



XVIII Seminário Nacional de Distribuição de Energia Elétrica

SENDI 2008 - 06 a 10 de outubro

Olinda - Pernambuco - Brasil

Identificação de Irregularidades em Sistemas de Baixa Tensão

Karla Figueiredo	Gustavo Chavez	Marley Vellasco
Departamento de Engenharia Eletrônica/Telecomunicações, UERJ	Inteligência Computacional Aplicada/Departamento de Engenharia Elétrica, PUC-Rio	Inteligência Computacional Aplicada/Departamento de Engenharia Elétrica, PUC-Rio
karlaf@uerj.br	gchavez@rdc.puc-rio.br	marley@ele.puc-rio.br
Cyro Assis	Ricardo Levi Menezes Campos	Marco Aurélio C. Pacheco
Inteligência Computacional Aplicada/Departamento de Engenharia Elétrica, PUC-Rio	Light Serviços de Eletricidade S.A	Inteligência Computacional Aplicada/Departamento de Engenharia Elétrica, PUC-Rio
cyro@ele.puc-rio.br	ricardo.levi@light.com.br	marco@ele.puc-rio.br

Palavras-chave

Fraude
Irregularidade
Recuperação de Perdas
Redes Neuais
Residencial

Resumo

O volume de perdas de energia elétrica que vem sofrendo as concessionárias no país por questões de furto, erros de medida e demais tipos de irregularidades, tem-se intensificado nos últimos anos. Por este motivo, existe uma grande preocupação das empresas distribuidoras em identificar o perfil de clientes irregulares com o intuito de reduzir este volume de perdas. Este artigo apresenta o desenvolvimento de um modelo, baseado em Redes Neurais, que tem por objetivo aumentar o grau de acertos na identificação de irregularidades em clientes de baixa tensão. O modelo utilizou dados da empresa Light S.A. e é composto por um comitê de 5 redes neurais artificiais. Para definição do modelo das redes neurais foi necessário pesquisar os atributos disponíveis na base de dados da empresa e identificar a topologia para as redes neurais que formam este comitê. Todos os modelos aqui apresentados foram desenvolvidos com o toolbox de redes neurais do programa Matlab. Os dados de clientes da Light utilizados nos testes dos modelos foram inicialmente pré-processados e normalizados, com a intenção de aumentar a precisão do método. Os resultados obtidos mostram significativo aumento no percentual de clientes irregulares indicados corretamente como suspeitos pelo comitê de redes neurais proposto.

1. Introdução

Atualmente, um dos maiores problemas das empresas brasileiras distribuidoras de energia elétrica é o de perdas comerciais, responsáveis pela maior parte das perdas do setor. A Light, por exemplo, é a terceira distribuidora com maiores perdas comerciais no Brasil, com 3,79 milhões de clientes de baixa tensão em 31 municípios do Estado do Rio de Janeiro. As perdas de energia elétrica por irregularidades têm impacto direto na tarifa paga pelos consumidores. Nos cálculos da agência, as contas de luz são até 17% mais caras, em alguns casos, unicamente em função do roubo e das fraudes de energia. Por isso é importante uma ação em conjunto de todos os setores envolvidos para tentar reduzir essas perdas (Junior, 2005).

As distribuidoras geralmente empregam um conjunto de metodologias heurísticas para identificar os clientes de baixa tensão suspeitos de cometerem algum tipo de irregularidade. Todavia, a média de acertos dessas metodologias ainda é bastante inferior ao desejado, acarretando prejuízos elevados para as distribuidoras brasileiras. No caso específico da Light, a média de acerto na comprovação de clientes fraudadores é de apenas 25%. Verifica-se, portanto, que o processo adotado não é eficiente.

Diversos trabalhos registraram as dificuldades deste tipo de investigação, como, por exemplo, a dificuldade em se avaliar os registros tidos como normais, os quais podem estar contaminados por irregularidades de medição que não foram detectadas durante a inspeção (Rauber et al., 2005). Em (Cabral et al. 2004) foram usados Rough Sets para o descobrimento de irregularidades de medição. Os trabalhos de (Rong et al. 2002) e (Rauber et al., 2005) abordam de forma semelhante o problema de classificação de irregularidades. Nestes dois trabalhos, foram usados somente consumos históricos. Como os consumos apresentam natureza temporal, nestes trabalhos foi necessário a aplicação de métodos de análise de séries temporais (Pollock, 1999) para extrair novas características invariantes.

Este artigo descreve o desenvolvimento de uma metodologia cujo objetivo é aumentar o número de acertos na identificação de clientes de baixa tensão que são irregulares, a partir de um conjunto de suspeitos selecionados. Neste trabalho foram usados outros atributos além do consumo, como temperatura, informações de irregularidade apontada por leituristas e consumo. Também foram consideradas diversas métricas heurísticas inspiradas no atual procedimento de detecção de irregularidade utilizado pela Light.

