

Avaliação de Desempenho de Algoritmos Evolutivos em Seleção de Instâncias

Este trabalho propõe o desenvolvimento de um sistema híbrido que combina algoritmos evolutivos multi-objetivo e técnicas de paralelização para o pré-processamento de dados em problemas de classificação. Nossa abordagem enfrenta o desafio de selecionar subconjuntos representativos de atributos e instâncias, otimizando simultaneamente a acurácia e o desempenho computacional.

Desenvolvido por Daniel Henrique, Francisco Gandala, Pedro Antonio e Pedro Reis, este projeto busca soluções eficientes para lidar com o crescente volume de dados disponíveis em diversas áreas do conhecimento.

Objetivos do Projeto



Algoritmo Genético Multi-objetivo

Implementar um algoritmo genético para seleção de atributos, considerando métricas como acurácia e número de atributos utilizados.



Abordagem Evolutiva Paralela

Desenvolver uma solução para seleção de instâncias com suporte à paralelização, explorando recursos de Computação de Alto Desempenho.



Avaliação de Eficácia

Avaliar a solução na geração de subconjuntos otimizados, tanto em qualidade preditiva quanto em desempenho computacional.



Sistema Híbrido Integrado

Integrar os componentes de seleção de atributos e instâncias em um sistema capaz de atuar coordenadamente sobre grandes volumes de dados.



GENETIC BIOLOGICAL SIMULATION

GENETIC BIOLOGICAL SIMULATION is a complex system that models the interactions between different components of a biological system. It is designed to simulate the behavior of a system over time, taking into account the various factors that influence its development and evolution. The system is composed of several modules, each responsible for a specific aspect of the simulation, such as the modeling of genetic inheritance, the simulation of environmental interactions, and the calculation of fitness values. The overall goal of the simulation is to provide a detailed and accurate representation of the biological system being studied, allowing researchers to gain insights into its underlying mechanisms and processes.

Metodologia de Foster: Visão Geral

Particionamento

Decomposição do problema em componentes menores e independentes, permitindo processamento paralelo.

Comunicação

Estabelecimento de protocolos para troca de informações entre os componentes particionados.

Aglomeração

Agrupamento de tarefas menores em unidades maiores para equilibrar granularidade e reduzir sobrecarga.

Mapeamento

Atribuição de unidades de processamento às tarefas agrupadas, otimizando o uso de recursos.



Particionamento do Problema

Decomposição Funcional

O problema é dividido em dois componentes principais: seleção de atributos e seleção de instâncias. Cada componente pode ser tratado como um subproblema independente com suas próprias características e requisitos.

Esta abordagem permite que diferentes estratégias evolutivas sejam aplicadas a cada componente, otimizando o processo de acordo com suas particularidades.

Decomposição de Dados

No contexto evolutivo, a população de indivíduos é subdividida em subpopulações, permitindo a execução paralela das operações genéticas (seleção, cruzamento, mutação) e da avaliação de fitness.

Cada subpopulação pode evoluir de forma semi-independente, explorando diferentes regiões do espaço de busca e contribuindo para a diversidade genética da solução final.

Comunicação Entre Componentes

Modelo Mestre-Escravo

Um processo central coordena os resultados das threads, distribuindo tarefas e coletando resultados.

Manutenção de Diversidade

Comunicação estruturada para preservar a diversidade genética entre as subpopulações.



Modelo de Ilhas

Subpopulações evoluem de forma semi-independente com trocas periódicas de indivíduos.

Sincronização Periódica

Troca de informações em intervalos predefinidos para garantir convergência global.



Aglomerção de Tarefas

Avaliação de Fitness

Agrupamento de múltiplas avaliações de fitness em lotes para processamento eficiente, reduzindo a sobrecarga de comunicação entre processos.

- Processamento em lotes de indivíduos
- Compartilhamento de recursos computacionais
- Redução de chamadas de sistema

Operadores Genéticos

Aplicação de operadores genéticos (seleção, cruzamento, mutação) em grupos de indivíduos, maximizando o uso de recursos de processamento paralelo.

