UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO CENTRO TECNOLÓGICO DEPARTAMENTO DE ENGENHARIA ELÉTRICA PROJETO DE GRADUAÇÃO

INGRID ANDRADE REIS

USO DE REDES NEURAIS RECORRENTES PARA ESTIMAÇÃO DO BTP EM UM PROCESSO DE SINTERIZAÇÃO

VITÓRIA – ES DEZEMBRO/2018 INGRID ANDRADE REIS

USO DE REDES NEURAIS RECORRENTES PARA ESTIMAÇÃO DO BTP EM UM PROCESSO DE SINTERIZAÇÃO

Parte manuscrita do Projeto de Graduação da aluna **Ingrid Andrade Reis**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheira Eletricista.

Orientador: Prof. Dr. Klaus Fabian Côco

VITÓRIA – ES DEZEMBRO/2018

USO DE REDES NEURAIS RECORRENTES PARA ESTIMAÇÃO DO BTP EM UM PROCESSO DE SINTERIZAÇÃO

Parte manuscrita do Projeto de Graduação da aluna **Ingrid Andrade Reis**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Aprovada em 10 de dezembro de 2018.

COMISSÃO EXAMINADORA:

ri Cèco

Klaus Fabian Côco, DSc Universidade Federal do Espírito Santo Orientador

Jorge Aehing Samatelo, DSc Universidade Federal do Espírito Santo Examinador

its of h

Vitor Faiçal Campana, MSc Instituto Federal do Espírito Santo Examinador

RESUMO

O *Burn Through Point* (BTP) é importante parâmetro para o processo de sinterização nas usinas de produção de aço e, tendo em vista a crescente demanda deste material nos processos de produção, novas técnicas que possibilitem a melhora na sua qualidade e na sua produtividade são estudadas e aplicadas continuamente na indústria. Trabalhos anteriores já utilizaram Redes Neurais Artificiais (RNA) para a predição do BTP e comprovaram sua eficiência. Dessa forma, este trabalho propõe a análise de dados reais de uma indústria siderúrgica empregando diferentes modelos de RNA para prever o BTP utilizando horizonte de previsão de 1h. Os modelos com e sem variáveis exógenas, selecionadas pelos especialistas de sinterização, foram desenvolvidos, bem como um estudo da relevância dessas variáveis, através de testes de poda, foram implementados visando a previsão do valor do BTP. Os resultados indicam que, nas condições dos modelos avaliados, a rede neural sem variáveis exógenas apresentou desempenho superior às outras. Isto posto, tal modelo pode ser de grande aplicabilidade no auxílio no controle do processo produtivo, contribuindo para maior assertividade do BTP.

LISTA DE FIGURAS

Figura 1 – Máquina de Dwaight-Lloyd10
Figura 2 – Representação de um Neurônio Típico12
Figura 3 - Redes Neurais Artificiais alimentadas adiante (Feedfoward): a) Com camada
única; b) Com múltiplas camadas13
Figura 4 – Rede Neural Artificial Recorrente com neurônios ocultos
Figura 5 – Diagrama de blocos dos modelos: a) NARX; b) NAR; c) NOE15
Figura 6 – Exemplo de representação da regressão17
Figura 7 – Exemplo de representação da autocorrelação do erro
Figura 8 – Exemplo do comportamento de 4 variáveis do processo de sinterização estudado:
a) BTP; b) Densidade da mistura; c) Temperatura na caixa de vento nº 19; d) Velocidade da
máquina de sínter
Figura 9 – Previsão do BTP antes da poda: a) Valores a cada amostra; b) Erro de predição a
cada amostra
Figura 10 – Previsão do BTP após poda: a) Valores a cada amostra; b) Erro de predição a
cada amostra
Figura 11 – Variável de saída antes e após tratamento de outliers
Figura 12 – Previsão do BTP com modelo NARX, 11 variáveis e horizonte de previsão de
1h
Figura 13 – Regressão para treinamento com modelo NARX, 11 variáveis e horizonte de
previsão de 1h
Figura 14 – Autocorrelação do erro para treinamento com modelo NARX, 11 variáveis e
horizonte de previsão de 1h
Figura 15 – Previsão do BTP com modelo NARX, 4 variáveis e horizonte de previsão de
1h
Figura 16 - Regressão para treinamento com modelo NARX, 4 variáveis e horizonte de
previsão de 1h
Figura 17 – Autocorrelação do erro para treinamento com modelo NARX, 4 variáveis e
horizonte de previsão de 1h
Figura 18 – Previsão do BTP com horizonte de previsão de 1h com modelo NAR35
Figura 19 – Regressão para o treinamento NAR
Figura 20 – Autocorrelação do erro para o treinamento NAR

LISTA DE TABELAS

Tabela 1 - Potenciais variáveis de entrada do modelo	21
Tabela 2 – Características dos treinamentos com poda2	26
Tabela 3 – Levantamento das variáveis selecionadas pelo algoritmo de poda em cada teste 2	29
Tabela 4 – Características dos testes da rede NARX com 11 variáveis exógenas	\$2
Tabela 5 – Características dos testes da rede NAR para 20 dias diferentes de previsão com o)
mesmo treinamento	37

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
BTP	Burn-Through Point
IEEE	Institute of Electrical and Electronics Engineers - Instituto de Engenheiros
	Eletricistas e Eletrônicos
MSE	Mean Squared Error – Erro Quadrático Médio
NAR	Nonlinear AutoRegressive Model – Modelo não linear autorregressivo
NARX	Nonlinear AutoRegressive Model with eXogenous inputs - Modelo não linear
	autorregressivo com entradas exógenas
NOE	Nonlinear Output Error Model – Modelo não linear de erro de saída
RNA	Rede Neural Artificial
UFES	Universidade Federal do Espírito Santo

LISTA DE SÍMBOLOS

Peso sináptico referente ao n-ésimo sinal de entrada no k-ésimo neurônio
Limiar correspondente ao neurônio k (bias)
Função de ativação
Saída da junção aditiva do neurônio k
Sinal de saída do neurônio k
<i>n</i> -ésima entrada
Saída estimada do modelo no instante k
Número de atrasos da memória de saída
Número de atrasos da memória de entrada
Erro Quadrático Médio
Função de autocorrelação do erro