Tal metodologia é baseada em um comitê de 5 redes neurais artificiais (Duin, 2000) (Lima, 2004) (Santos, 2001). Este comitê realiza uma votação: se pelo menos 3 redes do comitê identificarem um cliente como suspeito, este será classificado como suspeito de ser irregular pelo comitê. O número de redes que indicaram a irregularidade permite avaliar a probabilidade daquele cliente ser efetivamente irregular. Assim, o comitê fornece como resultado uma lista de clientes classificados como suspeitos de serem irregulares, ordenados de forma decrescente pelas suas probabilidades de irregularidade. Dessa forma, os clientes suspeitos com maior probabilidade de possuir alguma irregularidade terão maior prioridade na lista de clientes a serem inspecionados.

O restante deste trabalho está organizado em mais 4 seções: a seção 2 deste artigo descreve a metodologia atual usada pela Light para detecção de irregularidades de consumo; a seção 3 descreve a metodologia proposta; a seção 4 descreve os casos de estudo considerados; e, por último, na seção 5 são apresentadas as conclusões do trabalho.

2. Desenvolvimento

A empresa Light utiliza atualmente um conjunto de quatro metodologias de análise de consumo para selecionar clientes de baixa tensão que apresentem indícios de irregularidades, a saber: Variação Trimestral, Variação Anual, Fator de Ajuste e Fator de Tendência.

As metodologias chamadas de Variação Trimestral e Variação Anual consistem, respectivamente, na comparação dos consumos dos 3 últimos meses e dos 24 últimos meses do cliente. A metodologia

chamada Fator de Ajuste trata da comparação do consumo do último mês com o consumo do mesmo mês do ano anterior do cliente. A última metodologia, Fator de Tendência, consiste na comparação do consumo do último mês com o consumo esperado do mês (calculado pelo fator de tendência) do cliente. Em todos estes casos, se a variação de consumo for negativa, isto é, os casos em que o consumo foi inferior ao consumo esperado, a instalação (o cliente) é indicada para uma verificação de campo. O cálculo do consumo esperado (CE) de um dado mês é dado por (1):

$$CE = C_b \times FT \quad (1)$$

onde C_b é o consumo do mês de um ano base e FT é o fator de tendência. O fator de tendência é calculado da seguinte forma: compara-se o consumo do mês no ano atual (C_a) com o consumo do mesmo mês no ano base (C_b), segmentado por classe e por região, conforme equação (2):

$$FT = \frac{\sum C_a}{\sum C_b} \quad (2)$$

Por exemplo, considere 3 clientes de uma mesma classe em uma mesma região:

Cliente	C_b	C_a	CE
1	100	200	100
2	200	200	200
3	300	200	300
Σ	600	600	

O fator de tendência é dado por (3):

$$FT = \frac{600}{600} = 1 \quad (3)$$

Desta forma, o consumo esperado dos três clientes, (CE_1), (CE_2) e (CE_3), são:

$$\begin{aligned} CE_1 &= 100 \times 1 = 100. \\ CE_2 &= 200 \times 1 = 200. \\ CE_3 &= 300 \times 1 = 300. \end{aligned} \quad (4)$$

Como o consumo do mês atual (C_a) do terceiro cliente é menor do que o seu consumo esperado (CE_3), ele é identificado como irregular e está passível de verificação de campo.

Todos os clientes que forem capturados por este conjunto de metodologias compõem uma base de dados da Light de clientes de baixa tensão suspeitos de irregularidades de consumo. Os clientes desta base de dados ainda passam por um especialista, que através de filtros construídos a partir de suas experiências e conhecimentos, seleciona um número específico de clientes a serem inspecionados por peritos da empresa.

Utilizando este conjunto de metodologias, a Light tem atingido um valor preditivo positivo (VPP) médio de 25%. O VPP é a proporção de clientes comprovadamente irregulares entre todos os clientes que foram classificados como suspeitos de estarem cometendo alguma irregularidade. Isto é, expressa o percentual de clientes irregulares no conjunto de clientes suspeitos.

A Tabela 1 apresenta uma matriz de confusão típica para o caso de um problema de duas classes (por exemplo, clientes normais e irregulares). Deste modo, de acordo com a Tabela 1, o valor preditivo positivo é dado por (5):

$$VPP = \frac{a}{(a+b)} \quad (5)$$

Tabela 1. Clientes Suspeitos x Clientes Inspeccionados

		Clientes Inspeccionados	
		Irregular	Normal
Clientes Suspeitos	Irregular	a	b
	Normal	c	d

3. Metodologia Proposta

A metodologia proposta neste trabalho consiste na formação de um comitê definido por 5 redes neurais artificiais. O objetivo de se construir um comitê é melhorar a generalização a partir da combinação dos resultados individuais de cada Rede Neural.

Um cliente é classificado como suspeito de ser irregular pelo comitê se pelo menos 3 redes o identificarem como suspeito. Através desta metodologia de comitê, além de identificar clientes suspeitos de serem irregulares, também é possível obter, através da votação de cada rede neural, a probabilidade deles serem irregulares. O cálculo da probabilidade é baseado no número de redes que votam. Isto é, se todas as 5 redes do comitê classificarem um cliente como irregular, este cliente é classificado como 100% de probabilidade de ser irregular. Já um cliente em que apenas 3 redes em 5 o classificarem como irregular, este cliente é classificado com 60% de chance de ser irregular. Por este motivo, a lista fornecida pelo comitê contendo clientes suspeitos de serem irregulares é ordenada, de forma decrescente, pelas probabilidades de irregularidades destes clientes suspeitos, ou seja, o primeiro cliente da lista tem uma probabilidade maior de ser realmente irregular do que o último cliente desta lista. Desta forma, não é necessário que um especialista filtre estes clientes suspeitos até atingir um número específico para a inspeção.

Para se obter estes suspeitos, foram modeladas as redes neurais artificiais que compõem o comitê, sendo que todas as 5 redes deste comitê são do tipo MLP (multi-layer perceptron) (Haykin, 1999)(Bishop, 1995). A escolha da topologia de uma rede neural depende de uma série de fatores, mas o número de atributos de entrada e de saída da rede neural, bem como a quantidade de registros na base de dados utilizada no treinamento, fornece uma boa diretriz para esta escolha.

A subseção 3.1 apresenta o pré-processamento da base de dados e a subseção 3.2 apresenta o modelo proposto para as redes neurais dos comitês.

3.1. Pré-processamento dos dados

A figura 1 mostra um diagrama de blocos com as etapas da metodologia proposta. A etapa de pré-processamento é composta pelas seguintes fases: levantamento de dados, pré-seleção de dados, limpeza de dados e normalização.

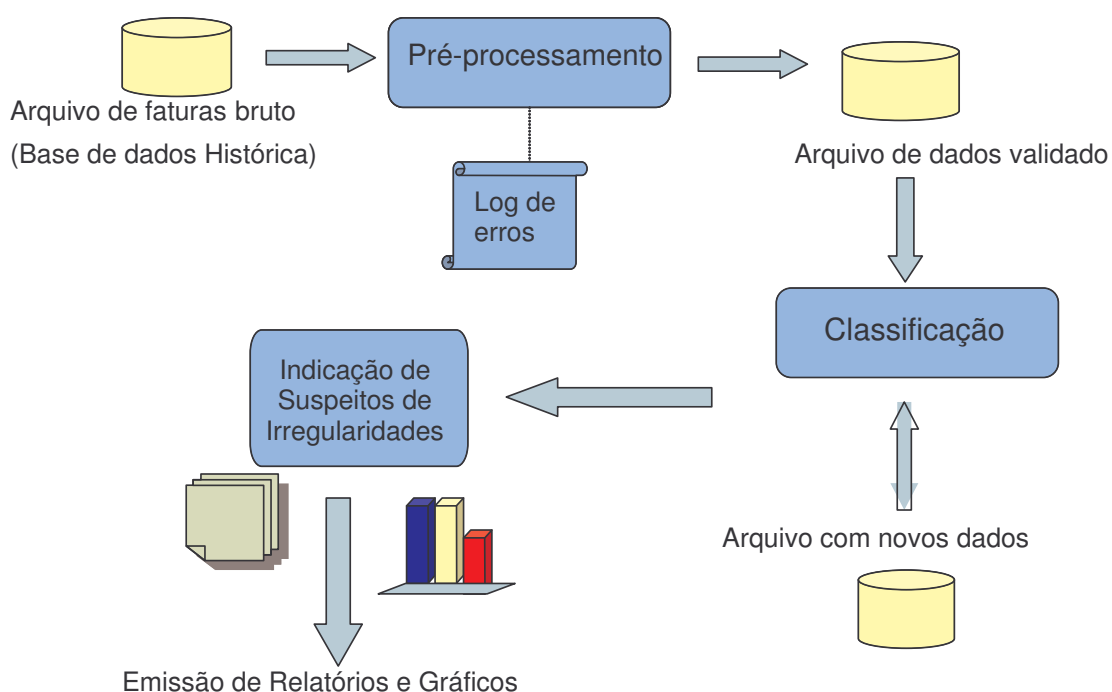


Figura 1 - Diagrama de Blocos com a descrição da metodologia proposta

Os dados utilizados neste trabalho foram fornecidos pela empresa Light S.A.. A Light possui, em sua área de concessão, mais de três milhões de unidades consumidoras. Ela divide sua área de concessão em cinco Superintendências Regionais: Litorânea, Leste, Oeste, Baixada e Interior (Figura 2). A regional Leste por sua vez é dividida em onze sub-regiões denominadas ZEIs (Zonas Elementares de Intervenção). Na Tabela 1 estão dispostos os códigos de cada uma das onze ZEIs da regional Leste e sua localização. Também foi colocada na Tabela 1 a quantidade de unidades consumidoras inspecionadas em cada ZEI, pois esses registros de unidades consumidoras inspecionadas contituiram a base de aprendizado e teste do modelo desenvolvido.

