

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

BRUNO GABRIEL FERNANDES

**Estudo da Aplicabilidade de
Sistemas Imunológicos Artificiais
Voltados ao Ajuste de Séries Temporais**

Monografia apresentada como requisito parcial para a obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Dante Augusto Couto Barone

Porto Alegre, Julho de 2014.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Graduação: Prof. Vladimir do Nascimento

Diretor do Instituto de Informática: Prof. Luís da Cunha Lamb

Coordenador do Curso de Ciência da Computação: Prof. Raul Fernando Weber

Bibliotecário-Chefe do Instituto de Informática: Alexander Borges Ribeiro

AGRADECIMENTOS

Gostaria de começar agradecendo aos Professores do Instituto de Informática, que nunca falham na sua excelência acadêmica e na sua capacidade de inspirar. Gostaria de agradecer em especial ao Prof. Dr. Dante Barone por me apresentar à área de Computação Evolutiva, que aproxima dois conceitos que me agradam muito.

Gostaria de agradecer à minha Família, e à minha mãe Tânia Fernandes por me dar os meios e incentivos que me permitiram seguir meus objetivos e não me desviar do caminho correto.

Gostaria fazer uma menção especial ao meu chefe e grande mentor Gustavo Silveira Neto, que me permitiu faltar horas do meu trabalho para me dedicar este trabalho.

RESUMO

O objetivo deste trabalho é apresentar o conceito de Sistemas Imunológicos Artificiais, dando ênfase na sua aplicação no problema de Ajuste de Séries Temporais. É feita uma introdução ao conceito de Séries Temporais, bem como de métodos tradicionais utilizados no problema de Previsão de Séries Temporais. Por fim, é feita uma análise de como o Algoritmo de Seleção Clonal pode ser adaptado ao problema de Ajuste, e uma implementação deste Algoritmo é proposta e analisada.

Palavras-chave: Sistemas Imunológicos Artificiais. Séries Temporais. Ajuste de Séries Temporais. Seleção Clonal. Computação Evolutiva. Algoritmo.

Artificial Immune Systems applied to Time Series Prediction

ABSTRACT

The purpose of this paper is to present the concept of Artificial Immune Systems, with special regard to its use in the problem of Time Series Fitting. An introduction to the concept of Time Series is made, as well as to traditional methods used in the problem of Time Series Fitting. Then, an analysis is made on how the Clonal Selection Algorithm may be used to Time Series Fitting, and an implementation of the Algorithm is proposed and analyzed.

Keywords: Artificial Immune Systems. Time Series. Time Series Fitting. Clonal Selection. Evolutionary Computation. Algorithm.

LISTA DE FIGURAS

Figura 2.1 – Exemplo de Série Estacionária.	15
Figura 2.2 - Exemplo de Série Não Estacionária.	16
Figura 2.3 - Gráfico da previsão de uma Série utilizando o Naive sem mudanças.	17
Figura 2.4 - Gráfico da previsão de uma Série utilizando o Naive com variação constante. .	17
Figura 2.5 – Aproximação Linear.	18
Figura 2.6 – Aproximação Parabólica.	18
Figura 2.7 - Rede neural típica.	21
Figura 2.8 - Fluxo para geração da previsão do próximo valor de uma série.	22
Figura 3.1 – Célula B e Anticorpo.	24
Figura 3.2 – Detalhe do Anticorpo.	25
Figura 3.3 – Reconhecimento do Antígeno pelas Células B.	26
Figura 3.4 – Proliferação e Diferenciação das Células B	27
Figura 3.5 – Utilizando XOR para medida de Afinidade.	29
Figura 3.6 – Exemplo de medida de Afinidade com variação de alinhamento.	30
Figura 4.1 – Exemplo de uma representação de uma equação em forma de Árvore.	33
Figura 5.1 – Exemplo de uma árvore contendo um modelo.	37
Figura 5.2 – Exemplo de um Anticorpo.	37
Figura 5.3 - $P_m(atc)$ para diferentes níveis de Afinidade.	41
Figura 5.4 – Exemplo de Série de Ruído Branco.	43
Figura 5.5 – Exemplo de Série de Ruído Branco Gaussiano.	44
Figura 5.6 – Exemplo de <i>Random Walk</i> (vermelho) e <i>Random Walk with Drift</i> (azul).	45
Figura 5.7 – Série <i>Random Walk with Drift</i> utilizada no experimento.	46
Figura 5.8 – Série <i>Random Walk with Drift</i> normalizada.	46
Figura 5.9 – Melhor Afinidade por Geração nas 5 execuções iniciais.	47
Figura 5.10 – Gráfico da Série Original (azul) e do modelo da Execução 3 (vermelho).	48
Figura 5.11 – Série IBovespa de Janeiro de 1993 a Dezembro de 1995.	49
Figura 5.12 – Série IBovespa Normalizada.	50
Figura 5.13 – Evolução da Afinidade do Melhor Anticorpo da População.	51
Figura 5.14 – Série Original Normalizada (azul) e Modelo da Execução 1 (vermelho).	52
Figura 5.15 – Série Original Normalizada (azul) e Modelo de Execução 2 (vermelho).	52
Figura 5.16 – Testando diferentes valores de m	54
Figura 5.17 – Testando Diferentes Valores de α	55

LISTA DE TABELAS

Tabela 5.1 – Erro Quadrático Médio e Afinidade por Execução	48
Tabela 5.2 – Erro Quadrático Médio e Afinidade por Execução	50

LISTA DE ABREVIATURAS E SIGLAS

ARIMA	<i>Autoregressive integrated moving average</i>
SI	Sistema Inmunológico
SIA	Sistema Inmunológico Artificial

SUMÁRIO

RESUMO	4
ABSTRACT	5
LISTA DE FIGURAS	6
LISTA DE TABELAS	7
LISTA DE ABREVIATURAS E SIGLAS	8
1 INTRODUÇÃO	11
2 SÉRIES TEMPORAIS.....	13
2.1 Notação	14
2.2 Função Conjunta de Distribuição	14
2.3 Séries Estacionárias e Não Estacionárias	14
2.4 Métodos de Análise de Séries Temporais.	16
2.4.1 Naive.....	16
2.4.2 Métodos de Regressão Polinomial	18
2.4.3 ARIMA.....	19
2.4.4 Redes Neurais.....	21
3 SISTEMAS IMUNOLÓGICOS ARTIFICIAIS	23
3.1 O Sistema Imunológico Natural	23
3.1.1 Linfócitos	23
3.1.2 Anticorpos	24
3.1.3 Reconhecimento Imunológico	25
3.1.4 Proliferação de Células B e Maturação de Afinidade	26
3.2 Algoritmo de Seleção Clonal	27
3.2.1 Representação dos Indivíduos	28
3.2.2 Medidas de Afinidade	28
3.2.3 Mecanismos de Seleção	30
3.2.4 Mecanismos de Mutação	31
4 APLICANDO SISTEMAS IMUNOLÓGICOS ARTIFICIAIS PARA O AJUSTE E PREVISÃO DE SÉRIES TEMPORAIS.....	32
4.1 Modelo	32
4.2 Representações das Moléculas de Anticorpo e Antígeno	33
4.3 Medidas de Afinidade	33
4.4 Mecanismos de Seleção	34
4.5 Mecanismos de Mutação	34
5 ESTUDO DE CASO	35
5.1 Detalhes da Implementação	35
5.1.1 Normalização da Série	35
5.1.2 Modelo	35
5.1.3 Representação do Anticorpo	35

5.1.4 Medida de Afinidade	38
5.1.5 Mecanismo de Seleção para Proliferação.....	39
5.1.6 Proliferação Clonal	40
5.1.7 Mecanismo de Mutação	40
5.1.8 Seleção dos Sobreviventes	42
5.1.9 Resumo da Implementação	42
5.1.10 Linguagens de Programação e Ferramentas	43
5.2 Experimentos	43
5.2.1 Ruído Branco e <i>Random Walk</i>	43
5.2.2 Mercado de Ações	49
6 CONCLUSÕES	56
REFERÊNCIAS	57

1 INTRODUÇÃO

No trabalho de (ASTIAZARA, 2012), o Autor propõem um método baseado em SIA para Predição de Fraudes e Furtos no sistema de energia, e em (OLIVEIRA, 2012) é apresentado um Classificador baseado em SIA.

Estudando Computação Evolutiva e Computação Bio-Inspirada, surgiu uma oportunidade para estudar melhor os métodos baseados em SIA e determinar qual é a gama de problemas que estes métodos são capazes de resolver.

Então, para este trabalho, foi decidido abordar outro problema ainda pouco explorado na Área de SIA, a Análise de Séries Temporais. Este trabalho busca descobrir qual é a aplicabilidade de Métodos de SIA no problema de Ajuste de dados de Séries Temporais, buscando explorar um Método que possa fazer o ajuste dos dados medidos de uma Série Temporal, com o objetivo de prever valores futuros nesta Série.

O objetivo deste SIA é evoluir um Modelo matemático que possa explicar os valores passados da Série e que possa ser extrapolado para valores futuros. O que isso é na prática é uma exploração do espaço de soluções possíveis, e o SIA se comporta como um mecanismo de busca neste espaço. Espera-se que com um SIA bem elaborado, o modelo do Melhor Anticorpo da população convirja para a solução ótima.

Esta solução “ótima” nem sempre é ideal, pois na área de Previsão de Séries Temporais, na prática é impossível encontrar um modelo perfeito em quase todos os casos de Séries existentes, com exceção das mais simples, pois normalmente as Séries que se deseja prever dependem de muitas variáveis de outras Séries Temporais desconhecidas ou impossíveis de medir.

Então, um modelo é considerado “ótimo” se ele se adequa à Série Temporal de acordo com os critérios de Afinidade descritos neste capítulo, mas a sua verdadeira qualidade depende da acurácia de suas previsões. O que normalmente é feito para medir esta qualidade é separar o(s) último(s) valor(es) da Série que se deseja prever para testar a qualidade da previsão, como foi feito nos trabalhos de (Y. Q. SHI et al, 2010) e (LUO E SHI, 2011).

Em (Y. Q. SHI et al, 2010), o autor define algumas métricas que podem ser utilizadas para medir a qualidade da solução proposta por um Método de Previsão de Séries temporais como uma relação entre os valores previstos pelo método e os valores reais medidos futuramente.

Sendo \hat{x}_t um valor previsto pelo método para o tempo t , e x_t o valor medido da Série em um tempo t , algumas métricas que podem ser utilizadas para medir a qualidade da solução são:

- Erro Absoluto: $|\hat{x}_t - x_t|$
- Erro Absoluto Médio: $\frac{\sum_{t=1}^n |\hat{x}_t - x_t|}{n}$
- Raiz do Erro Quadrático Médio: $\sqrt{\frac{\sum_{t=1}^n (\hat{x}_t - x_t)^2}{n}}$

Este Trabalho é organizado da seguinte maneira:

No Capítulo 2, é feita uma introdução ao conceito de Séries Temporais, e são apresentados alguns métodos utilizados para a Análise de Séries Temporais.

No Capítulo 3, é feita uma introdução a Sistemas Imunológicos Artificiais, com uma breve descrição do Sistema Imunológico Natural.

No Capítulo 4 são feitas considerações necessárias para adaptar Sistemas Imunológicos Artificiais para o problema de Previsão de Séries Temporais.

No Capítulo 5, é proposta uma implementação de método de Previsão de Séries temporais baseado em Sistemas Imunológicos Artificiais.

No Capítulo 6 são apresentadas algumas conclusões a respeito do trabalho.

2 SÉRIES TEMPORAIS

Uma Série Temporal é uma sequência ordenada de observações tomadas de uma ou mais variáveis, sendo esta ordem normalmente determinada pela tomada das observações ao longo do tempo (WEI, 2006). Estas Séries podem ser observadas em uma gama variada de campos, como por exemplo:

- Agricultura: produção anual e preços.
- Controle de Qualidade: medição de qualidade de um processo ao longo do tempo.
- Economia: preço de fechamento de ações, taxas flutuantes de juros e índices de preços.
- Engenharia: sinais sonoros e elétricos, voltagem.
- Geofísica: medição de atividades sísmicas e vulcânicas, marés.
- Medicina: medição de eletroencefalogramas e eletrocardiogramas.
- Meteorologia: velocidade de ventos, temperaturas.
- Sociologia: taxas de nascimento e mortalidade, de acidentes e criminalidade.

Séries Temporais que podem ser amostradas continuamente (tal como voltagens) são chamadas de Séries Temporais Contínuas, enquanto que séries que são obtidas em intervalos de tempo específicos (como preço de fechamento de ações) são chamadas de Séries Temporais Discretas (WEI, 2006).

Existem diversos motivos para o estudo de séries temporais, tal como o entendimento de processos complexos, a previsão de valores futuros e a otimização do controle de sistemas (WEI, 2006).

Este trabalho focará no problema de como encontrar uma equação que melhor se ajuste aos dados de uma Série Temporal e tentar realizar a previsão de valores futuros da Série, dado como entrada um histórico de valores passados de uma Série

Para que isso seja possível, neste trabalho assume-se a premissa de que cada valor observado é correlacionado e codependente aos valores anteriores, portanto a ordem em que os valores são apresentados é importante. Isso significa que os valores futuros da Série são na verdade dependentes dos valores passados, e então os valores da série poderiam ser explicados por alguma função dos valores anteriores.

A previsão de Séries Temporais tem diversas aplicações, entre elas (CHATFIELD 2000):

- Planejamento Econômico, previsão de Despesas e Lucros.
- Previsão de vendas.
- Controle de estoque, previsão de entrada/saída do estoque.
- Planejamento de produção e capacidade.
- Avaliação de estratégias econômicas alternativas.
- Criação de Orçamento, previsão de Gastos.
- Gerenciamento de risco

- Meteorologia, previsão de Temperaturas Máximas e Mínimas

Os métodos utilizados para a previsão de Séries Temporais podem ser classificados nas seguintes categorias (CHATFIELD 2000):

- *Previsões Subjetivas*, que são baseados na avaliação subjetiva ou intuitiva. Um método muito conhecido é o *Delphi Technique* (ROWE E WRIGHT, 2000), baseado na obtenção de um consenso entre *experts* através do uso de questionários.
- *Métodos Univariados*, onde as previsões dependem somente dos valores atuais e passados da série que se pretende avaliar, como por exemplo, obtendo um modelo que dependa somente dos valores passados da série e extrapolando-o para o futuro.
- *Métodos Multivariados*, onde as previsões dependem dos valores atuais e passados da Série, além de uma ou mais Séries Temporais adicionais, chamadas de Séries Predictoras ou Séries Explanatórias. Um exemplo seria obtendo um modelo que depende dos valores passados de duas ou mais séries, e extrapolando este modelo para valores futuros.

2.1 Notação

Ao longo deste texto, uma série temporal será definida como uma sequência de variáveis x_1, x_2, x_3, \dots , onde a variável x_1 denota o valor obtido da série temporal no instante de tempo t_1 , a variável x_2 denota o valor obtido no instante de tempo t_2 , a variável x_3 denota o valor obtido no instante de tempo t_3 , e assim em diante (SHUMWAY E STOFFER, 2010).

2.2 Função Conjunta de Distribuição

Outra maneira de descrever uma série temporal é pela função conjunta de distribuição. Dado os n valores de uma série temporal, obtidos nos tempos t_1, t_2, \dots, t_n , a função conjunta de distribuição é dada pela probabilidade conjunta que os valores da série sejam menores ou iguais aos valores de n constantes, c_1, c_2, \dots, c_n (SHUMWAY E STOFFER, 2010):

$$F(c_1, c_2, \dots, c_n) = P(x_1 \leq c_1, x_2 \leq c_2, \dots, x_n \leq c_n) \quad (2.1)$$

2.3 Séries Estacionárias e Não Estacionárias

Séries estritamente estacionárias ou fortemente estacionárias são séries em que o comportamento probabilístico de toda coleção de variáveis:

$$\{x_1, x_2, \dots, x_k\}$$

é idêntico ao do conjunto deslocado:

$$\{x_{1+h}, x_{2+h}, \dots, x_{k+h}\}$$

Ou seja,

$$P\{x_{t_1} \leq c_1, x_{t_2} \leq c_2, \dots, x_{t_k} \leq c_k\} = P\{x_{t_{1+h}} \leq c_1, x_{t_{2+h}} \leq c_1, \dots, x_{t_{k+h}} \leq c_k\} \quad (2.2)$$

para todo $k = 1, 2, \dots$, e todos os deslocamentos de tempo $h = \pm 1, \pm 2, \dots$ (SHUMWAY E STOFFER, 2010).

Já nas séries não estacionárias, a média das variáveis e o desvio padrão dos valores muda conforme o tempo passa (SPROTT 2003). As figuras 2.1 e 2.2 apresentam exemplos de duas Séries, uma estacionária e outra não estacionária, com as suas médias e variâncias.

Figura 2.1 – Exemplo de Série Estacionária (Vermelho), Média (Azul) e Variância (Verde).

