

1 Introdução

A explicação ou entendimento dos diversos fenômenos que ocorrem no mundo real é o propósito da ciência, a qual tem como objetivo remover a camada da aparência, perceptível aos sentidos, para revelar a essência da sua natureza subjacente. De fato, se aparência e essência fossem a mesma coisa, não haveria necessidade de ciência, pois todas as noções de senso comum teriam decifrado os seus segredos. Talvez a noção de senso comum mais profundamente arraigada acerca de nosso mundo seja a de que ele é tridimensional. Nem é preciso dizer que comprimento, largura e profundidade são suficientes para descrever todos os fenômenos de nosso universo visível. A inclusão do tempo como uma outra dimensão, e as conseqüentes quatro dimensões seriam suficientes para o registro de qualquer evento no universo. Não importa onde os “instrumentos” humanos tenham penetrado, desde as profundezas do átomo até os mais remotos confins do aglomerado galáctico, só encontra-se evidências dessas quatro dimensões. Entretanto, a Física afirma que existem mais dimensões, pois existem “eventos” que são impossíveis de ser explicados ou equacionados em quatro dimensões [1,2,3].

Além do escopo da Física teórica, os fenômenos do mundo real podem ser modelados genericamente por equações n -dimensionais, onde o conceito de dimensão é abstrato. Entretanto, mesmo nestes hiperespaços abstratos, a dimensão física do tempo está presente, pois no momento de registrar e quantificar os sinais desse fenômeno em dados, a dimensão física do tempo inevitavelmente deve ser considerada. Estas seqüências de dados, tipicamente, são denominadas como “séries temporais” [4,5] ou simplesmente “observações do fenômeno”, os quais descrevem padrões de comportamento temporal que podem ser complexos de serem modelados. Esta complexidade dificulta a dedução de equações matemáticas capazes de reproduzir esse fenômeno de forma artificial e com precisão absoluta.

Uma forma de contornar este problema é adaptar um modelo paramétrico $\mathbf{y}_t = \mathbf{f}(\boldsymbol{\theta}_t)$ às observações do fenômeno $\mathbf{Y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$, onde $\mathbf{y}_t \in \mathcal{R}^m$ é a variável aleatória que modela as observações do fenômeno e $\boldsymbol{\theta}_t \in \mathcal{R}^n$ modela os parâmetros adaptativos do modelo, os quais são calculados através de um processo de “aprendizado”. Ao longo deste trabalho o símbolo t será usado para representar o tempo, a não ser quando indicado de outra forma.

O termo aprendizado, usado na área de inteligência artificial, equivale ao termo “estimação” na nomenclatura estatística ou “filtragem do ruído” na nomenclatura de processamento de sinais. Independentemente da terminologia, todas elas se reduzem a um processo de amostragem de soluções ou a um processo de otimização dos parâmetros em relação ao erro produzido pelo modelo matemático [6]. Neste contexto surgiram dois paradigmas de aprendizado: o aprendizado *batch* ou em *batelada*, onde a otimização é realizada sobre todo o conjunto de treinamento simultaneamente; e o aprendizado *online* ou *seqüencial*, que ajusta os parâmetros do modelo seqüencialmente no tempo, após a apresentação de cada observação do fenômeno (exemplo ou padrão de treinamento).

Embora o aprendizado *batch* seja eficiente para pequenos e médios conjuntos de treinamento e para dimensões razoáveis do vetor de parâmetros do modelo, ele é ineficiente no caso de grandes conjuntos de treinamento e/ou quando a dimensão do vetor de parâmetros é elevada. Desta forma, o aprendizado *online* surge como uma necessidade para tratar este tipo de problema. Isto porque o método *online* representa uma abordagem natural para tratar processos dinâmicos não-estacionários, que evoluem no tempo, para os quais, com o aprendizado *batch*, o modelo teria que ser inevitavelmente retreinado cada vez que uma nova observação do fenômeno em estudo fosse concebida. Todavia, o aprendizado *online* exige sofisticados mecanismos de inferência no treinamento dos modelos pois o processo de otimização dos parâmetros é realizada sobre uma superfície de erro estocástica, ao invés de uma superfície de erro fixa, como é o caso do aprendizado *batch*.

