

# UMA ABORDAGEM DE CLASSIFICAÇÃO COM APRENDIZADO ATIVO PARA AVALIAÇÃO DE AMOSTRAS DE PAPEL ISOLANTE EMPREGANDO ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO

CLÁUDIA E. DA MATTA<sup>1</sup>; HENRIQUE M. PAIVA<sup>2</sup>; ROBERTO K. H. GALVÃO<sup>2</sup>; MARIA F. PIMENTEL<sup>3</sup>;  
EDMILSON O. DOS SANTOS<sup>3</sup>; WALACE D. FRAGOSO<sup>4</sup>

<sup>1</sup>*Divisão de Ciência da Computação, Instituto Tecnológico de Aeronáutica, São José dos Campos – SP*

<sup>2</sup>*Divisão de Engenharia Eletrônica, Instituto Tecnológico de Aeronáutica, São José dos Campos – SP*

<sup>3</sup>*Departamento de Engenharia Química, Universidade Federal de Pernambuco, Recife – PE*

<sup>4</sup>*Departamento de Química, Universidade Federal da Paraíba, João Pessoa – PB*

*E-mails: claudia\_da\_matta@hotmail.com, hmpaiva@ita.br, kawakami@ita.br, mfp@ufpe.br, edsantos77@gmail.com, wallace.fragoso@gmail.com*

**Abstract**— The degree of polymerization (DP) of the insulating paper employed in a transformer is an important index to assess the remaining useful life of the equipment. However, the reference method for DP determination is time-consuming and requires a large quantity of paper to obtain reliable results. The present work presents an alternative based on the use of near-infrared spectroscopy for classification of insulating paper in three categories (low, medium and high DP), according to the NBR 8148 ABNT standard. It is worth noting that the training of the classifier requires the availability of samples for which the DP has been previously determined. For this reason, an active learning approach is proposed to select, from a given set, the most appropriate samples for use in the training procedure. In this manner, the experimental effort involved in the direct measurement of DP can be reduced. The results obtained for a set of 75 real samples of insulating paper show that the active learning can provide better classification rates, as compared to methods that employ random selection of the training samples.

**Keywords**— Active learning, classification, insulating paper, degree of polymerization, near-infrared spectroscopy.

**Resumo**— O grau de polimerização (GP) do papel isolante usado em um transformador é um importante índice para se avaliar a vida útil remanescente do equipamento. Contudo, o método de referência para determinar o GP é demorado e necessita de grande quantidade de papel para obtenção de resultados confiáveis. Neste trabalho, apresenta-se uma alternativa baseada no uso da espectroscopia no infravermelho próximo para classificação de papel isolante em três categorias (baixo, médio e alto GP), conforme a norma NBR 8148 da ABNT. Vale salientar que o treinamento do classificador requer a disponibilidade de amostras cujo GP tenha sido previamente determinado. Por essa razão, propõe-se uma abordagem de aprendizado ativo para selecionar, de um dado conjunto, as amostras mais apropriadas para uso no treinamento. Com isso, é possível reduzir o esforço experimental envolvido na medida direta do GP. Os resultados obtidos com um conjunto de 75 amostras reais de papel isolante mostram que o aprendizado ativo pode proporcionar maiores taxas de acerto na classificação, em comparação com métodos que fazem uso de escolhas aleatórias das amostras de treinamento.

**Palavras-chave**— Aprendizagem ativa, classificação, papel isolante, grau de polimerização, espectroscopia no infravermelho próximo.

## 1 Introdução

A obtenção de classificadores apropriados requer que os dados empregados para treinamento sejam representativos da população, permitindo definir corretamente a fronteira de separação entre as classes [1]. Contudo, a rotulagem das amostras de treinamento pode ser custosa devido à necessidade de consultar um especialista da área ou realizar ensaios laboratoriais. Em tais casos, justifica-se a utilização de técnicas para seleção das amostras a serem rotuladas. Nesse contexto, uma possível abordagem consiste no uso de Aprendizado Ativo [2].

No Aprendizado Ativo o algoritmo de classificação passa de uma condição passiva, no qual só recebe amostras para treinamento, para uma condição ativa, de controle sobre o aprendizado. Nesse caso, o algoritmo passa a escolher as amostras cuja classe verdadeira deve ser determinada. Desse modo, espera-se reduzir o trabalho associado à classificação

prévia das amostras por um especialista ou método laboratorial de referência.

