

4 Conclusão

A discussão dos resultados se encontra no capítulo anterior, deixando este último restrito a algumas conclusões e observações finais.

Nesta tese foram propostas 3 técnicas de classificação de padrões, a saber: Classificador por Quantização Vetorial (CQV), um algoritmo que combina os ambientes supervisionado e não supervisionado, onde, após uma QV, um procedimento é realizado para células que contenham observações de ambas as classes; Zona de Risco Generalizada (ZRG), um método independente de modelo que seleciona observações representativas com o objetivo de melhorar a classificação; Quantizador Vetorial das Fronteiras de Decisão (QVFD), um algoritmo cujo objetivo é quantizar as fronteiras de decisão para obter-se uma boa performance de classificação. Além disso, foram desenvolvidas aplicações de CQV e ZRG para seleção de observações não representativas.

Em termos de classificação, os métodos CQV e QVFD apresentaram resultados competitivos com métodos clássicos, o que demonstra a potencialidade destes algoritmos. Como boa parte dos bancos de dados utilizados nos experimentos desta tese possuem as classes desbalanceadas, uma análise foi feita neste sentido. Mais uma vez o equilíbrio entre os métodos prevaleceu, e houve uma variação entre os métodos que foram bem sucedidos de acordo com o experimento. Em particular, o QVFD demonstrou bom desempenho (principalmente levando em consideração a classe minoritária) no banco de câncer de pulmão, um banco com poucas observações, muitos atributos de entrada e desbalanceado. Entretanto, em outros bancos desbalanceados o desempenho do QVFD foi inferior a outros algoritmos levando em consideração a classe minoritária.

A contribuição denominada ZRG apresentou resultados muito interessantes quando aplicadas a uma variedade de técnicas, mantendo bom desempenho de classificação além de uma considerável redução do conjunto de treinamento. Esta abordagem foi comparada com os métodos IB3 e DROP3 que possuem um

excelente trade-off entre porcentagem de observações selecionadas e desempenho. A comparação mostrou que, apesar de selecionar mais observações que os métodos citados, o desempenho de classificação ao usar a ZRG é, em geral, melhor.

Considerando a questão de seleção de observações não representativas, foram propostas aplicações dos métodos CQV e ZRG a este assunto. Estas propostas demonstraram um bom índice de reconhecimento de observações com rótulo invertido, valorizando sua utilização em problemas ruidosos. Mais uma vez comparações com IB3 e DROP3 foram feitas, pois estes algoritmos também possuem uma excelente taxa de reconhecimento de observações ruidosas. Um desempenho razoavelmente equivalente a estes algoritmos foi encontrado, o que demonstra a potencialidade dos métodos para este tipo de aplicação.

Como trabalhos futuros, esperamos: combinar diferentes critérios para a classificação local no algoritmo CQV, em especial a utilização de *Information Theoretic Learning* para extrair informações locais das observações; estudar mais profundamente o comportamento dos algoritmos CQV e QVFD em bancos desbalanceados; investigar o impacto de algoritmos de seleção de atributos para os algoritmos CQV, QVFD e ZRG e estudar aplicações reais para o reconhecimento de observações com rótulo invertido.