Incorporating Prior Knowledge in Multiple-Output Regression with Kernel-Based Vector Functions

Jan Verwaeren[1], Willem Waegeman[1] Tapio Pahikkala[2], Antti Airola[2], and Bernard De Baets[1]

[1] KERMIT, Ghent University, Coupure links 653, 9000 Ghent, Belgium,
Jan.Verwaeren@UGent.be

[2] University of Turku, Department of Information Technology Turku Centre for Computer Science (TUCS) Finland

When learning a single-output regression or classification function, it is well known that prior knowledge about the underlying stochastic process can be incorporated to reduce the hypothesis space and improve the predictive performance of the learned model. Think in this context at frameworks such as graphical models, where dependencies between variables can be encoded by means of edges, or Bayesian approaches, where domain knowledge is typically enforced by defining an informative prior on model parameters. For kernel-based approaches, prior knowledge is often incorporated by adding constraints to the mathematical program that is used for learning [3–5]. Alternatively, prior knowledge can be included by using a prior knowledge-based kernel. As the latter approach translates the problem of prior knowledge incorporation into one of designing an appropriate kernel function, it provides a flexible way of incorporating different forms of prior knowledge such as invariance [7], meta-information about features [2] or knowledge provided by unlabeled data in semi-supervised learning settings [1]. All these approaches somehow encode prior knowledge that is relevant for the task at hand into the learning procedure.

The incorporation of domain knowledge in multiple output regression settings has not been extensively studied in the machine learning community, despite manifesting a number of important challenges. When solving multiple-output problems by learning a model for each output independently, the incorporation of prior knowledge remains a trivial extension of the single task setting. However, the independent modelling approach has been often criticized in the literature for ignoring potentially important dependencies in the label space – also recall in this context similar observations for multi-label classification problems. Unfortunately, for state-of-the-art algorithms, the incorporation of prior knowledge is often nontrivial, especially concerning the occurrence of computational bottlenecks.

Apart from the computational challenges, multiple output regression problems are also characterized by an increased flexibility in specifying domain knowledge. As a first contribution of this paper, we distinguish three types of prior knowledge that can arise in multiple output regression problems.

Firstly, we distinguish output-specific input-output relations as prior knowledge of dependencies between (a subset of) the inputs and one (or several) outputs. As a typical example, such knowledge appears in image reconstruction problems, when an image has to be reconstructed from blurred versions of that image. A particular pixel in the deblurred target image (one output) is probably most influenced by neighborhood pixels. As such, each target pixel has its own subset of input pixels to which it is most likely related.

Secondly, we distinguish input-input relations as a raw description of dependencies between inputs. This knowledge is not output-specific. Such dependencies can for instance be used to reduce the dimensionality of the input space or derive an application-specific regularizer.

Thirdly, we distinguish output-output relations that describe knowledge about dependencies between outputs. As discussed in Section 2, state-of-the-art algorithms for multiple output regression and multi-task learning often take dependencies between outputs or tasks into account, for example by clustering outputs. When raw information about such a clustering is given, it can be incorporated in the learning algorithm.

To date, existing approaches for multiple output regression do not take all above types of domain knowledge into account, while remaining at the same time computationally tractable. As the second contribution of this paper, we discuss how the above three types of prior knowledge can be incorporated in the powerful framework of kernel-based vector functions, introduced in [6]. In addition, we also introduce a new optimization algorithm for learning the resulting models efficiently, and we present experimental results on synthetic and real-world data.

References

1. R. Johnson and T. Zhang. Graph-based semi-supervised learning and spectral kernel design. *IEEE Transactions on Information Theory*, 54:275–288, 2008.
2. E. Krupka and N. Tishby. Incorporating prior knowledge on features into learning. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, 2007.
3. G. Kunapuli, R. Maclin, and J. W. Shavlik. Advice refinement in knowledge-based SVMs. In *Advances in neural information processing systems*, 2011.
4. F. Lauer and G. Bloch. Incorporating prior knowledge in support vector regression. *Machine Learning*, 70:89–118, 2008.
5. Q. Le and A. Smola. Simpler knowledge-based support vector machines. In *Proceedings of the International Conference on Machine Learning*, 2006.
6. C. A. Micchelli and M. Pontil. On learning vector-valued functions. *Neural Computation*, 17:177–204, 2005.
7. B. Schölkopf, P. Simard, A. Smola, and V. Vapnik. Prior knowledge in support vector kernels. In *Advances in neural information processing systems*, 1998.