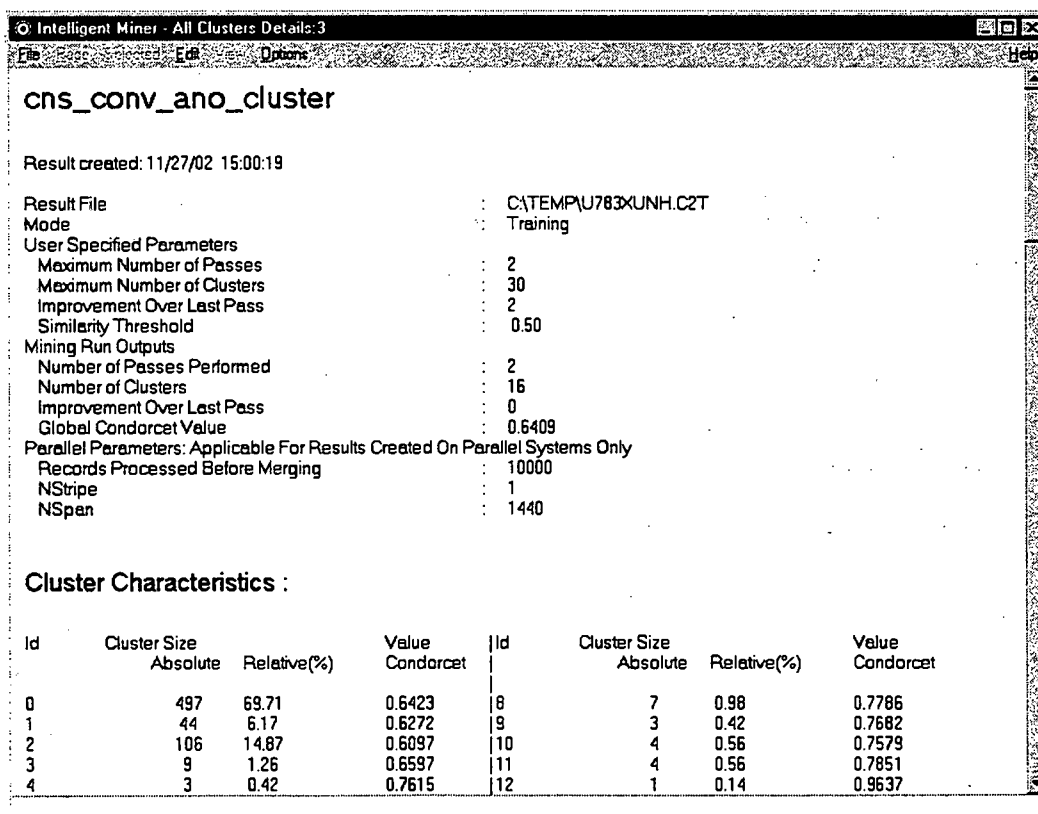


Figura 5.6 – Características e similaridades dos grupos

Intelligent Miner - All Clusters Details:3

File Page Selected Edit View Options Help

Cluster Characteristics :

Id	Cluster Size		Value Condorcet	Id	Cluster Size		Value Condorcet
	Absolute	Relative(%)			Absolute	Relative(%)	
0	497	69.71	0.6423	8	7	0.98	0.7786
1	44	6.17	0.6272	9	3	0.42	0.7682
2	106	14.87	0.6097	10	4	0.56	0.7579
3	9	1.26	0.6597	11	4	0.56	0.7851
4	3	0.42	0.7615	12	1	0.14	0.9637
5	14	1.96	0.6928	13	3	0.42	0.7833
6	14	1.96	0.7266	14	1	0.14	0.9544
7	2	0.28	0.8331	15	1	0.14	0.9629

Similarity Between Clusters

Similarity Filter: 0.25

Cluster 1	Cluster 2	Similarity	Cluster 1	Cluster 2	Similarity
0	2	0.36	4	9	0.41
0	6	0.41	4	13	0.42
0	8	0.48	4	15	0.52
0	10	0.44	5	6	0.40
0	11	0.40	5	7	0.47
0	12	0.27	5	9	0.38
0	14	0.40	5	10	0.27
1	2	0.37	5	12	0.31
1	3	0.43	5	13	0.43
1	4	0.46	5	14	0.25
1	5	0.49	5	15	0.32
1	6	0.31	6	7	0.38

Iniciar Intelligent Miner: cns_conv Intelligent Miner - All Cluste... Microsoft PowerPoint - [m... Intelligent Miner - All ... 15:09

5.5 Verificação de campo

Para realizar a verificação em campo de possíveis fraudadores, gerados a partir do uso da arquitetura proposta, foram selecionados 1.340 consumidores comerciais da parte continental de Florianópolis. A partir da interfase do aplicativo de gerência regional, a equipe de fiscalização gerou os relatórios, conforme mostra a Figura 4.17, para taxas de confiança entre 0.95 e 1.0, relacionando-se com isso 24 casos a serem averiguados.

