Fast Approximation Algorithms for Kernel Methods

M. Griebel[1] and D. R. Wissel[2]

Abstract: The Gauss transform is an important numerical tool with many applications in, e.g. image manipulation, option pricing, and data mining including classification, regression and density estimation.

The d-dimensional Gaussian kernel $k_h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is defined as

$$k_h(t, s) = e^{-\|t-s\|^2/h^2},$$

where the parameter $h \in \mathbb{R}$ denotes the bandwidth and $\| \cdot \|$ is the Euclidean distance.

The discrete Gauss transform (DGT) is then defined as

$$G(t_i) = \sum_{j=1}^{N} f_j \cdot k_h(t_i, s_j) \qquad i = 1, \ldots, M,$$

with $t_i \in \mathbb{R}^d$, $i = 1, \ldots, M$, $s_j \in \mathbb{R}^d$ and $f_j \in \mathbb{R}$, $j = 1, \ldots, N$.

Several methods for the rapid approximation of the DGT have been developed to reduce the runtime complexity from $\mathcal{O}(N \cdot M)$ to $\mathcal{O}(N + M)$. The "Fast Gauss Transform" was introduced by Greengard and Strain and applied for problems in up to 3 dimensions. For higher-dimensional problems it however shows severe limitations since the dimension enters exponentially into the order constant. Similar algorithms like the "Improved Fast Gauss Transform" by Yang, Duraiswami and Gumerov and the "Dual-Tree Fast Gauss Transform" (DFGT) proposed by Lee, Gray and Moore introduce new series expansions and clustering schemes and thus allow for efficient evaluation for some datasets in dimensions up 10 and 16, respectively.

We present a new approach based on the dual-tree method that comes with a refined series expansion and an efficient local error and cost control scheme, combining the advantages of the previous methods. Computational results with synthetic as well as real-world data for up to 60 dimensions show competitive performance over the whole range of tested bandwidth values h. Our method can easily be generalized to work with so-called power series kernels. Finally, we present some approaches to exploit intrinsically low-dimensional features in the data and thus break the curse of dimensionality to a certain extent.

[1,2] Institute for Numerical Simulation
University of Bonn
Wegelerstr. 6, 53115 Bonn, Germany
griebel@ins.uni-bonn.de, wissel@ins.uni-bonn.de