SUMÁRIO

1	INTRODUÇÃO	9
	1.1 Apresentação e objeto de pesquisa	9
	1.2 Objetivo geral e objetivos específicos	10
	1.3 Organização do trabalho	11
2	EMBASAMENTO TEÓRICO	12
	2.1 Redes Neurais Artificiais	12
	2.1.1 Algoritmo de poda	18
	2.2 Tratamento dos Dados	19
	2.2.1 Detecção de <i>outliers</i>	19
	2.2.2 Tratamento de <i>outliers</i>	20
3	CASO EM ESTUDO	21
	3.1 Principais variáveis do processo	21
4	DESENVOLVIMENTO E RESULTADOS	24
	4.1 Análise dos sinais	24
	4.2 Treinamento e poda com todas as variáveis	25
	4.3 Treinamento para as variáveis selecionadas pela fase anterior	29
5	CONCLUSÕES E TRABALHOS FUTUROS	
R	EFERÊNCIAS BIBLIOGRÁFICAS	40

1 INTRODUÇÃO

1.1 Apresentação e objeto de pesquisa

É de interesse da indústria, em um panorama de cada vez maior preocupação com eficiência das técnicas e redução de custo, a melhor utilização dos dados de processo obtidos para melhor tomada de decisões estratégicas. Isso não é diferente para as usinas siderúrgicas, nas quais é produzido o aço a partir do refino do ferro-gusa. O ferro-gusa, ou gusa, é formado nos altos-fornos, onde ocorre a redução do minério de ferro, sendo usado para isto diferentes materiais que são adicionados em mistura com o minério para promover essa redução.

Dentre estes produtos, o sínter é tido como o melhor provedor de matéria necessária ao processo de produção de ferro. A produção do sínter, conhecida como sinterização, requer atenção especial, pois além de servir para o reaproveitamento de finos da produção do ferro, ele demanda uma grande área de operação bem como tempo para seu preparo.

O processo de sinterização corresponde a um pré-processamento para o minério de ferro e fundentes (sílica e cal, por exemplo), cargas do alto-forno que tem grande efeito sobre as condições de funcionamento dele e, consequentemente, sobre a produção do ferro-gusa e de aço. O resultado final deste processo é o sínter, material que substitui o minério granulado, constituído de hematita compacta, que possui maior teor de óxido de ferro e está em processo de extinção (MOURA, 2008).

Existem basicamente duas estruturas de produção de sínter: a Máquina de Greenawalt e a Máquina de Dwight-Lloyd. Esta última é normalmente utilizada para a maioria dos processos industriais de sinterização (Figura 1).



Fonte: Adaptado de (USAMENTIAGA, 2013).

Na Dwight-Lloyd o calor é aplicado a partículas finas de minério de ferro para transformá-las em produto de granulação maior. Além das partículas de minério, a mistura bruta do sínter é composta também de coque e uma mistura de retorno (finos que não foram aproveitados logo depois do primeiro processamento). Essa mistura é umidificada e carregada em uma esteira em movimento e nivelada para formar uma camada (leito). A esteira é feita de uma rede de metal e refratário, que permite que um grande exaustor sugue o ar através do leito para as chamadas caixas de vento, onde a temperatura do ar sugado é medida, a matéria volátil é expelida e a mistura bruta é fundida para formar o sínter. Quando o sínter atinge o fim da esteira, ele é triturado e os produtos de tamanho aceitável são levados ao alto-forno, enquanto os de tamanho menores são refinados para reprocessamento (retorno).

1.2 Objetivo geral e objetivos específicos

O objeto do presente trabalho é a implementação de um sistema de predição para uma das variáveis do processo de sinterização, o *Burn Through Point* (BTP), com intuito de produzir estimativas para o BTP, e, assim, balizar as decisões dos operadores e possibilitar aumento da assertividade em relação ao processo de produção do sínter.

Os objetivos específicos deste trabalho são:

- Estudo e seleção de técnicas de análises de dados e de processamento de sinais em 36 variáveis do processo de sinterização;
- Seleção das variáveis do processo de sinterização mais relevantes aos modelos de solução do problema;
- Implementação-teste de propostas de solução para a predição do BTP.

1.3 Organização do trabalho

Este trabalho está dividido em 5 capítulos, onde o primeiro deles apresenta o contexto do problema, as justificativas para o desenvolvimento do trabalho e os objetivos do mesmo. O Capítulo 2 apresenta uma revisão bibliográfica dos métodos adotados para realização deste projeto. No Capítulo 3, são discutidos os aspectos do caso em estudo, enquanto no Capítulo 4 apresenta-se as metodologias desenvolvidas para análise das variáveis e, para cada uma das etapas do trabalho, seus resultados e discussão deles. As conclusões obtidas encontram-se no Capítulo 5.

2 EMBASAMENTO TEÓRICO

É importante para qualquer estudo o levantamento das ferramentas que podem ser utilizadas no desenvolvimento da solução desejada. Nesse sentido, este capítulo apresenta uma breve revisão bibliográfica dos métodos usados neste projeto.

2.1 Redes Neurais Artificiais

Segundo Monard e Baranauskas (2003), Aprendizado de Máquina (AM) é uma área de Inteligência Artificial que tem como objetivo tanto desenvolver técnicas computacionais sobre o aprendizado como também construir sistemas que sejam capazes de adquirir conhecimento de modo automático. A Rede Neural Artificial (RNA) é um dos modelos computacionais para tais finalidades.

Uma Rede Neural Artificial, usualmente denominada "rede neural" é um sistema inspirado no funcionamento extremamente não linear, complexo e paralelo que ocorre no cérebro. Ela é constituída de unidades de processamento básicas, chamadas de neurônios, que são interligados e propensos a acumular conhecimento experimental, a partir de um processo de aprendizagem, e a torná-lo disponível para uso (HAYKIN, 2001).

Na Figura 2 é representada a estrutura básica de um neurônio típico: ele possui um conjunto de *n* sinais de entrada $(u_1, u_2, ..., u_n)$ e seus respectivos pesos sinápticos $w_{k,n}$, em que *k* é o índice do neurônio ao qual o peso se refere. Além disso, b_k é o limiar correspondente ao neurônio *k* (bias), $\varphi(\cdot)$ é a função de ativação e y_k é o sinal de saída do neurônio *k*.



