Recent Advances in Machine Learning for Mathematical Reasoning

Steven Van Vaerenbergh

Universidad de Cantabria

Symposium on Artificial Intelligence
for Mathematics Education
CIEM, Castro Urdiales, February 2020

UNIVERSIDAD DE CANTABRIA

Balacheff revisited: Learner modeling

Balacheff, 1993: "There is a gap between the meaning the learner has constructed and the intended meaning. It is **essential that the machine can diagnose this gap** and that it can provide adequate feedback to students."[1]

[1] Balacheff, N. (1993). Artificial intelligence and mathematics education: Expectations and questions. In *14th biennal of the australian association of mathematics teachers*, Perth, Australia.

Balacheff revisited: Learner modeling

Several solutions are explored concerning this problem[2]:

- The implementation of a **catalogue of errors**: the machine try to match the gap it observes at the interface to errors a priori described in a catalogue. It then provides some ad hoc feedback .
- **Error generation**: a model is implemented which allows the reconstruction of conceptions which can be the source of the errors.
- Error reconstruction: using some machine learning algorithms, the machine attempts to **automatically deduce mal-rules** which might "explain" the observed gaps.

[2] Balacheff, N. (1993). Artificial intelligence and mathematics education: Expectations and questions. *14th biennal of the australian association of mathematics teachers*, Perth, Australia.

Motivation

- **Machine learning** is a sub-field of artificial intelligence in which several breakthroughs have been made in the past 10 years:
  - computer vision;
  - natural language understanding;
  - speech recognition;
  - ...
- We study the application of machine learning to mathematics education and learner modeling, in particular problems related to **mathematical reasoning**.

A definition

"Machine Learning is the study of computer algorithms that **improve automatically through experience**." – Tom Mitchell

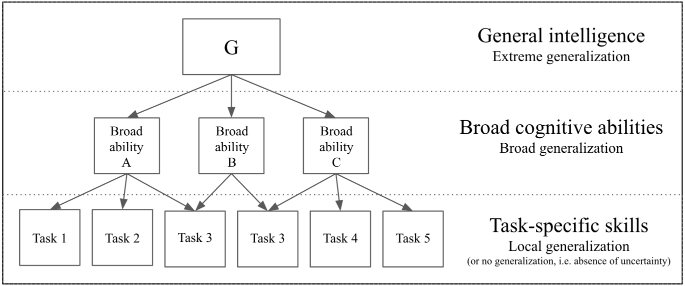
- "improve" $\rightarrow$ requires an evaluation metric
- "automatically" $\rightarrow$ without intervention
- "through experience" $\rightarrow$ by processing examples / data

Machine Learning

- Program logic is not explicitly modeled. Rather, framework to **learn model specifics** from data.
- Pattern recognition and more: primitives / building blocks include image analysis, audio analysis, but also sequence models, synthesis.

Current state of ML

Major theories of the structure of human intelligence organize cognitive abilities in a hierarchical fashion[3].

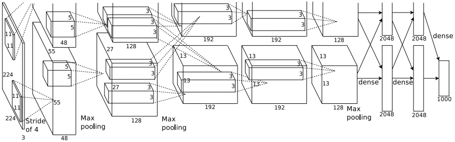


State-of-the-art machine learning achieves **task-specific skills**.

[3]Chollet, F. (2019). On the measure of intelligence. *arXiv preprint arXiv:1911.01547*.

Deep Learning

- A sub-field of machine learning in which the networks have a large amount of layers (from 5 to hundreds).
- Allows to model complex input-output relations.
- Requires lots of data and computational power. Improvements are often engineering feats.
- Deep learning "revolution" started around 2012[4].
- Past 2 years: growing interest in mathematical reasoning.



[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*.

Mathematical Reasoning

Target problems:

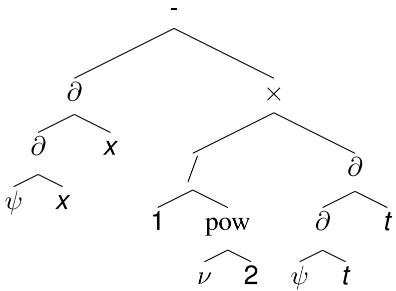
- Solve symbolic equations.
- Solve word problems.
- Automated proving.
- ...

Problem: ML and NN are “soft” algorithms that are best at **approximation**, while mathematical reasoning requires “hard”, precise algorithms.

