Planning to learn: Recent Developments and Future Directions

Filip Železný

Czech Technical University in Prague, Faculty of Electrical Engineering
Technická 2, 16627 Prague 6, Czech Republic
zelezny@fel.cvut.cz

The talk will cover my lab's recent research concerning planning to learn and discuss its relationships to relevant work of other researchers.

I will first introduce a machine-learning application that had motivated us to explore how knowledge-discovery workflows could be designed automatically using a data-mining ontology. In particular, we mined product-engineering data such as CAD documents for structural design patterns [1]. This task entailed the orchestration of numerous data-preprocessing and machine-learning algorithms in surprisingly complex workflows. The involved technique of sorted refinement [2] lead to non-linear, non-tree knowledge discovery workflows, in that the data flow was forked into individually processed data streams later reuniting as inputs to an inductive logic programming algorithm.

Given the circumstances above, we wanted to see if the user could be alleviated from composing such complex workflows manually. To this end, we started to develop a knowledge-discovery ontology to capture the functionalities, constraints and mutual relations among data mining algorithms. Building on established strategies for automatic planning, we implemented a planning algorithm that proposes a suitable workflow using knowledge represented in the ontology, accessed by the planner through the standard SPARQL querying interface. I will review the main concepts of the ontology and the planning task, following mainly the paper [3].

Next I will present how the mentioned developed concepts were integrated and exploited in the machine-learning suite Orange4WS in collaboration with the Dept. of Knowledge Technologies at the Jozef Stefan Institute in Slovenia. This part will be based mainly on the paper [4]. I will also review our latest developments in an alternative approach to workflow planning [5], which exploits the definitions of some domain-specific experience-proven workflows. Such established workflows are viewed as initial templates, which are algorithmically optimized by exploring their variations such as a replacement of an algorithm by a functionally similar algorithm or by a suitable sub-workflow. These variations are achieved by applying special *graph-rewriting* rules applicable to non-linear and non-tree workflow graphs.

As for related work, I will mainly discuss the relationships of the efforts above to the knowledge-discovery ontology project KDOnto [6], the more foundational (and less procedurally oriented) ontology OntoDM [7], and the relevant achievements of the recent eLico project. Here, the ontology-based workflow planning

concepts [8] were integrated in the popular software tools Taverna and RapidMiner.

Finally, the talk will cover some selected topics, which I argue to be worthwhile a concentrated investigation in the future. Most imporantly, I will outline the idea of *semantic meta-learning*, in which a meta-learner could learn and reason on the basis of the semantics of (object-level) data attributes. I will also suggest to elaborate a theoretical framework in which performance indicators for data models (such as classifier accuracies) would be jointly estimated from validation experiments with the currently analysed data, and from accuracies of similar models in similar previous experiments.

Acknowledgements. The talk covers my joint work with Monika Zemenová, Petr Křemen, Jiří Bělohradský, Radomír Černoch, Ondřej Kuželka, Matěj Holec, Vid Podpečan, Igor Trajkovski, Nada Lavrač, David Monge, and C.G. Garino. I am supported by the Czech Science Foundation through project P103/10/1875 Learning from Theories.

References

1. Zakova, M., Zelezny, F., Garcia-Sedano, J.A., Tissot, C.M., Lavrac, N., Kremen, P., Molina, J.: Relational data mining applied to virtual engineering of product designs. In: Procs. of the 16th Int. Conf. on Inductive Logic Programming (ILP'06), Springer (2006)
2. Zakova, M., Zelezny, F.: Exploiting term, predicate, and feature taxonomies in propositionalization and propositional rule learning. In: The 18th Eur. Conf. on Machine Learning and the 11th Eur. Conf. on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD'07), Springer (2007)
3. Zakova, M., Kremen, P., Zelezny, F., Lavrac, N.: Automatic knowledge discovery workflow composition through ontology-based planning. IEEE Transactions on Automation Science and Engineering **8** (2011) 253–264
4. Zakova, M., Podpecan, V., Zelezny, F., Lavrac, N.: Advancing data mining workflow construction: A framework and cases using the orange toolkit. In: Int. Workshop on Third Generation Data Mining: Towards Service-oriented Knowledge Discovery (SoKD'09). (2009)
5. Belohradsky, J., Monge, D., Zelezny, F., Holec, M., Garino, C.G.: Template-based semi-automatic workflow construction for gene expression data analysis. In: Proc. of the 24th Int. Sympos. on Computer-Based Medical Systems (CBMS'11), IEEE Computer Society (2011)
6. Diamantini, C., Potena, D., Storti, E.: Kddonto: An ontology for discovery and composition of kdd algorithms. In: Int. Workshop on Third Generation Data Mining: Towards Service-oriented Knowledge Discovery (SoKD'09). (2009)
7. Panov, P., Soldatova, L.N., Džeroski, S.: Towards an ontology of data mining investigations. In: Procs. of the 12th Int. Conf. on Discovery Science (DS'09), Springer (2009)
8. Hilario, M., Nguyen, P., Do, H., Woznica, A., Kalousis, A.: Ontology-based meta-mining of knowledge discovery workflows. In: Meta-Learning in Computational Intelligence. Springer (2011)