Research Seminar

Cohn Institute

for the History and Philosophy of Science and Ideas

Tel Aviv, Israel

The Language of Mathematics

Epistemological Consequences of Applying AI Methods to Mathematics

Juan Luis Gastaldi

ETH zürich

April 17th, 2023

This project has received funding from the
European Union's Horizon 2020 research and innovation programme
under grant agreement No 839730

Outline

Szegedy-Marcus Bet on Deep Mathematics

Christian Szegedy
@ChrSzegedy

I am happy to have a long bet with anyone including @MelMitchell1 or @GaryMarcus on the formalization + theorem proving capabilities of AIs by 2029.

I am fairly confident that we will have a system with comparable or stronger capabilities to/than strong human mathematicians.

> **Joscha Bach** @Plinz · 07.06.22
> I know less about the sota in modeling math problems, but natural language parsing of school and undergrad math problems into solvers is already beginning to work, and I don't really expect it to hit any walls before 2029.
> Show this thread

09:58 · 07.06.22 · Twitter Web App

23 Retweets **8** Quote Tweets **210** Likes

Gary Marcus 🇺🇦 @GaryMarcus · 07.06.22
Replying to @ChrSzegedy and @MelMitchell1
Ok @ErnestSDavis & I will take your action, up to $100. There is nothing yet we know that can read any kind of mathematical article or book with unformalized proofs and turn it into formalization. Gap between mathematics in English and mathematics in formal notation is enormous.

3 9

Christian Szegedy @ChrSzegedy · 08.06.22
Sounds fun! I am in. ;)

2 10

Christian Szegedy @ChrSzegedy · 07.06.22
I could give a precise definition along these lines:

A diverse set of 100 graduate text books are automatically formalize/verified in a popular proof assistant (eg Lean).

10% of problems from a preselected 100 open human conjectures is proved completely autonomously.

3 7 12

Outline

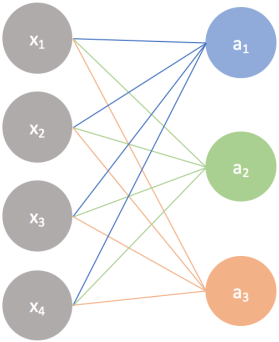
Overview of Artificial Neural Nets

Philosophical Significance of Neural Applications to Mathematics

Distributional Semantics

Distributional Arithmetics

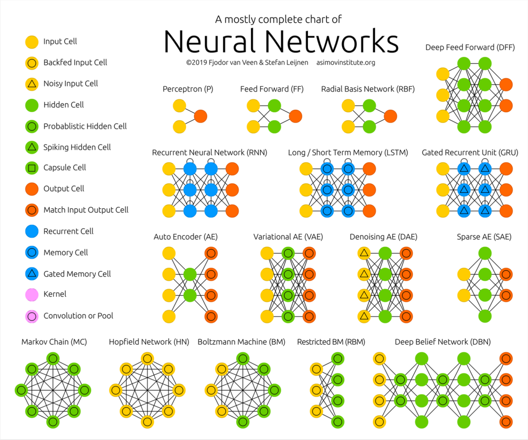
Neural Networks

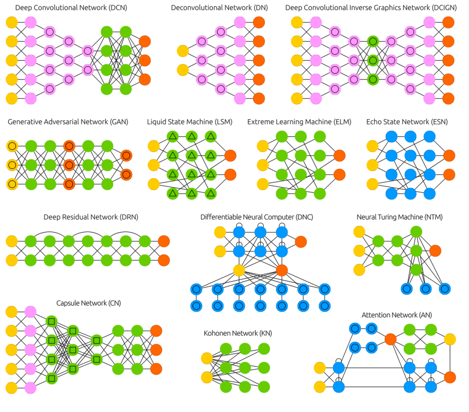


$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b \\ b \\ b \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \end{bmatrix} \xrightarrow[activation]{} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Credit: Jeremy Jordan
https://www.jeremyjordan.me/intro-to-neural-networks/

Deep Neural Nets (DNNs)



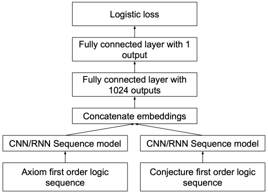


Source: https://www.asimovinstitute.org/neural-network-zoo/

Outline

Main Trends in Neural Applications to Mathematics

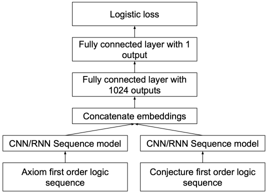
- Proof-Oriented
  - Bansal et al., 2019; Kaliszyk et al., 2017.



