



Universidade Federal de Pernambuco
Centro de Informática
Bacharelado em Sistemas de Informação



**Uma Abordagem Híbrida para a Predição de
Séries Temporais**
Trabalho de Conclusão de Curso

Aluno: Avyner Henrique Bezerra da Fonseca Lucena. (ahbfl@cin.ufpe.br)

Professor Orientador: Paulo Salgado Gomes de Mattos Neto. (psgmn@cin.ufpe.br)

Recife, 2016



Universidade Federal de Pernambuco
Centro de Informática
Bacharelado em Sistemas de Informação



Uma Abordagem Híbrida para a Predição de Séries Temporais
Trabalho de Conclusão de Curso

Trabalho de conclusão de curso, apresentado a coordenação de graduação do Centro de Informática da Universidade Federal de Pernambuco, como requisito obrigatório para obtenção do grau de Bacharel no Curso de Sistemas de Informação.

Aluno: Avyner Henrique Bezerra da Fonseca Lucena. (ahbfl@cin.ufpe.br)

Professor Orientador: Paulo Salgado Gomes de Mattos Neto. (psgmn@cin.ufpe.br)

Recife, 2016



Universidade Federal de Pernambuco
Centro de Informática
Bacharelado em Sistemas de Informação



Uma Abordagem Híbrida para a Predição de Séries Temporais
Trabalho de Conclusão de Curso

Trabalho de conclusão de curso, apresentado a coordenação de graduação do Centro de Informática da Universidade Federal de Pernambuco, como requisito obrigatório para obtenção do grau de Bacharel no Curso de Sistemas de Informação.

Trabalho Aprovado em: 20/02/2016.

Assinatura do Orientador:

Paulo Salgado Gomes de Mattos Neto.

Recife, 2016

*“Aerials in the sky, when you lose
small mind you free your life
Aerials so up high, when you free
your eyes eternal prize.”*

(Serj Tankian, Daron Malakian)

Agradecimentos

Gostaria de agradecer aos meus Pais Adelmo de Lucena da Silva e Rita de Cássia B. da Fonseca Lucena, por todo o apoio e confiança em mim depositados nos últimos 21 anos de convivência.

Ao meu Irmão, Aramys Hiago B. da Fonseca Lucena e namorada Jaqueline Miranda pelo incentivo e paciência.

Ao meu orientador, Professor Dr. Paulo Salgado Gomes de Mattos Neto, pela orientação e ajuda na elaboração deste trabalho.

E por fim a todos os meus colegas e professores, que tornaram os últimos quatro anos de convivência engrandecedores e agradáveis.

Resumo

Neste trabalho é proposta uma abordagem híbrida para previsão de séries temporais. A combinação de modelos preditivos é um assunto recorrente e amplamente discutido e estudado na área [9]. Este tipo de combinação visa atingir uma maior acurácia, em comparação com as previsões individuais dos modelos combinados. A abordagem aqui proposta, combinará as previsões pela modelagem do erro da previsão de uma determinada série temporal, onde este erro se trata da diferença entre a previsão de um modelo utilizado e um valor alvo a ser alcançado [33]. Combinações de preditores podem ser realizadas de diferentes maneiras [9], pode-se observar na literatura da área [10,21,22], que a combinação de modelos distintos vem se provando uma boa alternativa para a previsão de séries temporais advindas de observações do mundo real, pois tendo em vista que estas observações são compostas por padrões com diferentes características, a combinação de modelos de naturezas diferentes tende a capturar de maneira mais ampla tais características. O presente trabalho apresenta uma abordagem híbrida, onde foram separadas as correlações não lineares das lineares dos padrões apresentados, por meio da modelagem do erro, buscando-se uma melhor utilização de preditores assim como proposto em Zhang (2003) [10]. Porém tais previsões não foram combinadas pelo meio de sua soma ou média como realizado na literatura da área [10,15], sendo combinadas então pela técnica de aprendizagem de máquina SVR [23] e a rede neural artificial MLP [18], onde se obteve como resultado que a combinação realizada por estas técnicas atingiu melhoras nas previsões combinadas quando comparadas com a previsão individual dos modelos combinados.

Abstract

This work, proposes a hybrid approach for time series prediction. The combination between predictive models has been extensively discussed and studied in the field [9]. This kind of combination aims to increase the accuracy over the metrics applied to the forecast results, in comparison with the individual predictions of the combined predictors. The approach proposed, aims to combine the forecasts by modeling the series error, where this error is about the difference between a model's forecast and an target value [33]. There are several different ways to combine forecasters [9], looking into the field literature [10,21,22], it is possible to realize that the combination of dissimilar models, has been proved as a good way to predict time series, generated from real word observations, because those observations are composed by different kind of patterns, so is expected that the combination between non similar models to be able of catching a large range of characteristics. The current work presents a hybrid approach, where the non-linear correlations was separated from the linear correlations of the used patterns by using error modeling, aiming applying the best model over the different kind of patterns, just as proposed in Zhang(2003) [10]. But those predictions weren't combined by it's sum or average as proposed in field lectures [10,15], being combined by the machine learning technique SVR[23] and the artificial neural network MLP[18], where achieved gains over the combined predictions when compared with the individual forecast of each of combined models.

Sumário

Índice de Figuras

Índice de Tabelas

1. Contextualização e Justificativas	11
1.1 Objetivos	13
1.2 Organização do trabalho	14
2. Séries Temporais	14
2.1 Características Gerais.....	15
2.2 Objetivos de Análises em Séries Temporais.....	17
2.3 Séries Temporais Utilizadas.....	18
3. Modelos utilizados	20
3.1 ARIMA	20
3.2 Support Vector Regression	21
3.3 Redes Neurais Artificiais	24
Redes Multilayer Perceptron.....	25
4. Abordagens Híbridas	26
4.1 Estratégia Híbrida Proposta	27
5. Metodologia dos Experimentos Realizados	29
5.1 Tratamento dos Dados.....	29
5.2 Divisão dos Conjuntos de Dados	30
5.3 Ordem dos Experimentos.....	30
5.4 Seleção e Variação de Parâmetros:.....	31
5.5 Métricas Utilizadas	32
6. Resultados dos Experimentos	34
6.1 Experimentos para a Série Temporal British Pound / USD Exchange Rate.....	34
6.2 Experimentos para a Serie Temporal Canadian Lynx.....	38
6.3 Experimentos para a Série Temporal Wolf's Sunspot	43
6.4 Análise dos Resultados	47
7. Conclusões	49
7.1 Melhorias Propostas e Trabalhos Futuros.....	49
Referências.....	51

Índice de Figuras

Figura 1 - Série Temporal Índice Dow Jones.

Figura 2 - Exemplo de Série Temporal Estacionária

Figura 3 - Exemplo de Série Temporal Não Estacionária

Figura 4 - Exemplos de Sazonalidade e Tendência

Figura 5 - Série Temporal British Pound / US Dollars Exchange Rate

Figura 6 - Série Temporal Canadian Lynx

Figura 7 - Série Temporal Wolf's Sunspot

Figura 8 - Separação entre classes com margens SVM

Figura 9 - Representação da margem de separação não linear

Figura 10 - Mapeamento de conjuntos de dados para espaço de maior dimensionalidade.

Figura 11 - Representação de Single Layer Perceptron de Rosenblatt

Figura 12 - Representação da Topologia do Multilayer Perceptron

Figura 13 - Estratégia Híbrida Implementada

Figura 14 - Resultados das Previsões do Arima para a Série BP/USD Exchange Rate

Figura 15 - Resultados das Previsões do SVR para a Série BP/USD Exchange Rate

Figura 16 - Resultados da Combinação pelo MLP para a Série BP/USD Exchange Rate

Figura 17 - Resultados da Combinação pelo SVR para a Série BP/USD Exchange Rate

Figura 18 - Resultados da Combinação pela Soma para a Série BP/USD Exchange Rate

Figura 19 - Todas as Previsões para a Série BP/USD Exchange Rate

Figura 20 - Resultados das Previsões do Arima para a Série Canadian Lynx

Figura 21 - Resultados das Previsões do SVR para a Série Canadian Lynx

Figura 22 - Resultados da Combinação pelo MLP para a Série Canadian Lynx

Figura 23 - Resultados da Combinação pelo SVR para a Série Canadian Lynx

Figura 24 - Resultados da Combinação pela Soma para a Série Canadian Lynx

Figura 25 - Todas as Previsões para a Série Canadian Lynx

Figura 26 - Resultados das Previsões do Arima para a Série Wolf's Sunspot

Figura 27 - Resultados das Previsões do SVR para a Série Wolf's Sunspot

Figura 28 - Resultados da Combinação pelo MLP para a Série Wolf's Sunspot

Figura 29 - Resultados da Combinação pelo SVR para a Série Wolf's Sunspot

Figura 30 - Resultados da Combinação pela Soma para a Série Wolf's Sunspot

Figura 31 - Todas as Previsões para a Série Wolf's Sunspot

Índice de Tabelas

Tabela 1: Configuração dos Modelos para *BP/USD Exchange Rate*.

Tabela 2: Resultado dos Experimentos para *BP/USD Exchange Rate*

Tabela 3: Configuração dos Modelos para *Canadian Lynx*

Tabela 4: Resultados Dos Experimentos para *Canadian Lynx*

Tabela 5: Configuração dos Modelos para *Wolf's Sunspot*.

Tabela 6: Resultados dos Experimentos para *Wolf's Sunspot*

1. Contextualização e Justificativas

É evidente o atual crescimento exponencial no volume de dados gerados seja por empresas, pelo número crescente de sensores espalhados nas cidades, usuários de redes sociais etc. Devido a este crescimento surge a necessidade de se processar e analisar conjuntos de dados cada vez maiores e mais complexos, então novos algoritmos e técnicas vem sendo desenvolvidas para realização destas tarefas [26]. Muitos dos métodos aplicados sobre este tipo de conjunto de dados fazem parte das áreas que conhecemos como Mineração de Dados [24] e Aprendizado de Máquina [25]. No que tange a Mineração de Dados a ideia é o desenvolvimento e aplicação de técnicas para extração de informações desejadas sobre um determinado conjunto de dados. Existem duas grandes vertentes na mineração de dados sendo elas a mineração preditiva e a mineração descritiva [34]. As técnicas relacionadas a mineração descritiva tem por objetivo maior, entender e explicitar relações entre padrões apresentados e seu significado [27]. Por outro lado, a mineração preditiva tem como objetivo geral mapear e compreender as relações entre os conjuntos de dados explorados buscando prever estimativas para possíveis valores futuros. Para que seja possível a aplicação da mineração de dados, usualmente são utilizadas técnicas de aprendizado de máquina e computação inteligente ao invés de se utilizar técnicas convencionais [24] para o tratamento e recuperação de informações.

A previsão de séries temporais [4], pode ser considerada uma área de aplicação da mineração de dados preditiva, onde são coletadas observações de valores passados de uma ou mais variáveis e a partir da apresentação desses valores a um modelo, se tenta prever valores futuros para a série apresentada. Este campo de estudo é amplamente abordado e vem sendo aplicado por áreas como a estatística e contabilidade [2], possuindo diversas aplicações, tais como: a previsão de preços de venda de determinado produto, a previsão de valores em bolsas de valores e a previsão do lucro de uma empresa em um determinado período de tempo. A atividade de previsão de séries temporais se estende por muitas outras áreas de atuação e continua a crescer e se expandir para mais e mais setores [4], pois devido ao cenário atual com o crescimento inegável na quantidade de dados gerados, o número de problemas reais continua a crescer, sendo assim a necessidade de aplicações como a predição de valores futuros em conjuntos de dados relacionados a intervalos de tempo também tende a aumentar [26].

No decorrer dos anos, muitos modelos foram desenvolvidos e melhorados com o objetivo de prever séries temporais [3]. Dentre as muitas abordagens e estratégias desenvolvidas se encontra a combinação entre preditores. Esta ideia tem como objetivo a combinação da previsão de um grupo de modelos buscando-se assim uma melhora na previsão, quando comparada com as previsões individuais de cada modelo. Devido a sua relevância, trabalhos de autores como Bates e Granger (1969) [12] e Reid (1968) [13], são considerados por alguns como os pioneiros na área de combinação de previsores [9], mesmo existindo trabalhos anteriormente publicados na área. A combinação entre preditores pode ser realizada de muitas maneiras diferentes [9], tais como: a utilização da média entre as previsões de um grupo de preditores, a utilização de regressões lineares entre outras. A literatura da área possui um grande número de publicações e propostas de combinações, porém se observa que grande parte dessas combinações utilizam previsores com características muito semelhantes ou até iguais, porém com diferentes estruturas. Uma das abordagens desenvolvidas para a combinação de preditores, envolve a modelagem do erro da série temporal, sendo este erro a diferença entre o valor encontrado pela previsão de um modelo e o valor real da série a ser prevista [11]. Esta foi a abordagem utilizada neste trabalho, devido ao aumento na acurácia da previsão proporcionada por este tipo de modelagem.

A utilização de modelos lineares estatísticos combinados para predição de séries temporais, devido a relativa facilidade encontrada na sua implementação e experimentação, vem sendo amplamente abordada na área, possuindo assim um grande número de publicações recentes e antigas [8,14]. Outro tipo de combinação que possui muitos trabalhos publicados é a de conjuntos de redes neurais artificiais [15], que devido a características como capacidade de generalização e modelagem de padrões não linearmente separáveis, também vem sendo utilizada extensivamente na área.

Ao se combinar as previsões de diferentes modelos se espera que suas configurações sejam capazes de mapear e explorar diferentes relações entre os padrões a estes apresentados. O que torna a combinação das estimativas encontradas, mais amplas e provavelmente mais precisas, uma vez que não existe um melhor modelo com capacidade de generalização tão ampla que possa ser aplicado em qualquer situação.

É observado na literatura, que a utilização de modelos muito semelhantes pode não ser tão interessante, pois devido as limitações dos modelos utilizados, características relevantes dos padrões analisados podem vir a ser negligenciados. Devido a esse tipo de limitação, a utilização de abordagens híbridas onde são combinados modelos não semelhantes, surge como uma alternativa onde se faz possível a exploração dos pontos fortes de cada um dos modelos individualmente, e logo após a sua combinação podem ser encontradas melhores estimativas, mesmo em conjuntos de dados com características diversas [10].

Na abordagem híbrida proposta neste trabalho de conclusão de curso, serão utilizados alguns modelos. O modelo estatístico *Auto Regressive Integrated Moving Average* (ARIMA), que se trata de uma das técnicas desenvolvidas por Box e Jenkins [3], e vem sendo amplamente utilizada na previsão de séries temporais e também como *benchmark* para medir a acurácia da previsão de outros tipos de métodos de previsão, onde se espera que esta técnica devido a suas características, lide com os padrões lineares presentes nas séries temporais a ele apresentadas. Outra técnica que será utilizada será o *Support Vector Regression* [23], um modelo capaz de gerar boas previsões mesmo com poucos exemplos de entrada. Devido a suas características que o permitem uma boa adequação a padrões não lineares, ele ficará responsável pela previsão dos resíduos formados pelo erro resultante da diferença entre a previsão do ARIMA e o valor real de tais padrões. Por fim será utilizada a rede neural artificial *Multilayer Perceptron* [18] MLP e um modelo SVR dada sua capacidade de generalização, na combinação entre as previsões realizadas pelo modelo ARIMA e a previsão do erro dada pelo SVR.

