Multimodal fair learning

Cécile Capponi and Hachem Kadri

Qarma, pôle SD, LIS – Aix-Marseille University

SCIENTIFIC CONTEXT

On the one side machine learning (ML) can lead to unfair prediction models, i.e. models that rely on biases present in the datasets. Moreover, without any way to explain the predictions through interpretable models, that unfairness leads to untrustable models Barocas et al. (2019). The existence of these biases may be anticipated in the learning process, for example when certain so-called "sensitive" variables describe the data (e.g. ethical, gender), or when certain categories of examples or variables are under-represented, etc. In some insidious learning situations, unfairness is not expected but still arises: it mainly concerns the violation of the identically and independently distributed (i.i.d.). assumption of the learning algorithms. Various approaches exist to deal with non-equity situations in ML Mohri and Rostamizadeh (2008), but they never led to fully fair algorithms. A fundamental goal is to control the trade-off between the accuracy of the models and their capacity to eliminate those biases that lead to unfairness.

On the other side multi-modal learning (or more generally multi-view Capponi and Koço (2019); Sun et al. (2019)) considers that the input space $\mathcal{X}$ is made up of $v \geq 2$ modalities $\mathcal{X}^{(v)}$ describing examples, where some correlations are supposed to lie across them. For example, in any usual classification setting where $\mathcal{Y}$ is the label space, the dataset is $\mathcal{S} = \{(x_i \in \mathcal{X}, y_i \in \mathcal{Y})\}_{i=1}^n$, supposed to be i.i.d from the unknown joint distribution $\mathcal{D}_{\mathcal{X}\times\mathcal{Y}}$. Learning from $\mathcal{S}$ is finding the hypothesis $h : \mathcal{X} \to \mathcal{Y}$ that minimizes the true risk $R(h) = \int_{\mathcal{D}_{\mathcal{X}\times\mathcal{Y}}} \ell(h(x), y)$ (generalization goal, where ℓ is a loss function suited to the task at hand). In the multi-modal setting, we suppose that $\mathcal{X} = \mathcal{X}^{(1)} \times \cdots \times \mathcal{X}^{(m)}$ with $\forall v_1 \neq v_2 \in [1, v], \mathcal{D}_{\mathcal{X}^{v_1}\times\mathcal{Y}} \neq \mathcal{D}_{\mathcal{X}^{v_2}\times\mathcal{Y}}$ which implies that the Bayes error in each modality might be different. Another assumption in multi-modal learning is that modalities are not independent. Multi-modal learning is of interest when one modality of data observation is not sufficient on its own to carry out a learning task with good generalization properties, and/or when each modality carries its own information, sometimes semantically distant from that carried by the other modalities. Research works in multi-modal learning have grown internationally over the last decade Sun et al. (2019), within several learning frameworks, from (semi-)supervised to representation learning.

Focus of the thesis. Research about the potential benefits of multi-modal learning in terms of bias reduction and interpretability, and more generally of fairness considerations, are still at their early stage. This thesis aims at highlighting that multi-modal aspects would help to solve crucial drawbacks in tackling unfair learning, as long as cross-modalities aspects are considered, by exploring their complementarity. The multi-modal setting could improve mono-modal methods to avoid unfair models and to enhance cross-interpretability of them, once methods exploit cross-modalities. As far as we know, except few ad-hoc applied solutions, this work is fully exploratory.

PHD PROJECT

This PhD thesis will study cross-modal fair learning within the framework of kernel methods. The work is mainly organized around two main axis.

Selection of sources of unfairness

Needless to detail the fact that sources of unfairness are numerous. Only two categories of them will be considered in this thesis: biases in data that involve groups of attributes (inter- and extra-modalities), and imbalanced observations. Both need to be defined in a cross-modal context.

1. Biases in the data usually lead to unfair models, and often rely on some groups of "sensitive" variables describing the examples. Among such biases, a major concern of fairness is the demographic parity Calders et al. (2009); Feldman et al. (2015) which states that some sensitive variables (e.g. gender or ethnicity) should not influence a fair model. In other words, a prediction model must not reveal information about these sensitive variables. Demographic parity was generalized by equality of opportunity Hardt et al. (2016), which will be tackled in this thesis.

2. Imbalanced data is also a source of unfairness. Among the various imbalanced scenarii, the thesis will focus on the case where observed samples are not independently and identically distributed (i.i.d.) Gagnon-Bartsch and Shem-Tov (2019), which affects usual generalization bounds in statistical learning theory Mohri and Rostamizadeh (2008). For example in genomics, some genes are more explored/observed than some others because of their impact in health.

Starting from the mono-modal existing definitions of these sources of unfairness, it is necessary to accurately define their multi-modal perspective, especially when the modalities can be correlated. During that selection of unfair sources in learning, we might discover other theoretically important biases: they may lead to cross-modality definitions as well as the two first targeted.