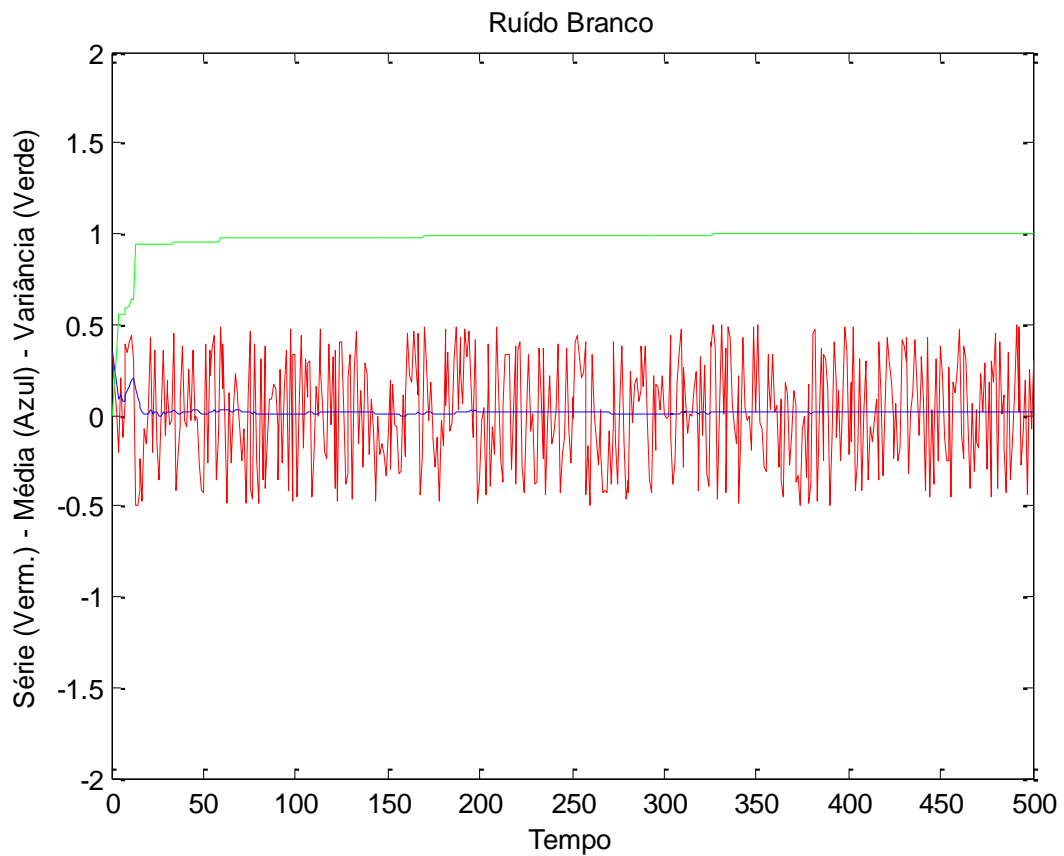
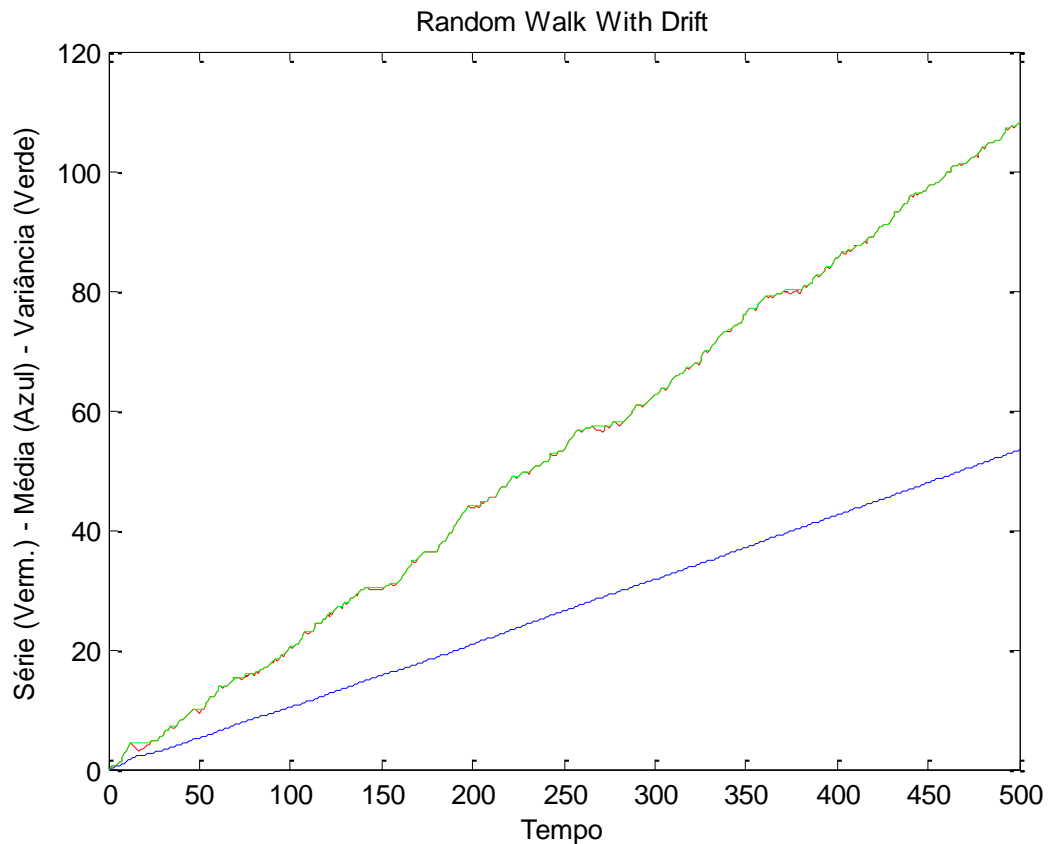


Figura 2.2 – Exemplo de Série Não Estacionária (Vermelho), Média (Azul) e Variância (Verde).



2.4 Métodos de Análise de Séries Temporais

Neste capítulo são apresentados alguns dos métodos clássicos utilizados na Análise, Ajuste e Previsão de Séries Temporais.

2.4.1 Naive

O método Naive (ou *ingênuo*) é o método de Previsão de Séries Temporais mais simples de todos, e é utilizado como base de comparação com outras técnicas. Existem duas variações mais utilizadas desta técnica (PROCHNOW 2013):

- **Naive sem mudanças:** neste método muito simples, os valores futuros serão sempre iguais ao último valor obtido da série.
- **Naive com variação constante:** neste método, a variação entre os valores futuros é sempre igual a variação do último intervalo medido.

Nas figuras 2.3 e 2.4 pode-se ver uma Série Temporal medida (em vermelho) e as previsões (em azul), usando as duas variações do Naive.

Figura 2.3 - Gráfico da previsão de uma Série utilizando o Naive sem mudanças.

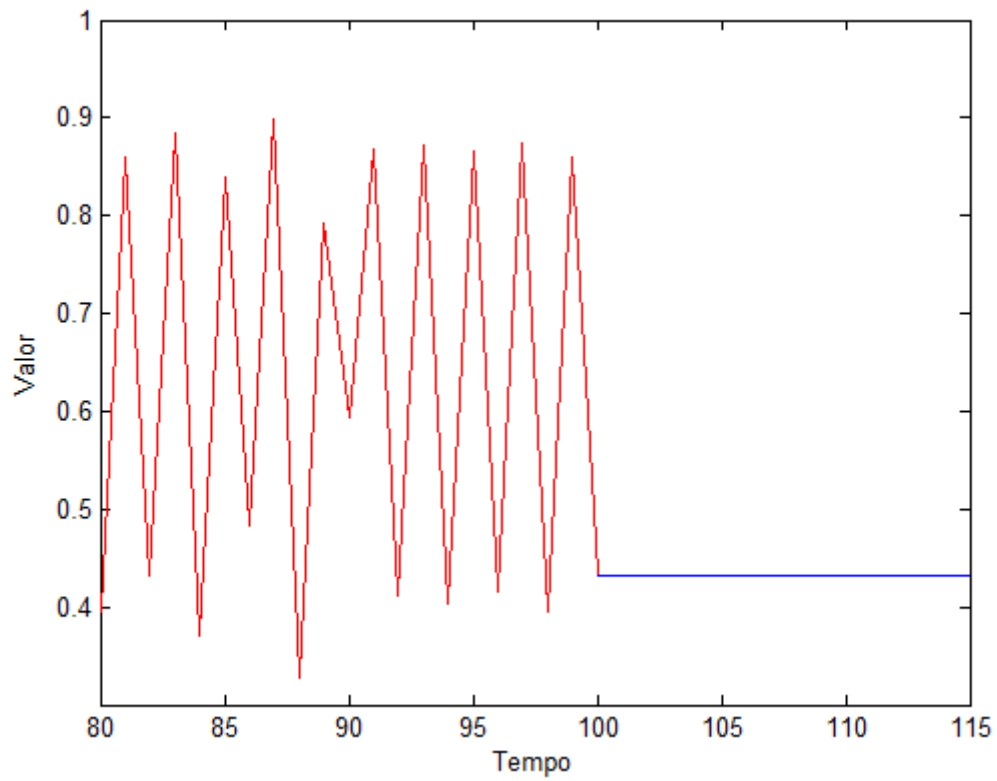
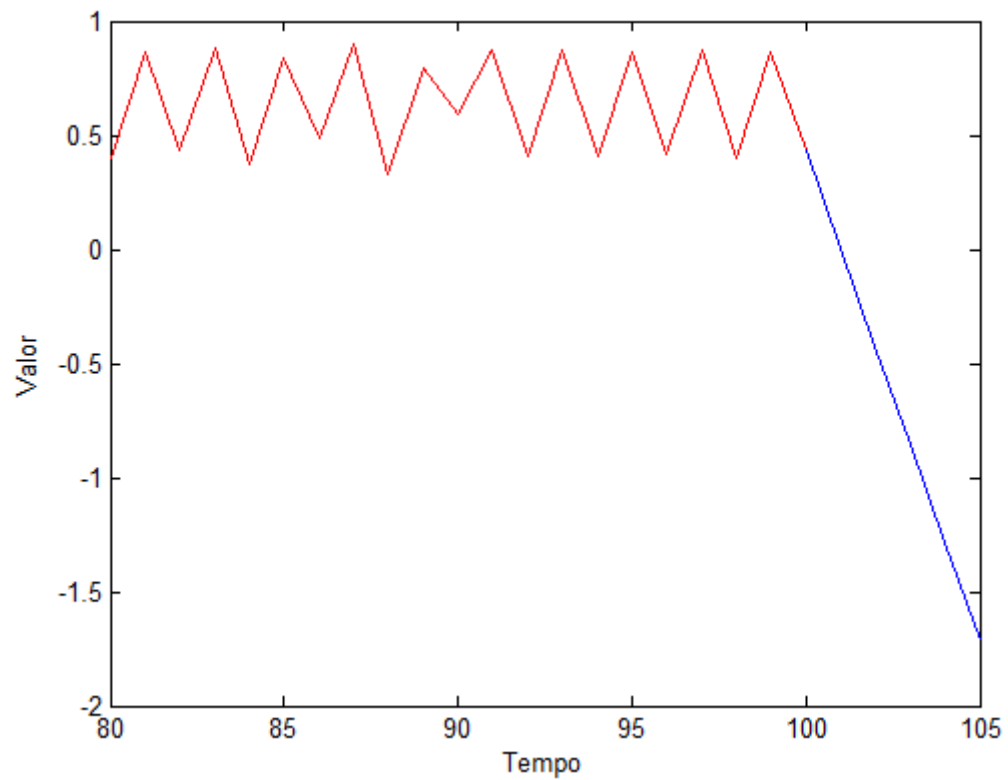


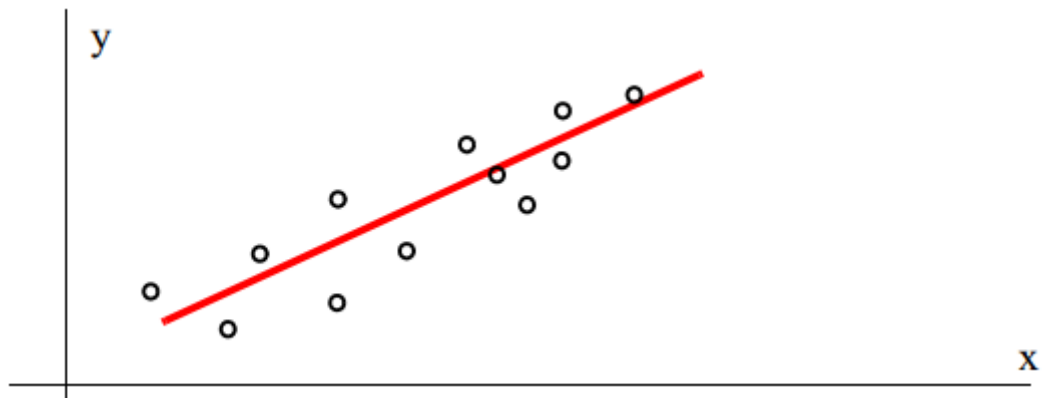
Figura 2.4 - Gráfico da previsão de uma Série utilizando o Naive com variação constante.



2.4.2 Métodos de Regressão Polinômial

Curve Fitting é o processo de encontrar linhas e curvas que melhor se ajustam a um conjunto de dados (KARRIS, 2007). Por exemplo, na figura 2.5, pode-se usar a equação de uma reta na forma $y = ax + b$:

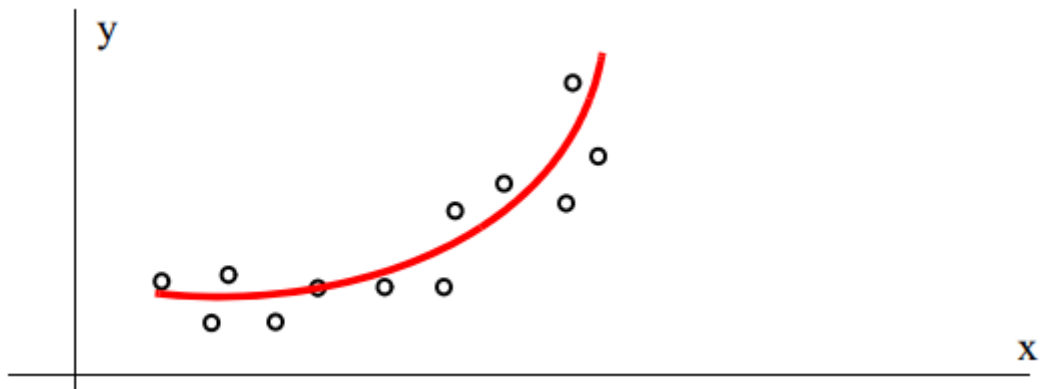
Figura 2.5 – Aproximação Linear.



Fonte: Karris (2007, p. 8-1).

Na figura 2.6, pode-se usar uma equação quadrática do tipo $y = ax^2 + bx + c$

Figura 2.6 – Aproximação Parabólica



Fonte: Karris, 2007 (p. 8-1).

Para encontrar a melhor equação, assume-se que os dados representam uma variável independente x , e a tarefa de encontrar uma equação que descreva a variável dependente y é chamada de **Regressão** (KARRIS, 2007).

No processo de Regressão, deseja-se encontrar a curva que melhor se assemelha aos dados. Para isso, queremos minimizar a equação:

$$d_1^2 + d_2^2 + \dots + d_n^2 \quad (2.3)$$

onde d_1 é a distância entre a medição x_1 e a curva, d_2 é a distância entre a medição x_2 e a curva, e assim por diante. Essa curva é chamada da curva de mínimos quadrados (KARRIS, 2007).

2.4.3 ARIMA

A regressão clássica frequentemente é insuficiente para explicar toda a dinâmica de uma série temporal. Métodos de regressão linear simples muitas vezes falham em capturar estruturas adicionais na série temporal (SHUMWAY E STOFFER, 2010).

O modelo ARIMA (*Autoregressive integrated moving average*, ou modelos autoregressivos de média móvel integrados) foi popularizado no trabalho de Box e Jenkins de 1970, que introduziu um método para identificar um modelo ARIMA plausível (SHUMWAY E STOFFER, 2010).

Nesta seção será feita uma breve introdução dos conceitos básicos por trás dos modelos ARIMA.

2.4.3.1 Modelos Autoregressivos

Modelos autoregressivos são baseados na ideia de que o valor atual de uma série, x_t , pode ser explicado como uma função dos últimos p valores da mesma série, onde p determina o número de passos no passado necessários para prever o valor atual (SHUMWAY E STOFFER, 2010).

Um modelo auto regressivo de ordem p , abreviado como $AR(p)$ é da forma:

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + w_t \quad (2.4)$$

Onde x é uma série estacionária, x_t é o valor da série no tempo t , $\varphi_1, \varphi_2, \dots, \varphi_p$ são constantes ($\varphi_p \neq 0$), e w_t é uma série de ruído branco.

2.4.3.2 Modelos de Média Móvel

Como uma alternativa aos modelos autoregressivos em que se assume que x_t no lado esquerdo da equação é combinado linearmente, o modelo de média móvel de ordem q , abreviado como $MA(q)$, assume que o ruído branco w_t no lado direito da equação é combinado linearmente para formar os dados observados (SHUMWAY E STOFFER, 2010).

Um modelo de média móvel de ordem q , ou $MA(q)$, é da forma:

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q} \quad (2.5)$$

onde existem q atrasos na média móvel e que $\theta_1, \theta_2, \dots, \theta_q$ ($\theta_q \neq 0$) são constantes, e w_t é uma série de ruído branco (SHUMWAY E STOFFER, 2010).

2.4.3.3 Modelo ARMA

O modelo misto auto regressivo e de média móvel, ARMA, é obtido como uma combinação dos dois modelos vistos anteriormente.

Uma série temporal $\{x_t; t = 0,1,2 \dots\}$ é *ARMA*(p, q) se for estacionária e se

$$x_t = \varphi_1 x_{t-1} + \dots + \varphi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q} \quad (2.6)$$

com $\varphi_p \neq 0$, $\theta_q \neq 0$ e a variância de w_t é maior que zero (SHUMWAY E STOFFER, 2010).

Os parâmetros p e q são chamados de ordem regressiva e ordem de média móvel, respectivamente. Quando $q = 0$ o modelo é chamado de autoregressivo de ordem p , ou *AR*(p), e quando $p = 0$ ele é chamado de modelo de média móvel de ordem q , *MA*(q) (SHUMWAY E STOFFER, 2010).

2.4.3.4 Modelo ARMA Integrado – ARIMA

Em muitas situações, uma série temporal pode ser pensada como sendo composta de dois componentes, um componente não estacionário de “tendência” e um componente estacionário de média zero (SHUMWAY E STOFFER, 2010).

Por exemplo, podemos considerar o modelo:

$$x_t = \mu_t + y_t \quad (2.7)$$

Onde μ_t é uma função linear $\mu_t = \beta_0 + \beta_1 t$ com $\beta_1 \neq 0$, e y_t é uma série estacionária de média zero (como por exemplo o ruído branco). A série x proveniente deste modelo é não estacionária pois tem sua média igual a μ_t que muda conforme t aumenta.

Diferenciando este modelo leva a um modelo estacionário:

$$\nabla x_t = x_t - x_{t-1} = \beta_1 + y_t - y_{t-1} = \beta_1 + \nabla y_t \quad (2.8)$$

com média igual a β_1 (SHUMWAY E STOFFER, 2010).

Se μ_t na equação 3.4 fosse um polinômio de ordem k , $\mu_t = \sum_{j=0}^k \beta_j t^j$, então a série diferenciada $\nabla^k y_t$ é estacionária. O modelo ARMA integrado, ou ARIMA, é uma ampliação da classe de modelos ARMA para incluir a diferenciação (SHUMWAY E STOFFER, 2010).