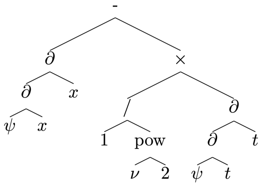
(Alemi et al., 2016)

Main Trends in Neural Applications to Mathematics

- Proof-Oriented
  - Bansal et al., 2019; Kaliszyk et al., 2017.
- Object-Oriented
  - Blechschmidt and Ernst, 2021; Charton, 2021; d'Ascoli et al., 2022; Lample and Charton, 2019; Li et al., 2021; Ryskina and Knight, 2021



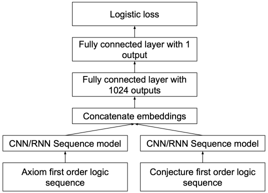
(Alemi et al., 2016)



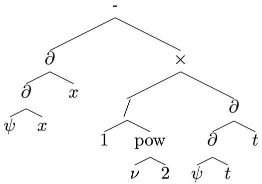
(Lample and Charton, 2019)

Main Trends in Neural Applications to Mathematics

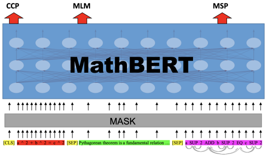
- Proof-Oriented
  - Bansal et al., 2019; Kaliszyk et al., 2017.
- Object-Oriented
  - Blechschmidt and Ernst, 2021; Charton, 2021; d'Ascoli et al., 2022; Lample and Charton, 2019; Li et al., 2021; Ryskina and Knight, 2021
- Skill-Oriented
  - Brown et al., 2020; Peng et al., 2021; Shen et al., 2021



(Alemi et al., 2016)



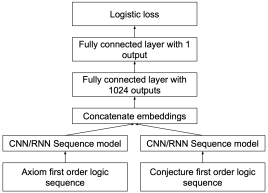
(Lample and Charton, 2019)



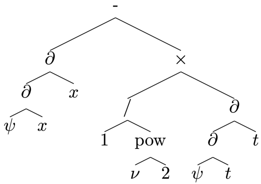
(Peng et al., 2021)

Main Trends in Neural Applications to Mathematics

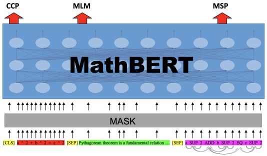
- Proof-Oriented
  - Bansal et al., 2019; Kaliszyk et al., 2017.
- Object-Oriented
  - Blechschmidt and Ernst, 2021; Charton, 2021; d'Ascoli et al., 2022; Lample and Charton, 2019; Li et al., 2021; Ryskina and Knight, 2021
- Skill-Oriented
  - Brown et al., 2020; Peng et al., 2021; Shen et al., 2021
- Heuristic-Oriented
  - Davies et al., 2021



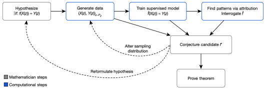
(Alemi et al., 2016)



(Lample and Charton, 2019)

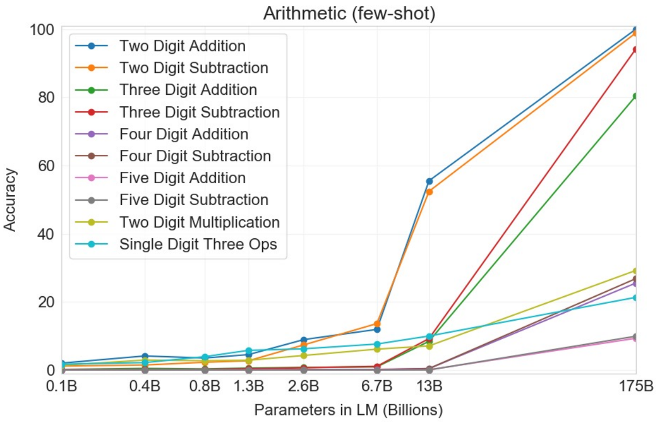


(Peng et al., 2021)



(Davies et al., 2021)

Arithmetic in Transformers



(Brown et al., 2020)

Arithmetic in Transformers

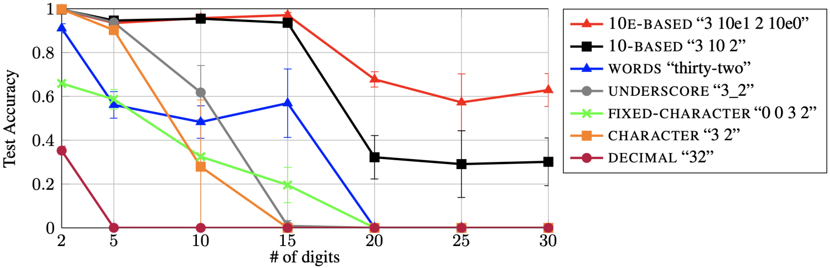


Figure 1: Accuracy of different number representations on the addition task.

(Nogueira et al., 2021)

Philosophical Significance

- The fact that mathematical properties can be addressed from the empirical perspective of current ML approaches should be enough to raise a whole series of philosophical questions.

- However, the fruitful encounter between the philosophy of mathematics and current machine learning practices has not yet taken place.

- First step in this direction:
  focus on the relation between mathematics and natural language (textuality).

- Question to be asked:
  What must mathematics be, given that models designed to analyze, reproduce and manipulate natural language are able to grasp some significant aspects of it.