Tabela 1 - ZEIs da Regional Leste e suas características

Código da ZEI	Nome da Localidade	Unidades Inspeccionadas
535	Méier	15655
538	Pilares	27268
540	Madureira	15963
544	Guadalupe	36049
545	Irajá	39765
550	Olaria	19219
551	Maré	4700
552	Ramos	29644
555	Penha	38
565	Ilha do Governador	247
566	Paquetá	5720

Para este trabalho foi considerado apenas dos clientes da regional Leste, que são um total de 816.078 unidades consumidoras. Assim, durante o levantamento de dados a empresa forneceu todos os registros citados acima (clientes de baixa tensão que foram inspeccionados) com os atributos de solicitados.

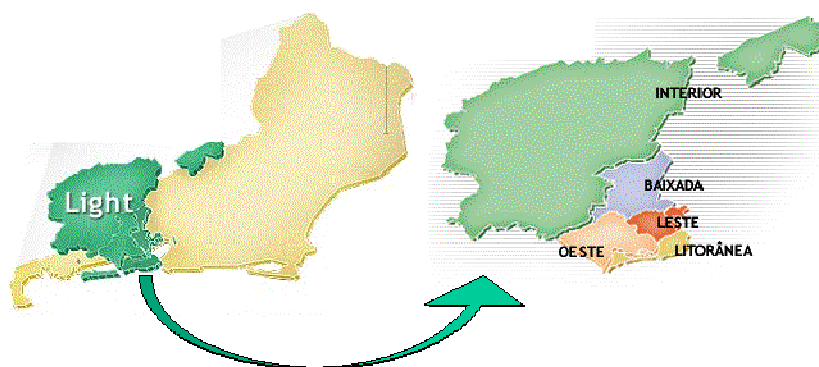


Figura 2- Área de Concessão da Light

Na fase de pré-seleção, após avaliação de todos os atributos fornecidos (mais de 500 atributos), foram selecionados todos os atributos que tinham relação com o consumo dos clientes. Foram descartados aqueles que não tinham relação direta com o perfil de consumo. Após esta pré-seleção, realizou-se uma análise estatística destes atributos (Dash e Liu, 1997) (Mardia, 1997) com o objetivo de aumentar o conhecimento sobre o conteúdo da base. Nesta fase, o atributo categórico bairro foi eliminado por apresentar um número muito grande de bairros, com baixa frequência na base de dados. A Tabela 2 apresenta os atributos selecionados nesta fase.

Tabela 2 - Atributos Selecionados das Bases de Dados

	Atributo	Descrição
1	Local	Código de identificação da unidade consumidora
2	Data de Inspeção	Data da realização da Inspeção
3	Mês de Referência	Mês utilizado como referência. Mês anterior ao da realização da inspeção. EXEMPLO inspeção: 05/07/2006 → mês referência: junho de 2006
4	ZEI	Localização geográfica da unidade consumidora.
5	Tipo de Cliente	Tipo do cliente (Residencial, Comercial, Industrial,...)
6	Equipamento Vital	Indica se há equipamento vital na unidade consumidora
7	Medidor	Identificador de modelo/fabricante do medidor
8	MonoBiTri	Indica se a instalação é mono, bi ou trifásica
9	Tipo de Medição	Indica se a medição feita no local é direta ou indireta
10	Origem	Motivo principal que ocasionou a inspeção.
11	Código de Irregularidade de Consumo	Última irregularidade de consumo detectada
12	Código de Irregularidade de Leitura	Última irregularidade constatada pelo leiturista
13	Qualidade Pagadora	Quantificação do atraso no pagamento de contas
14	Quantidade de Cortes	Número de cortes de fornecimento efetuados no mês, nesta unidade consumidora
15	Consumo no mês	Consumo da unidade consumidora no mês de referência
16	Consumo no ano anterior	Consumo registrado um ano antes do mês de referência. EXEMPLO - mês referência: junho de 2006 → ano anterior: junho de 2005.
17	Consumo no ano base	Consumo registrado no ano base (2001), no mesmo mês de referência. EXEMPLO – mês referência: junho de 2006 → ano base: junho de 2001
18	Média 3 meses	Média de consumo dos últimos 3 meses. Cálculo
19	Média 6 meses	Média de consumo dos últimos 6 meses.
20	Média 12 meses	Média de consumo dos últimos 12 meses.
21	Média dos meses 13 a 24	Média dos Consumos de 12 meses contando a partir do mês do ano anterior ao de referência.
22	Indicador Trimestral 1	indica o menor valor entre as seguintes variações: variação do consumo do último mês para o penúltimo mês e variação do consumo do penúltimo mês para o antepenúltimo mês
23	Indicador Trimestral 2	indica o menor valor entre as seguintes variações: variação do consumo do último mês para o antepenúltimo mês e variação do consumo do penúltimo mês para o antepenúltimo mês.
24	Indicador Trimestral 3	indica o menor valor entre as seguintes variações: variação do consumo do último mês para o penúltimo mês e variação do consumo do último mês para o antepenúltimo mês.
25	Indicador Anual	indica a variação entre as seguintes médias: média dos consumos dos 12 últimos meses e média dos consumos dos 12 meses anteriores aos últimos
26	Indicador de Ajuste	indica a variação entre o consumo do último mês e consumo do mesmo mês do ano anterior.
27	Indicador de Tendência	indica a variação entre o consumo do último mês e consumo esperado para o último mês, calculado pelo fator de tendência
28	Temperatura Mínima	Mínima temperatura registrada no mês
29	Temperatura Máxima	Máxima temperatura registrada no mês
30	Carga	Carga consumida no mês (somatório de todos os clientes Light)
31	Indicador de Fraude	Indica se o inspetor encontrou uma fraude, irregularidade ou normalidade

