Tutorial: Interactive Adaptive Learning

Mirko Bunse[1], Georg Krempl[2], Alaa Tharwat[3] and Amal Saadallah[1]

[1] *TU Dortmund University, Germany*
[2] *Utrecht University, Netherlands*
[3] *Hochschule Bielefeld, Germany*

Abstract
We summarize the contents of the tutorial we present as a part of the 7th Interactive Adaptive Learning workshop. This workshop is co-located with the ECML-PKDD conference, where it takes place on September 22nd, 2023 in Turin, Italy.

Keywords
active learning, active class selection, active feature acquisition, meta-learning

Introduction

Interactive adaptive learning comprises methods that improve the overall life-cycle of machine learning models, including interactions with human supervisors, interactions with other processing systems, and adaptations to different forms of data that become available at different points in time. Most importantly, interactive adaptive learning is concerned with different forms of *active learning* (AL) [1], a research area with many facets. We cover the most important facets of AL in this tutorial, before discussing recent progress in a workshop session.

This tutorial is structured into five parts, which we detail in the following sections. Their titles and presenters are as follows:

1. Foundations of Active Learning (A. Tharwat & G. Krempl)
2. Beyond Pool-Based Scenarios (G. Krempl & A. Tharwat)
3. Beyond Active Labeling (M. Bunse)
4. Towards Explainable Active Learning using Meta-Learning (A. Saadallah)
5. Practical Challenges and New Research Directions (A. Tharwat & G. Krempl)

Part 1—Foundations of Active Learning

Although a huge amount of unlabeled data has been collected recently, this data is still useless for developing learning algorithms that require labeled data. However, collecting labeled data might require an expert annotator, might be expensive (e.g., when a series of processes must

IAL@ECML-PKDD'23: 7th Intl. Worksh. & Tutorial on Interactive Adaptive Learning, Sep. 22nd, 2023, Torino, Italy
mirko.bunse@cs.tu-dortmund.de (M. Bunse); g.m.krempl@uu.nl (G. Krempl); alaa.othman@fh-bielefeld.de (A. Tharwat); amal.saadallah@cs.tu-dortmund.de (A. Saadallah)
0000-0002-5515-6278 (M. Bunse); 0000-0002-4153-2594 (G. Krempl); 0000-0003-2976-7574 (A. Saadallah)

be performed in laboratories to generate the label), might be time-consuming (e.g., when long documents need to be annotated), or might be difficult for several other reasons. In this case, the active learning technique provides a solution by querying a small set of informative and representative points from the available unlabeled points to annotate them. This selected set of points represents the training data for learning a model and yields promising results.

Active Labeling and Semi-Supervised Learning In the semi-supervised technique, the aim is to leverage a combination of labeled and unlabeled data to enhance the model's performance. To do this, the unlabeled data is utilized to further improve the supervised models, which has been learned from the labeled data. In contrast, AL aims to optimize the process of data labeling by querying the most informative samples from an unlabeled pool. So, the semi-supervised learning and AL could be solutions when the labeled data is limited, but the semi-supervised technique searches for the most certain points while ALs query the most uncertain ones [2].

Scenarios of Active Labeling There are three main scenarios of active labeling. First, the scenario of membership query synthesis, where AL creates synthetic instances within the space and then queries these new instances. Since there is no processing of unlabeled data, this scenario is fast compared to the other scenarios and suitable for finite problem domains [1]. The main limitation here is that the artificially created instances may not have a meaningful label. Second, in the stream-based selective sampling scenario, the unlabeled data instances are drawn iteratively, one at a time, and the learning model makes a decision whether to query the unlabeled point based on its information content. The third, last and most well-known AL scenario is the pool-based scenario, where a query strategy is used to evaluate the informativeness of some/all instances in the pool of unlabeled data to query the labels of one or more instances [2].

Other Forms of Active Learning There are different research directions of active learning. The best known is active labeling, where AL searches a pool of unlabeled instances to select the most informative and representative points to be labeled and added to the training data. Another direction is active feature selection, where the environment is searched for unobserved features to improve performance at the time of evaluation. The active class selection direction aims to actively select a class and ask the annotator to provide a sample/instance for that class to optimise classification performance with a small number of queries.

Part 2—Beyond Pool-Based Scenarios

In real-world applications, there are situations that go beyond the classical setup of the pool-based scenario, leading to more challenging and interesting AL scenarios:

Stream-based AL Here, data arrives in a continuous stream (e.g., social media posts). Therefore, the selection of instances for labeling becomes more dynamic, requiring the model to make decisions in real time as the stream evolves [2].

Batch-based AL This involves selecting a batch of instances for labeling at the same time, rather than selecting a single instance, which is relevant when labeling can be done in groups. However, the challenge here is to select a diverse and informative batch that effectively improves the performance of the model [2].

