Towards Deterministic Diverse Subset Sampling

J. Schreurs[1], M. Fanuel[1], and J.A.K. Suykens[1]

KU Leuven, Department of Electrical Engineering (ESAT),
STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics,
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
{joachim.schreurs,michael.fanuel,johan.suykens}@esat.kuleuven.be

1 Introduction

Selecting a diverse subset is an interesting problem for many applications like document or video summarization, image search tasks, pose estimation and many others. Diverse sampling algorithms have also shown their benefits to calculate a low-rank matrix approximations using the Nyström method [5], where using a dependent or *diverse* sampling algorithm for the Nyström approximation has shown to give better performance than independent sampling methods in [1,4]. An example of a diverse sampling algorithm is a determinantal point process (DPP) [3]. A DPP is a distribution over subsets of a fixed ground set, which can be seen as modeling a binary feature vector of length N. In DPPs, these binary variables are negatively correlated; that is, including one item makes the inclusion of other items less likely. As a result, DPPs assign higher probability to sets of items that are diverse. A k-DPP [2] is used when conditioning on a fixed cardinality k. This paper discusses a deterministic adaption of a k-DPP. Because the method is deterministic, there is no failure probability and the method always produces the same output.

2 Deterministic adaptation of a k-DPP

We discuss a deterministic adaptation of a k-DPP, by selecting iteratively landmarks with the highest probability. This corresponds to a greedy approach in finding a maximum a posteriori probability (MAP) estimate of a k-DPP. We successively maximize the probability over a nested sequence of sets $\mathcal{C}_0 \subseteq \mathcal{C}_1 \subseteq \cdots \subseteq \mathcal{C}_k$ starting with $\mathcal{C}_0 = \emptyset$ by adding one landmark at each iteration. Algorithmically, the proposed method corresponds to the DAS algorithm [1] with a different projector kernel matrix. More precise, DAS uses a smoothed projector kernel matrix with the ridge leverage scores on the diagonal. Where the proposed method has a *sharp projector kernel matrix* with the rank-k leverage scores on the diagonal. This has the added benefit that there is no regularization parameter to tune.

3 Numerical results and illustration

We evaluate the performance of the deterministic variant of the k-DPP on multiple real-life datasets, where we refer the reader to the original paper for a full description of te experiments. We observed that the proposed method samples a more diverse subset than the original k-DPP. The method shows superior accuracy in terms of the max norm of the Nyström approximation, along with better accuracy of the kernel approximation for the operator norm when there is fast decay of the eigenvalues. We demonstrate the use of the proposed method on a two image summarization tasks. An example is given on Figure 1. The dataset has 9 classes consisting of each 11 images. The method samples landmarks out every class, making it a desirable image summarization. This is supported by the projection of the landmarks on the 2 first principal components of the KPCA, where a landmark points is chosen out of every small cluster.

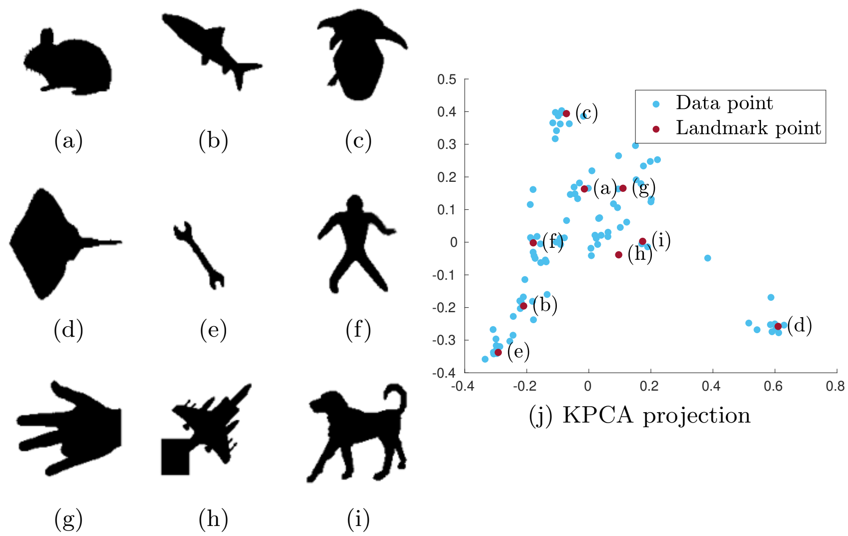

Fig. 1: Illustration of the proposed method with $k = 9$ on the `Kimia99` dataset. The selected landmark points are visualized on the left, the projection on the 2 first principal components of the KPCA on the right.

References

1. Fanuel, M., Schreurs, J., Suykens, J.A.K.: Nyström landmark sampling and regularized Christoffel functions (2019), `http://arxiv.org/abs/1905.12346`
2. Kulesza, A., Taskar, B.: k-dpps: Fixed-size determinantal point processes. In: ICML 11. pp. 1193–1200 (2011)
3. Kulesza, A., Taskar, B.: Determinantal point processes for machine learning. Foundations and Trends in Machine Learning **5**(2-3), 123–286 (2012)
4. Li, C., Jegelka, S., Sra, S.: Fast DPP sampling for Nyström with application to kernel methods. In: ICML 16. pp. 2061–2070 (2016)
5. Williams, C.K., Seeger, M.: Using the Nyström method to speed up kernel machines. In: Advances in neural information processing systems. pp. 682–688 (2001)