PhilML'23
Eberhard Karls University of Tübingen
Tübingen, Germany.
September 12-14, 2023

The Language of Mathematics

Epistemological Consequences of the Application of Neural Models to Mathematics

Juan Luis Gastaldi

ETH zürich

September 13, 2023

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

Reference Papers

- Gastaldi, J. L., & Pellissier, L. (2021). The calculus of language: Explicit representation of emergent linguistic structure through type-theoretical paradigms. *Interdisciplinary Science Reviews*. https://doi.org/10.1080/03080188.2021.1890484

- Gastaldi, J. L. (Forthcoming 2004b). How to Do Maths with Words. Neural Machine Learning Applications to Mathematics and Their Philosophical Significance. In B. Sriraman (Ed.), *Handbook of the History and Philosophy of Mathematical Practice*. Springer

- Bradley, T.-D., Gastaldi, J. L., & Terilla, J. (Forthcoming 2024a). The structure of meaning in language: parallel narratives in linear algebra and category theory. *Notices of the AMS*

- Gastaldi, J. L. (Forthcoming 2024c). Content from Expressions. The Place of Textuality in Deep Learning Approaches to Mathematics. *Synthese (under review)*

Outline

Neural ML and Mathematics

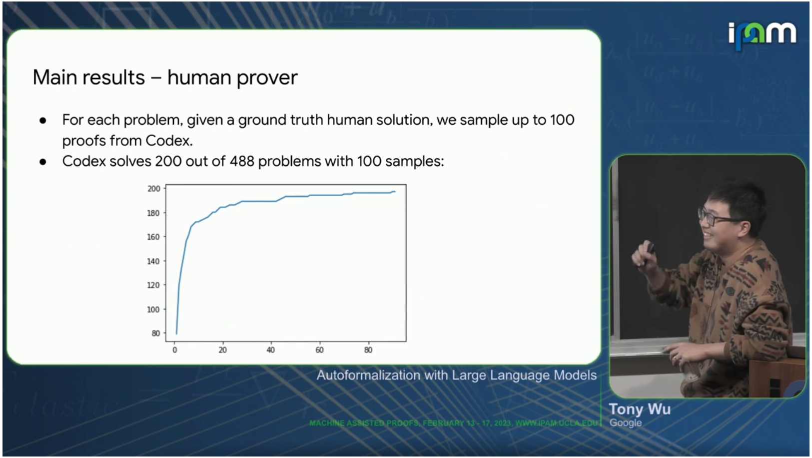
Epistemology of Natural Language Processing (NLP)

Distributional Arithmetic

Takeaways

Outline

Tony Wu at the IPAM



Tony Wu, *Autoformalization with Large Language Models* (IPAM (UCLA), Feb 15, 2023)

Melanie Mitchell on PaLM2

Melanie Mitchell
@MelMitchell1

Weird statement from Google's Palm 2 announcement.

(from blog.google/technology/ai/...)

- **Reasoning:** PaLM 2's wide-ranging dataset includes scientific papers and web pages that contain mathematical expressions. As a result, it demonstrates improved capabilities in logic, common sense reasoning, and mathematics.

6:23 PM · May 12, 2023 · **52.9K** Views

https://blog.google/technology/ai/google-palm-2-ai-large-language-model/

What’s So Funny?

- Proof assistants and automated theorem provers were supposed to prevent the frailty of human proving practices instead of reproducing them and relying on them as a gold standard.

What's So Funny?

- Proof assistants and automated theorem provers were supposed to prevent the frailty of human proving practices instead of reproducing them and relying on them as a gold standard.
- Mathematical competence was supposed to be a condition to write and read scientific papers and mathematical expressions instead of their miraculous effect.

What's So Funny?

- Proof assistants and automated theorem provers were supposed to prevent the frailty of human proving practices instead of reproducing them and relying on them as a gold standard.
- Mathematical competence was supposed to be a condition to write and read scientific papers and mathematical expressions instead of their miraculous effect.
- Natural language was considered the cause of rather than the solution to the multiple problems preventing mathematics from achieving higher degrees of precision.

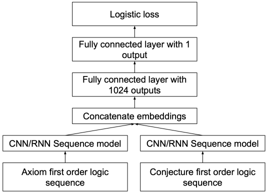
What's So Funny?

- Proof assistants and automated theorem provers were supposed to prevent the frailty of human proving practices instead of reproducing them and relying on them as a gold standard.
- Mathematical competence was supposed to be a condition to write and read scientific papers and mathematical expressions instead of their miraculous effect.
- Natural language was considered the cause of rather than the solution to the multiple problems preventing mathematics from achieving higher degrees of precision.
- More generally, the *formal* nature of mathematics was believed to make it impassive to the strong empirical position assumed by connectionist approaches guiding the application of DNNs.

Main Orientations in DNN Applications to Maths

(Gastaldi, Forthcoming 2004b)

- Proof-Oriented
  - Bansal et al., 2019; Polu and Sutskever, 2020; Wu et al., 2022.