Uma técnica de aprendizado ativo para treinamento de classificadores neurais foi proposta por [3]. Posteriormente, [4] apresentou um algoritmo que utiliza aprendizado ativo e modelos estatísticos para seleção de amostras. Um algoritmo que integra os benefícios do aprendizado ativo baseado na seleção de amostras por um algoritmo genético foi proposta por [2]. Vale ainda salientar que o algoritmo proposto permitia a utilização tanto das amostras rotuladas quanto das não-rotuladas para treinamento do classificador.

Neste artigo, investiga-se o uso de uma técnica de aprendizado ativo para classificação de amostras de papel isolante de acordo com o grau de polimerização (GP). O monitoramento do GP em transformadores de potência é importante para prevenir a ocorrência de falhas, sendo um procedimento exigido contratualmente por seguradoras [5]. Contudo, o processo para medida direta do GP é caro e laborioso, o que tem motivado o desenvolvimento de técnicas

alternativas, tais como a Espectroscopia no Infravermelho Próximo (*Near Infrared* – NIR), conforme (Santos et al., 2008).

A espectroscopia NIR é um método rápido e de baixo custo, que pode ser implementado em instrumentos portáteis para uso em campo [6]. Contudo, sua aplicação requer a obtenção de modelos que permitam relacionar o espectro NIR da amostra de papel com o seu grau de polimerização. Para isso, é necessário dispor de amostras de treinamento cujo GP tenha sido previamente determinado pelo método convencional.

Na abordagem aqui proposta, supõe-se que espectros NIR tenham sido obtidos para um dado conjunto de amostras. Com base nos espectros, uma técnica de aprendizado ativo é empregada para selecionar quais amostras deveriam ter seu GP medido. Ao final, um classificador é treinado para discriminar as amostras em três classes (GP baixo, médio e alto).

As próximas seções deste artigo encontram-se organizadas da seguinte forma: na seção 2, detalha-se o problema de monitoramento da condição de papéis isolantes e descreve-se como o mesmo será tratado em um contexto de classificação; na seção 3, apresenta-se o algoritmo de classificação empregando aprendizado ativo proposto neste trabalho; na seção 4, descrevem-se os materiais e métodos utilizados para aquisição e pré-tratamento dos dados empregados; na seção 5 são apresentados os resultados obtidos; e, por fim, é feita uma conclusão do estudo na seção 6.

## 2 Monitoramento da Condição de Papéis Isolantes em Transformadores

O transformador de potência é um equipamento vital em uma subestação distribuidora e conseqüentemente para o sistema elétrico como um todo [7]. Tal equipamento utiliza o papel elétrico como isolante sólido e o óleo mineral isolante (OMI) como fluido isolante, trocador de calor e meio de extinção.

Durante a operação do transformador ocorrem: a degradação do papel isolante, diminuindo sua resistência mecânica e o grau de polimerização (GP); a deterioração do óleo isolante, alterando suas características isolantes; e ainda há a formação de gases provenientes desses dois processos. Esse desgaste natural e contínuo é proveniente da interação entre os materiais de construção do equipamento e das condições de operação, ocasionando a diminuição da vida útil do equipamento e o aumento de custos operacionais.

Atualmente o acompanhamento periódico das propriedades do OMI é realizado com os ensaios físico-químicos de rotina, e a condição de operação é avaliada através da análise dos gases dissolvidos no óleo. Estes dois ensaios são fundamentais para operação do transformador por um longo período. No entanto, a vida útil deste equipamento depende do seu isolamento sólido e este atinge o fim de vida quando a sua resistência mecânica se reduz à metade do valor inicial,

o que corresponde a um nível baixo do Grau de polimerização (GP) [8].

As seguradoras de transformadores elétricos nos últimos anos passaram a exigir dos clientes o acompanhamento da vida útil do equipamento pela medida do GP. Esta exigência diminuiu o pagamento de quantias relevantes por um equipamento que estava em final de vida útil e que iria falhar de qualquer forma. Também reduziu os custos operacionais por uma falha inesperada, pois passou-se a determinar o final da vida do equipamento antes deste vir a falhar.

O método normalizado pela ABNT (NBR 8148) [9] quantifica GP pela determinação do número médio de glicose que constitui a molécula polimérica da celulose. O método é baseado na medida da viscosidade específica de uma amostra de papel dissolvida em etilenodiamina cúprica, a partir da qual a viscosidade intrínseca da solução é deduzida e o GP calculado. Este ensaio é demorado, destrutivo e muitas vezes necessita de uma grande quantidade de papel para obter resultados repetitivos e confiáveis.

