

IDENTIFICAÇÃO DE IRREGULARIDADES NO CONSUMO DE ENERGIA ELÉTRICA EM BAIXA TENSÃO

KARLA FIGUEIREDO¹, JUAN G. LAZO LAZO², MARLEY VELLASCO², MARCO PACHECO², LUCIANA CAMPOS²

¹UERJ: Depto. Eng. Eletrônica/Telecomunicações, Rua São Francisco Xavier, 524 - Rio de Janeiro, 20550-900 RJ Brasil

²ICA: Inteligência Computacional Aplicada/DEE-PUC-Rio, Rua Marquês de S. Vicente 225, Gávea, Rio de Janeiro, CEP 22453-900, RJ Brasil

karlaf@uerj.br, juan@ele.puc-rio.br, marley@ele.puc-rio.br, marco@ele.puc-rio.br, hall@ele.puc-rio.br, lcampos@ele.puc-rio.br

Abstract— The volume of energy loss that electric utilities have been dealing with in Brazil, due to theft, measurement mistakes and other types of irregularities, has been ever increasing. This is the greatest concern of the Electrical Energy companies, so they are trying to identify irregular clients in order to reduce this loss. This article presents the development of a Neural Network model that tries to enhance the identification of irregularities in low tension clients. This model uses data from the company Light S.A. and is composed of a committee of five artificial neural networks. In order to define the neural networks model it was necessary to analyse all attributes available in the company's database and to identify the topology for each neural network that forms the committee. Every model presented here has been developed with the Matlab neural networks toolbox. The clients' data that were used in the tests were initially pre-processed and normalized in order to increase the method's accuracy. The results obtained show a significant increase in the percentage of actually irregular clients within the suspicious group indicated by the networks.

Keywords— irregularities, electrical energy, low tension

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 de energia elétrica 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.

Palavras-chave— irregularidade, energia elétrica, baixa tensão

1. Introdução

As perdas comerciais de energia elétrica por irregularidades representam um prejuízo total para o país da ordem de R\$ 5 bilhões por ano. Essas perdas são de todos os agentes do setor elétrico, como as empresas distribuidoras e os Estados e Municípios, em termos de arrecadação de impostos. Segundo a ANEEL (Agência Nacional de Energia Elétrica), o montante dessas perdas chegam a 15 milhões de MWh por ano. Isso representa cerca de 5% da energia total consumida no país. Grande parte dessas perdas ocorre na rede da empresa Light S.A., que atende a 3,7 milhões de unidades consumidoras em 28 municípios 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].

A concessionária Light possui um conjunto de metodologias para identificar clientes de baixa tensão suspeitos de estarem cometendo algum tipo de irregularidade. Estes clientes, classificados como suspeitos por estas metodologias, são indicados para um especialista da empresa que, com base em experiência, seleciona um determinado conjunto de clientes para serem inspecionados. Através deste processo, a Light tem obtido uma média de 25% de acertos na comprovação de clientes irregulares. 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 [Raubert 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 [Raubert 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 do tipo residencial 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, pelo sistema e pelo 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. Este comitê realiza uma votação: se as 5 redes do comitê identificarem um cliente como suspeito, este será classificado como suspeito de ser irregular pelo comitê, caso contrário será considerado normal.

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. Metodologia atual para detecção de irregularidades de consumo

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:

$$CE_1 = 100 \times 1 = 100.$$

$$CE_2 = 200 \times 1 = 200. \quad (4)$$

$$CE_3 = 300 \times 1 = 300.$$

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

	<i>Irregular</i>	<i>Normal</i>
<i>Irregular</i>	<i>a</i>	<i>b</i>
<i>Normal</i>	<i>c</i>	<i>d</i>

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 5 redes o identificarem como suspeito. Através desta metodologia de comitê, pôde-se garantir o aumento percentual do VPP, quando este é comparado ao resultados obtidos por qualquer das redes que compõem o comitê.

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

Os dados utilizados neste trabalho foram fornecidos pela empresa Light S.A. e são originários de 3 bases de dados:

- Base Pegasus – base de dados principal da empresa que contém os registros e atributos gerais de todos os clientes da Light.
- Base STI – gerada a partir da base Pegasus, contém todos os registros e atributos de clientes de baixa tensão que foram identificados como suspeitos de estarem cometendo algum tipo de irregularidade pelas metodologias apresentadas na seção 2.
- Base IEN – gerada a partir da base STI, contém todos os registros e atributos de clientes de baixa tensão que foram inspeccionados.

