EPiC Series in Computing
Volume 51, 2017, Pages 33–36
ARCADE 2017. 1st International Workshop on Automated Reasoning: Challenges, Applications, Directions, Exemplary Achievements

AI at CADE/IJCAR

Josef Urban[1]*

Czech Technical University in Prague

Abstract

This text briefly discusses and criticizes the prevailing attitudes towards AI and Machine Learning research in the current Automated Reasoning (CADE/IJCAR) community.

"It is by logic that we prove, but by intuition that we discover."

– Henri Poincaré, Science and Method [4]

1 ARCADE Text

Has first-order ATP flattened out? This was a discussion topic at the PAAR workshop at IJCAR'14. I had two papers at the workshop. One about a 25% improvement of E prover's auto-mode by BliStr – an evolutionary strategy invention. Another a system description of MaLARea 0.4 and 0.5, a system that won the CASC 2013 LTB competition by 77% (the largest margin in CASC since 2000). Both papers were rejects from IJCAR'14. Nobody cared for these numbers and the methods involved. The reviews were rather explicit about it. This has been happening over and over with the CADE/IJCAR conferences.

Machine learning, evolutionary methods, probabilistic inference guidance, probabilistic parsing: topics often uninteresting and unknown to the small and quite unchanging circle of CADE/IJCAR reviewers. Studying the motivation, taking a look at some ML 101 course to avoid elementary pitfalls in reviews, or getting an expert subreviewer is often too much to ask. Sometimes the rejections say explicitly: "who cares about competitions and benchmarks?". No idea that ATP is actually useful and that bechmarks, challenges and competitions like CASC LTB have been targeting (for a decade!) huge applications like ITP. IJCAR'16 did not have the rebuttal phase, removing one of the very few weapons for answering to very uninformed reviews.

Bob Veroff – one of the few people who use ATPs to prove hard open conjectures – has been gone from CADE/IJCAR long ago. His ongoing research on the hints guidance and proof sketches in related theories has been crucial to his success. A very interesting and practically working kind of symbolic machine learning targeted at saturation-style ATP. Not interesting to CADE/IJCAR.

Max Planck said that science progresses by opponents dying out. This will most likely have to happen to the current petrified CADE/IJCAR, which revels in completeness proofs and

*Supported by the ERC Consolidator grant *AI4REASON* nr. 649043.

considers modern probabilistic AI methods to be “tuning” and “black magic”. Reasoning topics are now more and more being taken up by the broader AI/ML community, which is emboldened by its recent huge successes. Nobody has solved Go, and nobody can “prove” that AlphaGo’s strategy is “complete”. But nobody cares. Nobody cares about manually encoding specialized solvers for ten (or so) possible openings in Go - total waste of time and human brain power. At FLoC’06 in Seattle, I wanted to scream when I heard that one of the hot AR challenges was going to be SMT (specialized solvers for ten or so theories). This is really not the way to build general AI for all of math and reasoning![1] Wake up! AI conferences like NIPS have grown to an army of six thousand young excited people who are willing to experiment and attack old tasks in new ways. Very often this includes some sort of learning.

I have long argued that deduction and learning need to be combined to get to strong AI and ATP. It is very naive to think that we will soon be able to fully manually encode all aspects and procedures of thinking and theorem proving. Mathematicians (and thinking people in general) learn, use, and improve a huge bag of context-dependent heuristics during their studies and work. Some of them can be made “crisp” after a while, resulting e.g. in provably correct decision procedures for various fragments. But a vast majority of them are fuzzy. Today’s ATPs mostly lack the capability to learn and to use the learned knowledge. Most of today’s ATP research is done outside of large theories where learning is most rewarding.

The AR domain has strong and solid semantics. It is really cool an mostly unseen in other AI areas. We (humans) can sometimes turn “fuzzy” ideas into “crisp” ideas that have formally defined meaning and proofs, allowing computation, exact reasoning, proof verification, etc. But most of CADE/IJCAR has become so focused on producing crisp logic-oriented papers, that they forgot that 99% of human thinking and invention involves some fuzzy levels. One of the greatest opportunities for the AR field is today invention of methods that automatically go from the fuzzy understanding (e.g., in the form of some statistically learned knowledge), to its crisp formulation (in the form of a proved theorem or procedure working under exact assumptions). And also back: from crisp formulations to further fuzzy conjectures. This is going to be a huge AI topic that should be naturally taken up by the AR community. Unfortunately, the current AR community avoids statistical methods like a disease.

I am not going to argue and explain more here – I have written enough on this topic in the past fifteen years [5, 6, 7, 9, 11, 10, 2, 8, 3, 1]. Despite Dmitriy’s and Giles’s honest efforts, it is hard to believe that it would not be a wasted effort, as were my occasional replies to the most appalling reviews in the past. In 2004, I have tried something similar with the declining Mizar community, and it has been a complete failure. I wrote a wishlist[2], saying that they have a good chance to be a great project leading and benefiting the world (and themselves) IF they wake up to the 21st century. Instead, they have been doing the exact opposite of all of my points ever since, committing a slow honorable suicide in the diminishing circle of their small sect. It will take hundreds of years to get to the QED project’s goals, because our great and unmistakable leaders have said so. So let’s produce more incremental research.