Trabalhos têm sido publicados envolvendo a análise de papel isolante por espectroscopia NIR. Mudanças na intensidade dos picos espectrais estão correlacionadas a estruturas organizacionais diferentes do papel, que dependem do processo de fabricação empregado [10]. As mudanças ocorridas na estrutura cristalina de papel isolante de transformador quando envelhecido em laboratório foram estudadas por [11] e obtiveram resultados consistentes para um mesmo tipo de papel. Usando amostras de papel kraft novo envelhecidas por via térmica em laboratório [12], empregou-se calibração multivariada aos espectros NIR conseguindo bons modelos para GP. Embora poucos dados e nenhuma amostra real tenham sido incorporados nesse estudo, foi demonstrado o potencial da técnica para determinação espectroscópica do GP.

O uso da espectroscopia NIR na determinação do GP de forma mais abrangente e com maior confiabilidade foi avaliada por [5]. Esta abrangência foi assegurada selecionando-se um conjunto de dados que contém três diferentes tipos de papel, oriundos de 75 transformadores diferentes em vários estágios de degradação.

O presente trabalho aborda o problema de avaliação de papéis isolantes por espectroscopia NIR em um contexto de classificação. Nesse caso, dado o espectro de uma amostra de papel, deseja-se associar a amostra a uma de três possíveis classes, a saber: baixo GP (menor que 300 unidades), médio GP (entre 300 e 700 unidades) e alto GP (maior que 700 unidades). Conforme item 9.1.1 da norma NBR 8148, a classificação de uma amostra em um destes grupos já seria uma boa indicação do estado de degradação do papel.

O estudo será conduzido empregando o mesmo conjunto de dados levantado por [5].

### 3 Algoritmo de Aprendizagem Ativo

#### 3.1 Método de classificação adotado

Supõe-se que o conjunto de treinamento contenha amostras rotuladas (isto é, amostras com índice de classe conhecido) e não-rotuladas. Adota-se a notação  $\{l_1, l_2, \dots, l_L\}$  para as  $L$  amostras rotuladas e  $\{u_1, u_2, \dots, u_U\}$  para as  $U$  amostras não rotuladas. Por simplicidade de notação, representa-se por  $s_i$  uma amostra genérica, que pode ser rotulada ou não.

O índice de classe da amostra  $s_i$  é denotado por  $w(s_i)$ . Tal índice pode assumir os valores  $w_1, w_2, \dots, w_Q$ , sendo  $Q$  o número de classes envolvidas no problema. O valor estimado de  $w(s_i)$  é representado por  $\hat{w}(s_i)$ . Supõe-se que o valor correto de  $w(s_i)$  é conhecido apenas para as amostras rotuladas  $l_i$ . Para as amostras não-rotuladas  $u_i$ , é necessário obter os valores estimados  $\hat{w}(u_i)$ .

Considera-se que os  $n$  atributos que caracterizam a amostra  $s_i$  estejam dispostos em um vetor  $\mathbf{v}(s_i) \in \mathbf{R}^n$ . A distância  $d$  entre duas amostras  $s_i$  e  $s_j$  é então dada pela distância euclidiana entre  $\mathbf{v}(s_i)$  e  $\mathbf{v}(s_j)$ , ou seja:

$$d(s_i, s_j) = \|\mathbf{v}(s_i) - \mathbf{v}(s_j)\| \quad (1)$$

Para cada amostra  $s_i$ , um índice  $P(s_i, w_j)$  de pertinência à classe  $w_j$  pode ser calculado como:

$$P(s_i, w_j) = \sum_{\substack{k=1 \\ w(l_k)=w_j}}^L \exp\left(\frac{-(d(s_i, l_k))^2}{\sigma^2}\right) \quad (2)$$

Vale salientar que a expressão (2) faz uso de janelas de Parzen gaussianas [13] centralizadas sobre as amostras rotuladas. Neste trabalho, o ajuste do parâmetro  $\sigma$  foi realizado como em [14].