Nas unidades consumidoras relacionadas, a equipe realizou uma fiscalização mais detalhada, usando inclusive alguns equipamentos.

Após retornar do campo, a equipe registrou na base de dados, através da interface do aplicativo de gerência regional, os resultados obtidos:

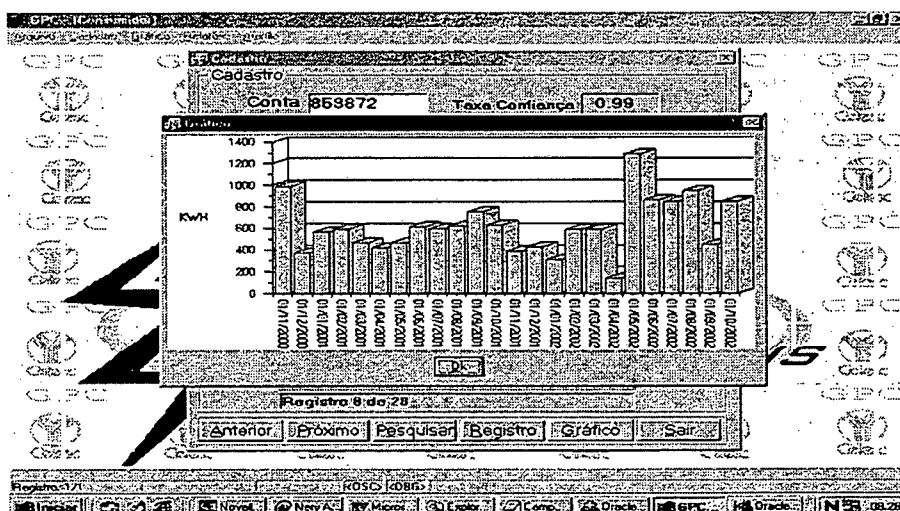
- Caso 1: mudança de uso da propriedade
- Caso 2: novo inquilino
- Caso 3: trocado medidor por ser antigo
- Caso 4: normal
- Caso 5: trocado o medidor. Registrador travado
- Caso 6: consumidor desligado
- Caso 7: imóvel desocupado
- Caso 8: normal
- Caso 9: imóvel desocupado
- Caso 10: imóvel desocupado
- Caso 11: normal
- Caso 12. normal
- Caso 13: leitura anotada maior do que a real
- Caso 14: diminuiu a carga
- Caso 15: diminuiu o consumo
- Caso 16: estava em reforma
- Caso 17: instalou novo equipamento
- Caso 18: estava desocupado
- Caso 19: desvio de energia após o disjuntor e antes do medidor
- Caso 20: normal
- Caso 21: normal

- Caso 22: normal
- Caso 23: normal
- Caso 24: desvio de energia após o disjuntor e antes do medidor

Como descrito acima, dois casos de fraudes foram constatados com um valor estimado de retorno de 18 mil reais.

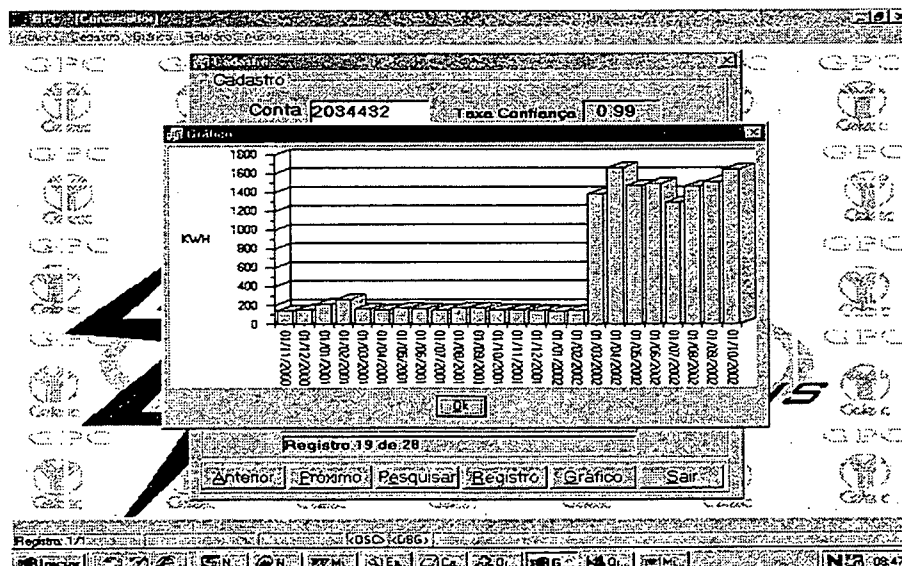
Na Figura 5.7, é apresentado o gráfico de consumo de energia do fraudador do caso 19, relacionado anteriormente. Como se pode observar pelo gráfico, o consumo do estabelecimento comercial é bastante irregular.