Semi-supervised AL Here, both labeled and unlabeled data are available, and AL strategies aim to use both data sources to improve model performance [3].

Transductive and Inductive AL The goal of these two different strategies is to find the most informative examples for labeling. The main difference between them lies in their goals and focus, where transductive AL aims to improve the model's performance on the current set of unlabeled instances, while inductive AL focuses on improving the model's generalization to new, unseen instances. Thus, the choice between them depends on whether the goal is to achieve immediate accuracy or broader generalization [4].

Part 3—Beyond Active Labeling

The term *active learning* is often identified with an *active labeling* of unlabeled instances from a pool or a stream. This understanding is limited by the assumptions *i)* that labels can be assigned in hindsight and to arbitrary instances and *ii)* that labels are the only relevant cost factor during data acquisition; use cases of AL might violate these assumptions, thereby rendering an active acquisition of labels infeasible. In the third part of the tutorial, we therefore broaden the idea of AL to settings where other parts of the data are queried, e.g., settings where individual features or complete instances have to be acquired instead of labels.

Active Class Selection One such setting is *active class selection* [5], where some class-conditional data generator $g : \mathcal{Y} \to \mathcal{X}$ is assumed. Strategies for active class selection query this data generator in terms of the class proportions which are to be generated during the next acquisition round. The generator then produces a batch of new data instances according to these proportions. Data generators of this kind appear in use cases as diverse as astrophysics [6], brain computer interaction [7, 8], and gas sensor arrays [5]. They are in contrast to the oracles $o : \mathcal{X} \to \mathcal{Y}$ that are required for an active labeling of pre-existing instances.

Most strategies for the active selection of classes [5, 9] consist of ad-hoc heuristics. A recent line of research, however, evolves around theoretical analyses of the implications that active class selection has on supervised learning. This line of research, put forward by the presenters of this tutorial, begins with a study on consistency [10] and evolves into a PAC analysis [6, 11]. The theoretical results therein give rise to an acquisition strategy [12] which leverages the prior knowledge of a practitioner instead of relying on heuristics.

Active Feature Acquisition Another setting of AL is *active feature acquisition* [13], where instances have missing features that can be queried individually. For this purpose, a feature value oracle is needed. Acquisition strategies have to choose missing feature values of specific instances to acquire—a problem that might occur at training time or at testing time of a supervised model.

In our tutorial, we introduce the problem of active feature acquisition and we revisit some of the most important strategies [14, 15, 16]. All in all, the third part of this tutorial demonstrates that AL comprises several important problems beyond the active acquisition of labels.

Part 4—Towards Explainable Active Learning using Meta-Learning

Explainable AI focuses on developing machine learning models that provide transparent and interpretable explanations for their predictions. A lack of interpretability can be a significant drawback in safety-critical applications, like healthcare, finance, and autonomous systems. Meta-learning aids in providing interpretable explanations in AL.

Meta-Learning The majority of popular AL approaches relies on heuristics, none of which clearly outperforms the others in all cases [17]. The primary objective of meta-active learning (meta-AL) is to develop a data-driven approach to AL, which is capable of selecting the optimal set of unlabeled items for labeling. The fundamental idea is to train a regressor that predicts the informativeness of a candidate sample in a specific learning state [18]. Recent examples include bandit algorithms [19] and reinforcement learning techniques [20] which, however, are limited to combining pre-existing hand-designed heuristics. This limitation is lifted in "Learning Active Learning" [17], a method which predicts the reduction in generalization error that is caused by labeling an instance. This method outperforms competing methods at a relatively low computational cost.

Explainability Explainable AL using meta-learning [18] involves leveraging the benefits of both AL and meta-learning techniques to obtain more informative labels while ensuring the interpretability of the AL process. This goal can be achieved through several measures: explainable model architectures, interpretable meta-models, explainable active sample selection, attention mechanisms, post-hoc explanation techniques, regularization towards explainable outcomes, and human-in-the-loop feedback.

By combining these strategies, researchers can develop a system that not only achieves high accuracy through AL and meta-learning but also provides transparent, interpretable, and trustworthy explanations for its predictions. The goal is to strike a balance between accuracy and interpretability, making the AL process more trustworthy and usable in real-world applications where human understanding is essential.

Part 5—Practical Challenges and New Research Directions

We conclude our tutorial with a discussion of common challenges that appear in real-world AL applications. We also propose several research directions that have not received a lot of attention yet.

Practical Challenges

Imbalanced Data This challenge stems from the difficulty of learning from minority classes. The problems of unequal labelling costs or unequal misclassification costs are also similar; these problems happen when the classification costs are different between classes [21]. Therefore, ALs should improve their exploration capability to scan the whole space including minority class subspaces [3, 22].