I wish CADE/IJCAR had a better fate, but I am not going to put my money on it. I can see many of the symptoms. But I have a great belief in the future of AI and AR.

[1] An ARCADE review of this text mentioned “the denigration of SMT throughout the text”. I should say that I am a pragmatic user of the best ATPs and SMTs that are around, consider them to be useful tools and building blocks, and I actively modify them by learning algorithms. Also, I do not oppose hard “human labor” (e.g., formalization of mathematics, but also “manual” encoding of theory reasoning) for the pragmatic reason that such efforts (will) ultimately provide more data for AI systems to learn from. But I have long believed [5] that inventing automation for theory after theory is really not a scalable plan for automating math.

[2] http://wiki.mizar.org/twiki/bin/view/Mizar/MizarWishlist

2 ARCADE Questions

I have submitted the following questions to the ARCADE discussion session:

1. How do humans gain the ability to solve problems? How do mathematicians gain the ability to prove theorems that no current ATP system can prove?

2. How many people have read Turing's 1950 paper including its last section? How many of the CADE/IJCAR PC members have read the AlphaGo and DeepStack papers? How many of those can confidently explain the main ideas and the main reasons for the superhuman performance of these systems? How many people can confidently explain the main ideas behind the current performance of image recognition and language translation systems?

3. How many people believe that we can in 100 years produce a system that will (using today's hardware) prove Fermat's Last theorem (FLT), by manually defining "theories", manually defining "classes of problems", manually "designing solvers" for them, manually defining their "strategies and calculi", and manually "combining solvers" for more and more complicated lemmas involved in the proof of FLT?

References

[1] Jan Jakubuv and Josef Urban. ENIGMA: efficient learning-based inference guiding machine. In Herman Geuvers, Matthew England, Osman Hasan, Florian Rabe, and Olaf Teschke, editors, *Intelligent Computer Mathematics - 10th International Conference, CICM 2017, Edinburgh, UK, July 17-21, 2017, Proceedings*, volume 10383 of *Lecture Notes in Computer Science*, pages 292–302. Springer, 2017.

[2] Cezary Kaliszyk and Josef Urban. Learning-assisted automated reasoning with Flyspeck. *J. Autom. Reasoning*, 53(2):173–213, 2014.

[3] Cezary Kaliszyk, Josef Urban, and Jirí Vyskocil. Learning to parse on aligned corpora (rough diamond). In Christian Urban and Xingyuan Zhang, editors, *Interactive Theorem Proving - 6th International Conference, ITP 2015, Nanjing, China, August 24-27, 2015, Proceedings*, volume 9236 of *Lecture Notes in Computer Science*, pages 227–233. Springer, 2015.

[4] Henri Poincaré. *Science and method*. London : T. Nelson, 1914.

[5] J. Urban. Experimenting with Machine Learning in Automatic Theorem Proving. Master's thesis, Charles University, Prague, 1998. https://www.ciirc.cvut.cz/~urbanjo3/MScThesisPaper.pdf.

[6] Josef Urban. MPTP - Motivation, Implementation, First Experiments. *Journal of Automated Reasoning*, 33(3-4):319–339, 2004.

[7] Josef Urban. MPTP 0.2: Design, Implementation, and Initial Experiments. *J. Autom. Reasoning*, 37(1-2):21–43, 2006.

[8] Josef Urban. BliStr: The Blind Strategymaker. In Georg Gottlob, Geoff Sutcliffe, and Andrei Voronkov, editors, *Global Conference on Artificial Intelligence, GCAI 2015, Tbilisi, Georgia, October 16-19, 2015*, volume 36 of *EPiC Series in Computing*, pages 312–319. EasyChair, 2015.

[9] Josef Urban, Geoff Sutcliffe, Petr Pudlák, and Jirí Vyskocil. MaLARea SG1–machine learner for automated reasoning with semantic guidance. In Alessandro Armando, Peter Baumgartner, and Gilles Dowek, editors, *IJCAR*, volume 5195 of *Lecture Notes in Computer Science*, pages 441–456. Springer, 2008.

[10] Josef Urban and Jirí Vyskocil. Theorem proving in large formal mathematics as an emerging AI field. In Maria Paola Bonacina and Mark E. Stickel, editors, *Automated Reasoning and Mathematics - Essays in Memory of William W. McCune*, volume 7788 of *Lecture Notes in Computer Science*, pages 240–257. Springer, 2013.

[11] Josef Urban, Jirí Vyskocil, and Petr Stepánek. MaLeCoP: Machine learning connection prover. In Kai Brünnler and George Metcalfe, editors, *TABLEAUX*, volume 6793 of *Lecture Notes in Computer Science*, pages 263–277. Springer, 2011.