Figura 5.7 - Gráfico do consumo de energia do fraudador do caso 19



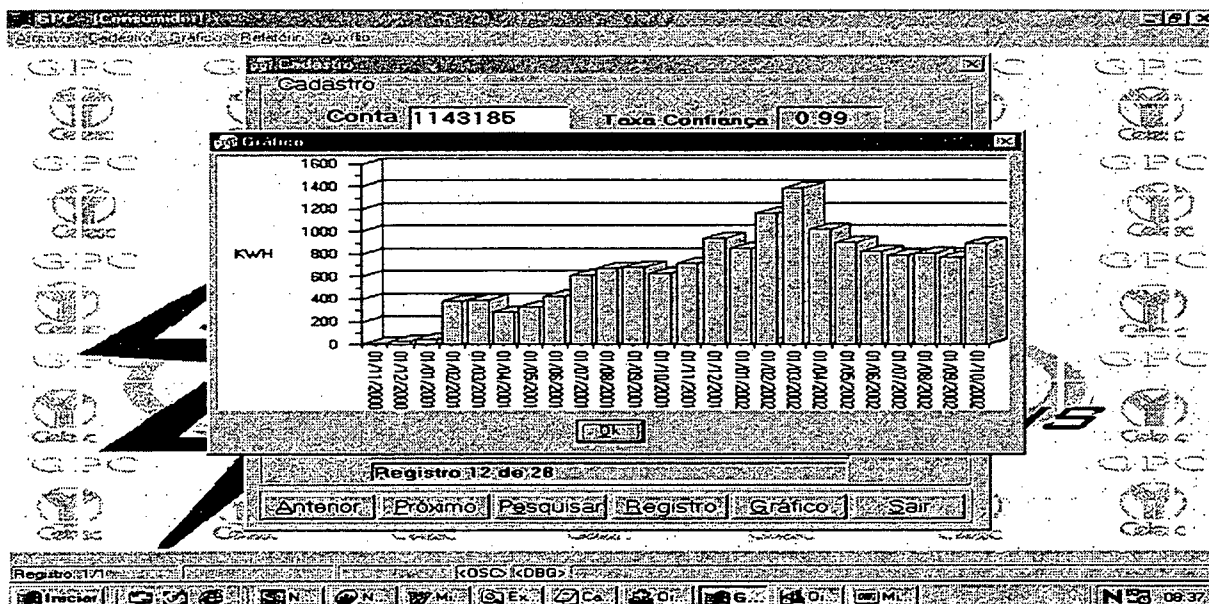
O outro caso de fraude é mostrado na Figura 5.8, na qual se pode notar que num determinado período o consumo era bastante baixo. Com o resultado da identificação da fraude, a medição real de consumo de energia foi restabelecida.

Figura 5.8 - Gráfico do consumo de energia do fraudador do caso 19



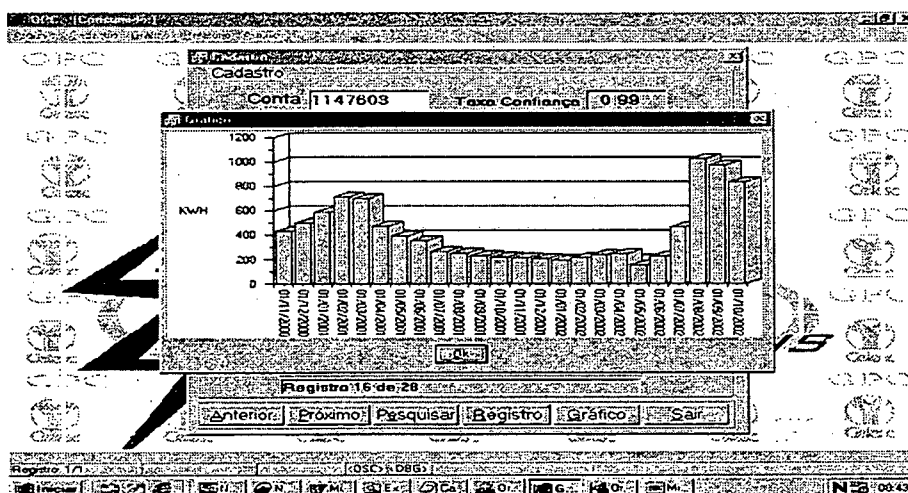
Um outro gráfico interessante é o da Figura 5.9. Esta figura retrata o caso 16, cujo imóvel estava passando por reformas. É possível observar que próximo à posição “0” do eixo “x”, o consumo estava realmente baixo, o que configura uma irregularidade.