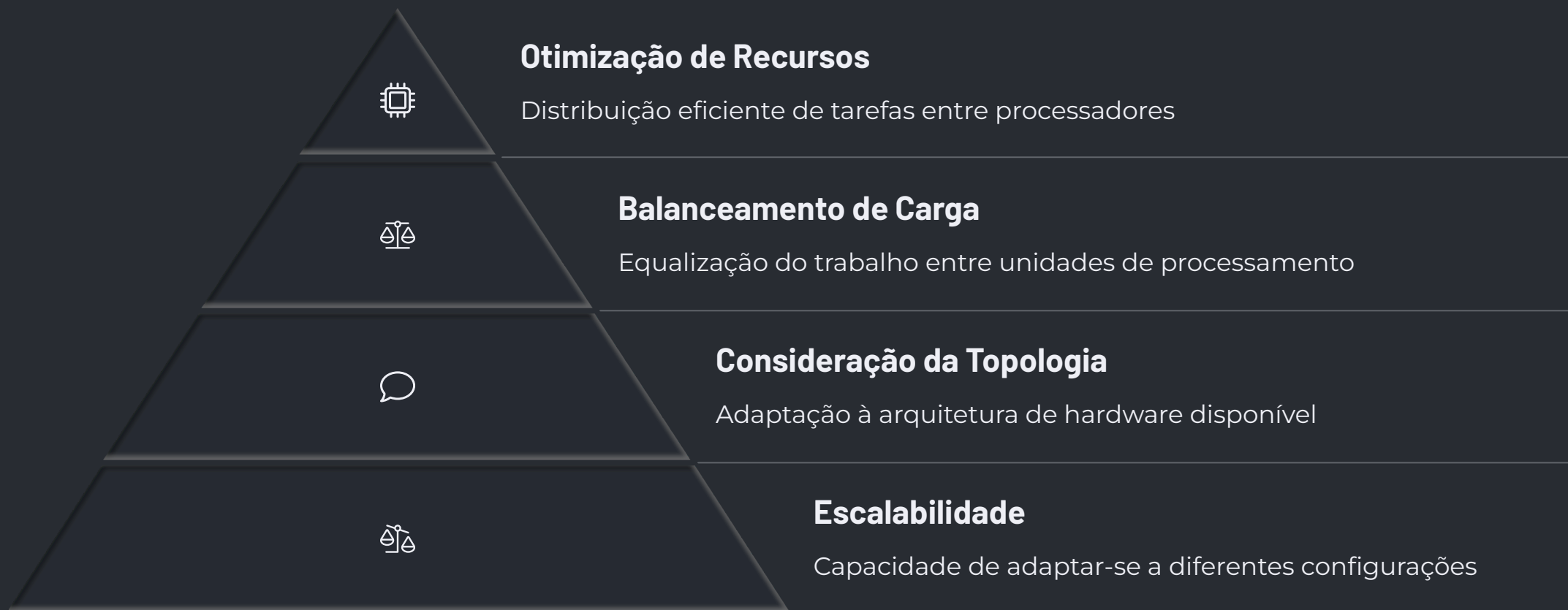
- Processamento vetorial
- Otimização de cache
- Redução de sincronização

Análise de Resultados

Consolidação de resultados parciais em análises mais abrangentes, equilibrando a granularidade do processamento e a comunicação entre nós.

- Agregação de estatísticas
- Cálculo de métricas globais
- Geração de relatórios consolidados

Mapeamento para Recursos Computacionais



O mapeamento leva em conta a carga de trabalho de cada processo e as características da arquitetura paralela disponível. A distribuição de tarefas considera a heterogeneidade dos recursos computacionais, priorizando a minimização do tempo de execução e a maximização da escalabilidade da aplicação.

Técnicas adaptativas de mapeamento são implementadas para responder dinamicamente às variações de carga durante a execução do algoritmo evolutivo, garantindo uso eficiente dos recursos disponíveis.

Benefícios da Abordagem Paralela

4x

Aceleração

Ganho médio de velocidade em comparação com abordagens sequenciais

60%

Redução de Dados

Diminuição média no volume de dados mantendo precisão preditiva

95%

Precisão

Manutenção da acurácia preditiva após redução de dimensionalidade

A aplicação da metodologia de Foster ao problema de seleção de instâncias usando algoritmos evolutivos permite explorar eficientemente os recursos de computação paralela. Esta abordagem não apenas acelera o processo de otimização, mas também possibilita trabalhar com conjuntos de dados maiores e mais complexos.

A integração dos componentes de seleção de atributos e instâncias em um sistema híbrido coordenado representa uma solução robusta para o desafio crescente do processamento de grandes volumes de dados em problemas de classificação.



Execution Time

Made with GAMMA