Uma série temporal x_t é *ARIMA*(p, d, q) se

$$\nabla^d x_t = (1 - B)^d x_t \quad (2.9)$$

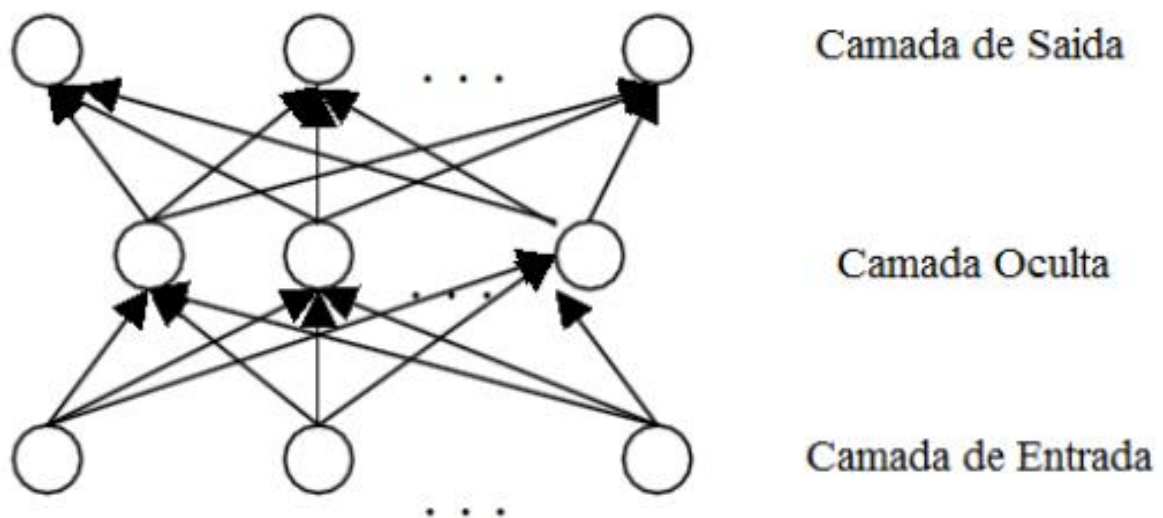
2.4.4 Redes Neurais

As redes neurais são um paradigma de computação inspirado no funcionamento do neurônio e do cérebro humano. Segundo (PROCHNOW, 2013), a qualidade dos resultados de soluções de redes neurais para o problema de previsão de séries temporais têm no mínimo se assemelhado a soluções da estatística clássica, como a aplicação de modelos ARIMA, e muitas vezes as redes neurais têm obtido resultados ainda melhores que estes métodos clássicos.

O modelo mais usado, o Perceptron Multicamadas, possui uma estrutura em três camadas distintas e interconectadas de neurônios artificiais, conforme pode ser visto na figura 2.7. A primeira camada é responsável por receber as entradas e propagá-las para a segunda camada. Cada neurônio da segunda camada realiza uma soma ponderada de todas as entradas e propaga um valor de saída para os neurônios da terceira camada, responsável por sintetizar uma saída através de outra soma ponderada (PROCHNOW, 2013).

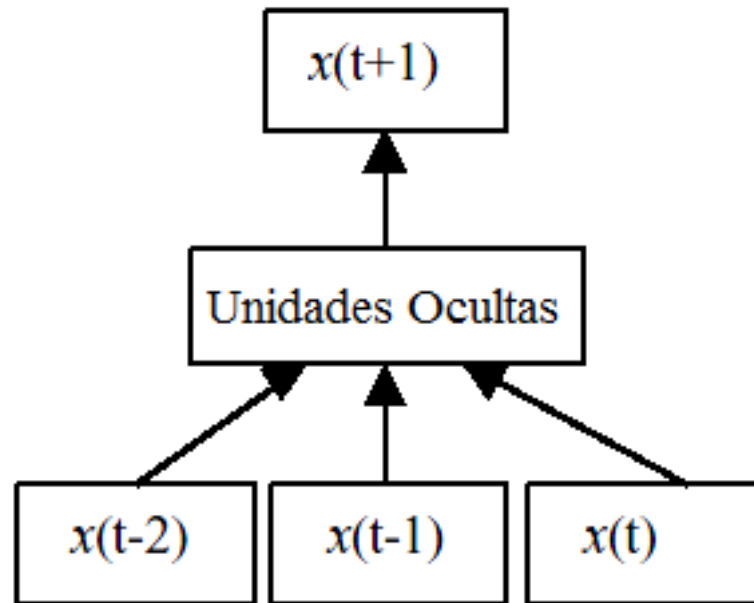
Para o uso na previsão de séries temporais, a rede neural recebe como entrada n medições x_{t-n}, \dots, x_{t-1} da série temporal analisada e repassa para os neurônios da camada oculta, que por sua vez passam para última camada, que gera o valor para x_t (PROCHNOW, 2013). Este processo está ilustrado na figura 2.8.

Figura 2.7 - Rede neural típica.



Fonte: Prochnow (2013, p. 33)

Figura 2.8 - Fluxo para geração da previsão do próximo valor de uma série.



Fonte: Prochnow (2013, p. 33)

Para determinar os pesos das somas ponderadas, a Rede Neural é calibrada em um processo de treinamento que faz uso de valores prévios da série temporal. Neste processo, os pesos são ajustados de forma a gerar o mínimo de erro possível para a previsão de cada valor do conjunto de treinamento. Este processo de utilizar as n medições anteriores para realizar uma previsão é chamado de **Técnica da Janela Deslizante**, e o tamanho de n influencia diretamente na qualidade das previsões feitas pela rede (PROCHNOW, 2013).

3 SISTEMAS IMUNOLÓGICOS ARTIFICIAIS

Segundo (CASTRO E TIMMIS, 2002), "Sistemas imunológicos artificiais são sistemas adaptativos inspirados pela imunologia teórica e pelas observadas funções, princípios e modelos imunológicos, e que são aplicados à solução de problemas."

O paradigma de computação de Sistemas Imunológicos Artificiais (SIA) é uma estratégia inspirada em modelos biológicos tal como Computação Evolutiva e Redes Neurais Artificiais. SIA têm como inspiração o modelo de funcionamento do sistema imunológico dos mamíferos.

Em (ASTIAZARA, 2012), o autor define uma gama de aplicações possíveis para SIA:

- Reconhecimento de Padrões.
- Detecção de Anomalias e Falhas.
- Análise de Dados (mineração, classificação, etc.).
- Sistemas baseados em Agentes.
- Escalonamento.
- Navegação e controle de Autônomos.
- Robótica.
- Métodos de busca e otimização.
- Segurança em sistemas de informação.

Dentre os algoritmos baseados em SIA, pode-se citar Redes Imunológicas, *Danger*, e Seleção Clonal. Neste trabalho o enfoque será dado a este último.

3.1 O Sistema Imunológico Natural

O Sistema Imunológico (SI) é um mecanismo complexo que envolve diferentes Órgãos e tipos de Células especializadas que cumprem funções distintas no processo de Identificação e Eliminação de Antígenos do corpo dos mamíferos.

Nesta seção será feita uma breve introdução a conceitos da Imunologia Teórica, e não inclui todos os processos que compõem este complexo sistema, somente as partes que são necessárias para compreender o funcionamento do Algoritmo de Seleção Clonal utilizado neste trabalho.

3.1.1 Linfócitos

As células brancas do sangue, também chamadas de Linfócitos, são um elemento importante para o reconhecimento de corpos estranhos, ou Patógenos, e são classificadas em diferentes tipos, dentre os principais deles as Células T e B. Estas células tem como principal funcionalidade a identificação de Patógenos, bem como manter uma "memória" de Patógenos adquirida (ASTIAZARA, 2012).

As células B são responsáveis por reconhecer antígenos particulares, e são uma parte importante do SI. Elas são produzidas medula óssea e são responsáveis por produzir e secretar

anticorpos no formato da letra Y que são capazes de se ligar aos antígenos (ASTIAZARA, 2012).

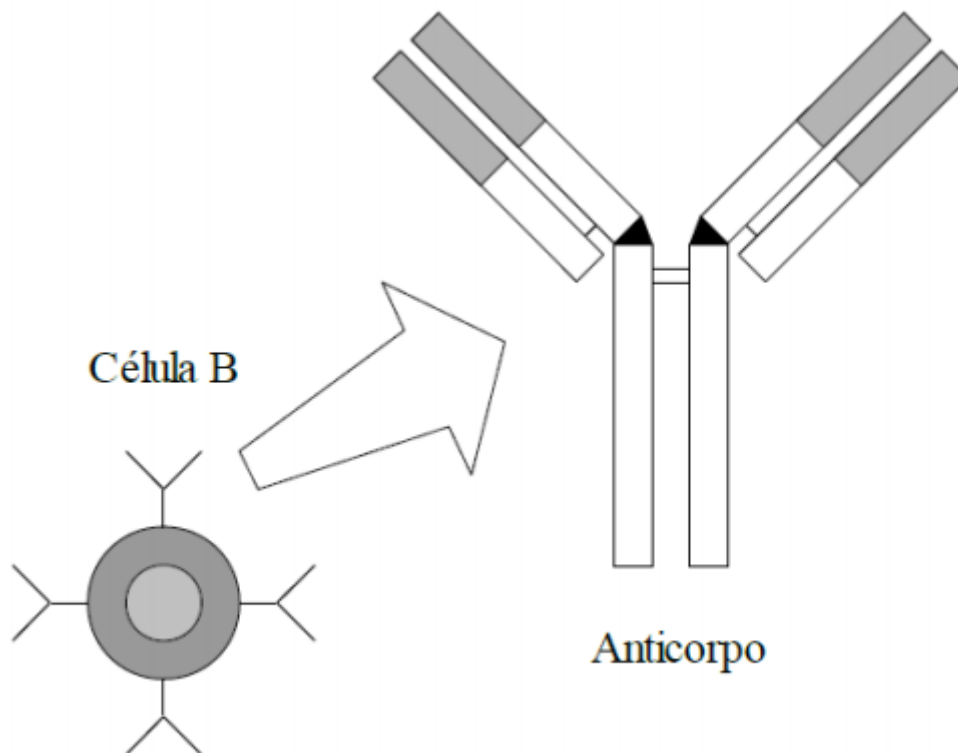
Cada célula B secreta múltiplas cópias de um tipo de anticorpo, e quando são ativadas (se ligam a determinado antígeno) elas se multiplicam e se diferenciam em Plasmócitos ou Células de Memória. Os Plasmócitos bombardeiam moléculas de anticorpos no Sangue. Cada Célula B produz somente um tipo de Anticorpo. (ASTIAZARA, 2012).

As células B são componentes importantes da Camada de Imunidade Adaptativa, também chamada de imunidade específica e imunidade adquirida, que é responsável por reconhecer e seletivamente eliminar micro-organismos e moléculas estranhas do corpo (ASTIAZARA, 2012).

3.1.2 Anticorpos

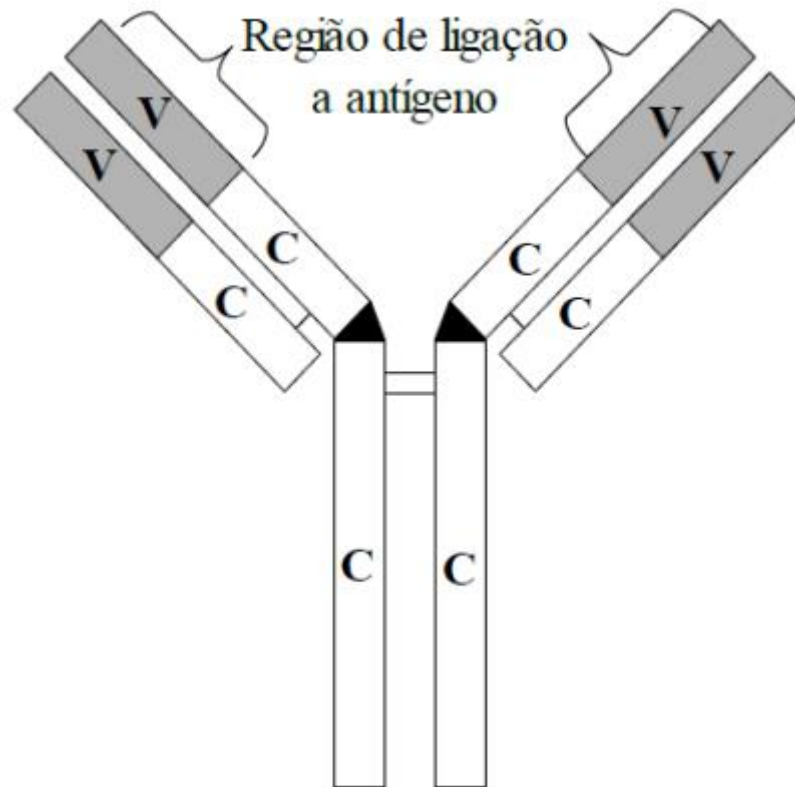
Os anticorpos são moléculas presentes na parede celular de Células B não maduras e Células B de Memória, e também são secretadas diretamente no sangue por Plasmócitos. Os Anticorpos são compostos de quatro cadeias polipeptídicas, cada cadeia possui uma região variável e uma região constante. Um peptídeo é uma cadeia de aminoácidos (ASTIAZARA, 2012).

Figura 3.1 – Célula B e Anticorpo.



Fonte: Astiazara (2012, p. 17).

Figura 3.2 – Detalhe do Anticorpo.



Fonte: Astiazara (2012, p. 17).

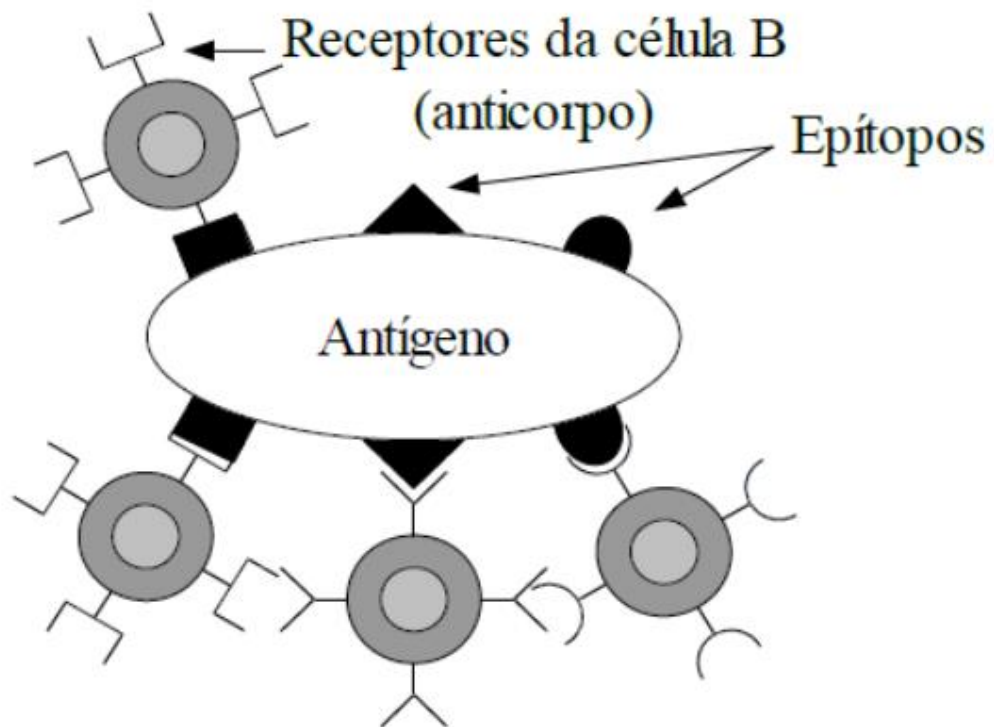
3.1.3 Reconhecimento Imunológico

Diversos processos imunológicos dependem do reconhecimento de corpos estranhos por um elemento do sistema imunológico. Um tipo de reconhecimento é realizado pelas Células B, que acontece quando ocorre uma ligação molecular entre o antígeno e os receptores na superfície da Célula B, os anticorpos (ASTIAZARA, 2012).

Somente pequenas partes dos Antígenos chamadas Epípetos ligam-se aos receptores das Células B. Múltiplas Células B diferentes podem se ligar a um mesmo Antígeno com diferentes níveis de afinidade. Esta afinidade depende da complementaridade entre as estruturas (Anticorpo e Epítipo) (ASTIAZARA, 2012).

A ligação não precisa ser perfeita para disparar o processo de reconhecimento. Uma ligação perfeita seria muito difícil de ocorrer. Devido à complexidade das estruturas, seria muito difícil evoluir Anticorpos que se adaptem perfeitamente a um determinado Epítipo. Se uma ligação perfeita fosse necessária, seria muito baixa a probabilidade de duas estruturas aleatórias ligarem-se. Uma consequência disto é que uma mesma Célula B pode reconhecer um conjunto de Antígenos diferentes, o que torna o SI eficiente em termos de tempo e memória (número de Células B diferentes necessárias) (ASTIAZARA, 2012).

Figura 3.3 – Reconhecimento do Antígeno pelas Células B.



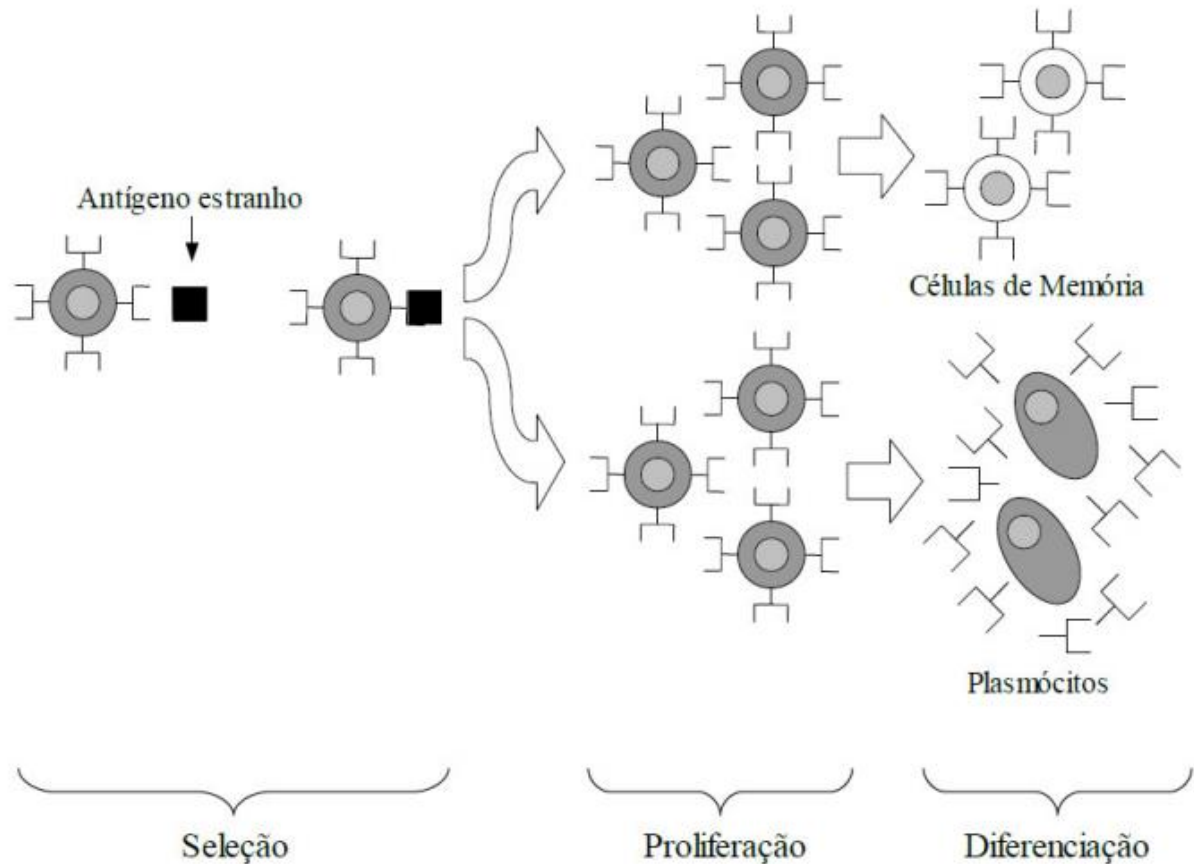
Fonte: Astiazara (2012, p.21).