1.1 Objetivos

O presente trabalho teve como objetivo verificar a eficácia de uma nova estratégia para a combinação de previsões, onde invés de se somar as previsões encontradas como apresentado em [10], é utilizada uma técnica de aprendizagem de máquina para modelar e combinar estas previsões. Para atingir este objetivo foi realizada a implementação de uma abordagem híbrida para a predição das séries temporais British pound/US dollar Exchange rate data, a série Canadian Lynx [16] e por fim a Wolf's sunspot data[17]. Onde foi utilizado o modelo ARIMA para prever as séries,

sendo seguido pela modelagem dos erros residuais desta predição por um modelo SVR tendo por fim ambas as predições combinadas por uma rede MLP e um modelo SVR.

1.2 Organização do trabalho

Este trabalho de conclusão de curso é formado por um total de 7 seções, na primeira seção é introduzido o tema, por meio de uma contextualização sobre a área e as justificativas deste trabalho, seus objetivos e seguido por uma pequena introdução deste e dos demais capítulos. Na segunda seção, séries temporais, são apresentadas características gerais de uma série temporal, alguns objetivos por trás de sua análise e as séries utilizadas na elaboração deste trabalho. A terceira seção, modelos utilizados, é composta por uma conceituação introdutória quanto aos modelos utilizados nos experimentos deste trabalho. Para a quarta seção Abordagens Híbridas, conceitos relativos ao tema foram explanados assim como a abordagem desenvolvida para este trabalho. A quinta seção apresenta de forma sequencial, a metodologia seguida pelos experimentos realizados. A sexta seção apresenta as configurações dos modelos utilizados nos experimentos, seguidos dos resultados alcançados nos experimentos conduzidos. A sétima e última seção deste trabalho apresenta as conclusões e um pequeno conjunto de possíveis melhorias e novos trabalhos.

2. Séries Temporais

Uma série temporal pode ser definida como um conjunto de dados sequencial, que se encontra relacionado a um determinado período sucessivo de tempo, se caracterizando normalmente como a captura de uma sequência de observações acerca de uma ou mais variáveis de interesse [5]. Em um determinado espaço de tempo onde $t = [1, \dots, T]$, uma série temporal Y é composta por uma sequência de observações onde $Y = [y_1 \dots y_T]$ [4], um exemplo de série temporal pode ser visto na Figura 1 composta por observações do índice Dow Jones.

Figura 1 - Série Temporal do Índice Dow Jones



Fonte: [www.google.com.br/googlefinance\(2016\)](http://www.google.com.br/googlefinance(2016))

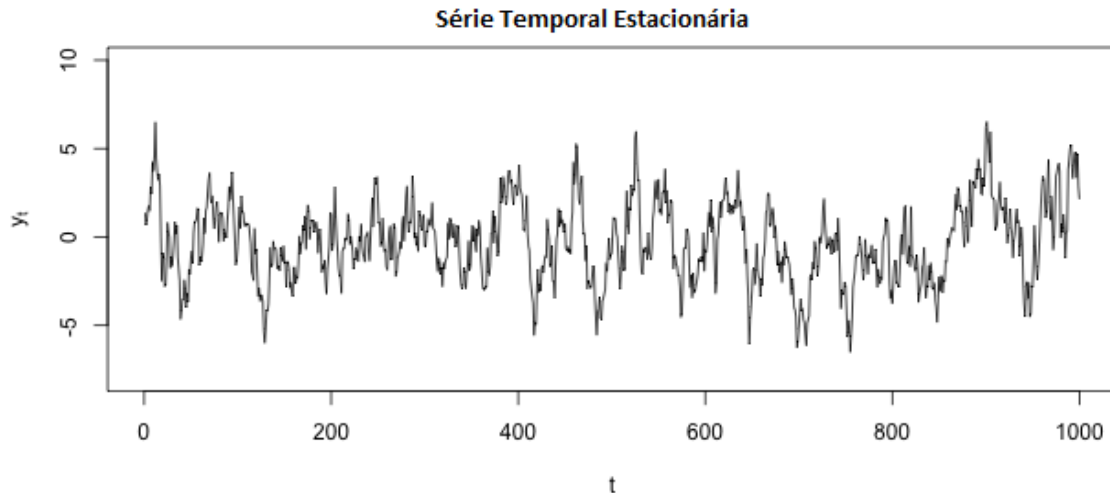
2.1 Características Gerais

Uma série temporal que contém registros de apenas uma variável é definida como univariada e caso sejam compostas por registros de mais de uma variável estas são chamadas de séries multivariadas. Outro ponto diz respeito aos intervalos de tempo relacionados a série, podendo ela ser discreta ou contínua, um exemplo de uma série temporal discreta é o número médio anual de manchas solares e um exemplo contínuo pode ser o registro das marés no porto de Santos. Por muitas vezes séries discretas são compostas por amostragens de séries contínuas em um determinado espaçamento temporal fixo entre as observações, como por exemplo, intervalos diários, semanais, mensais ou anuais [1].

Uma outra característica muito importante que é levada em conta na análise de qualquer série temporal é a estacionariedade. A estacionariedade de uma série temporal é relacionada ao comportamento estatístico desta série no decorrer do tempo, caracterizado por uma probabilidade de distribuição constante durante o tempo [4]. Porém, a maioria das séries temporais do mundo real apresenta alguma forma de não estacionariedade seja por curtos ou longos períodos de tempo, estas séries podem ser estacionárias por períodos de tempo longos e não estacionárias por períodos de tempo curto e vice e versa. Muitas técnicas estatísticas assumem que a série apresentada seja estacionária, pois utilizam estimativas como média, variância e auto correlações e como estas métricas não se mantêm constantes ou com valores próximos no decorrer do tempo em séries não estacionárias, estes modelos acabam

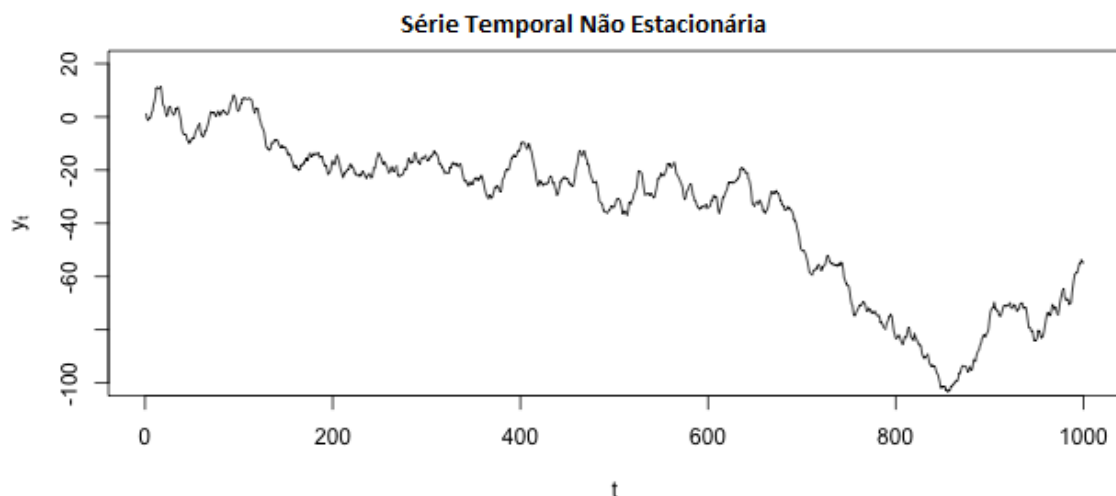
não podendo ser bem utilizados caso não seja realizada uma transformação desta série para que esta se torne estacionária. [1,2].

Figura 2 - Exemplo de Série Temporal Estacionária



Fonte: en.wikipedia.org/wiki/Stationary_process

Figura 3 - Exemplo de Série Temporal Não Estacionária

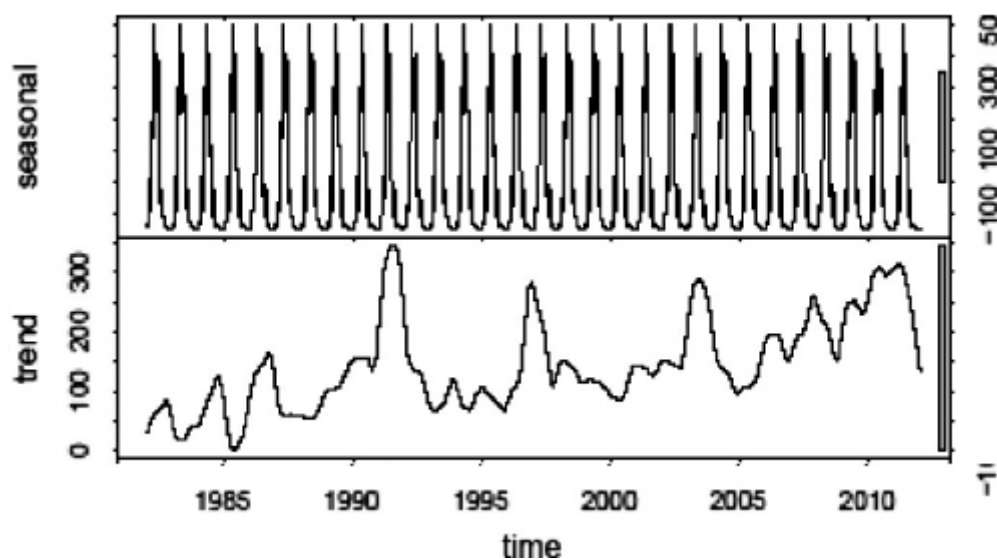


Fonte: en.wikipedia.org/wiki/Stationary_process

O conceito de sazonalidade é relacionado ao acontecimento de um determinado evento que se repete em um certo período de tempo. Existem muitas séries temporais que são afetadas por este tipo de evento principalmente aquelas que envolvem vendas, um exemplo simples se daria em uma observação acerca do número de casacos vendidos durante todo o ano, provavelmente se notaria um aumento de compras em meses mais frios. A sazonalidade é uma característica facilmente identificada assim que são obtidas mais informações sobre o problema, e

devido a este e outros fatores este tipo de comportamento em séries temporais tende a ser facilmente identificado e removido, a não ser que esta sazonalidade seja uma medida de interesse. Outra característica que pode ser facilmente confundida com a sazonalidade é a tendência, que em séries temporais pode ser definida como uma mudança de longo termo no nível da média. Normalmente, séries com este tipo de característica, apresentam variações semelhantes em um período de tempo considerado longo e aparentemente simétrico. Caso sejam utilizados dados com intervalo de tempo inferiores ao de ocorrência das alterações decorrentes de tendências se torna difícil identificá-las, porém em intervalos de tempo que abranjam e repitam este tipo de evento, se torna evidente e visível. Algo que deve ser salientado é a dificuldade de se escolher um período de tempo correto para que uma tendência não seja confundida com o conceito de sazonalidade, naturalmente a compreensão do contexto do problema tende a ser suficientemente esclarecedora [6].

Figura 4 - Exemplos de Sazonalidade e Tendência



Fonte: www.researchgate.net

2.2 Objetivos de Análises em Séries Temporais

Ao se analisar uma série temporal $Z(z_1 \dots z_t)$, compreendida entre os instantes $t_1 \dots t_n$, podem ser investigados muitos fatores tais como: o processo de geração da série analisada, a previsão de valores futuros, a modelagem de características tais como tendência e sazonalidade entre outras possíveis análises [1].

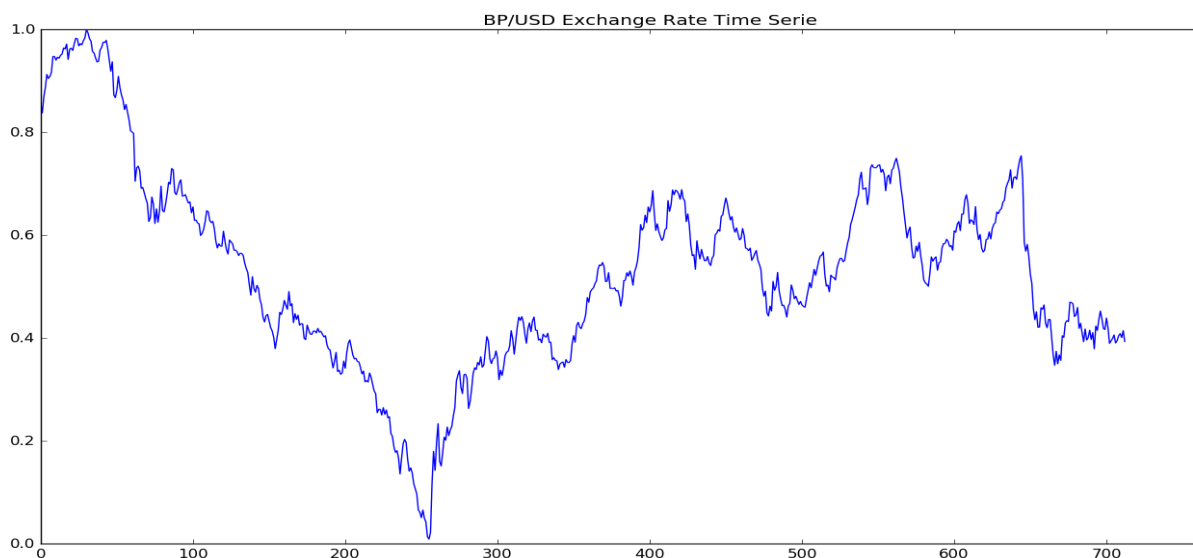
2.3 Séries Temporais Utilizadas

A escolha das séries temporais utilizadas neste trabalho se deu pelas características individuais e diferenças entre as escolhidas. Outro ponto levado em consideração foi sua utilização em trabalhos da área que também utilizam abordagens híbridas e modelagem de erro residual [10,11].

2.3.1 British Pound / US Dollar Exchange Rate

A taxa de câmbio se trata do valor de uma moeda em relação a alguma outra, no caso desta série temporal se trata de observações acerca da taxa de câmbio entre libras esterlinas e o dólar americano.

Figura 5 - Série Temporal British Pound / US Dollars Exchange Rate

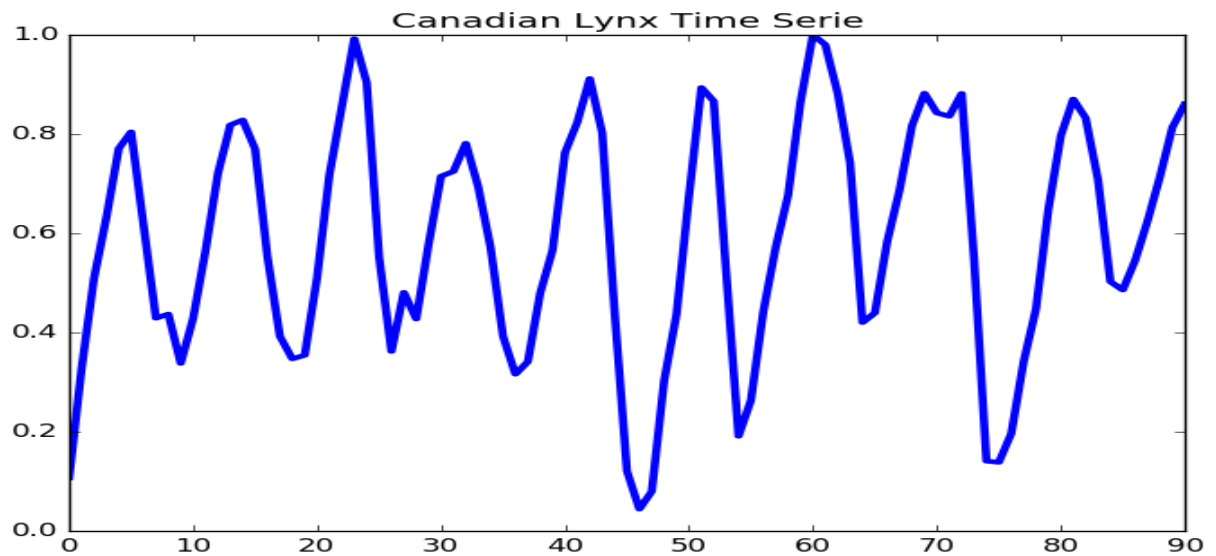


Fonte: O autor(2016).

