



Uso de Diagrama de Decisão Binária como acelerador para Random Forest

Olavo Barros*¹, Caio Von Rondow*² e Ricardo dos Santos Ferreira*³

*Departamento de Informática / UFV - ¹olavo.barros@ufv.br, ²caio.rondow@ufv.br e ³ricardo@ufv.br

Grande Área: Ciências Exatas e Tecnológicas - Área temática: Ciência da Computação - Categoria do Trabalho: Pesquisa

Palavras-Chave: Diagrama de Decisão Binária, Random Forest, Acelerador

Introdução

Uma Floresta Aleatória é um conjunto de árvores de decisão que votam na classe predominante após classificar uma amostra. Isso é feito usando um método de agregação, como o majority vote. **Apesar de ser fácil de entender, esse modelo é considerado uma "caixa preta" devido ao paralelismo envolvido na agregação de muitas árvores de decisão.**

Um Diagrama de Decisão Binária (DBB) é um grafo com dois tipos de nós: terminais (0 e 1) e não terminais (representando expressões booleanas). Os nós não terminais sempre têm dois nós filhos que representam resultados positivos e negativos da expressão. **DBBs têm a vantagem de representar funções booleanas de forma eficiente em termos de tamanho.** Sendo assim, um Diagrama de Decisão Binária (DBB), se mostra uma solução para sanar tais anteparos.

Objetivos

Analisar a possibilidade do uso de diagramas de decisão binária como representação de uma Floresta Aleatória e, além disso, apresentar um criador de DDB que realize essa transformação na floresta.

Ainda, como objetivos específicos, tem-se o estudo mais aprofundado dos **DBBs como meio de tornar o modelo de uma Random Forest mais "caixa branca"** e a criação de uma application programming interface (API) para a manipulação dos DBBs.

Material e Método

Foi usado estratégias de geração de códigos, como expressões lógicas para construção de operadores e padrões de projeto, juntamente com bibliotecas na linguagem de programação C++ e Python já amplamente conhecidas para manipulações de DBBs, sendo elas a **Colorado University Binary Decision Diagram (CUDD) e DD.**

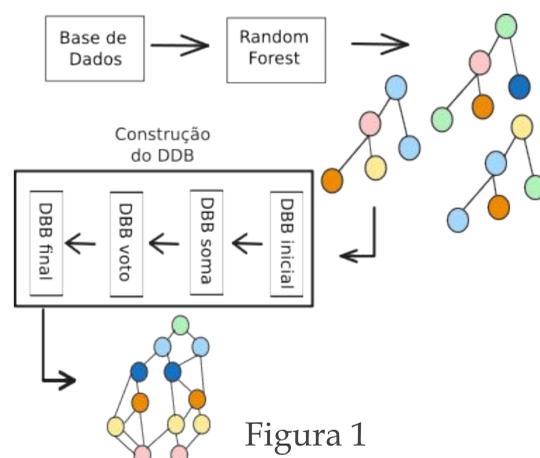


Figura 1

Foi modelado todo o caminho de construção para o DBB, como é apresentado na Figura 1, nela é visto: a formação da Random Forest usando a biblioteca em python Sklearn; a construção de um DBB inicial com as variáveis sendo as comparações das árvores; um segundo DBB com a soma das votações para cada classe; um terceiro para a votação; e um último para fazer a redução, no qual a saída já classifica em qual classe os dados de entrada são enquadrados.

Resultados e Discussão

Table 1: Dataset Thyroid, 4 classes, 20 atributos, and 9,172 instâncias.

Árvores	Profu.	Var	Inicial	Soma	Voto	Final
3	3	20	5	7	5	5
3	7	20	105	282	132	109
3	-	20	364	875	333	287
7	7	20	18	53	19	15
7	7	20	252	2,612	868	718
7	-	20	968	7,055	1,383	1162

A tabela 1 apresenta alguns dos resultados encontrados, nela as colunas representam, respectivamente, a configuração da Random Forest, como o número de árvores e a profundidade de cada uma, sendo o "-" representando a profundidade máxima; o número de variáveis do DDB; e o número de nós que o DDB possui em cada etapa de construção.

Portanto, é possível perceber que a estrutura apresentada consegue **representar uma Random Forest com somente 5 nós**, o que a torna um meio mais rápido de ser percorrido, acelerando a classificação.

Conclusões

É possível concluir, portanto, que o uso de DDBs como uma representação de um Random Forest é algo viável. Ainda, como estudos futuros, é importante investigar a escalabilidade do método, usando bases de dados que sejam mais desafiador.

O QR Code ao lado leva para uma visualização dos DBBs finais.



Bibliografia

GOSEN, F.; STEFFEN, B. **Algebraic aggregation of random forests: towards explainability and rapid evaluation.** International Journal on Software Tools for Technology Transfer, Springer, p. 1-19, 2021.
FERREIRA, R. S. **Power driven combinational synthesis: mapping and power estimation.** Tese (Doutorado) – UCL-Université Catholique de Louvain, 1999.

Agradecimentos e Apoio Financeiro

FAPEMIG (PIBIC, PQ-01577-22), NVIDIA, Xilinx, Funarbe, Laboratório Nacional de Computação Científica (Gen-RegAcc).

