

Classificação de dados de imagem em alta dimensionalidade empregando amostras semi-rotuladas e estimadores para as probabilidades a priori

Tayane Fiorelli Moreira¹, Delnir Monteiro de Lemos^{1*}
*Orientador

¹Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) -
Campus Rio Grande. Rio Grande, RS, Brasil.

Em cenas naturais, ocorrem com certa frequência classes espectralmente muito similares, isto é, os vetores média indicam valores muito próximos. Em situações como esta dados de baixa dimensionalidade (*LandSat-TM*, *Spot*) não permitem uma classificação acurada da cena. Por outro lado, sabe-se que dados em alta dimensionalidade tornam possível a separação destas classes, desde que as matrizes covariância sejam suficientemente distintas. Neste caso, o problema de natureza prática que surge é o da estimação dos parâmetros que caracterizam a distribuição de cada classe. Na medida em que a dimensionalidade dos dados cresce, aumenta o número de parâmetros a serem estimados, especialmente na matriz covariância. Contudo, é sabido que, no mundo real, a quantidade de amostras de treinamento disponíveis é frequentemente muito limitada, ocasionando problemas na estimação dos parâmetros necessários ao classificador, degradando, portanto a acurácia do processo de classificação, na medida em que a dimensionalidade dos dados aumenta. O *Efeito de Hughes*, como é chamado este fenômeno, já é bem conhecido no meio científico, e estudos vêm sendo realizados com o objetivo de mitigar este efeito. Entre as alternativas propostas com a finalidade de mitigar o *Efeito de Hughes*, encontram-se as técnicas que utilizam amostras não rotuladas e amostras semi-rotuladas para minimizar o problema do tamanho reduzido das amostras de treinamento. Com isso, utilizando-se funções de decisão mais ajustadas à realidade da cena analisada esperasse obter resultados mais acurados no processo de classificação. A contribuição desse trabalho consistirá na inclusão de uma etapa adicional, introduzindo a estimação das probabilidades a priori referentes às classes envolvidas para utilização no classificador Máxima Verossimilhança Gaussiana (MVG).

Palavras-chave: Sensoriamento Remoto. Reconhecimento de Padrões. Probabilidade a Priori. Amostras Semi-Rotuladas. Alta Dimensionalidade.

Trabalho executado com recursos do Edital PROPPi Nº 014/2015 – Fomento Interno 2016/2017, da Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação.