Research Statement

Antoine Ledent
School of Computing and Information Systems, Singapore Management University
Tel: (65) 6826-4914; Email: aledent@smu.edu.sg
27th of December 2022

Background

My research passion lies in **Machine Learning** (ML), a subfield of Artificial Intelligence (AI) which studies algorithms that learn from data without being explicitly programmed. I have worked with both Neural Networks (NN) and **Matrix Completion** (MC) models, proving statistical guarantees as well as proposing new algorithmic solutions for problems ranging from Computer Vision (CV) to **Recommender Systems** (RS). Having moved to the field of AI from a background in mathematics, I see a natural connection between the statistical theory of ML models and algorithm design: a starting point in much of my research is **Statistical Learning Theory** (SLT), which broadly concerns the estimation of the number of samples required to train an ML model effectively. In short, I am interested in understanding why ML algorithms work, and in using the insights derived to obtain gains in interpretability and new theoretically motivated algorithms. In addition, statistical guarantees can often help practitioners choose between several architectures for a given problem. In terms of specific classes of ML models, much of both my theoretical and practical research has recently focused on Matrix Completion and Recommender Systems.

Recommender Systems (RS) aim to guide users in a personalized way to items maximizing user satisfaction while providing business value (e.g. increased revenue, engagement) for the recommendation provider. For instance, streaming companies such as Netflix or Douban recommend movies to customers based on previous feedback (clicks, "likes", etc.), allowing users to discover and watch movies they are more likely to enjoy and which would otherwise have been difficult to single out from the multitude of possibilities. In the simplest case, the recommendation task consists in predicting the user-item ratings based only on the existing observed ratings. This task can be viewed as completing the missing entries of a partially observed matrix, a simple and elegant problem with a surprisingly rich (and yet not quite complete) mathematical theory. In practice, real-world recommender systems rely on much more than the explicit feedback contained in the observed ratings. The implicit feedback implied by the subset of items which each user has chosen to interact with arguably contains just as much information. In addition, in the current paradigm of Web 2.0, users contribute content and interact with each other in various ways in blogs, Twitter, social media websites and other platforms, providing a wealth of heterogeneous information which is relevant to the recommendation task. Incorporating such information into well-principled RS and MC models presents an unparalleled source of opportunities and challenges in both theory and practice.

I also have a strong interest in the theoretical properties of deep neural networks (NN). In general, classic methods struggle to fully explain why such models perform so well even when the number of parameters is much more than the number of samples: it is increasingly clear that underlying structure in the data has as much effect as algorithm construction on the statistical properties and behavior of deep learning models. For instance, Convolutional Neural Networks (CNN) exhibit especially strong success in image recognition due to the presence of repetitive patterns in the data. Defining and quantifying such effects is especially difficult due to the surprising generality of the phenomenon for many types of ‘natural’ data. The Neural Tangent Kernel (NTK) provides both partial answers to these theoretical questions and huge application potential.

My recent research can be categorized into three interconnected directions: Statistical Learning Theory, Matrix Completion/Recommender Systems, and other Applications (e.g. CV). During my PhD, I have also worked on Stochastic Analysis, the study of random processes in continuous time.

Research Areas

<u>On recommender systems, matrix completion and low-rank methods</u>

In classic MC, a typical assumption is that the ground truth matrix has an approximate low-rank structure: there is a small (unknown) number of unobserved hidden user and item features that determine the rating via a simple inner product function. This type of low-rank constraint is ubiquitous in ML applications beyond MC and RS.

The Inductive Matrix Completion (IMC) model is a generalization of the basic MC model where side information is provided for the users and the items in the form of a set of feature vectors corresponding to each user and to each item. The assumption is that users with similar feature vectors will also exhibit similar rating behaviors. Thus, this mathematical model represents a first step towards incorporating user information from the web into RS, assuming only that such information has been preprocessed into feature vectors. Interestingly, this similarity assumption coexists in parallel with the low-rank assumption. Until our paper [NeurIPS’21a], the theoretical analysis of MC and IMC relied on nearly distinct approaches, and sample complexity guarantees for IMC did not take the low-rank constraint into account at all.