3.1.4 Proliferação de Células B e Maturação de Afinidade

Quando uma Célula B se liga a um Antígeno com certo grau de afinidade, esta Célula é estimulada a passar por um processo de proliferação e diferenciação. Este processo é chamado de Seleção Clonal porque a ligação leva à Célula B a fazer uma multiplicação clonal. Parte das Células B geradas se tornam Células B de Memória, e asseguram que outra infecção pelo mesmo patógeno sejam identificadas rapidamente. As demais Células B se diferenciam em Plasmócitos que secretam grandes quantidades de um Anticorpo específico no sangue (ASTIAZARA, 2012). A figura 3.4 ilustra este processo.

Quando ocorre uma infecção por um novo Antígeno no organismo, a afinidade entre os Anticorpos e os Antígenos pode ser baixa, mas no processo de Proliferação, os clones da Célula B ativada sofrem uma mutação, chamado de Hipermutação Somática. Estas células por sua vez fazem o mesmo processo (Proliferação e Hipermutação) diversas vezes, e depois deste aperfeiçoamento elas amadurecem em Plasmócitos que produzem Anticorpos de alta afinidade com o Antígeno (ASTIAZARA, 2012).

Figura 3.4 – Proliferação e Diferenciação das Células B.



Fonte: Adaptado de Astiazara (2012, p. 26).

3.2 Algoritmo de Seleção Clonal

O Algoritmo de Seleção Clonal é baseado no Mecanismo de Proliferação e Maturação de Células B e no processo de Seleção Clonal destas células. (ASTIAZARA, 2012).

Para fins de simplificação, a palavra Anticorpo é utilizada para identificar um indivíduo da população, porém, o comportamento do indivíduo se assemelha mais com o comportamento de uma Célula B.

O algoritmo de Seleção Clonal tem como funcionamento básico (ASTIAZARA, 2012):

1. Gerar uma população inicial aleatória de indivíduos;
2. Apresentação antigênica: para cada padrão antigênico fazer:
 - a. Avaliação de afinidade: apresentá-lo à população e determinar a afinidade com cada elemento da população;
 - b. Seleção e expansão clonal: selecionar n elementos de maior afinidade na população e gerar clones destes indivíduos proporcionalmente à sua afinidade com o antígeno. Quanto maior a afinidade, maior o número de cópias;
 - c. Maturação da afinidade: realizar mutação em todas essas cópias com uma taxa inversamente proporcional à sua afinidade. Quanto maior a afinidade, menor a mutação. Adicionar estes indivíduos à população P e selecionar o

melhor indivíduo para ser mantido como a memória m do antígeno apresentado;

d. Metadinâmica: substituir um número p de indivíduos com baixa afinidade por novos gerados aleatoriamente;

3. Ciclo: repetir o passo 2 até que algum critério de parada seja atingido.

Os mecanismos de Seleção podem ser os mesmos utilizados por outros algoritmos de Computação Evolutiva, e os de mutação dependem do tipo de representação de anticorpos utilizado.

3.2.1 Representação dos Indivíduos

Os indivíduos da população (também chamados de Anticorpos neste texto) são os elementos que vão evoluir para adaptar ao problema, e necessitam de uma representação que defina suas características com o fim de determinar a afinidade do Anticorpo com um determinado Antígeno (Y. Q. SHI et al, 2010).

O conjunto de características da superfície da molécula que se liga ao antígeno é representado por uma cadeia de atributos que podem ser de qualquer tipo, valores inteiros, reais, símbolos de um alfabeto ou palavras de um dicionário, e normalmente o tipo de valor utilizado depende do domínio do problema que se deseja resolver (ASTIAZARA, 2012).

3.2.2 Medidas de Afinidade

A Medida de Afinidade é um valor real não negativo que representa o grau de combinação entre as moléculas. (ASTIAZARA, 2012).

Existem muitas maneiras de calcular a Afinidade entre as moléculas de Anticorpo e Antígeno, nesta subseção são apresentadas algumas, assumindo que uma molécula de Anticorpo fica definida por uma cadeia de atributos, $Atc = \langle Atc_1, Atc_2, \dots, Atc_{Latc} \rangle$, e uma molécula de Antígeno é dada por $Atg = \langle Atg_1, Atg_2, \dots, Atg_{Latg} \rangle$.

3.2.2.1 Similaridade como Medida de Afinidade

Uma maneira simples de medir a Afinidade entre o Anticorpo e o Antígeno é pelo grau de similaridade entre os atributos das duas cadeias. Neste caso, o objetivo é procurar Anticorpos com formatos mais similares ao invés de formatos complementares.

Uma maneira muito simples de realizar esta medida por Similaridade é utilizar a distância euclidiana:

$$Aff(Atc, Atg) = \sqrt{\sum_{i=1}^L (Atc_i - Atg_i)^2} \quad (3.1)$$

Claro que nem sempre é possível aplicar a distância euclidiana, como por exemplo no caso da molécula ter representação simbólica. Neste caso, é possível utilizar a distância de Hamming (ASTIAZARA, 2012):

$$Aff(Atc, Atg) = \sum_{i=1}^L \delta_i, \text{ onde } \delta_i = \begin{cases} 1, & \text{se } Atc_i \neq Atg_i \\ 0, & \text{se não} \end{cases} \quad (3.2)$$

3.2.2.2 Complementariedade como Medida de Afinidade

Neste caso, conforme maior a complementariedade das moléculas, maior a afinidade. Um exemplo de medida por complementariedade seria o inverso da Distância Euclidiana ou da de Hamming. Outro exemplo, no caso das moléculas serem representadas por cadeias de *bits*, uma medida de Afinidade por complemento seria utilizar a operação XOR e contar os *bits* 1 resultantes, conforme exemplificado na figura 3.5 (ASTIAZARA, 2012):

3.5 – Utilizando XOR para medida de Afinidade.

Molécula 1	0	0	1	1	0	0	1	1
Molécula 2	1	1	1	0	1	1	0	1
XOR:	1	1	0	1	1	1	1	0

Afinidade: 6

Fonte: Astiazara (2012, p. 37)

3.2.2.3 Variações de Alinhamento das Moléculas

Para alguns casos, em especial no caso em que as moléculas de Anticorpo e Antígeno são de tamanhos diferentes, pode-se variar a posição dos atributos das moléculas, comparando-as de maneiras diferentes. (ASTIAZARA, 2012). Um exemplo de medida de Afinidade deste tipo poderia ser a média da quantidade de *bits* 1 resultantes de uma operação XOR permutando todas as posições possíveis para a segunda molécula, conforme exemplificado na figura 3.6:

Figura 3.6 – Exemplo de medida de Afinidade com variação de alinhamento.

<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border: none;">Molécula 1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td></tr> <tr><td style="border: none;">Molécula 2</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: none;"></td><td style="border: none;"></td></tr> <tr><td style="border: none;">XOR</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: none;"></td><td style="border: none;"></td></tr> </table>	Molécula 1	0	0	1	0	1	1	Molécula 2	1	1	0	0			XOR	1	1	1	0			<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border: none;">Molécula 1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td></tr> <tr><td style="border: none;">Molécula 2</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: none;"></td><td style="border: none;"></td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td></tr> <tr><td style="border: none;">XOR</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: none;"></td><td style="border: none;"></td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td></tr> </table>	Molécula 1	0	0	1	0	1	1	Molécula 2	0			1	1	0	XOR	0			1	0	1
Molécula 1	0	0	1	0	1	1																																					
Molécula 2	1	1	0	0																																							
XOR	1	1	1	0																																							
Molécula 1	0	0	1	0	1	1																																					
Molécula 2	0			1	1	0																																					
XOR	0			1	0	1																																					
<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border: none;">Molécula 1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td></tr> <tr><td style="border: none;">Molécula 2</td><td style="border: none;"></td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: none;"></td></tr> <tr><td style="border: none;">XOR</td><td style="border: none;"></td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: none;"></td></tr> </table>	Molécula 1	0	0	1	0	1	1	Molécula 2		1	1	0	0		XOR		1	0	0	0		<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border: none;">Molécula 1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td></tr> <tr><td style="border: none;">Molécula 2</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: none;"></td><td style="border: none;"></td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td></tr> <tr><td style="border: none;">XOR</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: none;"></td><td style="border: none;"></td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td></tr> </table>	Molécula 1	0	0	1	0	1	1	Molécula 2	0	0			1	1	XOR	0	0			0	0
Molécula 1	0	0	1	0	1	1																																					
Molécula 2		1	1	0	0																																						
XOR		1	0	0	0																																						
Molécula 1	0	0	1	0	1	1																																					
Molécula 2	0	0			1	1																																					
XOR	0	0			0	0																																					
<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border: none;">Molécula 1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td></tr> <tr><td style="border: none;">Molécula 2</td><td style="border: none;"></td><td style="border: none;"></td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td></tr> <tr><td style="border: none;">XOR</td><td style="border: none;"></td><td style="border: none;"></td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td></tr> </table>	Molécula 1	0	0	1	0	1	1	Molécula 2			1	1	0	0	XOR			0	1	1	1	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border: none;">Molécula 1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">1</td></tr> <tr><td style="border: none;">Molécula 2</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: none;"></td><td style="border: none;"></td><td style="border: 1px solid black; padding: 2px;">1</td></tr> <tr><td style="border: none;">XOR</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: 1px solid black; padding: 2px;">0</td><td style="border: 1px solid black; padding: 2px;">1</td><td style="border: none;"></td><td style="border: none;"></td><td style="border: 1px solid black; padding: 2px;">0</td></tr> </table>	Molécula 1	0	0	1	0	1	1	Molécula 2	1	0	0			1	XOR	1	0	1			0
Molécula 1	0	0	1	0	1	1																																					
Molécula 2			1	1	0	0																																					
XOR			0	1	1	1																																					
Molécula 1	0	0	1	0	1	1																																					
Molécula 2	1	0	0			1																																					
XOR	1	0	1			0																																					

$$\text{Afinidade} = (3 + 1 + 3 + 2 + 0 + 2) / 6 = 1,8333$$

3.2.2.4 Outras Medidas de Afinidade

A medida de Afinidade é utilizada para expressar reconhecimento. É possível ver a medida de Afinidade que expressa a qualidade do Indivíduo em relação ao seu ambiente. Por exemplo, em um SIA aplicado para a otimização de uma função, o indivíduo do Sistema representa os parâmetros de entrada para a função, que resulta em um ponto da curva da função. Como o objetivo é maximizar o valor desta função, a medida de Afinidade de um Anticorpo poderia ser simplesmente o valor resultante da função para os parâmetros daquele Anticorpo, e essa medida seria plausível com a proposta de medir a qualidade do Indivíduo. (ASTIAZARA, 2012).

O conceito de Afinidade é parecido com o conceito de *Fitness*. Nos Algoritmos Evolutivos, o *Fitness* é uma medida que representa o quão adaptado um indivíduo está em relação ao seu ambiente. A diferença entre os conceitos de *Fitness* e Afinidade é que o *Fitness* representa o ambiente, e nos SIA isso nem sempre é verdade, já que muitas vezes é preciso medir a afinidade entre os próprios elementos do SIA, além da interação com o ambiente (ASTIAZARA, 2012).

3.2.3 Mecanismos de Seleção

O mecanismo de Seleção Imunológica serve para modificar a diversidade da população de modo a encontrar a solução global de otimização de um problema. Devido ao fato de que a

Seleção Imunológica precede o processo de evolução imunológica, esta etapa tem um forte efeito neste processo (Y. Q. SHI et al, 2010).

Diversos mecanismos de Seleção podem ser utilizados para selecionar células para clonagem e sobrevivência, porém, em SIA, todos devem depender de alguma maneira da Afinidade do Anticorpo com o Antígeno. Uma maneira de implementar a seleção de maneira probabilística seria utilizando o Método da Roleta. Outros Mecanismos de Seleção de Computação Evolutiva podem ser utilizados, como por exemplo, Seleção Bipolar, Seleção Elitista e por Ranking (ASTIAZARA, 2012).

No método da Roleta, os indivíduos são selecionados probabilisticamente com chance dependendo tipicamente da afinidade dos indivíduos. O método do Ranking é similar, porém a chance depende da posição do indivíduo em um Ranking, também baseado na Afinidade.

A seleção Bipolar pega os melhores e os piores indivíduos em relação à Afinidade, enquanto a seleção Elitista pega sempre os melhores indivíduos.

3.2.4 Mecanismos de Mutação

A operação de Mutação de Anticorpos é a mais crítica do processo de Evolução Imunológica. É a mutação que cria a diversidade da população de Anticorpos e permite a variação com o fim de se explorar o espaço de busca e se aproximar do Modelo Ótimo Global (Y. Q. SHI et al, 2010).

Na mutação também, diversas maneiras podem ser utilizadas para introduzir variabilidade nos clones, dependendo é claro, da representação escolhida para as moléculas. Em uma representação onde os atributos são bits, um ou mais atributos da molécula podem ser escolhidos aleatoriamente para trocarem de valor. Em diversas representações, como a simbólica e por números reais, a mutação pode ser realizada escolhendo duas posições de atributos e trocando o valor delas entre si (ASTIAZARA, 2012).

O importante a ressaltar aqui é que as mutações são sempre realizadas probabilisticamente com base em uma Taxa de Mutação, que em SIA é tipicamente variável. A Taxa de Mutação é sempre inversamente relacionada com a Afinidade do indivíduo (ASTIAZARA, 2012).

4 APLICANDO SISTEMAS IMUNOLÓGICOS ARTIFICIAIS PARA O AJUSTE E PREVISÃO DE SÉRIES TEMPORAIS

Neste capítulo são feitas algumas considerações necessárias para adaptar o Algoritmo de Seleção Clonal com a finalidade de Prever Valores Futuros de uma Série Temporal. As observações deste capítulo são genéricas e podem ser aplicadas em qualquer implementação de SIA com fins de prever Séries Temporais.

4.1 Modelo

Um tipo muito simples de modelo seria uma função do valor do instante atual de tempo da série:

$$x_t = F(t) \quad (4.1)$$

onde x_t é o valor da série no instante de tempo t .

Os modelos podem ser ainda uma função dos últimos q valores da série:

$$x_t = F(x_{t-1}, x_{t-2}, \dots, x_{t-q}) \quad (4.2)$$

O modelo poderia ainda incorporar valores de uma ou mais Séries Preditoras ou Explanatórias. Por exemplo, um modelo para prever os valores de uma série x poderia ser do tipo:

$$x_t = F(x_{t-1}, \dots, x_{t-n}, y_{t-1}, \dots, y_{t-n}, z_{t-1}, \dots, z_{t-n}) \quad (4.3)$$

onde y e z são Séries Temporais.

Em (Y. Q. SHI et al, 2010) e (LUO E SHI, 2011), os autores propõem um modelo em que se utiliza um método chamado de Reconstrução de Espaço de Fase (*Phase Space Reconstruction*) para encontrar uma medida chamada Dimensão de Inserção d (*Embedding Dimension*) e um atraso de tempo τ (*Time Delay*), de maneira que o modelo é uma função dos d valores anteriores da Série, com deslocamento τ entre esses valores, ou seja:

$$x_t = F(x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-d\tau}) \quad (4.4)$$

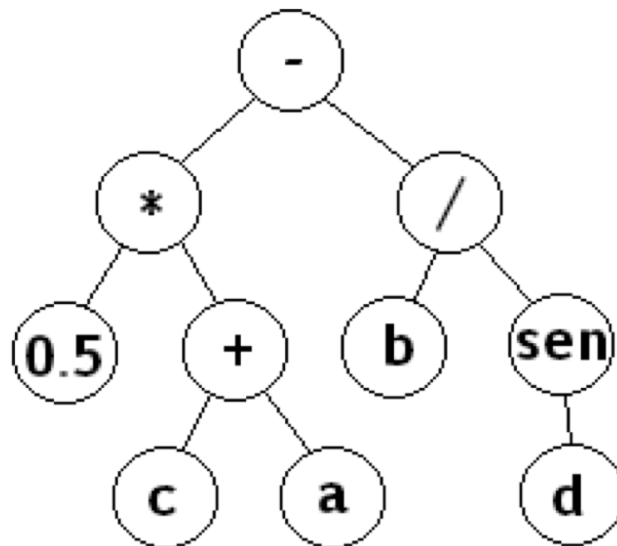
A ideia por trás deste modelo é encontrar alguma estrutura cíclica da Série Temporal e criar um modelo que se aproveita dessa informação. Um exemplo é uma Série Temporal que mede valores mensalmente e que apresenta uma natureza cíclica em relação aos meses do ano. Um modelo para medir esta série teria como atraso de tempo $\tau = 12$ e os valores de entrada do modelo para prever o valor do próximo mês Julho, por exemplo, seriam os valores dos últimos d meses de Julho anteriores.

4.2 Representações das Moléculas de Anticorpo e Antígeno

Em um SIA para Previsão de Séries Temporais, os indivíduos da população (Células ou Anticorpos) representam modelos matemáticos distintos que descrevem uma série. Muitas maneiras poderiam ser utilizadas para representar este modelo, uma forma mais intuitiva seria a representação em forma de árvore.

Desta forma, uma equação matemática qualquer pode ser vista como forma de uma árvore em que os nodos seriam operações matemáticas (adição, subtração, seno, cosseno, raiz quadrada, exponenciação) ou variáveis do modelo ou constantes. Os nodos folha dessa árvore devem ser necessariamente variáveis ou constantes. A figura 4.1 mostra um exemplo de uma equação modelada em forma de árvore.

Figura 5.1 – Exemplo de uma representação de uma equação em forma de Árvore.