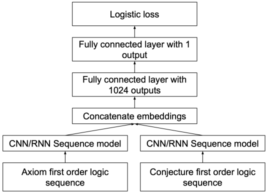


(Alemi et al., 2016)

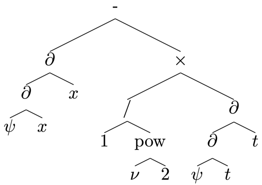
Main Orientations in DNN Applications to Maths

(Gastaldi, Forthcoming 2004b)

- Proof-Oriented
  - Bansal et al., 2019; Polu and Sutskever, 2020; Wu et al., 2022.
- Object-Oriented
  - Blechschmidt and Ernst, 2021; Charton, 2021; d'Ascoli et al., 2022; Lample and Charton, 2019; Li et al., 2021



(Alemi et al., 2016)

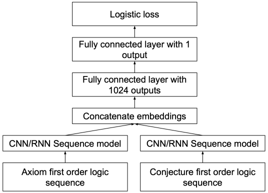


(Lample and Charton, 2019)

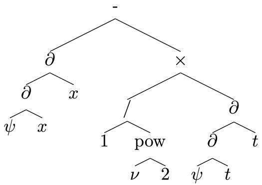
Main Orientations in DNN Applications to Maths

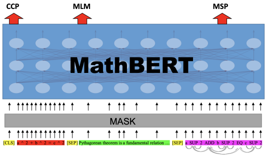
(Gastaldi, Forthcoming 2004b)

- Proof-Oriented
  - Bansal et al., 2019; Polu and Sutskever, 2020; Wu et al., 2022.
- Object-Oriented
  - Blechschmidt and Ernst, 2021; Charton, 2021; d'Ascoli et al., 2022; Lample and Charton, 2019; Li et al., 2021
- Skill-Oriented
  - Brown et al., 2020; Lewkowycz et al., 2022; Shen et al., 2021



(Alemi et al., 2016)


(Lample and Charton, 2019)

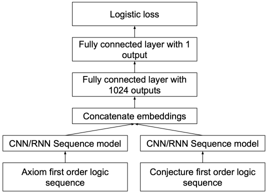


(Peng et al., 2021)

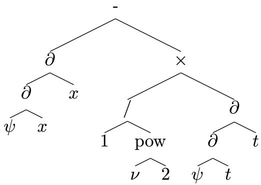
Main Orientations in DNN Applications to Maths

(Gastaldi, Forthcoming 2004b)

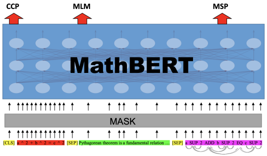
- Proof-Oriented
  - Bansal et al., 2019; Polu and Sutskever, 2020; Wu et al., 2022.
- Object-Oriented
  - Blechschmidt and Ernst, 2021; Charton, 2021; d'Ascoli et al., 2022; Lample and Charton, 2019; Li et al., 2021
- Skill-Oriented
  - Brown et al., 2020; Lewkowycz et al., 2022; Shen et al., 2021
- Heuristic-Oriented
  - Davies et al., 2021; Wagner, 2021



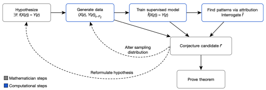
(Alemi et al., 2016)



(Lample and Charton, 2019)



(Peng et al., 2021)



(Davies et al., 2021)

Philosophical Significance: The Return of Language

- DNN applications to mathematics lack of practically any relation to philosophical inquiries.

Philosophical Significance: The Return of Language

- DNN applications to mathematics lack of practically any relation to philosophical inquiries.
- Research orientations tend to be spontaneously organized according to the AI researchers' implicit assumptions as to *what it is that we do when we do mathematics*.

Philosophical Significance: The Return of Language

- DNN applications to mathematics lack of practically any relation to philosophical inquiries.
- Research orientations tend to be spontaneously organized according to the AI researchers' implicit assumptions as to *what it is that we do when we do mathematics*.
- However, practically all applications share a common philosophical assumption: *Natural language* plays a critical role in the processing mathematical knowledge.

Philosophical Significance: The Return of Language

- DNN applications to mathematics lack of practically any relation to philosophical inquiries.
- Research orientations tend to be spontaneously organized according to the AI researchers' implicit assumptions as to *what it is that we do when we do mathematics*.
- However, practically all applications share a common philosophical assumption: *Natural language* plays a critical role in the processing mathematical knowledge.
- The potential success of DNN methods in mathematics is inseparable from a reorientation of the epistemology of mathematics from logic and formal systems to natural language practice.

Outline

Neural ML and Mathematics

Epistemology of Natural Language Processing (NLP)

Distributional Arithmetic

Takeaways

Stochastic Parrots vs. AI Consciousness

📉🦜

Language models are not like us,
therefore they do not and can not have any relation to meaning.

🤖🧠

Language models have a relation to meaning,
therefore they are like us.

Stochastic Parrots vs. AI Consciousness

📉🦜

Language models are not like us,
therefore they do not and can not have any relation to meaning.