In our paper [NeurIPS’21a] we have proved several generalization bounds for the IMC model and introduced an improved regularization strategy for IMC. The first contribution is a bound which incorporates both the low-rank constraint and the side information: ignoring logarithmic terms, the sample complexity is $\tilde{O}\left(d^{3/2}\sqrt{r}\right)$ in (the distribution-free case) where d is the dimension of the side information, and r is the rank of the ground truth matrix. Thus, the bound is non-trivial when the side information approaches the identity, unlike existing IMC bounds in this distribution-free setting. The second contribution is both theoretical and algorithmic: we have introduced a modification of the regularization strategy which counters the effects of non-uniformity. The strategy involves a data-dependent geometric transformation of predictors before the evaluation of the regularization term. Thus, our model relies on the empirical distribution of the observed entries to adapt the behavior of the regularization term. The solution to this modified optimization problem exhibits a sample complexity which is as good as in the uniform sampling case. Although the algorithm is theoretically motivated, it leads to gains in accuracy in the experiments we ran on classic RS datasets such as MovieLens, LastFM and Douban.

In [AAAI'23], we performed a more detailed theoretical analysis of the sample complexity of IMC where the noise in the observations is assumed to be small. So-called 'exact recovery results' show that if the entries of a matrix are observed exactly (from a uniform distribution), there is a high probability of recovering the full ground truth matrix exactly, so long as the number of samples exceeds a given threshold. This is in contrast to 'approximate recovery results' such as those of [NeurIPS'21a] which prove that the error decays at a certain rate (typically $N^{-1/2}$ where N denotes the sample size). In [AAAI'23], we bridge this gap between approximate and exact recovery for IMC by observing and characterizing a threshold phenomenon: the error drops sharply past the exact recovery threshold and continues to decay at the classic $N^{-1/2}$ rate afterwards, with a multiplicative constant proportional to the variance of the noise.

[NeurIPS'21b] deals with density estimation under the assumption that the density is approximately low-rank, i.e. the sampling distribution is well approximated by a mixture of processes each of which has independent components. We obtained sample complexity guarantees for a low-rank histogram model with polynomial dependence on the ambient dimension, whilst traditional guarantees for the standard histogram suffer from the curse of dimensionality. This work is closer to traditional mathematical statistics and not directly related to RSs, but it relies on matrix and tensor decompositions that are commonly used in RS and even other ML applications.

In [TNNLS'21], we introduced a novel matrix completion method for the situation where side information is available in the form of a partition of users and items into distinct categories, or 'communities' according to established terminology. For instance, movies might be categorized by genre (romance, comedy, action, etc.), and items might be divided by gender or nationality. Our model is a sum of several components, including a component for community effects and a component that corresponds to purely low-rank effects independent of the community structure. Thus, our model can disentangle community behavior from individual behavior. Differently from existing works, both community and individual behavior coexist and are combined additively. We also proposed an efficient and scalable optimization implementation based on iterative imputation which takes the 'sparse plus low-rank' structure of the iterates into account. In the spinoff paper [PMLR'21] we extended this paradigm to the situation where the communities must be learned by the model.

In [TNNLS'23], we have explored matrix completion in the presence of a user-level reweighting applied directly to the loss function, and its effects on the nuclear norm regularizer's generalization performance and ability to yield low-rank solutions. As it turns out, a weighted trace norm regularization term analogous to that found in [NeurIPS'21] is required to preserve favorable sample complexity. We also extensively explored the effects of this correction term on both real and synthetic data, demonstrating the superior generalization performance of the model which includes our correction term. In the real data experiments, the reweighting of the loss was applied based on an estimate of the 'reliability' of the user, derived from the level of comprehensiveness in their overall reviews.