Figura 5.9 - Gráfico do consumo de energia do caso 16 – imóvel em reformas



No gráfico da Figura 10, a informação de campo relata que o consumidor simplesmente diminuiu o consumo (caso 15). No entanto, o que se observa é que, após a fiscalização, o consumidor voltou a ter um nível maior de consumo, podendo-se configurar em uma fraude não descoberta.

Figura 5.10 - Gráfico do consumo de energia do caso 15 – imóvel em reformas



5.6 Considerações finais

Nos experimentos realizados com consumidores residenciais, pode-se verificar que outras variáveis necessitam ser incorporadas ao modelo para se tentar descobrir comportamentos típicos de consumidores fraudadores. Somente com a relação de consumo ao longo dos meses, tempo decorrido da última inspeção, número de fases, tipo do medidor e área do imóvel fica realmente difícil melhorar o nível de acertos da rede neural. Todavia, os resultados obtidos indicam que essa técnica deve continuar a ser explorada, buscando-se a melhoria da entrada da rede para que o comportamento do fraudador seja mais claro de ser percebido por ela. Alguns dados que poderiam ser fornecidos seriam, por exemplo, itens elétricos com a respectiva potência, características dos moradores (idade, sexo, renda, etc.), nível socioeconômico, informações de inadimplência, etc.

Com relação a consumidores comerciais, o resultado foi mais promissor, como pode ser constatado pela verificação de campo. No entanto, mais testes devem ser feitos, envolvendo um universo bem maior de consumidores, de modo que se possa verificar se existe particularidade regional. Este trabalho poderá ser realizado quando a codificação de endereço estiver concluída na empresa. Seria interessante também se ter detalhes da carga instalada do estabelecimento comercial, horário médio de funcionamento, fator de carga típico por ramo de negócio, fator de demanda típico por ramo de negócio, etc.

Consumidores industriais exigirão um esforço maior em estudos, pesquisas e testes, pelo volume de consumo envolvido e pelo impacto causado na arrecadação, dependendo do porte do negócio. Por outro lado, esse tipo de consumidor é vistoriado duas vezes no ano, na grande maioria por uma equipe de técnicos capacitados para realizar tal trabalho. Eles também são monitorados mensalmente através de alguns alertas do Sistema de Consumidores, quando ocorrem oscilações de consumo, demanda e fator de potência. Com isso, as ocorrências de fraudes são bem mais reduzidas. Paralelamente à utilização de redes neurais para identificação de possíveis fraudadores, estuda-se, a partir deste trabalho, um banco de regras que irá gerar um sistema especialista, e também a utilização do fator de carga típico, fator de demanda típico e fator de utilização típico (relaciona potência do transformador e consumo) para se buscarem possíveis fraudadores. Nos experimentos realizados com a segmentação desse tipo de consumidor, já ocorreu um avanço, pois se conseguiu isolar 30% de fraudadores em segmento único. No entanto, ainda existe muito espaço para que o modelo seja aperfeiçoado e, para contribuir com essa melhoria, novas informações poderiam ser agregadas: detalhes da carga instalada, dados contábeis da empresa, etc.

6 CONCLUSÕES E TRABALHOS FUTUROS

6.1 Conclusões

A perda da arrecadação proveniente das perdas comerciais de energia, e mais especificamente da fraude, atinge cifras preocupantes para empresas do setor de distribuição de energia elétrica, como foi apresentado nos itens 1.1 e 1.3 deste trabalho.

Em períodos de racionamento de energia elétrica, como aconteceu recentemente devido à estiagem, e em momentos em que a energia atinge um percentual elevado na planilha de custo mensal de uma residência, comércio e indústria, a tendência é a de se ter uma elevação acentuada no consumo ilícito de energia elétrica. Uma outra constatação, segundo alguns fiscais da Celesc, é que a ousadia e a esperteza dos fraudadores não têm limites, encontrando-se desde casos rudimentares em termos de método de burlar até mecanismos realmente engenhosos.

Em muitos casos, a medição eletrônica poderá ser uma grande aliada no combate à fraude; no entanto, não a exterminará, pois a capacidade humana de criar mecanismos para pagar menos energia, ou não pagar, é realmente considerável.

