



Universidade Federal de Pernambuco
Centro de Informática

Graduação em Ciência da Computação

Árvore de Decomposição Binária: Um novo paradigma para a classificação de múltiplas classes

Proposta de Trabalho de Graduação

Aluno: Caio Rocha Pereira

Índice

Índice	2
Informações sobre o aluno e o orientador	3
Contexto	4
Resumo	5
Cronograma	6
Referências	7
Assinaturas	8

Informações sobre o aluno e o orientador

- Aluno: Caio Rocha Pereira
- Curso: Ciência da Computação
- E-mail: crp3@cin.ufpe.br
- Área: Inteligência Computacional
- Orientador: Renata Maria Cardoso Rodrigues de Souza (rmcrs@cin.ufpe.br)
- Possíveis avaliadores:
 - Ricardo Bastos Cavalcante Prudêncio(rbcpr@cin.ufpe.br)

Contexto

Com o surgimento de novas tecnologias e grandes avanços na geração e armazenamento de dados, cresce progressivamente o interesse em extrair *insights* e informações valiosas da grande massa de dados gerados. A aprendizagem de máquina vem ganhando massivo interesse desde então, vista a grande capacidade da área em abranger várias técnicas capazes de extrair alguma informação valiosa para quem possui os dados.

O problema da classificação de múltiplas classes é um problema chave da área, visto que é bastante aplicado em toda a indústria, seja para prever fraudes, prever o tipo de um determinado cliente uma vez que o mesmo acione o seguro, dentre os mais diversos casos. Para isso, utiliza-se normalmente técnicas do estado da arte como *One-vs-One* e *One-vs-Rest*.(Bishop, 2006) Porém, recentemente surgiram estudos de algoritmos de esquematização de classificadores binários baseados (Fei & Liu, 2006) em estruturas de dados, tal qual um grafo ou uma árvore(Fernandéz, Barrachena, Bustince & Herrera, 2012). Isso permitiu que esquemas mais eficientes pudessem ser gerados, usando menos classificadores com perda mínima de desempenho.

Resumo

Consequente de diversos estudos sugerindo-a como técnica padrão para problemas de múltiplas classes, vistas a performance e complexidade da mesma comparada a abordagens nativas do próprio classificador, a binarização vem sendo bastante abordada por pesquisadores da área nos últimos anos.

Este trabalho propõe a árvore de decomposição binária como uma nova técnica de binarização, além de comparar a performance da mesma com técnicas do estado da arte desse problema. A árvore consiste em uma esquematizar o problema com base no dendrograma do agrupamento aglomerativo(Duda, Hart & Stork, 1995), fazendo com que o esquema consista em $N-1$ classificadores binários treinados, onde N é o número de classes, ao passo que *One-vs-One* e *One-vs-All* consistem em $\frac{n(n-1)}{2}$ e N classificadores binários, respectivamente.

Para isso, será feito um estudo das técnicas, e aplicação das mesmas sobre as bases de dados:

- abalone: <https://archive.ics.uci.edu/ml/datasets/Abalone>
- absenteeism: <https://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work>
- wholesale customers:
<https://archive.ics.uci.edu/ml/datasets/Wholesale+customers>
- iris: <https://archive.ics.uci.edu/ml/datasets/Iris>
- covertype: <https://archive.ics.uci.edu/ml/datasets/Covertype>
- seeds: <https://archive.ics.uci.edu/ml/datasets/seeds>
- leaf: <https://archive.ics.uci.edu/ml/datasets/Leaf>
- wine: <https://archive.ics.uci.edu/ml/datasets/Wine>
- image segmentation:
<https://archive.ics.uci.edu/ml/datasets/Image+Segmentation>
- multiple features: <https://archive.ics.uci.edu/ml/datasets/Multiple+Features>

visando analisar cada caso detalhadamente.

Este trabalho também toma como objetivo, de forma extra curricular, a escrita de um artigo analisando o desempenho do algoritmo, bem como comparando-o com outras técnicas.

Cronograma

Atividades	Agosto	Agosto	Setembro	Setembro	Outubro	Novembro	Novembro
Revisão Bibliográfica							
Implementação							
Experimentos							
Preparação da defesa							
Defesa							

Referências

Mitchell, T. M. (1997). Machine learning. WCB.

Fei, B., & Liu, J. (2006). Binary tree of SVM: a new fast multiclass training and classification algorithm. *IEEE transactions on neural networks*, 17(3), 696-704.

Duda, R. O., Hart, P. E., & Stork, D. G. (1995). Pattern classification and scene analysis 2nd ed. *ed: Wiley Interscience*.

Bishop, C. M. (2012). Pattern recognition and machine learning, 2006. *대한토목학회지*, 60(1), 78-78.

Galar, M., Fernández, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, 44(8), 1761-1776.

Cheong, S., Oh, S. H., & Lee, S. Y. (2004). Support vector machines with binary tree architecture for multi-class classification. *Neural Information Processing-Letters and Reviews*, 2(3), 47-51.

Aly, M. (2005). Survey on multiclass classification methods. *Neural Netw*, 19, 1-9.

Assinaturas

Caio Rocha Pereira

Renata Maria Cardoso Rodrigues de Souza