From proofs to theorems*

Karel Chvalovský and Josef Urban

Czech Technical University in Prague, Czech republic,
karel@chvalovsky.cz and josef.urban@gmail.com

In automated theorem proving (ATP) the essential task is, not surprisingly, to produce a proof for a given theorem. However, for human mathematicians such a task usually involves also producing various conjectures that are proved, refuted, or more likely modified and which help to clarify the problem in question. Clearly, to mimic such an approach in ATP is a very challenging task. Moreover, the approaches proposed for computer generated conjecturing have produced mostly toy or domain specific conjectures, see e.g. [9, 5, 4, 8, 6].

One of the core problems is to even decide whether a produced conjecture is fruitful. This clearly depends on the particular task for which we want to use the conjectures. An activity usually rich on generating conjectures is reading mathematical texts. For example, the reader may anticipate the flow of the paper by guessing the next theorem based on the previous text. This seems to be an interesting machine learning task that requires a non-trivial understanding of a mathematical text. Another reading of this task is given a proof attempt what is a useful lemma that helps to complete the proof. Hence the task is hard and it is probably better to start with a related and more approachable problem namely the inverse task to what automated theorem provers do—to produce a theorem given a proof. Clearly, even this can be, especially without a proper context of the rest of the paper and for informal proofs, an extremely hard to impossible task. Nevertheless, at least there are data available for learning. Hence we are here interested in the problem of transforming a proof into a corresponding theorem.

It is much easier if we use a formalized mathematical library, in many cases it can be even trivial to produce such a transformation. However, usually it requires some, at least statistical, insight. For example, take the following tokenized[1] proof from Mizar Mathematical Library [2] (MML, contains over 50K theorems)

```
proof let L , M be non empty RelStr such that A1 : L , M are_isomorphic and A2 : L is
reflexive ; let x be Element of M ; M , L are_isomorphic by A1 , WAYBEL_1 : 6 ; then
consider f being Function of M , L such that A3 : f is isomorphic ; reconsider
fx = f . x as Element of L ; fx <= fx by A2 ; hence thesis by A3 , WAYBEL_0 : 66 ; end ;
```

as an input. The corresponding theorem is

```
theorem for L , M being non empty RelStr st L , M are_isomorphic & L is reflexive holds
M is reflexive
```

Our system is able to correctly produce this theorem. Although it is easy to extract the assumptions from the proof, in this particular case, a bit of work is required to statistically infer that we want to know that `M is reflexive` holds, because this fact does not occur explicitly in the proof. Moreover, the system provides the correct output only as its third option with `for x being Element of M holds x is reflexive` being the top candidate.

A preliminary version of our very simple system is based on a popular neural machine translation toolkit OpenNMT-py and basically follows an approach [7] developed for text summarization, because we can loosely speaking understand our task as a summarization task.

*Supported by the ERC Consolidator grant no. 649043 AI4REASON and by the Czech project AI&Reasoning CZ.02.1.01/0.0/0.0/15_003/0000466 and the European Regional Development Fund.

[1] We tokenize all the inputs and outputs to make the task better suited to our tools.

The model we employ is based on the sequence to sequence approach using a variant [1] of the attention mechanism and importantly it is able to copy words directly from the input to the output. Hence it can handle even words that it did not see during the training phase. We also experimented with the Transformer [10] model, but the results have been slightly worse. However, it is well known that this model is sensitive to the right choice of hyperparameters and after tuning them it is likely to outperform the former model.

However, we should emphasize that although natural language processing (NLP) tools have proven to be useful, they still suffer from several problems. Among them is the problem that sentences produced by such systems are in many cases logically inconsistent. Hence it may seem a bit silly to use the exactly same approach to produce mathematical theorems. However, as our task boils down basically to extracting correct sub-sequences from a proof and adding a bit of statistically plausible knowledge, it seems powerful enough for our purposes. Moreover, we do not claim that such a simple approach should provide surprisingly complicated results, but it has been experimentally shown [11] that NMT can be used to produce statistically plausible results for MML. Hence it is not so surprising that for MML we get decent results: on a test set the success rate, which means that we produce an exact match, is 0.28 (0.39 if compared against the ten most probable outputs).