Neural networks for symbolic reasoning

Lample, G., & Charton, F. (2019). Deep learning for symbolic mathematics. *arXiv preprint arXiv:1912.01412*:

- Treats complex equations like sentences in a language.
- Tree for $\frac{\partial^2\psi}{\partial x^2} - \frac{1}{\nu^2}\frac{\partial^2\psi}{\partial t^2} \rightarrow$



Neural networks for symbolic reasoning

- Motivation: humans rely on some kind of intuition for symbolic mathematics.
- E.g. if an expression is of the form $yy'(y^2+1)^{-1/2}$ suggests that its primitive will contain $\sqrt{y^2+1}$.
- Architecture: seq2seq transformer model with eight attention heads and six layers.
- Trained on data set of more than 100M paired equations and solutions (generated).

Neural networks for symbolic reasoning

Equation	Solution[5]
$y' = \frac{16x^3-42x^2+2x}{(-16x^8+112x^7-204x^6+28x^5-x^4+1)^{1/2}}$	$y = \sin^{-1}(4x^4 - 14x^3 + x^2)$
$3xy\cos(x) - \sqrt{9x^2\sin(x)^2+1}\,y' + 3y\sin(x) = 0$	$y = c\exp\left(\sinh^{-1}(3x\sin(x))\right)$
$4x^4yy'' - 8x^4y'^2 - 8x^3yy' - 3x^3y'' - 8x^2y^2 - 6x^2y' - 3x^2y'' - 9xy' - 3y = 0$	$y = \frac{c_1+3x+3\log(x)}{x(c_2+4x)}$

- Mathematica and Matlab: no solution for these problems.
- NN model: 99.7% and 81.2% success on integration problems and 2nd order differential equations, respectively. Mathematica: 84% and 77.2%.

[5] Lample, G., & Charton, F. (2019). Deep learning for symbolic mathematics. *arXiv preprint arXiv:1912.01412*.

Mathematical Reasoning in Latent Space

Lee, D., Szegedy, C., Rabe, M. N., Loos, S. M., & Bansal, K. (2019). Mathematical reasoning in latent space. *arXiv preprint arXiv:1909.11851*:

- Neural network maps mathematical formulas into a latent space of fixed dimension.
- This network is trained by predicting whether a given rewrite is going to succeed (i.e. returns with a new formula).
- Architecture: Combination of Graph neural networks.
- Trained on 19591 theorems from HOList database.
- First result: NN can perform several steps of approximate reasoning in latent space.

Word problems

Problem: Dan has 2 pens, Jessica has 4 pens. How many pens do they have in total?
Equation: $x = 4 + 2$
Solution: 6

Wag, Y., Liu, X., & Shi, S. (2017). Deep neural solver for math word problems. In *Proc. of the 2017 conf. on empirical methods in natural language processing*

- Recurrent neural network (seq2seq-based, GRU+LSTM).

Wang, L., Zhang, D., Gao, L., Song, J., Guo, L., & Shen, H. T. (2018). Mathdqn: Solving arithmetic word problems via deep reinforcement learning. In *Thirty-second AAAI conference on artificial intelligence*:

- Deep Q-network (two-layer feed-forward neural network).

Other works (2015-2019)

- Prediction of the next step of a proof, which is executed with a “hard” algorithm: Bansal et al., 2019; Gauthier and Kaliszyk, 2015; Lederman et al., 2018; Loos et al., 2017.
- RNN to simplify complex symbolic expressions: Zaremba et al., 2014.
- Verify the correctness of given symbolic entities using tree-structured neural networks: Arabshahi et al., 2018.
- Data set of wide range of mathematical questions and answers (symbolic, word-based, etc.): Saxton et al., 2019.

Automated Reasoning

ML for AR: exploit statistical inference of previous proofs (inductive reasoning) in the classical deductive reasoning used in ATP and ITP[6]:

- Building systems that are helpful for developers and users.
- Premise selection techniques by learning premise relevance. (Kühlwein, 2014 combines random-hill climbing based strategy finding with strategy scheduling via learned runtime predictions.)
- ML for tuning automated theorem prover to find good search strategies.

[6] Kühlwein, D. A. (2014). *Machine learning for automated reasoning* (Doctoral dissertation). Radboud Universiteit Nijmegen.

Some contributions in ARCADE 2019

Schon et al., 2019:

- Treats common-sense reasoning problems
- Background knowledge graphs are combined with target formulae and fed theorem prover. Afterwards, a machine learning component is used to predict the relevance of different models obtained.

