INF

FAKULTÄT FÜR
INFORMATIK

Seminar and project topics Winter Term 2014/2015

Tommy Hielscher

Medical Data Mining

Discovering knowledge in medical data

- Finding risk factors associated with specific diseases / disorders
- Identifying subpopulations with common characteristics
- Modeling the evolution of individuals and subpopulations over time
- Prediction of medical outcomes

BS Topic 1: Feature Selection in Medical Data Mining

Finding subsets of features which are highly correlated to the target concept but exhibit only weak associations with each other

- Using redundant or irrelevant data can lead to bad classification performance

- Can support the identification of reasonable (disease-specific) medical assessments, associated characteristics and predictive risk-factors

BS Topic 1: Feature Selection in Medical Data Mining

Summarize the paper presenting the Feature Selection method:

R. Abraham, J.B. Simha and I.S.S., ”Medical datamining with a new algorithm for Feature Selection and Naïve Bayesian classifier,” in Proc. of the 10th Int. Conf. on Information Technology, pp. 44–49, 2007

- Why is Feature Selection generally important and especially in a medical context?
- Distinguish between Filter and Wrapper methods and assign the presented method accordingly.
- Explain in detail how the presented Feature Selection method works.
- How was the method evaluated?
- Elaborate on advantages and drawbacks of the method and compare it to at least one other Feature Selection method

Results of the seminar works

- Oral presentation
- Accompanying set of slides
- The written paper/seminar contribution

SW–TP 1: Visualization of distributions for classification–rule induced subpopulations

How can we visually explore the common characteristics of disorder–/disease–specific subpopulations ?

TODO

- Implementation of a tool to generate classification rules (for a given patient data set)
- Visualization of induced subpopulation statistics in a parallel coordinate plot

Master–Seminar 1: Survey on patient similarity measures

Create a survey on supervised and unsupervised measures for patient similarity assessment

TODO

- Literature has to be reviewed and compared: Find and describe methods in detail.
- Discuss the necessity of different patient representations associated with the different similarity measures.
- Compare the methods and state their unique characteristics, advantages and disadvantages

Master–Seminar 1: Survey on patient similarity measures

Entry points:

S. Klenk, J. Dippon, P. Fritz, and G. Heidemann, "Determining patient similarity in medical social networks," in Proc. of 1st Int. Workshop on Web Science and Information Exchange in the Medical Web, vol. 572. Raleigh, NC, USA: CEUR–WS, April 2010, pp. 6–14.

C. Li and H. Li, "A survey of distance metrics for nominal attributes," J. SW., vol. 5, no. 11, pp. 1262–1269, 2010.

D. R. Wilson and T. R. Martinez, "Improved heterogeneous distance functions," J. Artif. Int. Res., vol. 6, no. 1, pp. 1–34, 1997.

Master–Seminar 2: Survey on trajectory clustering methods

Create a survey on trajectory clustering methods.

TODO

- Literature has to be reviewed and compared: Find and describe different trajectory clustering methods in detail.
- Why could trajectory clustering be useful in a medical context.
- Compare the methods and state their unique characteristics, advantages and disadvantages.

Results of the seminar works

- Oral presentation
- Accompanying set of slides
- The written paper/seminar contribution