🤖🧠

Language models have a relation to meaning,
therefore they are like us.

- Two tools to resist to this alternative:
  - Conceptual: Operational notion of *formal content*
  - Technical: Distributionalism through *algebraic word vector representations*

Formal Content

(Gastaldi and Pellissier, 2021;Gastaldi, Forthcoming 2024c)

Form ~~vs.~~ and ~~Meaning~~ Content

Kant, Hegel, Frege, Saussure, Hjelmslev, etc.

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

Illustration of Formal Contents

(Gastaldi and Pellissier, 2021;Gastaldi, Forthcoming 2024c)

Characteristic Content

{cat, dog, spider, gavagai}

Clustering

Class

Syntactic Content

"the gavagai is on the mat"

Type Theory

Type

Informational Content

{cat:0.059%, dog:0.012%, spider:0.009% gavagai:0.000%}

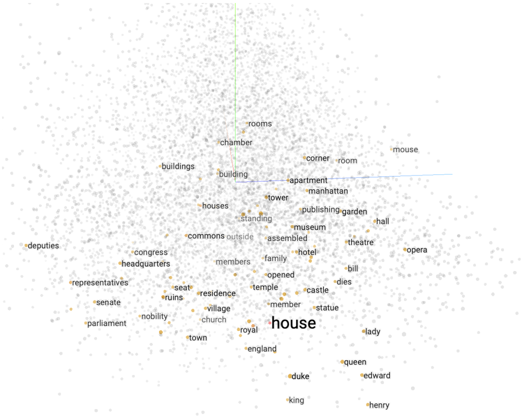
Probability and Information Theory

Probability Distribution

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)

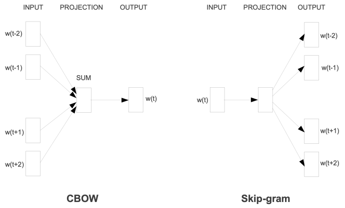
- Computational interpretation: Word Embeddings

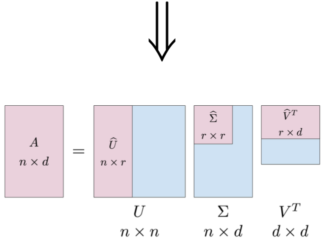


(https://projector.tensorflow.org)

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

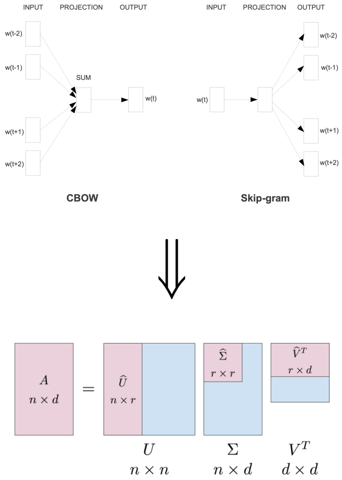




Word Embeddings as Matrix Factorization

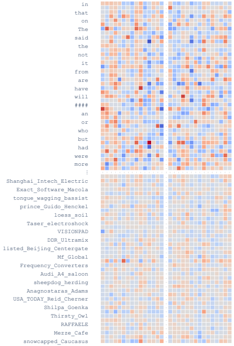
- Word2vec performs an implicit factorization of a word-context matrix (Levy and Goldberg, 2014)
  - (shifted) pointwise mutual information (PMI)
  - Truncated SVD to reduce dimensionality

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



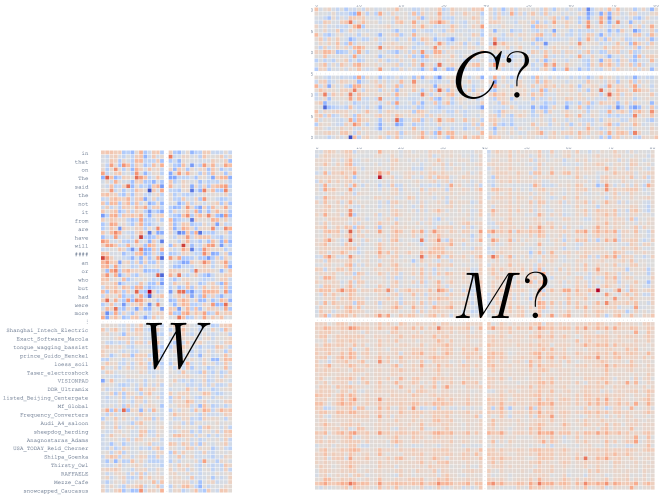
Word Embeddings as Matrix Factorization

(Levy and Goldberg, 2014)



