

01

Salvando e carregando um modelo

Transcrição

Vamos supor que para um caso real, fizemos o treinamento com o algoritmo IBk, e o melhor valor de classificação que conseguimos chegar foi de 71%.

Queremos salvar esse modelo treinado para poder mandar para outra pessoa usar com o Weka em outro lugar. Para fazer isso, abriremos outra base de dados, a base "banco_corte.arff". Como o nome diz, cortamos alguns dos dados, pois pegamos esses dados a ser utilizados novamente para fingir que são dados novos que com certeza não usamos durante o treinamento do algoritmo.

Realizaremos o treinamento com essa base de dados e após isso salvaremos o modelo treinado clicando com o botão direito sobre "lazy IBk" em "Result list" e em "Save model". Salvaremos o arquivo dentro da pasta "weka".

Agora vamos carregar o modelo salvo simulando uma situação em que só temos esse modelo e precisamos dele para fazer classificações novas. Clicaremos com o botão direito em qualquer lugar em "Result list" e em "Load model". Precisaremos agora mexer em algumas opções. Clicaremos em "Test options > More options" e vamos desselecionar todas as opções.

Em "Output prediction" clicaremos em "Choose" e selecionaremos "PlainText". Feito isso, daremos "Ok", marcaremos o campo "Supplied test set" e abriremos o arquivo "clientes.arff". Na sequência clicaremos com o botão direito no modelo carregado e escolheremos "Re-evaluate model on current test set", para reavaliar o modelo que acabamos de carregar no atual conjunto de testes.

Agora a avaliação do modelo será realizada de forma muito mais rápida, com base nos dados de testes que eram de clientes que não foram vistos durante o treinamento.

Conseguimos que fosse mostrado quais eram os valores reais na base de dados e quais foram colocados para a classificação. Ocorreram erros numa determinada porcentagem de valores, conforme o esperado, mas ele acertou a maioria. Quando aparecer o sinal de mais "+" na saída do classificador, será uma indicação de que houve um erro.

Se procurarmos a base de dados "clientes.arff", acessando na máquina via "weka > data > clientes.arff" podemos abri-la com o Notepad++ e ver que os valores "yes" ou "no" constarão nela. Num caso real, eles não existiriam, nós os colocamos apenas para fazer a comparação.

Poderíamos, por exemplo, ter um sinal de interrogação entre aspas simples ("?") no lugar desses valores, e mesmo assim o algoritmo conseguiria fazer a classificação, pois é o que nosso modelo aprendeu a fazer.