Learning Semantics for Automated Reasoning

Daniel Kühlwein, Evgeni Tsivtsivadze, Josef Urban, and Tom Heskes
Institute for Computing and Information Sciences
Radboud University Nijmegen
The Netherlands
`firstname.lastname@science.ru.nl`

Abstract

Learning reasoning techniques from previous knowledge is a largely underdeveloped area of automated reasoning. As large bodies of formal knowledge are becoming available, state-of-the-art machine learning methods, particularly the ones that are able to leverage semantics from the mathematical libraries, provide a new avenue for problem-specific detection of relevant knowledge contained in the mathematical data. We present automated reasoning as a novel application area for machine learning and briefly describe promising results we have recently obtained.

1 Automated Reasoning and Machine Learning

In the last fifteen years, the body of formally expressed mathematics has grown substantially. Interactive Theorem Provers (ITPs) like Coq, Isabelle, Mizar, and HOL [15] have been used for advanced formal theory developments and verification of non-trivial theorems, like the Four Color Theorem and Jordan Curve Theorem, and also for advanced verification of software and hardware models. The large Mizar mathematical library (MML)[1] contains today nearly 1100 formal mathematical articles, covering a substantial part of standard undergraduate mathematical knowledge. The library has about 50000 theorems, proved with about 2.5 million lines of mathematical proofs.

Such proofs often contain nontrivial mathematical ideas, sometimes precised over decades and centuries of development of mathematics and abstract formal thinking. Having this kind of a "knowledge base of abstract human thinking" in a completely machine-processable and machine-understandable way, presents very interesting opportunities for application and development of novel artificial intelligence methods that make use of the *semantic* knowledge in various ways. An example is the MaLARea meta-system [14] combining deductive proof finding (automated theorem proving) with learning from new proofs in a closed feedback loop, boosting over time the performance of both the deductive and inductive components. A concrete and pressing task is the selection of relevant premises from the large formal knowledge bases, when one is presented with a new conjecture that needs to be proved. Providing good solution to this problem is important both for the mathematicians, and also for the existing tools for automated theorem proving (ATP) that typically cannot be successfully used directly with tens or hundreds of thousands of axioms. Experiments like the MPTP Challenge[2] and the LTB (Large Theory Batch) division of the CASC competition[3] that smart premise selection can significantly boost the performance of existing ATP techniques in large domains [14].

We aim to solve the following *premise selection* problem in large real-world mathematics: given a large knowledge base $\mathcal{P}$ of thousands of premises and proofs, and a new conjecture x, find the

[1] `http://www.mizar.org`
[2] `http://www.tptp.org/MPTPChallenge`
[3] `http://www.tptp.org/CASC`

premises $\mathcal{P}_x$ that are most relevant for proving x. We have the following setting: Let Γ be the set of all first order formulas over a fixed countable alphabet, $\mathcal{X} = \{x_i \mid 1 \leq i \leq n\} \subset \Gamma$ be the set of conjectures, $\mathcal{P} = \{p_j \mid 1 \leq j \leq m\} \subset \Gamma$ be the set of premises, and $\mathcal{Y} : \mathcal{X} \times \mathcal{P} \to \{0, 1\}$ be the indicator function such that $y_{x_i,p_j} = 1$ if p_j is used to prove x_i and $y_{x_i,p_j} = 0$ if p_j is not used to prove x_i. For each premise $p \in \mathcal{P}$ we can construct a dataset $\mathcal{D}_p = \{(x, y_{x,p}) \mid x \in \mathcal{X}\}$. Based on $\mathcal{D}_p$, a suitable algorithm can *learn* a classifier $C_p(\cdot) : \Gamma \to \mathbb{R}$ which, given a formula x as input, can *predict* whether the premise p is relevant for proving x. Typically, classifiers give a graded output. Having learned classifiers for all premises p, the classifier predictions $C_p(x)$ can be ranked: the premises that are predicted to be most relevant will have the highest output $C_p(x)$.

2 Results

We address described above problem by proposing semantic graph kernel that is tailored for the premise selection task [12]. Using an extension of the RLSC algorithm (also known as kernel ridge regression [9], proximal svm [4], ls-svm [11]) we compare it with other kernels in [12]. Figure 2 shows the result of one of the experiments performed in the paper. We evaluate the classification performance of five kernel functions: geometric graph [3], linear, latent semantic [2], Gaussian [10], and semantic graph [3]. The semantic graph performs better compared to other kernel functions, in particular when the same fact is introduced with different syntactic representations.

In [5] we introduce two kernel-based learning algorithms and compare them with a state-of-the-art automated reasoning tool. The average AUC of the proposed methods is up to 23% higher than that of automated reasoning tool. Furthermore, we have recently compared several kernel based classifiers, naive Bayes, and state-of-the art automated reasoning heuristics on a part of the MML library, the so called MPTP2078 benchmark, which contains 2078 problems from 33 Mizar articles [1]. On this dataset, prediction obtained using naive Bayes algorithm leads to a 30,8% more problems solved compared to the state-of-the-art automated reasoning method. Our kernel-based classifier leads to a 40,7% improvement.