Algorithms and theoretical analysis of selected sources of unfairness wrt learning tasks

One major concern in fair learning is to be able to maximize fairness without affecting the accuracy of the models too much: a trade-off between accuracy and fairness must be quantified so optimized (see e.g. Zafar et al. (2019)). To do so, this thesis is intended to define, in a multi-modal setting, metrics/losses for evaluating unfair learning within the above sources of unfairness. These definitions will be inspired by mono-modal approaches that have to be leveraged to the cross-modal setting. This step is of importance, because once formalized, these ways to deal with fairness in multi-modal learning will be the starting points for deriving relevant and theoretically founded algorithms for processing them before or during the learning process.

The cross-modal fair metrics that will be defined will be used to design fair learning algorithms. The focus here will be on building fairer ML systems by adapting the optimization scheme to catch the most useful information while avoiding some known biases Tan et al. (2020). Moreover, above new loss functions and regularization terms (for optimizing multi-modal metrics) are related to focused fairness aspects: they will be added to the defined optimization problems. It will thus require new constraints on solvers to be tackled for making computable the resulting objective functions. From a computational point of view, greedy methods and reweighting approaches, which can be relevant for class-imbalance source of unfairness ⋆Capponi and Koço (2019), as well as multi-modal kernel approximations Huusari et al. (2018) will be studied.

The generalization capabilities of the proposed learning algorithms will be studied during the thesis. For that purpose, Rademacher complexity, with useful generalization bounds, is actually a promising setting for cross-modal fair learning, because (i) it is well-suited for multiple kernel-based approaches ⋆Huusari et al. (2018)Cao et al. (2016), and (ii) it has been largely studied with regard to different types of biases (the assumption that a dataset may not be i.i.d. Mohri and

Rostamizadeh (2008), the demographic parity Tan et al. (2020), etc.). In order to deal with the non i.i.d. assumption in the multi-modal setting within the kernel theory, we will steadily relax that assumption in each modality, as it pertains to be recovered through transformations of all modalities according to optimization constraints to be defined. More specifically, we will study some kernel-based properties when considering fairness losses/metrics previously defined. As a result, we expect to derive relevant cross-modal generalization bounds.

Experimental validation

The designed algorithms will have to be experimentally validated through artificial and benchmarked datasets. Datasets in biology and neuro-imaging will be available by the end of 2023, and will be used to assess the performance of the proposed fair learning algorithms. This will be performed in a collaboration with INT [1] and CENTURI [2].

ACADEMIC CONTEXT AND REQUIRED SKILLS

The thesis will be driven in the team QARMA, in the Data Science pole of the Computer and Systems Lab. at Aix-Marseille University, France. Teaching duties are possible, in addition to the proposed salary.

The position is opened to Graduated Students in Computer Science or Applied Mathematics, ideally both. Strong programming skills are required, as well as knowledge in Machine Learning, Statistics and Linear Algebra.

REFERENCES

Barocas, S., Hardt, M., and Narayanan, A. (2019). *Fairness and Machine Learning*. fairmlbook.org.

Calders, T., Kamiran, F., and Pechenizkiy, M. (2009). Building classifiers with independency constraints. In *IEEE ICDM*.

Cao, B., Zhou, H., Li, G., and Yu, P. S. (2016). Multi-View Machines. In *ACM WSDM*, WSDM '16, pages 427—436.

Capponi, C. and Koço, S. (2019). Learning from Imbalanced Datasets with Cross-View Cooperation-Based Ensemble Methods. In Springer, editor, *Linking and Mining Heterogeneous and Multi-view Data*.

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., and Venkatasubramanian, S. (2015). Certifying and removing disparate impact. In *ACM SIGKDD Int. Conf. Knowl.*

Gagnon-Bartsch, J. and Shem-Tov, Y. (2019). The classification permutation test: A flexible approach to testing for covariate imbalance in observational studies. *The Annals of Applied Statistics*, 13(3):1464–1483.

Hardt, M., Price, E., and Srebro, N. (2016). Equality of opportunity in supervised learning. volume 29.

Huusari, R., Kadri, H., and Capponi, C. (2018). Multi-view Metric Learning in Vector-valued Kernel Spaces. In *AISTATS*, volume 84 of *Proceedings of Machine Learning Research*, pages 415–424. PMLR.

[1] https://www.int.univ-amu.fr/

[2] https://centuri-livingsystems.org/

Mohri, M. and Rostamizadeh, A. (2008). Rademacher Complexity Bounds for Non-I.I.D. Processes. In Koller, D., Schuurmans, D., Bengio, Y., and Bottou, L., editors, *NeurIPS*. Curran Associates, Inc.

Sun, S., Mao, L., Ziang, D., and Lidan, W. (2019). *Multiview machine learning*. Springer.

Tan, Z., Yeom, S., Fredrikson, M., and Talwalkar, A. (2020). Learning Fair Representations for Kernel Models. In *AISTATS*.

Zafar, M. B., Valera, I., Gomez-Rodriguez, M., and Gummadi, K. P. (2019). Fairness Constraints: A Flexible Approach for Fair Classification. *J Mach Learn Res*, 20(75).