

RELATÓRIO DO DESAFIO 1 DO PROCESSO SELETIVO DA VIRTURIAN

Candidato: Diego Silva Caldeira Rocha {diegoscr.ec@gmail.com}

Abordagem formulada:

Nós selecionamos um dataset clássico de mineração de dados. Gostaríamos que você nos apresentasse uma solução de classificação dos sinais que nos indique claramente qual grupo de onda representa cada trecho dos dados.

Metodologia realizada

Para fazer esta classificação foi utilizada o método supervisionado, através do arquivo de dados *waveforme.date*, foi utilizada as primeiras 104 instâncias das 5000, ou seja 2% do total de amostras do sinal para tabelar os atributos(parâmetros) das três formas de sinais, havendo assim 37 conjuntos de atributos da onda da classe “zero”, 30 conjuntos de atributos da onda da classe “um” e 37 conjuntos de atributos da onda da classe “dois”. Estes conjuntos resultaram em 3 matrizes: MS0[37][21], MS1[30][21] e MS2[37][21].

Na etapa seguinte de classificação para cada uma das demais instâncias do banco, foi realizado para cada conjunto das matrizes foi calculada a média da distancia euclidiana dos atributos, comparando o valor resultante com a três matrizes e classificando para cada grupo a menor média da diferença euclidiana dos atributos.

Este método foi utilizado a fim de utilizar menos custo computacional e a media dos atributos ocasiona com um filtro passa baixa fazendo com que diminua o impacto dos ruídos na classificação.

RESULTADOS OBTIDOS

Para um total de 5129 amostras, obteve-se um aproveitamento de 57.4%. O resultado foi bom para utilização de apenas 2% de instâncias coletadas, sendo assim o aproveitamento pode melhorar bastante com utilização de mais coletas.