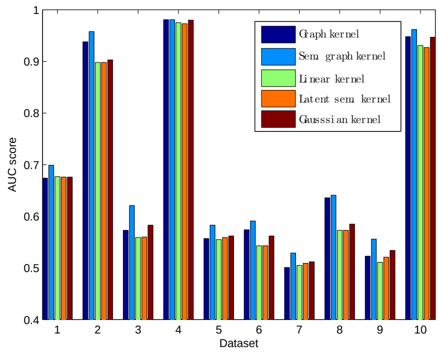

Figure 1: Semantic graph kernel outperforms geometric graph [3], linear, latent semantic [2], and Gaussian [10] kernels on 10 datasets.

3 Significance and Impact

So far premise selection methods for automated reasoning provide initial solutions over medium-size problems [6], and heuristics like SiNE [13] perform reasonably well over large ontologies consisting mainly of definitions, like SUMO [7] and Cyc [8]. However, these methods are still relatively weak (15% success in proving the Mizar theorems, see [13]) in comparison with human-based premise selection when used on very large libraries with complicated proof structure and many complicated theorems. Structured and complicated large mathematical libraries provide an interesting challenge and application field for development of machine learning methods that are aware of the libraries' contents, semantics, and proof structure. The work described here is a first serious attempt to develop kernel-based learning algorithms that aim to solve premise selection problem by taking into account *structure and semantics* of the theorems contained in the mathematical libraries.

4 Conclusion

Obtained results indicate that the performance of automated theorem proving for real-world mathematics can be notably improved by using machine learning methods that can leverage semantics and structural information from the mathematical data. We have also incorporated proposed methods into open source ATP system [5] which leads to notable benefits for the end users both in terms of accuracy and efficiency.

Acknowledgments

We acknowledge support from the Netherlands Organization for Scientific Research (NWO), in particular `Learning2Reason` and `Vici` grants (639.023.604).

References

[1] Jesse Alama, Daniel Kühlwein, Evgeni Tsivtsivadze, Josef Urban, and Tom Heskes. Premise selection for mathematics by corpus analysis and kernel methods. *CoRR*, abs/1108.3446, 2011.

[2] Nello Cristianini, John Shawe-Taylor, and Huma Lodhi. Latent semantic kernels. *Journal of Intelligent Information Systems*, 18(2-3):127–152, 2002.

[3] Scott C. Deerwester, Susan T. Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. Indexing by latent semantic analysis. *JASIS*, 41(6):391–407, 1990.

[4] Glenn Fung and Olvi L. Mangasarian. Proximal support vector machine classifiers. In *ACM SIGKDD*, pages 77–86. ACM, 2001.

[5] Daniel Kühlwein, Evgeni Tsivtsivadze, Josef Urban, Tom Heskes, and Herman Geuvers. Multi-Output Ranking for Automated Reasoning. In *KDIR*, 2011.

[6] Jia Meng and Lawrence C. Paulson. Lightweight relevance filtering for machine-generated resolution problems. *Journal of Applied Logic*, 7(1):41–57, 2009.

[7] I. Niles and A. Pease. Towards A Standard Upper Ontology. In C. Welty and B. Smith, editors, *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems*, pages 2–9, 2001.

[8] K. Panton, C. Matuszek, D. Lenat, D. Schneider, M. Witbrock, N. Siegel, and B. Shepard. Common Sense Reasoning - From Cyc to Intelligent Assistant. In Y. Cai and J. Abascal, editors, *Ambient Intelligence in Everyday Life*, pages 1–31. Springer-Verlag, 2006.

[9] Craig Saunders, Alexander Gammerman, and Volodya Vovk. Ridge regression learning algorithm in dual variables. In *ICML*, pages 515–521. Morgan Kaufmann Publishers Inc., 1998.

[10] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.

[11] Johan A K Suykens, Tony Van Gestel, Jos De Brabanter, Bart De Moor, and Joos Vandewalle. *Least Squares Support Vector Machines*. World Scientific Publishing, Singapore, 2002.

[12] Evgeni Tsivtsivadze, Josef Urban, Herman Geuvers, and Tom Heskes. Semantic graph kernels for automated reasoning. In *SDM*, pages 795–803, 2011.

[13] Josef Urban, Krystof Hoder, and Andrei Voronkov. Evaluation of automated theorem proving on the mizar mathematical library. In *ICMS*, pages 155–166, 2010.

[14] Josef Urban, Geoff Sutcliffe, Petr Pudlák, and Jirí Vyskocil. MaLARea SG1- Machine Learner for Automated Reasoning with Semantic Guidance. In *IJCAR*, pages 441–456. Springer, 2008.

[15] Freek Wiedijk, editor. *The Seventeen Provers of the World, Foreword by Dana S. Scott*, volume 3600. Springer, 2006.