2.3.2 Canadian Lynx

Esta série temporal é composta por registros relativos a quantidade de Lynx canadenses capturados no distrito de Rio Mackenzie no Canadá, entre os anos de 1821 e 1934 [16].

Figura 6 - Série Temporal Canadian Lynx

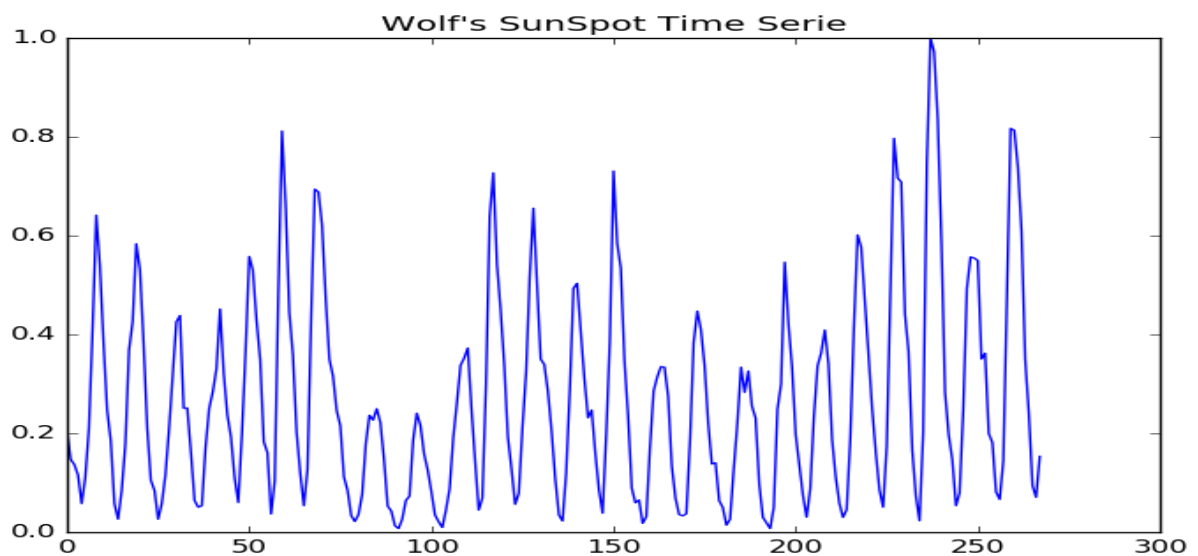


Fonte: O autor (2016).

2.3.3 Wolf's Sunspot

Manchas solares são áreas no sol que possuem temperatura inferior que a média local, sendo geradas por fortes campos magnéticos. A série temporal em questão é composta por um conjunto de médias anuais de manchas solares geradas a partir de um método desenvolvido por Rudolf Wolf para quantificar esses eventos [17].

Figura 7 - Série Temporal Wolf's Sun Spot



Fonte: O autor (2016).

3. Modelos utilizados

A seleção do modelo apropriado de acordo com as características e tipos de padrões que compõem as séries temporais estudadas é de suma importância, pois como já citado, não existe modelo ideal e geral o suficiente para qualquer problema proposto [10]. Devido a abordagem utilizada neste trabalho, se fez necessária a seleção de modelos que obtivessem uma boa performance no tratamento de padrões lineares e não lineares, tendo isto em vista, a seguir serão apresentados de maneira geral os modelos utilizados nos experimentos realizados.

3.1 ARIMA

O modelo Autoregressive Integrated Moving Average (ARIMA) é um modelo linear proposto por Box e Jenkins, sendo amplamente utilizado na literatura devido a sua boa performance na modelagem de séries temporais [3]. A popularidade de modelos lineares se dá pela relativa facilidade encontrada quanto a sua compreensão e implementação. Um problema encontrado quanto a utilização de grande parte dos modelos lineares propostos na literatura, se dá pela sua não adequação a séries não estacionárias. Neste contexto é proposto o modelo ARIMA, que é uma generalização de um modelo anteriormente proposto o modelo Auto Regressivo de Médias Móveis, e possui a capacidade de lidar igualmente com séries estacionárias e não estacionárias por meio da diferenciação sucessiva dos parâmetros encontrados nesta série de acordo com seus valores passados. Este modelo é representado da seguinte maneira: ARIMA (p,d,q). Onde p representa o número de termos auto regressivos, d representa o número de diferenças e q representa o número de termos da média móvel [1].

O processo de previsão de valores futuros para uma determinada série temporal é dada pela função linear que relaciona observações passadas com erros aleatórios expressa abaixo.

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1)$$

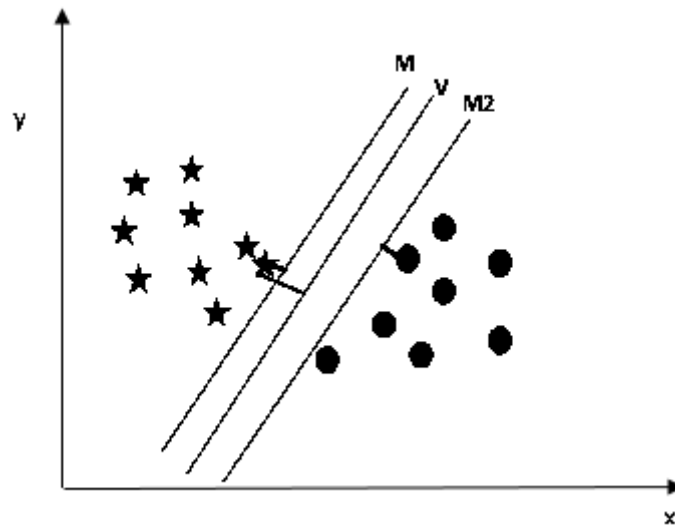
Onde y_t e ε_t são as representações do valor atual na série e do erro aleatório em um determinado período de tempo t e todos e as outras notações são representações de parâmetros do modelo [10].

3.2 Support Vector Regression

Support Vector Regression (SVR), foi proposto por Vapnik e equipe na década de 90 [36], com o propósito de adaptar o conceito de *Support Vector Machines* de Vapnik (1995) para realização de regressões, conceitos já abordados pela técnica a qual se deriva, como minimização do erro geral por maximização de margens e utilização de funções de *kernel* foram mantidos [23]. As principais alterações são relativas ao tipo de problema ao que a técnica se propõe a solucionar, como por exemplo os formatos e tipos de entrada e saída. Ao se aplicar SVM para uma classificação binária, as entradas são basicamente um grande vetor com os padrões a serem classificados e um marcador relativo a classe de $[0,1]$, como saída os valores de zero ou um, relativos a classificação. Já para problemas de regressão, as saídas são números reais relativos as previsões quanto aos padrões apresentados no teste. Onde saída = $(p_1 \dots p_n)$, e p são as previsões obtidas. As entradas são organizadas da seguinte maneira: entrada = $(x_1, y_1) \dots (x_n, y_n)$, onde x se refere aos atributos dos dados apresentados, no caso específico deste trabalho se refere aos valores passados de uma série temporal em um determinado intervalo de tempo. E y se refere aos valores alvos a serem utilizados durante o treinamento.

Uma das características mais fortes deste modelo é a sua capacidade de generalização herdada das SVM. Ao se aplicar SVM para atividades relativas a classificação, se tenta traçar o vetor de suporte que melhor separe as classes, mantendo a maior margem de separação possível relativa as classes separadas, a partir de funções estatísticas aplicadas. A imagem abaixo (Figura 8) apresentada a aplicação padrão de um SVM para classificação linear de duas classes, onde o melhor vetor é o vetor V , que consegue dividir corretamente as duas classes com maior distância entre as classes apresentadas.

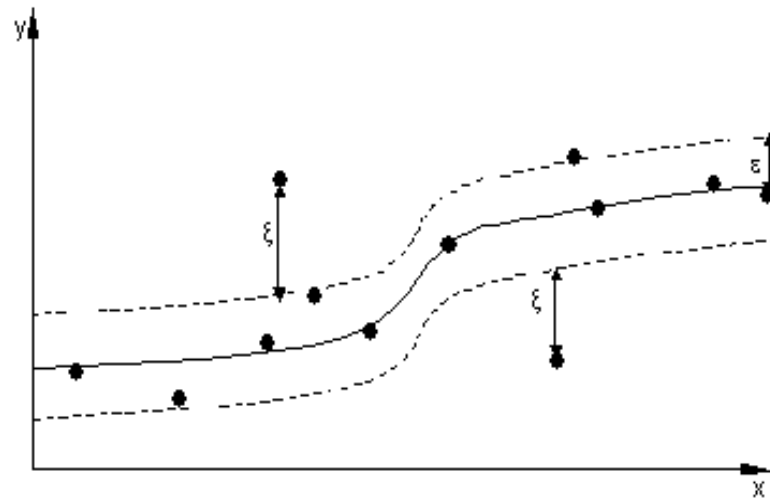
Figura 8 – Separação entre classes com margens SVM



Fonte: O autor (2016).

No exemplo anterior, fica evidente que podem ser traçados ou testados um grande número de vetores, então além de se convergir para uma melhor solução, durante o processo são encontradas muitas outras soluções. Esta ideia de maximização de margens também é mantida para SVR, porém pela diferença na natureza das tarefas a serem atingidas, uma abordagem um pouco diferente é abordada. Basicamente seria inviável senão impossível, tentar se prever todos os possíveis números no espaço de busca próximo ao alvo de uma solução, pois em se tratando de números reais, se sabe que teríamos infinitos números. Para solucionar este problema e ao mesmo tempo manter o algoritmo geral o suficiente para não ficar preso em soluções não ótimas, durante a procura pelo melhor valor de uma predição, um dos parâmetros utilizados permite que a função utilizada na busca por uma melhor solução, permita uma margem de erro definida por um parâmetro ϵ , sendo assim se é mantida a generalização pois no final do processo de treinamento se termina com várias possíveis soluções, porém é utilizada aquela que apresenta o menor erro. Na Figura 9 segue um exemplo da utilização da “*Loss Function*” [6], utilizada pelo SVR, sobre um problema não linear onde os círculos dentro da margem também são aceitos como possíveis soluções.

Figura 9 - Representação da margem de separação não linear



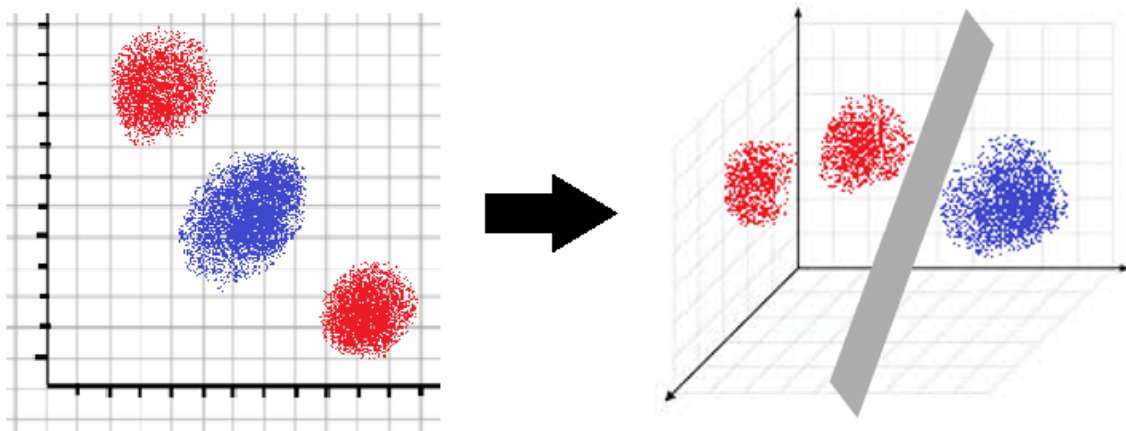
Fonte: kernelsvm.tripod.com

Uma das características mais interessantes das SVM e SVR se dá na utilização das funções de *kernel*, para mapear os dados apresentados a partir de funções não lineares em um espaço n -dimensional, então o novo hiperplano ótimo pode ser construído neste novo espaço. Basicamente, este mapeamento realizado transforma problemas não lineares em problemas lineares, permitindo assim a aplicação de regressões lineares. Por fim ao se colocar os dados em um espaço de alta dimensionalidade e aplicar regressões lineares, se tem o mesmo resultado da utilização de modelagens mais complexas e não lineares, evitando assim problemas como a maldição da dimensionalidade [28]. De acordo com o problema específico a ser solucionado podem ser selecionadas diferentes opções de funções de *kernel* que podem ser utilizadas para SVM/SVR, algumas das mais populares são [29]:

- Polinomial : $K(a,b) = (a \cdot b + 1)^d$ (2)
- *Radial Basis Function* (RBF): $K(a,b) = \exp(-\gamma \|a - b\|^2)$ (3)
- Sigmoidal: $K(a,b) = \tanh(ka \cdot b + c)$ (4)

Na Figura 10 segue um exemplo de mapeamento de um conjunto de dados não lineares para um espaço de maior dimensão.

Figura 10 - Mapeamento de conjuntos de dados para espaço de maior dimensionalidade.



Fonte: O autor (2016).

3.3 Redes Neurais Artificiais

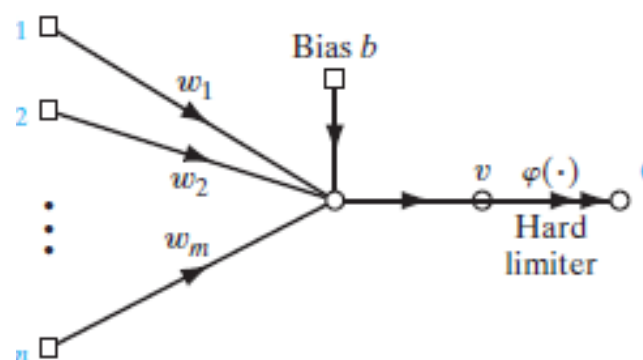
Assim como SVR apresentado na seção anterior, as redes neurais artificiais vêm sendo amplamente utilizadas na previsão de séries temporais devido a sua capacidade de modelar e prever padrões não lineares. A capacidade de generalização que permite tais modelagens, se dá pela estrutura das redes neurais, que por sua vez são formadas em sua grande maioria por simples unidades de processamento paralelas, interligadas por conexões criadas a partir dos padrões a ela apresentados, onde estas conexões na maioria dos casos possuem pesos associados que funcionam como um tipo de memória para aquela rede [25]. A inspiração por trás desta técnica é o funcionamento do cérebro humano e sua capacidade de plasticidade e aprendizado, onde a analogia a capacidade de aprendizagem por meio da apresentação de exemplos, se dá pela adaptação e realização de conexões por meio de pesos e a capacidade de plasticidade, que diz respeito a como o cérebro pode se recuperar de danos sofridos, se pelo processamento paralelo por unidades simples, que em alguns modelos são tratados como neurônios, onde mesmo que uma unidade venha a falhar, outra pode substituir sua função não prejudicando tanto a performance da rede como um todo.