Outline

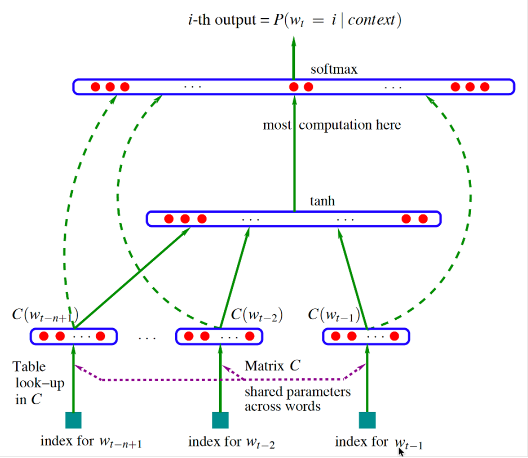
Distributionalism and Word Embeddings

- Distributional Hypothesis (Harris, 1960; Saussure, 1959)
  - "You shall know a word by the company it keeps!" (Firth, 1935)
  - The content of a linguistic unit is determined by its distribution over a corpus (i.e., the other units appearing in its context)

Distributionalism and Word Embeddings

- Distributional Hypothesis (Harris, 1960; Saussure, 1959)
  - "You shall know a word by the company it keeps!" (Firth, 1935)
  - The content of a linguistic unit is determined by its distribution over a corpus (i.e., the other units appearing in its context)

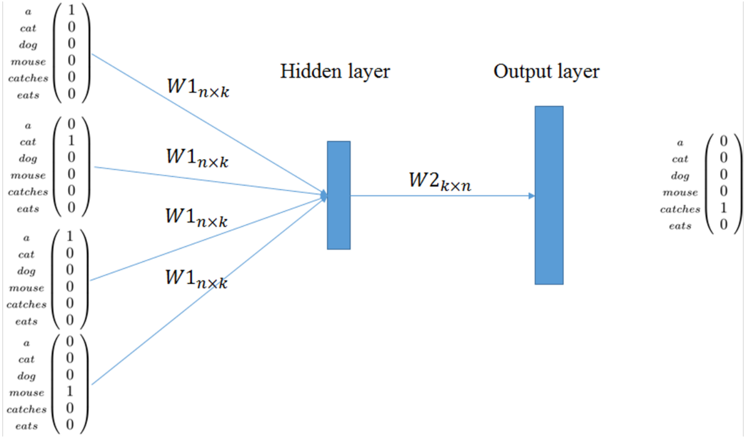
- Computational interpretation: Word Embeddings



(Bengio et al., 2003)

Word Embeddings: word2vec

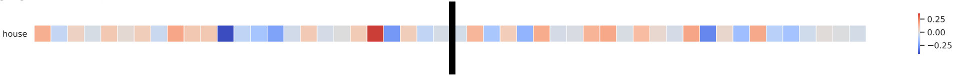
a cat <u>catches</u> a mouse



Source: Ferrone et al., 2017

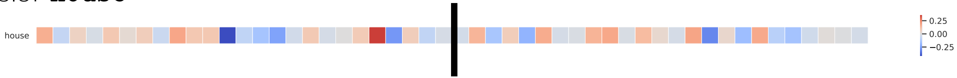
Word Embeddings: Example

◇ Example: `house`

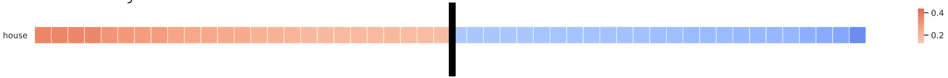


Word Embeddings: Example

- Example: `house`

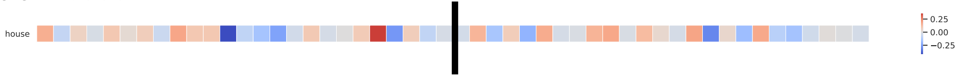


- Syntactic and semantic properties
  - Similarity

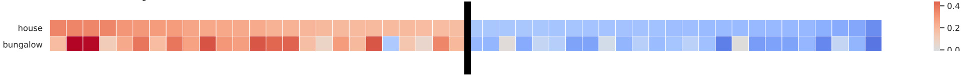


Word Embeddings: Example

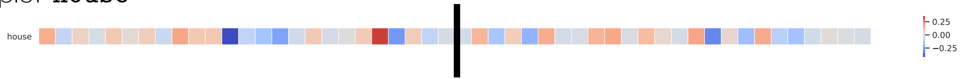
- Example: `house`



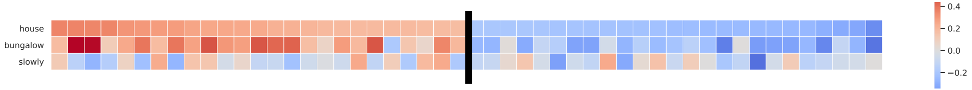
- Syntactic and semantic properties
  - Similarity