O valor estimado  $\hat{w}(s_i)$  é aquele que maximiza  $P(s_i, w_j)$ , isto é:

$$\hat{w}(s_i) = \arg \max_{w_j \in \{w_1, w_2, \dots, w_Q\}} P(s_i, w_j) \quad (3)$$

#### 3.2 Algoritmo de Aprendizagem Ativo Proposto

Supõe-se que o conjunto de treinamento possua inicialmente um pequeno número de amostras rotuladas. No algoritmo de aprendizagem ativo aqui proposto, a próxima amostra a ser rotulada depende da confiança na estimativa para o índice de classe dessa amostra. Para esse propósito, define-se um grau de confiança  $C(\hat{w}(s_i))$  na estimativa  $\hat{w}(s_i)$  como

$$C(\hat{w}(s_i)) = \frac{P(s_i, \hat{w}(s_i))}{\bar{P}(s_i, \hat{w}(s_i))} \quad (4)$$

sendo

$$\bar{P}(s_i, w_j) = \sum_{\substack{k=1 \\ k \neq j}}^Q P(s_i, w_k) \quad (5)$$

De acordo com as expressões (4) e (5), a confiança na classificação será grande se o índice de pertinência de  $s_i$  à classe atribuída for elevado em comparação com os índices de pertinência de  $s_i$  às demais classes.

O algoritmo de aprendizagem ativo pode então ser descrito da seguinte forma. As amostras não-rotuladas  $u_i$  do conjunto de treinamento são classificadas, obtendo um valor estimado  $\hat{w}(u_i)$  de acordo com a expressão (3). O grau de confiança na classificação  $C(\hat{w}(u_i))$  é então calculado empregando-se as expressões (4) e (5). Por fim, escolhe-se a próxima amostra  $u^*$  a ser rotulada como sendo aquela cuja classificação estimada  $\hat{w}(u_i)$  apresentar a menor confiança, ou seja:

$$u^* = \arg \min_{u_i \in \{u_1, u_2, \dots, u_U\}} C(\hat{w}(u_i)) \quad (6)$$

Tal procedimento objetiva a seleção de amostras que estejam próximas à superfície de separação entre as classes, de modo a melhor caracterizar a região de fronteira. Tendo-se rotulado  $u^*$ , a mesma é movida para o conjunto de amostras rotuladas e o procedimento pode ser repetido para a seleção de uma nova amostra.

### 4 Metodologia

As amostras de papel isolante, oriundas de 75 transformadores, são de três tipos: kraft, crepe e papelão. As amostras de papel isolante foram preparadas conforme procedimentos indicados pela norma NBR 8148, depois foram feitas as medidas espectroscópicas e por fim foi utilizada a técnica de aprendizagem ativa para classificação. Os métodos e instrumentos utilizados são detalhados a seguir.

#### 4.1 Preparação das Amostras de Papéis

As amostras de papéis isolantes foram obtidas ao longo de três anos de dois laboratórios que analisaram o grau de polimerização, segundo a norma NBR 8148. Os resultados obtidos variaram na faixa de 200 a 900 unidades de grau de polimerização. Utilizou-se clorofórmio como desengraxante para remoção do OMI impregnado no papel, em extrator tipo Soxhlet por 8 horas. Em seguida, as 75 amostras foram picotadas.

#### 4.2 Medidas espectroscópicas

Os espectros no infravermelho próximo foram adquiridos com as amostras picotadas, diretamente em um espectrômetro Bomem MB 160, equipado com um acessório de reflectância difusa com resolução de  $8 \text{ cm}^{-1}$  com 50 varreduras. A região espectral utilizada foi de 1265 a 2500 nm. A quantidade de amostra utilizada em cada medida foi a suficiente para recobrir a janela do acessório.

#### 4.3 Pré-processamento dos espectros

Para eliminar variações na linha de base dos espectros, optou-se por trabalhar com espectros derivativos. Para isso, o cálculo da derivada foi efetuado empregando-se um filtro

de Savitzky-Golay com polinômio de segundo grau e janela de 23 pontos, assim como em [5]. O resultado está apresentado na Fig. 1.

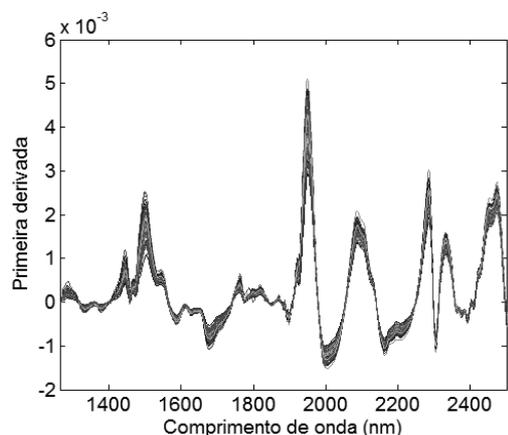


Figura 1 - Espectros derivativos das 75 amostras de papel empregadas no estudo.