Alguns exemplos de redes neurais são: Redes Elman [38] e Redes RBF [30].

Redes Multilayer Perceptron

É dita como Multilayer Perceptron (MLP), uma rede neural baseada na representação de neurônio artificial perceptron, que foi proposto por Rosenblatt (1958), como um modelo de aprendizagem supervisionada e anteriormente por McCulloch e Pitts (1943), onde foi introduzida a ideia de redes neurais como máquinas [25]. Rosenblatt utiliza em sua rede, o neurônio de McCulloch e Pitts, que mesmo quando utilizado em grupo com outros perceptrons, continua apenas resolvendo problemas lineares. A figura 11 representa a topologia do perceptron simples que possui n padrões de entradas, que são relacionados a um peso, e é apresentado a uma unidade intermediária de associação que possui um *bias* relacionado e por fim após essas entradas serem mapeadas por uma função linear, onde o resultado é aplicado em uma saída, que dá como saída -1 caso o resultado da combinação seja negativa e +1 caso positiva [18].

Figura 11 - Representação de Single Layer Perceptron de Rosenblatt

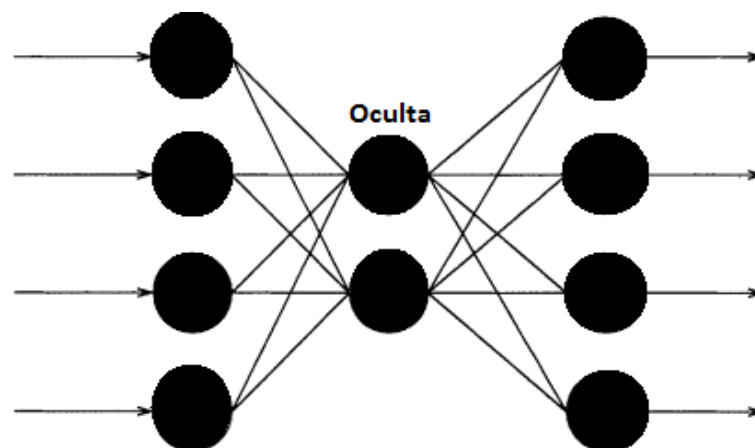


Fonte: S. Haykin, *Neural Networks - A Comprehensive Foundation*.

Como observado o perceptron de uma camada ou simples, não é capaz de atuar sobre padrões não linearmente separáveis [35]. A solução encontrada para que fosse possível o tratamento de padrões não lineares foi uma nova arquitetura denominada *Multilayer Perceptron*, pois esta combina um número definido de perceptrons com funções de ativação não lineares, em no mínimo três camadas, por isso o nome *Multilayer* que pode ser traduzido como “de muitas camadas”. Uma das características do MLP é que este possui uma ou mais camadas escondidas, que não entram em contato direto com os padrões de entrada ou de saída. Em suma as camadas escondidas gradativamente mapeiam por meio de uma função não linear os

valores a estes passados em um espaço denominado “*Feature Space*”, que possui uma maior dimensionalidade diminuindo assim a complexidade do problema. Para predição de séries temporais a rede MLP, vêm se mostrando uma boa opção quanto a modelagem e previsão de padrões com características não lineares, e isto se dá devido a sua capacidade de generalização, que é proveniente de sua estratégia de mapeamento de padrões não lineares para um espaço de alta dimensionalidade seguida por ajustes realizados sobre os pesos que definem as sinapses da rede de acordo com os erros encontrados durante o treinamento. Por se tratar de uma técnica de aprendizado supervisionado, assim como o SVR apresentado na seção anterior, a maneira com que os dados são apresentados a uma rede MLP, é basicamente a mesma, sendo apresentados valores passados e valores alvos a serem relacionados com estes. Na Figura 12 é apresentada a topologia do MLP na sua forma mais simples, com apenas uma camada escondida ou oculta.

Figura 12 - Representação da Topologia do Multilayer Perceptron



Fonte: R Beale, T Jackson. Neural Computing: An Introduction

Os modelos apresentados nesta seção foram explanados de maneira geral e introdutória, para uma visão mais aprofundada seguem as referências: [3,23,18,25].

4. Abordagens Híbridas

Como já introduzido em sessões anteriores, a utilização de estratégias híbridas para a previsão de séries temporais, vem se apresentando válida e com resultados muito interessantes como visto em [10,11,22]. A ideia de combinação da previsão de modelos, tem como objetivo a melhoria da previsão acerca dos valores futuros de uma determinada série temporal quando comparada à obtida individualmente pelos

modelos combinados. A melhoria é possível a partir da captura de uma quantidade maior de especificidades dos padrões apresentados, devido as diferentes configurações encontradas nos modelos combinados. A combinação de preditores já é um assunto explanado a um longo período de tempo [9], porém muitas das combinações apresentadas acabam por utilizar técnicas muito semelhantes como em [14,15], o que segundo alguns trabalhos como Zhang (2003) [10], não aparenta ser uma abordagem muito eficaz em se tratando da modelagem de problemas que envolvem padrões do mundo real. Sendo assim a combinação de preditores muito semelhantes, pode acabar negligenciando relações importantes entre os padrões utilizados. Tendo em vista que estes padrões podem ter características lineares e não lineares, e a maioria dos modelos encontrados na literatura não possuem capacidades tão amplas de generalização, que os permita a modelagem de padrões tão diferentes.

A utilização de uma abordagem híbrida visa permitir o aproveitamento dos pontos fortes de modelos diferentes como apresentado em Zhang (2003) [10] e Bijari (2008) [22]. Na abordagem híbrida desenvolvida neste trabalho, se pretende realizar a separação das características não lineares das lineares, permitindo assim que estas características possam ser modeladas por preditores adequados.

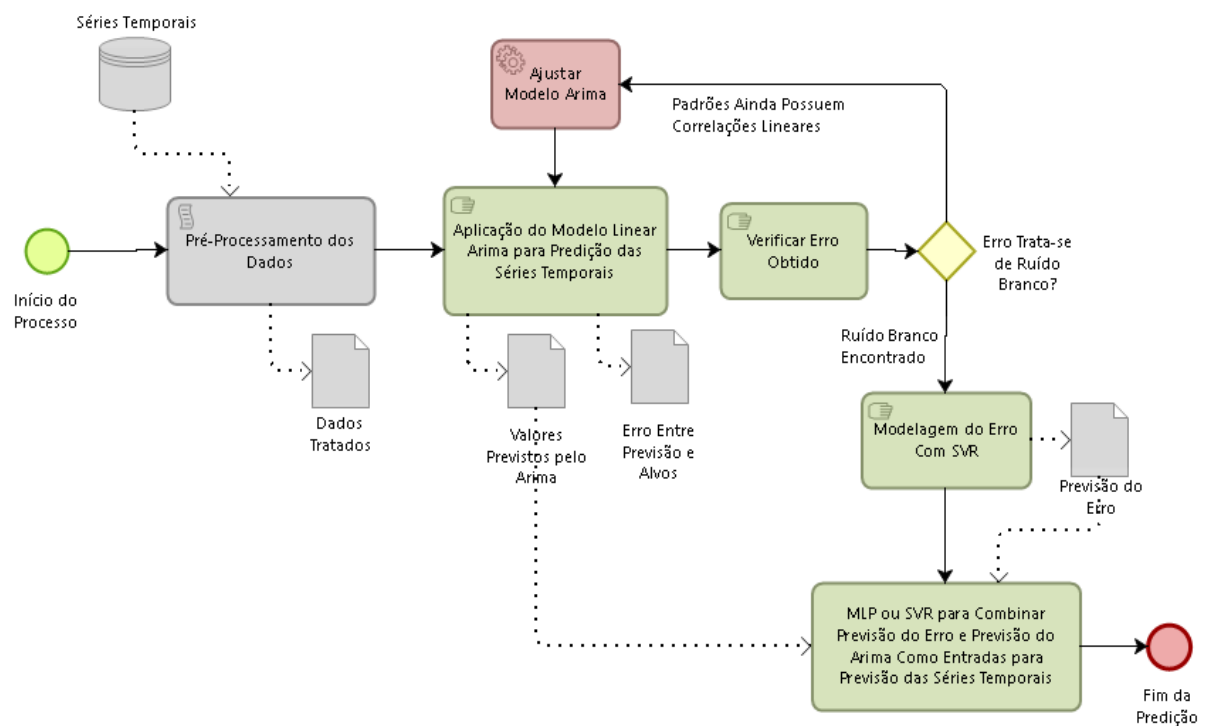
Em [10], é visto que em uma série temporal pode ser formada por padrões com correlações lineares e não lineares, caso seja utilizado um modelo linear para a predição desta série de maneira correta, é possível realizar a separação das correlações não lineares das lineares, a partir da diferença entre os valores alvos e os valores previstos por este modelo linear, onde a $diferença_t = alvo_t - previsão_t$. Em posse desta diferença, que por definição contém apenas correlações não lineares, é aplicado um modelo com uma boa capacidade na previsão de padrões não lineares, como por exemplo uma rede neural artificial. No trabalho proposto por Zhang (2003), após possuir a previsão do modelo linear e a previsão advinda da modelagem dos resíduos do modelo linear, é realizada uma soma destes valores obtendo-se assim a previsão final da abordagem híbrida que se provou melhor que a previsão tanto do modelo linear como da rede neural artificial utilizadas individualmente.

4.1 Estratégia Híbrida Proposta

Neste trabalho, é aplicada uma alternativa para a combinação dos padrões lineares e não lineares a partir de uma técnica de aprendizado de máquina. O objetivo

é usar uma rede neural para encontrar a melhor função de combinação entre a previsão da série temporal e a previsão dos resíduos. Nesse contexto, foram investigadas duas técnicas de aprendizado de máquina: uma rede neural do tipo MLP e a SVR, para combinar as previsões da série temporal e dos resíduos. Na Figura 13 segue um esquema geral sobre a abordagem híbrida utilizada neste trabalho.

Figura 13 - Estratégia Híbrida Implementada



Fonte: O autor (2016).

Uma descrição do modelo proposto é apresentada pelos seguintes passos:

- 1- A princípio as séries temporais são pré-processadas e podem ser definidas como $Z_t = L_t + NL_t$. Onde a série temporal é formada por L_t que são as correlações lineares e por NL_t as correlações não lineares.
- 2- Após o processamento dos dados estes são apresentados ao modelo linear ARIMA se obtendo então uma previsão adequada para as correlações lineares dos dados apresentados, tornando possível a separação das características lineares das não lineares pela diferença da previsão do ARIMA para o valor real da série. Então $NL_t = Z_t - L_t$.

- 3- Após se obter NL_t ou $erro_t = alvo_t - previsão_t$, o erro ou resíduos que formam o conjunto de características não lineares caso na etapa de verificação do erro obtido seja encontrado ruído branco. Um conjunto de dados é determinado como ruído branco quando não existe nenhuma correlação linear entre os seus padrões e estes padrões encontrados possuem média zero em diferentes janelas aplicadas, para descrições mais completas sobre ruído branco [31].
- 4- Então é utilizado o modelo SVR, na modelagem e previsão de NL_t ou $erro_t$, que é composto pelas correlações não lineares.
- 5- Após se obter a previsão da parte não linear do problema pelo SVR e a parte linear pelo ARIMA. É desenvolvido um modelo MLP e SVR com a função de realizar a combinação entre estas previsões obtidas.

5. Metodologia dos Experimentos Realizados

Esta seção apresenta a metodologia seguida durante a realização de todos os experimentos relativos a este trabalho de graduação.

5.1 Tratamento dos Dados

Como se espera de qualquer trabalho relativo à análise de dados, sejam eles relacionados a tempo ou não, o sucesso de um método de previsão se encontra intrinsecamente relacionado com a maneira que os dados são pré-processados e divididos. Muitos preditores e métricas utilizadas neste experimento, possuem uma dependência relacionada a normalização dos dados com sua distribuição ao redor de zero, e além disso outro fator importante é a facilitação quanto a comparação com outros trabalhos que utilizem a mesma normalização. Tendo estes fatores em vista, as três series temporais utilizadas, foram normalizadas entre os intervalos de zero a um, pela utilização da seguinte equação:

$$V_{normalizado} = \frac{v - v_{min}}{v_{max} - v_{min}} \quad (5)$$

Devido às características das séries utilizadas, se fez necessária uma suavização de seus valores para logaritmo de base 10 para *Canadian Lynx* e logaritmo natural para *BP/USD Exchange Rate*.

5.2 Divisão dos Conjuntos de Dados

Durante as etapas dos experimentos, os dados foram divididos da seguinte maneira para cada uma das técnicas:

- ARIMA: 70% dos dados para treinamento e 30% para teste.
- SVR: 50% dos dados para treinamento, 20% para validação e 30% para testes.
- MLP: 50% dos dados para treinamento, 20% para validação e 30% para testes.

5.3 Ordem dos Experimentos

Abaixo será apresentada a ordem dos experimentos realizados para elaboração do presente trabalho, tanto dos modelos individualmente como da abordagem proposta.

1- A primeira etapa do experimento, foi a predição das três séries anteriormente citadas e já tratadas pelo SVR. Foram definidas as janelas utilizadas durante todo o experimento, tais janelas foram escolhidas de acordo com observações relativas a performance quanto as configurações selecionadas. Os resultados desta etapa foram todos armazenados e utilizados apenas para comparação futura com combinações realizadas.

2- A segunda etapa é a previsão das séries utilizando o modelo ARIMA, após realizada a previsão das séries é observada a diferença entre a série real e a predição do ARIMA, verificando se os padrões caracterizam um ruído branco[31], então caso se trate sim esta etapa pode ser dada como concluída, e caso não deve se continuar realizando alterações em modelos até que seja encontrado um conjunto de dados com características de ruído branco. A importância de se chegar neste ponto é de se atestar que os padrões lineares já tenham sido capturados na predição realizada utilizando o modelo ARIMA.

3- A terceira etapa do experimento se dá pela predição do erro do ARIMA pela técnica de aprendizagem de máquina SVR, onde se espera que sejam tratados os padrões não lineares das séries temporais previamente tratadas.

4- A quarta etapa se trata da soma da predição do erro do ARIMA pelo SVR com a Predição da série do ARIMA que é a abordagem já utilizada na literatura, tendo seus valores armazenados para futura comparação.

5- A quinta etapa se trata da combinação das predição do erro do ARIMA pelo SVR com a predição da série realizada pelo ARIMA, através de MLP e SVR.

