

Machine Learning and the Semantic Web

a Half-Day Tutorial at ICML-2005

Andreas Hotho
KDE Group, University of Kassel
hotho@cs.uni-kassel.de

Steffen Staab
ISWeb, University of Koblenz-Landau
staab@uni-koblenz.de

1 Summary of Tutorial Topic

Machine Learning and the Semantic Web are two fields of research that allow for rich and intriguing interactions.

Firstly, the Semantic Web is an effort to put ontologies as well as explicitly and richly structured semantic data on the WWW for purposes such as query answering, data integration or intelligent reasoning. While some of its core building blocks have achieved a quite sophisticated level, e.g. its representation languages, there is an urgent need to facilitate the creation of ontologies as well as the creation of corresponding data. Machine learning here serves as an important vehicle to achieve this purpose by:

- Wrapper learning
- Semi-automatic semantic annotation
- Ontology learning by classification, clustering and information extraction

Secondly, such ontologies and semantic data give rise to new possibilities for improving existing machine learning tasks such as

- Supervised classification
- Unsupervised clustering

Application of such approaches lie in typical applications like text classification and clustering, but also in improvements and new possibilities for data integration.

Thirdly, the availability of ontologies and semantic data is not only a means to improve existing tasks, it also offers chances for defining new machine learning challenges and applications, i.e. the mining of semantic data such as

- Friend-of-a-friend files
- Semantic peer-to-peer knowledge bases

The goal of this tutorial is to acquaint the reader with the basics of the Semantic Web and then to introduce him to the different interactions between Semantic Web and Machine Learning that are currently explored.

In addition we will try to address efforts which can be provided by the machine learning community to bring up the semantic web. There is a need to support tasks like ontology construction, evolution and mapping or the filling of the knowledge base. Further resulting semantic data should to be analyzed by machine learning algorithm. Both direction can be seen as a challenge for machine learning and can end up in the development of new algorithm. In the tutorial we will try to address this issue and we will show possible direction for a fruitful combination of both areas and future research.

2 Prerequisite Knowledge

We expect basic knowledge about machine learning such as is taught in a graduate course.

3 Intended Audience

This tutorial targets a broad audience in the machine learning community. It is not the target of the tutorial to analyze very specific algorithms and features currently used in the area of “Semantic Web and Machine Learning”, but to acquaint the reader with the principal problems and solutions that are currently tackled. It is thus the purpose of the tutorial to show the machine learning audience how they could potentially exploit the Semantic Web for the research challenges they investigate and to inspire them to apply their methods to challenges of the Semantic Web. The tutorial speakers will take particular care in order to point to the crucial needs that such a transfer requires.

4 Tutorial Content

4.1 Foundations of the Semantic Web

- Semantic Web
- Ontologies
- Ontology Representation (RDF, OWL)

4.2 Ontology Learning

- Evaluation of Ontology Learning approaches
- Feature extraction for learning
 - Lexical entries
 - Concepts
 - Relations
- Unsupervised approaches
- Supervised approaches
- Ontology Pruning

4.3 Semantic Annotation

- Web Information Extraction
- Hearst Patterns

4.4 Using Ontologies - Applying Ontologies in Machine Learning Tasks

- Clustering with Background Knowledge
- Classification with Background Knowledge
- Visualization of Text Cluster