Fonte: Adaptado de (ZANETTI, 2007).

A ação de um neurônio pode ser descrita pela Equação (1):

$$y_{k} = \varphi(\cdot) \left(\sum_{j=1}^{n} u_{j} \cdot w_{k,j} v_{k} + b_{k} \right).$$
(1)

Quanto a sua estrutura, as redes neurais são organizadas em camadas. Na Figura 3a, está ilustrada a chamada *rede de camada única*, com 4 neurônios e 4 nós de entrada. Nessa topologia de rede, em apenas uma camada ocorre alguma computação. A complexidade das redes neurais cresce conforme aumenta o número de camadas, as chamadas *camadas ocultas*, como ilustrada na rede da Figura 3b.

Figura 3 – Redes Neurais Artificiais alimentadas adiante (*Feedfoward*): a) Com camada única; b) Com múltiplas camadas.



Fonte: Adaptado de (HAYKIN, 2001).

As redes neurais apresentadas na Figura 3 são caracterizadas como alimentadas adiante, também chamada de *feedfoward*, pois a camada de entrada é projetada na camada de saída de neurônios, e não vice-versa (HAYKIN, 2001). Na Figura 4 são apresentadas Redes Recorrentes, que se diferenciam das Redes *Feedfoward* por apresentarem pelo menos um laço de

realimentação, possibilitando que a saída também seja influenciada por informações passadas (DINIZ, 2018).



Figura 4 - Rede Neural Artificial Recorrente com neurônios ocultos

Fonte: Adaptado de (HAYKIN, 2001).

Redes recorrentes podem ser aplicadas através de um modelo não linear autorregressivo com entradas exógenas (*Nonlinear AutoRegressive model with eXogenous inputs* – NARX), com a inclusão de operadores de atraso também em sua saída. Assim, a saída do modelo NARX pode ser representado pela Equação (2):

$$\hat{y}(k) = \varphi \left[u(k-1), u(k-2), \dots, u(k-n_u), y(k-1), y(k-2), \dots, y(k-n_y) \right]^T.$$
(2)

Nela, $\hat{y}(k)$ é a saída estimada do modelo no instante k, u(k) representa a entrada e y(k), a saída do modelo, n_y é o número de atrasos da memória de saída, n_u é o número de atrasos da memória de entrada e $\varphi(\cdot)$ representa a função de ativação.

Quando usado somente os valores passados da saída para a previsão dos valores futuros, temse o modelo não linear autorregressivo (*Nonlinear AutoRegressive model* – NAR), representado na Equação (3), considerado uma simplificação do modelo NARX (DINIZ, 2018):

$$\hat{y}(k) = \varphi [y(k-1), y(k-2), ..., y(k-n_y)]^T.$$
 (3)

Similar ao modelo NARX, há também o modelo não linear de erro de saída (*Nonlinear Ouput Error* – NOE), que inclui predições passadas em vez de saídas medidas, representado na Equação (4):

$$\hat{\mathbf{y}}(k) = \varphi \left[u(k-1), u(k-2), \dots, u(k-n_u), \hat{\mathbf{y}}(k-1), \hat{\mathbf{y}}(k-2), \dots, \hat{\mathbf{y}}(k-n_y) \right]^T.$$
(4)

Na Figura 5 estão representados em diagrama de blocos os modelos citados.



Figura 5 - Diagrama de blocos dos modelos: a) NARX; b) NAR; c) NOE

(c) Fonte: Adaptado de (RANKOVIĆ e NIKOLIĆ, 2008).

 $u(k - n_u)$

Modelo NOE

O processo de aprendizagem de uma rede neural ocorre por indução e pode ser classificado, de forma simplificada, em supervisionado, não supervisionado e por reforço. Ao aprendizado supervisionado é fornecido um conjunto de exemplos de entrada-saída, que possibilita uma resposta desejada, a partir de treinamento e classificação desses. É proporcionado um rótulo de categoria ou custo para cada padrão em um conjunto de treinamento e é almejado reduzir a soma dos custos desses padrões (DUDA; HART; STORK, 2012). Esse custo, neste trabalho, é representado pelo Erro Quadrático Médio (*Mean Square Error – MSE*), conforme a Equação (5):

$$\varepsilon = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$
(5)

em que *N* é o número de amostras para treinamento, y_i é a saída do sistema real e \hat{y}_i é a saída do modelo de predição.

O desafio a superar no aprendizado supervisionado é a generalização, ou seja, a capacidade de produzir saídas adequadas para entradas diferentes daquelas usadas durante a aprendizagem (HAYKIN, 2001). Nesse contexto, o sobre ajuste (ou *overfitting*) é um termo que caracteriza a falta de generalização de um modelo, visto que ele descreve quando um modelo se adequa em demasia aos dados de treinamento.

A fim de melhorar a generalização, a técnica de *early stopping* é utilizada neste trabalho. Esta técnica divide o conjunto de dados em 3 subconjuntos: de treinamento, de validação e de teste. O primeiro subconjunto é usado para atualizar os pesos sinápticos e a polarização e o segundo tem o erro calculado e monitorado durante o treinamento. Na fase inicial, esse erro tende a diminuir e quando ocorre *overfitting*, a aumentar. Quando isso ocorre, o treinamento é encerrado e os pesos sinápticos e as polarizações referentes ao menor erro de validação são retornados ao algoritmo da rede. O terceiro subconjunto é usado para comparação entre modelos e pode indicar uma má distribuição dos dados nos subconjuntos.

Com o intuito de verificar se o modelo de uma rede neural é adequado, é possível, dentre outras, empregar a técnica de regressão e a análise dos coeficientes da autocorrelação do erro. A técnica de regressão é utilizada a fim de detalhar a resposta da rede neural em questão, na qual se compara as saídas da rede neural (valores da predição) e os valores esperados. A partir dela, é ilustrada a relação entre esses dois valores, na qual o melhor ajuste é indicado pela linha contínua e os dados são indicados por círculos, conforme mostrado na Figura 6. Quanto mais próximos esses círculos se apresentam da linha, o coeficiente de correlação é muito próximo de 1 e, por isso, indica bom ajuste.