A clearly challenging task is to test a similar approach on LaTeX documents from arXiv.org, where it is in many cases easy to identify theorems and proofs, but the format of proofs vary widely. We have performed a few preliminary experiments using the Stacks project, which provides a curated and coherent playground suitable for our purposes. Not surprisingly, our simple approach produces very poor results in this context. A relatively, for our purposes, small size of the dataset (ca. 12K theorems) may contribute to this, but more likely the main problem is that in such a general setting the task is no longer about selecting the right sub-sequences and guessing a statistically plausible conclusion. It requires at least a superficial understating of the problem in its entirety. Moreover, it would be helpful to take a bit more of context into account, for example, use the previous theorems as a part of the input. In fact, it helps slightly to see the previous theorems, because they provide additional sources of data.[2] Although it is possible to modify our task in many such ways, it is unlikely that such modifications will be sufficient to produce reasonable results on informal texts.

We believe that producing conjectures based on a mathematical text is an important task. And although narrowing it down to producing theorems from proofs looks much less interesting, it is a task that connects both these notions in a non-proof theoretical way, a potentially useful viewpoint on its own. Moreover, the attention mechanism makes it possible to weight the contribution of tokens in the input and use this knowledge elsewhere, for example, for fingerprinting mathematical object, cf. [3].

References

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[2] Grzegorz Bancerek, Czeslaw Bylinski, Adam Grabowski, Artur Kornilowicz, Roman Matuszewski, Adam Naumowicz, Karol Pak, and Josef Urban. Mizar: State-of-the-art and beyond. In Manfred

[2] However, it is important to carefully split our dataset into training, validation, and testing sets. Clearly, the results are much "better" if an output theorem in the test set appears (literally) also among the previous theorems in the training set.

Kerber, Jacques Carette, Cezary Kaliszyk, Florian Rabe, and Volker Sorge, editors, *Intelligent Computer Mathematics - International Conference, CICM 2015, Washington, DC, USA, July 13-17, 2015, Proceedings*, volume 9150 of *Lecture Notes in Computer Science*, pages 261–279. Springer, 2015.

[3] Sara C. Billey and Bridget E. Tenner. Fingerprint databases for theorems. *Notices Amer. Math. Soc.*, 60(8):1034–1039, 2013.

[4] Simon Colton. *Automated Theory Formation in Pure Mathematics*. Distinguished Dissertations. Springer London, 2012.

[5] Siemion Fajtlowicz. On conjectures of Graffiti. *Annals of Discrete Mathematics*, 72(1–3):113–118, 1988.

[6] Thibault Gauthier, Cezary Kaliszyk, and Josef Urban. Initial experiments with statistical conjecturing over large formal corpora. In Andrea Kohlhase, Paul Libbrecht, Bruce R. Miller, Adam Naumowicz, Walther Neuper, Pedro Quaresma, Frank Wm. Tompa, and Martin Suda, editors, *Joint Proceedings of the FM4M, MathUI, and ThEdu Workshops, Doctoral Program, and Work in Progress at the Conference on Intelligent Computer Mathematics 2016 co-located with the 9th Conference on Intelligent Computer Mathematics (CICM 2016), Bialystok, Poland, July 25-29, 2016.*, volume 1785 of *CEUR Workshop Proceedings*, pages 219–228. CEUR-WS.org, 2016.

[7] Sebastian Gehrmann, Yuntian Deng, and Alexander Rush. Bottom-up abstractive summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4098–4109, 2018.

[8] Moa Johansson, Dan Rosén, Nicholas Smallbone, and Koen Claessen. Hipster: Integrating theory exploration in a proof assistant. In Stephen M. Watt, James H. Davenport, Alan P. Sexton, Petr Sojka, and Josef Urban, editors, *Intelligent Computer Mathematics - International Conference, CICM 2014, Coimbra, Portugal, July 7-11, 2014. Proceedings*, volume 8543 of *Lecture Notes in Computer Science*, pages 108–122. Springer, 2014.

[9] Douglas Bruce Lenat. *AM: An Artificial Intelligence Approach to Discovery in Mathematics as Heuristic Search*. PhD thesis, Stanford, 1976.

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 5998–6008, 2017.

[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, *Intelligent Computer Mathematics - 11th International Conference, CICM 2018, Hagenberg, Austria, August 13-17, 2018, Proceedings*, volume 11006 of *Lecture Notes in Computer Science*, pages 255–270. Springer, 2018.