Anais do IX Congresso Brasileiro de Redes Neurais / Inteligência Computacional (IX CBRN)
Ouro Preto 25-28 de Outubro de 2009

FCANN - EXTRAÇÃO DE CONHECIMENTO DE REDES NEURAIS ARTIFICIAIS ATRAVÉS DA ANÁLISE FORMAL DE CONCEITOS

Sérgio M. Dias[†] Luiz E. Zárate[*]

[*] *Pontifícia Universidade Católica de Minas Gerais - Av. Dom José Gaspar, 500 - Tel: (31) 33194407 - CEP 30535610 - Belo Horizonte, MG*

[†] *Universidade Federal de Minas Gerais - Av. Antônio Carlos 6627 - ICEx - 4010 - Pampulha - CEP 31270-010 - Belo Horizonte, Minas - Brasil*

Email: mariano@dcc.ufmg.br, zarate@pucminas.br

Resumo— Devido à sua capacidade de lidar com problemas não-lineares, as Redes Neurais Artificiais são aplicadas em diversas áreas da ciência. Uma vez treinada, a rede neural é capaz de resolver situações não vistas em seu processo de treinamento, mantendo um erro tolerável em suas saídas. Entretanto, os elementos humanos não conseguem assimilar o conhecimento mantido nessas redes, uma vez que tal conhecimento é implicitamente representado por suas conexões e os respectivos pesos numéricos. Neste trabalho, apresenta-se a metodologia FCANN, baseado na Análise Formal de Conceitos para extração de conhecimento de redes neurais previamente treinadas. Como estudo de caso, considera-se dados relativos a um Processo Siderúrgico de Laminação a Frio, uma vez que, devido à sua natureza real e complexa, envolve diversas variáveis em seu domínio. Desta forma, busca-se que, com o conhecimento extraído, o usuário consiga compreender de maneira mais clara o domínio do problema.

Keywords— Redes Neurais Artificiais, Análise Formal de Conceitos, Extração de Conhecimento.

1 Introdução

O entendimento de processos energéticos, econômicos, industriais ou climáticos pode ser obtido pelo estudo das variáveis envolvidas no processo. Uma vez que muitos problemas do mundo real envolvem uma alta complexidade no seu domínio, entender esses processos pode ser muito complexo. Uma alternativa é estudá-los através de representações simbólicas de causa e efeito. Neste contexto, métodos de Inteligência Artificial (IA) têm sido propostos como uma alternativa para representar o conhecimento de processos do mundo real, sem a necessidade de muitos detalhes ou estudos sobre seus princípios físicos.

Uma subárea da IA é o aprendizado de máquinas que busca desenvolver sistemas capazes de aprender, adquirir e representar o conhecimento. O aprendizado de máquinas possui dois principais paradigmas, o simbólico e o conexionista. Um grande representante do paradigma conexionista são as Redes Neurais Artificiais. Uma rede neural tem a capacidade de obter as relações entre atributos de entrada e saída como uma função entre esses parâmetros, na qual parâmetros de entrada são mapeados em parâmetros de saída. Entretanto, a natureza de uma rede neural é de uma "*caixa preta*"(Towell and Shavlik, 1993), em que o conhecimento é implícito e nenhuma explicação pode ser retirada para o processo de tomada de decisão. Desta forma, faz-se necessária a utilização de métodos de extração de conhecimento de redes neurais. Esse conhecimento pode ser extraído na forma de regras acerca do domínio do problema, que demonstram o conhecimento do problema no seu contexto real por intermédio das relações entre as variáveis de seu domínio.

Muitos pesquisadores têm se concentrado na extração de conhecimento das redes neurais (Hilario, 2000; Towell and Shavlik, 1993; Craven, 1996; Andrews and Geva, 2002). Diferentes metodologias têm sido propostas para a extração de regras. Alguns métodos extraem regras analisando a estrutura da rede (pesos, topologia, dentre outros), enquanto outros métodos extraem regras analisando o conjunto de dados utilizado no treinamento da rede. Existem, ainda, métodos que usam ambos, tanto a estrutura da rede quanto o conjunto de dados para extrair regras.

Nos últimos anos uma nova abordagem para extração e representação do conhecimento de redes neurais, baseada na Análise Formal de Conceitos (AFC), foi proposta. Em Vimieiro et al. (2004), ela foi utilizada para extrair regras de redes neurais através da análise visual do diagrama de linhas(Ganter and Wille, 1999). Esse método não foi considerado uma abordagem satisfatória, uma vez que a complexidade desses diagramas pode ser muito alta.

Já em Zárate et al. (2008) a AFC foi utilizada em uma metodologia chamada FCANN *(Formal Concept Artificial Neural Networks)* para extração e representação do conhecimento de redes neurais previamente treinadas. Em Zárate et al. (2008) o método FCANN foi comparado com técnicas convencionais de extração e representação de conhecimento. Essas comparações mostraram a relevância do método para extrair conhecimento qualitativo de processos aprendidos por redes neurais através

de regras de fácil entendimento. O método FCANN extrai relações qualitativas aprendidas por redes neurais, independente das entradas, saídas ou da estrutura da rede.

Neste trabalho a metodologia FCANN é apresentada e aplicada no Processo Siderúrgico de Laminação a Frio, com o objetivo de extrair conhecimento qualitativo que descreva o comportamento do processo. O presente trabalho está dividido da seguinte forma: na segunda seção os principais conceitos da Análise Formal de Conceitos são discutidos, na terceira seção a metodologia FCANN é apresentada, na quarta seção um estudo de caso é realizado, e por último as contribuições deste trabalho são discutidas.

2 Análise Formal de Conceitos

A Análise Formal de Conceitos (AFC) é uma técnica da matemática, baseada na matematização do conceito e na hierarquia conceitual (Ganter and Wille, 1999). Seu objetivo principal envolve a representação do conhecimento em um diagrama específico chamado diagrama de linhas.

A primeira definição importante na AFC é a de contextos formais. Um contexto formal consiste de dois conjuntos G e M e uma relação de incidência I entre G e M; e possui a seguinte notação *(G, M, I)*, sendo $I \subseteq G\ x\ M$ uma relação binária chamada incidência. Os elementos do conjunto G (linhas) são chamados de objetos, enquanto os elementos do conjunto M (colunas) são chamados de atributos. Todo elemento da incidência tem a notação *gIm*, que denota uma relação entre um objeto e um atributo, que deve ser lida como "o objeto g possui o atributo m".

Outra definição importante é a definição de conceitos formais, que formalmente são pares ordenados *(A, B)*, em que A e B são subconjunto do conjunto de objetos e atributos, ou seja, $A \subseteq G$ (extensão) e $B \subseteq M$ (intenção). Cada elemento da extensão (Objetos) possui todos os elementos da intenção (atributos) e conseqüentemente, cada elemento da intenção é um atributo de todos os objetos da extensão.

Quando o conjunto de todos os conceitos formais, de um contexto formal *(G, M, I)*, são ordenados hierarquicamente de acordo com a Teoria do Reticulados Completos (Ganter and Wille, 1999), ele é chamado reticulado conceitual ou diagrama de linhas e recebe a notação *B (G, M, I)*. Os conceitos formais são relacionados como *(A1, B1)* $\leq$ *(A2, B2)* quando *A1* $\subseteq$ *A2* e *B2* $\subseteq$ *B1*, então o par *(A1, B1)* será chamado sub-conceito e o par *(A2, B2)* super conceito.

A análise de dados de forma estruturada é muito mais intuitiva. Entretanto, dependendo da densidade e do número de objetos e atributos do contexto formal, o reticulado pode se tornar muito complexo, dificultando sua análise visual e conseqüentemente, provocando a perda de relações importantes. Felizmente, é possível aplicar algoritmos para encontrar todas as implicações existentes ou implicações com características específicas. Uma implicação f é uma relação entre dois conjuntos (Ganter and Wille, 1999), ou seja, uma regra do tipo $X \mapsto Y$, onde $(X, Y) \subseteq G$. Um conjunto de implicações $\pounds$ de um contexto formal *(G, M, I)* é chamado base de implicações. Um algoritmo capaz de extrair essa base é o algoritmo *Next Closure* (Ganter and Wille, 1999).