A análise do modelo por meio dos coeficientes de autocorrelação do erro é baseada na função de autocorrelação do erro $R_{\varepsilon\varepsilon}$ entre o instante t_k e seu atraso (*lag*) t_{k-1} pode ser calculada por meio da Equação (6):

$$R_{\varepsilon\varepsilon}(t_{k-1}, t_k) \triangleq E[\varepsilon(t_{k-1})\varepsilon^*(t_k)]$$
(6)

em que E é o valor esperado entre o MSE (ε) e seu conjugado.

Quando os coeficientes são nulos, então pode-se afirmar que não há correlação entre o erro nesses instantes e que, portanto, o modelo da rede neural é apropriado (TÁPIA, 2000). Para casos práticos, indica-se um intervalo de confiança dentro do qual pode-se afirmar que o modelo se apresenta satisfatoriamente, conforme ilustrado na

Fonte: O próprio autor.



Figura 7 – Exemplo de representação da autocorrelação do erro

Fonte: O próprio autor.

As redes neurais se destacam por serem técnicas de modelagem de dados estatísticos não lineares, que fazem parte de sistemas com relações complexas entre entradas e saídas, a fim de, principalmente, encontrar padrões. Desde 1986, elas vêm sido aplicadas para a identificação de sistemas dinâmicos não lineares (RANKOVIĆ e NIKOLIĆ, 2008).

Em conjunto com outras técnicas, como os Algoritmos Genéticos, essa é a mais usada para a predição do BTP, como pode ser observado nos trabalhos de Er, Liao e Lin (2000), Wushan (2005), Laitinen e Saxén (2007) e Wang (2014).

2.1.1 Algoritmo de poda

No contexto de aplicações reais, as redes neurais apresentam tamanho grande, para que possuam graus de liberdade suficientes para acomodar rapidamente as características gerais dos dados de entrada, de uma forma pouco sensível às condições iniciais e aos mínimos locais (ESPENCHITT, 2001).

Tendo isso em vista, é desejável minimizar o tamanho da rede mantendo bom desempenho e uma alternativa para isso é a utilização da poda da rede. O algoritmo de poda deve atuar em uma rede neural inicialmente grande e então reduzir ou eliminar certos pesos sinápticos de forma seletiva e ordenada (HAYKIN, 2001).

A estratégia para poda das redes neurais utilizadas neste trabalho é a de aplicação de algoritmos fundamentados no cálculo da matriz hessiana (NORGAARD *et al.*, 1996). Para esta técnica, busca-se, por meio do cálculo da sensibilidade (ou saliência) de um peso sináptico, uma relação entre a complexidade da rede e o desempenho do erro de treinamento (HAYKIN, 2001).

O tipo de poda empregado utiliza um modelo local da superfície do erro para realizar uma previsão analítica do efeito no erro da saída da rede quando determinado peso sináptico é retirado (DINIZ, 2018). No método intitulado Cirurgião Cerebral Ótimo, o procedimento consiste em treinar uma arquitetura de rede totalmente conectada, calcular a sensibilidade de cada peso sináptico utilizando a matriz hessiana, eliminar aqueles com menor sensibilidade e atualizar os pesos e treinar novamente a rede. Esse método é encerrado quando mais nenhum peso sináptico puder ser retirado da rede sem que isso provoque um grande aumento no erro médio quadrático.

2.2 Tratamento dos Dados

A presença de valores atípicos (*outliers*) em um conjunto de dados pode influenciar diretamente na homogeneidade de séries (DINIZ, 2018). Esse fenômeno ocorre por diversos fatores, como falhas de medição comuns em processos industriais, e deve ser tratado de forma que torne os dados consistentes para a continuidade do projeto.

2.2.1 Detecção de outliers

A ausência de tendência no comportamento de uma série é explicada por erros de medições que devem ser tratados como valores discrepantes, ou *outliers*. Para identificação deles, considerouse aqueles cujos valores são maiores que 3 vezes o desvio absoluto mediano (DAM), calculado pela Equação (7), em que N é a quantidade de amostras.

$$DAM = \frac{-1}{\sqrt{2}*(-0.4769)} * mediana(|A_i - mediana(A)|), i = 1, 2, \dots N.$$
(7)

2.2.2 Tratamento de *outliers*

A fim de tornar tais amostras coerentes, elas são substituídas pela interpolação dos valores vizinhos não considerados *outliers*. A fim de manter o comportamento da série, neste trabalho foram utilizadas a interpolação linear ou a interpolação de Hermite seccionalmente cúbica (*pchip*).

3 CASO EM ESTUDO

Durante todo o processo de produção do sínter ocorrem várias medições com objetivo de monitorar e controlar todo o processo da produção do sínter. Dentre essas medições é de fundamental interesse ter a informação de onde ocorre o BTP, ou seja, do ponto em que ocorre a maior temperatura ao longo da esteira em que a mistura, que formará o sínter, passa.

A relevância se deve ao fato de que o BTP indica em que local, ao longo da máquina de sínter, o calor cedido pelo forno de ignição penetrou toda a camada de material da mistura e houve a adequada fusão dela. Para que isso ocorra, é necessário cuidadoso ajuste, de forma que esse local indicado pelo BTP não seja muito anterior ao fim da máquina de sínter, prejudicando a produtividade, nem posterior a ela, danificando os equipamentos resfriadores instalados no processo. Na usina em estudo, a máquina de sínter é um equipamento de importância considerável, possuindo 424 metros quadrados de área e 8 caixas de vento.

3.1 Principais variáveis do processo

A partir da avaliação dos engenheiros especialistas da Sinterização, foram selecionadas 36 variáveis potencialmente significativas para a predição do BTP, apresentadas na Tabela 1. Além disso, são informadas a unidade de medida da variável e a periodicidade de amostragem com que cada variável era armazenada no sistema de monitoramento da usina.