A árvore da figura 4.1 pode ser traduzida na equação:

$$(0.5 * (c + a)) - (b / \text{sen}(d)) \quad (4.5)$$

O antígeno é o elemento contra qual queremos adaptar os indivíduos, então ele representa a Série Temporal que queremos encontrar um modelo. O antígeno então é formado como uma cadeia de valores Reais, em que cada valor corresponde a uma medida da Série.

4.3 Medidas de Afinidade

Na aplicação de Previsão de Séries Temporais, a afinidade entre o Anticorpo e o Antígeno sempre será uma medida de adequação do modelo em relação à série. Esta medida pode ser feita de várias maneiras, como por exemplo utilizando o inverso da média da Distância Euclidiana entre os valores previstos pelo Modelo e pela Série Temporal que se deseja prever.

Também é possível aplicar outros tipos de Afinidade, como afinidade dos Anticorpos entre si. Nesse caso, um exemplo de Medida de Afinidade seria o inverso da distância de Hamming entre os atributos dos Anticorpos.

4.4 Mecanismos de Seleção

Para previsão de séries temporais, todos os mecanismos de seleção da seção 3.2.3 se aplicam.

4.5 Mecanismos de Mutação

Para mutação, no caso da representação do modelo por árvore, pode-se utilizar vários tipos de mutação, como por exemplo:

- Mudança aleatória em um nodo da árvore: escolhe-se um nodo aleatoriamente e substitui o tipo do nodo por outro aleatoriamente.
- Troca de um nodo por outro: escolhem-se dois nodos aleatoriamente e trocam-se os tipos de nodo um pelo outro.
- Geração de sub-árvore aleatória: escolhe-se um nodo aleatoriamente e substitui o nodo por uma nova sub-árvore aleatória.
- Troca de sub-árvore: escolhem-se dois nodos aleatoriamente e troca-se de posição as sub-árvores tomando estes nodos como as raízes das sub-árvores.

Um SIA pode usar uma ou combinar duas ou mais técnicas de mutação ao mesmo tempo. A quantidade/intensidade dessas mutações é probabilística e depende de um valor chamado de Taxa de Mutação. Essa Taxa, no Algoritmo de Seleção Clonal, é variável, por se inspirar no mecanismo de Proliferação e Maturação das Células B (ver seção 3.1.4).

5 ESTUDO DE CASO

Neste capítulo é apresentada em detalhes uma solução de SIA utilizando o Algoritmo de Seleção Clonal para a Previsão de Séries Temporais. Esta solução foi baseada nos artigos de (Y. Q. SHI et al, 2010) e (LUO E SHI, 2011).

Estes artigos foram escolhidos por apresentarem um Método Baseado em Seleção Clonal para o Ajuste de Séries Caóticas, com o fim de Prever os Valores futuros desta Série. A implementação que é descrita nesse capítulo toma emprestado destes artigos o formato da representação do Anticorpo, bem como as medidas de Afinidade entre o Anticorpo e o Antígeno e a medida de Afinidade entre os Anticorpos, bem como o mecanismo de Mutaçao em duas estratégias (Local e Global).

Na primeira seção do capítulo são apresentados detalhes da implementação utilizada, e na segunda seção são apresentados resultados obtidos, bem como detalhes da calibração do algoritmo.

5.1 Detalhes da Implementação

5.1.1 Normalização da Série

Antes de iniciar o processo, a Série Temporal é normalizada utilizando a seguinte equação:

$$\bar{x}_t = \frac{x_t - x_{min}}{x_{max} - x_{min}} \quad (5.1)$$

Dessa maneira, todos os valores da série são deslocados para o intervalo de 0 a 1. Essa normalização coloca todos os valores da série em um intervalo fixo, o que permite que as constantes do modelo possam pertencer à esse mesmo intervalo.

5.1.2 Modelo

As soluções do método são modelos da série em função do tempo, do tipo $y_t = F(t)$, onde y_t representa o valor do modelo no tempo t , e a sua única variável representa o índice de tempo atual da Série Temporal.

5.1.3 Representação do Anticorpo

O anticorpo nesta implementação será a representação de uma árvore, que por sua vez representa um modelo matemático de uma função, que é utilizado como modelo da série temporal. A modelagem desse Anticorpo é dada como uma sequência de caracteres e números, dividida em 3 regiões: Hipervariável ou H_r , Framework ou F_r e Constantes ou C_r .

O anticorpo é modelado então, da seguinte maneira:

$$atc = \{ \langle H_r, F_r, C_r \rangle \mid H_r \in \{Opr_s, Folhas\}^{L_h} \mid F_r \in \{Folhas\}^{L_f} \mid C_r \in \mathbb{N}^{L_c} \} \quad (5.2)$$

A região H_r é composta de operações matemáticas e marcadores de variáveis e constantes. O tamanho L_h dessa região determina qual será a complexidade máxima do modelo do anticorpo determinando o número máximo de operações matemáticas dentro do modelo.

O conjunto Opr_s de operações é definido como $Opr_s = \{+, -, *, /, q, p, s, c\}$, onde $+, -, *, /$ representam a operação de soma, subtração, multiplicação e divisão, respectivamente, q representa a operação de raiz quadrada, p a de exponenciação, s a de seno e c a de cosseno.

O conjunto $Folhas$ é definido como $Folhas = \{v, ?\}$, sendo que v é um marcador de uma variável, e $?$ é o marcador de uma constante.

A região Framework é composta exclusivamente por marcadores para constante e variável. O tamanho desta região é determinado pela equação:

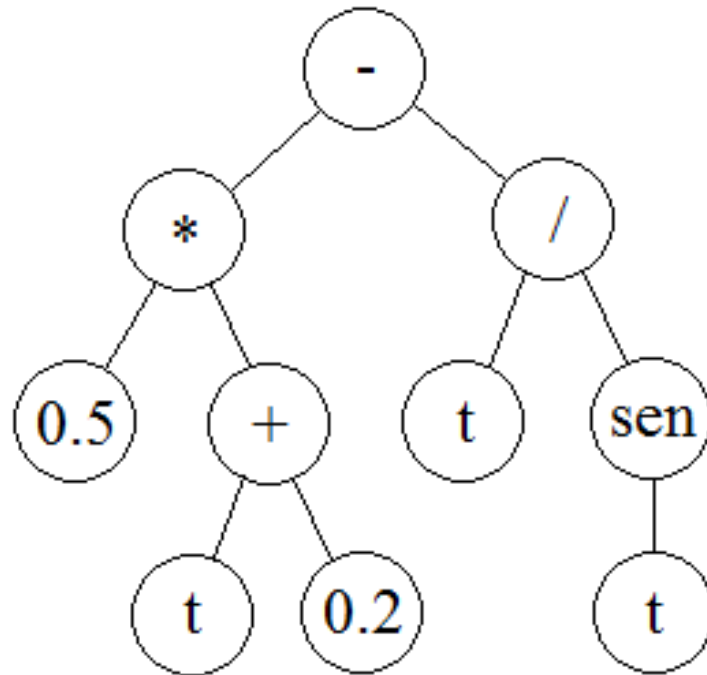
$$N_f = (m - 1) * N_{n_f} + 1 \quad (5.3)$$

Onde N_f é o número de nodos folha de uma árvore, m é o número máximo de filhos de um nodo não folha, e N_{n_f} é o número de nodos não folha de uma árvore. A prova desta equação pode ser encontrada em (Y. Q. SHI et al, 2010). Como o número máximo de nodos não folha da árvore é determinado por L_h e todas as operações são unárias ou binárias, o tamanho dessa região é $L_f = L_h + 1$.

A região de Constantes é composta por números de 0 a $n_c - 1$, onde n_c é o tamanho do Conjunto de Constantes. O Conjunto de Constantes consiste em todas as constantes utilizadas nos modelos, e para os testes feitos neste trabalho, o conjunto de constantes utilizados foi $\{0, 0.1, 0.2, 0.3, \dots, 0.9, 1\}$. Essas constantes foram escolhidas de dentro do intervalo $[0, 1]$ em que os valores normalizados da Série se encontram.

A expressão matemática que representa o modelo de previsão da Série Temporal é criada num formato de árvore, como exemplificado na figura 5.1.

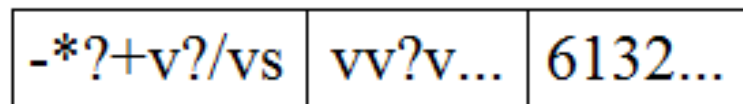
Figura 5.1 – Exemplo de uma árvore contendo um modelo.



A árvore da figura 5.1 seria traduzida na equação $(0.5 * (t + 0.2)) - (t/\text{sen}(t))$.

Esta árvore de exemplo seria montada a partir do anticorpo da figura 5.2 e do conjunto de constantes $\{0, 0.1, 0.2, \dots, 0.9, 1\}$.

Figura 5.2 – Exemplo de um Anticorpo.



A árvore é montada utilizando uma função recursiva. O trecho de pseudocódigo a seguir demonstra o funcionamento:

```

//estes índices são globais e são zerados antes de criar uma árvore para um
//anticorpo
indice_oprsvs = 0;
indice_const = 0;

function cria_nodo (Anticorpo atc, Double ConjConstantes[Nc])
{
    Nodo novo_nodo;

    //Determina qual é o tipo do nodo
    if (indice_oprsvs < Lh)
        novo_nodo.tipo = atc.Rh[indice_oprsvs];
    else

```

```

        novo_nodo.tipo = atc.Rf[indice_oprsvars - Lh];

    indice_oprsvars++;

    //Se o nodo é do tipo Constante, determina qual é a constante
    if (novo_nodo.tipo == '?')
    {
        novo_nodo.constante = ConjConstantes[atc[indice_const]];
        indice_const++;
        if (indice_const > Lc)
            indice_const = 0;
    }

    //Chama recursivamente a função dependendo do número de filhos:
    //operações binárias
    if (novo_nodo.tipo == '+' or '-' or '*' or '/' or 'p')
    {
        novo_nodo.filho1 = cria_nodo(atc, ConjConstantes);
        novo_nodo.filho2 = cria_nodo(atc, ConjConstantes);
    }

    //operações unárias
    If (novo_nodo.tipo == 'q' or 's' or 'c')
    {
        novo_nodo.filho1 = cria_nodo(atc, ConjConstantes);
    }
}

```

5.1.4 Medida de Afinidade

A Afinidade com o Antígeno é definida na equação 5.4:

$$Aff_{atg}(atc, atg) = \frac{1}{1 + \sqrt{\frac{1}{m} \sum_{j=1}^m (p_j(atc, j) - atg_j)^2}} \quad (5.4)$$

Onde m é o tamanho da Série, atg_j é o j -ésimo elemento do Antígeno (no modelo utilizado, corresponde ao valor da série no tempo j), e $p_j(atc, j)$ é o valor predito pelo modelo do anticorpo atc no tempo j . O valor da raiz quadrada corresponde à raiz quadrada do erro médio ao quadrado entre os valores preditos e reais da série. Quanto menor for esse valor, maior será a afinidade do anticorpo com a série.

A afinidade entre Anticorpos serve para medir a similaridade entre dois anticorpos membros da população, e é definida pelas equações:

$$Aff_{atc}(atc_s, atc_t) = \frac{\sum_{i=1}^{L_h+L_f+L_c} f_{comb}(atc_s(i), atc_t(i))}{L_h+L_f+L_c} \quad (5.5)$$

$$f_{comb}(atc_s(i), atc_t(i)) = \begin{cases} 0, & \text{se } atc_s(i) \neq atc_t(i) \\ 1, & \text{caso contrário} \end{cases} \quad (5.6)$$

Onde L_h, L_f, L_c são os tamanhos das regiões Hipervariáveis, de Framework, de Constantes, respectivamente. A afinidade entre anticorpos é determinada pelo inverso da distância de Hamming entre os atributos do Anticorpo, normalizada para um valor real de 0 a 1.

5.1.5 Mecanismo de Seleção para Proliferação

Baseado na Teoria de Redes Imunológicas, os autores de (Y. Q. SHI et al, 2010) e (LUO E SHI, 2011) propuseram um novo mecanismo de seleção que será adotado nesta implementação.

Como esse mecanismo utiliza ao mesmo tempo a Afinidade entre os Anticorpos e Antígenos e Afinidade dos Anticorpos entre si, este mecanismo propicia não só a seleção dos anticorpos com alta afinidade com o Antígeno bem como também os Anticorpos de baixa Afinidade com o fim de manter a diversidade da população.

Utilizando as equações 5.4 e 5.5, calculamos um valor que corresponde à expectativa de um Anticorpo ser selecionado baseado na sua afinidade com o Antígeno e sua similaridade com os demais Anticorpos da população.

Esta medida de similaridade de um Anticorpo com os demais é chamada de Valor de Estímulo e é calculada através das equações 5.7 e 5.8.

$$Acs_{sub}(atc_s) = \{atc | Aff_{atc}(atc_s, atc) \geq \alpha, \forall atc \in Acs - atc_s\} \quad (5.7)$$

$$Est(atc_s) = \frac{|Abs_{sub}(atc_s)|}{|Acs|} \quad (5.8)$$

Onde Acs é a população de Anticorpos, atc indica um Anticorpo, e α é um valor limite de similaridade definido como constante no algoritmo.

Utilizando a equação de Estímulo, os autores em (Y. Q. SHI et al, 2010) e (LUO E SHI, 2011) definem a Expectativa de Seleção de um Anticorpo $P_{sel}(atc_s)$ como sendo:

$$P_{sel}(atc_s) = \frac{Aff_{atg}(atc_s, atg)}{\sum_{i=1}^n Aff_{atg}(atc_i, atg)} * e^{-Est(atc_s)} \quad (5.9)$$

Onde n é o número de Anticorpos na população.

Sendo assim, o processo de Seleção Imunológica fica definido como sendo:

1. Utilizar a equação 5.4 para determinar a Afinidade com o Antígeno dos membros da população Acs ;
2. calcular o Valor de Estímulo de cada membro atc_s da população Acs :
 - a. utilizar a equação 5.7 para obter o subconjunto $Acs_{sub}(atc_s)$ da população Acs ;
 - b. utilizar a equação 5.8 para calcular o Valor de Estímulo $Est(atc_s)$ do anticorpo;
3. Utilizar a equação 5.9 para calcular a Expectativa de Seleção $P_{sel}(atc_s)$ de todos os anticorpos da população Acs .

Usando agora os valores de Expectativa de Seleção P_{sel} , a o mecanismo de seleção utilizado nesta implementação é elitista. Os anticorpos da população Acs são colocados em ordem descendente de acordo com o seu valor de Expectativa de Seleção, e os m Anticorpos com maior P_{sel} formam o subconjunto de Anticorpos Acs_{sel} que serão utilizados para a operação de Proliferação Clonal.

5.1.6 Proliferação Clonal

O conjunto Acs_{sel} é ordenado de forma descendente de acordo com seu Valor de Expectativa P_{sel} , e para cada membro atc_i onde i é um índice indo de 1 a m , onde o Anticorpo atc_1 é o de maior Expectativa de Seleção e atc_m é o Anticorpo de menor Expectativa de Seleção, o número de clones é:

$$N_{clones}(atc_i) = \lfloor (n - i)/i \rfloor \quad (5.10)$$

Onde n é o tamanho da população de Anticorpos. Nesta etapa, todos os clones são cópias idênticas de seus pais, e o conjunto de clones gerados nesse processo é chamado de Acs_{clones} .

5.1.7 Mecanismo de Mutação

De acordo com as características da modelagem dos Anticorpos, os autores de (Y. Q. SHI et al, 2010) apresenta duas Estratégias de Mutação, a chamada “Mutação Local” e a chamada “Mutação Global”.

A Mutação Local consiste da Operação de Mutação somente dos elementos da Região Hipervariável dos Anticorpos de baixa Afinidade com o Antígeno, e permite uma busca mais ampla no espaço de soluções possíveis.

Para os elementos de maior Afinidade com o Antígeno, a Mutação Global atua nas Regiões de Framework, Constantes e Variáveis, de modo a realizar pequenas mudanças no modelo do Anticorpo de maneira a se aproximar de maneira suave em direção à Solução Ótima global.

Para determinar se a Mutação de um clone será do tipo Local ou Global, é comparado a Afinidade com o Antígeno do clone com a média das Afinidades do subconjunto Acs_{sel} .

$$\beta = \frac{\sum_{i=1}^m Aff_{atg}(atc_i)}{m} \quad (5.11)$$

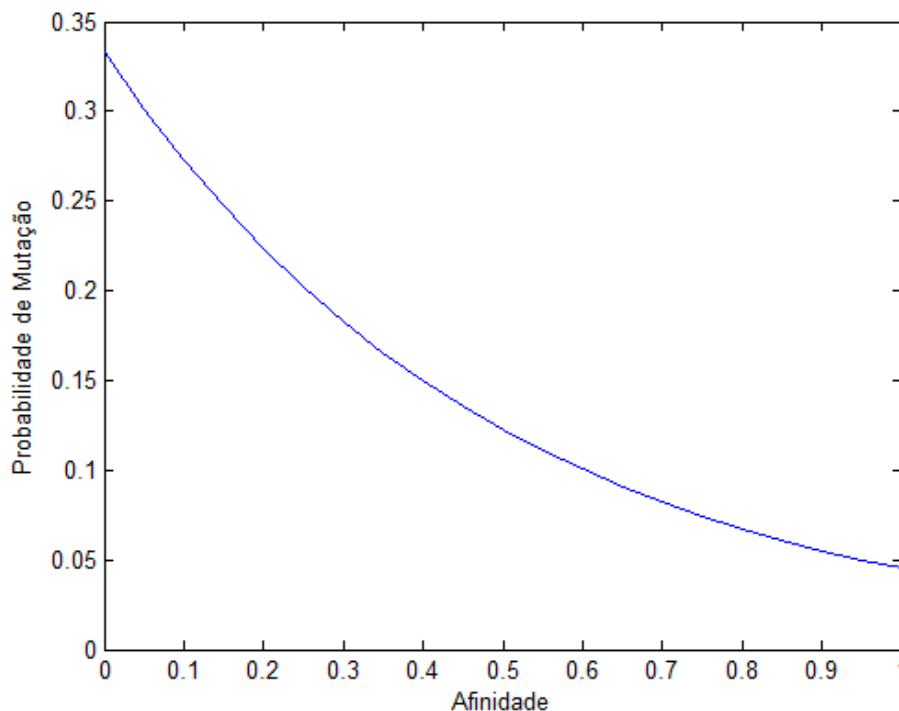
Onde m é o tamanho do subconjunto Acs_{sel} , e atc_i são membros deste subconjunto. Se a Afinidade do clone é menor ou igual a β , o clone sofre o processo de Mutação Local, caso contrário, o clone sofre o Processo de Mutação Global.