Word Embeddings as Matrix Factorization

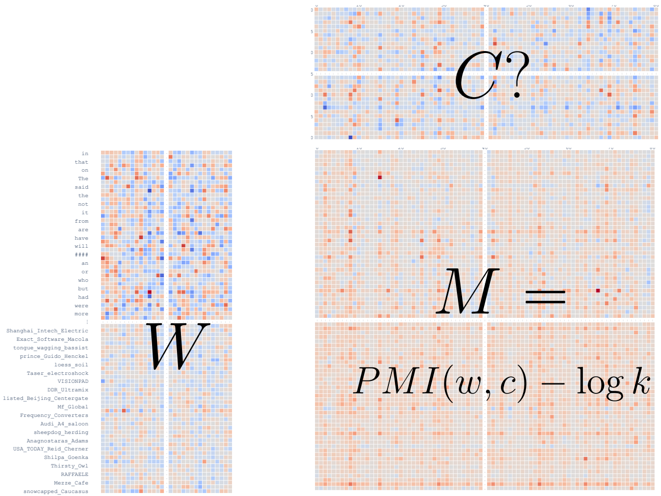
(Levy and Goldberg, 2014)



$$W \times C \approx M$$

Word Embeddings as Matrix Factorization

(Levy and Goldberg, 2014)



$$W \times C \approx M$$

$$PMI(w,c) = \log \frac{p(w,c)}{p(w)p(c)}$$

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

Recursion through Peano Axioms

1. 0 is a number.

2. If n is a number, the successor of n is a number.

3. 0 is not the successor of a number.

4. Two numbers of which the successors are equal are themselves equal.

5. If a set $\mathbf{S}$ of numbers contains 0 and also the successor of every number in $\mathbf{S}$, then every number is in $\mathbf{S}$ (induction axiom).

Recursion through Peano Axioms

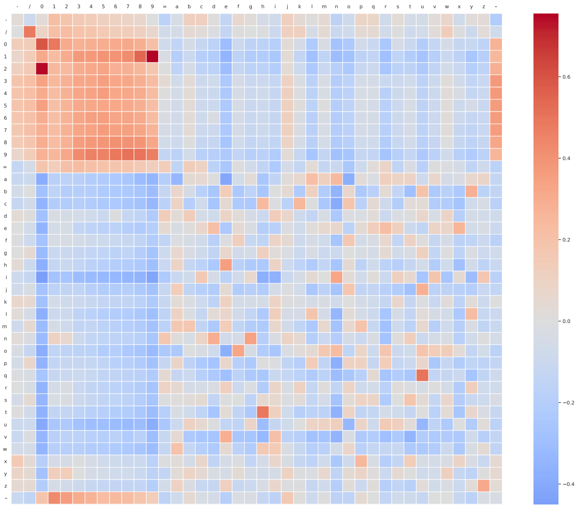
1. 0 is a number.
   $0 \in \mathbb{N}$
2. If n is a number, the successor of n is a number.
   $n \in \mathbb{N} \implies \text{succ}(n) \in \mathbb{N}$
3. 0 is not the successor of a number.
   $\forall n \in \mathbb{N}, 0 \neq \text{succ}(n)$
4. Two numbers of which the successors are equal are themselves equal.
   $\forall n, m \in N, \text{succ}(x) = \text{succ}(y) \implies x = y$
5. If a set $\mathbf{S}$ of numbers contains 0 and also the successor of every number in $\mathbf{S}$, then every number is in $\mathbf{S}$ (induction axiom).
   $0 \in S \wedge (\forall n, n \in S \implies \text{succ}(s) \in \mathbf{S}) \implies \forall n \in \mathbf{S}, n \in \mathbb{N}$

Recursion through Peano Axioms

1. 0 is a number.
   $0 \in \mathbb{N}$
2. If n is a number, the successor of n is a number.
   $n \in \mathbb{N} \implies \text{succ}(n) \in \mathbb{N}$
3. 0 is not the successor of a number.
   $\forall n \in \mathbb{N}, 0 \neq \text{succ}(n)$
4. Two numbers of which the successors are equal are themselves equal.
   $\forall n, m \in N, \text{succ}(x) = \text{succ}(y) \implies x = y$
5. If a set $\mathbf{S}$ of numbers contains $\mathbf{0}$ and also the successor of every number in $\mathbf{S}$, then every number is in $\mathbf{S}$ (induction axiom).
   $0 \in S \wedge (\forall n, n \in S \implies \text{succ}(s) \in \mathbf{S}) \implies \forall n \in \mathbf{S}, n \in \mathbb{N}$

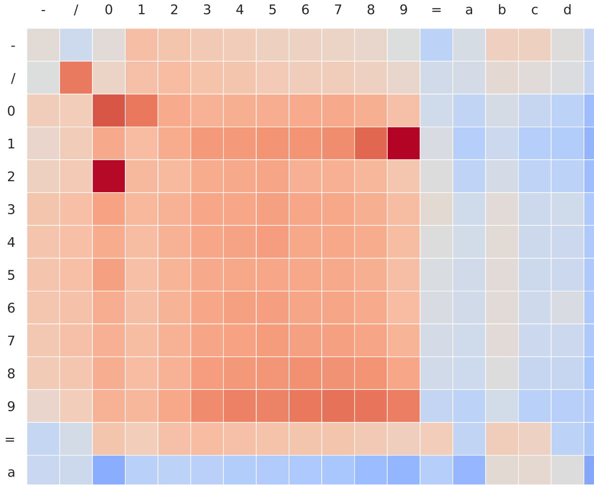
The Distributional Properties of Characters

