

Learning in Big Data Analytics Introduction

Alexander Schönhuth



Bielefeld University
November 3, 2021

WHO ARE WE?

- ▶ Research group “Genome Data Science”
<https://gds.techfak.uni-bielefeld.de>
- ▶ Coordinates:
Prof. Dr. Alexander Schönhuth
email: aschoen@cebitec.uni-bielefeld.de
office: UHG U10-128

PREREQUISITES AND TOPICS

- ▶ Course prerequisites:
 - ▶ Basic knowledge machine learning
 - ▶ Big data analytics
- ▶ Important topics:
 - ▶ Data streams (covered in Big Data Analytics)
 - ▶ Recommendation systems (covered in Big Data Analytics)
 - ▶ Clustering (covered in other courses)
 - ▶ Decision trees, nearest neighbor (covered in other courses)
 - ▶ Support vector machines (covered here)
 - ▶ Social network analysis (covered in Big Data Analytics)
 - ▶ Advertizing in the Web (covered here)

MODULES

- ▶ Lecture part of modules
 - ▶ *31-M-ASM2 Advanced Statistical Methods II* (graded, “benotete Prüfungsleistung”)
 - ▶ *39-Inf-AB Algorithmen der Bioinformatik* (ungraded, “Studienleistung”)
 - ▶ *39-Inf-SAB_a Spezielle Algorithmen der Bioinformatik* (ungraded, “Studienleistung”)
 - ▶ *39-M-Inf-ABDA Advanced Big Data Analytics* (ungraded)
 - ▶ New module, see here <https://ekvv.uni-bielefeld.de/sinfo/publ/modul/308598306>

PRESENTATION, PAPERS

- ▶ Presentations:
 - ▶ Presentations in groups of 2
 - ▶ Each presentation to last for approximately 2×30 minutes, including discussions
 - ▶ Groups can be formed via LernraumPlus:
<https://lernraumplus.uni-bielefeld.de/course/view.php?id=12174>, move to “Group Assignment”
- ▶ Papers:
 - ▶ Papers for presentations to follow (some now, some later)
 - ▶ Papers available via LernraumPlus:
<https://lernraumplus.uni-bielefeld.de/course/view.php?id=12174>
 - ▶ Papers available via Wiki: <https://gds.techfak.uni-bielefeld.de/teaching/2021winter/lbda>

DATA STREAMS: GENERAL READING

- ▶ Mining Massive Datasets, Chapter 4:
<http://infolab.stanford.edu/~ullman/mmds/ch4.pdf>
- ▶ Machine Learning for Data Streams: with Practical Examples in MOA,
[https://direct.mit.edu/books/book/4475/
Machine-Learning-for-Data-Streamswith-Practical](https://direct.mit.edu/books/book/4475/Machine-Learning-for-Data-Streamswith-Practical)
- ▶ Learning under Concept Drift: A Review.
<https://arxiv.org/pdf/2004.05785.pdf>
- ▶ Learning in Nonstationary Environments: A Survey.
[https://www.researchgate.net/publication/282907128_
Learning_in_Nonstationary_Environments_A_Survey](https://www.researchgate.net/publication/282907128_Learning_in_Nonstationary_Environments_A_Survey)
- ▶ Incremental on-line learning: A review and comparison of state of the art algorithms. <https://pub.uni-bielefeld.de/download/2914730/2914731>

DATA STREAMS: PAPERS

- ▶ “Mondrian Forests: Efficient Online Random Forests”.
<https://arxiv.org/pdf/1406.2673.pdf>
- ▶ “Nonconvex online support vector machines”. <http://clgiles.ist.psu.edu/pubs/TPAMI11-Nonconvex.pdf>
- ▶ “Adaptive random forests for evolving data stream classification”.
<https://link.springer.com/article/10.1007/s10994-017-5642-8>
- ▶ “An ensemble of cluster-based classifiers for semi-supervised classification of non-stationary data streams”. https://www.researchgate.net/publication/276134371_An_ensemble_of_cluster-based_classifiers_for_semi-supervised_classification_of_non-stationary_data_streams
- ▶ “Extreme learning machine for multilayer perceptron”.
- ▶ Any suggestions equally welcome!

RECOMMENDATION SYSTEMS, WEB ADVERTIZING: GENERAL READING

- ▶ “Deep Learning based Recommender System: A Survey and New Perspectives”. <https://arxiv.org/pdf/1707.07435.pdf>
- ▶ “A review on deep learning for recommender systems: challenges and remedies”. https://daiwk.github.io/assets/Batmaz2018_Article_AReviewOnDeepLearningForRecomm.pdf
- ▶ “Reinforcement Learning”.
<http://incompleteideas.net/book/RLbook2020.pdf>

RECOMMENDATION SYSTEMS, WEB ADVERTIZING: PAPERS

- ▶ “Wide & Deep Learning for Recommender Systems”.
<https://dl.acm.org/doi/pdf/10.1145/2988450.2988454>
- ▶ “DeepFM: A Factorization-Machine based Neural Network for CTR Prediction”. <https://arxiv.org/pdf/1703.04247.pdf>
- ▶ “Collaborative deep learning for recommender systems”.
<https://arxiv.org/pdf/1409.2944.pdf>
- ▶ “Bidding Machine: Learning to Bid for Directly Optimizing Profits in Display Advertising”. <https://arxiv.org/pdf/1803.02194.pdf>
- ▶ “Time-Aware Prospective Modeling of Users for Online Display Advertising”. <http://papers.adkdd.org/2019/papers/adkdd19-gligorijevic-time.pdf>
- ▶ Alternative suggestions?

PREDICTING LINKS IN SOCIAL NETWORKS

General Reading

- ▶ Mining Massive Datasets, Chapter 10:
<http://infolab.stanford.edu/~ullman/mmds/ch10.pdf>

Papers

- ▶ “Gravity-inspired graph autoencoders for directed link prediction.”
<https://arxiv.org/pdf/1905.09570.pdf>
- ▶ “Link prediction based on graph neural networks.”
<https://papers.nips.cc/paper/2018/file/53f0d7c537d99b3824f0f99d62ea2428-Paper.pdf>
- ▶ “GCN-GAN: A non-linear temporal link prediction model for weighted dynamic networks.”. <https://arxiv.org/pdf/1901.09165.pdf>
- ▶ Again, any suggestions welcome.

NEXT LECTURE; GENERAL SCHEDULE

Yet to be determined, soon to be communicated.

Thanks for your attention!