- Example: `house`

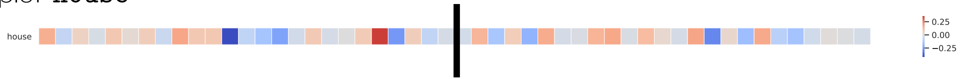


- Syntactic and semantic properties
  - Similarity



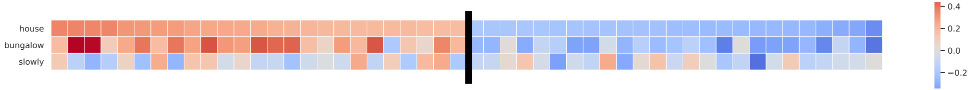
Word Embeddings: Example

◇ Example: `house`

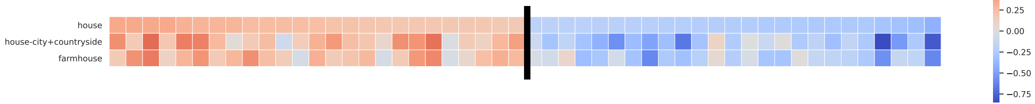


◇ Syntactic and semantic properties

– Similarity

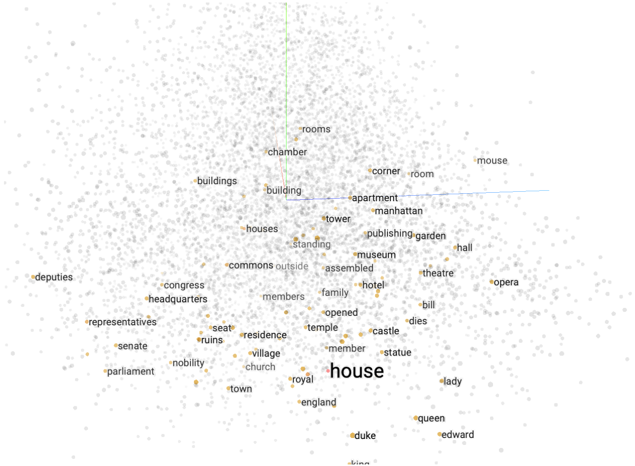


– Analogy



Word Embeddings: Similarity

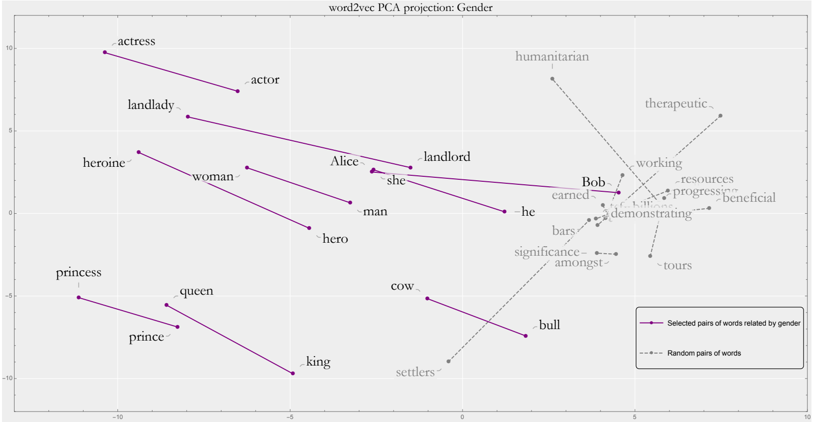
house	cosine distance
houses	0.292761
bungalow	0.312144
apartment	0.3371
bedroom	0.350306
townhouse	0.361592
residence	0.380158
mansion	0.394181
farmhouse	0.414243
duplex	0.424206
homes	0.43802



(https://projector.tensorflow.org)

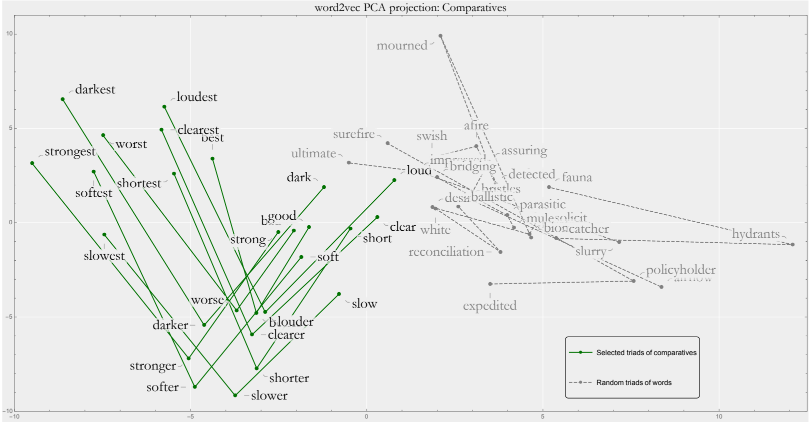
Word Embeddings: Analogy

$$v_{king} - v_{queen} \approx v_{hero} - v_{heroine}$$

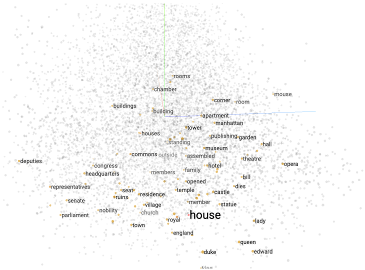


Word Embeddings: Analogy

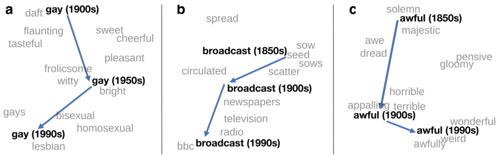
$$v_{good} - v_{better} \approx v_{soft} - v_{softer}$$



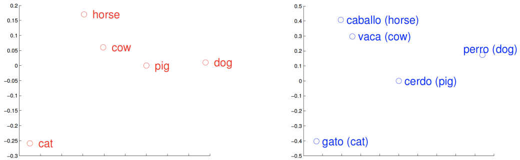
Embedding Space



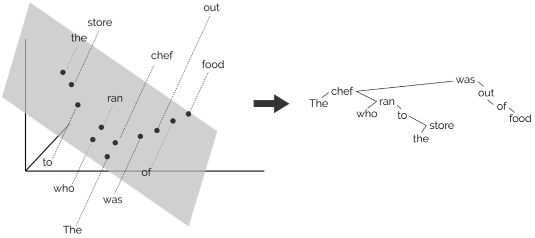
(https://projector.tensorflow.org)