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



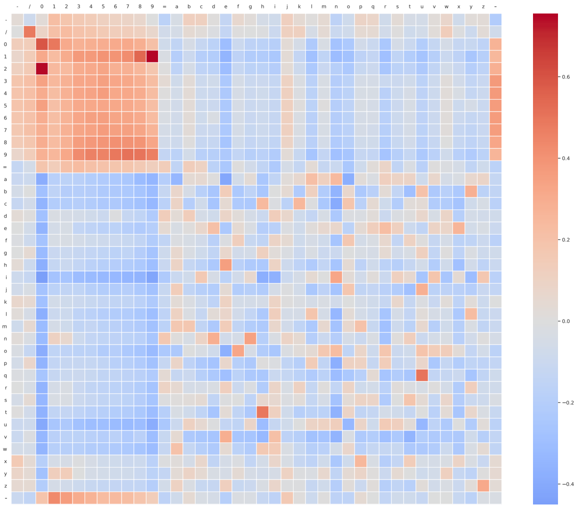
The Distributional Properties of Characters

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

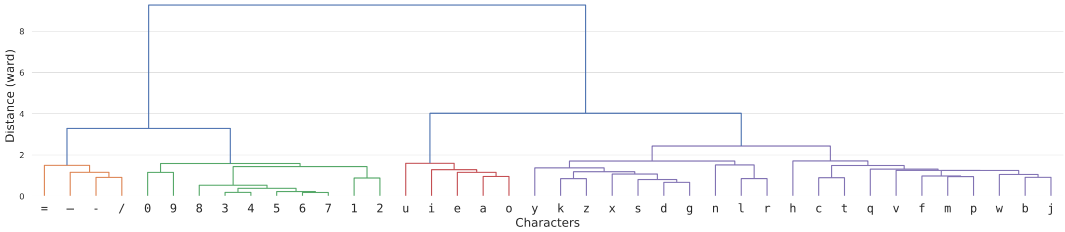


The Distributional Properties of Characters

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



The Characteristic Content of Digits



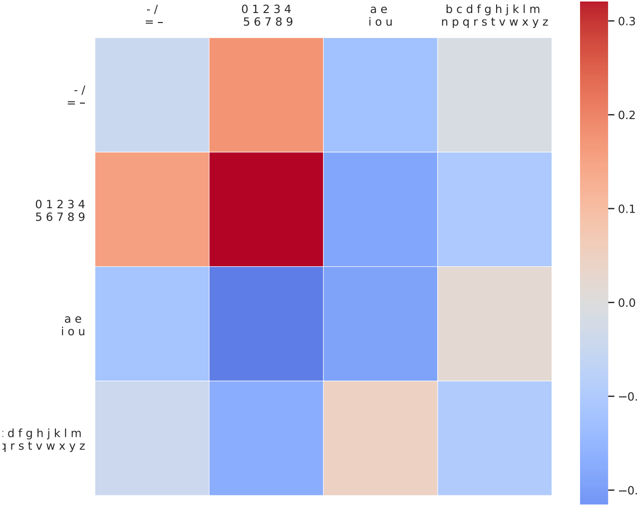
$$O := \{=, -, \text{-}, /\}$$

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

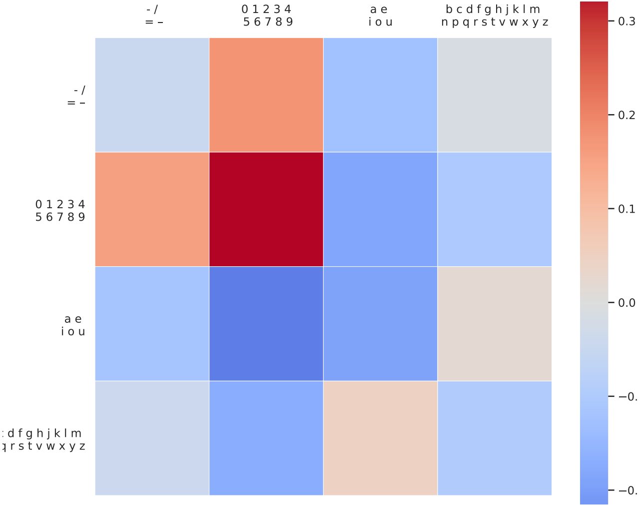
$$V := \{\mathtt{u}, \mathtt{i}, \mathtt{e}, \mathtt{a}, \mathtt{o}\}$$

