Exploration of Machine Translation Techniques in Auto-formalization of Mathematics: Three Experiments

Qingxiang, Wang[12], Cezary Kaliszyk[1], Josef Urban[2]

[1] University of Innsbruck
[2] Czech Technical University in Prague

Formalizing all existing mathematical knowledge into a computer-based database and formally checking the correctness of all the mathematical proofs have for a long time been a dream for researchers in interactive and automated theorem proving [1, 2]. Achieving such goals will promote the dissemination of mathematical ideas [3], confirm the validity of complex mathematical proofs [4, 5], and the size of such a database can be an invaluable data source for using machine learning techniques in automated theorem proving [6, 7].

However, despite the existence of various formalization libraries (e.g. Mizar, HOL family, Coq, etc) which contain only a portion of mathematics, extracting information from all mathematics literature is still a task that is too costly to be done by manual labor. We believe that machine learning techniques themselves, especially artificial neural networks, can be adapted to facilitate and automate formalization of mathematics.

As machine learning techniques require training data but the purpose of formalization itself is to obtain data, we run into a dilemma of having to gather enough training data at the very beginning. Previous approaches circumvented the issue by synthesizing informal mathematical statements directly from formal statements [8, 9]. Thanks to the work of Bancerek et al. [10] which generated over 1 million latex-to-mizar aligned statement pairs, we were able to train a recurrent neural network model that achieved promising results for informal-to-formal translation [11].

Our previous work proves that machine learning techniques can be of use in accelerating formalization of mathematics, but the issue of obtaining textbook mathematical data remains. As we have found later, the trained neural network model does not generalize well for arbitrary mathematical statement. Based on this, we propose three experiments to further increase the quality of translation:

1. **Increase generalization power by employing unsupervised machine learning**. Given two corpora which may not need to be aligned, the auto-denoising and back-translation technique described in [12, 13] can transform an unsupervised learning problem into a series of supervised machine learning problems. Through a fixed word embedding [14] during the whole training process, the unsupervised learning is known to achieve reasonable translation in natural language. This paradigm is suitable for our scenario as there are reasonably large datasets in either informal latex (e.g. Proofwiki) or language from a formal proof assistant (e.g. Mizar, HOL-Light). The experiment can be conducted by first only aligning similar subjects such as point-set topology, then proceeding further to all subjects available in the corpora. We have conducted several initial experiments, showing that deploying unsupervised methods is an interesting direction.

2. **Improve translation quality by adding type-checking mechanism during training**. Translation from a trained neural network is well-known to be fluent, but since we are translating to a formal language, it is possible to type-check the decoded formal statements. Successfully type-checked statements can give preference on the weight space

therefore affecting the training process, or the generated correct statements can be added to the training data for a new training pass. For Mizar in particular, the translated statement could be first transformed into an intermediate statement by a probabilistic CKY parser as described in [9] or another pre-trained sequence-to-sequence network, then a custom Mizar elaborator [15] can be invoked to fill in type information on all involving notions of the intermediate statement. Successful elaboration amounts to successful type-checking of the Mizar statement. We have already plugged in our custom elaborator into the learning toolchain, and experimented with several suitable formats (lisp-based parse trees, TPTP, prefix notation) for the neural task that precedes the elaboration. The biggest issue seems to be the verbosity of some of the formats, that seems unsuitable for unmodified neural sequence-to-sequence methods. The best representation however achieves a reasonable perplexity (less than 2) on a dataset of 50000 aligned statements, indicating that it will be possible to set up interesting feedback loops based on elaboration.

3. **Explore various input-output formats and figure out new evaluation metric**. Progress in tree neural networks [16] has made it tempting to be adapted in informal-to-formal translation, as both informal and formal statements are suitable to be represented in tree format. We will see the gain from different input-output combinations and incorporate the result into the above two experiments. Currently we evaluate the quality of translated statements using standard NLP metrics (e.g. BLEU rate, perplexity), new metrics that are more akin to logical statements need to be explored and adapted when the output formats are changed.

