

Nyström Methods for Efficient Kernel-Based Methods for Community Question Answering

Danilo Croce, Simone Filice, and Roberto Basili

Department of Enterprise Engineering,
University of Roma Tor Vergata, Roma Italy
{croce,filice,basili}@info.uniroma2.it,

Kernel methods have been employed in several Machine Learning algorithms achieving state-of-the-art performances in many classification tasks. Recently, the kernel based approach presented in [4] has been applied in the community Question Answering (cQA) challenge at SemEval 2016 obtaining state-of-the-art results. Unfortunately, when large data volumes are involved, time and space complexity required in learning and classification may prevent the adoption of expressive but complex kernel functions, such as Sequence or Tree kernels. In [1] it has been shown that a viable and general solution to the above scalability issues is the Nyström methodology: it been designed to approximate the Gram Matrix derivable by a kernel function, enabling the projections of examples into low-dimensional spaces. Once each example is projected in such dense spaces, the application of efficient linear learning methods is enabled preserving at the same time the expressiveness and effectiveness of the original kernel methods.

In this paper, we show that the Nyström method can be effectively used in the cQA task, by adopting the same kernel functions proposed in [4] and obtaining the same results w.r.t. the metrics adopted in the SemEval task, by reducing the computational cost of orders of magnitude¹. The full version of this paper is provided in [2] and it is available in the proceedings of the *Third Italian Conference on Computational Linguistics*, CLIC-it 2016.

References

1. Croce, D., Basili, R.: Large-scale kernel-based language learning through the ensemble nystrom methods. In: 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20-23, 2016. Proceedings. pp. 100–112 (2016)
2. Croce, D., Filice, S., Basili, R.: Nyström methods for efficient kernel-based methods for community question answering. In: Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016), Napoli, Italy, December 5-7, 2016. (2016)
3. Filice, S., Castellucci, G., Croce, D., Basili, R.: Kelp: a kernel-based learning platform for natural language processing. In: Proceedings of ACL: System Demonstrations. Beijing, China (July 2015)
4. Filice, S., Croce, D., Moschitti, A., Basili, R.: Kelp at semeval-2016 task 3: Learning semantic relations between questions and answers. In: Proceedings of the SemEval-2016. pp. 1116–1123. San Diego, California (June 2016)

¹ All the above methods are implemented in the Kernel-based Learning Platform (KeLP), presented in [3] and available at: www.kelp-ml.org