5.4 Seleção e Variação de Parâmetros:

- O modelo ARIMA utilizado foi baseado no modelo utilizado por Zhang (2003) [10].
- As predições de qualquer modelo estão muito relacionadas com a seleção de seus parâmetros, levando isto em consideração para a seleção dos parâmetros do SVR durante o treinamento, foi utilizado o script “GridSearchCV”, que é componente do módulo scikit-learn que faz parte da biblioteca em python Scipy. O que este script permite, é uma busca exaustiva sobre a combinação de um conjunto de parâmetros fornecidos, onde por fim podem ser evidenciados parâmetros ótimos para o conjunto de treinamento apresentado. No caso do SVR foram variados três parâmetros: C, ϵ e γ ; sendo mantido inalterado apenas o kernel. Após realizado o treinamento os parâmetros do modelo criado foram checados e ajustados com a utilização do conjunto de validação.
- A rede neural Multilayer Perceptron foi utilizada apenas na combinação da previsão do erro do ARIMA pelo SVR e a previsão das séries do ARIMA. O algoritmo de treinamento escolhido foi o backpropagation, para otimização dos pesos foi escolhido o “stochastic gradient descent” e a função de ativação foi a identidade. Os parâmetros variados foram

basicamente a taxa de aprendizagem, que foi mantida constante durante o treinamento, a quantidade de camadas ocultas e a quantidade de neurônios nessas camadas escondidas.

(As previsões em todos estes modelos foram de um passo a frente, assim sendo, todas as variações de parâmetros mantiveram uma única saída.)

5.5 Métricas Utilizadas

A escolha das medidas utilizadas para verificar a acurácia das previsões de uma série temporal no decorrer do tempo, são extremamente importantes tendo em vista que os modelos desenvolvidos serão ajustados buscando melhorar cada vez mais estas medidas. Após a observação de alguns trabalhos na área de predição de séries temporais [19], pode ser vista a utilização de diferentes métricas. Neste trabalho serão utilizadas seis métricas que serão listadas a seguir [20].

Mean Squared Error (MSE)

Esta métrica é relativamente simples, sendo composta apenas pelo somatório da diferença entre o valor alvo e o valor previsto, elevado a dois e dividido pela quantidade de padrões apresentados.

$$MSE = \frac{1}{N} \sum_{j=1}^N (target_j - output_j)^2 \quad (6)$$

Mean Absolute Percentage Error (MAPE)

O MAPE como o nome sugere se trata de uma média do valor absoluto do erro percentual entre a previsão de um modelo e o seu respectivo alvo.

$$MAPE = \frac{1}{N} \sum_{j=1}^N \left| \frac{target_j - output_j}{target_j} \right| \quad (7)$$

Index of Agreement (IA)

Esta métrica é uma métrica padronizada que serve para medir o grau do erro na predição de um modelo, onde seus resultados devem ser encontrados entre 0 e 1. Em

que um valor de 1 indica uma combinação perfeita enquanto 0 não indica nenhuma concordância.

$$IA = 1 - \left(\frac{\sum_{j=1}^N (target_j - output_j)^2}{\sum_{j=1}^N (|target_j - output_j| + |target_j - \overline{target}|)} \right) \quad (8)$$

Prediction of change in direction (POCID)

Esta medida avalia a capacidade do modelo em prever as variações quanto ao crescimento ou decrescimento dos valores futuros da série prevista, capturando assim o que seria a tendência da série estudada.

$$POCID = 100 \frac{\sum_{j=1}^N D_j}{N}$$

$$D_j = \begin{cases} 1, & \text{if } (target_j - target_{j-1})(output_j - output_{j-1}) > 0, \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Average relative Variance (ARV)

Esta medida compara a performance do modelo com a média da série temporal, onde caso seu resultado seja maior que um, a performance do modelo em prever valores da série são piores do que a média da série, caso este seja menor a acurácia é melhor e caso seja igual a 1 significa que o resultado é exatamente igual a média.

$$ARV = \frac{1}{N} \frac{\sum_{j=1}^N (output_j - target_j)^2}{\sum_{j=1}^N (output_j - \overline{target})^2}, \quad (10)$$

Estatística U de Theil

Esta medida se trata de uma medida estatística que compara o valor da predição de um modelo com o conhecido modelo random-walk [11], em termos gerais um modelo random-walk sempre considera como melhor valor para o dia seguinte, o valor para o dia atual. Quanto a seus valores, caso esta métrica seja menor que um,

significa que a mesma é melhor que um random walk ou caso seja maior indica que esta predição é pior.

$$\text{Theil} = \frac{\sum_{j=1}^N (\text{target}_j - \text{output})^2}{\sum_{j=1}^N (\text{output}_j - \text{output}_{j-1})^2}, \quad (11)$$

6. Resultados dos Experimentos

Nesta seção as configurações dos modelos utilizados em cada etapa dos experimentos, são apresentadas por meio das tabelas: 1, 3 e 5. Onde em cada uma das tabelas, são detalhados os parâmetros escolhidos para a realização de cada um dos experimentos. Para todos os modelos e etapas do experimento as janelas de tempo definidas foram respectivamente, tamanho 3 para Canadian Lynx e Wolf's Sunspot e 10 para BP/USD Exchange Rate. Os resultados de cada etapa dos experimentos foram apresentados de duas maneiras sendo a primeira por meio das figuras compreendidas entre a Figura 14 e a Figura 31 e a segunda maneira por meio das tabelas: 2, 4 e 5 onde estas apresentam as métricas obtidas após a utilização dos modelos e estratégias para previsão, onde os valores das métricas que se encontram em **negrito** representam os melhores valores daquela métrica dentre os preditores.

6.1 Experimentos para a Série Temporal *British Pound / USD Exchange Rate*

As melhores configurações encontradas para os modelos utilizados nos experimentos realizados para a série *BP/USD Exchange Rate*, estão todos descritos na Tabela 1.

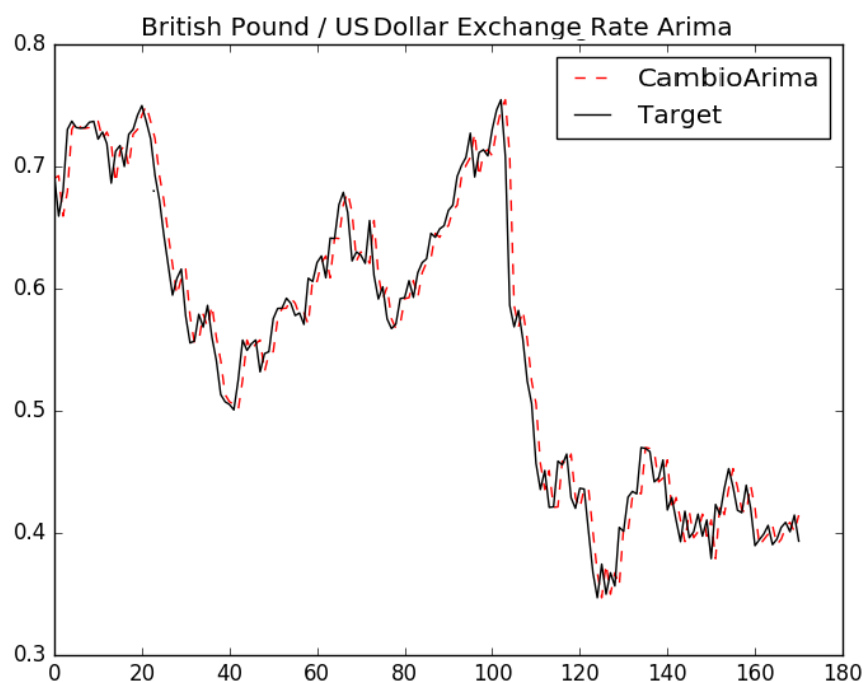
Configurações dos Modelos SVR nos Experimentos				
Experimento	Kernel	C	Gamma	Épsilon
Previsão do Erro	RBF	1000	0.9	0.001
Combinação da Previsão do Erro e do ARIMA	RBF	100	0.001	0.001
Previsão da Série	RBF	100	0.001	0.001

Configurações da Rede MLP para Combinação da Previsão Do ARIMA e do Erro	
Nº de Camadas Escondidas	Duas
Nº de Neurônios por Camada Escondida	(10,10)
Função de Ativação	Identidade
Taxa de Aprendizado	0.01
Otimização dos Pesos	Gradiente Descendente Estocástico

Tabela 1: Configuração dos Modelos para BP/USD Exchange Rate.

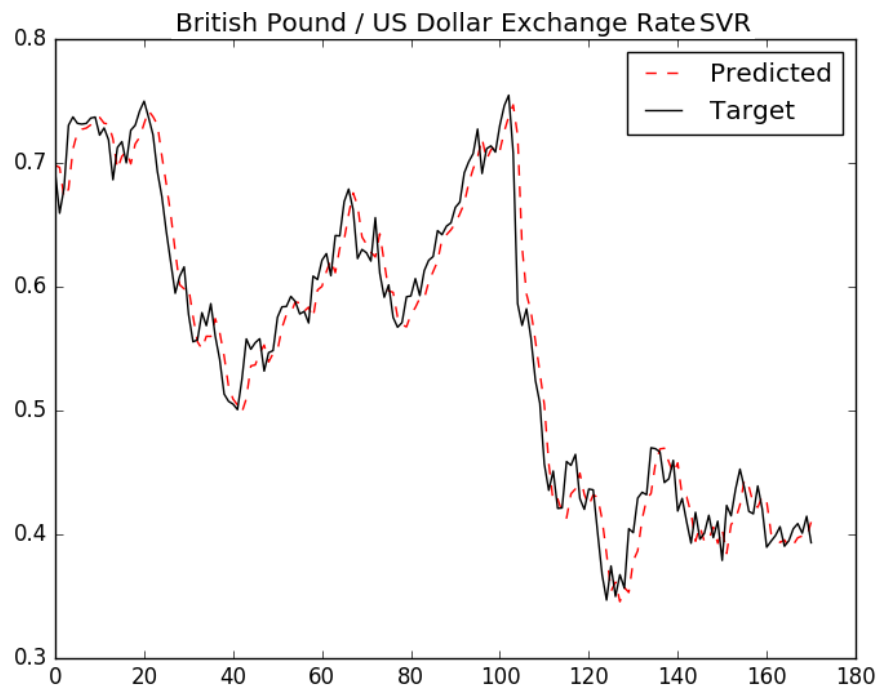
Uma das melhores maneiras de se verificar quão boa é a previsão realizada por um modelo é a construção de um gráfico comparativo entre os valores previstos e os valores alvos para o modelo em questão. Da Figura 14 até a Figura 19 são apresentadas tais comparações para todos os experimentos realizados para a série *BP/USD Exchange Rate*. As cinco primeiras figuras apresentam comparações individuais das etapas enquanto a Figura 19 unifica os resultados das previsões de todos os experimentos em um só local.

Figura 14 - Resultados das Previsões do ARIMA para a Série *BP/USD Exchange Rate*



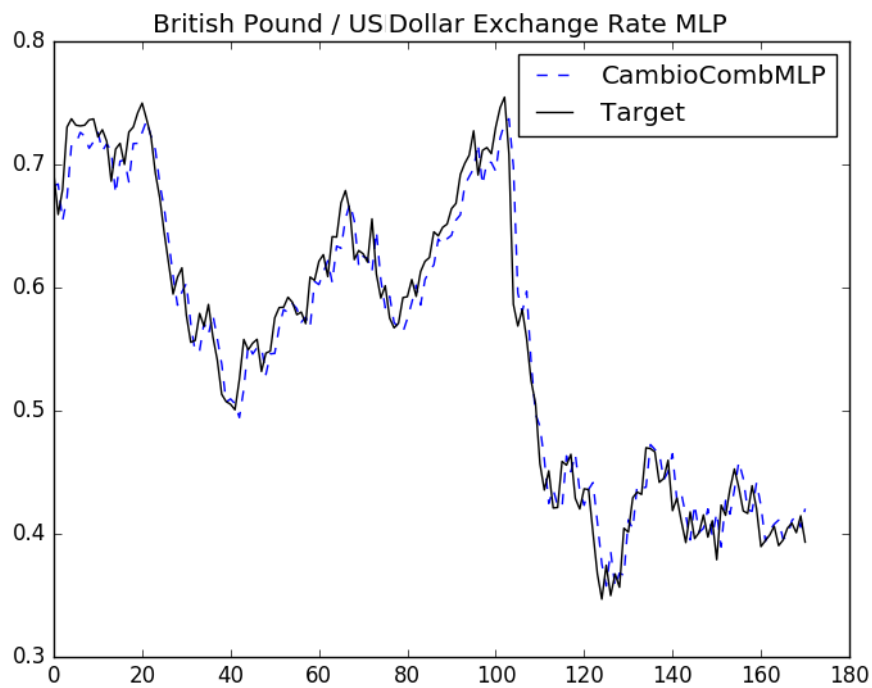
Fonte: O autor (2016).

Figura 15 - Resultados das Previsões do SVR para a Série BP/USD Exchange Rate



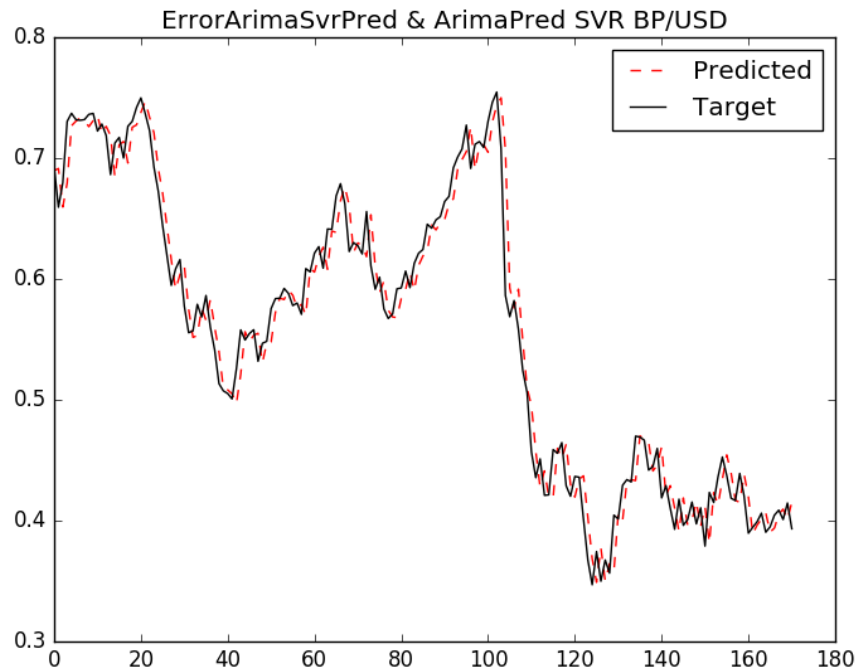
Fonte: O autor (2016).

