Relational Data Mining Applied to Virtual Engineering of Product Designs

Monika Žáková[1], Filip Železný[1], Javier A. Garcia-Sedano[2], Cyril Masia Tissot[2] and Nada Lavrač[3,4]

[1] Czech Technical University, Prague, Czech Republic
[2] Semantic Systems, Av. Del Txoriherri 9, Derio, Spain
[3] Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia
[4] University of Nova Gorica, Nova Gorica, Slovenia

Abstract. Contemporary product design based on 3D CAD tools aims at improved efficiency using integrated engineering environments with access to databases of existing designs, associated documents and enterprise resource planning (ERP). One of the goals of the SEVENPRO project is to achieve design process improvements through the utilization of relational data mining (RDM), utilizing past designs and commonly agreed design ontologies. In this paper we discuss the applicability of state-of-the-art ILP systems to the RDM tasks in this application domain as well as the implied challenges to ILP research and system implementations.

1 Introduction

Despite considerable successes of ILP in various knowledge discovery problems such as in bioinformatics [7] , industrial applications of ILP have been relatively rare. Although the usefulness of ILP has been demonstrated in areas such as finite element mesh design [8], we are not aware of industrial software employing ILP technologies in real-life regular use. The SEVENPRO[5] project aims at developing a semantic virtual engineering environment for product design, extending traditional CAD tools with semantic web, virtual reality and RDM technologies.

Engineering is one of the most knowledge-intensive activities that exist. Product engineering involves diverse knowledge types including CAD structures, technical specifications, and standards. Moreover, knowledge about the electrical, mechanical and thermodynamic behaviour sometimes comes supported by means of empirical data and simulation models. Such background motivates a number of relational data mining tasks. For example, discovering design substructures frequently occurring in a corporate CAD repository would allow to establish their easily invocable templates, eliminating repetitive designing work. Similarly,

[5] SEVENPRO, Semantic Virtual Engineering Environment for Product Design, is the project IST-027473 (2006-2008) funded under the 6th Framework Programme of the European Commission. The authors acknowledge the support by this project.

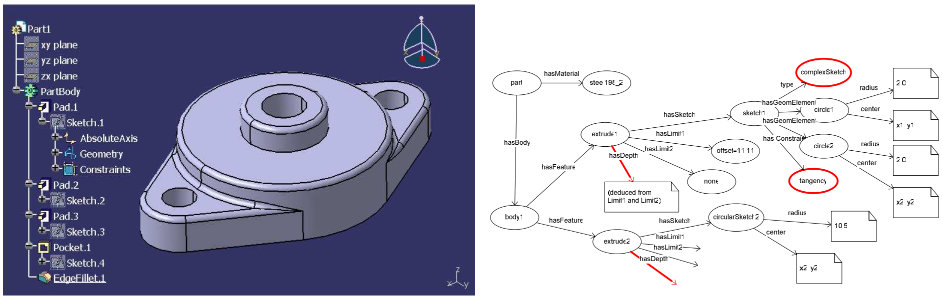

Fig. 1. Left: Example of a CAD design including commands history. Right: Part of a semantic annotation of the design.

typical patterns may be searched in design command chains[6] conducted for a product of a certain class, to make explicit the tacit knowledge of an experienced engineer. A number of other objectives involving relational classification, clustering or outlier detection are also motivated in this domain, including rather unorthodox tasks such as learning to match between a formalized product requirement set with an appropriate product design, where both the requirements and designs are represented through relational databases.

2 Problem Setting and Challenges to ILP

The information available in CAD files and other data sources is formalized and combined by means of a semantic annotation based on ontologies. An example of CAD design can be seen in Fig. 1 (left). Semantic annotation of the CAD design is generated automatically from the commands history available via the API of CAD tools. The annotation is based on an CAD ontology developed in SEVENPRO and available in the RDF format. Fig. 1 (right) shows an example of a design annotation part. Annotation including ontology of CAD items and axioms is automatically translated to Prolog.

