MaSh: Machine Learning for Sledgehammer

Daniel Kühlwein DANIEL.KUEHLWEIN@GMAIL.COM
ICIS, Radboud Universiteit Nijmegen, The Netherlands

Jasmin Christian Blanchette BLANCHETTE@IN.TUM.DE
Fakultät für Informatik, Technische Universität München, Germany

Cezary Kaliszyk CEZARY.KALISZYK@UIBK.AC.AT
Institut für Informatik, Universität Innsbruck, Austria

Josef Urban JOSEF.URBAN@GMAIL.COM
ICIS, Radboud Universiteit Nijmegen, The Netherlands

Keywords: Machine Learning, Interactive Theorem Proving, Automatic Theorem Proving

Sledgehammer (Paulson & Blanchette, 2010) is a subsystem of the proof assistant Isabelle/HOL (Nipkow et al., 2002) that discharges interactive goals by harnessing external automatic theorem provers (ATPs). It heuristically selects a number of relevant facts (axioms, definitions, or lemmas) from the thousands available in background libraries and the user's formalization, translates the problem to the external provers' logics, and reconstructs any machine-found proof in Isabelle. The tool is popular with both novices and experts.

Meng and Paulson (Meng & Paulson, 2009) designed a filter, MePo, that iteratively ranks and selects facts similar to the current goal, based on the symbols they contain. Despite its simplicity, and despite advances in prover technology (Hoder & Voronkov, 2011; Schulz, 2011), this filter greatly increases the success rate: Most provers cannot cope with tens of thousands of formulas, and translating so many formulas would also put a heavy burden on Sledgehammer.

MaSh is a learning-based alternative to MePo. It learns from successful proofs, whether human-written or machine-generated. MaSh's heart is a Python program that implements a custom version of a weighted sparse naive Bayes algorithm that is faster than the algorithms used in previous studies (Alama et al., 2011). The program maintains a persistent state and supports incremental, nonmonotonic updates. Although distributed with Isabelle, it is fully independent and could be used by other proof assistants or applications with similar requirements.

The full paper is submitted to ITP 2013, the fourth conference on interactive theorem proving.

Acknowledgments

Daniel Kühlwein is supported by the Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO) project Learning2Reason. Jasmin Blanchette is supported by the Deutsche Forschungsgemeinschaft (DFG) project Hardening the Hammer (grant Ni 491/14-1).

References

Alama, J., Heskes, T., Kühlwein, D., Tsivtsivadze, E., & Urban, J. (2011). Premise selection for mathematics by corpus analysis and kernel methods. *CoRR*, *abs/1108.3446*. `http://arxiv.org/abs/1108.3446`.

Hoder, K., & Voronkov, A. (2011). Sine qua non for large theory reasoning. *CADE-23* (pp. 299–314). Springer.

Meng, J., & Paulson, L. C. (2009). Lightweight relevance filtering for machine-generated resolution problems. *J. Applied Logic*, *7*, 41–57.

Nipkow, T., Paulson, L. C., & Wenzel, M. (2002). *Isabelle/HOL: A proof assistant for higher-order logic*, vol. 2283 of *LNCS*. Springer.

Paulson, L. C., & Blanchette, J. C. (2010). Three years of experience with Sledgehammer, a practical link between automatic and interactive theorem provers. *IWIL-2010*.

Schulz, S. (2011). First-order deduction for large knowledge bases. Presentation at Deduction at Scale 2011 Seminar, Ringberg Castle.