Figura 16- Resultados da Combinação pelo MLP para a Série BP/USD Exchange Rate



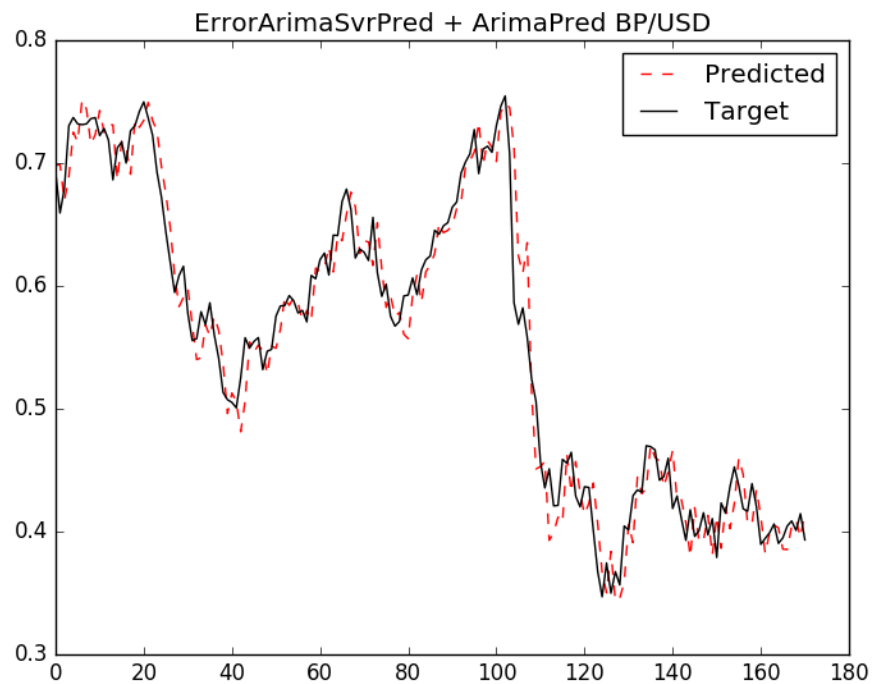
Fonte: O autor (2016).

Figura 17 - Resultados da Combinação pelo SVR para a Série *BP/USD Exchange Rate*

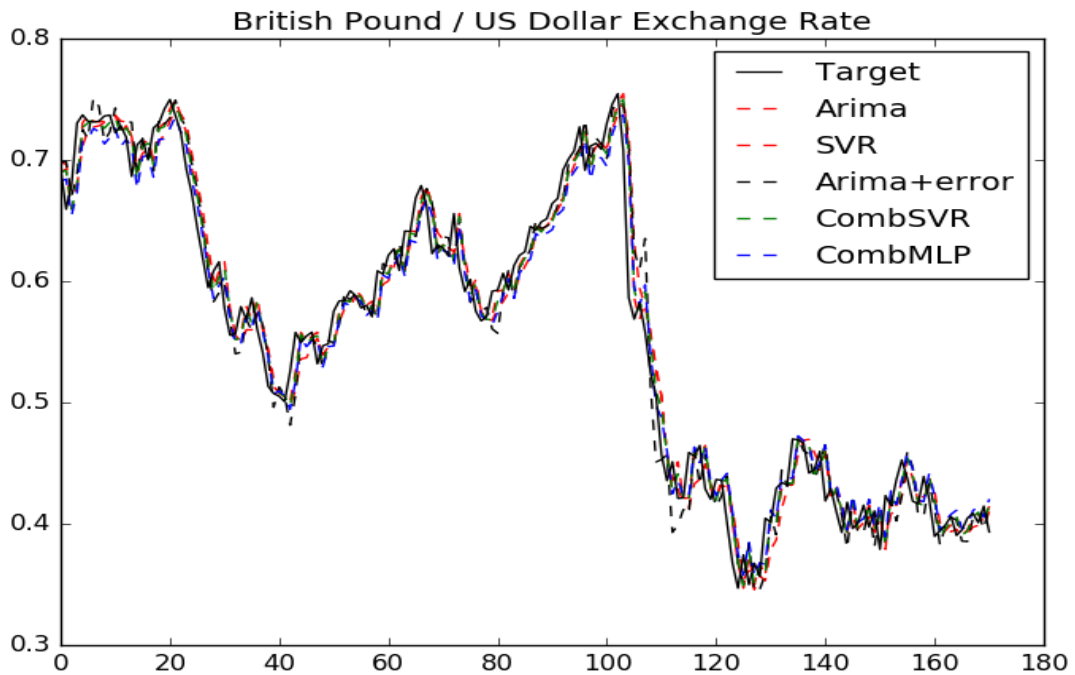


Fonte: O autor (2016).

Figura 18 - Resultados da Combinação pela Soma para a Série *BP/USD Exchange Rate*



Fonte: O autor (2016)

Figura 19 - Todas as Previsões para a Série *BP/USD Exchange Rate*

Fonte: O autor (2016).

Para uma melhor compreensão acerca dos resultados dos experimentos para a série temporal *BP/USD Exchange Rate*, foi elaborada a Tabela 2, que apresenta os valores encontrados pelas métricas aplicadas sobre o resultado dos modelos utilizados nos experimentos.

Resultados das Métricas Utilizadas para a Série <i>BP/USD Exchange Rate</i>						
Modelo	MSE	MAPE	POCID	Theil	ARV	IA
ARIMA	0.00048	3.20	46.47	1.00	0.03	0.99
SVR	0.00058	3.55	49.42	2.22	0.04	0.99
ARIMA+SVR	0.00066	3.82	45.88	0.97	0.04	0.99
MLP(ARIMA, SVR)	0.00052	3.40	44.11	1.14	0.04	0.99
SVR(ARIMA, SVR)	0.00048	3.23	42.94	1.00	0.03	0.99

6.2

Tabela 2: Resultado dos Experimentos para *BP/USD Exchange Rate*

Experimentos para a Serie Temporal Canadian Lynx

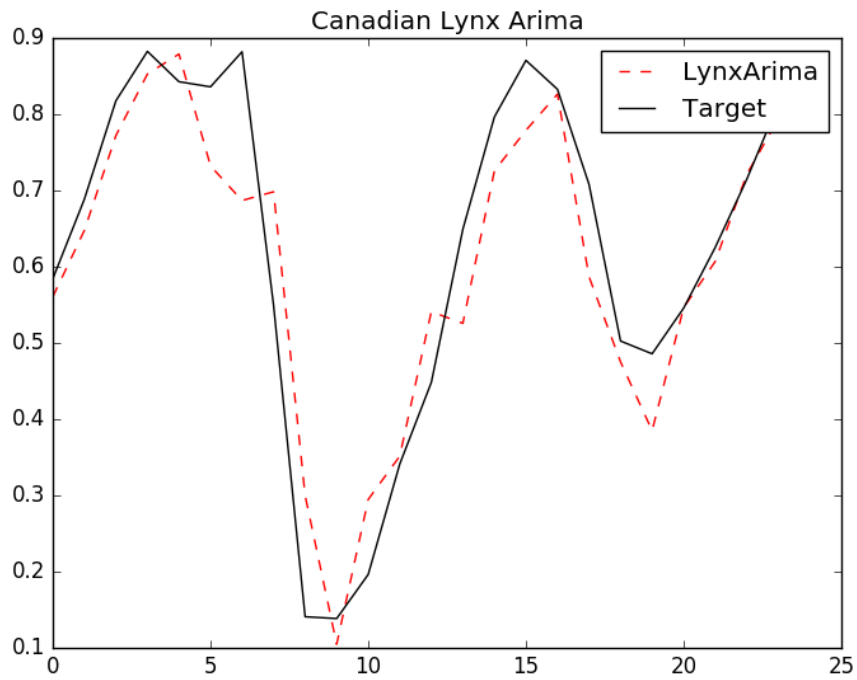
A Tabela 3 apresenta os parâmetros escolhidos para os modelos que apresentaram melhores resultados nas fases dos experimentos realizados para a série temporal *Canadian Lynx*.

Configurações dos Modelos SVR nos Experimentos				
Experimento	Kernel	C	Gamma	Épsilon
Previsão do Erro	RBF	600	0.9	0.001
Combinação da Previsão do Erro e do ARIMA	RBF	200	0.01	0.001
Previsão da Série	RBF	700	0.1	0.001
Configurações da Rede MLP para Combinação da Previsão Do ARIMA e do Erro				
Nº de Camadas Escondidas		Duas		
Nº de Neurônios por Camada Escondida		(5,5)		
Função de Ativação		Identidade		
Taxa de Aprendizado		0.1		
Otimização dos Pesos		Gradiente Descendente Estocástico		

Tabela 3: Configuração dos Modelos para Canadian Lynx

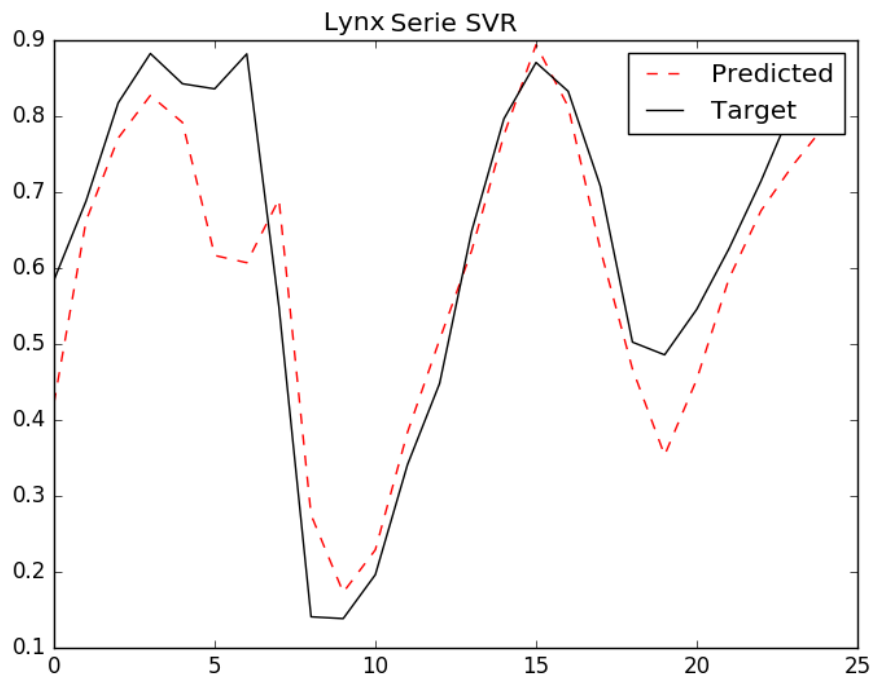
Nesta seção são apresentados para os experimentos realizados, os gráficos dos resultados comparativos entre as previsões e os valores reais para o conjunto de teste da série *Canadian Lynx*, onde a Figura 20 e 21 apresentam os resultados das previsões da série realizada individualmente pelos modelos Arima e SVR, nas figuras 22, 23 e 24 são respectivamente apresentados os resultados da combinação de previsões com MLP, SVR e pela soma das previsões, tendo por último a Figura 25 que centraliza todos os resultados das previsões em uma só figura.

Figura 20 - Resultados das Previsões do ARIMA para a Série *Canadian Lynx*



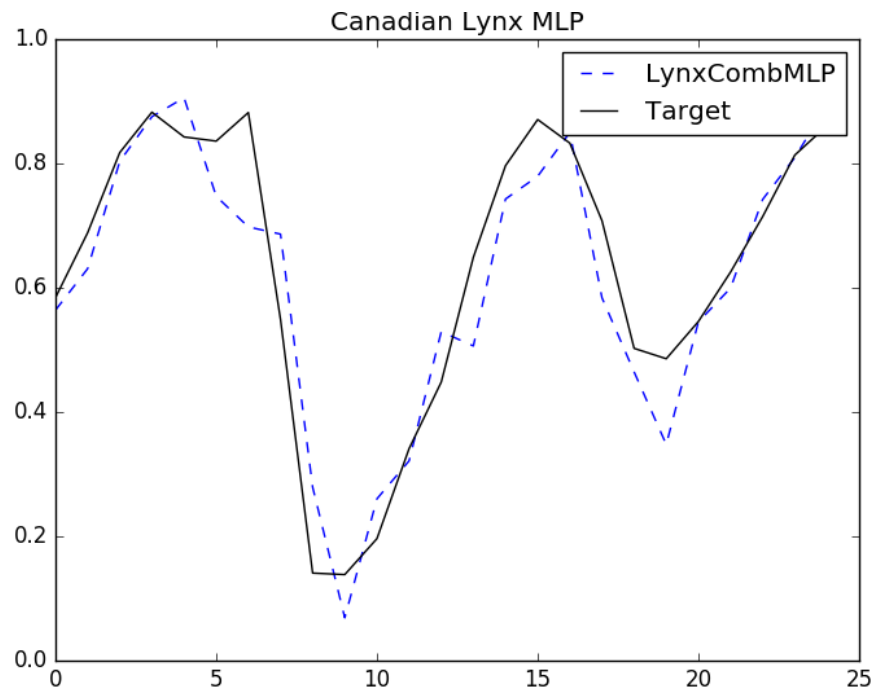
Fonte: O autor (2016)

Figura 21 - Resultados das Previsões do SVR para a Série *Canadian Lynx*



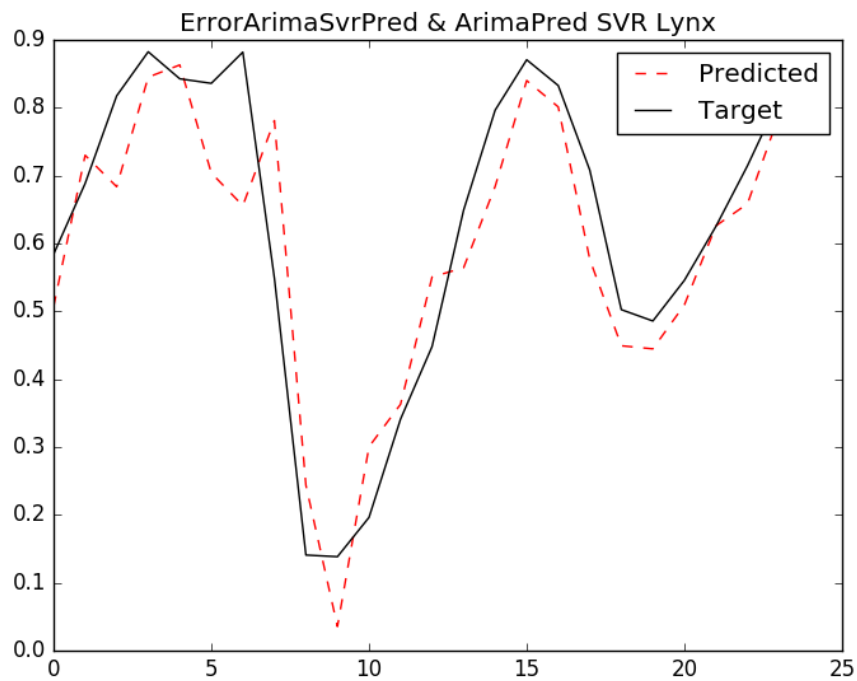
Fonte: O autor (2016).