References

- Antoniou, G. & van Harmelen, F. (2004). *A Semantic Web Primer*. MIT Press.
- Bloehdorn, S. & Hotho, A. (2004). Text classification by boosting weak learners based on terms and concepts. In (ICDM-04, 2004), pages 331–334.
- Bussler, C., Davies, J., Fensel, D., & Studer, R. (Eds.) (2004). *The Semantic Web: Research and Applications, First European Semantic Web Symposium, ESWS 2004, Heraklion, Crete, Greece, May 10-12, 2004, Proceedings*, volume 3053 of *Lecture Notes in Computer Science*. Springer.
- Cimiano, P., Handschuh, S., & Staab, S. (2004a). Towards the self-annotating web. In (Feldman et al., 2004), pages 462–471.
- Cimiano, P., Hotho, A., & Staab, S. (2004b). Comparing conceptual, divisive and agglomerative clustering for learning taxonomies from text. In (de Mántaras & Saitta, 2004), pages 435–439.
- Ciravegna, F. (2001). Adaptive information extraction from text by rule induction and generalisation. In (Nebel, 2001), pages 1251–1256.
- Ciravegna, F., Chapman, S., Dingli, A., & Wilks, Y. (2004). Learning to harvest information for the semantic web. In (Bussler et al., 2004), pages 312–326.
- Ciravegna, F., Dingli, A., Petrelli, D., & Wilks, Y. (2002). User-system cooperation in document annotation based on information extraction. In (Gómez-Pérez & Benjamins, 2002), pages 122–137.
- de Mántaras, R. L. & Saitta, L. (Eds.) (2004). *Proceedings of the 16th European Conference on Artificial Intelligence, ECAI'2004, including Prestigious Applicants of Intelligent Systems, PAIS 2004, Valencia, Spain, August 22-27, 2004*. IOS Press.
- Doan, A., Madhavan, J., Domingos, P., & Halevy, A. Y. (2004). Ontology matching: A machine learning approach. In (Staab & Studer, 2004), pages 385–404.
- Feldman, S. I., Uretsky, M., Najork, M., & Wills, C. E. (Eds.) (2004). *Proceedings of the 13th international conference on World Wide Web, WWW 2004, New York, NY, USA, May 17-20, 2004*. ACM.
- Fensel, D., Sycara, K. P., & Mylopoulos, J. (Eds.) (2003). *The Semantic Web - ISWC 2003, Second International Semantic Web Conference, Sanibel Island, FL, USA, October 20-23, 2003, Proceedings*, volume 2870 of *Lecture Notes in Computer Science*. Springer.
- Gómez-Pérez, A. & Benjamins, V. R. (Eds.) (2002). *Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, 13th International Conference, EKAW 2002, Sigüenza, Spain, October 1-4, 2002, Proceedings*, volume 2473 of *Lecture Notes in Computer Science*. Springer.
- Heß, A. & Kushmerick, N. (2003). Learning to attach semantic metadata to web services. In (Fensel et al., 2003), pages 258–273.
- Hotho, A., Staab, S., & Stumme, G. (2003a). Explaining text clustering results using semantic structures. In (Lavrac et al., 2003), pages 217–228.
- Hotho, A., Staab, S., & Stumme, G. (2003b). Ontologies improve text document clustering. In (ICDM-03, 2003), pages 541–544.
- ICDM-03 (2003). *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM 2003), 19-22 December 2003, Melbourne, Florida, USA*. IEEE Computer Society.
- ICDM-04 (2004). *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM 2004), 1-4 November 2004, Brighton, UK*. IEEE Computer Society.
- Lavrac, N., Gamberger, D., Blockeel, H., & Todorovski, L. (Eds.) (2003). *Knowledge Discovery in Data-*

- bases: *PKDD 2003, 7th European Conference on Principles and Practice of Knowledge Discovery in Databases, Cavtat-Dubrovnik, Croatia, September 22-26, 2003, Proceedings*, volume 2838 of *Lecture Notes in Computer Science*. Springer.
- McIlraith, S. A., Plexousakis, D., & van Harmelen, F. (Eds.) (2004). *The Semantic Web - ISWC 2004: Third International Semantic Web Conference, Hiroshima, Japan, November 7-11, 2004. Proceedings*, volume 3298 of *Lecture Notes in Computer Science*. Springer.
- Nebel, B. (Ed.) (2001). *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI 2001, Seattle, Washington, USA, August 4-10, 2001*. Morgan Kaufmann.
- Sabou, M. (2004). From software apis to web service ontologies: A semi-automatic extraction method. In (McIlraith et al., 2004), pages 410–424.
- Staab, S. & Studer, R. (Eds.) (2004). *Handbook on Ontologies*. International Handbooks on Information Systems. Springer.