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

The Syntactic Content of Digits

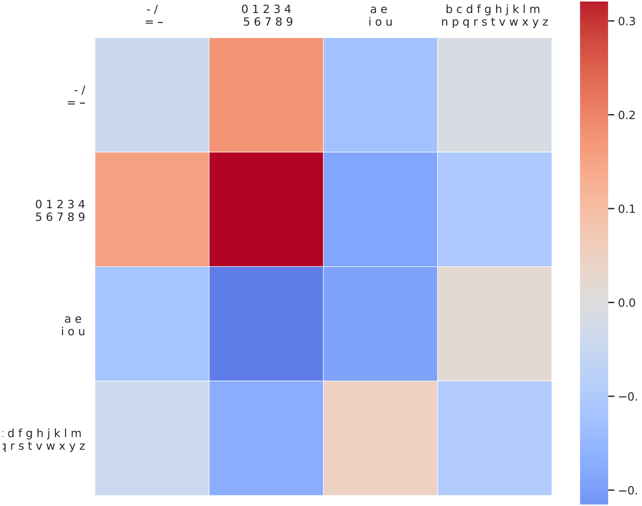


The Syntactic Content of Digits



$$f(c_n) = c_{n+1}$$

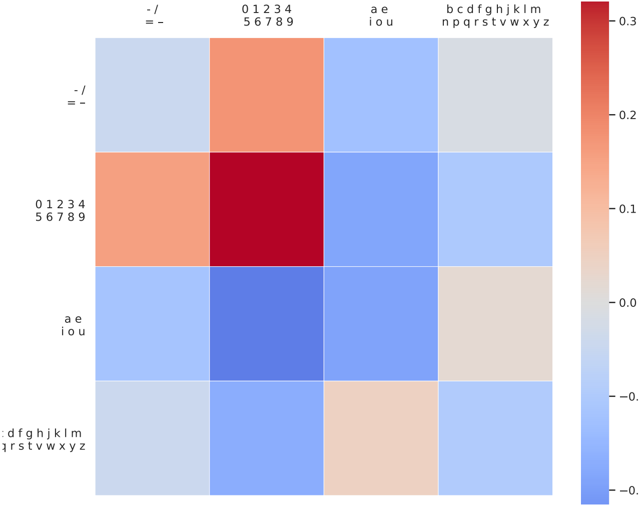
The Syntactic Content of Digits



$$f(c_n) = c_{n+1}$$

$$f(D) = D$$

The Syntactic Content of Digits



$$f(c_n) = c_{n+1}$$

$$f(D) = D$$

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

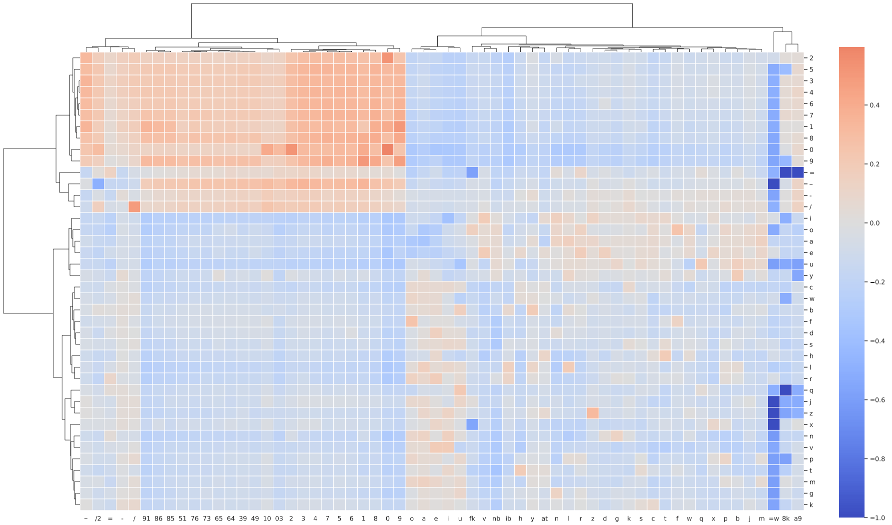
$$f = T \circ t$$

$$T(D) = D$$

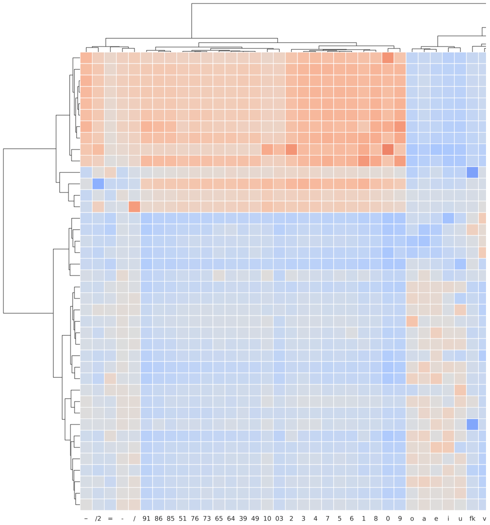
Concatenation and Characteristic Content of $\mathbb{N}$

