

Abstract

Mohamed S El-Mahallawy

Soft Vector Quantization with Inverse Power-Function Distributions for Machine Learning Applications

This paper discusses the positive impact of soft vector quantization on the performance of machine-learning systems that include one more vector quantization modules. The most impactful gains here are 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 paper then introduces a soft vector quantization scheme with inverse power-function distributions, and analytically derives an upper bound of its relative quantization noise energy to that of typical (hard-deciding) vector quantization. This relative noise is expressed as a closed-form function of the power in order to allow the Selection of its optimal values of that compromise both a soft enough vector quantization with a stable performance via small enough relative quantization noise. Finally, we present empirical evidence obtained via experimenting with two versions of the best reported OCR system for cursive scripts - that happened to deploy discrete HMMs - one version with hard vector quantization and the other with our herein presented soft quantization. Test samples of real-life scanned Arabic text pages are used to challenge both versions hence the recognition error margins are compared.