Figura 22- Resultados da Combinação pelo MLP para a Série *Canadian Lynx*



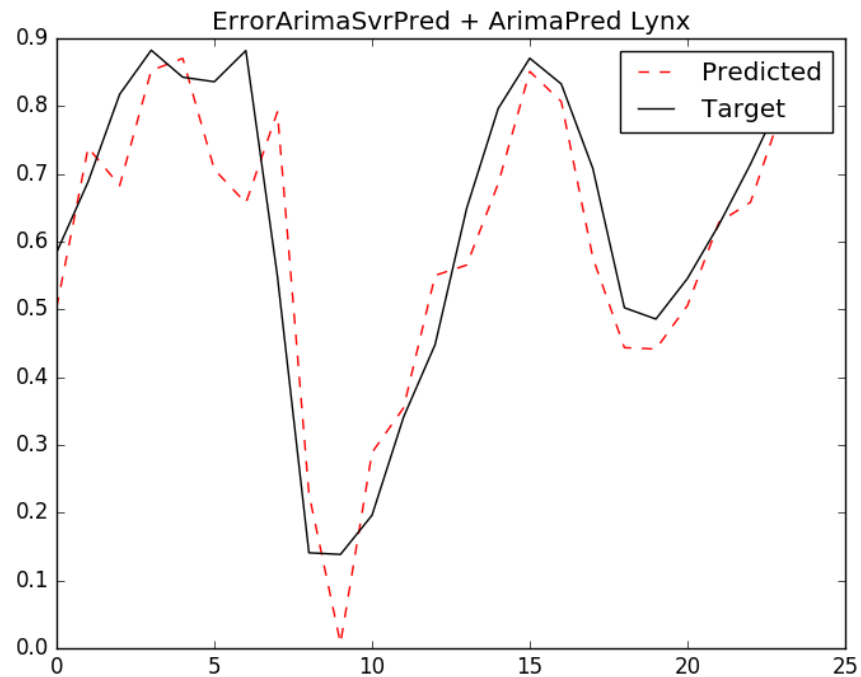
Fonte: O autor (2016).

Figura 23- Resultados da Combinação pelo SVR para a Série *Canadian Lynx*



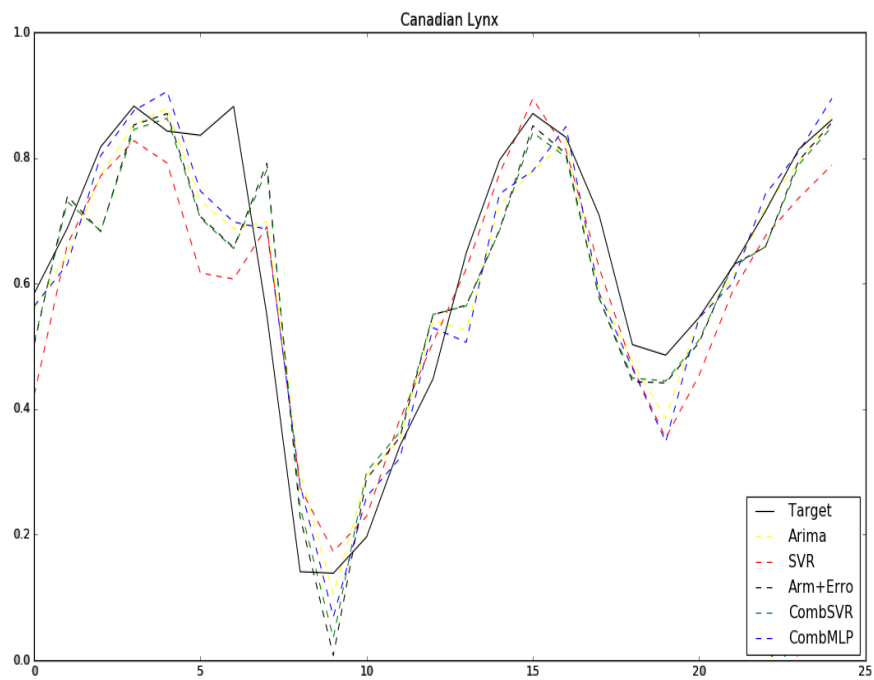
Fonte: O autor (2016).

Figura 24- Resultados da Combinação pela Soma para a Série *Canadian Lynx*



Fonte: O autor (2016).

Figura 25- Todas as Previsões para a Série *Canadian Lynx*



Fonte: O autor (2016).

Além da apresentação dos resultados dos experimentos por meio de gráficos, para se realizar uma análise sobre os resultados obtidos, também se faz necessária a utilização de métricas como as descritas na Tabela 4, que apresenta os valores das métricas para os experimentos realizados para a série *Canadian Lynx*.

Resultados das Métricas Utilizadas para a Série <i>Canadian Lynx</i>						
Modelo	MSE	MAPE	POCID	Theil	ARV	IA
ARIMA	0.007	15.43	79.16	0.35	0.17	0.96
SVR	0.010	15.97	91.66	0.48	0.25	0.83
ARIMA + SVR	0.009	18.00	83.33	0.31	0.21	0.91
MLP(ARIMA, SVR)	0.006	15.61	83.33	0.28	0.14	0.95
SVR(ARIMA, SVR)	0.009	17.86	83.33	0.34	0.22	0.92

Tabela 4: Resultados Dos Experimentos para Canadian Lynx

6.3 Experimentos para a Série Temporal *Wolf's Sunspot*

Estão apresentadas na Tabela 5, as configurações utilizadas nos modelos que demonstraram melhor performance dentre os experimentos realizados para a série temporal *Wolf's Sunspot*.

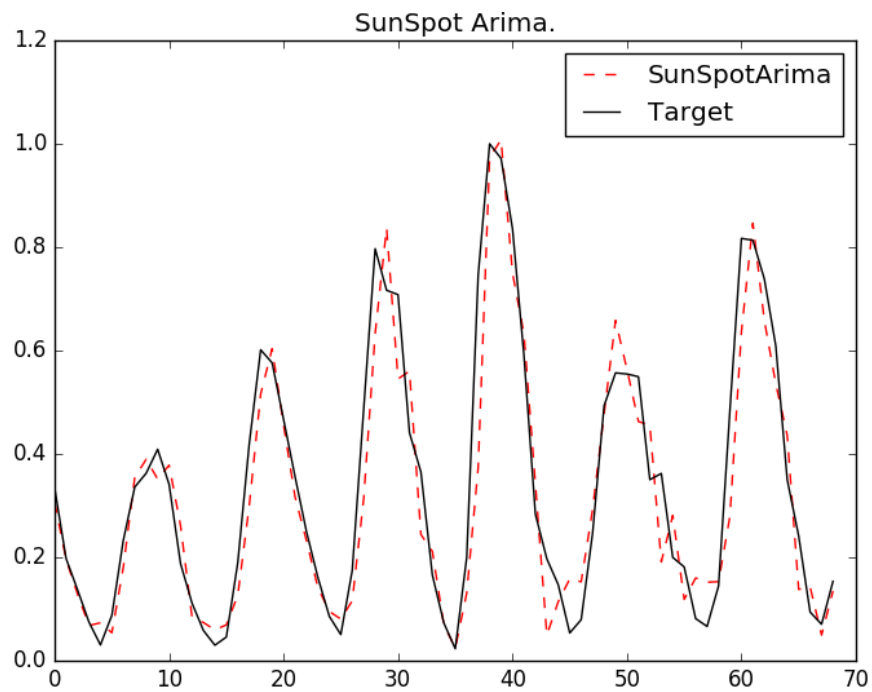
Configurações dos Modelos SVR nos Experimentos				
Experimento	Kernel	C	Gamma	Épsilon
Previsão do Erro	RBF	600	0.7	0.0001
Combinação da Previsão do Erro e do ARIMA	RBF	100	0.76	0.001
Previsão da Série	RBF	100	0.1	0.001
Configurações da Rede MLP para Combinação da Previsão Do ARIMA e do Erro				
Nº de Camadas Escondidas		Duas		
Nº de Neurônios por Camada Escondida		(15,15)		
Função de Ativação		Identidade		
Taxa de Aprendizado		0.1		

Otimização dos Pesos	Gradiente Descendente Estocástico
----------------------	-----------------------------------

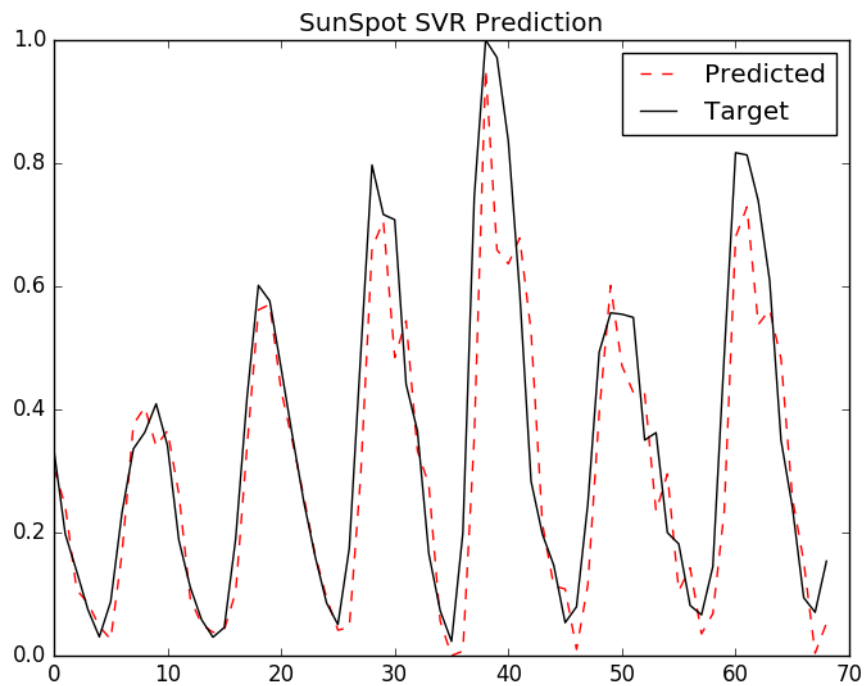
Tabela 5: Configuração dos Modelos para Wolf's Sunspot.

Entre as Figuras 26 e 31 são apresentadas por meio de gráficos, as comparações entre as previsões dos modelos utilizados nos experimentos, e os valores reais do conjunto de teste da série temporal *Wolf's Sunspot*.

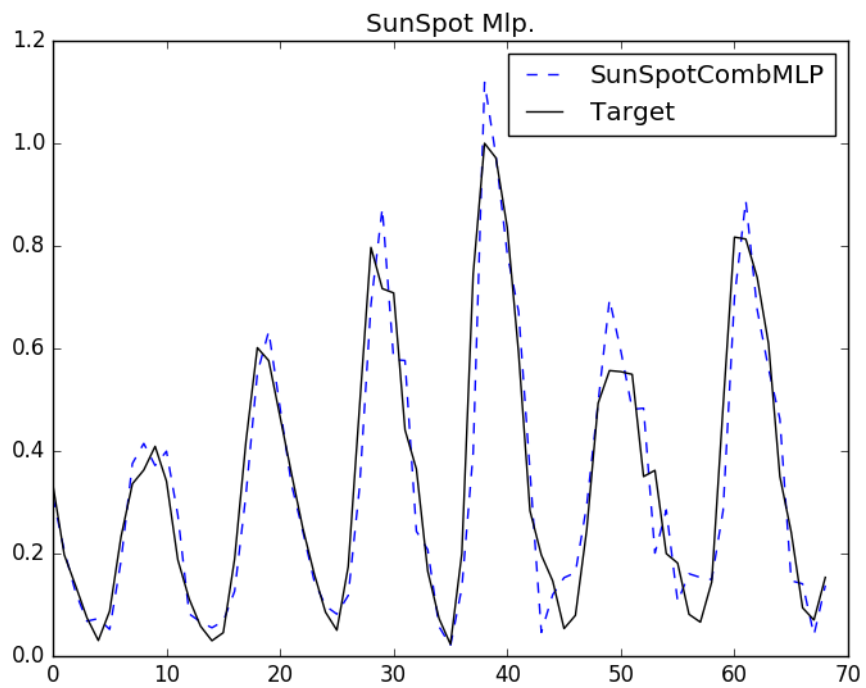
Figura 26 - Resultados das Previsões do ARIMA para a Série *Wolf's Sun Spot*



Fonte: O autor (2016)

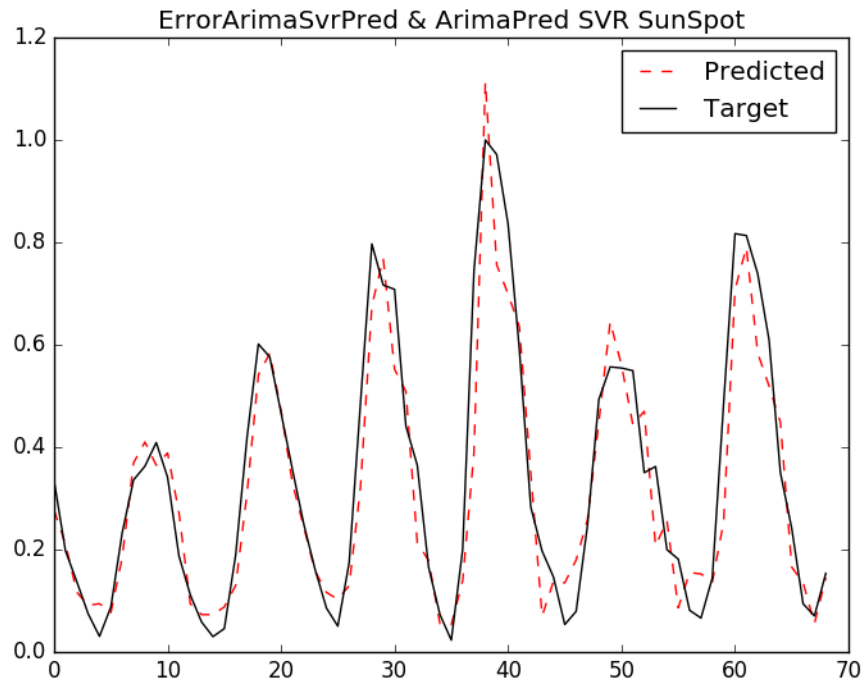
Figura 27 - Resultados das Previsões do SVR para a Série *Wolf's Sun Spot*

Fonte: O autor (2016).