Em seguida, realizou-se uma Análise de Componentes Principais, retendo-se oito componentes que explicam 99% da variância dos dados. Os escores dessas oito componentes foram utilizados como atributos para descrever as amostras no restante do trabalho.

#### 4.4 Separação das Amostras

As 75 amostras disponíveis foram separadas em um conjunto de treinamento (2/3 do número total de amostras) e outro de teste (1/3), conforme a Tabela 1. Para isso, utilizou-se o algoritmo de amostragem uniforme de [15]. O conjunto de teste não foi utilizado no algoritmo de aprendizado ativo, tendo sido reservado para avaliar a taxa de acerto do classificador.

Tabela 1: Conjuntos de treinamento e teste

Classe	Treinamento	Teste
GP Baixo	7	4
GP Médio	26	13
GP Alto	17	8
<b>Total</b>	<b>50</b>	<b>25</b>

#### 4.5 Avaliação do algoritmo de aprendizado ativo

As amostras de treinamento a serem rotuladas foram selecionadas de acordo com o algoritmo de aprendizado ativo descrito na Seção 3.2. O algoritmo foi inicializado com uma única amostra rotulada, escolhida de forma aleatória. Tendo-se rotulado um dado número de amostras de treinamento, as amostras de teste foram classificadas de acordo com a expressão (3).

Para fins de comparação, foram também utilizados dois métodos de classificação empregando seleção aleatória das

amostras a serem rotuladas. No primeiro método, a classificação foi realizada empregando a expressão (3), assim como no algoritmo de aprendizado ativo. No segundo método, a classificação foi efetuada empregando-se os  $k$ -vizinhos mais próximos ( $k$ -Nearest-Neighbours, kNN).

Vale salientar que o resultado do aprendizado ativo depende da escolha da primeira amostra. Da mesma forma, os resultados obtidos pelos outros dois métodos dependem da realização do processo de seleção aleatória de amostras. Por essa razão, a aplicação de cada método foi repetida 100 vezes, com diferentes sementes para o gerador de números aleatórios. A taxa de acerto média obtida no conjunto de teste foi então considerada para fins de comparação.

## 5 Resultados e Discussão

Para permitir uma comparação apropriada com o kNN, o número  $k$  de vizinhos mais próximos foi escolhido de modo a otimizar a taxa de acerto no conjunto de teste. Como se pode notar na Fig. 2, o melhor resultado é obtido para  $k = 1$ .

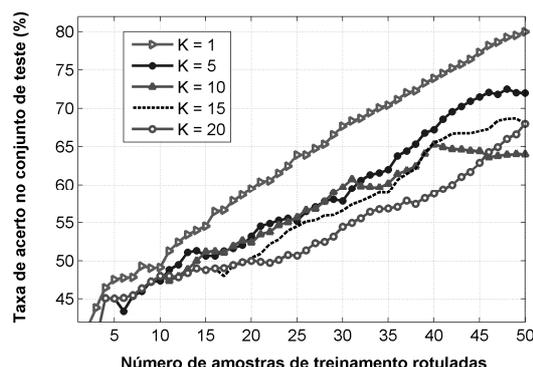


Figura 2 - Efeito do número de vizinhos mais próximos ( $k$ ) na classificação por kNN.

Os resultados obtidos pelas três técnicas empregadas são apresentados na Fig. 3. Sem o uso de aprendizado ativo, a classificação empregando janelas de Parzen teve desempenho similar ao do kNN, apresentando uma ligeira superioridade quando todas as amostras de treinamento são rotuladas. Com a introdução do aprendizado ativo, a taxa de acerto foi substancialmente melhorada ao longo do processo de seleção das amostras a serem rotuladas. Para a classificação empregando janelas de Parzen, vale notar que as curvas correspondentes ao aprendizado ativo e à seleção aleatória convergem para um mesmo ponto final. Tal ponto corresponde à utilização de todas as 50 amostras do conjunto de treinamento para classificar as amostras do conjunto de teste. O principal mérito do aprendizado ativo, em comparação com a seleção aleatória, consiste em reduzir o número de amostras de treinamento a serem rotuladas para obter uma dada taxa de acerto no conjunto de teste.

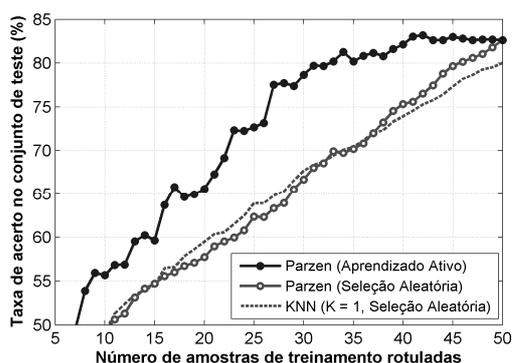


Figura 3 - Comparação dos resultados obtidos pelas três técnicas de classificação empregadas.

## 6 Conclusão

Neste trabalho, foi apresentada uma estratégia de aprendizado ativo para classificação de papéis isolantes com respeito ao grau de polimerização (GP). A classificação é efetuada empregando espectros de infravermelho próximo, que podem ser adquiridos de forma comparativamente simples. Contudo, o treinamento do classificador ainda requer a disponibilidade de amostras cujo GP tenha sido determinado de forma convencional. Por essa razão, foi proposto um algoritmo de aprendizado ativo para seleção judiciosa de tais amostras.

Como resultado, observou-se que o aprendizado ativo proporcionou maiores taxas de acerto, em comparação com escolhas aleatórias das amostras a serem rotuladas. Desse modo, conclui-se que o algoritmo proposto seria vantajoso caso restrições de tempo ou custo imponham limites ao número de amostras que possam ser analisadas pelo método convencional.

## Agradecimentos

CNPq, Procad/CAPES, FAPESP (processo 2006/58850-6), FINEP/CTPETRO, FACEPE, SM Controle de Qualidade Ltda, CHESF.

## Referências Bibliográficas

[1] Spiegel, M. R. *Estatística*. 3ª Ed. São Paulo: Makron Books, 1993.  
 [2] Kothari, R.; Jain, V. Learning From Labeled and Unlabeled Data Using a Minimal Number of Queries. *IEEE Trans. Neural Networks*, v.14, n. 6, p. 1496-1505, 2003.  
 [3] Cohn, D.A.; Ghahramani, Z.; Jordan, M.I. Active learning with statistical models. *J. Artificial Intelligence Research*, v.4, p.129-145, 1996.  
 [4] Cohn, D.A. Neural Network Exploration Using Optimal Experiment Design. *Advances in Neural Information Processing Systems*, Morgan Kaufmann, v.6, p. 679-686, 1994.

[5] Santos, E.O. ; Silva, A.M.S. ; Frago, W.D. ; Pasquini, C. ; Pimentel, M. Avaliação de técnicas de processamento para predição do grau de polimerização em papel isolante empregando a espectroscopia NIR e modelagem empírica multivariada. In: *Anais do XVII Congresso Brasileiro de Engenharia Química*, Recife, 2008.

[6] Pasquini, C., Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications, *J. Braz. Chem. Soc.*, v. 4, n. 2, p. 198-219, 2003.

[7] Barreto Jr., J.T.; Langone, J.; Barros, J.R.F.; Pesenti, G.R. ; Lisboa, C.A.W. e Cantanhede, F.P. *Potencialidade total do processo de tratamento de óleo isolante com o transformador energizado*. Abramam, 1996.

[8] Darveniza, M.; Saha, T. K.; Hill, D. J. T. et al. Studies of the insulation in aged power transformers - Part 2: Fundamental electrical and chemical considerations properties. In: *Proc. International Conference on Properties of Dielectric Materials*, 1991.

[9] ABNT NBR 8148. Papéis celulósicos novos e envelhecidos, para fins elétricos: Medida do grau de polimerização viscosimétrico médio. 2000.

[10] Silva, E.L.; Wypych, F; Ramos, L.P. Caracterização do efeito cumulativo das etapas de branqueamento (O-Q-OP-ZQ-PO) sobre a composição centesimal e estrutura fina de polpas Kraft. In: *Congreso Iberoamericano de Investigación em Celulosa y Papel*, 2000.

[11] Ali M.; Emsley A. M.; Herman H.; Heywood R. J. *Spectroscopic studies of the ageing of cellulosic paper*. Polymer. 2001.

[12] Sá, A.L.C.R. de; Nepomuceno, P. R. Avaliação do grau de polimerização antes e após secagem de transformadores em fábrica. In: *Anais do XIV SNPTEE - Seminário Nacional de Produção e Transmissão de Energia Elétrica*, Belém, 1997.

[13] Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*, 2nd ed, John Wiley, New York, 2001.

[14] Bengio, Y.; Chapados, N. Extensions to metric-based model selection. *Journal of Machine Learning Research*, v. 3, p. 1209-1227, 2003.

[15] Kennard, R.W.; Stone, L.A. Computer aided design of experiments. *Technometrics*, v. 11, p. 137-148, 1969.