O processo de descoberta de fraudadores através do estudo de padrões de consumo ilícito, utilizando-se técnicas de mineração de dados, poderá ser de grande valia no combate ao consumo irregular de energia elétrica. Nos experimentos realizados e nas verificações em campo, constatou-se que o potencial da arquitetura que se está propondo deve ser levado em consideração pela empresa. No entanto, urge que se aprimore mais o modelo e que se busquem mais informações para que a descoberta de padrões comportamentais fraudulentos seja mais fácil de ser percebidos pela rede. Deve-se também realizar treinamento/teste/aplicação/verificação em campo de um volume mais abrangente de casos que envolvam, por exemplo, toda uma agência.

A metodologia para se calcular as perdas técnica e comercial por alimentador terá uma importância bastante expressiva, pois tornará possível a classificação dos alimentadores pelo índice de perda comercial, possibilitando dessa forma a geração de possíveis fraudadores para os alimentadores com os piores índices de perda comercial. Esta opção também propicia um melhor aproveitamento das equipes de fiscalização, haja vista que elas poderão atuar de forma mais concentrada nos locais onde a incidência de fraudes/irregularidades de medição são mais expressivas. O modelo de cálculo de perda por alimentador apresentado neste trabalho é

apenas um embrião. Com o seu aperfeiçoamento, através de trabalhos de campo e melhoramentos nos algoritmos e cálculos matemáticos, a taxa de erro da perda comercial poderá ser bastante reduzida, tornando-se dessa forma um bom mecanismo auxiliar para identificação de consumidores fraudadores.

6.2 Contribuições

As contribuições que serão enumeradas a seguir levam em conta o aspecto comercial da arquitetura de informação proposta e também a contribuição acadêmica.

Para uma empresa de distribuição de energia elétrica como, por exemplo, a Celesc, que serviu como laboratório de estudos e implementações, destacam-se as seguintes contribuições:

- um teste bastante rudimentar da nova arquitetura de informação possibilitou um nível de acertos de 8,3% nos prognósticos emitidos para consumidores comerciais. Para 192.283 fiscalizações, realizadas em consumidores comerciais em 2002 (BEC, 2002), foram encontrados 462 casos de fraudes, o que equivale a um índice de descoberta de fraudes de 0,24%. Como pode ser observado pelos valores apresentados, a contribuição nesse item pode ser bastante significativa;
- atualmente, o Sistema de Consumidores da Celesc é baseado em *mainframes*, tendo como interface os terminais da família 3270 (*character mode*), da IBM. O que se está propondo são soluções voltadas para a WEB, tendo como principal componente de manipulação das informações ferramentas baseadas em geoprocessamento;
- hoje, quando os fiscais estão em campo, o universo de informações dos consumidores se resume a simples relatórios em papel. Na nova arquitetura, esse universo amplia-se consideravelmente, através da tecnologia *wireless*. O computador, nesse caso, é *handheld*, e a comunicação explora a telefonia celular. Num primeiro momento, poder-se-á trabalhar *off-line*, através do processo de sincronização;
- propõe-se também uma metodologia para se calcular perda técnica e comercial por alimentador, de forma a auxiliar a busca mais efetiva de fraudadores;

- poderá ocorrer um aumento da arrecadação pela possibilidade de se descobrir um maior número de consumos fraudulentos;
- os prognósticos de fraudes emitidos através da arquitetura de informação proposta poderão constituir na realidade casos de defeitos na medição. Ter-se-á então, de forma indireta, mais uma possibilidade de aumentar a arrecadação com consumos não faturados;
- possibilidade de melhorar os prognósticos de fraudes, com o uso mais intenso da arquitetura proposta.

Para a comunidade acadêmica, destacam-se os seguintes pontos:

- utilização com algum sucesso de um conhecimento científico (RNA) para resolver um problema complexo, abrangente e de grande impacto em uma empresa de distribuição de energia elétrica;
- realizou-se um teste com o alimentador BCU02, localizado em Biguaçu, durante o mês de novembro de 2002, colocando-se um medidor de demanda e consumo na saída do alimentador, dentro da subestação. Esse alimentador possui 228 consumidores secundários (residências e comércios, basicamente) e oito primários (indústrias). O cálculo das perdas realizado segundo o item 4.4 teve praticamente o mesmo resultado que o valor conseguido com as medições (perda comercial = 1,6%; perda técnica = 1,43%; perda global = 3,03%). Isso indica que a metodologia que se está desenvolvendo para o cálculo de perdas técnicas e comerciais, se aprimorada ainda mais, poderá ter um grande potencial na descoberta de consumidores fraudadores;
- a utilização de conhecimentos técnicos e científicos – IA, geoprocessamento, WEB, *wireless communication*, estatística, etc – possibilitou o desenvolvimento de uma arquitetura de informação para descobrir consumidores fraudadores, com o potencial de se tornar uma ferramenta de trabalho do dia-a-dia de uma empresa do setor;
- utilização de técnicas de IA (RNA, MLP e KOHONEN) para classificar e segmentar consumidores como também utilização de métodos estatísticos (variância, desvio-padrão, etc.) e outros algoritmos (*demographic clustering*) para tratar adequadamente cada um dos grupos de consumidores trabalhados