Figura 28 - Resultados da Combinação pelo MLP para a Série *Wolf's Sun Spot*

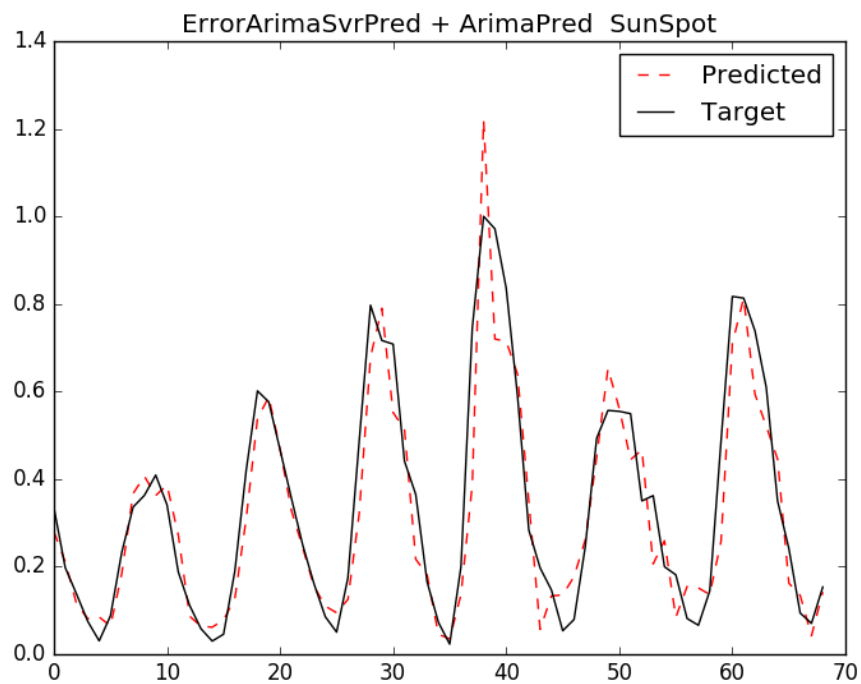
Fonte: O autor (2016)

Figura 29 - Resultados da Combinação pelo SVR para a Série *Wolf's Sun Spot*



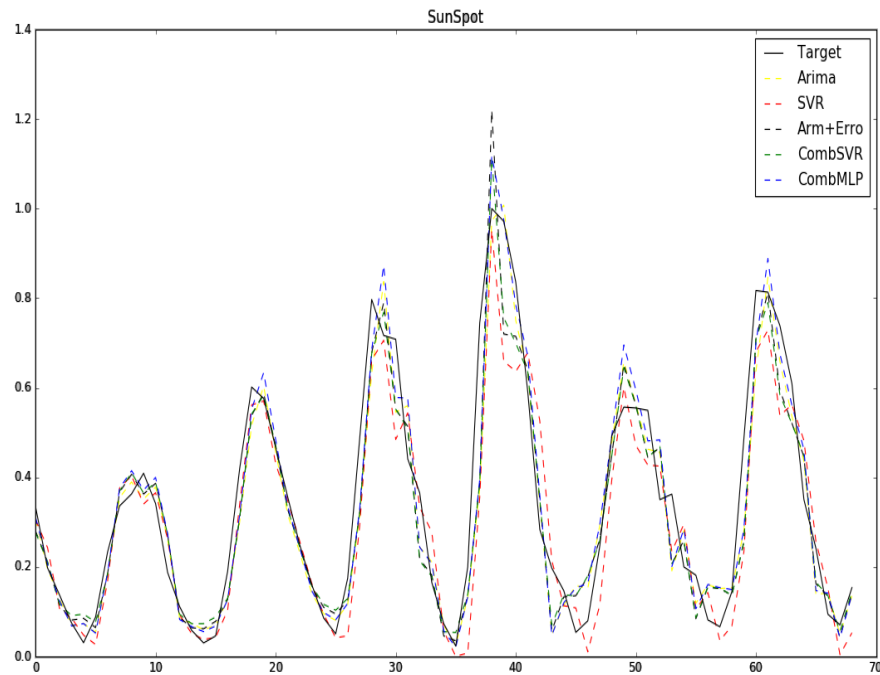
Fonte: O autor (2016).

Figura 30 - Resultados da Combinação pela Soma para a Série *Wolf's Sun Spot*



Fonte: O autor (2016).

Figura 31- Todas as Previsões para a Série *Wolf's Sun Spot*



Fonte: O autor (2016).

A Tabela 6, apresenta os valores obtidos pelas métricas aplicadas sobre a previsão obtida a partir dos modelos preditores utilizados em cada uma das etapas dos experimentos realizados para a série temporal Wolf's Sunspot.

Resultados das Métricas Utilizadas para a Série <i>Wolf's Sunspot</i>						
Modelo	MSE	MAPE	POCID	Theil	ARV	IA
ARIMA	0.008	30.73	76.47	0.30	0.14	0.89
SVR	0.012	32.21	80.88	0.43	0.23	0.58
ARIMA + SVR	0.009	32.63	79.41	0.27	0.16	0.88
MLP(ARIMA,SVR)	0.007	31.29	79.41	0.24	0.11	0.98
SVR(ARIMA, SVR)	0.008	34.64	79.41	0.29	0.16	0.88

Tabela 6: Resultados dos Experimentos para Wolf's Sunspot

6.4 Análise dos Resultados

Os resultados dos experimentos foram representados de duas maneiras, uma por meio dos gráficos comparativos entre os valores previstos e os valores para o conjunto

de teste das séries temporais utilizadas e por meio das tabelas contendo os valores das métricas aplicadas sobre cada modelo. Quanto aos gráficos apresentados, as Figuras: 14, 20 e 26, apresentam a comparação entre as previsões do modelo ARIMA. As Figuras: 15, 21 e 27 demonstram a previsão de um modelo SVR sobre as séries temporais utilizadas, este experimento não faz parte da abordagem híbrida proposta e aplicada, porém serve como um meio de comparação juntamente com o modelo ARIMA, para se verificar se a combinação das previsões destes pela abordagem híbrida, melhora ou não os valores encontrados nas métricas resultantes destes experimentos. As Figuras: 16, 22 e 28 e as Figuras: 17, 23 e 29 representam respectivamente a previsão realizada a partir da combinação da previsão da série do erro, com a previsão do modelo ARIMA utilizando modelagens do MLP e SVR. Outro experimento que não faz parte da abordagem proposta, porém foi realizado como meio de comparação, foi a realização da combinação pela soma das previsões do ARIMA e do erro pelo SVR, com seus resultados apresentados nas Figuras: 18, 24 e 30. As últimas utilizadas, são as Figuras: 19, 25 e 31 que apenas permitem a visualização de todas as previsões dos modelos utilizados para cada uma das séries temporais reunidas em um só gráfico.

Após a análise dos resultados obtidos nas figuras citadas anteriormente e nos valores das métricas encontradas das Tabelas: 2, 4 e 6. Pode se inferir que:

Os resultados dos experimentos realizados para a série *BP/USD Exchange Rate*, demonstraram que a utilização da combinação das previsões do modelo híbrido proposto, por meio dos modelos MLP e SVR acabaram não resultando em melhorias significativas para esta série, pois devido a avaliação tanto dos gráficos como dos resultados encontrados pelas métricas, as previsões foram extremamente próximas ou até iguais a um *random walk*. O modelo que atingiu as melhores métricas quanto as suas previsões foi o modelo ARIMA, estando este à frente em duas das métricas observadas quando comparado com o segundo melhor modelo, que foi a combinação realizada com o SVR. O modelo ARIMA obteve melhores resultados em seu POCID e MAPE, tendo todas as demais métricas encontrado valores iguais quando comparado a combinação pelo SVR.

Já para os resultados obtidos com os experimentos sobre a série temporal *Canadian Lynx*, a estratégia híbrida proposta, demonstrou melhorias nos resultados das métricas ao ser utilizada a combinação das previsões pelo MLP obtendo os

melhores resultados de MSE, Theil's U e ARV, quando comparado a previsão das séries realizada por todos os demais modelos utilizados.

E por fim quanto aos resultados obtidos pelos experimentos realizados para a série temporal *Wolf's Sunspot*, a abordagem híbrida proposta juntamente com a combinação das previsões pelo modelo MLP, demonstrou melhorias em suas previsões nas métricas: MSE, Theil's U, ARV e IA. Quando comparado as previsões de todos os demais modelos utilizados nos experimentos realizados.

7. Conclusões

Após análise dos experimentos, foi possível observar que os resultados empíricos obtidos, confirmaram a proposição de que a realização da combinação das previsões da abordagem híbrida implementada, por meio de sua modelagem pela rede neural MLP e a técnica SVR aumentaram a acurácia das predições. Porém alguns pontos devem ser considerados, o primeiro acerca da análise dos resultados da previsão da série BP/USD Exchange Rate, que se trata de uma série temporal do setor financeiro, e possui características que prejudicam a sua previsão, fazendo com que se caia em resultados típicos de modelos *Random Walk* [32]. Outra questão se dá na relação entre as métricas utilizadas para medir a acurácia das predições, pois é muito difícil se conseguir melhorar todas as métricas abordadas ao mesmo tempo, então em alguns casos as abordagens combinadas podem acabar perdendo em alguma das métricas para os modelos individuais [37]. Uma última observação a ser realizada diz respeito a separação entre as correlações lineares e não lineares, que como visto em [10,20], mesmo que o erro encontrado a partir da diferença entre a previsão e o valor real por um modelo linear venha a possuir características de ruído branco, existe uma grande possibilidade de que diferentes fatores como falta de atributos, inconsistências ou até a limitação de um modelo linear, possam fazer com que mesmo se tendo passado pela verificação que detecta se o erro tem características de ruído branco, estes resíduos não sejam apropriados para modelagem.

7.1 Melhorias Propostas e Trabalhos Futuros

Devido às limitações observadas na seção anterior, uma primeira proposição de melhoria se dá no desenvolvimento de uma estratégia de correção para a previsão de séries temporais do domínio financeiro e econômico evitando características de *random walk*. Uma outra possível melhoria seria a utilização de alguma estratégia que corrigisse ou eliminasse problemas relativos a separação entre as correlações entre os padrões utilizados e por último uma outra melhoria, poderia se dar na aplicação de estratégias evolutivas para otimização na seleção de parâmetros e geração de modelos.

Referências

- [1] P. A. Morettin, C. M. Toloi. *Análise de Séries Temporais*. Editora Blucher, São Paulo. 2006.
- [2] Franses H. Philip, Dijk V. Dick, Opschoor Anne. *Time Series Models for Bysiness and Economic Forecasting*. Cambrige University Press, 2014 Second Edition, 77-104.
- [3] Box, G.E. e Jenkins, G.M. *Time series analysis, forecasting and control - Segunda Edição*. Prentice Hall San Francisco, 1976.
- [4] Montgomery, Douglas; Jennings, Cheryl; Kulahci, Murat. *Introduction To Time Series Analysis And Forecasting*, Second Edition. Willey 2015.
- [5] Planas Christophe, *Applied Time Series Analysis Modeling, Forecasting, Unovserved Components Analysis & the Wiener-Kolmogorov filter*, Luxembourg, 15 Outubro de 1997.
- [6] Ratnadip Adhikari; R. K. Agrawal. *An Introductory Study on Time Series Modeling and Forecasting*, LAP Lambert Academic Publishing, Germany, 2013.
- [7] C. Chatfield. *The Analysis of Time Series – An Introduction*. Chapman & Hall, quarta edição, 1989.
- [8] P. Newbold and C. W. J. Granger. Experience with Forecasting Univariate Time Series and the combination of Forecasts. *Journal of the Royal Statistical Society. Series A (General)* Vol. 137, No. 2 (1974), pp. 131-165
- [9] R. T. Clemen, Combining forecasts : a review and annotated bibliography ,*International Forecasting Journal* 5.(1989),559–583.
- [10] Zhang, G.Peter. *Time series Forecasting using a hybrid ARIMA and neural network model*. Elsevier, *Neurocomputing* 50, pg.159-175, 2003.
- [11] Firmino, Paulo Renato A., Paulo SG de Mattos Neto, and Tiago AE Ferreira. "Error modeling approach to improve time series forecasters." *Neurocomputing* 153 (2015): 242-254.
- [12] Bordley, R.F. (1982) "The combination of forecasts: A Bayesian approach", *Journal of the Operational Research Society*, 33, 171-174.
- [13] Reid, D.J. (1968) "Combining three estimates of gross domestic product", *Economica*, 35, 431-444.

- [14] Hui, Zou. Yuhong, Yang. Combining time series models for forecasting. International Forecasting Journal, Volume 20, issue 1, pg. 69-89, .
- [15] Kourentzes Nikolaos, Barrow K. Devon, Crone F. Sven. Neural, network ensemble operators for time series forecasting, Expert Systems with Applications Volume 41 issue 9, pg. 4235-4244, Jul 2014.
- [16] D. F. Andrews, A. M. Herzberg, Canadian Lynx Trappings Part of the series Springer Series in Statistics pp 13-16.
- [17] Friedli, T.K. Sol Phy Sunspot Observations of Rudolf Wolf from 1849 – 1893 (2016) 291: 2505
- [18] S. Haykin, Neural Networks - A Comprehensive Foundation. Prentice Hall, Terceira edição ed., 2008.
- [19] Davydenko, A., & Fildes, R. (2016). Forecast Error Measures: Critical Review and Practical Recommendations. In Business Forecasting: Practical Problems and Solutions. John Wiley & Sons Inc. 2016.
- [20] S. de A. Filho Francisco, Madeiro Francisco, M.M.F. Sérgio, S.G. de Mattos Neto Paulo, A.E. Ferreira Tiago. Time series forecasting of pollutant concentration levels using particle swarm optimization and artificial neural networks. Quím. Nova vol.36 no.6 São Paulo 2013
- [21] A. M. Kumar, Ashu Jain. Hybrid neural network models for hydrologic time series forecasting. Applied Soft Computing, Volume 7, March 2007, Pages 585-592.
- [22] M. Khashei, S. R. Hejazi, M. Bijari. A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. Fuzzy Sets and System, Volume 159, Issue 7, 1 april 2008, Pages 769-786.
- [23] Smola J. Alex, Schölkopf Bernhard. A tutorial on support vector regression Statistics and Computing (2004) 14. Pages 199-222
- [24] J. Han, J. Pei, M. Kamber. Data Mining Concepts and Techniques. Third Edition. Elsevier 2012.
- [25] T. Mitchell. Machine Learning. McGraw-Hill Science/Engineering/Math. March, 1997.

- [26] Jared Dean. Big Data, Data Mining and Machine Learning : Value Creation for Business Leaders and Practitioners. Wiley 2014.
- [27] Lakhal L, Stumme G. Efficient mining of association rules based on formal concept analysis. In: Formal Concept Analysis. Heidelberg: Springer Verlag; 2005, 180–195.
- [28] Y.Bengio, O. Delalleau, N. Le Roux. The Curse of Dimensionality for Local Kernel Machines. Technical Report, Dept. IRO, Université de Montréal 2005.
- [29] H Yu, S Kim - Svm tutorial—classification, regression and ranking. Handbook of Natural computing, Springer 2012.
- [30] Khadir T. Mohamed, Ghazi Sabri. Radial Basis Function and Elman Networks for Pollutant's Parameter Prediction in the Region of Annaba Algeria. ADVCOMP 2010 : The Fourth International Conference on Advanced Engineering Computing and Applications in Sciences.
- [31] C. W. Gardiner Handbook of Stochastic Methods: for Physics, Chemistry and Natural Sciences Second Edition. Springer.
- [32] Kilian Lutz, Taylor P. Mark. Why is so difficult to beat the random walk forecast of exchange rates? Empirical Exchange Rate Models. Volume 60, Issue 1, Pages 85-107.
- [33] Mattos Neto S.G. Paulo, Cavalcanti D.C. George, Madeiro Francisco, Ferreira A. E. Tiago. Na Approach to Improve the Performance of PM Forecasters. PLOS ONE September 28, 2015.
- [34] Sondwale P. Pradnya. Overview of Predictive and Descriptive Data Mining Techniques. International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 4, April 2015.
- [35] Beale R, Jackson T. Neural Computing : An Introduction 1990.
- [36] VAPNIK, Vladimir et al. Support vector method for function approximation, regression estimation, and signal processing. Advances in neural information processing systems, p. 281-287, 1997.
- [37] Silva, David A., et al. "Measurement of fitness function efficiency using data envelopment analysis." *Expert Systems with Applications* 41.16 (2014): 7147-7160.
- [38] GAO, X. Z.; GAO, X. M.; OVASKA, S. J. A modified Elman neural network model with application to dynamical systems identification. In: Systems, Man, and Cybernetics, 1996., IEEE International Conference on. IEEE, 1996. p. 1376-1381.