Advances in Semantic Data Mining

Nada Lavrač[1][2]

[1] Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
[2] University of Nova Gorica, 5000 Nova Gorica, Slovenia

In data mining and knowledge discovery, the expert is usually confronted with a task of constructing a predictive model or a set of descriptive patterns from the available data. In early approaches to data mining, simple tabular data was used as input to data analysis. Recent approaches, however, address much more complex data mining tasks, including relational data mining, text mining and mining of heterogeneous information networks. The availability of large amounts of semantically annotated data in all domains of science, and biology in particular, poses requirements for new data mining approaches which need to deal with increased data complexity, the relational character of semantic representations, as well as the reasoning capacities of the underlying ontologies. This talk addresses semantic data mining [1, 5], an emerging data mining task in which semantic data in the form of domain ontologies is used as background knowledge in data analytics. The talk introduces a general framework for semantic data mining, followed by the presentation of the SEGS method for semantic subgroup discovery from microarray data [2, 4], and two general methods for semantic subgroup discovery SDM-SEGS and SDM-Aleph [5] implemented as reusable workflows in the Orange4WS data mining environment [3]. The use of described methods and tools is illustrated on selected biomedical applications.

References

1. N. Lavrač, A. Vavpetič, M. Hilario, A. Kalousis, A. Lawrynowicz, and J. Potoniec. "Tutorial on Semantic Data Mining" at ECMLPKDD 2011, September 2011.
2. V. Podpecan, N. Lavrac, I. Mozetic, P. K. Novak, I. Trajkovski, L. Langohr, K. Kulovesi, H. Toivonen, M. Petek, H. Motaln, and K. Gruden. SegMine workflows for semantic microarray data analysis in Orange4WS. *BMC Bioinformatics*, 12:416, 2011.
3. V. Podpecan, M. Zemenova, and N. Lavrac. Orange4WS Environment for Service-Oriented Data Mining. *Comput. J.*, 55(1):82–98, 2012.
4. I. Trajkovski, N. Lavrac, and J. Tolar. SEGS: Search for enriched gene sets in microarray data. *Journal of Biomedical Informatics*, 41(4):588–601, 2008.
5. A. Vavpetič and N. Lavrač. Semantic Subgroup Discovery Systems and Workflows in the SDM-Toolkit. *Comput. J.*, 2012.