There are three main challenges for ILP due to ontologies in the background knowledge: hierarchies of concepts, hierarchies of relations and representation conversion (between Prolog and other knowledge representation languages). *SubclassOf* is a core ontological relation correponding to a taxonomy on terms. Therefore an efficient handling of term taxonomies has to be integrated into the RDM system. The RDF formalism also allows to define hierarchies on relations by means of the *subpropertyOf* relation. To exploit the subproperty relation directly, the RDM system would have to deal with taxonomies of predicates. Also, since the induced rules should be available to the semantic server for search and reasoning, a conversion of rules into SWRL [9] has to be developed.

[6] A design is obtained by successive applications of CAD commands, such as *extrusion*, *rotation*, etc., which are mutually parameterically related. Various command sequences may lead to the same design, while differing greatly in quality respects, such as complexity, reusability, etc.

3 Ongoing Work

Horn rules are not expressive enough to model taxonomies on terms and predicates [1]. Our currently tested baseline approach for integration of taxonomies into RDM is based on refinement operator proposed in [2]. We are using the RSD system [6] which implements a propositionalization approach to relational subgroup discovery. A propositionalized representation of classified relational data is generated by constructing first-order features. During the feature generation a table of mutual feature subsumptions is maintained. The subsumption relation combines θ-subsumption with taxonomies on terms and predicates. This subsumption is exploited in propositional search which prunes any conjunctions of a subsumer with its subsumee and specializes a conjunction not only by extending it but also by replacing an included feature with its subsumee.

For a closer integration of hierarchical background knowledge into RSD we are considering the following approaches: integrating a special refinement operator into RSD as in [2], using a hybrid language integrating description logics and Horn logic e.g. $\mathcal{AL}$-log [5] and CARIN [1] or using a more expressive description language such as ψ-terms [3] or antecedent description grammars [4]. The first approach, i.e. extending RSD's refinement operators would give us the ability to deal with other requirements on specialized refinement in future. The hybrid languages mentioned above offer well-defined approaches based on DLs popular for ontology development. However, reasoning in hybrid languages generally harder than in separate languages. Also $\mathcal{AL}$-log does not support predicate taxonomies. The third approach would require modules for automatic transformation to and from these languages, but it would provide a general framework supporting not only subsumption on terms and predicates but also e.g. partial descriptions for ψ-terms and constraints on predicate use for antecedent description grammars.

References

1. Levy, A. and Rousset, M.-C.: Combining Horn rules and description logics in CARIN. Artificial Intelligence, 104:165209, 1998.
2. Frisch, A.: Sorted downward refinement: Building background knowledge into a refinement operator for ILP. In ILP, LNAI 1634:104115. Springer, 1999.
3. Sasaki, Y.: Induction of Logic Programs based on psi-terms, In proc. of ALT-99, Tokyo, LNAI 1720:169-181, Springer, 1999.
4. Cohen, W. W.: Grammatically biased learning: learning logic programs using an explicit antecedent description language. Artificial Intelligence 68(2): 303-366, 1994.
5. Lisi, F.A., Malerba, D.: Ideal Refinement of Descriptions in AL-Log. In proceedings of ILP 2003, Szeged, Hungary. LNCS 2835: 215-232, Springer 2003.
6. Železný, F., Lavrač, N.: Propositionalization-based relational subgroup discovery with RSD. Machine Learning 62: 33-63, 2006.
7. R. D. King et. al.: Functional genomic hypothesis generation and experimentation by a robot scientist.*Nature*, 427:247–252, 2004.
8. B. Dolsak et al.: Finite element mesh design: An engineering domain for ILP application. In proc. of ILP 1994, GMD-Studien 237: 305-320, 1994.
9. Horrocks, I. et al.: SWRL: A semantic web rule language combining OWL and RuleML. 21 May 2004. Available at http://www.w3.org/Submission/SWRL/.