3 Metodologia FCANN

A metodologia FCANN (Zárate et al., 2008) extrai e representa o conhecimento de processos previamente treinados por rede neurais de forma simbólica e de fácil compreensão. Considerando uma rede com estrutura multicamada, *feedforward*, e totalmente conectada com N entradas e M saídas, pode-se assim descrever os passos da metodologia FCANN:

1 - Selecione um conjunto de dados representativos de um processo e treine uma rede neural. Esses dados são representados como: $X = [x_{ij}]m \times n$, onde: n é o número de parâmetros; x_{ij} para todo i= 1, ...,m e j=1, ..., n-1 são os parâmetros de entrada e x_{in} para i=1, ..., m é o parâmetro de saída. Esse parâmetro de saída deve ter uma distribuição de probabilidade conhecida e poderá ser uma distribuição normal $N_1(\overline{x}, S(X))$.

2 - Defina a estrutura da rede neural multicamadas (com N parâmetros de entradas; H camadas escondidas e M saídas) e treine a rede.

3 - Construa uma base de dados sintética para a extração de conhecimento da rede neural. Esta base de dados deverá ser gerada considerando o intervalo de domínio dos parâmetros de entrada, podendo ser definida como: $Y = [y_{ij}]_{p \times n-1}$; onde: Y tem apenas elementos gerados com o propósito de extração de conhecimento, P representa o número de registros e n o número de parâmetros. Cada parâmetro de entrada possui um valor mínimo e máximo, que define o domínio do intervalo do parâmetro. Os valores mínimos e máximos são vetores que são, respectivamente, definidos como: $inf = \{u_1, .., u_{n-1}\}; u_j = min[x_{ij}]; i = 1, .., m$ e $sup = \{v_1, .., v_{n-1}\}; v_j = max[x_{ij}]; i = 1, ., m$.

O vetor W define o número de dados que será gerado por cada parâmetro, entre os valores mínimos

e máximos: $W = [W_j]; j = 1, ..., n-1$.

Assim, o intervalo da variação de cada parâmetro, que irá compor os dados sintéticos, é representado pela seguinte expressão: $Int = \{I_1, ..., I_{n-1}\}$; $I_j = \frac{|v_j - u_j|}{w_i}$; $\forall\, j = 1, ..., n-1$

É possível observar que o número de parâmetros P do conjunto que será gerado depende do número de dados de cada parâmetro. Assim P pode ser definido como: $p = W_1 \times W_2 \times, ..., \times W_{n-1} = \prod_{i=1}^{n-1} W_i$

4 - Aplique a base de dados sintética Y na rede neural treinada para obter o parâmetro de saída $Z = [z_{ij}]_{p\times 1}$ o qual se espera ter uma distribuição probabilidade $N_2(\overline{z}, S(Z))$ como a de um parâmetro real. Para verificar a generalização da rede, a comparação entre as distribuições $N_2(\overline{x}, S(X))$ e $N_2(\overline{z}, S(Z))$ pode ser feita. Se $e_{x,z} = |\overline{x}_1 - \overline{z}_2|$ e $e_{S(x,z)} = |S(x)_1 - S(z)_2|$ são erros significantes, isto significa que os dados do treinamento não são representantes para o processo. Neste caso, retorne ao passo 1.

5 - Classifique os parâmetros (colunas) da matriz $U = [Y, Z]_{p\times n}$ em intervalos.

6 - Construa o contexto formal, tabela cruzada.

7 - Obtenha os conceitos formais e construa o diagrama de linhas.

8 - Aplique o algoritmo *Next Closure* para obter um conjunto mínimo de regras do tipo: **Se** ... **Então** ... .

Durante o estudo de um processo físico pode ser necessária a análise de um subconjunto de parâmetros envolvidos no mesmo. Nesse caso pode se atribuir um valor constante para alguns parâmetros e variar os demais procurando assim determinar o comportamento qualitativo do processo. Esse estudo pode ser realizado através do método FCANN alterando os passos 3 e 6 que compõem o método:

Seja C o conjunto de todos os índices dos parâmetros que se pretende fixar, neste caso, os passos 3 e 6 podem ser redefinidos como:

3 - De acordo com o passo 3 deve-se construir uma base de dados sintética [S]. Para atribuir um valor constante a um subconjunto C de parâmetros faça: $Se\ j \in C$ Então $S_{kj} = q; for\ k = 1 \ldots, w_j;\ onde\ u_j \leq q \leq v_j$

6 - De acordo com a etapa 6 deve-se construir o contexto formal *K:=(G, M, I)* neste caso faça $M = M'$, onde: $M' = \{m_i | m_i \in M; onde j \in C\ e\ i \neq j\}$

Para aplicação da metodologia FCANN, a ferramenta SOPHIANN (Zárate et al., 2008) será utilizada.