N°	Variável de Entrada	Unidade de medida	Amostragem
1	Geração de Finos de Retorno	ton/dia	1h
2	Produtividade	ton/dia*m ²	6h
3	Altura da Camada	mm	6h
4	Retorno	%	1h
5	Densidade da Mistura	t/m³	1h
6	Velocidade de Sinterização	mm/min	1h
7	Temperatura da Caixa de Vento nº11	°C	6h
8	Temperatura da Caixa de Vento nº13	°C	6h
9	Temperatura da Caixa de Vento nº18	°C	1h

Tabela 1 - Potenciais variáveis de entrada do modelo

10	Temperatura da Caixa de Vento nº19	°C	1h
11	Temperatura da Caixa de Vento nº20	°C	1h
12	Temperatura da Caixa de Vento nº21	°C	1h
13	Temperatura da Caixa de Vento nº22	°C	1h
14	Temperatura da Caixa de Vento nº23	°C	1h
15	Pressão na Caixa de Vento nº22	mmH ₂ O	1h
16	Pressão na Caixa de Vento nº23	mmH ₂ O	1h
17	Velocidade da Máquina de Sínter	m/min	1h
18	Temperatura BTP	°C	1h
19	Vazão do Exaustor Nº 1	Nm ³ /min	1h
20	Vazão do Exaustor Nº 2	Nm ³ /min	1h
21	Pressão do Exaustor	mmH ₂ O	1h
22	СО	mg/Nm ³	1h
23	Cal	%	24h
24	Base de Coque	%	24h
25	Pilha Homogeneizada A	ton	1h
26	Pilha Homogeneizada B	ton	1h
27	Total Pilha Homogeneizada	ton	1h
28	RDEG na Pilha	%	24h
29	SFBR no sínter feed	%	24h
30	STFA no sínter feed	%	24h
31	+1mm	%	24h
32	Relação +1/-0,105 na Pilha	-	24h
33	Fração +1mm	-	24h
34	Fração -0,105mm	-	24h
35	Relação +1/-0,105mm na Mistura	-	24h
36	Total RDEG	%	24h

Fonte: Produção do próprio autor.

A Geração de Finos de Retorno é a quantidade de material que deverá ser reutilizada no processo de sinterização, pois foi totalmente agregada ao sínter produzido, enquanto a variável Retorno é a porcentagem dessa quantidade em relação a mistura total. A produtividade referese à quantidade de sínter produzido por dia, dividido pela área da máquina de sínter. A

velocidade de sinterização é a velocidade aproximada com que o calor se move pela altura da camada, provocando a sinterização da mistura. Diferentemente, a velocidade da máquina de sínter é a velocidade com que a máquina move a esteira. A variável Temperatura BTP é a temperatura estimada no ponto que ocorre a queima total. CO é a quantidade de monóxido de carbono produzida na reação que ocorre por conta do forno de ignição.

As variáveis Cal, Coque, Fração +1mm, Fração -0,105mm, Relação +1/-0,105mm na Mistura e Total RDEG são calculadas em relação à mistura total, ou seja, à mistura composta pela pilha homogeneizada e pelo sínter *feed*, enquanto as variáveis RDEG na Pilha, +1mm e Relação +1/-0,105 na Pilha são referentes à composição da pilha homogeneizada somente.

RDEG é a sigla referente aos finos provenientes do peneiramento da mistura, que são reutilizados na composição de uma pilha. SFBR e STFA são siglas utilizadas para o minério proveniente das minas de Brucutu e de Fábrica, respectivamente. Em conjunto com outras fontes de minério, formam o sínter *feed*.

A granulometria é um parâmetro que afeta a permeabilidade do leito e, consequentemente, a produtividade (HONORATO, 2015). Por isso, foram elencadas as variáveis +1mm, Relação +1/-0,105 na Pilha, Fração +1mm, Fração -0,105mm e Relação +1/-0,105mm na Mistura.

4 DESENVOLVIMENTO E RESULTADOS

Para o desenvolvimento deste estudo, foram utilizados dados provenientes de uma planta de sinterização em siderúrgica localizada na região da Grande Vitória, Espírito Santo, coletados durante o período de 24/08/2017 a 19/08/2018.

Além do BTP, foram captados dados de 36 variáveis que descrevem o processo de sinterização e que, baseado na análise de especialistas da área da empresa, possivelmente mais o influenciam. Todos eles foram tratados, para supressão de períodos em que houve anomalias, paradas no processo ou ausência de medições e para correção de *outliers*.

Após análise preliminar dos sinais compostos pelos dados coletados, foram utilizadas redes neurais artificiais com a aplicação de um algoritmo de poda, de forma a selecionar as variáveis mais relevantesPara isso, foram utilizados algoritmos adaptados de um *toolbox* para MATLAB, desenvolvido por Magnus Nogaard e disponibilizado para download no site da Universidade Técnica da Dinamarca (THE NNSYSID, 2003). De posse das variáveis selecionadas pelo algoritmo de poda, foram novamente avaliados modelos de predição usando redes neurais e seus resultados finais são discutidos.

4.1 Análise dos sinais

Em virtude das diferenças entre os períodos de amostragem de cada variável de entrada, conforme mostrado na Tabela 1, foi feita uma reamostragem dos dados, de forma que todos apresentassem valores médios a cada hora. Para isso, foi utilizada a replicação dos valores, tendo em vista que, devido às características do processo, em geral, estes parâmetros não sofrem grandes variações em curto prazo.

Foram plotadas as variáveis a fim de avaliar visualmente algum aspecto de correlação entre variáveis. Na Figura 8, apresentam-se todas as amostras das variáveis de entrada ao longo do tempo, após a exclusão dos períodos de anomalias no processo. As amostras do BTP selecionadas foram aquelas referentes aos mesmos períodos em que as demais variáveis do processo não apresentaram desvios.



Figura 8 – Exemplo do comportamento de 4 variáveis do processo de sinterização estudado: a) BTP; b) Densidade da mistura; c) Temperatura na caixa de vento nº 19; d) Velocidade da máquina de sínter

Fonte: Produção do próprio autor.

Por meio da análise visual apenas não foi possível verificar correlação direta cada uma das variáveis exógenas e o BTP. Dessa forma, nenhuma das 36 variáveis foi descartada para as etapas posteriores.

4.2 Treinamento e poda com todas as variáveis

Nesta seção são apresentados os resultados de testes com redes de topologias NNOE e NARX e posterior aplicação do Algoritmo de Poda. Esse algoritmo será utilizado para definir as variáveis mais relevantes para treinamento de uma rede neural que visa a predição do BTP.

