

APLICAÇÕES DE MODELOS ARIMA E REDES NEURAIIS PARA PREVISÃO DE DEMANDA EM CURTO PRAZO

Lee Vinagre Monteiro

Pontifícia Universidade Católica do Paraná
leemonteiro14@gmail.com

Marcia Gemari

Pontifícia Universidade Católica do Paraná
gemari.marcia@gmail.com

Resumo

A preocupação com preço de bens e serviços é uma constante no dia a dia organizacional. Gerir de forma correta entradas e saídas de produtos, a fim de evitar todo e qualquer desperdício, é um fator importante na redução de gastos de produção. Por isso, se faz pertinente a aplicação de métodos para prever a demanda, a fim de evitar prejuízos em estoques de matérias-primas e itens acabados. O presente trabalho tem como objetivo principal a aplicação de técnicas de previsão, a fim de prever a demanda de um produto x de uma indústria alimentícia. As técnicas utilizadas foram: (i) Redes Neurais Artificiais e (ii) Modelos ARIMA. Estas técnicas foram comparadas por de estatísticas de mensuração de erro e assim foi verificada a superioridade do modelo neural em comparação aos modelos ARIMA.

Palavras-Chave: Redes Neurais Artificiais; Modelos ARIMA; Previsão de demanda.

Abstract

A daily concern with prices of goods and services are a constant in the organizational environment. Manage properly the products inputs and outputs in order to avoid any waste are an important factor in the outlay reductions of the production. Therefore is relevant the use of methods for demand forecasting, to prevent losses with row material storage and finished products. This work has as main objective to apply prediction techniques in order to forecast the demand of a product x from a food factory. The techniques used were: (i) Neural Networks and (ii) ARIMA models. These techniques were apply and compared with statistical methods of error mensuration and were observed the superiority of the neural model in comparison with the ARIMA models.

Key-Words: Neural Networks; ARIMA Models; Demand Forecasting.

1. INTRODUÇÃO

A preocupação com preço de bens e serviços é uma constante no dia a dia organizacional. Gerir de forma correta entradas e saídas de produtos, a fim de evitar todo e qualquer desperdício, é um fator importante na redução de gastos de produção. Por isso, se faz pertinente a aplicação de métodos para prever a demanda, a fim de evitar prejuízos em estoques de matérias-primas e itens acabados.

O objetivo desse trabalho é prever a demanda de produção de um produto em uma indústria alimentícia, utilizando duas técnicas de previsão: (i) Redes Neurais Artificiais e (ii) Modelos ARIMA, para realizar uma previsão a curto prazo, ou seja, um passo a frente. Posteriormente, será realizada a comparação dos resultados obtidos com os dados reais, a fim de mensurar o erro desta previsão. Desta forma é possível saber a eficiência destas técnicas como meio de previsão para o caso em estudo.

Em decorrência de sua importância, são muitas as publicações que tratam de previsão de demanda. Podemos citar os seguintes trabalhos: Pellegrini e Fogliatto (2011), que

apresentam em seu trabalho um passo a passo de como realizar o estudo e demonstram alguns dos métodos existentes, finalizando com um estudo de caso; Werner e Ribeiro (2006) que apresentam uma nova técnica para prever a demanda, que parte da combinação de previsões e ajuste por opinião, onde o modelo proposto apresentou excelentes resultados e, dentre tantos outros, Calôba *et al.* (2003) é o que mais se aproxima desse artigo, pois também faz aplicação das Redes Neurais Artificiais para prever a demanda de uma indústria de bebidas.

2. REFERENCIAL TEORICO

2.1. PREVISÃO DE DEMANDA

Demanda é a disponibilidade de compra de bens e serviços pelos clientes (LUSTOSA *et al.*, 2008). Para prever a demanda, há considerável quantidade de métodos disponíveis, que se baseiam em dados de vendas anteriores – séries temporais - e, com o auxílio de métodos numéricos, determinam quais serão as próximas quantidades desejadas pelo público alvo. Devido aos fatores que geram certa variação nesta previsão, os cálculos acabam por apresentar um valor erro.