Num contexto genérico, a diferença fundamental entre os modos de operação *batch* e *online* pode ser ilustrada através de um exemplo. Suponha que

se deseje calcular a média das observações de um fenômeno $\mathbf{Y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$. Este valor pode ser estimado através da seguinte equação

$$\hat{\mathbf{a}}_t = \frac{1}{t} \sum_{i=1}^t \mathbf{y}_i \quad (1)$$

Suponha-se agora que uma nova observação \mathbf{y}_{t+1} seja concebida. Considerando uma análise no modo *batch*, $\hat{\mathbf{a}}_{t+1}$ seria calculada usando uma fórmula equivalente à equação (1). Entretanto, no caso de uma análise no modo *online*, $\hat{\mathbf{a}}_{t+1}$ seria calculada através da equação:

$$\hat{\mathbf{a}}_{t+1} = \frac{t}{t+1} \hat{\mathbf{a}}_t + \frac{1}{t+1} \mathbf{y}_{t+1} \quad (2)$$

Embora os dois métodos forneçam o mesmo resultado neste caso, o método sequencial possui duas vantagens. A primeira é a redução no custo computacional devido a sua estrutura recursiva; a segunda vantagem é a redução dos recursos computacionais (memória) pois não é necessário armazenar toda a série histórica dos dados. A principal desvantagem do método *online*, entretanto, é o fato de nem sempre ser possível expressar as fórmulas matemáticas num formato recursivo exato. Usualmente são feitas aproximações recursivas que acumulam erros ao longo do tempo, podendo gerar problemas de convergência ou instabilidade.

A fim de tratar todas estas peculiaridades dos algoritmos de aprendizado *batch* e *online* no cálculo dos parâmetros do modelo é necessário primeiramente expressar a diversidade de modelos matemáticos existentes num formato comum, visando estabelecer uma plataforma que permita conceber estratégias ou paradigmas de aprendizado. Este é o assunto abordado na próxima seção.

1.1. Abstração do Modelo de Espaço de Estados

A maioria das formulações matemáticas que tentam modelar os fenômenos reais podem ser expressas num formato comum usando a abstração do Modelo de Espaço de Estados (MEE) [7,8]. Esta padronização permite expressar o modelo original através de duas novas equações matemáticas: a *equação de estado* e a *equação de observação*. A *equação de estado* modela os sinais não-observáveis do fenômeno e a *equação de observação* transforma estes sinais não-observáveis em sinais observáveis, os quais podem ser mensurados.

De forma genérica, a *equação de estado* e a *equação de observação* podem ser descritas, probabilisticamente, através das seguintes expressões matemáticas:

$$p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) \quad (3)$$

$$p(\mathbf{y}_t | \boldsymbol{\theta}_t) \quad (4)$$

onde a *equação de estado* $p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$ tipicamente é modelada como um processo estocástico Markoviano, ou seja, $p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_{t-2}, \dots, \boldsymbol{\theta}_0) = p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$, e a *equação de observação* $p(\mathbf{y}_t | \boldsymbol{\theta}_t)$ é modelada de forma que as observações sejam condicionalmente independentes aos parâmetros adaptativos $\boldsymbol{\theta}_t$, os quais são denominados tipicamente como “estados” do MEE.

1.2. As Estratégias de Aprendizado

As estratégias ou paradigmas de aprendizado são os diversos mecanismos de inferência utilizados no cálculo dos estados do MEE. Tipicamente, estes mecanismos de inferência são baseados em algum teorema matemático, princípio filosófico, princípio teórico, princípio físico etc., tais como o teorema de Bayes, o princípio evolucionário, princípio de máxima verossimilhança, o princípio de máxima entropia, entre outros.