Todas as redes possuirão uma camada de entrada, uma camada oculta e uma camada de saída única. A camada de entrada terá tamanho fixo igual ao número de variáveis de entrada e o tamanho da camada oculta será determinado testando-se redes com estruturas variando de 4, 5, 10 e 30 neurônios ocultos. Para cada rede, será calculado o MSE obtido. Os resultados de cada treinamento são apresentados na Tabela 2, ordenados em ordem decrescente de MSE durante o teste.

Nos testes da Tabela 2 não foram tratados os *outliers*, pois o objetivo é verificar a complexidade do comportamento das variáveis exógenas em relação ao BTP.

Nº da		Quantidade de			
IN UU	Topologia	lopologia neurônios na		MSE Teste	
leste		camada oculta			
1	NARX	4	0,14	0,55	
2	NARX	4	0,22	0,55	
3	NARX	4	0,14	0,7	
4	NOE	10	0,17	0,71	
5	NOE	4	0,2	0,73	
6	NOE	30	0,03	0,8	
7	NOE	10	0,17	0,84	
8	NOE	4	0,21	0,87	
9	NOE	30	0,03	0,97	
10	NOE	10	0,28	1,03	
11	NOE	30	0,124	1,103	
12	NOE	5	0,124	1,103	
13	NOE	5	0,124	1,103	
14	NOE	4	0,124	1,103	
15	NOE	4	0,124	1,103	
16	NARX	30	0,03	1,25	
17	NOE	30	0,04	1,81	
18	NOE	30	0,04	1,81	
19	NOE	30	0,04	1,81	
20	NARX	4	0,14	0,55	

Tabela 2 - Características dos treinamentos com poda

Fonte: Produção do próprio autor.

A fim de ilustrar o resultado dos testes, toma-se como exemplo o teste nº 1. Utilizando o modelo NARX de rede neural com 4 neurônios na camada oculta, foi realizado o treinamento e na

Figura 9 é mostrada saída para a etapa de teste da rede neural. O erro de predição é dado pela diferença entre o valor da saída real (esperado) e o valor da saída do modelo. Na Figura 9, notase que o modelo da rede não obteve bom desempenho, visto que os erros de predição foram altos e, em alguns pontos, a saída do modelo não correspondeu ao comportamento da saída real.





Fonte: O próprio autor.

Logo após esse teste, foi aplicado do algoritmo de poda e realizado novo teste, cujos resultados são apresentados na Figura 10. É possível notar que o modelo foi capaz de prever a tendência do comportamento do BTP, no entanto, os erros de predição foram ainda altos.



Figura 10 - Previsão do BTP após poda: a) Valores a cada amostra; b) Erro de predição a cada amostra

Fonte: O próprio autor.

Para cada um dos 20 testes, foram guardadas as variáveis cujos pesos sinápticos não foram iguais a zero ao fim da aplicação do algoritmo de poda. Isso indica que tais variáveis não foram eliminadas e, por isso, apresentam maior relevância para a rede neural. Na Tabela 3 foram elencadas aquelas variáveis que apresentaram mais de 1 ocorrência durante os testes. Essas 11 variáveis restantes foram separadas e utilizadas na etapa seguinte.

	Quantidade									Nº	do t	este	da 🛛	Гаbe	ela 2						
	de ocorrências		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Variável 13	5																				
Variável 15	5																				
Variável 10	4																				
Variável 11	4																				
Variável 1	3																				
Variável 9																					
Variável 12	2																				
Variável 14	2																				
Variável 17	2																				
Variável 18	2																				
Variável 19	2																				

Tabela 3 - Levantamento das variáveis selecionadas pelo algoritmo de poda em cada teste

Fonte: Produção do próprio autor.

4.3 Treinamento para as variáveis selecionadas pela fase anterior

De posse das 11 variáveis de entrada e da variável de saída, foram tratados os *outliers* de forma que as correlações e tendências no conjunto dos dados sejam mais facilmente detectadas pela rede neural. Na Figura 11, observa-se a variável de saída, o BTP, antes e após o tratamento dos valores considerados *outliers*.



Figura 11 - Variável de saída antes e após tratamento de outliers

Fonte: O próprio autor

Além disso, tanto para as variáveis exógenas quanto para o BTP foram inseridas amostras entre cada par de amostras do sinal original após a aplicação da interpolação de Hermite seccionalmente cúbica. Isso foi feito com o intuito de suavizar o sinal e possibilitar melhor aprendizado a rede neural.

Nos treinamentos realizados foram variados os seguintes parâmetros a fim de verificar a consistência do modelo da rede neural: quantidade de amostras adicionadas na interpolação, número de neurônios na camada oculta e número de atrasos, além de quais variáveis seriam utilizadas.

Empregando as 11 variáveis apresentadas na Tabela 3, foi feito um treinamento com modelo NARX, interpolação de 7 amostras, 3000 amostras iniciais, 30 neurônios na camada oculta e 21 atrasos (correspondendo, portanto, a 3 amostras reais, tendo em vista que as outras eram provenientes da interpolação). O resultado da predição do BTP utilizando a rede neural treinada com tais características é apresentada na Figura 12, na Figura 13, a regressão e na Figura 14, a autocorrelação do erro referentes a estre treinamento.



Figura 12 - Previsão do BTP com modelo NARX, 11 variáveis e horizonte de previsão de 1h



Fonte: O próprio autor.



Figura 13 - Regressão para treinamento com modelo NARX, 11 variáveis e horizonte de previsão de 1h

Figura 14 - Autocorrelação do erro para treinamento com modelo NARX, 11 variáveis e horizonte de previsão

de 1h



Fonte: O próprio autor.

As Figuras 12 a 14 mostram que o modelo provém um bom resultado, devido ao valor da regressão bastante próximo de 1 e a autocorrelação do erro dentro do intervalo de confiança para alguns *lags*. A fim de verificar, foram feitos testes com base nas 20 horas anteriores e horizontes de previsão igual a 1h e a 2h, cujos resultados são apresentados na Tabela 4. O erro médio quadrático apresentado nesses testes foi de 5,06%.