(residencial, comercial e industrial). A comprovação da validade do uso das técnicas se deu com verificação em campo dos prognósticos emitidos pela arquitetura.

- o estudo das variáveis endógenas e exógenas aos ambiente Celesc, possibilitou melhorar o prognóstico do *engine IA*. Através desta constatação, poder-se-á buscar outros atributos que possam melhorar ainda mais a taxa de acertos dos prognósticos.

6.3 Trabalhos futuros

Como sugestão para trabalhos futuros, o mais importante, com toda a certeza, será o desenvolvimento de uma ferramenta que contemple toda a arquitetura descrita neste trabalho.

Um outro aspecto que merecerá uma atenção especial diz respeito ao Kernel IA, sobre o qual deverão ser realizados estudos mais abrangentes, envolvendo talvez outros algoritmos e redes, para que o nível de acertos de fraudadores seja ainda maior, principalmente aqueles verificados em campo. Tentar identificar o peso e a importância de cada um dos atributos no conjunto de acertos de indicativos de fraudes, levando-se em consideração um volume maior de verificações de campo, é outro tema que merecerá atenção.

Acredita-se que seja bastante óbvio que a quantidade de variáveis trabalhadas para buscar comportamentos suspeitos é bastante reduzida, principalmente envolvendo consumidores residenciais, visto que todos possuem uma única denominação e várias possibilidades comportamentais de consumo de energia. Isso faz com que seja necessário buscar outras fontes de informações, internas ou externas à empresa (Celesc), para que o modelo seja aperfeiçoado.

Um outro item que poderá gerar vários trabalhos, pois está só começando, é o cálculo de perdas comerciais nos alimentadores. A maioria das subestações da Celesc encontram-se atualmente automatizadas, produzindo, minuto a minuto, valores de grandezas elétricas, tais como corrente, tensão, etc. O valor de demanda e energia na saída do alimentador é muito tranquilo de se calcular no início de cada mês, pois todas as variáveis estão presentes e em fase. O mesmo não acontece com o valor medido nas unidades de medição, nas quais a leitura é realizada ao longo do mês, ocorrendo, portanto, uma defasagem. Esse é apenas um dos problemas a serem resolvidos. Um outro, por exemplo, diz respeito à constituição de famílias

que identifique cada um dos transformadores em um alimentador. A construção de curvas de carga-padrão também merecerá estudos mais aprofundados.

O resumo de alguns desses trabalhos é o seguinte:

- trabalhar mais variáveis na camada de entrada das redes e também para os algoritmos;
- melhorar o processo de segmentação de consumidores industriais;
- testar outros algoritmos/redes para as diversas categorias de consumidores;
- melhorar a qualidade dos dados (Sistema de Consumidores);
- estudar maneiras de melhorar a precisão das inspeções;
- realizar testes mais abrangentes (uma ou mais agências) para refinar o modelo;
- concluir a metodologia de perda por alimentador e confrontá-lo com medições de campo.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRADEE. **Memória da eletricidade**. Disponível em: <<http://www.abradee.com.br>> Acesso em: 23 out. 2002.

AGRAWAL, R.; IMICLINSKI, T., SWAMI, A. Mining association rules between sets of items in large database. In: Proc. ACM SIGMOD International Conference on Management of Data. Washington, DC, USA, 1993.
Disponível em: <<http://CiteSeer.nj.nec.com>> Acesso em: 23 abr. 2002.

AGRAWAL, R. et al. Automatic subspace clustering of high dimensional data for data mining applications. In: Proc. ACM SIGMOD International Conference on Management of Data. Washington, DC, USA, p. 94-105, 1998.
Disponível em: <<http://CiteSeer.nj.nec.com>> Acesso em: 23 abr. 2002.

AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules. In: Proc. of the 20th Conference on Very Large Databases, Santiago, Chile, 1994.
Disponível em: <<http://www.almaden.ibm.com/cs/people/srikant/publications.html>> Acesso em: 23 abr. 2002.

