Optimal transport for the classification of structured data

IRISA Vannes, OBELIX Team, France

PhD position to be filled in September 2017

Key-words structured data, machine learning, optimal transport, regularisation.

Context In numerous applications, data are not in a vectorial form but are rather structured: they are described by a set of parts that have relationships or constraints between them. For instance, an image can be represented at different scales by a hierarchical representation; time series have an intrinsic internal structure that must be taken into account. A consequence of the presence of structure in the data is that the classical machine learning techniques can not be directly applied. Two solutions are usually implemented to solve this problem:

- data are first transformed in order to bring back to a vectorial form (*e.g.* thanks to a feature extraction step in the time series context [1, 2] or by stacking all the nodes attributes when dealing with a tree [3]). Nevertheless, providing meaningful features is not straightforward;
- similarity measures between the different subparts are computed, then combined together (thanks to a convolutional kernel for instance [4, 5]. It usually suffers from high computational costs, preventing the method to be used in a large scale context.

In both cases, the solution is problem dependent, depending on the type of the structure, the type of features etc.

In the meantime, optimal transport (OT) [6] has emerged as a powerful tool to compute distances(a.k.a. Wasserstein or earth mover's distances) between empirical distribution of data, thanks to new computational schemes that make the transport computation tractable [7]. It has wide applications in computer vision, statistics, imaging and has been recently introduced in the machine learning community to efficiently solve classification or transfer learning problems [8]. The advantage of OT is that it can compare possibly high dimensional empirical probability measures, taking into account the geometry of the underlying metric spaces and dealing with discrete measures.

Scientific objectives and expected achievements The objective of the PhD is to define a new unified paradigm for classification of structured data by leveraging on the theory of optimal transport. Two directions will be explored:

- integration of the information carried out by the structure directly in the OT problem. In particular, the lead of defining a dedicated regularization term shall be explored;
- integration fo the structure directly inside the distance matrix between the data, building upon the notion of Gromov-Wasserstein distances for instance [9].

The aim is to produce an unified framework for many types of structured data, integrating problem specificities within the shape of the regularization or distances. A particular emphasis will be put on the developpement of efficient solutions, able to deal with large datasets.

From an application point of view, a particular attention will be given on remote sensing datasets. Indeed, hierarchical representations are more and more used to model the content of an image, providing

an effective framework for image classification. In addition, with the launch of new satellites, spatial and temporal resolution of remote sensing images has considerably increased, thus calling for the development of efficient algorithms.

Research environnement/Location The research will take place in the context of the IRISA laboratory (`http://www.irisa.fr/`), which is a joint research unit between CNRS, INRIA and several Universities and Engineering schools. IRISA conducts research in computer science, applied mathematics and signal and image processing. More specifically, the post-doc will be located in the OBELIX team (Environment Observation by Complex Imagery, `https://www-obelix.irisa.fr/`, which focuses on image analysis, machine learning and data mining, mostly for environmental data and remote sensing, and that is colocated between Vannes and Rennes (France). The PhD will take place in Vannes, a beautiful medieval city of medium size close to the sea (2h30 in train from Paris).
The PhD topic is at the interface of several research themes of the team: classification of hierarchical representation (with the PhD of Yanwei Qui), time series classification (PhD of Adeline Bailly) and optimal transport (with a post doc to be recruited and projects to be launched).

Technical aspects The applied part of the PhD will lead to development in Python. The candidate will build upon the python toolbox for optimal transport (POT: `https://github.com/rflamary/POT`) developed by members of the team among others. He/she will benefit from the expertise of the other members of the team, as well as ongoing collaborations with other academic partners on this subject.

Candidate profile Applicants are expected to be graduated in computer science and/or machine learning and/or signal & image processing and/or applied mathematics/statistics, and show an excellent academic profile. Beyond, good programming skills are expected.

Application procedure Send a resume to Laetitia Chapel (laetitia.chapel@irisa.fr), Nicolas Courty (nicolas.courty@irisa.fr) and Romain Tavenard (romain.tavenard@univ-rennes2.fr).

References

[1] M. G. Baydogan, G. Runger, and E. Tuv, "A bag-of-features framework to classify time series," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2796–2802, 2013.

[2] A. Bailly, S. Malinowski, R. Tavenard, L. Chapel, and T. Guyet, "Dense bag-of-temporal-sift-words for time series classification," in *Lecture Notes in Artificial Intelligence*. Springer, 2016.

[3] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," *IEEE Trans. Geosci. Remote Sens*, vol. 44, no. 9, pp. 2587–2600, 2006.

[4] D. Haussler, "Convolution kernels on discrete structures," Department of Computer Science, University of California at Santa Cruz, Tech. Rep., 1999.

[5] Y. Cui, L. Chapel, and S. Lefèvre, "Scalable bag of subpaths kernel for learning on hierarchical image representations and multi-source remote sensing data classification," *Remote sensing*, vol. 9, no. 3, 2017.

[6] C. Villani, *Optimal transport: old and new*. Springer Science, 2008.

[7] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Advances in Neural Information Processing Systems*, 2013, pp. 2292–2300.

[8] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Optimal transport for domain adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

[9] G. Peyré, M. Cuturi, and J. Solomon, "Gromov-wasserstein averaging of kernel and distance matrices," in *ICML*, vol. 16, 2016, pp. 2664–2672.