Tanto para a Mutação Local como a Global, o processo de mutação é o mesmo, o que muda são os elementos que sofrerão o processo de Mutação. Cada elemento (caractere da Região que sofre a mutação) é mutado com probabilidade P_m que é definida para cada Anticorpo com base na sua Afinidade com o Antígeno:

$$P_m(atc) = \frac{e^{-2*Aff_{atg}(atc)}}{3} \quad (5.12)$$

Se um elemento do Anticorpo é mutado, ele é substituído por outro elemento de maneira aleatória, de acordo com a Região do Elemento. Se o elemento pertence à Região Hipervariável, ele é substituído por outro elemento escolhido aleatoriamente do conjunto $Orps \cup Folhas$, se é da Região de Framework, do conjunto $Folhas$, se da Região de Constantes, do conjunto $\{0, \dots, N_{cons} - 1\}$ onde N_{cons} é o tamanho do Conjunto de Constantes.

Figura 5.3 - $P_m(atc)$ para diferentes níveis de Afinidade.



5.1.8 Seleção dos Sobreviventes

A nova população que irá constituir a próxima geração é construída da seguinte maneira:

1. O melhor anticorpo em termos de Afinidade com o Antígeno é selecionado para a nova população;
2. Os demais $n - 1$ elementos da nova população (sendo n o tamanho da população) são selecionados dentre os membros de Acs e Acs_{clones} através do método da roleta baseado na Afinidade com o Antígeno.

O Método da Roleta é utilizado para dar mais diversidade à nova população.

5.1.9 Resumo da Implementação

Assim, o processo todo de Evolução Imunológica através do Algoritmo de Seleção Clonal para Previsão de Séries Temporais fica definido como:

1. Obter e Normalizar uma Série Temporal;
2. gerar uma população Acs aleatória de Anticorpos inicial;
3. até que o critério de parada ocorra:
 - a. determinar a Afinidade com o Antígeno dos Anticorpos da população Acs utilizando a equação 5.4;
 - b. determinar a Afinidade entre os Anticorpos e o Valor de Estímulo utilizando as equações 5.5 e 5.8;
 - c. calcular a Expectativa de Seleção para os Anticorpos da População Acs utilizando a equação 5.9;
 - d. selecionar m Anticorpos da população para formar o conjunto Acs_{sel} de acordo com o processo descrito na seção 5.1.5;
 - e. realizar o processo de Proliferação Clonal sobre Acs_{sel} para gerar o conjunto de Clones Acs_{clones} conforme descrito na seção 5.1.6;
 - f. realizar a mutação sobre a população de Clones Acs_{clones} conforme descrito na seção 5.1.7;
 - g. Formar a nova população Acs conforme o processo descrito na seção 5.1.8;
4. selecionar o melhor Anticorpo da última população População gerada para criar o Modelo de Previsão da Série.

O Critério de Parada utilizado é um número fixo de gerações. O algoritmo pára depois de processar este número de gerações. Como o melhor indivíduo de uma geração sempre passa para a geração posterior, é seguro afirmar que o melhor indivíduo da última geração tem o melhor modelo gerado no processo todo.

5.1.10 Linguagens de Programação e Ferramentas

Este método foi implementado na Linguagem C utilizando o Microsoft Visual Studio Express 2010, e os gráficos foram gerados utilizando o MATLAB R2013a.

5.2 Experimentos

Para a realização dos experimentos, as Séries de *Random Walk With Drift* e IBOVESPA foram selecionadas para testar a implementação.

5.2.1 Ruído Branco e *Random Walk*

O Ruído Branco é um exemplo simples de série temporal gerada que consiste em uma coleção de variáveis aleatórias não correlacionadas, com média 0 e variância finita. Esta série temporal é muito utilizada como modelo de ruído para aplicações de engenharia. O Ruído Branco pode também ser gerado utilizando variáveis aleatórias distribuídas de forma normal. Nesse caso, a série gerada é chamada de Ruído Branco Gaussiano (SHUMWAY E STOFFER, 2010).

Nas figuras 5.4 e 5.5 encontram-se exemplos de Séries de Ruído Branco e Ruído Branco Gaussiano, respectivamente.

Figura 5.4 – Exemplo de Série de Ruído Branco.

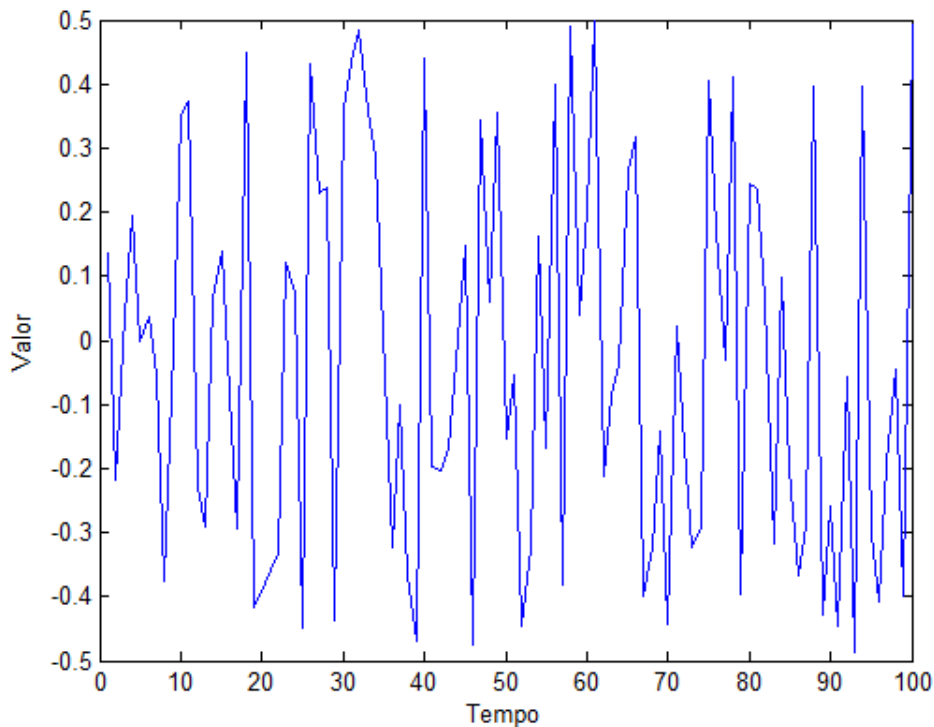
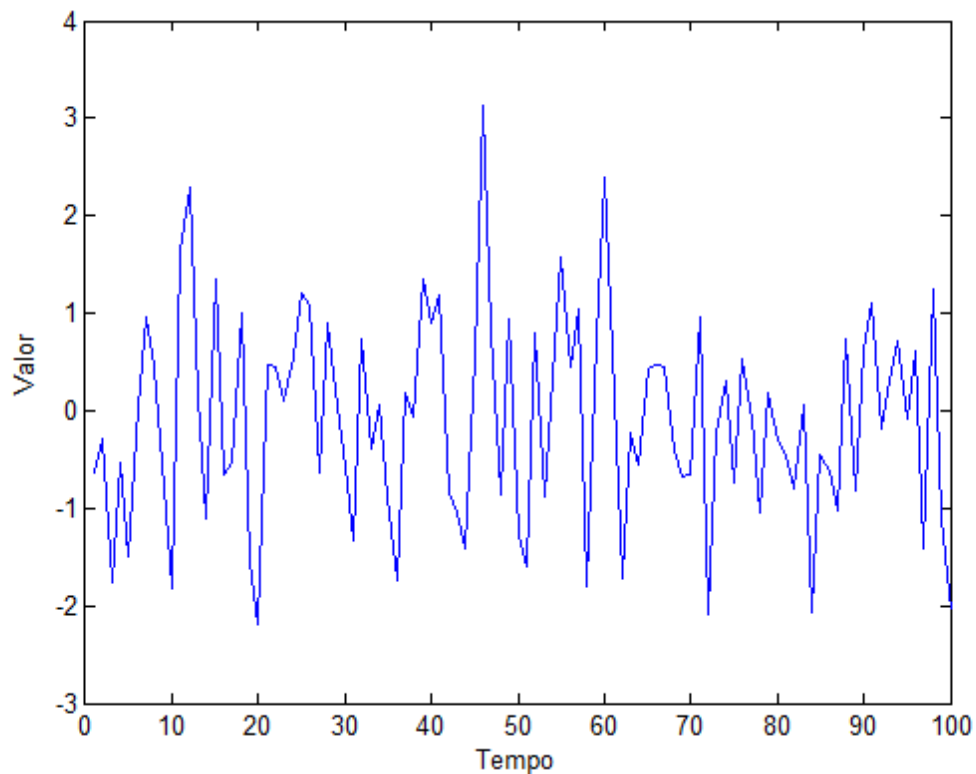


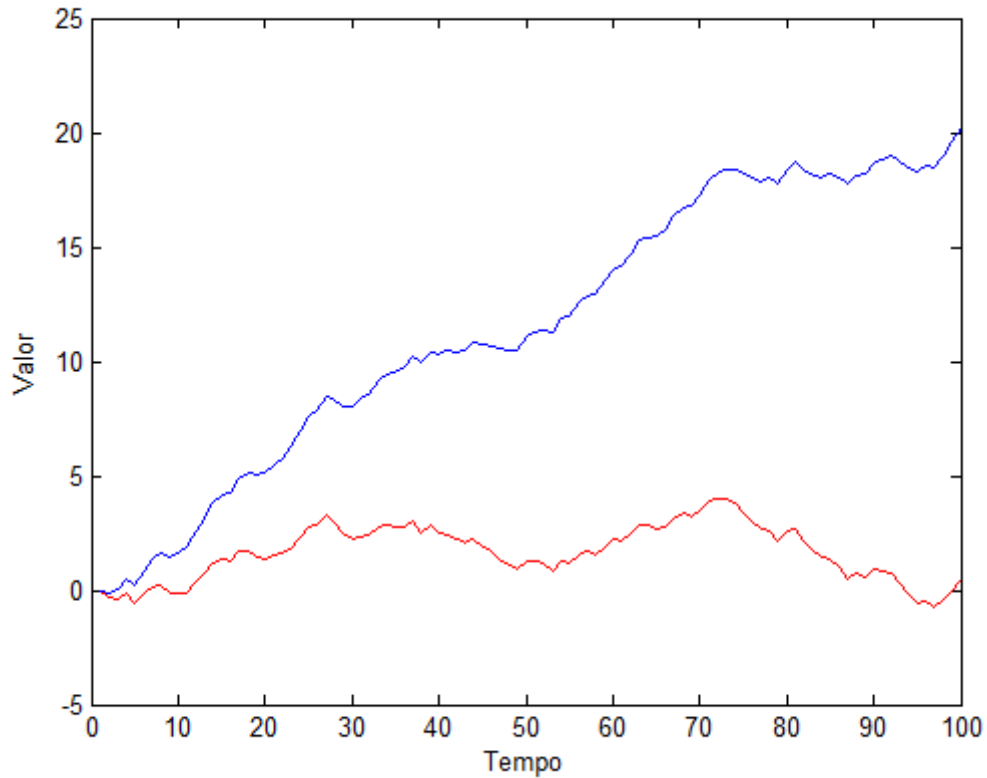
Figura 5.5 – Exemplo de Série de Ruído Branco Gaussiano.



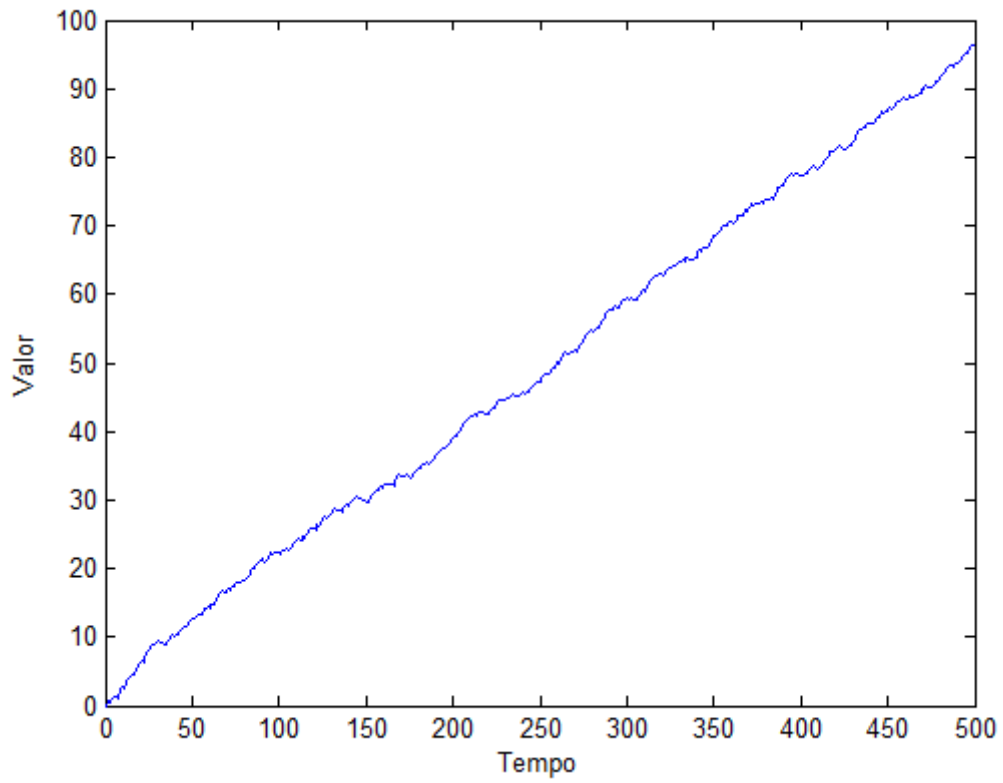
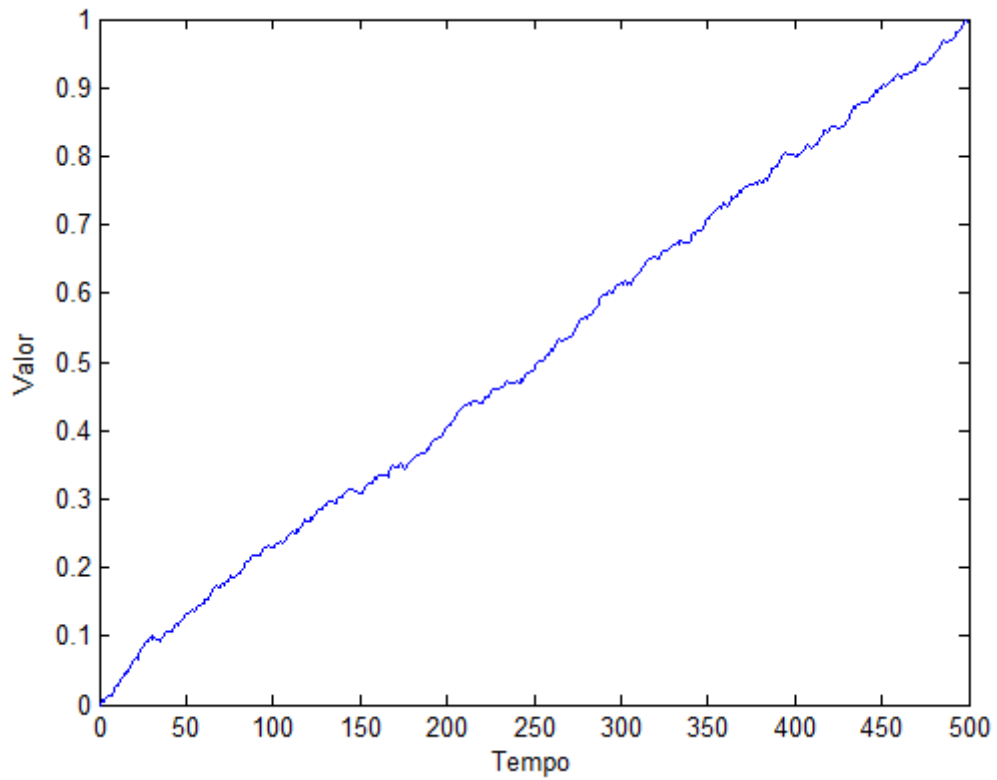
Outro exemplo de série gerada que pode ser muito útil para testar a eficácia de algoritmos de previsão de séries temporais é chamado de *Random Walk with Drift*, e é gerada usando o Ruído Branco. Cada valor x_t da série no tempo t é gerado com base na equação:

$$x_t = \delta + x_{t-1} + w_t \quad (5.13)$$

onde w_t é uma Série de Ruído Branco. A constante δ na equação é o valor chamado de *drift*, e no caso em que $\delta = 0$ a série é chamada simplesmente de *Random Walk* (SHUMWAY E STOFFER, 2010).

Figura 5.6 – Exemplo de *Random Walk* (vermelho) e *Random Walk with Drift* (azul).

A série de *Random Walk with Drift* utilizada para testar a eficácia do algoritmo foi gerada utilizando Ruído Branco com variância igual a 1, *drift* $\delta = 0.2$, valor inicial $x_1 = 0$ e com 500 medições. Na figura 5.7 pode-se ver um gráfico da Série original. Na figura 5.8 encontra-se um gráfico da Série normalizada de acordo com o procedimento da seção 5.1.1.

Figura 5.7 – Série *Random Walk with Drift* utilizada no experimento.Figura 5.8 – Série *Random Walk with Drift* normalizada.