2.2. MODELOS DE PREVISÃO

Este trabalho tem como objetivo prever a demanda de uma indústria do setor alimentício para tanto foram utilizadas duas ferramentas de previsão: (i) Redes Neurais Artificiais e (ii) modelos ARIMA de Box e Jenkins.

2.2.1. Redes Neurais Artificiais (RNA)

Redes Neurais Artificiais são, de acordo com Haykin (2001), sistemas que manipula a informação pela interação de unidades básicas de processamento denominadas de neurônios artificiais. Neurônios Artificiais, por sua vez, são unidades básicas de processamento que recebem sinais externos e estes são processados pelos componentes do neurônio artificial gerando saídas.

Segundo Haykin (2001), um neurônio artificial é formado pelos seguintes componentes: (i) Um conjunto de pesos sinápticos, que tem como função ponderar os sinais de entrada; (ii) Um termo de limiar ou Bias, que tem por finalidade de diminuir ou aumentar o efeito das entradas no funcionamento do neurônio; (iii) Um somador, que soma o valor das entradas ponderadas pelos pesos sinápticos e (iv) Uma função de ativação que restringe o sinal da saída. A Figura 1 apresenta um modelo de neurônio artificial:

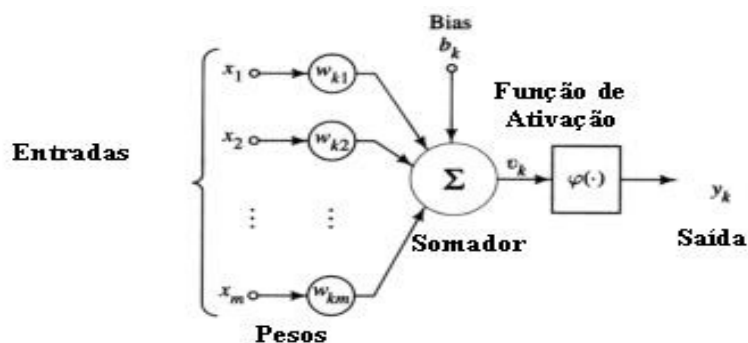


Figura 1 - Modelo de Neurônio Artificial
Fonte: Adaptado de Haykin (2001)

Haykin (2001) descreve matematicamente o neurônio artificial da Figura 1 como:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (01)$$

$$y_k = \varphi(v_k) \quad (02)$$

$$v_k = u_k + b_k \quad (03)$$

Onde: x_1, x_2, \dots, x_m são os sinais de entrada; $w_{k1}, w_{k2}, \dots, w_{km}$ são os pesos sinápticos; u_k é a saída do combinador linear; b_k é um polarizador ou bias; $\varphi(.)$ é a função de ativação; y_k é a saída do neurônio. As funções de ativação geralmente utilizadas são as funções de ativação logística, linear e tangente hiperbólica. Estas funções de ativação são escolhidas por sua fácil derivação o que simplifica a aplicação dos algoritmos de aprendizado como será visto a seguir. A Figura 2 apresenta a três funções de ativação comentadas:

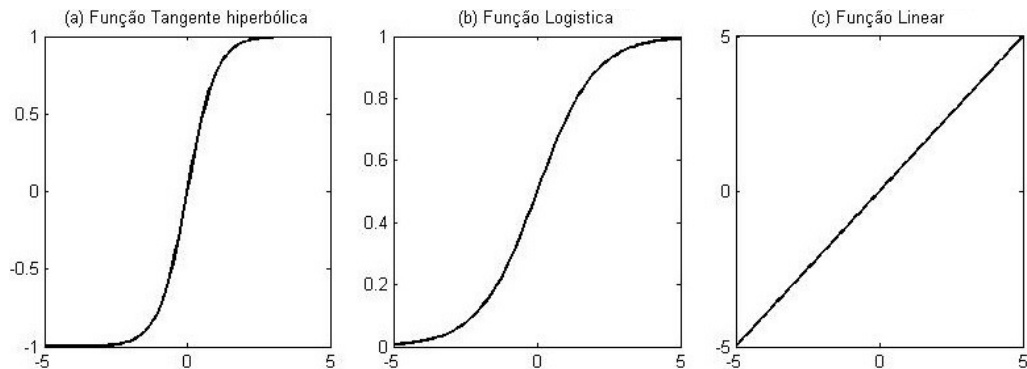


Figura 2 – (a) Função de ativação tangente hiperbólica; (b) Função de ativação logística; (c) Função de ativação linear. Fonte: Autor (2013)

Segundo Braga *et al.*(2007) a capacidade de processamento de um neurônio é muito limitada, no entanto quando neurônios são agrupados a capacidade processamento de dados aumenta. Existem diversos modelos de RNA, que variam de acordo com a forma que os neurônios são organizados e o tipo de algoritmo de aprendizado utilizado. Uma dos modelos tradicionais de RNA consiste em organizar os neurônios artificiais em camadas. A Figura 3 apresenta um modelo de RNA multicamadas:

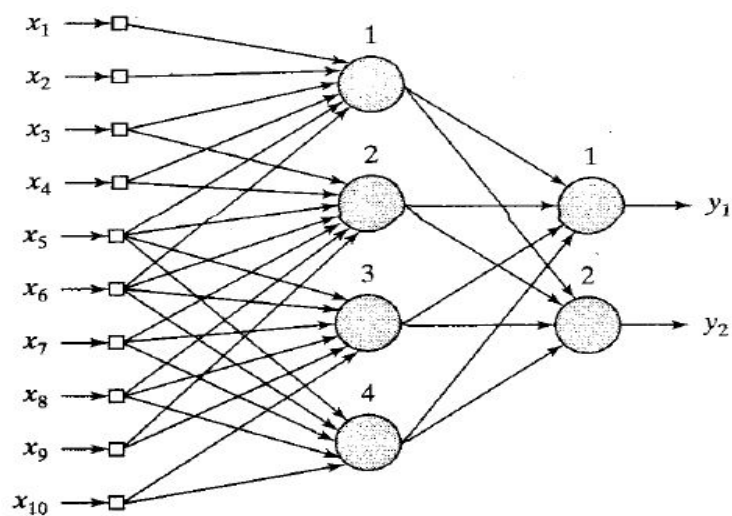


Figura 3 – Exemplo de Rede Neural Artificial
Fonte: Adaptado de Haykin (2001)

O algoritmo de aprendizado consiste na forma em que os pesos sinápticos dos

neurônios artificiais são ajustados. Uma das formas convencionais para realizar o ajuste dos pesos sinápticos de uma RNA é através do método de correção de erro. Matematicamente podemos expressar o ajuste dos pesos sinápticos de uma RN pela seguinte formulação:

$$w_{kj}(n+1) = w_{kj}(n) - \Delta w_{kj}(n) \quad (04)$$

Onde: $w_{kj}(n+1)$ e $w_{kj}(n)$ são os pesos sinápticos ativados pela entrada x_j nos momentos n e $n+1$, respectivamente, e $\Delta w_{kj}(n)$ são os ajustes dos pesos. Existem diversas abordagens para o calculo de $\Delta w_{kj}(n)$ uma abordagem usual é a utilização da regra Delta ou regra de Widrow-Hoff (HAYKIN, 2001), definida como:

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n) \quad (05)$$

$$e_k(n) = d_k(n) - y_k(n) \quad (06)$$

Onde η é a taxa de aprendizado, $d_k(n)$ é a saída gerada pela rede, $y_k(n)$ é a saída real e $e_k(n)$ é o erro calculado da rede. Assim a regra delta busca minimizar o erro da rede neural a partir da minimização de uma "função de custo" que para este é a função do erro quadrático médio:

$$\varepsilon(n) = \frac{1}{n} e_k^2 \quad (07)$$

A partir de $\varepsilon(n)$ é possível mapear os conjuntos de entrada/saída e gerar uma superfície de desempenho de erro multidimensional ou superfície de erro, assim para que a rede consiga melhorar seu desempenho os valores de $\varepsilon(n)$ devem seguir na direção de um valor mínimo (BRAGA *et al.*, 2008). Este tipo aproximação é eficiente quando se trata de redes com apenas uma camada, para treinar RNAs com múltiplas camadas foi desenvolvido o algoritmo de Retropropagação.

A grande dificuldade, segundo Braga *et al.* (2007), para o treinamento por correção de erro de RNAs de múltiplas camadas é que, diferentemente das redes de camada única, onde o erro da rede era obtido pela diferença entre a saída gerada e a saída desejada da rede, não existem saídas desejadas para as camadas intermediárias. Este problema foi resolvido com o desenvolvimento do algoritmo de Retropropagação (ou *Backpropagation*).

O algoritmo de Retropropagação pode ser definido como segue para uma RNA - MLP com uma camada oculta:

Fase *Forward*:

- Inicializar os pesos sinápticos w_{ij}, w_{hj} e bias θ_i, θ_h com valores aleatórios;
- Executar o somatório $t_i^p = \sum w_{ij} x_j^p + \theta_i$ para cada neurônio da camada i (oculta), onde x_j^p representa os dados de entrada da rede;
- Calcular a função de transferência $\varphi(v(i))$ para cada neurônio da camada i ;
- Executar o somatório $t_h^p = \sum w_{hi} a_i^p + \theta_h$ para cada neurônio da camada h (saída);
- Calcular a função de transferência $\varphi(v(h))$ para cada neurônio da camada h ;

Fase *Backward*:

- Calcular o erro da camada h : $\delta_h^p = (d^p - a_h^p) \varphi'_h(v(h))$; onde d^p é o valor real;
- Calcular variação dos pesos sinápticos: $w_{hi}(t) = w_{hi}(t-1) + \Delta^p w_{hi}(t)$, onde $\Delta^p w_{hi}(t) = \eta a_i^p \delta_h^p + \alpha \Delta^p w_{hi}(t-1)$, para η sendo a taxa de aprendizado e α o momentum;

- Calcular o erro da camada oculta: $\delta_i^p = \phi'_h(v(i)); \sum \delta_h^p w_{hi}$;
- Calcular a variação dos pesos: $\Delta^p w_{ji}(t) = \eta x_j^p \delta_i^p + \alpha \Delta^p w_{ji}(t-1)$.

Terminado o ajuste dos pesos, o processo reinicia, assim fazendo com que o erro mínimo tenda a diminuir até que chegue ao valor especificado. O nome dado a este ciclo é época, ou seja, toda vez que a rede completa um ciclo passando por todas as fases do algoritmo é dito que a rede completou uma época. O número de épocas, assim como o número de neurônios, camadas e funções de ativação é definido nas especificações da rede. O ciclo de otimização dos parâmetros finaliza quando um critério de parada preestabelecido é alcançado, como critério de parada pode ser escolhido um número determinado de épocas ou valor de erro médio quadrático.

2.2.2. Modelos ARIMA

Segundo Gujarati (2008) e Montgomery *et al.* (2008) os modelos da família ARIMA são os seguintes:

- Processos Autorregressivo (AR)

Um processo autorregressivo pode definido de acordo com Gujarati (2008), como um modelo que se utiliza da relação entre os termos passados de uma variável mais um termo aleatório. Para elaboração do modelo podem ser utilizados p termos passados, assim o modelo a seguir apresenta a estrutura de um modelo AR(p):

$$Y_t = \mu + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t \quad (08)$$

Onde: μ é média do processo, $\alpha_1 Y_{t-1}$ a $\alpha_p Y_{t-p}$ são proporções dos valores das observações anteriores e ε_t é um erro aleatório.

