CfP - Privacy Aware Machine Learning (PAML) for Health Data Science

Organized by
Andreas HOLZINGER, Peter KIESEBERG,
Edgar WEIPPL & A Min TJOA
in the context of ARES/CD-ARES 2016 in Salzburg,
August 29 – September, 2, 2016
Papers due to April, 30, 2016
http://hci-kdd.org/privacy-aware-machine-learning-for-data-science

Machine learning is the fastest growing field in computer science, and health informatics is amongst the greatest challenges, e.g. large-scale aggregate analyses of anonymized data can yield valuable insights addressing public health challenges and provide new starting points for scientific discovery. Privacy issues are becoming a major concern for machine learning tasks, which often operate on personal and sensitive data. Consequently, privacy, data protection, safety, information security and fair use of data is of utmost importance for health data science.

The amount of patient-related data produced in today's clinical settings poses many challenges with respect to collection, storage and responsible use. For example, in research and public health care analysis, data must be anonymized before transfer, for which the k-anonymity measure was introduced and successively enhanced by further criteria. As k-anonymity is an NP-hard problem, which cannot be solved by automatic machine learning (aML) approaches we must often make use of approximation and heuristics. As data security is not guaranteed given a certain k-anonymity degree, additional measures have been introduced in order to refine results (l-diversity, t-closeness, delta-presence). This motivates methods, methodologies and algorithmic machine learning approaches to tackle the problem. As the resulting data set will be a tradeoff between utility, usability and individual privacy and security, we need to optimize those measures to individual (subjective) standards. Moreover, the efficacy of an algorithm strongly depends on the background knowledge of a potential attacker as well as the underlying problem domain. One possible solution is to make use of interactive machine learning (iML) approaches and put a human-in-the-loop where the central question remains open: "could human intelligence lead to general heuristics we can use to improve heuristics?"

Research topics covered by this special session include but are not limited to the following topics:
- Production of Open Data Sets
- Synthetic data sets for machine learning algorithm testing
- Privacy preserving machine learning, data mining and knowledge discovery
- Data leak detection
- Data citation
- Differential privacy
- Anonymization and pseudonymization
- Securing expert-in-the-loop machine learning systems
- Evaluation and benchmarking

This special session will bring together scientists with diverse backgrounds, interested in both the underlying theoretical principles as well as the application of such methods for practical use in the biomedical, life sciences and health care domain. The cross-domain integration and appraisal of different fields will provide an atmosphere to foster different perspectives and opinions; it will offer a platform for crazy ideas and a fresh look on the methods to put these ideas into business.

Information about submission:
http://cd-ares-conference.eu/?page_id=43