We hope the proposed experiments could reveal insights on informal-to-formal translation. Positive results from the above experiments can provide us with more confidence in using machine learning in auto-formalization of mathematics. The talk will report on the several experiments done so far and the results achieved in them.

References

[1] The qed manifesto. In *Proceedings of the 12th International Conference on Automated Deduction*, CADE-12, pages 238–251, London, UK, UK, 1994. Springer-Verlag.

[2] Freek Wiedijk. The QED manifesto revisited. *Studies in Logic, Grammar and Rhetoric*, 10(23):121–133, 2007.

[3] Michael Kohlhase and Florian Rabe. Qed reloaded: Towards a pluralistic formal library of mathematical knowledge. *J. Formalized Reasoning*, 9:201–234, 2016.

[4] Thomas Hales, Mark Adams, Gertrud Bauer, Tat Dat Dang, John Harrison, Le Truong Hoang, Cezary Kaliszyk, Victor Magron, Sean Mclaughlin, Tat Thang Nguyen, Quang Truong Nguyen, Tobias Nipkow, Steven Obua, Joseph Pleso, Jason Rute, Alexey Solovyev, Thi Hoai An Ta, Nam Trung Tran, Thi Diep Trieu, Josef Urban, Ky Vu, and Roland Zumkeller. A formal proof of the Kepler conjecture. *Forum of Mathematics, Pi*, 5, 2017.

[5] Georges Gonthier. Computer mathematics. chapter The Four Colour Theorem: Engineering of a Formal Proof, pages 333–333. Springer-Verlag, Berlin, Heidelberg, 2008.

[6] Geoffrey Irving, Christian Szegedy, Alexander A. Alemi, Niklas Eén, François Chollet, and Josef Urban. Deepmath - deep sequence models for premise selection. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 2235–2243, 2016.

[7] Bartosz Piotrowski and Josef Urban. Atpboost: Learning premise selection in binary setting with ATP feedback. *CoRR*, abs/1802.03375, 2018.

[8] Cezary Kaliszyk, Josef Urban, and Jiří Vyskočil. Automating formalization by statistical and semantic parsing of mathematics. In Mauricio Ayala-Rincón and César A. Muñoz, editors, *8th International Conference on Interactive Theorem Proving (ITP 2017)*, volume 10499 of *Lecture Notes in Computer Science*, pages 12–27. Springer, 2017.

[9] Cezary Kaliszyk, Josef Urban, and Jiří Vyskočil. System description: Statistical parsing of informalized mizar formulas. 2017. to appear.

[10] Grzegorz Bancerek, Adam Naumowicz, and Josef Urban. System description: Xsl-based translator of mizar to latex. In *Intelligent Computer Mathematics - 11th International Conference, CICM 2018, Hagenberg, Austria, August 13-17, 2018, Proceedings*, pages 1–6, 2018.

[11] Qingxiang Wang, Cezary Kaliszyk, and Josef Urban. First experiments with neural translation of informal to formal mathematics. In Florian Rabe, William M. Farmer, Grant O. Passmore, and Abdou Youssef, editors, *11th International Conference on Intelligent Computer Mathematics (CICM 2018)*, volume 11006 of *LNCS*, pages 255–270. Springer, 2018.

[12] Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. Phrase-based & neural unsupervised machine translation. *CoRR*, abs/1804.07755, 2018.

[13] Mikel Artetxe, Gorka Labaka, Eneko Agirre, and Kyunghyun Cho. Unsupervised neural machine translation. *CoRR*, abs/1710.11041, 2017.

[14] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc., 2013.

[15] Chad E. Brown and Josef Urban. Extracting higher-order goals from the mizar mathematical library. In *Intelligent Computer Mathematics - 9th International Conference, CICM 2016, Bialystok, Poland, July 25-29, 2016, Proceedings*, pages 99–114, 2016.

[16] Kai Sheng Tai, Richard Socher, and Christopher D. Manning. Improved semantic representations from tree-structured long short-term memory networks. *CoRR*, abs/1503.00075, 2015.