In [RecSys'23], a recent project of my Visiting Research Student, we constructed a joint model with both a low rank matrix completion component and a graph neural network component with dual aims: the matrix completion module estimates the ratings, whilst the graph neural networks component estimates the uncertainty in the ratings. Both modules are trained jointly, with a loss function inspired from the literature on heteroskedastic regression. Extensive experiments demonstrate our method's ability to provide increased performance compared to a single low-rank module and increased interpretability compared to a single graph neural network module. In addition, our model can detect anomalous ratings both in the training set and at test time.

Statistical Learning Theory for neural networks and kernel methods

I have proved generalization bounds for various machine learning models, including CNNs and kernel methods in various scenarios matching popular ML models. I have focused on the so-called "norm-based" approach, which is better suited to the near-infinite parameter spaces ubiquitous in state-of-the-art ML models. In addition, I have paid especially strong attention to applications with extremely large output spaces: extreme multi-class classification, extreme multi-label learning and structured output prediction. The settings are particularly adapted to applications such as handwriting recognition of words, document tagging and Natural Language Processing (NLP) respectively. In all cases, a central concern is to obtain generalization bounds with a dependency on the number of classes that is as mild as possible.

In [AAAI'21a], we have proved norm-based bounds for CNNs with two main improvements over the state of the art: (1) they take the weight sharing into account, and (2) they have near optimal dependence on the number of classes. Up to logarithmic terms, the bounds do not change much if an identical set of trainable weights is applied to many different areas of an input image. This work was thus the first one to provide statistical guarantees for neural networks which take the translational invariance of CNNs into account in a norm-based context. In addition, in a class-balanced scenario, the dependence on the number of classes is linear (up to log factors) when expressed in terms of sample complexity, which is clearly optimal (as one needs to observe at least one sample from each class in that case). There are plenty of other refined bounds in the paper, including insights on the effect of local Lipschitzness and appropriate batch normalization on generalization performance.

In [AAAI'21b], the multi-class aspect of the above work was extended to the multi-label case, where each input can be associated with an arbitrary number of labels. In [IJCAI'21a], we pushed similar techniques to the extreme, providing bounds with optimal dependence on the dictionary size for structured output prediction for both kernel methods and simple NN architectures. In [NeurIPS'20], we provided model-agnostic generalization bounds for the Stochastic Gradient Descent algorithm.

In Computer Vision (CV)

In video tracking, target appearance can vary over time. For instance, a person may be slowly turning, sometimes facing the camera and sometimes having their back to it. Furthermore, since models must rely on the previous positions of the target (which can only be estimated by the model and are not labeled), errors can propagate and amplify through time. In [AAAI'21c], we proposed a video tracker based on a Siamese architecture that adapts to such situations by constructing an aggregate feature representation based on a representative sample of previous frames. The model also incorporates an estimation of uncertainty which is then used by different model components to improve robustness. In [IJCAI'21b] we investigated various notions of interpretability in NN models for computer vision: for instance, some neurons can be shown to learn concepts corresponding to scenes or objects or colors, or to be scouting for certain specific patterns. The main novelty in our work is that we proposed several regularization strategies that induce greater interpretability during training.

In Stochastic Analysis (PhD)

During my PhD, I proved upper bounds for the densities of low-dimensional projections of high-dimensional Stochastic Differential Equations (SDE) under conditions on the driving vector fields which are analogous to ellipticity or hypoellipticity. The idea is that no matter how complex a system of SDEs is, 'low-dimensional snapshots' of its solution still have 'almost as well-behaved' densities as those of solutions to SDEs directly constructed in the low-dimensional target space. Although the thesis itself is technical due to the great generality of the results, specific systems of SDEs have been used to model both biological neural networks and financial markets.

Selected Publications and Outputs

On Recommender Systems, Matrix Completion and low-rank methods:

[NeurIPS'21a] LEDENT, Antoine; ALVES, Rodrigo; LEI, Yunwen and KLOFT (2021). Fine-grained generalisation analysis of inductive matrix completion. Advances in Neural Information Processing Systems (NeurIPS) 34, 2021. 25540—25552.