ANEEL. Resolução n. 456 de 29 de novembro de 2000. Dispõe sobre condições gerais de fornecimento de energia elétrica. **Diário Oficial da República Federativa do Brasil**, Brasília, 27 mar. 2001.

_____. **Lei de criação/regimento interno**. Disponível em: <www.aneel.gov.br>. Acesso em: 01 maio 2002.

ARAÚJO, D. L. A.; LOPES, H.S.; FREITAS, A.A. A parallel genetic algorithm for rule discovery in large databases. In: IEE SYSTEMS, MAN AND CYBERNETICS CONFERENCE, 1999, Tokyo. Disponível em:
<<http://www.ppgia.pucpr.br/~alex/papers.html>> Acesso em: 22 set. 2002.

BARAGOIN, C. et al. Mining your own business in telecoms using DB2 intelligent mining for data. IBM, California, p. 1-144, 2001.

BEC. **Boletim estatístico comercial de dezembro de 2001** – Celesc. Florianópolis, 2001

_____. **Boletim estatístico comercial de dezembro de 2002** – Celesc. Florianópolis, 2002

BERRY, M. J. A.; LINOFF, G. **Data mining techniques – for marketing, sales and customer support**. New York: John Wiley & Sons, 1997.

BETTIO, J. A. de. **Constatação de procedimentos irregulares e deficiência no medidor ou demais equipamentos de medição**. CELESC, 2001. Apostila de curso.

BISHOP, C. M. **Neural networks for pattern recognition**. New York: Oxford University Press, 1997.

BONCHI, F.; GIANNOTI, F.; MAINETTO, G.; PEDRESCHI, D. Using data mining techniques in fiscal fraud detection. In: PROC-DAWORK '99, FIRST INT. CONFERENCE ON DATA WAREHOUSING AND KNOWLEDGE DISCOVERY, 1999, Itália. Disponível em: <<http://CiteSeer.nj.nec.com>> Acesso em: 22 set 2002.

BRAGA, A. P.; LUDEMIR, T. B.; LEON CARVALHO, A. C. P. de. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: Livros Técnicos e Científicos, 2000. 262 p.

CELESC. **Quem somos: histórico**. Disponível em: <<http://www.celesc.com.br/quemsomos/historico.php>>. Acesso em: 01 nov. 2002.

CEPEL. **Processo de estabelecimento do preço MAE**. Florianópolis: CEPEL/ASMAE/UFSC, UFSC/EEL/LABPLAN, 1999. 98 p.

CHAN, P. K.; STOLFO, S. J. Toward scalable learning with non-uniform class and cost distributions: a case study in credit card fraud detection. In Proc. of the Fourth International Conference on Knowledge Discovery and Data Mining. Washington, DC, USA, p. 164-168, 1998. Disponível em: <<http://CiteSeer.nj.nec.com>> Acesso em: 22 out 2002.

CHEN, M-S.; HAN, J.; YU, P. S. Data mining: an overview from database perspective. In: IEEE Transactions on Knowledge and Data Engineering, v. 8, n. 6, p. 866-883, 1996. Disponível em: <<http://master.cpe.ku.ac.th/mcpe/204562/survey97.pdf>> Acesso em: 15 jun. 2002.

CHEUNG et al. Efficient mining of association rules in distributed databases. In: IEEE Transactions on Knowledge and Data Engineering, v.8, n. 6, p. 911-922, 1996. Disponível em: <<http://CiteSeer.nj.nec.com>> Acesso em: 15 Jun. 2002.

CHOU, P. B. et al. Identifying prospective customers. In. SIGKDD 2000. 2000. Disponível em: <<http://dblab.cs.ucr.edu/papers/KDD2000-dg.pdf>> Acesso em: 01 nov. 2002.

CODI. **Documento técnico CODI-08.05**. Brasília, 1997.

CODI. **Método para determinação, análise e otimização das perdas técnicas em sistemas de distribuição – CODI-3.2.19.34.0**. Brasília, 1996.

DARIANO, D. **Jornal do Brasil**, Rio de Janeiro, 5 dez. 2001. Caderno Cidade.

FAUSETT, L. V. **Fundamentals of neural networks**. New Jersey: Prentice-Hall, 1994.

FAYYAD, U.; PIATETSKY-SHAPIRO, G; SMITH, P. From data mining to knowledge discovery: An overview. In: FAYYAD, U.; PIATETSKY-SHAPIRO, G; SMITH, P.; UTHURUSAMY, R.(Ed.). **Knowledge Discovery and Data Mining**. California: AAAI/MIT, 1996. P. 1-34.