Os dois primeiros atributos foram utilizados apenas para identificação de cada registro, não servindo como entrada para o sistema. O atributo Mês de Referência serve como referência para a extração do restante dos atributos, e esses atributos foram relativos a este mês. Os atributos ZEI e Tipo de Cliente foram utilizados para dividir o conjunto de dados, de forma que a sensibilidade do sistema aumente. Os nove atributos seguintes detalham aspectos técnicos da instalação e de comportamento do cliente. O código de irregularidade do leiturista é preenchido pelo leiturista no momento da leitura, ou seja, é uma indicação baseada em avaliação informal e superficial. O código de irregularidade de consumo é gerado a partir de verificação dos consumos passados do cliente (esta irregularidade é gerada a partir de análise feita no sistema). Esses atributos foram codificados de forma que, se em um cliente ocorreu algum tipo de irregularidade, o seu atributo apresentará o valor 1; caso contrário, conterá o valor 0. Os três atributos: Consumo no mês, Consumo no ano anterior e Consumo no ano base detalham o perfil de consumo do cliente. Os próximos dez atributos não foram extraídos da base diretamente, foram calculados utilizando outros atributos presentes na base. Além da informação do consumo do mês anterior ao mês que houve a inspeção, também foram calculadas 4 médias móveis: consumo médio trimestral, semestral e anual e anual do do ano anterior ao de referência de cada cliente, a partir do mês que houve inspeção. Também foram criados atributos a partir de indicadores com base nas metodologias utilizadas pela Light (apresentadas na seção 2). Por fim, acrescentou-se ao conjunto de atributos as temperaturas ambientais mínimas e máximas do mês, referentes à região geográfica dos clientes.

O atributo Indicador de Fraude indica o resultado da inspeção realizada, indicando se no cliente foi encontrada uma irregularidade ou não. Este atributo foi utilizado como a saída (alvo) para o treinamento das redes neurais do comitê.

Através do atributo Tipo de Cliente, verificou-se que a grande maioria dos clientes de baixa tensão é do tipo residencial, mas também existem clientes não residenciais (comerciais, industriais, etc.). Geralmente, o comportamento de clientes residenciais e não residenciais são distintos: enquanto os residenciais apresentam picos de tensão no período noturno, os não residenciais apresentam picos no período diurno. Por estes motivos, decidiu-se separar estes dois tipos de clientes – residenciais e não residenciais – e criar um comitê de redes neurais específico para cada um. Portanto, este atributo foi utilizado apenas para separar as duas bases de dados - uma de clientes residenciais e outra de clientes não residenciais, e não foi utilizado como entrada das redes neurais dos comitês. Este trabalho aborda somente o caso de clientes residenciais.

Desta forma, a base de dados utilizada no treinamento das redes neurais do comitê é composta pelos atributos destacados em cinza na Tabela 2, todos descritos acima, em um total de 25 atributos de entrada e uma variável de saída (indicação de irregularidade).

Todos os dados foram filtrados (foram retirados dados espúrios, redundantes e incompletos) e em seguida normalizados e codificados (Dash e Liu, 1997). Através da normalização, garante-se que todos os atributos apresentem o mesmo intervalo de variação, o que melhora o desempenho das redes neurais artificiais.

3.2. Modelo de Redes Neurais

As redes neurais que compõem o comitê foram desenvolvidos no toolbox de redes neurais do programa Matlab. Todas as redes são do tipo MLP (multi-layer perceptron) com 15 entradas, uma camada escondida com um número específico de neurônios (definido nos testes) e com uma camada de saída com 2 neurônios.

A Figura 3 apresenta o modelo proposto, que usa 25 variáveis de entradas e 1 neurônio na camada de saída. De acordo com este modelo, a variável de saída representa 2 classes: clientes suspeitos de irregularidade (medição ou técnica) e a de clientes sem irregularidades. Assim, os clientes suspeitos de

apresentar algum tipo de irregularidade são representados pelo código 1 e os clientes não suspeitos são representados pelo código 0.

O treinamento das redes neurais foi realizado com o algoritmo Back Propagation com Levenberg-Marquardt (Haykin, 1999). O treinamento das redes neurais com este algoritmo é da ordem de 10 a 100 vezes mais rápido que o Back Propagation original, mas é restrito quanto ao tamanho da rede neural (que deve conter uma quantidade moderada de pesos) e ao tamanho da base de dados, que não pode ser muito extensa.