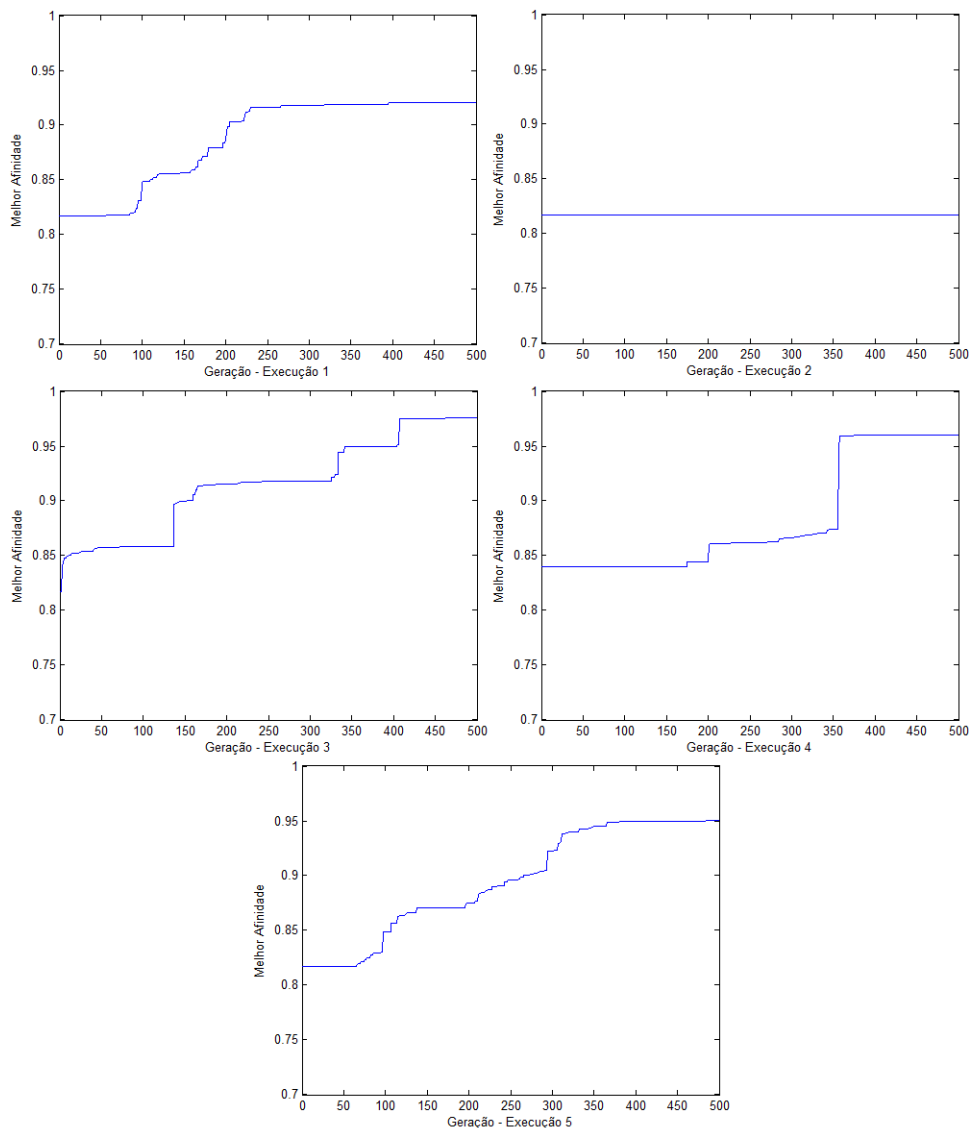
Para o experimento inicial, foram feitas 5 execuções do método, com os seguintes parâmetros:

- Número de Gerações: 500.
- Tamanho da População de Anticorpos: 80.
- Número de Anticorpos selecionados m : 40.
- Limite α de similaridade entre os anticorpos: 0.9.
- Tamanho da Região Hipervariável: 25.
- Tamanho da Região de Constantes: 9.

Da Série Original, 80% dos valores (400 medições) foram utilizados para criar o modelo, as outras medições foram separadas para medir a qualidade dos modelos encontrados. Essas últimas 100 medições não foram consideradas pelo algoritmo.

O gráfico da melhor afinidade do melhor anticorpo por geração demonstra a evolução da afinidade da população. Na figura 5.9 pode ser visto o gráfico de Melhor Afinidade por Geração.

Figura 5.9 – Melhor Afinidade por Geração nas 5 execuções iniciais.



Pode-se verificar que, com exceção da segunda execução, houve uma melhora da melhor afinidade da população ao longo das 500 gerações. Devido ao processo ser em grande parte probabilístico, pode acontecer um caso como na Execução 2, em que as mudanças aleatórias através do processo de mutação não conseguiram encontrar um anticorpo com melhor afinidade.

Utilizando a equação de Raiz do Erro Quadrático Médio, definida no Capítulo 5, determina-se a qualidade das soluções de cada execução, e pode-se fazer uma comparação em relação à afinidade da solução:

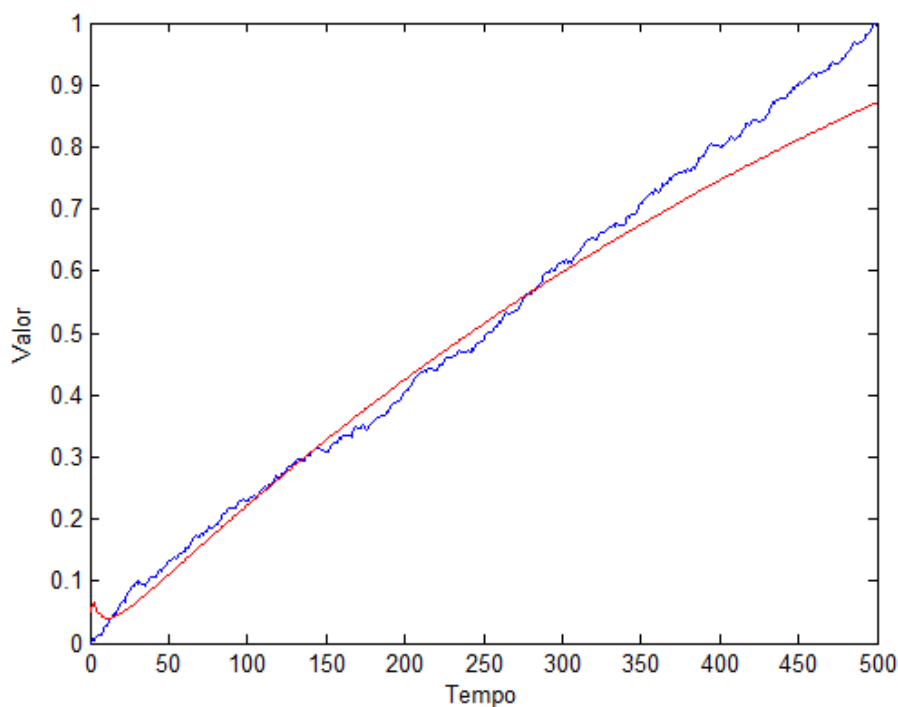
Tabela 5.1 – Erro Quadrático Médio e Afinidade por Execução.

	Raiz do Erro Quadrático Médio	Afinidade
Execução 1	0,2581	0,920985
Execução 2	0,4993	0,817068
Execução 3	0,0870	0,975881
Execução 4	0,1137	0,960284
Execução 5	0,1705	0,950344

Pode-se verificar que existe uma relação inversa entre a Raiz do Erro Quadrático Médio e a Afinidade da solução encontrada. Porém, essa relação nem sempre ocorre, pois é possível que um modelo se adapte melhor aos valores passados da série, porém preveja os valores futuros com mais erro do que outro que tem uma afinidade menor com os elementos passados da Série.

Utilizando a equação com menor erro (a solução da Execução 3) e fazendo o plot da série, podemos ver que a solução encontrada é bem próxima da série:

Figura 5.10 – Gráfico da Série Original (azul) e do modelo da Execução 3 (vermelho).



5.2.2 Mercado de Ações

Um dos exemplos mais comuns utilizados no estudo de Séries Temporais são valores obtidos do Mercado de Ações.

Os índices de retorno, ou *Stock Market Returns*, são medidos diariamente em diversas bolsas ao redor do mundo, e são indicativos dos valores que os investidores retiram do mercado de ações. Esses valores podem corresponder ao lucro através da compra e venda de ações ou através do pagamento de dividendos. (ECONOMY WATCH, 2013).

O índice Ibovespa é um índice de retorno total, tem como objetivo ser o indicador do desempenho médio das cotações dos ativos de maior negociabilidade e representatividade do mercado de ações brasileiro. (BM&FBOVESPA, 2014).

A série utilizada por este experimento (Fonte: BM&FBOVESPA, 2014) são os valores de fechamento diários do Índice Bovespa no período de Janeiro de 1993 a Dezembro de 1995. Este índice só é cotado em dias úteis, e neste intervalo houve 736 medições. Destas, as 588 iniciais serão usadas pelo algoritmo, e as restantes serão utilizadas para testar a qualidade das soluções.

Na figura 5.11 encontra-se um gráfico da Série e na figura 5.12 encontra-se um gráfico na Série normalizada de acordo com a seção 5.1.1.

Figura 5.11 – Série Ibovespa de Janeiro de 1993 a Dezembro de 1995.

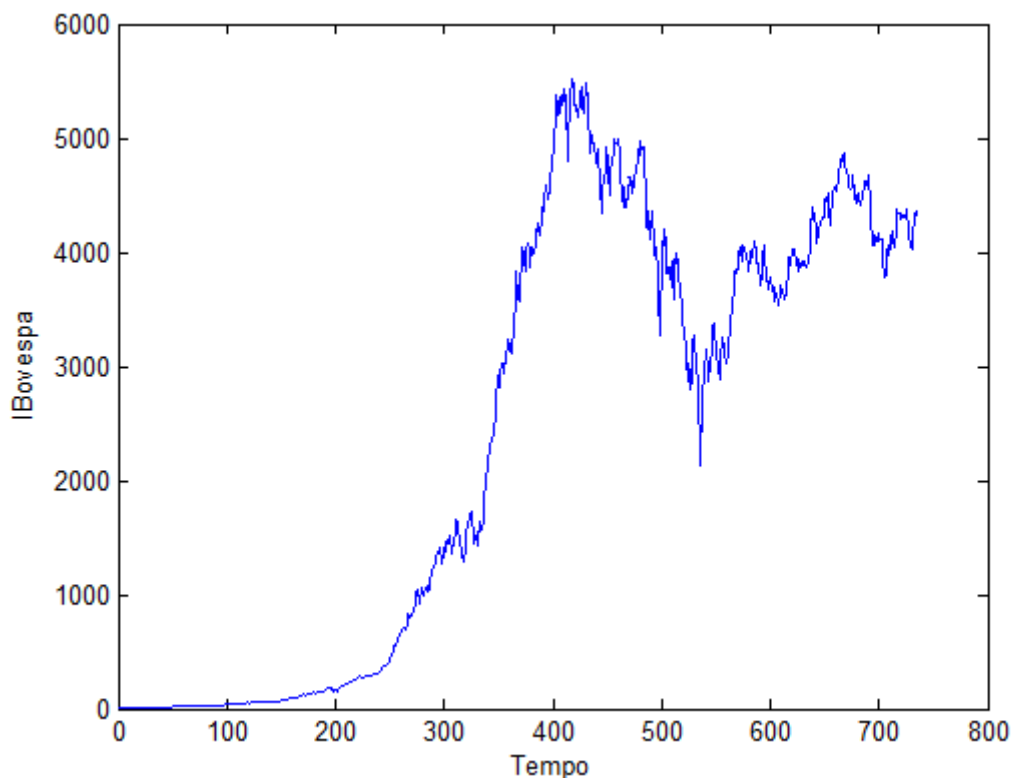
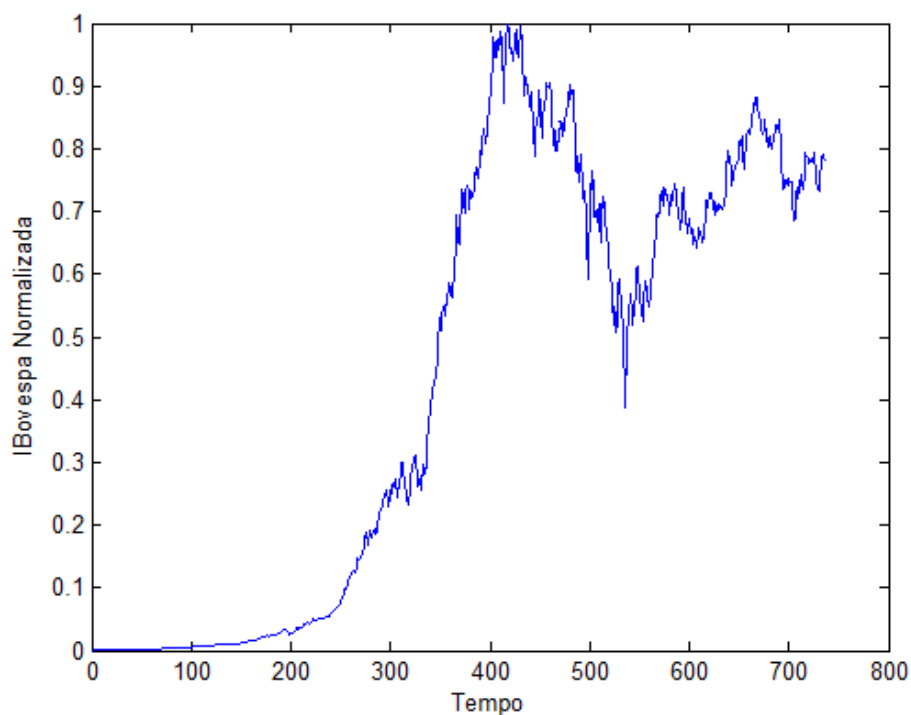


Figura 5.12 – Série IBovespa Normalizada.



Para este experimento, foram feitas 5 execuções do método, com os seguintes parâmetros:

- Número de Gerações: 500.
- Tamanho da População de Anticorpos: 80.
- Número de Anticorpos selecionados m : 40.
- Limite α de similaridade entre os anticorpos: 0.9.
- Tamanho da Região Hipervariável: 25.
- Tamanho da Região de Constantes: 9.

Na figura 5.13, pode-se ver os gráficos das evoluções das afinidades ao longo das gerações nas 5 execuções da implementação.

Na tabela 5.2 apresenta-se a qualidade das soluções de cada execução utilizando a equação da Raiz do Erro Quadrático Médio comparando as últimas 148 medições da Série e valores previstos pelos modelos, e comparando com a Afinidade de cada solução.

Tabela 5.2 – Erro Quadrático Médio e Afinidade por Execução.

	Raiz do Erro Quadrático Médio	Afinidade
Execução 1	0,0800	0,881033
Execução 2	0,2770	0,911769
Execução 3	0,3110	0,799312
Execução 4	0,1946	0,869913
Execução 5	0,4141	0,823247

Neste experimento, ocorre uma situação diferente. Nesse caso, a execução 1 produziu um modelo que prevê melhor os valores futuros da série, porém o modelo apresenta afinidade mais baixa com as medições anteriores da série. Nas figuras 5.14 e 5.15 encontram-se gráficos dos modelos das execuções 1 (com menor erro) e 2 (com maior afinidade) (em vermelho) e da Série original normalizada (em azul).

Figura 5.13 – Evolução da Afinidade do Melhor Anticorpo da População.

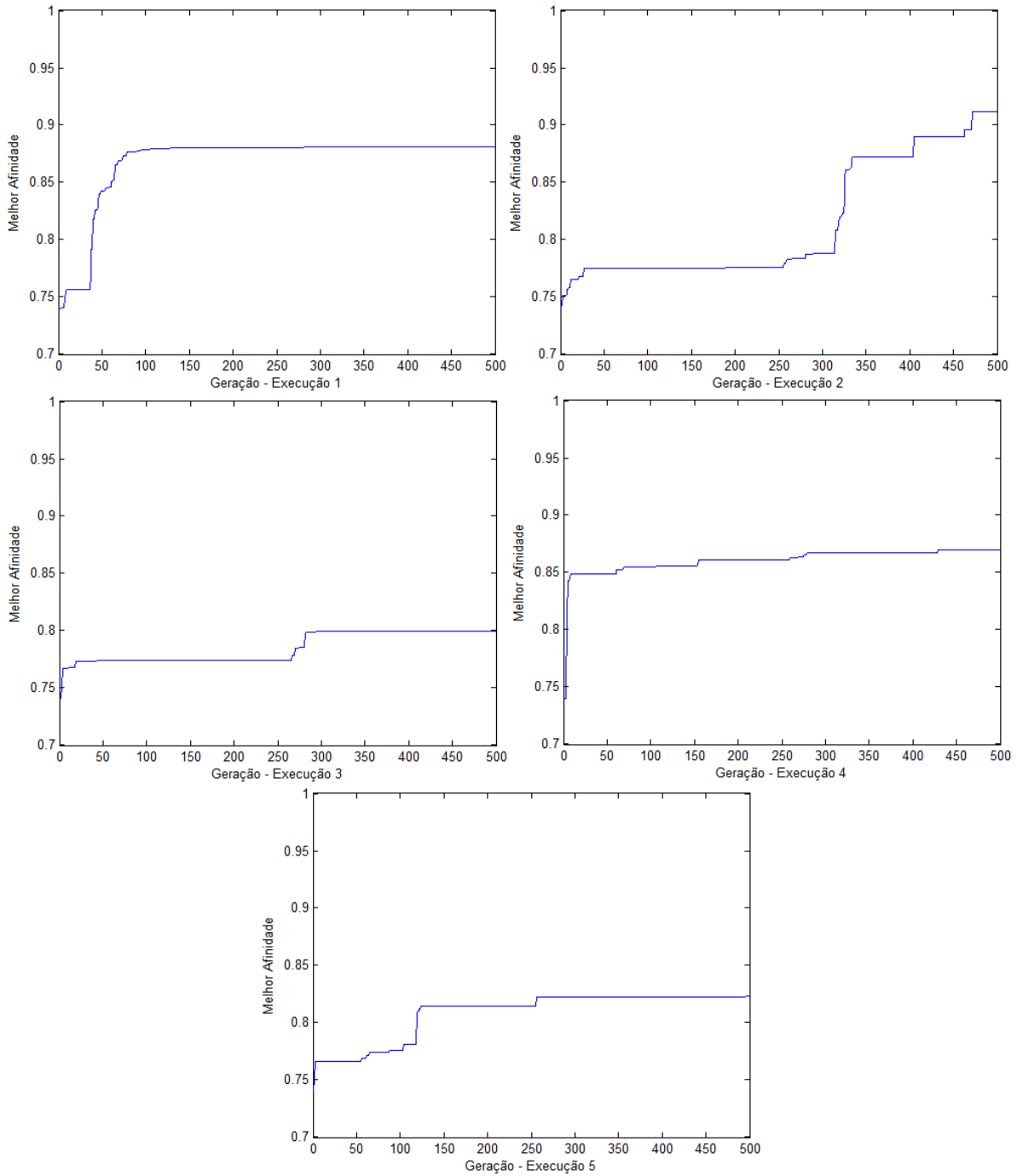


Figura 5.14 – Série Original Normalizada (azul) e Modelo da Execução 1 (vermelho).