Diversity of Samples Non-representative selections of points (e.g., when samples are concentrated in a small region) reduce generalization capabilities and lead to biased models. Hence, a consideration of diversity has to compensate for the lack of exploration in uncertainty methods [23].

Outliers Outliers can appear as representing an informative region. Instead, however, they deviate AL techniques from exploring truly uncertain regions [24].

High Dimensionality High-dimensional spaces challenge AL not only because they challenge the learning method, but also because of the computational time required [25].

Crowdsourcing Non-expert annotators have the potential of cheaply labeling large amounts of data. However, their labels might be noisy, leading to negative effects that can be more harmful than having only small training sets [26, 27].

Small Query Budgets With a small budget, AL may not be able to explore the entire space perfectly. This situation can lead to sub-optimal performances [28].

Stopping Criteria AL continues querying until a termination condition is met [2]. Termination conditions can be based on sampling complexity or on fixed budgets [29].

New Research Directions

Deep AL Deep learning achieves impressive results, especially with large training sets. However, collecting these sets is often challenging [30]. Deep AL promises reasonable performance with small but highly informative training sets [31].

AL with Evolutionary Algorithms The most expensive part of evolutionary optimization algorithms is the fitness evaluation. Here, AL can be used to select the most informative points to build a surrogate model that simulates the original fitness function [32].

AL with Simulation In large-scale simulation models, the calibration of large numbers of parameters is expensive. AL can be used to reduce the number of simulations required by finding the most informative regions within the space and the most important parameters [33].

AL with Design of Experiments Design of experiments allows researchers to optimize processes by identifying important factors and drawing reliable conclusions with a minimum number of trials [34]. Here, AL can reduce the number of experiments by finding and performing only the most informative ones.

Acknowledgments

The work of M.B. and A.S. was partly funded by the Federal Ministry of Education and Research of Germany and the state of North Rhine-Westphalia as part of the Lamarr Institute for Machine Learning and Artificial Intelligence. The work of A.T. was conducted within the framework of the project "SAIL: SustAInable Lifecycle of Intelligent SocioTechnical Systems" (grant no. NW21-059B). SAIL is receiving funding from the programme "Netzwerke 2021", an initiative of the Ministry of Culture and Science of the State of North Rhine-Westphalia. The sole responsibility for the content of this publication lies with the authors.

References

[1] B. Settles, Active learning literature survey, Technical Report 1648, University of Wisconsin-Madison Department of Computer Sciences, 2009.

[2] A. Tharwat, W. Schenck, A survey on active learning: State-of-the-art, practical challenges and research directions, Mathematics 11 (2023) 820.

[3] A. Tharwat, W. Schenck, Balancing exploration and exploitation: A novel active learner for imbalanced data, Knowledge-Based Systems 210 (2020) 106500.

[4] G. Ciano, A. Rossi, M. Bianchini, F. Scarselli, On inductive–transductive learning with graph neural networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (2021) 758–769.

[5] R. Lomasky, C. E. Brodley, M. Aernecke, D. Walt, M. A. Friedl, Active class selection, in: Europ. Conf. on Mach. Learn., volume 4701 of *Lecture Notes in Comput. Sci.*, Springer, 2007, pp. 640–647. doi:10.1007/978-3-540-74958-5_63.

[6] M. Bunse, K. Morik, Certification of model robustness in active class selection, in: Europ. Conf. on Mach. Learn. and Knowl. Discov. in Databases, Springer, 2021, pp. 266–281. doi:10.1007/978-3-030-86520-7_17.

[7] D. Wu, B. J. Lance, T. D. Parsons, Collaborative filtering for brain-computer interaction using transfer learning and active class selection, PloS one 8 (2013). doi:10.1371/journal.pone.0056624.

[8] I. Hossain, A. Khosravi, S. Nahavandi, Weighted informative inverse active class selection for motor imagery brain computer interface, in: Canad. Conf. on Electr. and Comput. Engin., IEEE, 2017, pp. 1–5. doi:10.1109/CCECE.2017.7946613.

[9] D. Kottke, G. Krempl, M. Stecklina, C. S. von Rekowski, T. Sabsch, T. P. Minh, M. Deliano, M. Spiliopoulou, B. Sick, Probabilistic active learning for active class selection, in: Proc. of the NeurIPS Worksh. on the Future of Interact. Learn. Mach., 2016.

[10] M. Bunse, D. Weichert, A. Kister, K. Morik, Optimal probabilistic classification in active class selection, in: Int. Conf. on Data Mining, IEEE, 2020, pp. 942–947. doi:10.1109/ICDM50108.2020.00106.

[11] M. Senz, M. Bunse, K. Morik, Certifiable active class selection in multi-class classification, in: Worksh. on Interact. Adapt. Learn., CEUR Worksh. Proc., 2022, pp. 68–76.