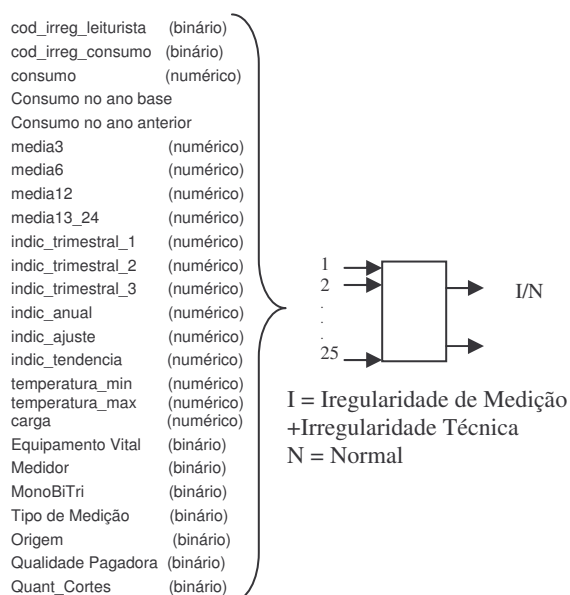


Figura 3. Topologia de Rede Neural MLP com 1 saída

4. Resultados Experimentais

Após o processamento dos dados selecionados (vide subseção 3.1) obteve-se três bases de dados. A Tabela 3 e 4 identificam as 3 bases (clientes com irregularidade de medição, clientes com irregularidade técnica e clientes normais), mostrando o número de clientes por base para a fase de aprendizado e teste, respectivamente.

Tabela 3 - Base Treinamento/Validação para Clientes Residenciais - ZEIS 540/544 da Regional Leste

Base Treinamento	Irregularidade de Medição	Irregularidade Técnica	Normal
540	4630	3182	28414
544	6698	7330	24946

Tabela 4 - Base Teste para Clientes Residenciais das ZEIS 540/544 da Regional Leste

Base Treinamento	Irregularidade de Medição	Irregularidade Técnica	Normal
540	776	272	1893
544	1096	435	3132

Os dados relativos ao Treinamento/Validação são do período entre março de 2002 e dezembro de 2005 e a base de testes é de clientes investigados nos meses de fevereiro/março e julho/agosto de 2006.

Como existe uma grande discrepância entre o número de registros de clientes com irregularidade de medição e o número de registros de clientes normais na base de Treinamento/Validação, decidiu-se criar amostras de dados das bases maiores com a mesma quantidade de registros da menor base (no caso, a de clientes com irregularidade de medição para o caso da ZEI 544 e clientes com irregularidade técnica para o caso 544), para não deixar o treinamento das redes neurais do comitê tendencioso. Assim, para a ZEI 540 utilizou-se uma base de dados com os 3182 registros de clientes com irregularidade de técnica, 3182 registros, escolhidos aleatoriamente, de clientes com irregularidade de medição, e 6364 amostras (também escolhidas aleatoriamente) de clientes normais e para a ZEI 544 foram: 6698 registros de clientes com irregularidade de medição, 6698 registros, escolhidos aleatoriamente, de clientes com irregularidade técnica, e 13396 amostras (também escolhidas aleatoriamente) de clientes normais, para que as duas classes de saída tenham o mesmo número de registros. A base de treinamento foi composta com 75% das amostras para treinamento e o restante da base para a validação.

Foram criadas 5 bases de dados diferentes, cada uma dando origem a cada uma das cinco redes que compõem o comitê sem o uso de *bootstrap*.

Para cada base de dados criada, foram também vários experimentos a fim de determinar o número de neurônios da camada escondida das redes, para cada estrutura avaliada (número de neurônios na camada escondida) realizou-se 30 treinamentos e selecionou-se, para compor o comitê, a rede que obteve o melhor desempenho na fase de validação. Durante a etapa de treinamento, o erro do conjunto de validação é monitorado para garantir a capacidade de generalização da rede. Este é o método *early stopping* que determina um ponto de parada no treinamento, tentando encontrar um ponto ótimo de generalização. Quando o erro do conjunto de validação começa a aumentar o processo de treinamento é interrompido. Assim, pode-se selecionar a melhor rede para a melhor configuração de neurônios na camada escondida.

Como o objetivo deste trabalho é aumentar o VPP de clientes com irregularidade, foram feitos dois experimentos: um considerando o VPP como métrica de erro na validação do treinamento das redes neurais dos comitês; e outro considerando o erro de classificação geral dos dados. Observou-se que a métrica de erro de classificação obteve melhores resultados, pois ao minimizar o erro de classificação de ambas as classes, indiretamente ele também maximiza o VPP na identificação de clientes com irregularidade de medição. Por este motivo, nos resultados experimentais seguintes, utilizou-se a métrica de erro de classificação.

4.1. Resultados do Teste com a Base de Clientes Residenciais

Segundo os resultados obtidos nos experimentos, a melhor topologia para as redes deste modelo, para o comitê de clientes residenciais, é a que contém 8 neurônios na camada escondida.