Moser and Winkler, 2019

- Search for right granularity of features in machine learning for term rewriting and theorem proving.
- Proposes more complex, structural features to learn from.
- Applied to term rewriting.

Abstract Reasoning

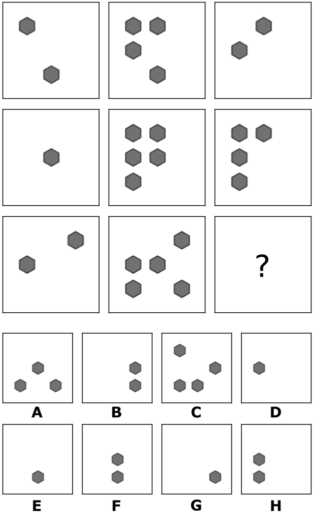
- Deep learning is often mostly **memorization**.
- In order to learn from fewer data, **generalization capabilities** are required.
- This may require basic **abstract reasoning** skills.
- Current work on abstract reasoning is inspired by examples from IQ tests, similar to Raven progressive matrices[7].

[7] Raven, J. Et al. (2003). Raven progressive matrices. *Handbook of nonverbal assessment*. Springer.

How to measure reasoning skills?

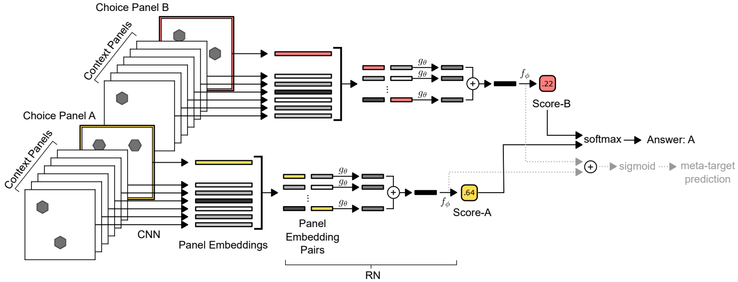
Santoro et al., 2018

- Current systems struggle on apparently simple tasks, especially when an abstract concept needs to be discovered and reapplied in a new setting.



Santoro et al., 2018:

- Architectures based on standard pattern recognition component.
- “Intepolation” of tasks is possible in some cases.
- “Extrapolation” is not possible yet. E.g. puzzles that contain dark colored objects during training and light colored objects during testing.

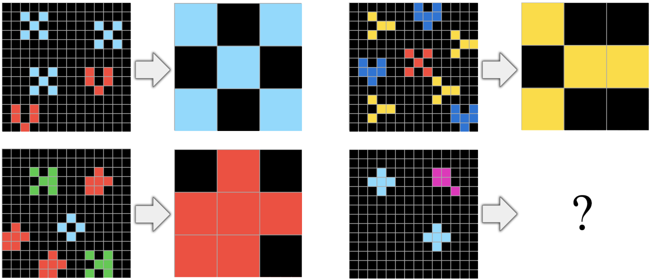


Abstraction and Reasoning Corpus

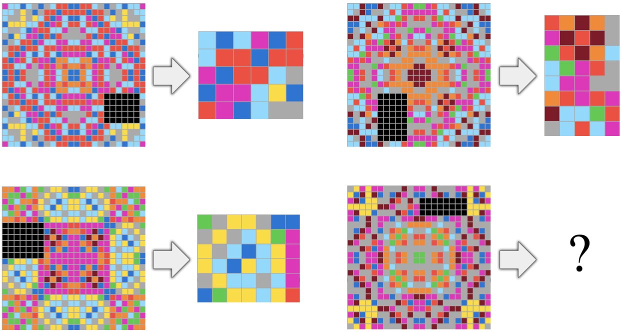
- Need for a measure of intelligence in AI context.
- Guidelines for a benchmark: reproducible, should establish validity, measure broad abilities and developer-aware generalization, description of priors, among others.
- "Abstraction and Reasoning Corpus" (ARC)[8].

[8] Chollet, F. (2019). On the measure of intelligence. *arXiv preprint arXiv:1911.01547*.

