Machine Learning, 32, 83–84 (1998)

Guest Editors' Introduction

GERHARD WIDMER gerhard@ai.univie.ac.at
Department of Medical Cybernetics and AI, University of Vienna, and Austrian Research Institute for Artificial Intelligence, Freyung 6/2, A-1010 Vienna, Austria

MIROSLAV KUBAT mkubat@cacs.usl.edu
Center for Advanced Computer Studies, University of Southwestern Louisiana, Lafayette, LA 70504-4330, and Department of Computer Science, Southern University, Baton Rouge, LA 70813-0400

In the last few years we have seen growing interest in learning settings where the learning target and/or the context of learning change over time. This special issue was motivated by the success of a workshop on "Learning in Context-Sensitive Domains" (Kubat & Widmer, 1996) that was held just prior to the 13th International Conference on Machine Learning (ICML'96). Our intention is to raise the awareness, in the machine learning community, of the role of context in learning and its relation to the phenomenon of concept drift.

Context, in the sense treated here, can informally be defined as any information that, while relevant to learning, is not made explicit to the learning agent. This situation is not uncommon — examples from the history of machine learning and pattern recognition include omitting illumination features in computer vision and ignoring language accents in speech recognition systems. To provide all aspects of a concept is often beyond the means of the "teacher"; and even if all requisite features are included, complications may arise in applications where the training examples do not permit the agent to detect their relevance. By way of illustration, suppose that all accent-related features have been included in a speech recognition system. If the training examples are from a single homogeneous group of speakers, then the importance of these contextual features will not be recognized.

The context can *change with time;* the accuracy of an on-line learner that has been trained on single-context examples may drop when the context changes. Indeed, it is this drop that may allow the agent to become aware of the existence of a hidden context. Gradual or abrupt context changes often become apparent in the form of *concept drift* (Schlimmer & Granger, 1986). Tracking concept drift on-line requires that the learner continually monitor its performance and adjust its hypotheses whenever necessary. Some domains even call for learners that are able to "forget" old, outdated information.

In batch learning, problems arise if the training data are collected in batches that pertain to different contexts — for example, groups of digital images that have been made under different lighting conditions. Effective learning algorithms should be able to recognize these kinds of discontinuities and to adapt their hypotheses to different conditions.

The articles collected in this special issue cover quite a spectrum of context-related issues, from theoretical as well as from practical perspectives. The first paper, "Robust Sensor Fusion: Analysis and Application to Audio Visual Speech Recognition" by Movellan and Mineiro, investigates the effects of a *changed context* on the classification performance of a multimodal recognition system. The authors show that individual recognition modules that were trained separately may have a detrimental effect on the overall system behavior when they are used in conditions that differ from their training context. They propose a principled solution to this problem that builds on the ideas of competitive models and Bayesian robustification. The effectiveness of the approach is demonstrated in a realistic audio-visual speech recognition task.

In the second paper, “Extracting Hidden Context,” Harries, Sammut and Horn present a novel approach to detecting *hidden context changes* in temporally ordered data. They describe a meta-learning algorithm that exploits a decision tree generator to perform what the authors call ‘contextual clustering’: to identify hidden contexts and locally stable representations associated with these contexts. In this way they set up an interesting bridge between off-line learning and on-line recognition of context changes. The algorithm was applied to a complex control task, and it successfully identified context changes that up to now had to be marked by the user.

The next two papers study the problem of *concept drift* in on-line learning from the perspective of Computational Learning Theory (COLT). In “Tracking the Best Disjunction”, Auer and Warmuth describe an extension of the well-known WINNOW algorithm (Littlestone, 1988) for learning k-literal disjunctions in drifting domains. The algorithm is proven to predict nearly as well as the best possible sequence of disjunctions over the sequence of training examples.

In a similar vein, “Tracking the Best Expert”, by Herbster and Warmuth, develops various drift-tracking variants of a weighted-majority algorithm (Littlestone & Warmuth, 1994) in a ‘multiple experts’ scenario, where a master algorithm learns to combine the predictions of a fixed number of individual experts. Again, the guiding assumption is that the identity of the best expert may change over time. The algorithms aim at minimizing the loss relative to the best possible sequence of individual experts. The solutions are simple and efficient, and they have guaranteed relative performance bounds for arbitrary sequences.

The final paper in this issue, “Statistical Mechanics of Online Learning of Drifting Concepts: A Variational Approach” by Vicente, Kinouchi, and Caticha, also addresses the problem of concept drift, but with a different set of analytical tools. Methods from statistical mechanics are applied to the formal analysis of on-line learning in feed-forward neural networks. Several idealized learning algorithms are studied under various types of concept drift. The authors make suggestions on how to turn these ideal algorithms into practicable methods.

We would like to express our gratitude to Tom Dietterich for his assistance and guidance throughout the entire editorial process, and also for improving and clarifying our problem formulation in the call for papers. Many thanks are due to the reviewers for their thorough work and, of course, to the authors who contributed to what we hope will be accepted as a stimulating collection of papers.

References

Kubat, M., & Widmer, G. (Eds.) (1996). *Learning in context-sensitive domains (Workshop Notes)*. 13th International Conference on Machine Learning, Bari, Italy.

Littlestone, N. (1988). Learning quickly when irrelevant attributes abound: a new linear-threshold algorithm. *Machine Learning, 2(4)*, 285–318.

Littlestone, N., & Warmuth, M. (1994). The Weighted Majority Algorithm. *Information and Computation, 108(2)*, 212–261.

Schlimmer, J.C., & Granger, R.H. (1986). Incremental learning from noisy data. *Machine Learning, 1(3)*, 317–354.