Nº do tosto	Erro Absoluto [%]						
IN UU LESLE	Horizonte de previsão de 1h	Horizonte de previsão de 2h					
1	0,81	0,17					
2	0,06	0,39					
3	4,55	3,75					
4	-0,49	-0,06					
5	-0,82	2,78					
6	0,31	2,49					
7	2,42	2,38					
8	2,65	4,21					
9	2,80	4,99					
10	0,51	0,74					
11	-3,34	-4,29					
12	0,51	-1,21					
13	0,92	1,58					
14	-3,63	-3,39					
15	1,25	-0,39					
16	-0,93	1,12					
17	-1,05	-1,20					
18	3,77	2,64					
19	3,07	5,96					
20	-2,31	1,09					

Tabela 4 - Características dos testes da rede NARX com 11 variáveis exógenas

Fonte: O próprio autor.

O erro quadrático para horizonte de previsão de 1h foi de 5,06%, enquanto para 2h foi de 218,46%. Isso demonstra a dificuldade para generalização do modelo, visto que eram utilizadas amostras diferentes a cada teste.

A fim de empregar menor quantidade de variáveis exógenas e possivelmente melhorar a generalização, foi testado novo modelo NARX, desta vez com as 4 mais recorrentes: as variáveis 13, 15, 10 e 11, que corresponderam a mais de 54% das ocorrências, conforme Tabela 3. Esse modelo contou com 10 amostras, 30 neurônios na camada oculta e 30 atrasos. O resultado da predição do BTP utilizando a rede neural treinada com tais características e com 5000 amostras iniciais apresenta-se na Figura 15. A regressão está representada na Figura 16 e a autocorrelação do erro do mesmo treinamento é mostrada na Figura 17.





Fonte: O próprio autor.





Figura 17 - Autocorrelação do erro para treinamento com modelo NARX, 4 variáveis e horizonte de previsão de



É importante analisar das Figuras 15 a 17 que, ainda que a regressão tenha apresentado bom resultado, a autocorrelação do erro não ficou dentro dos limites do intervalo de confiança para

1h

nenhum dos *lags* apresentados. Isso demonstra que a rede neural treinada com tais parâmetros interpolou bem os dados, mas não foi capaz de generalizar para o horizonte de previsão de 1h, como mostrou a Figura 12.

Um treinamento feito apenas com a série do BTP, usando uma rede NAR, que contou com interpolação de 7 amostras, 40 neurônios na camada oculta e 21 amostras de atraso (ou seja, 3 amostras reais, tendo em vista que as outras eram provenientes da interpolação), com 3000 amostras iniciais resultou na predição mostrada na Figura 18. A regressão está representada na Figura 19 e autocorrelação do erro do é mostrada na Figura 20.



Figura 18 - Previsão do BTP com horizonte de previsão de 1h com modelo NAR

Fonte: O próprio autor.

Figura 19 - Regressão para o treinamento NAR



Fonte: O próprio autor.





Pode-se observar que a ausência de variáveis exógenas, característica do modelo NAR, possibilitou maior assertividade da rede neural. Ainda que com menor valor de regressão na

fase de teste, a autocorrelação do erro foi mais aceitável, em relação às redes NARX, para o mesmo horizonte de previsão.

Assim como para o modelo NARX, averiguou-se a assertividade deste modelo por meio de testes feitos com base nas 20 horas anteriores, para horizonte de previsão igual a 1h e para horizonte de previsão igual a 2h, cujos resultados são apresentados na Tabela 5. O erro médio quadrático para horizonte de previsão de 1h apresentado foi de 3,42%, enquanto para 2h foi de 7,87%.

	Nº do tosto	Erro Abs	oluto [%]
	IN UU LESLE	Horizonte de previsão de 1h	Horizonte de previsão de 2h
-	1	0,55	0,17
	2	-0,24	0,39
	3	3,07	3,75
	4	0,85	-0,06
	5	0,97	2,78
	6	-0,78	2,49
	7	1,57	2,38
	8	2,49	4,21
	9	1,67	4,99
	10	-0,22	0,74
	11	-3,28	-4,29
	12	-0,24	-1,21
	13	0,97	1,58
	14	-3,76	-3,39
	15	0,28	-0,39
	16	0,64	1,12
	17	-0,45	-1,20
	18	2,95	2,64
	19	2,91	5,96
	20	-1,07	1,09

Tabela 5 – Características dos testes da rede NAR para 20 dias diferentes de previsão com o mesmo treinamento. ---

. .

.

.....

Fonte: Produção do próprio autor.

Como esperado, a partir de inferências sobre as Figuras 18, 19 e 20, a rede neural artificial com modelo NAR se mostrou mais hábil em generalizar a partir das amostras apresentadas. O erro quadrático médio manteve-se abaixo de 8% para ambos os horizontes de previsão estudados, bem abaixo que aqueles apresentados pela rede NARX com 11 variáveis. Isso mostra que o modelo NAR se configura como uma boa alternativa para a predição do BTP.

5 CONCLUSÕES E TRABALHOS FUTUROS

Anterior à aplicação de técnicas de análise de dados, foi feito estudo do processo de sinterização e feito com os especialistas da área o levantamento das possíveis variáveis que mais impactassem no BTP, sendo esta uma variável de grande criticidade ao processo. Foram estudadas também as relações entre essas variáveis.

Com os dados brutos coletados do sistema de supervisão da empresa, foi feita uma análise preliminar visual das 36 variáveis elencadas. Tendo em vista a quantidade de variáveis, não foi possível encontrar correlações diretas entre elas e, portanto, optou-se por não excluir nenhuma. Dessa forma, a utilização do algoritmo de poda tornou-se necessário a fim de indicar que variáveis possivelmente mais influenciavam o BTP. Foi proposto utilizar RNA com algoritmo de poda a fim de selecionar as variáveis exógenas que influenciam fortemente no resultado da estimação do BTP.

Com os dados brutos, foram realizados 20 treinamentos com diferentes configurações para as redes NARX e NOE e eles demonstraram, após a poda, a recorrência de 11 variáveis, tomadas como as mais relevantes em relação ao BTP. Essas passaram por um processo de detecção de *outliers* e interpolação e foram feitos novos treinamentos com as 4 e com as 3 variáveis mais frequentemente selecionada pelo algoritmo de poda em uma rede NARX, além de treinamentos apenas com o modelo NAR. Todavia, o que se observou, foi que não houve uma boa correlação dessas variáveis exógenas da qual se pudesse concluir sua influência de maneira significante na previsão do BTP considerando o modelo de RNA utilizado. A rede NAR mostrou que apenas a série do BTP foi suficiente para obter-se previsão com baixo valor de erro.

