Kernel Machine Classification Using Universal Embeddings

Petros T. Boufounos and Hassan Mansour
Mitsubishi Electric Research Laboratories,
Cambridge, MA 02139, USA, \{petrosb,mansour\}@merl.com

Abstract: Visual inference over a transmission channel is becoming an important problem in a variety of applications where, low latency and bit-rate consumption are often critical performance metrics, making data compression necessary. In this paper, we examine feature compression for support vector machine (SVM)-based inference using quantized randomized embeddings.

Universal Embeddings for Kernel Machines
We consider universal embeddings [1], namely transformations of the form \( \phi(x) = Q(Ax + e) \), where \( A \in \mathbb{R}^{M \times N} \) is a randomly generated matrix with i.i.d. standard normal entries, \( e \in \mathbb{R}^M \) is a random dither with elements drawn from an i.i.d. distribution uniform in \([0, \Delta]\), \( Q(y) \) is a non-monotonic scalar quantizer applied element-wise to its vector input, mapping \( y \) to 1 if \( y \in [2k, 2k+1) \) and to -1 otherwise, \( \Delta \) is a scaling parameter, and \( x \in \mathbb{R}^N \) is the vector being embedded—typically a feature vector or a signal to be classified. Universal embeddings have been shown to satisfy

\[
g \left( \|x - x'\|_2 \right) - \tau \leq d_H \left( \phi(x), \phi(x') \right) \leq g \left( \|x - x'\|_2 \right) + \tau,
\]

with overwhelming probability, where \( \tau \) decreases as \( 1/\sqrt{M} \), \( d_H(\cdot, \cdot) \) is the Hamming distance of the embedded signals and \( g(d) \) is the map

\[
g(d) = \frac{1}{2} - \sum_{i=0}^{\infty} \left( \pi(i + 1/2) \right)^{-2} e^{-\frac{(\pi i + 1/2)^2}{2 \Delta^2}} \approx \frac{d}{\Delta} \sqrt{\frac{2}{\pi}}, \text{ if } d \leq \frac{\Delta}{2} \sqrt{\frac{\pi}{2}} \text{ or } 0.5 \text{ otherwise.}
\]

We demonstrate that SVM kernels based on universal embeddings are very good approximations of radial basis function (RBF) kernels commonly used in classification. Thus, embedding features to a lower dimensional space is equivalent to using the SVM kernel trick with a kernel that approximates an RBF kernel.

Proposition. Let \( \phi(x) : \mathbb{R}^N \to \{-1, 1\}^M \) be a mapping function defined as above, with \( q = \phi(x) \). The kernel function \( K(x, x') \) given by \( K(x, x') = \frac{1}{2M} q^T q' \) is shift invariant and approximates the radial basis function \( K(x, x') \approx \frac{1}{2} - g \left( \|x - x'\|_2 \right) \), with \( g(d) \), as defined in (2). Furthermore, this RBF approximates the Gaussian RBF.

Our experimental results on an 8-class image database using histogram-of-gradients (HOG) features demonstrate that universal embeddings achieve 50% rate reduction over scalar quantization of the feature vectors, while maintaining the same inference performance.

References