Mach Learn (2010) 79: 257–259
DOI 10.1007/s10994-010-5184-9

EDITORIAL

Machine learning algorithms for event detection

A special issue of Machine Learning

Dragos Margineantu · Weng-Keen Wong · Denver Dash

Received: 30 March 2010 / Revised: 30 March 2010 / Accepted: 30 March 2010 /
Published online: 23 April 2010

A common task in many machine learning application domains involves monitoring routinely collected data for 'interesting' events. This task is prevalent in surveillance, but also in tasks ranging from the analysis of scientific data to the monitoring of naturally occurring events, and from supervising industrial processes to observing human behavior. We will refer to this monitoring process with the purpose of identifying interesting occurrences, as event detection. We put together this special issue of the Machine Learning journal with the belief that principled machine learning approaches can and will be a differentiator in addressing event detection tasks, and that theoretical and practical advances of machine learning in this area have the potential to impact a wide range of important real-world applications such as security, public health and medicine, biology, environmental sciences, manufacturing, astrophysics, business, and economics. In the recent past, domain experts in these areas have had the laborious job of manually examining the collected data for events of interest. With the emergence of computers, many efforts have been made to replace manual inspection with an automated process. Data, however, have become increasingly complex, and the quantities of collected data have become extremely large in recent years. Multivariate records, images, video footage, audio recordings, spatial and spatio-temporal data, text documents, and even relational data are now routinely collected.

We all expect that advances in machine learning would be well-suited for this class of tasks. However, in practice, the peculiarities of the application often grossly violate the

D. Margineantu (✉)
Boeing Research & Technology, Seattle, WA, USA
e-mail: dragos.d.margineantu@boeing.com

W.-K. Wong
Oregon State University, Corvallis, OR, USA
e-mail: wong@eecs.oregonstate.edu

D. Dash
Intel Labs, Pittsburgh, PA, USA
e-mail: denver.h.dash@intel.com

assumptions of standard machine learning algorithms. For instance, the events of interest are typically a miniscule fraction of the data and introduce an extreme case of class imbalance. As a result, to address event detection problems, new approaches have to be created in order to address their particular issues and to fill the corresponding gaps in machine learning research.

As the papers selected for this issue demonstrate, machine learning based solutions have the potential to address and impact some of the most pressing real-world applications being studied today.

The main focus of this special issue of Machine Learning is on event detection tasks and on the machine learning algorithms that address these tasks. With regard to the algorithms, we were particularly interested in how event detection algorithms handle complex forms of data, incorporate useful domain knowledge, and on how these event detection algorithms are tested. We also looked for approaches that researchers have employed for event detection task decomposition and at identifying how the proposed approaches can be generalized for different types of tasks. Another goal of ours was to inspire new directions of research.

One of the greatest challenges for machine learning when applied to event detection tasks is the fact that data is not sufficient if the detection of anomalous events is of interest. Most of the approaches so far incorporate pieces of domain expertise in an ad-hoc manner while implementing the learning algorithms or while training the models on the available data. Typically such solutions require significant re-engineering if deployed or if employed on (even slightly) different tasks. Neill and Cooper (2010) extend a univariate Bayesian detection framework into a principled multivariate Bayesian approach that integrates prior domain knowledge for a highly powerful detector of emerging patterns. Their multivariate Bayesian scan statistic (MBSS) approach stands out because of its flexibility and applicability to a wide range of multivariate detection problems.

Nikovski and Jain (2010) introduce novel, scalable algorithms for detecting abrupt change in streaming data. Unlike typical changepoint detection algorithms which assume the data are drawn from a distribution of a known parametric form, the algorithms in Nikovski and Jain (2010) make no distributional assumptions of the data. The authors describe dynamic programming approaches to making their non-parametric approaches scalable and also provide a theoretical analysis of their algorithms.

Next, Tandon and Chan (2010) look specifically at addressing one of the most challenging practical event detection tasks, namely computer network intrusions (network and host-based). Their approaches produce rules that can be edited and assessed by experts. The paper carefully analyzes the tradeoff between pruning and different rule coverage augmentation techniques. The proposed Hybrid coverage augmentation leads to the highest detection rate while maintaining a false positive rate of less than 1%—both network and host-based detection.

Finally, Singliar and Hauskrecht (2010) defined detection requirements, developed, and analyzed over different parameter ranges, a series of detectors for traffic incidents. They specifically address two of the most important aspects of handling of streaming data that is collected from sensors: data affected by noise and data that is not aligned. Their learning of the Tree-Augmented Naïve Bayes (TAN) approach addresses the alignment issue in a very elegant manner, and clearly improves detection.

Many open event detection research questions remain open, such as:

- How can we discover spatio-temporal events efficiently?
- How can events in relational data be discovered?

– How should learners interface with domain experts for retraining or refocusing on different event detection tasks, via a high level language? How can highly interactive accurate event detection be achieved?
– How should complex data be handled efficiently?
– What are the best ways of handling confidence estimates of a detector?
– How can additional knowledge that is more complex and structured than probabilistic priors be incorporated into event detectors?
– What are the best statistical tests for assessing anomaly and event detectors?

Meanwhile, the papers selected for this issue discuss some of the challenges raised by these tasks, and present ideas on how their proposed techniques can be extended to address those problems. We hope that this issue of the Machine Learning journal will inspire your thinking and your research, and that it represents a small step towards generic deployed machine learning based solutions for event detection.

Acknowledgements We would like to acknowledge the contributions of all the authors who submitted articles for this special issue. Meanwhile, this issue would not have been possible without the outstanding and tireless effort of the reviewers, Melissa Fearon of Springer, and Foster Provost, the editor in chief of the journal. Thank you!

References

Neill, D., & Cooper, G. (2010). A multivariate Bayesian scan statistic for early event detection and characterization. *Machine Learning* (this issue), doi:10.1007/s10994-009-5144-4

Nikovski, D., & Jain, A. (2010). Fast adaptive algorithms for abrupt change detection. *Machine Learning* (this issue), doi:10.1007/s10994-009-5122-x

Singliar, T., & Hauskrecht, M. (2010). Learning to detect incidents from noisily labeled data. *Machine Learning* (this issue), doi:10.1007/s10994-009-5141-7

Tandon, G., & Chan, P. (2010). Increasing coverage to improve detection of network and host anomalies. *Machine Learning* (this issue), doi:10.1007/s10994-009-5145-3