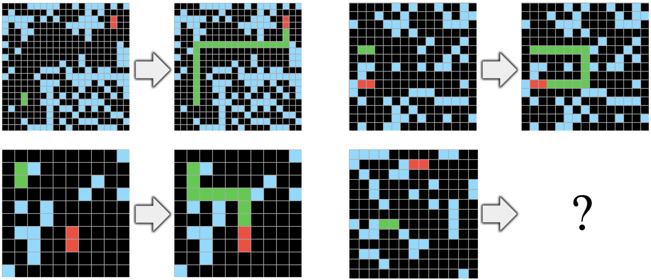
ARC data set



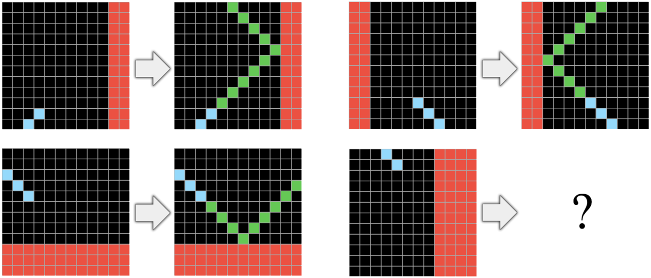
ARC data set



ARC data set



ARC data set



Conclusions

- Current ML approaches are based on **existing pattern recognition** methods.
- To level up: Either **rethink ML approach** or introduce **primitives for reasoning** into current architectures.
- Current research is working on **narrow or simplified problems**. NN methods took 20+ years to go from simple problems to solutions useful in the real world.
- Future **application in mathematics education**: student modeling, companion for learning mathematical reasoning: DGSs, but also in algebra, engineering, etc.

References I

- Arabshahi, F., Singh, S., & Anandkumar, A. (2018). Towards solving differential equations through neural programming.
- Balacheff, N. (1993). Artificial intelligence and mathematics education: Expectations and questions. In *14th biennal of the australian association of mathematics teachers*, Perth, Australia.
- Bansal, K., Loos, S. M., Rabe, M. N., Szegedy, C., & Wilcox, S. (2019). HOList: An environment for machine learning of higher-order theorem proving. *ICML 2019. International Conference on Machine Learning*.
- Chollet, F. (2019). On the measure of intelligence. *arXiv preprint arXiv:1911.01547*.

References II

- Gauthier, T., & Kaliszyk, C. (2015). Premise selection and external provers for HOL4. In *Proceedings of the 2015 conference on certified programs and proofs*. ACM.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*.
- Kühlwein, D. A. (2014). *Machine learning for automated reasoning* (Doctoral dissertation). Radboud Universiteit Nijmegen.
- Lample, G., & Charton, F. (2019). Deep learning for symbolic mathematics. *arXiv preprint arXiv:1912.01412*.
- Lederman, G., Rabe, M. N., & Seshia, S. A. (2018). Learning heuristics for automated reasoning through deep reinforcement learning. *CoRR*, *abs/1807.08058*arXiv 1807.08058. http://arxiv.org/abs/1807.08058

References III

- Lee, D., Szegedy, C., Rabe, M. N., Loos, S. M., & Bansal, K. (2019). Mathematical reasoning in latent space. *arXiv preprint arXiv:1909.11851*.
- Loos, S., Irving, G., Szegedy, C., & Kaliszyk, C. (2017). Deep network guided proof search. *LPAR-21. 21st International Conference on Logic for Programming, Artificial Intelligence and Reasoning*.
- Moser, G., & Winkler, S. (2019). Smarter features, simpler learning?, In *Proceedings arcade 2019*.
- Raven, J. Et al. (2003). Raven progressive matrices. In *Handbook of nonverbal assessment*. Springer.
- Santoro, A., Hill, F., Barrett, D., Morcos, A., & Lillicrap, T. (2018). Measuring abstract reasoning in neural networks. In *International conference on machine learning*.

References IV

- Saxton, D., Grefenstette, E., Hill, F., & Kohli, P. (2019). Analysing mathematical reasoning abilities of neural models. *arXiv preprint arXiv:1904.01557*.
- Schon, C., Siebert, S., & Stolzenburg, F. (2019). Using conceptnet to teach common sense to an automated theorem prover. In *Proceedings arcade 2019*.
- Wag, Y., Liu, X., & Shi, S. (2017). Deep neural solver for math word problems. In *Proc. of the 2017 conf. on empirical methods in natural language processing*.
- Wang, L., Zhang, D., Gao, L., Song, J., Guo, L., & Shen, H. T. (2018). Mathdqn: Solving arithmetic word problems via deep reinforcement learning. In *Thirty-second AAAI conference on artificial intelligence*.

References V

Zaremba, W., Kurach, K., & Fergus, R. (2014). Learning to discover efficient mathematical identities. In *Proceedings of the 27th international conference on neural information processing systems - volume 1*, Montreal, Canada, MIT Press.