$$
\begin{aligned}
&D \times f(D) \\
&= D \times D
\end{aligned}
$$

Concatenation and Characteristic Content of ℕ

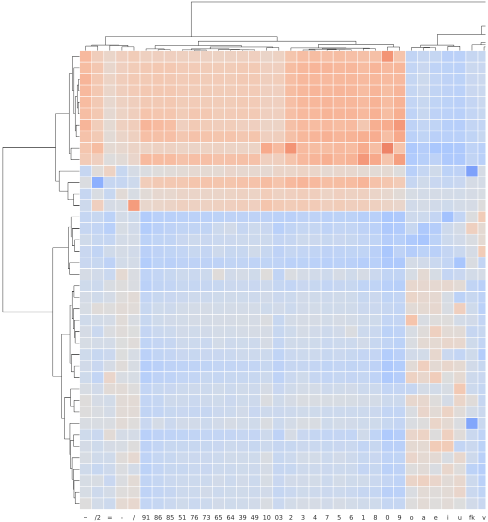


Concatenation and Characteristic Content of $\mathbb{N}$



$$D \times f(D)$$
$$= D \times D$$

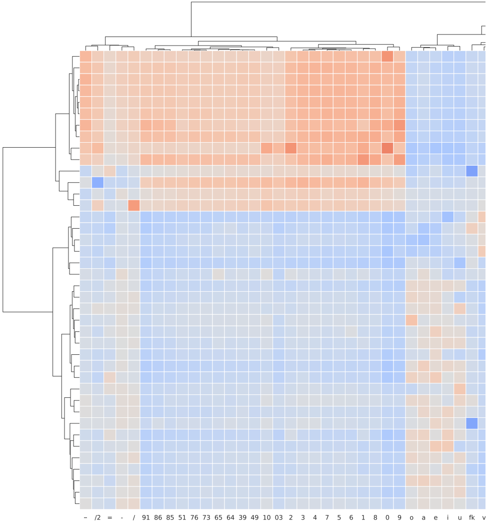
Concatenation and Characteristic Content of $\mathbb{N}$



