Mach Learn (2010) 79: 1–3
DOI 10.1007/s10994-010-5169-8

EDITORIAL

Guest Editorial: Learning from multiple sources

Nicolò Cesa-Bianchi · David R. Hardoon · Gayle Leen

Received: 22 January 2010 / Revised: 22 January 2010 / Accepted: 26 January 2010 /
Published online: 23 February 2010
© The Author(s) 2010

The increasing availability of numerous heterogeneous but related sources of data arises in many fields, such as bioinformatics, robotics, computer vision, information retrieval, and many others. Multiple sources of data can be viewed as different, related views of the same learning problem, where dependencies between the views could potentially take on complex structures. This gives rise to interesting and challenging machine learning problems where data sources are combined for learning. This framework encompasses several data fusion tasks and related topics, such as transfer learning, multitask learning, multiview learning, and learning under covariate shift. Often the general concept is to augment training data for each problem of interest with data from other partially related learning problems. Several approaches for inferring and exploiting complex relationships between sources have been presented.

The advantages of the multiple source learning paradigm can be seen in situations where individual data sources are noisy, incomplete, and learning from more than one source can filter out problem-independent noise. Additionally, another aspect of multiple source learning is that data sources need only contain partially relevant information to the desired learning problem, making it possible to exploit a large number of data sources.

This special issue on "Learning from Multiple Sources" was inspired by a successful (eponymous) NIPS workshop which took place in Whistler (Canada) on December 13, 2008. The workshop received twenty-three submissions, from which seven contributed talks and

N. Cesa-Bianchi (✉)
Università degli Studi di Milano, Milano, Italy
e-mail: nicolo.cesa-bianchi@unimi.it

D.R. Hardoon
Institute for Infocomm Research, Singapore, Singapore
e-mail: drhardoon@i2r.a-star.edu.sg

G. Leen
Aalto University School of Science and Technology, Aalto, Finland
e-mail: gleen@cis.hut.fi

twelve posters were selected. The program also included two invited talks: "Learning from multiple sources by matching their distributions" by Tobias Scheffer, and "Challenges of supervised learning from multiple sources" by Francis Bach.

The papers presented in this special issue give an idea of the diversity of the problems, challenges, and methodological diversity associated with the topic.

A first group of papers looks at the data fusion problem from a kernel methods perspective. Kernel methods provide a framework for representing many different types of data structures, such as vectors, strings, and graphs, as shown in the wide range of applications addressed here: fMRI data, multichannel recordings, electrode readings and webpage data. The combination of the kernels to infer the shared structure of the data sources can be formulated in terms of an optimization problem. The two main themes emerging from this approach are kernel canonical correlation analysis and multiple kernel learning. The paper "Temporal kernel CCA and its application in multimodal neural data analysis", by Biessmann et al., proposes a generic framework for the analysis of simultaneously acquired signals of different spatial and temporal resolutions whose dependencies exhibit temporal dynamics. Their algorithm extends Kernel Canonical Correlation Analysis (kCCA) to take into account time-dependent information, where filters for each data source are optimized using the aggregate information over a window on the faster varying source. In the paper "Decomposing the tensor kernel support vector machine for neuroscience data with structured labels", Hardoon and Shawe-Taylor present a method to combine paired image stimulus and fMRI data, based on the tensor kernel SVM. After learning the decision function on image labels using the tensor kernel SVM, the weight matrix is decomposed into two components, each representing a data source, without accessing the original feature space. This new feature representation has the advantage of increased interpretability. The paper "Multi-view kernel construction", by de Sa et al., applies a kernel combination technique to integrate multiple data sources (as in multi-view learning) for solving a clustering problem. The kernels are induced by a multipartite graph built exploiting co-occurrences of patterns among the different views. The resulting spectral clustering algorithm is tested on synthetic and real-world datasets. In the paper "Composite kernel learning", Szafranski et al. develop the composite kernel learning approach, which is dedicated to learning the kernel when there is a known group structure among a set of candidate kernels. This framework applies to learning problems arising from a single data source when the input variables have a group structure, and it is also particularly well suited to the problem of learning from multiple sources.

A second group of papers studies multiple source learning from the viewpoint of domain adaptation and transfer learning. This includes problems such as multitask learning, where the knowledge of classifiers trained on related tasks is shared in order to improve performance, or problems where a learning algorithm is tested on data drawn from a distribution different from the one it was trained on. The paper "A co-classification approach to learning from multilingual corpora", by Amini and Goutte, investigates a multi-view text categorization problem in which the views are represented by translations of the same text in different languages. The work shows that training a categorizer for each language using a shared loss, and enforcing a similarity constraint among the produced outputs, effectively outperforms classifiers trained on monolingual data only. In the paper "Multi-domain learning by confidence-weighted parameter combination", Dredze et al. combine properties of domain adaptation and multi-task learning in order to: adapt multiple source domain classifiers to a new target domain, learn across multiple similar domains, and learn across a large number of disparate domains. The algorithms are based on Confidence-Weighted learning, an online algorithm for learning linear classifiers that incorporates confidence about each parameter into the update. Experiments are reported on sentiment classification and spam filtering datasets. Finally, the paper "A Theory of learning from different domains", by Ben-David et al., studies the domain adaptation problem from a theoretical standpoint. They analyze the conditions ensuring that a classifier trained on a source domain perform well on a different target domain. Moreover, they devise a principled method for combining a large amount of source training data with a small amount of target data. The main theoretical result bounds the risk of a classifier on the target domain in terms of its performance on the source domain and the divergence between the two domain distributions. This divergence can be estimated using unlabeled data from both domains.

The final group of papers explores the viewpoint of learning from multiple sources from a clustering and Bayesian perspective. This includes problems such as clustering for multi-source data, where results from different clustering algorithms or models are aggregated together, or problems based on learning Bayesian networks by incorporating auxiliary sources of information to improve performance. In the paper "Ensemble clustering using semidefinite programming with applications", Singh et al. propose a semi-definite programming relaxation for the combinatorial problem of aggregating the results of different clustering algorithms into a clustering that maximizes agreement in the input ensemble. They demonstrate that this can be achieved using a 2D string encoding rather than the common voting strategy. In the paper "Infinite factorization of multiple non-parametric views", Rogers et al. present a nested probability clustering model for multi-source data. The authors propose to extend the rationale of classical canonical correlation analysis into a flexible, generative and non-parametric clustering setting by introducing a novel non-parametric hierarchical mixture model. In the paper "Inductive transfer for learning Bayesian network", the authors Sucar and Morales present a method for aiding learning a Bayesian network by incorporating auxiliary sources of information. The methodology focuses on both structure learning, using conditional independence tests, and parameter learning, based on linear pooling for probability aggregation.

We conclude this editorial by thanking all reviewers for their careful and timely work, which helped us throughout the selection process and resulted in a set of papers of remarkable quality and interest.