



UNIVERSIDADE
FEDERAL
DE PERNAMBUCO



**Universidade Federal de Pernambuco
Graduação em Engenharia da Computação
Centro de Informática**

**AVALIAÇÃO DE UM CLASSIFICADOR MULTICLASSE COM
ABORDAGEM ONE-VS-ONE USANDO DIFERENTES
CLASSIFICADORES BINÁRIOS
PROPOSTA DE TRABALHO DE GRADUAÇÃO**

Aluno: Victor Carriço Santos (vcs2@cin.ufpe.br)

Orientador: Cleber Zanchettin (cz@cin.ufpe.br)

Recife, 18 de setembro de 2016

Sumário

1. Contexto.....	3
2. Objetivos.....	4
3. Metodologia.....	4
4. Cronograma.....	5
5. Referências Bibliográficas.....	6
Assinaturas.....	7

1. Contexto

Classificação é um tipo de problema recorrente na área de aprendizagem de máquina que consiste em inferir em qual classe um determinado objeto pertence dentro de um número finito de classes. Suas aplicações são diversas e de grande importância quando se deseja que máquinas desempenhem funções que habitualmente são feitas por humanos, como reconhecimento de voz, imagem, detecção de SPAM na caixa de e-mail [3] e categorização de células cancerígenas [4].

Nesses problemas, quando o número de classes é maior que dois, ele é chamado de problemas de classificação multiclasse. Podemos dividir os métodos para a resolução desses problemas em duas categorias, alguns deles são métodos inicialmente elaborados para problemas de classificação binária mas que podem ser naturalmente estendidos para o caso multiclasse e outros precisam de uma formulação especial para se adequar ao caso de várias classes [6].

A primeira categoria inclui árvores de decisão [7], redes neurais [8], k-NN [9], Naive Bayes [10] e SVM [11] que são estendidos para o caso multiclasse. A segunda abordagem parte do princípio que problemas multiclasse são mais difíceis do que problemas de duas classes, pois as regiões de fronteiras de decisão tendem a ser mais complexas [5]. Logo, a técnica consiste em decompor o problema original em um conjunto de problemas de classificação binária que são resolvidos eficientemente com classificadores binários.

Uma das abordagens mais utilizadas para esse tipo de decomposição é a “um-contra-um”, one-against-one (OAO). Essa solução consiste basicamente em utilizar um algoritmo de classificação binário para treinar cada par de classes pertencentes ao problema e na predição analisar o resultado por todos eles para então inferir a classe do objeto.

Uma outra abordagem, bastante utilizada recentemente nos problemas de classificação, é a combinação de classificadores. Ela se baseia em uma variedade de métodos de classificação que podem obter diferentes taxas de acerto e a partir disso gerar resultados mais certos e precisos. Recentemente, o método chamado *diversified one-against-one (DOAO)* [12] foi proposto, ele consiste numa variação do *one-against-one* utilizando o conceito de combinação de classificadores, em que basicamente, para cada par de classe será utilizado o melhor classificador dentre um grupo de classificadores candidatos, tendo como objetivo o aprimoramento do seu desempenho.

2. Objetivos

Este trabalho de graduação tem como objetivo geral a realização de uma análise experimental e comparativa do método de classificação DOAO com outros métodos de combinação de classificadores e métodos de redução de problemas de classificação multiclasse para classificação binária.

Dentre os possíveis métodos que serão juntamente analisados com o DOAO estará o *Random Forests* e o *One-vs-All*. Utilizando como classificadores binários Rede Neural Artificial, Árvore de decisão, k-NN, Análise discriminante linear, Regressão logística e SVM.

A motivação deste trabalho é verificar a eficácia do método de classificação DOAO, proposto recentemente, com outros métodos mais conhecidos e de abordagem semelhante ao método em questão.

4. Metodologia

Os experimentos a serem realizados para a análise comparativa dos métodos serão feitos com o auxílio do *software*, feito em Java, *Weka 3* e da

biblioteca *Scikit Learn* para *Python*. Os métodos já conhecidos serão executados no *Weka* enquanto que o método proposto (DOAO) será desenvolvido em *Python 3* com o *Scikit Learn*.

Para comparar os classificadores as bases serão divididas pelo método *k-Fold Validation*, onde as bases são divididas em *k* conjuntos disjuntos e de mesmo tamanho. Usando *k-1* conjuntos para o treinamento e 1 conjunto para o teste, onde o número de conjuntos (*k*) ainda será definido. O processo é repetido de modo que cada conjunto será utilizado uma vez para teste.

Para todos os experimentos serão utilizadas as bases de dados do UCI, como *Zoo*, *Iris*, *Wine*, *Seed*, *Glass*, *Ecoli*, entre outras a serem escolhidas.

5. Cronograma

	Setembro	Outubro	Novembro	Dezembro
Criação da Proposta	■	■	■	
Levantamento bibliográfico	■	■	■	
Análise e estudo dos métodos		■	■	
Realização dos experimentos		■	■	
Análise dos resultados			■	
Escrita do relatório			■	■
Preparação da apresentação				■

6. Referências Bibliográficas

[1] Kang, S.; Cho, S.; Kang, P. Constructing a multi-class classifier one-against-one approach with different binary classifiers, *Neurocomputing*, Volume 149, Parte B, 3 de Fevereiro de 2015, Pág. 677-682

[2] Dietterich, T. G. (2000). *Ensemble Methods in Machine Learning*. In J. Kittler and F. Roli (Ed.) *First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science* (pp. 1-15). New York: Springer Verlag.

[3] Awad, W.A.; Elseuofi, S.M. (2011). Machine learning methods for spam e-mail classification, *International Journal of Computer Science & Information Technology*, Vol 3, No 1.

[4] Belekar, V.; Karthik, L.; Prabha, G. (2015) Classification of breast cancer resistant protein and non-Inhibitors using Machine learning approaches. *Combinatorial Chemistry & High Throughput screening*, Vol. 18 (pp. 476-485)

- [5] 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 Recognit.* (pp. 1761–1776)
- [6] Aly, M. *Survey on Multiclass Classification Methods*, The University of Utah, School of Computing, 2005.
- [7] Breiman L.; Friedman J.; Olshen R. A. and Stone C. J.. *Classification and Regression Trees*. Chapman and Hall, 1984.
- [8] Bishop C. M.; *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [9] Bay, S.D. Combining nearest neighbor classifiers through multiple feature subsets. In *Proceedings of the 17th International Conference on Machine Learning*, pages 37–45, Madison, WI, 1998.
- [10] Rish, I. An empirical study of the naive bayes classifier. In *IJCAI Workshop on Empirical Methods in Artificial Intelligence*, 2001.
- [11] Cortes, C.; Vapnik, V. Support-vector networks. *Machine Learning*, pages 273–297, 1995.