



OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

INF

FAKULTÄT FÜR
INFORMATIK

Topics for Seminars

Prof. Myra Spiliopoulou

myra@iti.cs.uni-magdeburg.de



Timeline for seminars and teamprojects

- REGISTRATION:
 - You decide for a topic, **in agreement with the supervisor.**
 - You sign the form and submit to the KMD team by
Friday 15.04.16, 13:00
- PRESENTATIONS for seminars: 27.06.2016, 09:00–12:00
- REPORT for seminars: 8.7.2016 till 13:00 s.t.
- PRESENTATIONS/REPORT for projects: Arrange with supervisor

ALL MEETINGS FROM NOW ON: **KMD** LAB (R 021)

Advanced Topics in KMD: Seminar (Master level)

Clustering in equisized clusters

Two topics on semi-supervised stream classification

Chapelle et al (2006) say that "Probably the earliest idea about using unlabeled data in classification is self-learning, which is also known as self-training, self-labeling, or decision-directed learning. This is a wrapper-algorithm that repeatedly uses a supervised learning method. It starts by training on the labeled data only. In each step a part of the unlabeled points is labeled according to the current decision function; then the supervised method is retrained using its own predictions as additional labeled points . . . "

Chapelle et al (2006) "Semi-supervised learning" by Chapelle, Olivier and Schölkopf, Bernhard and Zien, Alexander, MIT Press Cambridge

In the streaming context, semi-supervised learners have the inherent problem of dealing with concept drift. This leads to two fundamentally different types of approaches: some methods assume that some freshly labeled do arrive and are used to adapt to drift, while other methods assume that no labeled instances arrive after a first initialization of the stream.

Semi-supervised stream classification with clustering (1)

TASK: Write a literature overview of methods that combine freshly arriving labeled and unlabeled instances to learn and maintain a classification model over an evolving stream.

SUBTASKS:

1. Describe the formal problem solved by each method, focussing on commonalities among the methods.
2. Briefly describe how each method works.
3. Specify at least two criteria that allow you to compare the methods.
4. For each criterion, explain what method is best and which methods are not performing well.

Your overview must contain at least 10 methods published after 2007.

LITERATURE TO BEGIN WITH:

- Masud et al. (2008) "A practical approach to classify evolving data streams: Training with limited amount of labeled data" by Masud, Mohammad M and Gao, Jing and Khan, Latifur and Han, Jiawei and Thuraisingham, Bhavani, in *8th IEEE Int. Conf. on Data Mining (ICDM 2008)*, IEEE
- Hosseini et al. (2015) "An ensemble of cluster-based classifiers for semi-supervised classification of non-stationary data stream" by Hosseini, Mohammad Javad and Gholipur, Ameneh and Beigy, Hamid, in *Knowledge and Information Systems Journal (KAIS)*, SPRINGER.

Semi-supervised stream classification with clustering (2)

TASK: Write a literature overview of methods that use only an initial set of labeled instances, and thereafter exploit only unlabeled instances.

SUBTASKS:

1. Describe the formal problem solved by each method, focussing on commonalities among the methods.
2. Briefly describe how each method works.
3. Specify at least two criteria that allow you to compare the methods.
4. For each criterion, explain what method is best and which methods are not performing well.

Your overview must contain at least 10 methods published after 2010.

LITERATURE TO BEGIN WITH:

- Gan et al. (2013) "Using Clustering Analysis to Improve Semi-supervised Classification" by Gan, Haitao and Sang, Nong and Huang, Rui and Tong, Xiaojun and Dan, Zhiping, in *Neurocomputing Journal*, 101, 290–298, 2013, ELSEVIER
- Souza et al (2015) "Data Stream Classification Guided by Clustering on Nonstationary Environments and Extreme Verification Latency" by {Souza, Vinicius MA and Silva, Diego F and Gama, Joao and Batista, Gustavo, in *Proc. of the 2015 SIAM Data Mining Conf. (SDM 2015)*, SIAM.

Thank you very much!

Questions ?