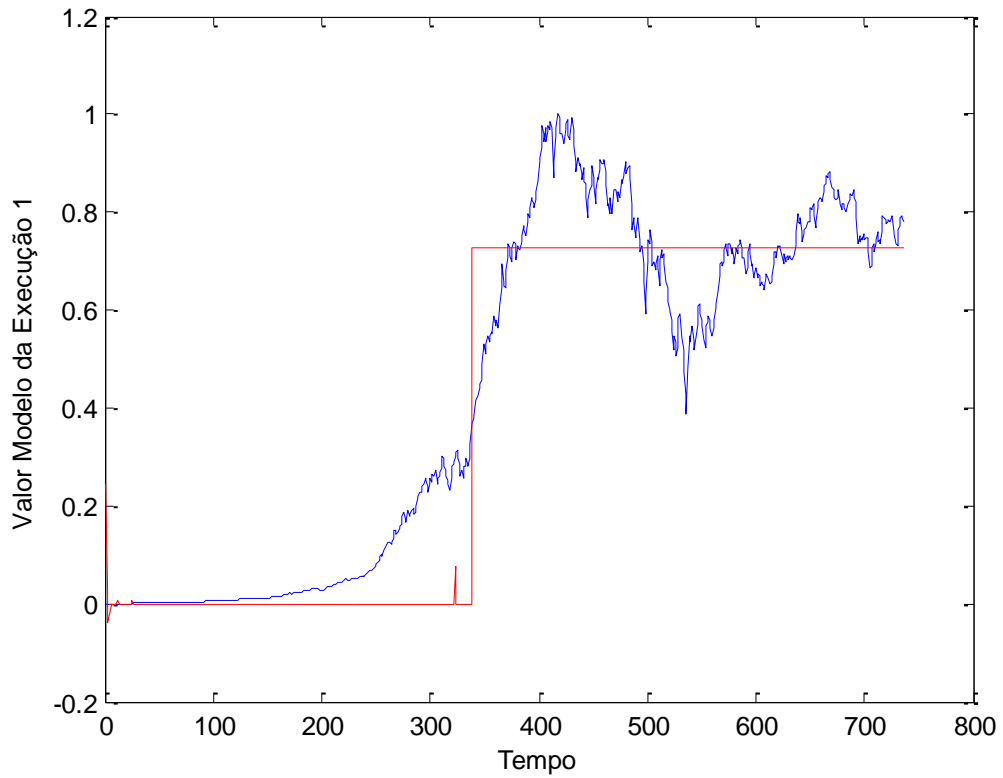
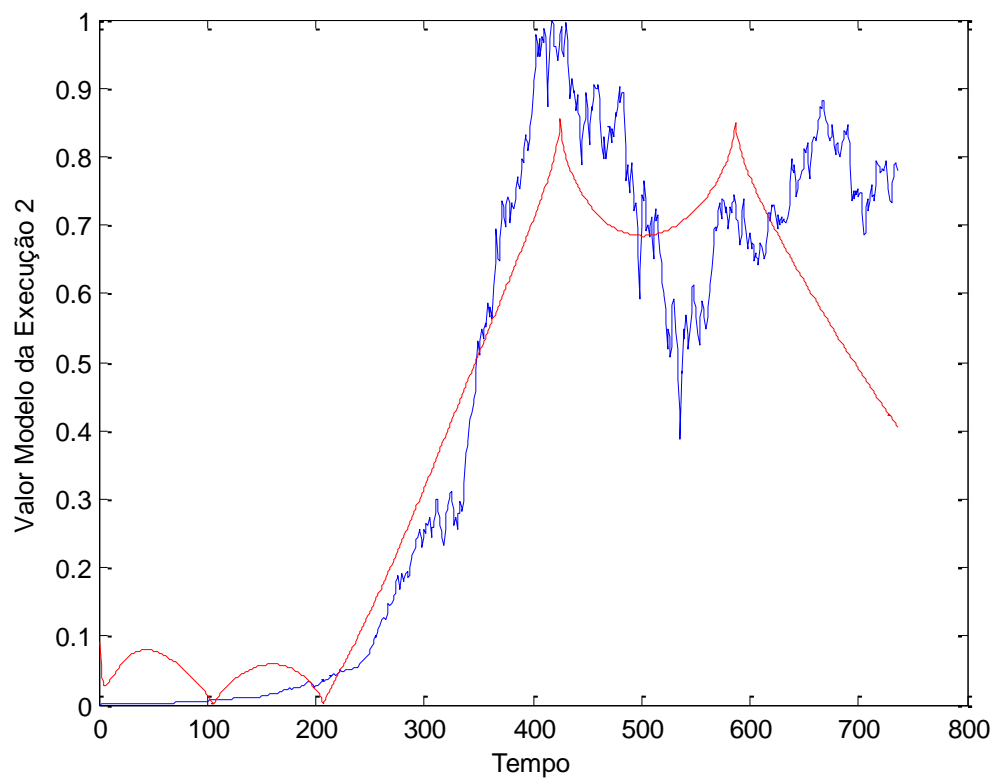


Figura 5.15 – Série Original Normalizada (azul) e Modelo de Execução 2 (vermelho).



5.2.2.1 Impacto dos Parâmetros na Implementação

É possível verificar nas figuras 5.9 e 5.13 que a afinidade do melhor Anticorpo da população nunca piora, sempre fica o mesmo ou aumenta. Então se pode esperar que aumentando o número de gerações sempre resulte em um Anticorpo com mesma afinidade ou melhor.

Aumentando o tamanho da população de Anticorpos também se pode esperar que melhorasse a afinidade do resultado, dado o mesmo número de gerações, por permitir maior variabilidade na população e com isso possa chegar a uma solução melhor mais rapidamente.

Já o número de anticorpos selecionados m gera alguma dúvida. A figura 5.16 mostra os gráficos da evolução da Afinidade do melhor membro da população variando os valores de m .

Com um número m menor levaria a uma pressão evolucionária maior, enquanto que um valor m maior aumentaria a variabilidade dos anticorpos selecionados. Os resultados da figura 6.16 levam a crer que um valor de $m = 30$ é um valor melhor, porém seriam necessários mais testes para ter uma precisão melhor.

O limite α de similaridade é outro valor que não é tão fácil de determinar tão facilmente. Um valor de α muito alto faria com que mais Anticorpos muito similares fossem selecionados para Proliferação, enquanto que com um valor muito baixo, bons Anticorpos não seriam selecionados porque são parecidos demais com outros. A figura 5.17 mostra alguns gráficos a evolução da Afinidade do melhor membro da população ao longo de 500 gerações com diferentes valores de α .

Os tamanhos Lh , Lf e Lc também afetam na qualidade do resultado. Com os tamanhos maiores, o espaço de busca aumenta, mas também aumenta a quantidade de modelos representáveis. Isso pode fazer com a melhor Afinidade da população demore mais a convergir, porém, o resultado que sairá deve ser de melhor qualidade.

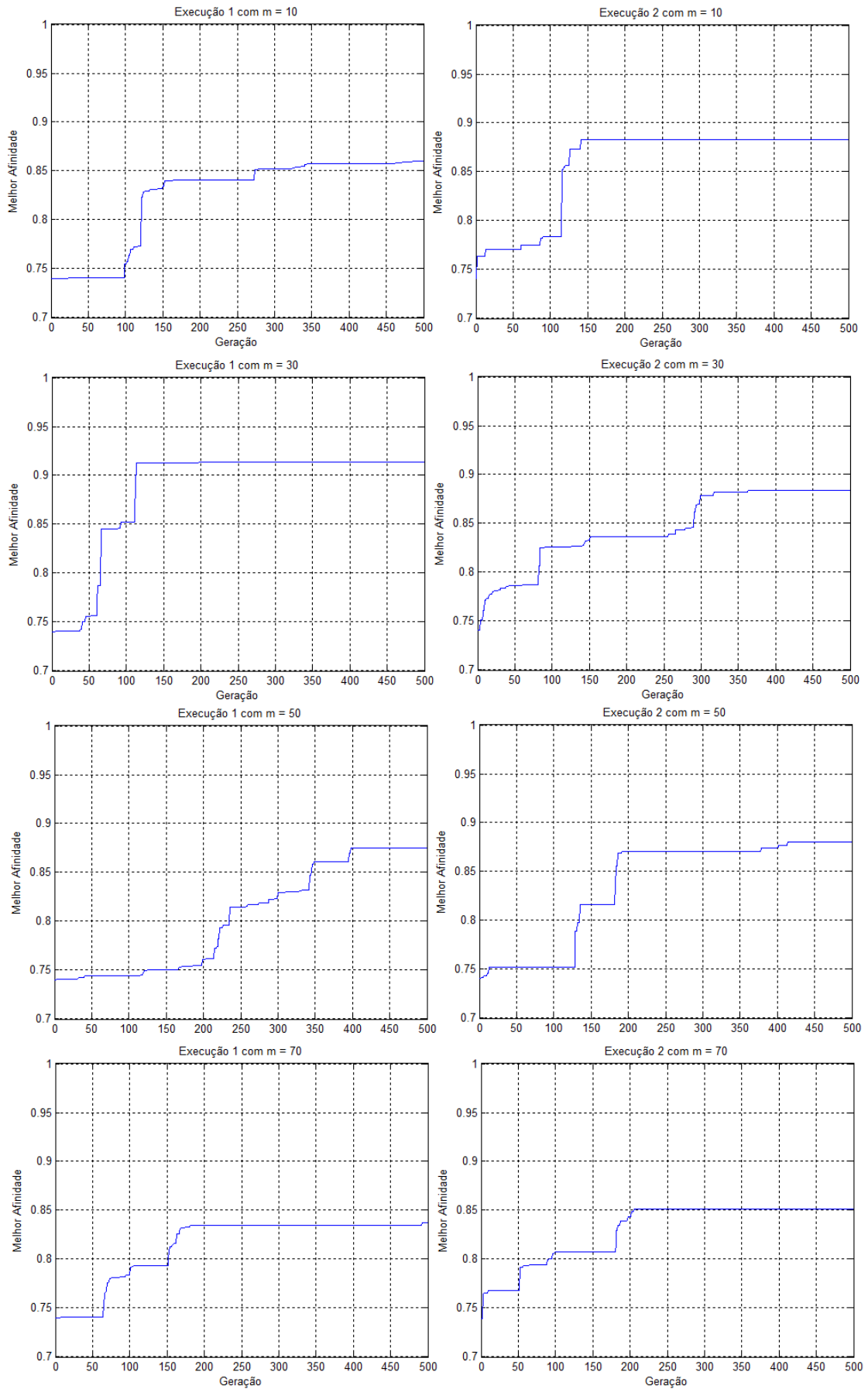
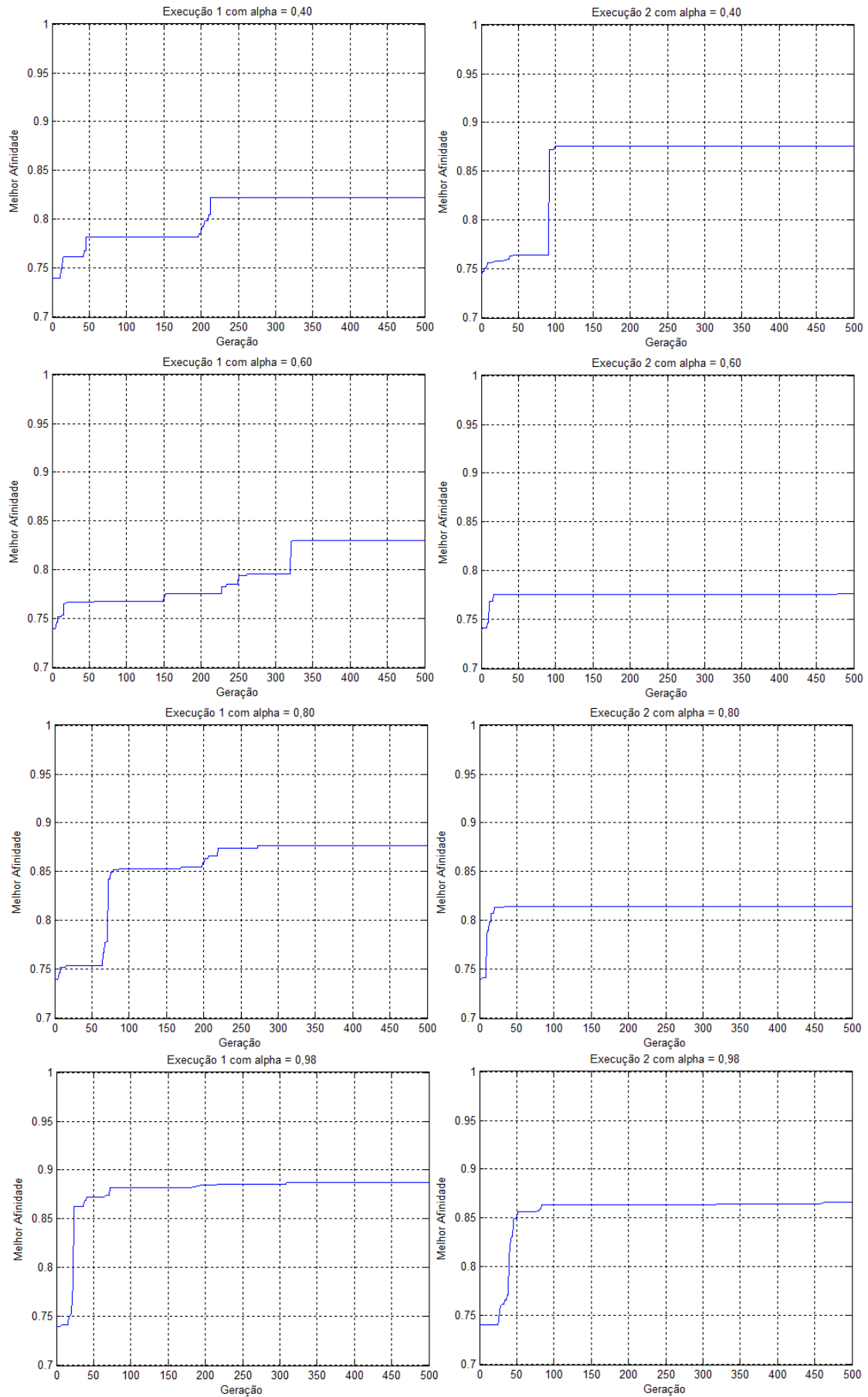
Figura 5.16 – Testando diferentes valores de m .

Figura 5.17 – Testando Diferentes Valores de α .



7 CONCLUSÕES

Neste trabalho vimos alguns exemplos de métodos de Previsão de Séries Temporais e uma introdução aos conceitos de SIA e Algoritmo de Seleção Clonal, e maneiras de como implementar estes conceitos com o objetivo de Prever Séries Temporais.

O Algoritmo de Seleção Clonal se assemelha muito a outras técnicas de Computação Evolutiva, como Algoritmos Genéticos e Programação Genética, e muitos dos conceitos a respeito da adaptação do Algoritmo de Seleção Clonal para Prever Séries Temporais poderiam ser facilmente aplicados à um algoritmo baseado em Programação Genética, apenas trocando alguns conceitos como:

- Afinidade: a medida de afinidade passaria a ser o *Fitness*.
- Mutação: a mutação passaria a ter uma Taxa constante e ter um papel secundário na criação de diversidade da população.
- Reprodução: deixaria de ser por clonagem e passaria a ser Sexuada, encontrando algum operador de *Cross-Over* para a recombinação dos pais.

Os resultados obtidos nesse trabalho levam a crer Algoritmo de Seleção Clonal é fraco em relação a Programação Genética. Um fator que possa ser relevante a isso é o tipo de reprodução, que é assexuada. A própria solução de Reprodução Sexuada surgiu e vingou na natureza por trazer vantagens evolutivas. Na reprodução Sexuada, partes dos pais se recombinam, o que cria a possibilidade de criar um herdeiro que possua somente as melhores partes dos pais. Essa probabilidade pode ser baixa, mas com repetição suficiente, a probabilidade desses eventos ocorrerem é alta demais para ser desprezada.

Um algoritmo de Programação Genética, utilizando a mesma representação em formato de árvore, traria muitas vantagens para um Sistema de Previsão de Séries Temporais se permitisse o *Cross-Over* entre os genes dos pais de maneira que sub-árvores de ambos os pais pudessem se recombinar de alguma maneira. A meu ver isto aceleraria o processo de evolução de uma solução.

Uma proposta muito interessante para um trabalho futuro seria implementar dois sistemas para Previsão de Séries Temporais, um de SIA e outro de Programação Genética, e comparar a velocidade que estes levam para chegar a uma solução de qualidade semelhante.

De qualquer jeito, as abordagens de Computação Evolutiva para Previsão de Séries Temporais parecem ser promissoras. Hoje em dia a tendência é a paralelização e difusão do processamento, e os algoritmos de Computação Evolutiva são facilmente paralelizáveis. As tarefas de busca probabilística que demorariam muito tempo para serem executadas podem ser paralelizadas de modo a aproveitar melhor os processadores de múltiplos núcleos hoje existentes no mercado.

REFERÊNCIAS

- Astiazara, M. V. **Sistema Imunológico Artificial para Predição de Fraudes e Furtos de Energia Elétrica**. Instituto de Informática - Universidade Federal do Rio Grande do Sul 2012.
- Bm&fBovespa, **Índice Bovespa**.
<http://www.bmfbovespa.com.br/indices/ResumoIndice.aspx?Indice=Ibovespa&Idioma=pt-br>.
 Acessado em Junho/2014.
- Castro, L. N. de; Timmis, J. **Artificial Immune Systems: a new computational intelligence approach**. London, UK: Springer, 2002.
- Chatfield, Christopher. **Time-series Forecasting**. Chapman & Hall/CRC 2000.
- Economy Watch, **Stock Market Returns**. <http://www.economywatch.com/stock-markets-in-world/returns.html>. Acessado em Dezembro/2013.
- Isaac Kohane, MD, PhD, The Children's Hospital, Boston, MA,
<http://archive.ics.uci.edu/ml/datasets/ICU>. Acessado em Maio/2014.
- Oliveira, L. O. V. B. **CSCDR: Um Classificador Baseado em Seleção Clonal com Redução de Células de Memória**. Instituto de Informática - Universidade Federal do Rio Grande do Sul 2012.
- Prochnow, F. A. **Programação Genética para Predição de Séries Temporais Aplicados a Mercados Financeiros**. Instituto de Informática - Universidade Federal do Rio Grande do Sul 2013.
- Rowe, G. e Wright, G. **The Delphi technique as a forecasting tool: issues and analysis**. Int. J. Forecasting, 15, 353-375.[1.1] 1999.
- Shumway, R. H. e Stoffer, D. S. **Time Series Analysis and its Applications With R Examples**. 3rd Edition Springer Texts in Statistics 2010.
- Sprott, J. C. **Chaos and Time-Series Analysis**. Oxford University Press, 2003.
- Yan Luo, Yuanquan Shi, **Immunity-based Nonlinear Time Series Prediction**. Journal of Information & Computational Science 8, 2011.
- Karris, Steven T. **Numerical Analysis Using MATLAB and Excel**. 3ed Orchard Publications, 2007.
- Yuanquan Shi, Xiaojie Liu, Tao Li, Xiaoning Peng, Wen Chen, Ruirui Zhang, Yanming Fu. **Chaotic Time Series Prediction Using Immune Optimization Theory**. International Journal of Computational Intelligence Systems, Suppl. 1 (December, 2010).
- Wei, William W. S. **Time Series Analysis: Univariate and Multivariate Methods**. 2nd Edition Addison Wesley 2006.