- Processo de Média Móvel (MA)

Segundo Gurjarati (2008), um modelo de média móvel consiste em uma constante somada à média móvel do erro aleatório das observações anteriores. Em um modelo MA podem ser utilizados q elementos de erro aleatório. Um modelo MA(q) pode ser expresso pela seguinte formulação:

$$Y_t = \mu + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} \quad (09)$$

Onde: $\beta_0 \varepsilon_t$ a $\beta_q \varepsilon_{t-q}$ são ponderações dos erros aleatórios.

- Processo Autorregressivo de Médias Móveis (ARMA)

Segundo Montgomery *et al.* (2008) um modelo autorregressivo de médias móveis nada mais é do que uma combinação entre modelos AR e MA onde são considerados termos anteriores mais o choque causado pelos média móvel dos erros aleatórios. Um modelo de processo ARMA de ordem p e q ou ARMA(p,q) pode ser observado a seguir:

$$Y_t = \theta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta \quad (10)$$

Onde: θ é uma constante.

- Processo Autorregressivo Integrado de Médias Móveis (ARIMA)

Segundo Gujarati (2008), um modelo ARIMA é utilizado quando os dados não apresentam comportamento estacionário, assim sendo necessário um número d de diferenciações para tornar o processo estacionário. A seguir é apresentado um modelo ARIMA com uma diferenciação, um termo autorregressivo e um termo de médias móveis, ou ARIMA (1,1,1):

$$Y_t = (Y_{t-2} - Y_{t-1}) + \alpha_1 Y_{t-1} + \beta_0 \varepsilon_1 \quad (11)$$

2.2.3. Medidas Estatísticas de minimização de Erro

Para Lustosa *et al.* (2008), o indicador básico de erro de previsão para um período de tempo t (e_t) é a diferença entre o valor real e o valor previsto no tempo correspondente, assim podendo formular a seguinte equação:

$$e_t = y_t - \hat{y}_t \quad (12)$$

Onde y_t é a observação real do momento t e \hat{y}_t é o valor previsto pelo modelo para no momento t . Segundo Montgomery *et al.* (2008) existem diversas ferramentas para avaliar o desempenho de um modelo de previsão entre elas podemos citar: Raiz do erro médio quadrático (RMSE) e o Erro Médio Absoluto (MAE). Estas medidas podem ser calculadas da seguinte forma:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (e_t)^2} \quad (13)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (14)$$

Quanto menor forem estas medidas, melhor será o desempenho do modelo de previsão em vista que estas medidas avaliam o quanto o modelo esta errando.

3. MATERIAIS E MÉTODOS

3.1. MATERIAIS

Para este trabalho foi utilizada uma base de dados referentes à saída semanal de um produto x de uma empresa no setor de alimentos, esta base contém 89 observações que foram utilizadas para a construção dos modelos de previsão. Para o desenvolvimento dos modelos de previsão foi utilizado o software MATLAB®. A Figura 4 apresenta o comportamento dos dados utilizados neste trabalho:

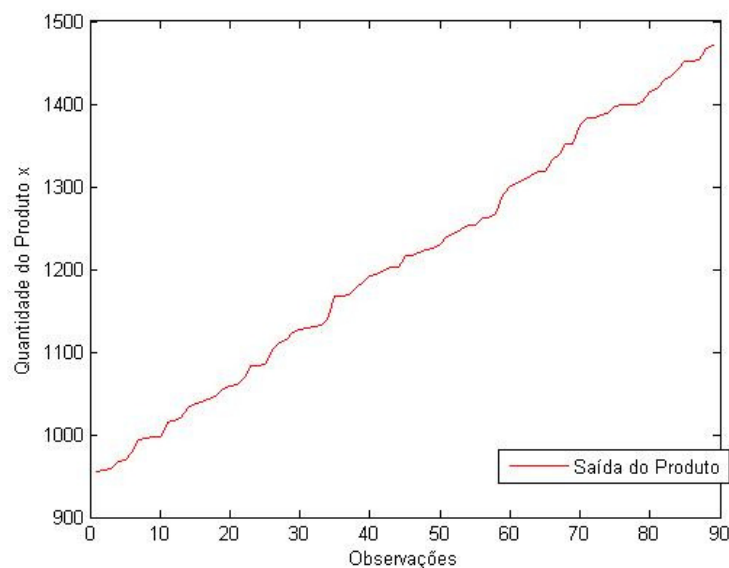


Figura 4 - Retirada semanal do produto x .
Fonte: Autor (2013)

Como podem ser observados os dados se apresentam como uma série temporal não estacionária, apresentando um comportamento crescente da saída do produto x. Este comportamento justifica a utilização das técnicas de previsão selecionadas para o estudo.

3.2. MÉTODOS

O método utilizado para esta pesquisa consiste de duas etapas: (i) Estimação dos parâmetros das técnicas de previsão e (ii) avaliação dos erros de previsão. Como as duas técnicas apresentam metodologias próprias para estimação dos parâmetros, estas serão apresentadas separadamente. Tendo seus parâmetros estimados, as técnicas foram comparadas utilizando as medidas estatísticas MAE e RMSE.

3.2.1. Construção da RNA

Para a construção da RNA foram selecionadas 59 observações que compõe um conjunto de treinamento e 30 para compor um conjunto de teste. Estes dados foram ajustados no intuito de facilitar o processo de aprendizado do algoritmo para um intervalo entre -1 e 1. Para realizar este ajuste foi utilizada a seguinte expressão:

$$t_s = \frac{t_u - t_{min}}{t_{max} - t_{min}} (t_{s,max} - t_{s,min}) + t_{s,min} \quad (15)$$

Como variáveis de entrada foram selecionados três termos autorregressivos assim a RNA pode ser expressa da seguinte forma:

$$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}) + \varepsilon \quad (16)$$

Os parâmetros utilizados para o treinamento da rede foram os seguintes:

- Número de Neurônios: 13;
- Número de camadas: 1 camada oculta;
- Número de Épocas: 3 mil;
- Funções de ativação: tangente hiperbólica para os neurônios da camada oculta e linear na camada de saída.

3.2.2. Estimação do Modelo ARIMA

Como puderam ser observados os dados apresentados apresentam um comportamento não estacionário, logo o modelo de previsão de séries temporais que deve ser selecionado é um modelo ARIMA. No intuito de tornar a comparação entre as duas ferramentas justa, foram utilizados apenas três termos autorregressivos para a construção do modelo ARIMA. Assim o modelo ARIMA estimado para prever a série temporal foi o seguinte:

$$Y_t = 0,3259 + 0,18Y_{t-1} + 0,106Y_{t-2} + 0,57Y_{t-3} + 0,40\varepsilon_1 + 3,85\varepsilon_{t-1} \quad (17)$$

4. RESULTADOS

Após a construção dos modelos foi possível estimar o erro dos modelos a partir das estatísticas MAE e RMSE. As Figuras 5 e 6 apresenta os valores dos erros de previsão para os dados em estudo:

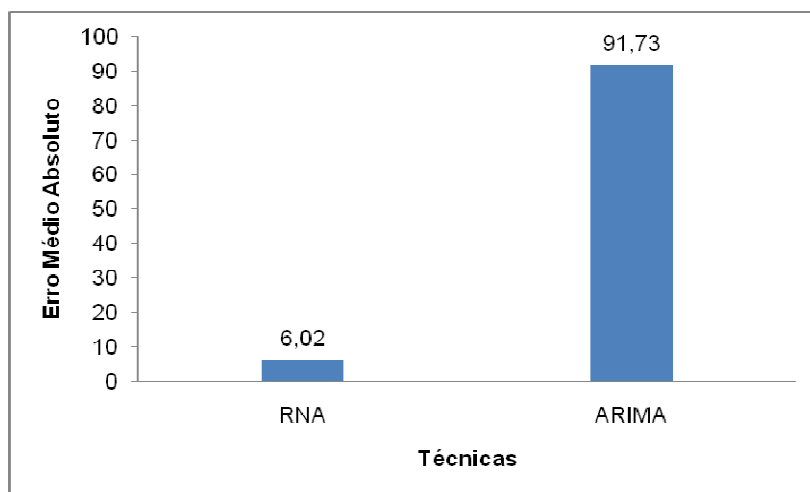


Figura 5– Comparação entre o MAE das técnicas utilizadas.
 Fonte: Autor (2013)

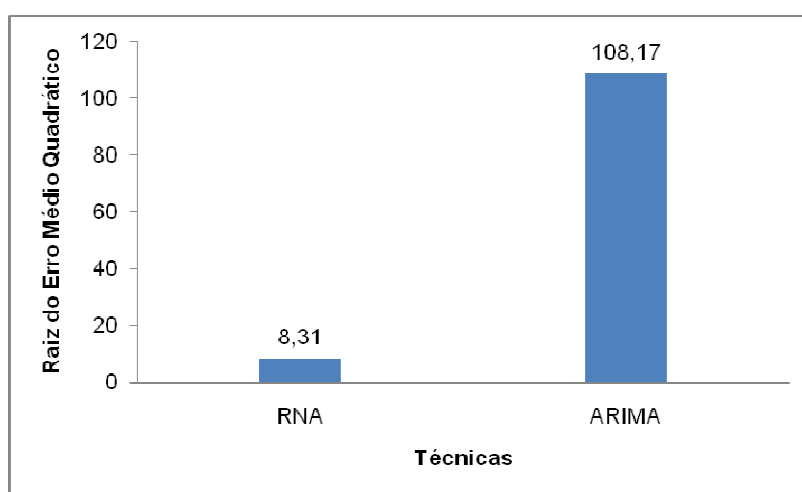


Figura 6– Comparação entre o RMSE das técnicas utilizadas.
 Fonte: Autor (2013)

Observando os dois gráficos é perceptível a eficiência do modelo neural em comparação a técnica clássica de previsão. Isso se deve a grande capacidade da RNA em aprender os padrões dos dados em estudo, a Figura 7 apresenta uma comparação entre as previsões estimadas pela RNA e pelos modelos ARIMA com os dados reais:

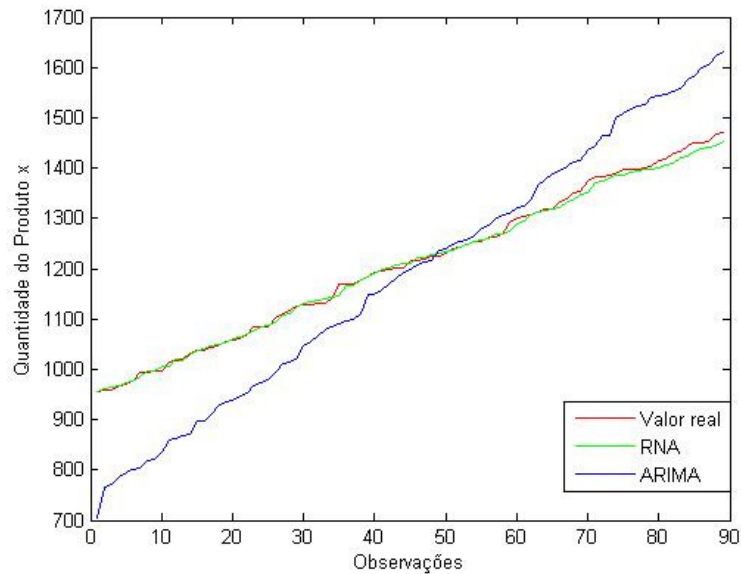


Figura 7– Valores reais comparados aos valores estimados pelas técnicas.
 Fonte: Autor (2013)

É visível na figura apresentada que a RNA conseguiu capturar melhor o comportamento dos dados em comparação ao modelo ARIMA utilizado. Um ponto importante a salientar é que o modelo ARIMA estimado não foi o melhor possível, pois foram utilizados poucos termos autorregressivos e de médias móveis, levando em conta que o objetivo era tornar a comparação entre os dois modelos mais justa.