(Hamilton et al., 2016)



(Mikolov et al., 2013)



(https://nlp.stanford.edu/~johnhew/structural-probe.html)

Word Embeddings as Matrix Factorization

- Word2vec performs an implicit factorization of a word-context matrix (Levy and Goldberg, 2014)
  - (shifted) pointwise mutual information (PMI)
  - Truncated SVD to reduce dimensionality

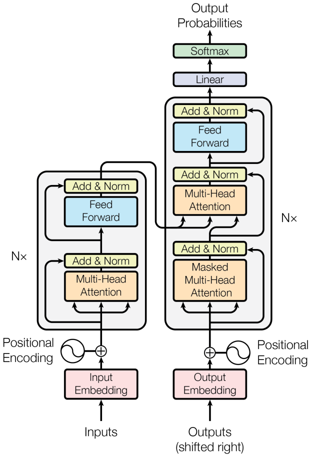
- Equivalent results can be achieved with explicit vector representations (Levy et al., 2015)

Word Embeddings as Matrix Factorization

- Word2vec performs an implicit factorization of a word-context matrix (Levy and Goldberg, 2014)
  - (shifted) pointwise mutual information (PMI)
  - Truncated SVD to reduce dimensionality

- Equivalent results can be achieved with explicit vector representations (Levy et al., 2015)

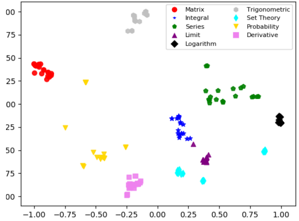
- More complex architectures (e.g. Transformers, Vaswani et al., 2017) are based on these representations for elementary units.



(Vaswani et al., 2017)

Mathematical Embeddings

- Several works on mathematical embeddings: (Gao et al., 2017; Greiner-Petter et al., 2019, 2020; Krstovski and Blei, 2018; Mansouri et al., 2019; Naik et al., 2019; Purgał et al., 2021; Ryskina and Knight, 2021; Thawani et al., 2021)



(Mansouri et al., 2019)

- At least two reasons why it seems insufficient
  - Lack of focus on the operational content of expressions.
    - 406 added to 326 equals 732
    - $A \wedge B$ is likely to be a premise in the proof of some given logical statement
    - $y'' - y = 0$ accepts the solution $y(x) = c_1 e^x + c_2 e^{-x}$
  - Embedding techniques are adopted uncritically

Dimensions of Formal Content

Formal Content: the dimension of content which finds its source in the internal relations holding between the expressions of a language.

Dimensions of Formal Content

Formal Content: the dimension of content which finds its source in the internal relations holding between the expressions of a language.

- *Syntactic Content*: the content a unit receives as a result of the multiple dependencies it can maintain with respect to other units in its context

Dimensions of Formal Content

Formal Content: the dimension of content which finds its source in the internal relations holding between the expressions of a language.

- *Syntactic Content*: the content a unit receives as a result of the multiple dependencies it can maintain with respect to other units in its context
- *Characteristic Content*: the content resulting from the inclusion of a unit in a class of other units by which it accepts to be substituted in given contexts

Dimensions of Formal Content

Formal Content: the dimension of content which finds its source in the internal relations holding between the expressions of a language.

- *Syntactic Content*: the content a unit receives as a result of the multiple dependencies it can maintain with respect to other units in its context
- *Characteristic Content*: the content resulting from the inclusion of a unit in a class of other units by which it accepts to be substituted in given contexts
- *Informational Content*: the content related to the non-uniform distribution of units within those substitutability classes

Dimensions of Formal Content

Syntactic Content	Characteristic Content	Informational Content
"the gavagai is on the mat"	{cat, dog, spider, gavagai}	{cat:0.059%, dog:0.012%, spider:0.009% gavagai:0.000%}
Type Theory	Clustering	Probability and Information Theory
Type	Class	Probability Distribution

Outline

Arithmetical Content

- How is it possible that a distributional approach to (natural) language can account for the mathematical content of mathematical expressions?

Arithmetical Content

- How is it possible that a distributional approach to (natural) language can account for the mathematical content of mathematical expressions?

- Illustration: recursive structure and total order of natural numbers

Arithmetical Content

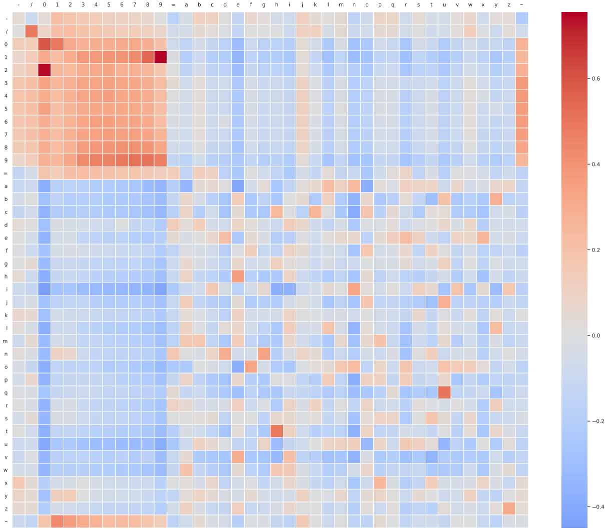
- How is it possible that a distributional approach to (natural) language can account for the mathematical content of mathematical expressions?