[TNNLS'21] LEDENT, Antoine; ALVES, Rodrigo; and KLOFT, Marius (2021). Orthogonal Inductive Matrix Completion. IEEE Transactions of Neural Networks and Learning Systems. Pages : 1-12, DOI : 10.1109/TNNLS.2021.3106155.

[AAAI'23] LEDENT, Antoine; ALVES, Rodrigo; LEI, Yunwen; GUERMEUR, Yann; and KLOFT, Marius (2023). Generalization Bounds for Inductive Matrix Completion in Low-Noise Regimes. Proceedings of the AAAI Conference, to appear.

[NeurIPS'21b] VANDERMEULEN, Rob; and LEDENT, Antoine (2021). Beyond Smoothness : Incorporating Low-Rank Analysis into Nonparametric Density Estimation. Advances in Neural Information Processing Systems (NeurIPS) 34, 12180—12193.

[RecSys'21] ALVES, Rodrigo; LEDENT, Antoine; and KLOFT, Marius (2021). Burst-induced Multi-Armed Bandit for Learning Recommendation. Recommender Systems Conference (RecSys) 2021, 292-301. DOI: https://doi.org/10.1145/3460231.3474250

[PMLR'21] ALVES, Rodrigo; LEDENT, Antoine; ASSUNÇÃO, Renato; and KLOFT, Marius (2021). An Empirical Study of the Discreteness Prior in Low-Rank Matrix Completion. NeurIPS 2020 Preregistration Workshop. Proceedings of Machine Learning Research (PMLR) 148:111-125, 2021.

[TNNLS'23] ALVES, Rodrigo; LEDENT, Antoine; and KLOFT, Marius. Uncertainty-Adjusted Recommendation via Matrix Factorization With Weighted Losses. IEEE Transactions of Neural Networks and Learning Systems. Pages: 1-14. DOI: 10.1109/TNNLS.2023.3288769.

[RecSys'23] KASALICKY, Petr; LEDENT, Antoine; and ALVES, Rodrigo. Uncertainty-adjusted Inductive Matrix Completion with Graph Neural Networks. Proceedings of the 17th ACM Conference on Recommender Systems, 1169-1174 (Late Breaking Results).

On the statistical theory of neural networks and kernel methods:

[AAAI'21a] LEDENT, Antoine; MUSTAFA, Waleed; LEI, Yunwen; and KLOFT, Marius (2021). Norm-based generalisation bounds for convolutional neural networks, by Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 35(9), 8279-8287.

[AAAI'21b] Liang WU, Antoine LEDENT, Yunwen LEI and Marius KLOFT (2021). Fine-grained Generalization Analysis of Vector-valued Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12): 10338-10346.

[IJCAI'21a] MUSTAFA, Waleed; LEI, Yunwen; LEDENT, Antoine; and KLOFT, Marius (2021). Fine-grained Analysis of Structured Output Prediction. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21. 2841-2847.

[NeurIPS'20] LEI, Yunwen; LEDENT, Antoine; and KLOFT, Marius (2020). Sharper generalisation bounds for pairwise learning. Advances in Neural Information Processing Systems (NeurIPS) 33, 21236–21246.

In Computer Vision:

[AAAI'21c] ZHOU, Lijun; LEDENT, Antoine; HU, Qintao; LIU, Ting; ZHANG, Jianlin; and KLOFT, Marius (2021). Model Uncertainty Guides Visual Object Tracking. Proceedings of the AAAI Conference on Artificial Intelligence, 35(4): 3581-3589.

[IJCAI'21b] VARSHNEYA, Saurabh; LEDENT, Antoine; VANDERMEULEN , Rob; LEI, Yunwen; ENDERS Matthias; BORTH Damian; and KLOFT, Marius (2021). Learning Interpretable Concept Groups in CNNs. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, 1061-1067.

In Stochastic Analysis:

[PhD'17] LEDENT, Antoine. Sharper Kusuoka-Stroock type bounds for densities related to low-dimensional projections of high dimensional SDEs. PhD Thesis, University of Luxembourg.