A Tabela 5 apresenta os menores erros de classificação encontrados para cada rede neural do comitê para treinamento /validação para as ZEIs 540 e 544.

Após o treinamento das redes neurais do comitê, foi realizado um teste com uma base composta por clientes investigados nos meses de fevereiro, março e abril de 2006. Assim, a base teste relativa a a ZEI 540 possui 2941 registros de clientes residenciais, sendo da classe de clientes com irregularidade são 1048 clientes e da classe normal são 1893. A ZEI 544 possui 4663 registros de clientes residenciais, sendo 1531 da classe de clientes com irregularidade de e 3132 da classe de clientes normais.

As Tabelas 6 e 7 ilustram as matrizes de confusão de cada uma das cinco redes que compõem os comitês para o teste realizado respectivamente para as ZEIs 540 e 544.

Tabela 5. Menor Erro (%) de Classificação das redes neurais para o comitê de clientes residenciais

	Rede	Treinamento	Validação
ZEI 540	1	30,86	34,73
	2	30,21	31,41
	3	29,79	32,42
	4	28,48	31,41
	5	30,76	30,75
ZEI 544	1	30,11	35,21
	2	34,00	37,78
	3	33,40	36,82
	4	35,00	38,99
	5	34,10	36,50

Tabela 6. Matriz de confusão das Redes que compõem o comitê para a ZEI 540

		Clientes Inspeccionados (Target)		VPP SIIPERCOM	Sensibilidade	Erro Classificação	VPP Light	
		Irregular(1)	Normal(0)					
Rede 1	Comitê de Classificação	Irregular(1)	705	842	45,57	67,27	40,29	35,63
		Normal (0)	343	1051				
Rede 2	Comitê de Classificação	Irregular(1)	770	1051	42,28	73,47	45,19	35,63
		Normal (0)	278	842				
Rede 3	Comitê de Classificação	Irregular(1)	705	984	41,74	67,27	45,12	35,63
		Normal (0)	343	909				
Rede 4	Comitê de Classificação	Irregular(1)	689	868	44,25	65,74	41,72	35,63
		Normal (0)	359	1025				
Rede 5	Comitê de Classificação	Irregular(1)	635	847	42,85	60,59	42,84	35,63
		Normal (0)	413	1046				

Tabela 7. Matriz de confusão das Redes que compõem o comitê para a ZEI 544

		Clientes Inspeccionados (Target)		VPP SIIPERCOM	Sensibilidade	Erro Classificação	VPP Light	
		Irregular(1)	Normal(0)					
Rede 1	Comitê de Classificação	Irregular(1)	1006	1584	38,84	65,71	45,23	32,83
		Normal (0)	525	1548				
Rede 2	Comitê de Classificação	Irregular(1)	1057	1620	39,48	69,04	44,91	32,83
		Normal (0)	474	1512				
Rede 3	Comitê de Classificação	Irregular(1)	1141	1892	37,62	74,53	48,94	32,83
		Normal (0)	390	1240				
Rede 4	Comitê de Classificação	Irregular(1)	1091	1621	40,23	71,26	44,20	32,83
		Normal (0)	440	1511				
Rede 5	Comitê de Classificação	Irregular(1)	1015	1564	39,36	66,30	44,61	32,83
		Normal (0)	516	1568				

Na Tabela 8 pode-se observar a matriz de confusão do resultado obtido com o comitê de classificação. A votação é contabilizada (por rede neural) para cada cliente de teste da seguinte forma: os clientes

que receberam 5 votos são classificados como irregulares. Os que receberam 4 ou menos votos são classificados como normais. O VPP da Light para a ZEI 540 nesse teste foi de 35,6 % e o VPP do comitê foi de 54,6 %. No caso da ZEI 544 o VPP da Light foi de 32,8 % e o obtido pelo comitê foi de 49,3 %. Outra métrica utilizada para avaliar o desempenho do modelo proposto foi a sensibilidade (equação 6). Esta métrica é definida como o número de clientes classificados pela rede (ou pelo comitê) como irregulares e o número total de irregularidades comprovadas. Segundo a tabela 1, a sensibilidade é expressa por:

$$\text{Sensibilidade} = \frac{a}{(a + c)} \quad (6)$$

Tabela 8. Resultados obtidos pelo Comitê de redes do comitê de clientes residenciais

Matriz de Confusão				Métricas	
ZEI 540	Clientes Inspeccionados			VPP _{RN} : 54,6% Sensibilidade: 36,2% Erro Classific.:33,5% VPP _{Light} = 35,6 %	
Saída do Comitê			Irregular		Normal
	Irregular	379	315		
	Normal	669	1578		
ZEI 544	Clientes Inspeccionados			VPP _{RN} : 49,3% Sensibilidade: 25,9% Erro Classific.:33,1% VPP _{Light} = 32,8 %	
Saída do Comitê			Irregular		Normal
	Irregular	397	408		
	Normal	1134	2724		

Também se pode observar que esta metodologia de classificação (votação unânime considerada pelo comitê), resultou em uma quantidade maior de clientes irregulares classificados (pelo comitê) como normais (669 para o caso da ZEI 540 e 1134 para a ZEI 544), quando esta é comparada aos resultados individuais de cada rede do comitê (Tabelas 6 e 7), reduzindo um pouco a sensibilidade. No entanto, como o objetivo principal deste trabalho era o aumento do VPP, pode-se dizer que a constituição de um comitê de redes neurais atendeu a este objetivo, uma vez que o VPP do comitê de redes neurais, para ambas as ZEI, apresentou valor superior aos individuais obtidos pelas redes neurais de cada comitê.