- Illustration: recursive structure and total order of natural numbers

- The task is to identify:
  - Class of numerals as an autonomous class among all character strings (characteristic content)
  - Iterative construction principle and self-similar syntactic embedding (syntactic content)
  - Probability distribution characterizing the order of all elements in the class of numerals (informational content)

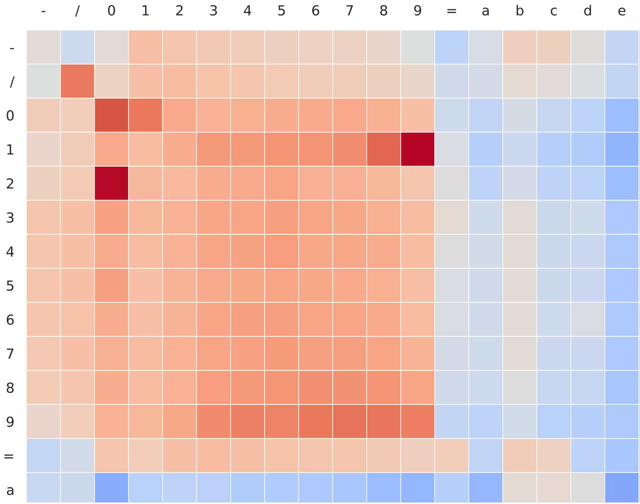
The Class of Numerals

$$A_{i,j} = pmi(c_i; c_j) = \log \frac{p(c_i, c_j)}{p(c_i)p(c_j)}$$

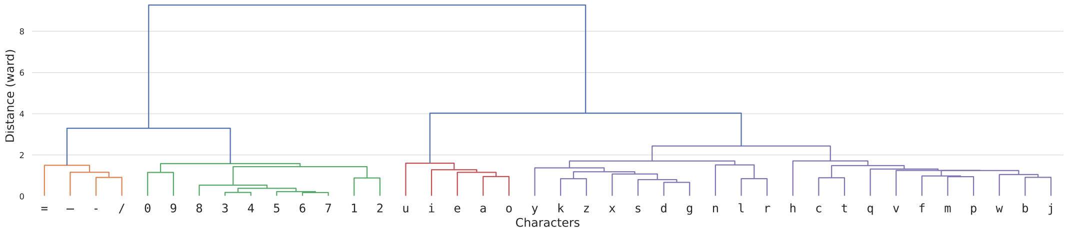


The Class of Numerals

$$A_{i,j} = pmi(c_i; c_j) = \log \frac{p(c_i, c_j)}{p(c_i)p(c_j)}$$



Clustering



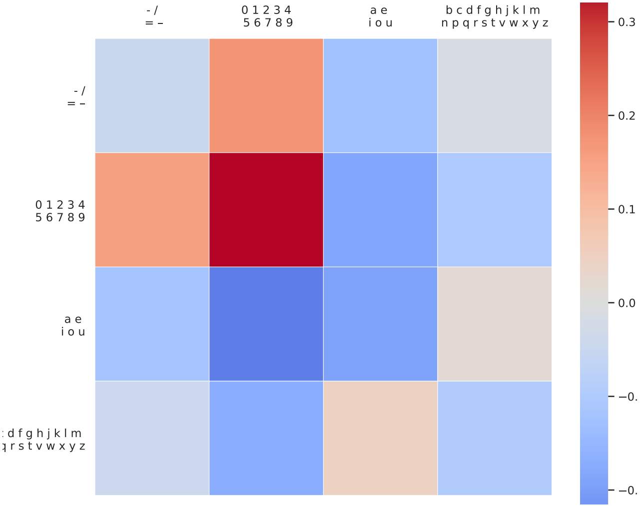
$$O := \{=, -, \text{-}, /\}$$

$$D := \{0, 9, 8, 3, 4, 5, 6, 7, 1, 2\}$$

$$V := \{\mathtt{u}, \mathtt{i}, \mathtt{e}, \mathtt{a}, \mathtt{o}\}$$

$$C := \{\mathtt{y}, \mathtt{k}, \mathtt{z}, \mathtt{x}, \mathtt{s}, \mathtt{d}, \mathtt{g}, \mathtt{n}, \mathtt{l}, \mathtt{r}, \mathtt{h}, \mathtt{c}, \mathtt{t}, \mathtt{q}, \mathtt{v}, \mathtt{f}, \mathtt{m}, \mathtt{p}, \mathtt{w}, \mathtt{b}, \mathtt{j}\}$$

Compressed Matrix



$$\overrightarrow{|\mathtt{d}|}_{\mathtt{d}\in D} := \overrightarrow{\mathtt{d}} - \overrightarrow{D}$$