Os dados obtidos pertencem a apenas duas subestações. Esta etapa de pré-processamento é composta pelas seguintes fases: pré-seleção de dados, limpeza de dados e normalização. Na fase de pré-seleção, após avaliação de todos os atributos das três bases (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

Base Pegasus	
Atributo	Descrição
*TIPO_CLIENTE	Residencial, Comercial, Industrial e Rural
CODIGO_ANOMALIA_1	Código de irregularidade do leiturista.
CODIGO_ANOMALIA_2	Código de irregularidade do sistema.
CODIGO_ANOMALIA_3	Código de irregularidade de consumo
Base STI	
Atributo	Descrição
Indica_Irregularidade_Cliente	Indica se o cliente já teve irregularidade ou não
Base IEN	
Atributo	Descrição
Consumo	Valor do consumo do local no mês em referência.
Média trimestral	valores mensais definidos por localidade e cliente.
Média semestral	valores mensais definidos por localidade e cliente.
Média anual	valores mensais definidos por localidade e cliente.

O primeiro atributo proveniente da base de dados Pegasus diz respeito ao tipo do cliente, cadastrado como residencial, comercial, industrial ou rural. Os outros 3 atributos selecionados dessa base (CODIGO_ANOMALIA1, CODIGO_ANOMALIA2 e CODIGO_ANOMALIA3) estão relacionados, respectivamente, ao código de irregularidade do leiturista, do sistema e do consumo. 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 do sistema é um código de irregularidade que é inferido por um sistema desenvolvido pela Light, com o objetivo de detectar inconsistências de preenchimento e valores. O código de irregularidade de consumo é gerado a partir de verificação dos consumos passados do cliente. Os atributos CODIGO_ANOMALIA (1, 2 e 3) 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.

A base de dados IEN contribuiu com informações sobre o consumo faturado de energia do cliente. Além da informação do consumo do mês anterior ao mês que houve a inspeção, também foram calculadas 3 médias móveis: consumo médio trimestral, semestral e anual de cada cliente, a partir do mês que houve inspeção.

O atributo da base de dados STI (INDICA_IRREGULARIDADE_CLIENTE), que indica se no cliente já foi encontrada uma irregularidade ou não, foi utilizado como a saída (alvo) para o treinamento das redes neurais do comitê.

Além destes atributos, foram criados indicadores com base nas metodologias utilizadas pela Light (apresentadas na seção 2):

- 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.
- 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.
- 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.
- 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.
- Indicador de Ajuste: indica a variação entre o consumo do último mês e consumo do mesmo mês do ano anterior.
- Indicador de Tendência: indica a variação entre o consumo do último mês e consumo esperado para o último mês, calculado pelo fator de tendência.

Por fim, acrescentou-se ao conjunto de atributos as temperaturas ambientais mínimas e máximas do mês, referentes à região geográfica dos clientes.

Através do atributo TIPO_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 da Tabela 2 com a exceção do atributo TIPO_CLIENTE, mais os indicadores e as temperaturas ambientais (mínima e máxima), todos descritos acima, em um total de 15 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 [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 1 apresenta o modelo proposto, que usa 15 variáveis de entradas e 2 neurônios na camada de saída. De acordo com este modelo, a variável de saída (INDICA_IRREGULARIDADE_CLIENTE) representa 2 classes: clientes suspeitos de irregularidade (medição ou técnica) e a de clientes sem irregularidades. Apenas um neurônio da camada de saída poderá ser ativado por vez, de forma a representar cada uma das classes. Assim, os clientes suspeitos de apresentar algum tipo de irregularidade são representados pelo código 10 (ativando o primeiro neurônio da camada de saída) e os clientes não suspeitos são representados pelo código 01 (ativando o segundo neurônio da camada de saída).

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.

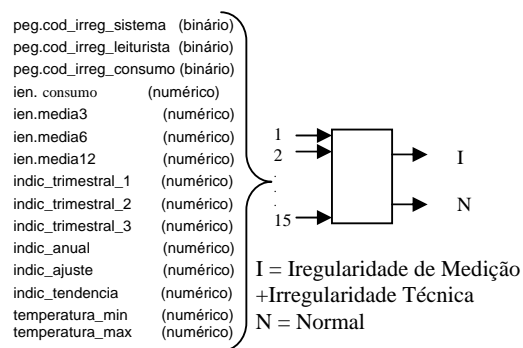


Figura 1. Topologia de Rede Neural MLP com 3 saídas

4. Resultados Experimentais

Após o processamento dos dados selecionados (vide subseção 3.1) obteve-se três bases de dados. A Tabela 3 identifica as 3 bases (clientes com

irregularidade de medição, clientes com irregularidade técnica e clientes normais), mostrando o número de clientes por base.

Tabela 3. Composição da Base de Dados

Clientes	Treinamento/Validação (registros)
irregularidade de medição	1023
irregularidade técnica	2244
normais	4462

Os dados relativos ao Treinamento/Validação são do período entre março de 2002 e dezembro de 2005 e a base de testes são de clientes investigados nos meses de fevereiro, março e abril 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 não deixar o treinamento das redes neurais do comitê tendencioso. Assim, utilizou-se uma base de dados com os 256 registros de clientes com irregularidade de medição, 256 registros, escolhidos aleatoriamente, de clientes com irregularidade técnica, e 512 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 por *bootstrap*.