[12] M. Bunse, K. Morik, Active class selection with uncertain deployment class proportions, in: Worksh. on Interact. Adapt. Learn., CEUR Worksh. Proc., 2021, pp. 70–79.

[13] M. Saar-Tsechansky, P. Melville, F. J. Provost, Active feature-value acquisition, Manag. Sci. 55 (2009) 664–684. doi:10.1287/mnsc.1080.0952.

[14] M. desJardins, J. MacGlashan, K. L. Wagstaff, Confidence-based feature acquisition to minimize training and test costs, in: SIAM Int. Conf. on Data Mining, SIAM, 2010, pp. 514–524. doi:10.1137/1.9781611972801.45.

[15] S. Huang, M. Xu, M. Xie, M. Sugiyama, G. Niu, S. Chen, Active feature acquisition with supervised matrix completion, in: Int. Conf. on Knowl. Discov. & Data Mining, ACM, 2018, pp. 1571–1579. doi:10.1145/3219819.3220084.

[16] Z. Zheng, B. Padmanabhan, On active learning for data acquisition, in: Int. Conf. on Data Mining, IEEE, 2002, pp. 562–569. doi:10.1109/ICDM.2002.1184002.

[17] K. Konyushkova, R. Sznitman, P. Fua, Learning active learning from data, Advances in neural information processing systems 30 (2017).

[18] Y. Taguchi, K. Kameyama, H. Hino, Active learning with interpretable predictor, in: 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019, pp. 1–8.

[19] K. Pang, M. Dong, Y. Wu, T. M. Hospedales, Dynamic ensemble active learning: A non-stationary bandit with expert advice, in: 2018 24th International Conference on Pattern Recognition (ICPR), IEEE, 2018, pp. 2269–2276.

[20] K. Pang, M. Dong, Y. Wu, T. Hospedales, Meta-learning transferable active learning policies by deep reinforcement learning, arXiv preprint arXiv:1806.04798 (2018).

[21] D. Kottke, A Holistic, Decision-Theoretic Framework for Pool-Based Active Learning, PhD thesis dissertation, Kassel University, 2021.

[22] A. Tharwat, W. Schenck, A novel low-query-budget active learner with pseudo-labels for imbalanced data, Mathematics 10 (2022) 1068.

[23] J. T. Ash, C. Zhang, A. Krishnamurthy, J. Langford, A. Agarwal, Deep batch active learning by diverse, uncertain gradient lower bounds, arXiv preprint arXiv:1906.03671 (2019).

[24] S. Karamcheti, R. Krishna, L. Fei-Fei, C. D. Manning, Mind your outliers! investigating the negative impact of outliers on active learning for visual question answering, arXiv preprint arXiv:2107.02331 (2021).

[25] T. Tran, T.-T. Do, I. Reid, G. Carneiro, Bayesian generative active deep learning, in: International Conference on Machine Learning, PMLR, 2019, pp. 6295–6304.

[26] Z. Shu, V. S. Sheng, J. Li, Learning from crowds with active learning and self-healing, Neural Computing and Applications 30 (2018) 2883–2894.

[27] A. Calma, Active Learning with Uncertain Annotators: Towards Dedicated Collaborative Interactive Learning, PhD thesis dissertation, Kassel University, 2020.

[28] Y. Zhang, P. Zhao, S. Niu, Q. Wu, J. Cao, J. Huang, M. Tan, Online adaptive asymmetric active learning with limited budgets, IEEE Transactions on Knowledge and Data Engineering 33 (2019) 2680–2692.

[29] V.-L. Nguyen, M. H. Shaker, E. Hüllermeier, How to measure uncertainty in uncertainty sampling for active learning, Machine Learning 111 (2022) 89–122.

[30] K. O. Lye, S. Mishra, D. Ray, P. Chandrashekar, Iterative surrogate model optimization (ismo): An active learning algorithm for pde constrained optimization with deep neural networks, Computer Methods in Applied Mechanics and Engineering 374 (2021) 113575.

[31] Y. Gal, R. Islam, Z. Ghahramani, Deep bayesian active learning with image data, in: International conference on machine learning, PMLR, 2017, pp. 1183–1192.

[32] N. Zemmal, N. Azizi, M. Sellami, S. Cheriguene, A. Ziani, A new hybrid system combining active learning and particle swarm optimisation for medical data classification, International Journal of Bio-Inspired Computation 18 (2021) 59–68.
[33] T. Lookman, P. V. Balachandran, D. Xue, R. Yuan, Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design, npj Computational Materials 5 (2019) 21.
[34] C.-T. Chen, G. X. Gu, Generative deep neural networks for inverse materials design using backpropagation and active learning, Advanced Science 7 (2020) 1902607.