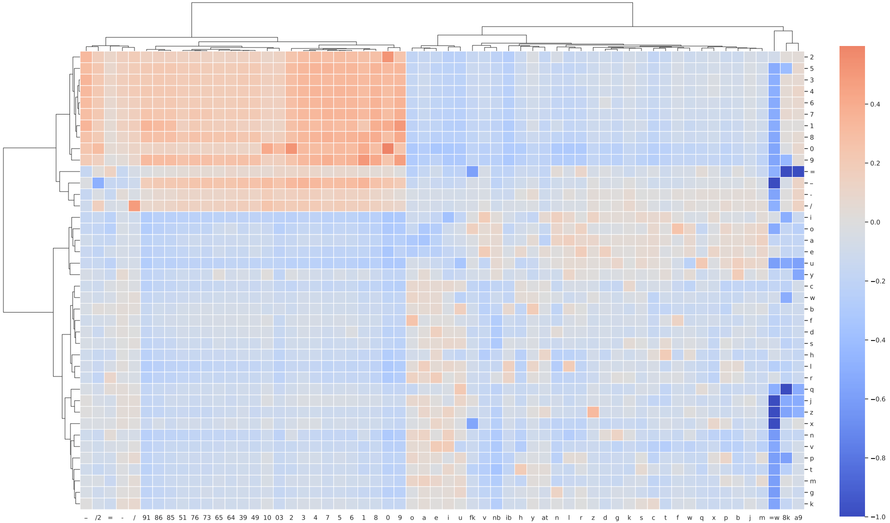
$$\overrightarrow{\mathtt{d}} = \overrightarrow{D} + \overrightarrow{|\mathtt{d}|}$$

$$f(\overrightarrow{D} + \overrightarrow{|\mathtt{d}_0|}) = \overrightarrow{D} + \overrightarrow{|\mathtt{d}_1|}$$

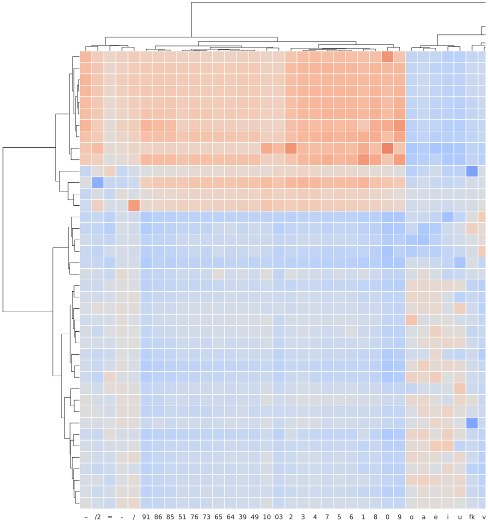
$$f = T \circ t$$

$$T(\overrightarrow{D}) = \overrightarrow{D}$$

Self-Similar Syntactic Embedding

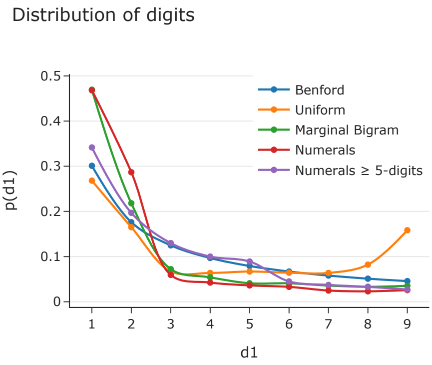


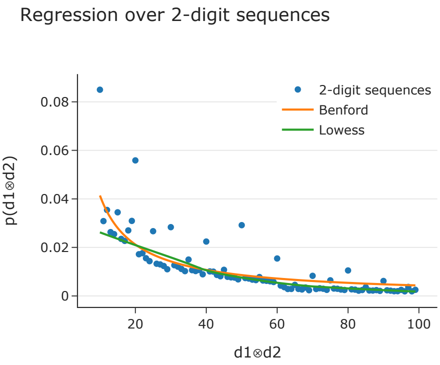
Self-Similar Syntactic Embedding



$$\overrightarrow{D \otimes D} \backsimeq \overrightarrow{D}$$

Total Order Through Benford's Law





Conclusions

- Semantic features of natural numbers could be derived from the distributional properties of syntax by means of tools associated to natural language processing
  - Maybe also other mathematical contents?
- Distributional approaches provide an original perspective on mathematical contents, unseen within the philosophy of mathematics
- Potentially useful for the history and the philosophy of scientific practices, due to the central role of the analysis of corpora
- A philosophical account of ML results can articulate the need for the explicit derivation of structural features underlying the syntactic data. We need to move from a distributional to a structuralist conception of language.

Reference Paper

Gastaldi, J. L., Content from Expressions: The Place of Textuality in Deep Learning Approaches to Mathematics. Under review at *Synthese. SI: Linguistically Informed Philosophy of Mathematics.* Fisseni, B., Kant, D., Sarikaya, D. and Schröder, B. (Eds.).

References I

Alemi, A. A., Chollet, F., Een, N., Irving, G., Szegedy, C., & Urban, J. (2016). Deepmath - deep sequence models for premise selection. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2243–2251.

Bansal, K., Loos, S. M., Rabe, M. N., Szegedy, C., & Wilcox, S. (2019). Holist: An environment for machine learning of higher-order theorem proving (extended version). *CoRR*, *abs/1904.03241*. http://arxiv.org/abs/1904.03241

Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A neural probabilistic language model. *J. Mach. Learn. Res.*, *3*, 1137–1155.

