

Trabalho Final Parte-1

Integrantes do Grupo

- DANIEL HENRIQUE PINHEIRO SILVA - 202306405
 - FRANCISCO GANDALA RODRIGUES NETO - 202303333
 - PEDRO ANTONIO MACIEL SARAIVA - 202303358
 - PEDRO REIS PIMENTA - 202303359
-

Tema - **Avaliação de desempenho de Algoritmos Evolutivos em seleção de instâncias**

Introdução

Com o constante aumento na geração e disponibilidade de dados em diversas áreas, surge a necessidade de técnicas eficientes para redução de dimensionalidade e volume de dados em problemas de classificação, mantendo a precisão dos modelos preditivos. Este trabalho propõe o desenvolvimento de um sistema híbrido que une algoritmos evolutivos multi-objetivo e técnicas de paralelização para o pré-processamento de dados em problemas de classificação. A proposta busca enfrentar o desafio de selecionar subconjuntos representativos de atributos e instâncias, otimizando a acurácia e o desempenho computacional.

Objetivo

- Implementar um algoritmo genético multi-objetivo para a seleção de atributos, considerando métricas como acurácia e número de atributos utilizados.
- Desenvolver uma abordagem evolutiva para a seleção de instâncias com suporte à paralelização, a fim de explorar recursos de Computação de Alto Desempenho.
- Avaliar a eficácia da solução em gerar subconjuntos de dados otimizados, tanto em termos de qualidade preditiva quanto de desempenho computacional.
- Integrar os dois componentes (seleção de atributos e instâncias) em um sistema híbrido capaz de atuar de forma coordenada sobre grandes volumes de dados.

Modelagem (Metodologia de Foster)

A modelagem da solução segue as etapas propostas pela Metodologia de Foster para o desenvolvimento de aplicações paralelas:

- **Particionamento:** O problema é decomposto em dois principais componentes — seleção de atributos e seleção de instâncias. No contexto evolutivo, a população de indivíduos pode ser subdividida em sub-população, permitindo a execução paralela das operações genéticas (seleção, cruzamento, mutação) e da avaliação de fitness.

- **Comunicação:** As subpopulações necessitam trocar informações periódicas para garantir diversidade genética e convergência global. Essa comunicação pode seguir modelos como mestre-escravo (onde um processo central coordena os resultados das threads) ou ilhas (em que as subpopulações evoluem de forma semi-independente com trocas periódicas de indivíduos).
- **Aglomerção:** As tarefas menores, como avaliação de indivíduos ou aplicação de operadores genéticos, são agrupadas em unidades maiores para equilibrar a granularidade e reduzir a sobrecarga de comunicação. Isso permite maximizar a utilização dos recursos computacionais disponíveis.
- **Mapeamento:** Por fim, as unidades de processamento são atribuídas às tarefas agrupadas. O mapeamento leva em conta a carga de trabalho de cada processo e as características da arquitetura paralela, visando minimizar o tempo de execução e maximizar a escalabilidade da aplicação.