Para cada base de dados criada, realizou-se 10 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.

Foram também realizados vários experimentos a fim de determinar o número de neurônios da camada escondida das redes.

Como o objetivo deste trabalho é aumentar o VPP de clientes com irregularidade de medição, 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 10 neurônios na camada escondida.

A Tabela 4 apresenta os menores erros de classificação encontrados para cada rede neural do comitê.

Tabela 4. Menor Erro (%) de Classificação das redes neurais para o comitê de clientes residenciais

Rede	Treinamento	Validação
1	24,23	28,51
2	25,60	26,88
3	33,01	34,52
4	28,96	33,58
5	27,28	32,31

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 possui com 1066 registros de clientes residenciais, sendo 482 da classe de clientes com irregularidade de medição, 288 da classe de clientes com irregularidade técnica e 296 da classe de clientes normais.

O erro de classificação do comitê neste teste foi de 25,4% e o VPP do comitê foi de 78%. A Tabela 5 ilustra a matriz de confusão de cada rede deste comitê no teste realizado. O código C1 indica os clientes com irregularidade de medição e técnica e o código C2 os clientes normais.

Outra métrica utilizada para avaliar o desempenho do modelo proposto foi a sensibilidade. 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:

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

A Tabela 6 ilustra a votação do comitê. 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. Os resultados da tabela 6 mostram que, ao se considerar esta metodologia para a classificação do cliente pelo comitê, obteve-se um aumento percentual do VPP, quando este é comparado ao VPP obtido por cada rede individualmente.

Também pode-se observar que esta metodologia de classificação (votação unânime), resultou em uma quantidade maior de clientes irregulares classificados (pelo comitê) como normais, reduzindo um pouco a sensibilidade (quando esta é comparada aos

resultados individuais de cada rede do comitê). 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.

Tabela 5. Resultados obtidos individualmente pelas redes do comitê de clientes residenciais

Rede	Matriz de Confusão				Métricas
1	Saída		Target		VPP: 74,5% Sensibilidade: 98,1% Erro Classific.: 25,6%
	Rede	C1	C1	C2	
			755	258	
		C2	15	38	
2	Saída		Target		VPP: 73,6% Sensibilidade: 96,9% Erro Classific.: 27,4%
	Rede	C1	C1	C2	
			746	268	
		C2	24	28	
3	Saída		Target		VPP: 74,4% Sensibilidade: 98,7% Erro Classific.: 25,4%
	Rede	C1	C1	C2	
			760	261	
		C2	10	35	
4	Saída		Target		VPP: 75,2% Sensibilidade: 96,9% Erro Classific.: 25,3%
	Rede	C1	C1	C2	
			746	246	
		C2	24	50	
5	Saída		Target		VPP: 75,1% Sensibilidade: 96,8% Erro Classific.: 25,5%
	Rede	C1	C1	C2	
			745	247	
		C2	25	49	

Tabela 6. Resultados obtidos com o comitê

Saída		Target		VPP: 78 % Sensibilidade: 90,0% Erro Classific.: 25,4%
Comitê	C1	C1	C2	
		695	196	
	C2	75	100	

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 duas saídas 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 78%, valor superior aos normalmente obtidos pela Light usando os métodos apresentados na seção 2.

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.

Referências Bibliográficas

- Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press.
- Cabral, J. E., Gontijo, E. M., Pinto, J. O. P., and Filho, J. R. (2004). Fraud detection in electrical energy consumers using rough sets. In 2004 IEEE International Conference on Systems, Man and Cybernetics, volume 4, pages 3625–3629.
- Dash, M. and Liu, H. (1997). Feature Selection for Classification, *Intelligent Data Analysis*, 1(3), 1997. Disponível em: http://www.tnt.uni-hannover.de/~rost/muster_erk/article.html
- Haykin S. (1999). *Neural Networks - A Comprehensive Foundation*. 2 ed. New Jersey: Prentice-Hall, 1999. ISBN 0-13-273350-1.
- Junior, J. E. C. (2005) Detecção de Fraudes em Unidades Consumidoras de Energia Elétrica Utilizando Rough Sets, Diss. de Mestrado, UFMS – Fund.Univ. Fed. Mato Grosso do Sul.
- Mardia, K.V; Kent, J.T; Bibby J.M. (1979) *Multivariate Analysis*. Academic Press Inc. London.
- Pollock, D. S. G. (1999). *A Handbook of Time-Series Analysis, Signal Processing and Dynamics*. Academic Press, New York, San Diego Edition.
- Rauber, T.; Drago, I., Varejão, F. e Queiroga, R. (2005) Extração e Seleção de Características na Identificação de Perdas Comerciais na Distribuição de Energia Elétrica, XXV Cong. Soc. Bras. Comp.
- Rong, J., Tagaris, H., Lachsz, A., and Jeffrey, M. (2002). 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, pp. 2251–2256.