4 Estudo de caso - Processo Siderúrgico de Laminação a Frio

Como estudo de caso o Processo Siderúrgico de Laminação a Frio foi escolhido. Esse processo tem como objetivo a produção de tiras de aço, cuja qualidade é medida principalmente pela sua espessura final. Uma motivação para esse estudo de caso é a captura de conhecimento através de representações simbólicas, diagrama de linhas e regras, que permitam o entendimento de operadores menos experientes, além de fornecer informação sobre o processo sendo analisado que possam ser utilizadas no projeto de sistemas de controle e supervisão online.

4.1 Representação Neural

Em Zárate (1998) uma representação para o Processo Siderúrgico de Laminação a Frio foi discutida e os resultados serão considerados neste trabalho. O modelo teórico do processo de laminação calcula a carga P, pressão exercida pelos cilindros, por intermédio da expressão não linear. Essa equação envolve relações não lineares, que conduzem geralmente a uma equação não linear de difícil análise ou solução numérica descrita por $P = f(h_i, h_o, \mu, t_b, t_f, \bar{y}, E, R)$, onde: h_i = espessura de entrada, h_o = espessura de saída, μ = coeficiente de atrito, t_b = tensão à re, t_f = tensão à frente, $\bar{y}$ = tensão escoamento, E = o módulo de Young da tira, R = raio do cilindro.

Dentre os parâmetros, o seguinte modelo neural foi considerado para representar o processo de laminação: $(h_i, h_o, \mu, t_b, t_f, \bar{y}) \overset{RNA}{\rightarrow} (P)$

A arquitetura da rede neural considerada foi uma rede neural multicamada, *feedforward*, totalmente conectada com *N=6* entradas e *M=1* saída e 13 (2N+1), como sugerido por Kovács (1996), neurônios na camada oculta. Como função de ativação, a função não linear *log sigmóide* foi escolhida e a rede foi treinada com o algoritmo *backpropagation*. Para facilitar a convergência do processo de treinamento, os dados para o treinamento e validação foram normalizados no intervalo de [0.2, 0.8][1], utilizando as seguintes equações: $Ln = \frac{(Lo - Lmin)}{(Lmax - Lmin)}$ e $Lo = Ln * Lmax + (1 - Ln) * Lmin$; onde Ln é o valor normalizado, Lo é o valor para

[1]Quando a rede neural contém nodos com a função de ativação log-sigmoide, melhores resultados são obtidos se os dados estão normalizados no intervalo [0.2, 0.8](Altincay and Demirekler, 2002).

normalizar, $Lmin$ e $Lmax$ são os valores mínimo e máximo respectivamente. $Lmin$ e $Lmax$ foram calculados como: $Lmin = (4 * LimInf - LimSup)/3$ e $Lmax = (LimInf - 0.8 * Lmin)/0.2$

Para obtenção de uma base de dados para o processo de treinamento e validação, o Modelo de Alexander (1972) foi utilizado. Esse modelo é considerado o mais completo na área e, apesar disso, é caracterizado pela demanda de um grande esforço computacional, o que motiva pesquisas de novas representações, nesse caso através de redes neurais.

Com o objetivo de obter uma base de dados que represente adequadamente o processo de laminação a frio, dez milhões de registros foram gerados. Essa grande quantidade de registros é possível devido a variação de cada parâmetro de entrada e da combinação entre eles, o que resulta em diferentes valores de carga P. Entretanto, o treinamento da rede neural com essa quantidade de dados demanda um grande esforço computacional. Portanto, é necessário obter um conjunto de dados representativos para o processo de treinamento e validação.

Em Zárate et al. (2006) algumas técnicas estatísticas são sugeridas para a construção de conjuntos de dados melhores, mantendo a capacidade de generalização da rede neural. A partir dessas sugestões, a seguinte Equação foi utilizada: $n = (z/e)^2 \times (f \times (1 - f))$; onde: n é o tamanho da amostra; z é nível de confiança; e é o erro em torno da média e f é a proporção da população.

Fixando z em 90% (1.645), f em 0.5 e e em 3%, $n \cong 752$ registros. Essa amostra foi selecionada utilizando o seguinte procedimento: (1º) Para cada parâmetro, um registro que continha o valor máximo, mínimo e próximo da média foi selecionado; (2º) O ponto de operação utilizado pelo Modelo de Alexander foi selecionado; (3º) Os elementos restantes foram selecionados aleatoriamente, com a finalidade de se alcançar o total de 752 registros.

Para o processo de treinamento um erro de 98 N/mm^2 foi aplicado e 752 dados foram utilizados. Já no processo de validação todos os dados que não foram vistos no processo de treinamento foram utilizados, isto é 999248 dados. O treinamento apresentou resultados satisfatórios e pode ser considerado capaz de representar o processo. Na Tabela 1 as informações estatísticas do erro obtido no processo de treinamento e validação são apresentadas.

Tabela 1: Resultado do treinamento e validação.

	Treinamento	**Validação**
Número de dados	752	999248
Erro mínimo	0.045	1.94E-05
Erro máximo	96.831	159.965
Erro médio	15.576	15.42
Desvio padrão	9.4123	9.3477

4.2 Extração de conhecimento através do método FCANN

Após um satisfatório processo de treinamento pode se aplicar o método FCANN. Utilizando 2 dado por parâmetro e 2 intervalos de discretização[2], que resulta em uma base de dados sintética com 64 registros, (veja seção 3), uma base de implicações $\pounds$, com 22 regras, foi obtida. Dois exemplos dessas regras e demonstrado a seguir: **Se** $\mu = 1$ **e** $\bar{y} = 1$ **Então** $P = 1$ e **Se** $h_i = 2$ **e** $h_o = 1$ **e** $\mu = 1$ **e** $t_b = 2$ **e** $\bar{y} = 2$ **e** $P = 2$ **Então** $t_f = 1$.

Em Zárate (1998), foi observado que o parâmetro tensão à ré (t_b) tem pouca influência no processo de laminação. Dentre as regras da base de implicações obtida nenhuma apresentou esse parâmetro, confirmando essa prévia observação. A eliminação desse parâmetro deve-se as características da base de implicações extraída pelo algoritmo *Next Closure*, que é mínima, completa e não redundante.

Para facilitar a análise do comportamento qualitativo do processo em estudo, pode analisar um subconjunto de parâmetros envolvidos no mesmo. Por exemplo, analisar o comportamento da carga de laminação P em função da variação da espessura de saída h_o. Aplicando 15 intervalos de discretização, 15 dados por parâmetros e fixando os demais parâmetros em valores médios, percebe-se esse comportamento através do diagrama de linhas (Figura 1). No diagrama observam-se dois grupos de conceitos formais (representando diferentes pontos de operação do processo), o grupo da esquerda apresenta um valor de carga P superior e associados a esse grupo encontram-se os menores valores de espessura de saída h_o. Já o grupo da direita apresenta um valor de carga P inferior, associados a ele encontram-se os maiores valores de espessura de saída h_o. Este comportamento é esperado para esse tipo de processo e nota se claramente a representação do comportamento qualitativo das variáveis analisadas.

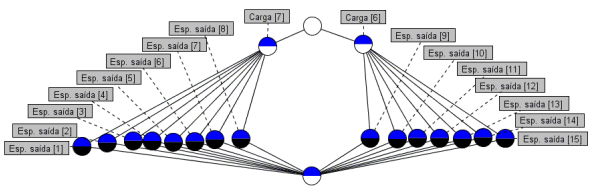

Figura 1: Diagrama de linhas para análise da carga P e espessura de saída h_o.

Além da análise do diagrama de linhas, pode-se obter o comportamento do processo de laminação

[2]O número de pontos de corte do processo de discretização ainda é uma questão em aberto. Já o número de dados por parâmetro deve ser suficiente para preencher todo contexto formal, evitando assim a perda de sua representatividade ou a criação de objetos repetidos.

a frio através da base de implicações $\pounds$. Novamente utilizando 15 intervalos de discretização e analisando o comportamento da carga P e da espessura de saída h_o, para diferentes valores de coeficiente de atritos μ, é possível reconstruir a curva característica da carga de laminação.

