

Abstract

Mohamed S El-Mahallawy

Post-Clustering Soft Vector Quantization with Inverse Power-Function Distribution, and Application on Discrete HMM-Based Machine Learning

-In this paper, we introduce a soft vector quantization scheme with inverse power-function distribution, and analytically derive an upper bound of the resulting quantization noise energy in comparison to that of typical (hard-deciding) vector quantization. We also discuss the positive impact of this kind of soft vector quantization on the performance of machine-learning systems that include one more vector quantization modules. Moreover, we provide experimental evidence on the advantage of avoiding over-fitting and boosting the robustness of such systems in the presence of considerable parasitic variance e.g. noise, in the runtime inputs. The experiments have been conducted with two versions of one of the best reported discrete HMM-based Arabic OCR systems one version deploying hard vector quantization and the other deploying our herein presented soft vector quantization. Test samples of real-life scanned pages are used to challenge both versions hence the recognition error margins are compared.