No entanto, aumentando a complexidade do modelo se torna necessário uma quantidade maior de informação para realizar as previsões. Uma vantagem que se apresentou ao utilizar RNAs foi poder usar uma quantidade menor de dados para realizar uma previsão eficiente, justamente pela alta capacidade da RNA em capturar o comportamento dos dados observados.

5. CONCLUSÕES

A aplicação das técnicas de Redes Neurais Artificiais para prever a demanda do produto x de uma indústria alimentícia foi muito satisfatória. Calôbaet *al.* (2002) conclui seu artigo com algumas proposições. Uma delas especificamente é inadequada: “Redes Neurais não são a melhor solução sempre”. Fica evidente neste artigo que o uso de RNA é válido e o erro encontrado quando comparado com os dados reais da série temporal é baixo. Em comparação com a aplicação do método ARIMA na série temporal, é observável a diferença entre valor real e a previsão obtida, que demonstra bem a superioridade da RNA.

É, também, válido o uso de meios clássicos para prever demandas, porém, tratando-se de uma produtora de alimentos que são, por definição, perecíveis em prazos relativamente curtos, não é interessante a sobra de ingredientes ou estocagem. Portanto, o investimento nesta técnica não usual traz grande vantagem, principalmente do ponto de vista econômico.

Levando em consideração o melhor desempenho das RNAs para realizar a previsão no caso estudado, uma opção para o desenvolvimento de trabalhos futuros seria verificar o desempenho dos modelos de RNA para realizar previsões em horizontes de tempo maiores, em médio prazo por exemplo. Outra opção é utilização de algoritmos mais eficientes para a otimização dos pesos sinápticos, tendo em mente aumentar o horizonte de previsão e reduzir a quantidade de dados utilizados para estimação do modelo neural.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] PELLEGRINI, F. R.; FOGLIATTO, F. S. Passos para Implantação de Sistema de Previsão de Demanda - Técnicas e Estudo de Caso. **Revista Produção**, v.11, n.1, p.43-64, 2001.
- [2] WERNER, L.; RIBEIRO, J. L. D. Modelo Composto para Prever Demanda Através da Integração de Previsões. **Revista Produção**, v.16, n.3, p. 493-509, 2006.
- [3] CALÔBA, G. M.; CALÔBA, LP; SALIBY, E. Cooperação entre Redes Neurais Artificiais e Técnicas ‘Clássicas’ para Previsão de Demanda de uma Série de Vendas de Cerveja na Austrália. **Pesquisa Operacional**, v.22, n.3, p. 345-358, 2002.
- [4] LUSTOSA, L.; MESQUITA, M.A.; QUELHAS, O.; OLIVEIRA, R. **Planejamento e Controle da Produção**. 1ed. Rio de Janeiro: Elsevier, 2008. 2ª reimpressão il.;
- [5] MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. **Introduction to time series analysis and forecasting**. New Jersey: John Wiley& Sons. Inc, 2008.
- [6] HAYKIN, S. **Redes Neurais - Princípios e Práticas**. 2 ed. Porto Alegre: Bookman, 2001.
- [7] GUJARATI, D. N. **Econometria básica**. 5. ed. Porto Alegre: AMGH, 2011.
- [8] BRAGA, A.P; CARVALHO, A.F.; LUDERMIR, T.B. **Redes Neurais Artificiais: Teorias e Aplicações**. 2.ed. Rio de Janeiro: LTC, 2007.