Na Figura 2 um esboço do comportamento da carga de laminação P para variações de $\pm$ 20% do valor nominal do coeficiente de atrito μ é apresentado. O Eixo "y" representa os 15 intervalos de discretização possíveis de carga, já o eixo "x" representa o valor da espessura de saída h_o, para um mesmo valor de espessura de entrada h_i. Para um valor mínimo de atrito μ a curva de carga se desloca abaixo da curva de carga nominal (condição ideal). Para esse caso todos os valores de carga P foram mapeado nos intervalos 5 e 6. Para um valor médio de atrito μ a curva foi mapeada nos intervalos 6 e 7. Já para um aumento no valor do atrito μ a curva de desempenho se deslocou para cima da curva nominal como esperado, e todas as regras foram mapeadas nos intervalos de 7 a 9.

A variação do coeficiente de atrito μ provoca uma alteração nos valores característicos da carga P e esse comportamento foi mapeado pelo método FCANN através da base de implicações. É necessário destacar que para níveis de discretização pequenos, menores que 10 intervalos, esse comportamento não pode ser percebido, pois os intervalos de discretização não conseguem perceber a variação da carga P.

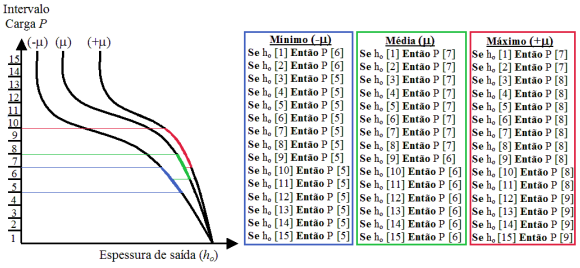

Figura 2: Esboço do comportamento da carga P para variações de $\pm$ 20% do valor nominal do coeficiente de atrito μ

5 Conclusões

Neste trabalho, a metodologia FCANN, baseada na Análise Formal de Conceitos, capaz de extrair e representar o conhecimento de Redes Neurais Artificiais, foi apresentada e aplicada no Processo Siderúrgico de Laminação a Frio. Desta forma, buscou-se que, com o conhecimento extraído, o usuário conseguisse compreender de maneira mais clara o domínio do problema.

Os resultados mostraram que a representação neural do processo foi satisfatória e que o método FCANN é capaz de representar o comportamento qualitativo do processo.

Referências

Alexander, J. M. (1972). On the theory of rolling, *Proc. of the Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 326, pp. 535–563.

Altincay, H. and Demirekler, M. (2002). Why does output normalization create problems in multiple classifier systems?, pp. 775–778.

Andrews, R. and Geva, S. (2002). Rule extraction from local cluster neural nets, *Neurocomputing* **47**: 1–17.

Craven, M. W. (1996). *Extracting Comprehensible Models from Trained Neural Networks*, PhD thesis, University of Wisconsin-Madison.

Ganter, B. and Wille, R. (1999). *Formal Concept Analysis: Mathematical Foundations*, Springer-Verlag, Germany.

Hilario, M. (2000). An overview of strategies for neurosymbolic integration, *in* S. Wermter and R. Sun (eds), *Hybrid Neural Systems*, 1 edn, Springer-Verlag, chapter 2.

Kovács, Z. L. (1996). *Redes Neurais Artificiais*, Edição Acadêmica São Paulo, Brasil.

Towell, G. G. and Shavlik, J. W. (1993). The extraction of refined rules from knowledge-based neural networks, *Machine Learning* **13**(1): 71–101.

Vimieiro, R., Zárate, L., J. P. D. Silva, E. M. D. P. and Diniz, A. (2004). Rule extraction from trained neural networks via formal concept analysis, *8th IASTED Inter. Conf. on Arti. Intelli. and Soft Comp.* pp. 334–339.

Zárate, L. E. (1998). *Um Método para Análise da Laminação Tandem a Frio*, PhD thesis, Universidade Federal de Minas Gerais, Escola de Engenharia Metalurgica de Minas, Belo Horizonte, MG, Brasil.

Zárate, L. E., Dias, S. M. and Song, M. A. J. (2008). Fcann: A new approach for extraction and representation of knowledge from ann trained via formal concept analysis, *Neurocomputing* **71**: 2670–2684.

Zárate, L. E., Pereira, E. M. D., Oliveira, L. A. R., Gil, V. P., Santos, T. R. A. and Nogueira, B. (2006). Techniques for training sets selection in the representation of a thermosiphon system via ann, in Proc IJCNN, pp. 2736– 2741.