Embora todos estes mecanismos de inferência tenham origens distintas, a eficiência computacional dos seus correspondentes “algoritmos de aprendizado” podem ser perfeitamente comparadas em termos de sua exatidão e do seu custo computacional.

Neste trabalho, estudou-se com ênfase o mecanismo de inferência Bayesiana, bem como o mecanismo de *inferência evolucionária* que foi desenvolvido neste trabalho.

1.2.1. A Inferência Bayesiana

A inferência Bayesiana [9-11] usa o conhecimento prévio disponível sobre o fenômeno para formular uma distribuição de probabilidade *a priori* sobre os estados do MEE e para formular uma função de verossimilhança que relaciona os estados com as observações do fenômeno. Desta forma, o cálculo dos estados é

realizado em função da distribuição de probabilidade *a posteriori*, a qual é obtida através do teorema de Bayes [12].

1.2.2. A Inferência Evolucionária

A inferência evolucionária, apresentada no capítulo 4, usa também o conhecimento prévio disponível sobre o fenômeno mas, neste caso, para criar uma população de indivíduos que representem soluções potenciais para os estados do MEE. Além disso, o conhecimento prévio do fenômeno é usado também para formular uma *função de avaliação* que relacione estas soluções com as observações do fenômeno. Desta forma, a solução final é obtida através de um processo evolutivo guiado pelo princípio Darwiniano da “sobrevivência dos mais aptos” [13].

1.3. Objetivos e Contribuições deste Trabalho

Na atualidade, o teorema de Bayes, o âmago da inferência Bayesiana, representa uma das estratégias ou paradigmas de aprendizado promissoras, uma vez que ele é imune a problemas caracterizados por comportamentos não-lineares, não-gaussianos e não-estacionários do fenômeno em estudo.

Infelizmente, nem sempre é possível obter uma solução analítica exata para esta distribuição *a posteriori*. Graças ao advento de um formidável poder computacional a baixo custo, em conjunto com os recentes desenvolvimentos na área de simulações estocásticas, este problema tem sido superado, uma vez que esta distribuição *a posteriori* pode ser aproximada numericamente através de uma distribuição discreta, formada por um conjunto de amostras.

Neste contexto, a seguir detalham-se os objetivos e contribuições deste trabalho:

1.3.1. Objetivos

Os principais objetivos deste trabalho são:

- Abordar o campo de simulações estocásticas sob a ótica da genética Mendeliana e do princípio evolucionário da “sobrevivência dos mais aptos”, de forma que o conjunto de amostras que aproxima a distribuição *a posteriori*, seja vista como uma população de indivíduos que tentam sobreviver num ambiente Darwiniano, sendo o indivíduo mais forte, aquele que possui maior probabilidade. Com base nesta analogia o objetivo é introduzir na área de simulações estocásticas (a) novas definições de *núcleos de transição* inspirados nos operadores genéticos de *cruzamento* e *mutação* e (b) novas definições para a *probabilidade de aceitação*, inspirados no *esquema de seleção*, presente nos Algoritmos Genéticos (AGs).
- Desenvolver a análise teórica das condições suficientes a serem satisfeitas pelas novas definições de *núcleos de transição* e pelas novas definições para a *probabilidade de aceitação* de forma a garantir a convergência do processo de aproximação numérica da *distribuição a posteriori*.
- Comprovar através de estudo de casos a eficiência e robustez dos novos algoritmos de aprendizado do MEE concebidos num enfoque Bayesiano e implementados na prática através de simulações estocásticas evolutivas.

1.3.2. Contribuições

- O estabelecimento de uma equivalência entre o teorema de Bayes e o princípio evolucionário, permitindo, assim, o desenvolvimento de um novo mecanismo de busca da solução ótima das quantidades desconhecidas, denominado de *inferência evolucionária*.
- O desenvolvimento do *Filtro de Partículas Genéticas* (FPG) e do *Filtro Evolutivo* (FE), que são algoritmos de aprendizado *online* e *batch* evolucionários, respectivamente. Além disso, mostra-se também que o *Filtro Evolutivo* é, em essência, um AG pois, além da sua capacidade de convergência à distribuições de probabilidade, o *Filtro Evolutivo* converge também a sua moda global.

- Em consequência da contribuição anterior, a fundamentação teórica do FE demonstra, analiticamente, a convergência dos AGs em espaços contínuos, demonstração esta de grande interesse da comunidade de pesquisadores de computação evolucionária.
- Em consequência da equivalência entre o teorema de Bayes e do princípio evolucionário, foi estabelecido a equivalência entre o FPG e o método de simulação estocástica denominado de simulação de *Monte Carlo Seqüencial* (MCS), uma vez que ambos métodos são usados na amostragem de distribuições de probabilidade variantes no tempo. Além disso, foi estabelecido a equivalência entre o FE e o método de simulação estocástica denominado simulação de *Monte Carlo via Cadeias de Markov* (MCCM), uma vez que ambos são usados na amostragem de distribuições de probabilidade estacionárias.
- Na comprovação da convergência dos algoritmos *online/batch* evolucionários, propostos neste trabalho, foram desenvolvidas aplicações em problemas de filtragem dos estados de um modelo não linear unidimensional, problemas de otimização, e do aprendizado de hiper-parâmetros de um modelo de volatilidade estocástica. Além disso, foram feitas diversas comparações com os respectivos algoritmos de aprendizado Bayesiano.

1.4. **Organização da Tese**

Esta tese está organizada em mais cinco capítulos, resumidos a seguir.

O Capítulo 2 descreve a estratégia Bayesiana de aprendizado, dando ênfase às implementações baseadas em simulações estocásticas, em especial os métodos MCS e MCCM.

O Capítulo 3 descreve a aplicação do princípio evolucionário em problemas de otimização. Basicamente é apresentado o estado da arte dos AGs, assim como são apresentadas as fundamentações teóricas que investigam seu processo de convergência.

O Capítulo 4 descreve a abordagem evolucionária do campo de simulações estocásticas sob a ótica da genética Mendeliana e do princípio evolucionário da

“sobrevivência dos mais aptos”. Também, são desenvolvidos a análise teórica de convergência do FPG e do FE. Cabe sublinhar que a análise teórica de convergência do FE demonstra também a convergência dos AGs em espaços contínuos, uma vez que o FE é em essência um AG pois além da sua capacidade de convergência à distribuições de probabilidade, converge também a sua moda global. Finalmente, com base nessa relação entre o FE e o AG é estabelecida uma equivalência entre o teorema de Bayes e o princípio evolucionário.

O Capítulo 5 refere-se ao estudo de casos, como a aplicação do princípio evolucionário na filtragem dos estados de um modelo não-linear unidimensional, em problemas de otimização e do aprendizado de hiper-parâmetros de um modelo de volatilidade estocástica.

No Capítulo 6 são apresentadas considerações finais e sugestões de trabalhos futuros.

1.5. Resumo

Neste capítulo foi apresentado o escopo do problema tratado, isto é o processamento de sinais de algum fenômeno em estudo. Este problema de forma genérica trata da análise seqüencial de dados, a qual envolve estimar quantidades desconhecidas em cada observação concebida do fenômeno. Também foi apresentado o MEE o qual permite modelar as quantidades desconhecidas do fenômeno como os estados de um processo Markoviano. Em seguida, foi apresentada brevemente as estratégias Bayesiana e Evolucionária de aprendizado dos estados do MEE. Neste contexto, foram apresentados os objetivos e contribuições desta tese e, finalmente, foi apresentada uma descrição dos capítulos restantes desta tese.

No próximo capítulo será abordada a estratégia Bayesiana de aprendizado dos estados do MEE, dando maior ênfase às implementações por métodos de simulações estocásticas, como é o caso dos métodos SMC e MCMC.