5. Conclusões

O assunto de que trata este artigo é de grande importância estratégica para as empresas distribuidoras de energia elétrica em todo o Brasil, uma vez que as perdas comerciais de energia elétrica por irregularidades representam um enorme prejuízo.

Este artigo apresentou o desenvolvimento de uma metodologia baseada em redes neurais. O modelo utilizado é formado por um comitê de 5 redes neurais, onde cada rede possui uma saída para classificar os clientes em duas categorias: classe com irregularidade (técnica e de medição) e classe de clientes normais. O modelo avaliado alcançou bom desempenho para clientes de baixa tensão do tipo residencial. Ressaltamos a relevância do trabalho na identificação dos atributos relativos às médias móveis, indicadores trimestrais, anual, tendência e ajuste, além da temperatura média, que introduz a sazonalidade do consumo.

Segundo os resultados dos testes utilizando-se o modelo comitê para os clientes do tipo residencial, observa-se que o modelo atinge um VPP (relação entre o número de casos em que se detecta a irregularidade de medição dividido pelo número de casos indicados para inspeção) superior a 54% para a ZEI 540 e 49 % para a ZEI 544, valores superiores aos normalmente obtidos pela Light usando

os métodos apresentados na seção 2. Também fica comprovada a eficácia no uso de comitês de redes neurais em casos como estes em detrimento ao uso de redes neurais individuais.

Dessa forma, pode-se concluir que os resultados apresentados mostram-se muito promissores no problema de identificação de irregularidades em baixa tensão para todos os clientes da Light.

Como trabalho futuro pretende-se estender a análise aos clientes não residenciais. Além disso, pretende aplicar métodos de clusterização da base de dados de forma a criar grupos mais coesos sob a ótica do perfil de consumo.

6. Referências Bibliográficas

BISHOP, C.M. *Neural Networks for Pattern Recognition*, Oxford University Press.

CABRAL, J. E., Gontijo, M., Pinto, J. O. P., & Filho, J. R. Fraud detection in electrical energy consumers using rough sets. In 2004 IEEE International Conference on Systems, Man and Cybernetics, volume 4, 2004, pp. 3625–3629.

DASH, M. & LIU, H. Feature Selection for Classification, *Intelligent Data Analysis*, 1(3), 1997. Disponível em: http://www.tnt.uni-hannover.de/~rost/muster_erk/article.html

DUIN, R. P. W. & TAX, D. M. J. Experiments With Classifier Combining Rules. In: Kittler, J.; Roli, F., (Eds). *Multiple Classifier Systems: v. 1897 of Lecture Notes in Computer Science*. Berlin Heidelberg: Springer-Verlag, 2000, pp. 16-29.

HAYKIN S. *Neural Networks - A Comprehensive Foundation*. 2 ed. New Jersey: Prentice-Hall, 1999. ISBN 0-13-273350-1.

JUNIOR, J. E. C. Detecção de Fraudes em Unidades Consumidoras de Energia Elétrica Utilizando Rough Sets, Diss. de Mestrado, UFMS – Fund.Univ. Fed. Mato Grosso do Sul, 2005.

LIMA, C.A.M. Comitê de Máquinas: Uma Abordagem Unificada Empregando Máquinas de Vetores-Suporte. Dissertação de Doutorado, Universidade Estadual de Campinas, 2004.

MARDIA, K.V., Kent, J.T & Bibby J.M. *Multivariate Analysis*. Academic Press Inc. London, 1979.

POLLOCK, D.S.G.A. *Handbook of Time-Series Analysis, Signal Processing and Dynamics*. Academic Press, New York, San Diego Edition, 1999.

RAUBER, T., DRAGO, I., VAREJÃO, F. & QUEIROGA, R. Extração e Seleção de Características na Identificação de Perdas Comerciais na Distribuição de Energia Elétrica, XXV Cong. Soc. Bras. Comp, 2005.

RONG, J., TAGARIS, H., LACHSZ, A., & JEFFREY, M. Wavelet based feature extraction and multiple classifiers for electricity fraud detection. In 2002 Trans. and Distribution Conf. and Exhibition 2002: Asia Pacific. IEEE/PES, vol. 3, 2002, pp. 2251–2256.

SANTOS, R. O. V. Combinação de Redes Neurais MLP em Problemas de Classificação. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro, 2001.