$$D \times f(D)$$
$$= D \times D$$

$$f(\overrightarrow{D \times D}) \approx D$$

Concatenation and Characteristic Content of $\mathbb{N}$

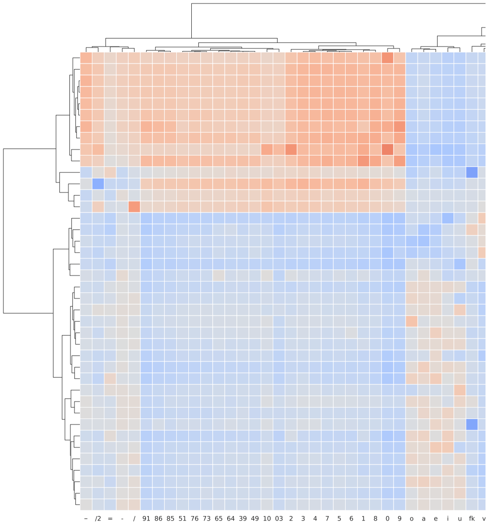


$$D \times f(D)$$
$$= D \times D$$

$$f(\overrightarrow{D \times D}) \approx D$$

$$D^2 = (D \times D) \cup D$$

Concatenation and Characteristic Content of $\mathbb{N}$



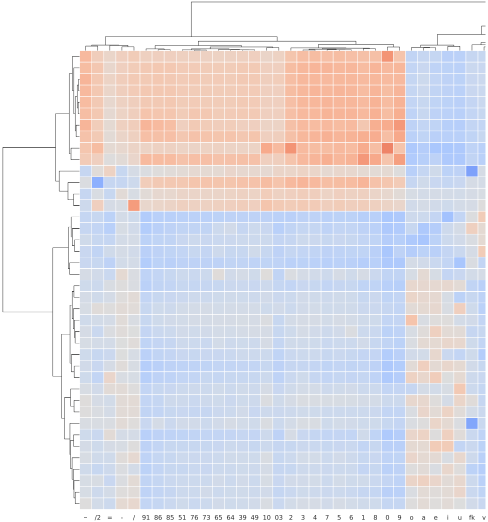
$$D \times f(D)$$
$$= D \times D$$

$$f(\overrightarrow{D \times D}) \approx D$$

$$D^2 = (D \times D) \cup D$$

$$f_2(D^2) = D^2$$

Concatenation and Characteristic Content of $\mathbb{N}$



$$D \times f(D)$$
$$= D \times D$$

$$f(\overrightarrow{D \times D}) \approx D$$

$$D^2 = (D \times D) \cup D$$

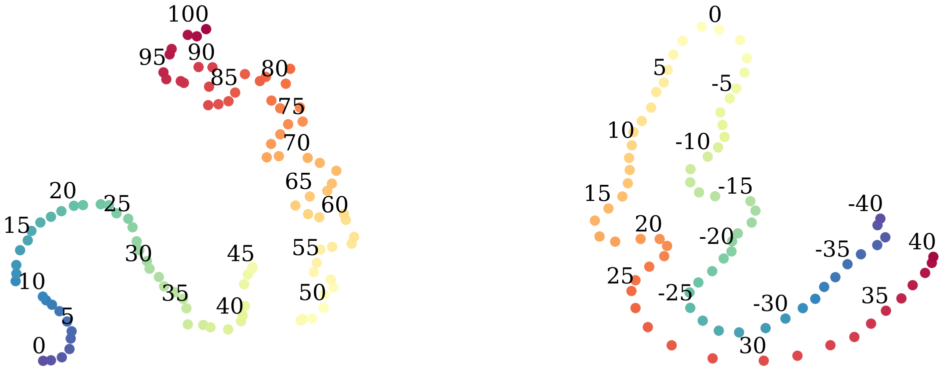
$$f_2(D^2) = D^2$$

$$\vdots$$

$$f_*(D^*) = D^*$$

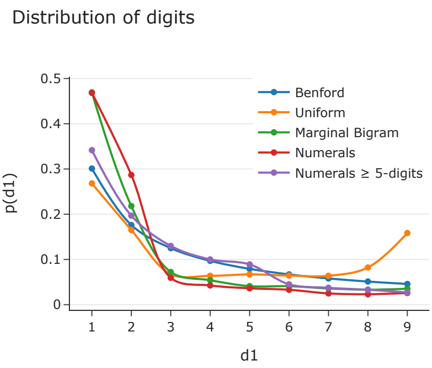
$$\mathbb{N} = D^* \approx D^n$$

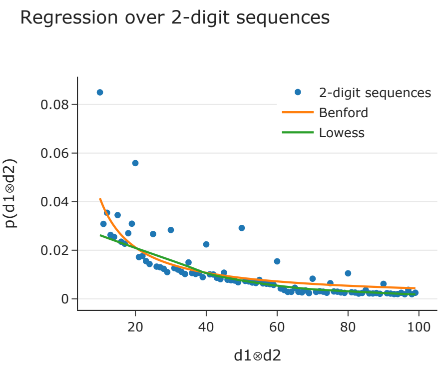
Number Embeddings



t-SNE of the embeddings of the integer model (left) and
the exponent embeddings of the float model (right) in d'Ascoli et al., 2022.

Total Order Through Benford's Law





Outline

Neural ML and Mathematics

Epistemology of Natural Language Processing (NLP)

Distributional Arithmetic

Takeaways

Takeways

- Epistemology of neural ML applications in science through case studies rather than general speculative principles.

- In the case of mathematics: New and original role of natural language.

- Remove the magic showing how what was assumed to be impossible is possible in principle and to some degree independent of neural techniques.

- The hope is that this perspective at the crossroads of many disciplines (mathematics, linguistics, ML) can provide epistemological insights to all of them:
  - Maths: Distributional account of mathematical objects.
  - NLP: Reorientation of philosophical debate, away from 📉🦜 vs 🤖🧠 debate.
  - ML: Interpretability tools and new links between statistical and symbolic features.

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

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

Bradley, T.-D., Gastaldi, J. L., & Terilla, J. (Forthcoming 2024a). The structure of meaning in language: parallel narratives in linear algebra and category theory. *Notices of the AMS*.

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

Gastaldi, J. L. (Forthcoming 2024c). Content from Expressions. The Place of Textuality in Deep Learning Approaches to Mathematics. *Synthese (under review)*.

Gastaldi, J. L. (Forthcoming 2004b). How to Do Maths with Words. Neural Machine Learning Applications to Mathematics and Their Philosophical Significance. In B. Sriraman (Ed.), *Handbook of the History and Philosophy of Mathematical Practice*. Springer.

Gastaldi, J. L., & Pellissier, L. (2021). The calculus of language: Explicit representation of emergent linguistic structure through type-theoretical paradigms. *Interdisciplinary Science Reviews*. https://doi.org/10.1080/03080188.2021.1890484

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

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

Lewkowycz, A., Andreassen, A., Dohan, D., Dyer, E., Michalewski, H., Ramasesh, V., Slone, A., Anil, C., Schlag, I., Gutman-Solo, T., Wu, Y., Neyshabur, B., Gur-Ari, G., & Misra, V. (2022). Solving quantitative reasoning problems with language models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in neural information processing systems* (pp. 3843–3857, Vol. 35). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2022/file/18abbeef8cfe9203fdf9053c9c4fe191-Paper-Conference.pdf

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

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

Polu, S., & Sutskever, I. (2020). Generative language modeling for automated theorem proving. *CoRR*, *abs/2009.03393*. https://arxiv.org/abs/2009.03393

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*.

Wagner, A. Z. (2021). Constructions in combinatorics via neural networks.

Wu, Y., Jiang, A. Q., Li, W., Rabe, M. N., Staats, C. E., Jamnik, M., & Szegedy, C. (2022). Autoformalization with large language models. In A. H. Oh, A. Agarwal, D. Belgrave, & K. Cho (Eds.), *Advances in neural information processing systems*. https://openreview.net/forum?id=IUikebJ1Bf0

PhilML’23

Eberhard Karls University of Tübingen

Tübingen, Germany.

September 12-14, 2023

The Language of Mathematics

Epistemological Consequences of the Application of Neural Models to Mathematics

Juan Luis Gastaldi

ETH zürich

September 13, 2023

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