FRANCISCO, L. **Jornal Folha de São Paulo**, 01 fev. 2001. Folha Dinheiro.

FREITAS, A. A. Understanding the crucial differences between classification and discovery of association rules – a position paper. In: SIGKDD Explorations, v.2, 2000. Disponível em: <<http://CiteSeer.nj.nec.com>> Acesso em: 06 jul. 2002.

- GOEBEL, M.; GRUENWALD, L. A survey of data mining and knowledge discovery software tools. In: SIGKDD Explorations, v. 1, p. 20-33, 1999.
Disponível em: <<http://CiteSeer.nj.nec.com>> Acesso em: 06 jul. 2002.
- HAIR JR., J. F. et al. **Multivariate data analysis**. 5 ed. New Jersey: Prentice-Hall, 1998.
- HARRISON, T. H. **Intranet data warehouse**. São Paulo: Berkeley Brasil, 1998.
- HAYKIN, S. **Neural networks: a comprehensive fundation**. New Jersey: Macmillan Publishing Company, 1994.
- HYNNINEN, J. Experiences in mobile phone fraud. In: Semminar on Network Security. 2000. Disponível em: <<http://www.niksula.cs.hut.fi/~jthynnin/mobfra.html>> Acesso em: 23 maio 2002.
- JDP. **Jornal Diário de Pernambuco**, Pernambuco, 29 abr. 2001. Caderno de Economia.
- JOESP. **Jornal O Estado de São Paulo**, São Paulo, 01 jun. 2001. Editorial, p.2.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4. ed. New Jersey: Prentice-Hall, 1998.
- METHA, M. et al. SLIQ: a fast scalable classifier for data mining. In: EDBT'96. Avignon, França, 1996. Disponível em:
<http://www.almaden.ibm.com/u/ragrawal/papers/edbt96_sliq.ps> Acesso em: 23 jul. 2002.
- MOREAU, Y.; VANDEWALLE, J. Detection of mobile phone fraud using supervised neural networks: a first prototype. In: ICANN'97. 1997. Disponível em:
<<http://CiteSeer.nj.nec.com>> Acesso em: 22 set 2002.
- NATORE, M. S. M. A. **Concepção, desenvolvimeno e análise de um sistema de gerência de segurança para redes de telecomunicações**. Florianópolis, 2000. Tese (Doutorado em Ciência da Computação) – Computação, UFSC.
- NUNES FILHO, F. B. Otimização do mwh recuperado através da inspeção técnica. In: SEMINÁRIO NACIONAL DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA, 14., 2000, Foz do Iguaçu. **Anais...** Paraná: SENDI, 2000.
- PANDYA, A. S.; MACY, R. B. **Pattern recognition with neural networks in c++**. Florida: CRC Press, 1995.
- PAVEI, N. Gatos fraudam SC em R\$ 4 milhões. **Jornal Diário Catarinense**, Florianópolis, 17 jun. 2001. Caderno de economia., p. 18.
- PYLE, D. **Data preparation for data mining**. San Diego: Academic Press, 1999. 540 p.
- ROCHA, A. A. F. et al. Recuperação de perdas comerciais de energia elétrica ativa. In: SEMINÁRIO NACIONAL DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA, 13., 1997, São Paulo. **Anais...** São Paulo: SENDI, 1997.

SHAFER, J. C. et al. **SPRINT**: a scalable parallel classifier for data mining. In: Proc. International Conference on Very Large Databases. 1996.
Disponível em: <<http://CiteSeer.nj.nec.com>> Acesso em: 02 set. 2002.

SKAPURA, D. M. **Building neural networks**. New York: ACM Press, 1996.

SFREDO, M.. **Jornal Zero Hora**, Rio Grande do Sul, 05 jun. 2001. Caderno de Economia.

STOLFO, S. J. et al. **Agent-base fraud and intrusion detection in financial information systems**. Disponível em: <<http://CiteSeer.nj.nec.com>> Acesso em: 2 set 2002.

TODESCO, José L. **Reconhecimento de padrões usando rede neuronal artificial com uma função de base radial**: uma aplicação na classificação de cromossomos humanos. 1995. 112 páginas. Tese (Doutorado em Engenharia de Produção) - Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina, Florianópolis, 1995.

WITTEN, I. H.; FRANK, E. **Data mining**: practical machine learning tools and techniques with java implementations. California: Morgan Kaufmann, 2000. P-1-116.

WHITLEY, D. **A genetic algorithm tutorial**. Technical Report, 1993.

ZHANG, T. et al. **BIRCH**: na efficient data clustering method for very large database. Disponível em: <<http://CiteSeer.nj.nec.com>> Acesso em: 01 nov. 2002.