Trabalhos futuros podem acrescentar análises mais profundas sobre as séries exógenas de dados, com objetivo de melhorar a detecção das relações entre essas variáveis e o BTP. Tendo em vista que o controle do BTP é uma importante medida para o processo de sinterização, podese considerar que um modelo que possibilite um horizonte de previsão de 1h é adequado ao processo e pode possibilitar maior assertividade do BTP e consequente maior qualidade da produtividade e redução de custos minimizando desperdícios.

REFERÊNCIAS BIBLIOGRÁFICAS

DINIZ, A. P. M. **Previsão do Conteúdo de Silício no Ferro-Gusa Usando Redes Neurais**. 2018. Dissertação (Mestrado em Engenharia Elétrica) – Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal do Espírito Santo, Vitória, 2018.

ER, M. J.; LIAO, J.; LIN, J. Fuzzy neural networks-based quality prediction system for sintering process. In: **IEEE Transactions on Fuzzy Systems**, v. 8, n. 3, p. 314-324, 2000.

ESPENCHITT, D. G., CUNHA, A. G. G., LACHTERMACHER, G. Cuidados no Uso do Algoritmo de Poda em Redes Neurais. In: **V Congresso Brasileiro de Redes Neurais**, p. 151–153, 2001.

FIRMINO, P. R. A. **Redes Bayesianas para a parametrização da confiabilidade em sistemas complexos**. 2004. Dissertação (Mestrado em Engenharia de Produção) - Programa de Pós Graduação em Engenharia de Produção, Universidade Federal de Pernambuco, Recife, 2004.

FRIEDMAN, N.; GEIGER, D.; GOLDSZMIDT, M. Bayesian network classifiers. In: **Machine learning**, v. 29, n. 2-3, p. 131-163, 1997.

GOODFELLOW, I., BENGIO, Y., COURVILLE, A. Deep learning. Cambridge: MIT press, 2016.

HAYKIN, S. Redes neurais: princípios e prática. 2ª. ed. Porto Alegre: Bookman, 2001.

HONORATO, E. P. Adequação granulométrica das matérias-primas e do sistema de segregação contínua (ISF), para melhorias na produtividade e qualidade do sínter para os altos-fornos. 2005. Dissertação (Mestrado em Engenharia Metalúrgica e de Minas) - Pós-graduação em Engenharia Metalúrgica e de Minas, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

LAITINEN, P. J.; SAXÉN, H. A neural network based model of sinter quality and sinter plant performance indices. In: **Ironmaking & steelmaking**, v. 34, n. 2, p. 109-114, 2007.

MENEGOTTO, J. L. **Arquitetura e Inteligência Artificial**. 2005. Dissertação (Mestrado em Ciências em Arquitetura) - Programa de Pós Graduação em Arquitetura, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005.

MENEZES JR, J. M. P; BARRETO, G. A. A new look at nonlinear time series prediction with NARX recurrent neural network. In: **2006 Ninth Brazilian Symposium on Neural Networks (SBRN'06)**. IEEE, 2006. p. 160-165.

MITCHELL, T. Machine Learning. 1. ed. Nova Iorque: McGraw Hill, 1997.

MOHAMED, Khaled Salah. Machine Learning for Model Order Reduction. Heliopolis: Springer, 2018.

MONARD, M. C., BARANAUSKAS, J. A. **Conceitos Sobre Aprendizado de Máquina**. Sistemas Inteligentes Fundamentos e Aplicações. 1 ed. Barueri-SP: Manole Ltda, p. 57-74;89-114, 2003.

MOURÃO, J. M. **Estudo Prospectivo do Setor Siderúrgico**: 2008-2025. Brasília: Centro de Gestão e Estudos Estratégicos, 2008. 50p. [Nota técnica]

NETO, J. M.; MOITA, G. C. Uma introdução à análise exploratória de dados multivariados. **Química nova**, v. 21, n. 4, p. 467-469, 1998.

NORGAARD, M. et al. The NNSYSID toolbox-a MATLAB (R) toolbox for system identification with neural networks. **Computer-Aided Control System Design, 1996.**, **Proceedings of the 1996 IEEE International Symposium on. IEEE**, 1996. p. 374-379.

OLIVEIRA, A. A. B. **Inventário das emissões atmosféricas na indústria siderúrgica**. 2014. Trabalho de Conclusão de Curso (Graduação em Engenharia Metalúrgica) – Escola Politécnica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.

ROJAS, R. Neural Networks: A Systematic Introduction. Berlin: Springer-Verlag, 1996.

TÁPIA, M. **Redes Neurais Artificiais: Uma Aplicação na Previsão de Preços de Ovos**. 2000. Dissertação (Mestrado em Ciências da Computação) - Programa de Pós Graduação em Ciências da Computação, Universidade Federal de Santa Catarina, Santa Catarina, 2000, p. 24. USAMENTIAGA, R. *et al.* Monitoring sintering burn-through point using infrared thermography. In: **Sensors**, v. 13, n. 8, p. 10287-10305, 2013.

WANG, J. *et al.* BTP prediction of sintering process by using multiple models. In: **Control and Decision Conference**, IEEE, p. 4008-4012, 2014.

WHITLEY, D. A genetic algorithm tutorial. In: **Statistics and computing**, v. 4, n. 2, p. 65-85, 1994.

WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. In: Chemometrics and intelligent laboratory systems, v. 2, n. 1-3, p. 37-52, 1987.

WUSHAN, C. The Application of Hybrid neural network for Burning Through Point Control. In: International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce. IEEE, p. 1160-1164, 2005.

RANKOVIĆ, V. M.; NIKOLIĆ, I. Z. Identification of Nonlinear Models with Feedforward Neural Network and Digital Recurrent Network. In: **FME Transactions**, v. 35, n. 2, p. 87-92, 2008.

ZANETTI, S. *et al.* Estimação da evapotranspiração de referência no Estado do Rio de Janeiro usando redes neurais artificiais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 12, n. 2, p. 174-180, 2008.