Blechschmidt, J., & Ernst, O. G. (2021). Three ways to solve partial differential equations with neural networks — a review. *GAMM-Mitteilungen*, *44*(2), e202100006. https://doi.org/https://doi.org/10.1002/gamm.202100006

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners.

Charton, F. (2021). Linear algebra with transformers. *CoRR*, *abs/2112.01898*. https://arxiv.org/abs/2112.01898

d'Ascoli, S., Kamienny, P., Lample, G., & Charton, F. (2022). Deep symbolic regression for recurrent sequences. *CoRR*, *abs/2201.04600*.

Davies, A., Veličković, P., Buesing, L., Blackwell, S., Zheng, D., Tomašev, N., Tanburn, R., Battaglia, P., Blundell, C., Juhász, A., Lackenby, M., Williamson, G., Hassabis, D., & Kohli, P. (2021). Advancing mathematics by guiding human intuition with AI. *Nature*, *600*(7887), 70–74. https://doi.org/10.1038/s41586-021-04086-x

Firth, J. R. (1935). The technique of semantics. *Transactions of the Philological Society*, *34*(1), 36–73. https://doi.org/10.1111/j.1467-968X.1935.tb01254.x

References II

Gao, L., Jiang, Z., Yin, Y., Yuan, K., Yan, Z., & Tang, Z. (2017). Preliminary exploration of formula embedding for mathematical information retrieval: Can mathematical formulae be embedded like a natural language? *CoRR*, *abs/1707.05154*. http://arxiv.org/abs/1707.05154

Greiner-Petter, A., Ruas, T., Schubotz, M., Aizawa, A., Grosky, W. I., & Gipp, B. (2019). Why machines cannot learn mathematics, yet. *CoRR*, *abs/1905.08359*. http://arxiv.org/abs/1905.08359

Greiner-Petter, A., Youssef, A., Ruas, T., Miller, B. R., Schubotz, M., Aizawa, A., & Gipp, B. (2020). Math-word embedding in math search and semantic extraction. *Scientometrics*, *125*(3), 3017–3046. https://doi.org/10.1007/s11192-020-03502-9

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. *CoRR*, *abs/1605.09096*.

Harris, Z. (1960). *Structural linguistics*. University of Chicago Press.

Kaliszyk, C., Chollet, F., & Szegedy, C. (2017). Holstep: A machine learning dataset for higher-order logic theorem proving. *CoRR*, *abs/1703.00426*. http://arxiv.org/abs/1703.00426

Krstovski, K., & Blei, D. M. (2018). Equation embeddings.

Lample, G., & Charton, F. (2019). Deep learning for symbolic mathematics.

Levy, O., & Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, 2177–2185.

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, *3*, 211–225. https://doi.org/10.1162/tacl_a_00134

References III

Li, Z., Kovachki, N. B., Azizzadenesheli, K., liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2021). Fourier neural operator for parametric partial differential equations. *International Conference on Learning Representations*. https://openreview.net/forum?id=c8P9NQVtmnO

Mansouri, B., Rohatgi, S., Oard, D. W., Wu, J., Giles, C. L., & Zanibbi, R. (2019). Tangent-cft: An embedding model for mathematical formulas. *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, 11–18. https://doi.org/10.1145/3341981.3344235

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J., Le, Q., & Strohmann, T. (2013). *Learning representations of text using neural networks. NIPS deep learning workshop 2013 slides*.

Naik, A., Ravichander, A., Rose, C., & Hovy, E. (2019). Exploring numeracy in word embeddings. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3374–3380. https://doi.org/10.18653/v1/P19-1329

Nogueira, R., Jiang, Z., & Lin, J. (2021). Investigating the limitations of the transformers with simple arithmetic tasks. *CoRR*, *abs/2102.13019*. https://arxiv.org/abs/2102.13019

Peng, S., Yuan, K., Gao, L., & Tang, Z. (2021). Mathbert: A pre-trained model for mathematical formula understanding. *CoRR*, *abs/2105.00377*. https://arxiv.org/abs/2105.00377

Purgał, S., Parsert, J., & Kaliszyk, C. (2021). A study of continuous vector representations for theorem proving. *Journal of Logic and Computation*, *31*(8), 2057–2083. https://doi.org/10.1093/logcom/exab006

Ryskina, M., & Knight, K. (2021). Learning mathematical properties of integers. *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, 389–395. https://doi.org/10.18653/v1/2021.blackboxnlp-1.30

References IV

Saussure, F. d. (1959). *Course in general linguistics* [Translated by Wade Baskin]. McGraw-Hill.

Shen, J. T., Yamashita, M., Prihar, E., Heffernan, N. T., Wu, X., & Lee, D. (2021). Mathbert: A pre-trained language model for general NLP tasks in mathematics education. *CoRR*, *abs/2106.07340*.

Thawani, A., Pujara, J., Ilievski, F., & Szekely, P. (2021). Representing numbers in NLP: A survey and a vision. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 644–656. https://doi.org/10.18653/v1/2021.naacl-main.53

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need.

Research Seminar
Cohn Institute
for the History and Philosophy of Science and Ideas
Tel Aviv, Israel

The Language of Mathematics

Epistemological Consequences of Applying AI Methods to Mathematics

Juan Luis Gastaldi

ETH zürich

April 17th, 2023

This project has received funding from the
European Union's Horizon 2020 research and innovation programme
under grant agreement No 839730