CIS 522: Lecture 12

Automated Decision Making

Lyle Ungar

Penn Engineering

Today

- **The 3rd type of machine learning**
  - Supervised
    - Classification/Prediction
  - Unsupervised
    - Auto-encoders
    - Generative models
  - Reinforcement Learning
    - Automated decision making
- **Optimally training big models**
- **Projects: how to write a paper**

Automated Decision Making

Learn policy $a = f(X)$ to maximize reward

	Offline	**Online**
Immediate	Supervised learning	Contextual bandit
Sequential	RL	classic RL

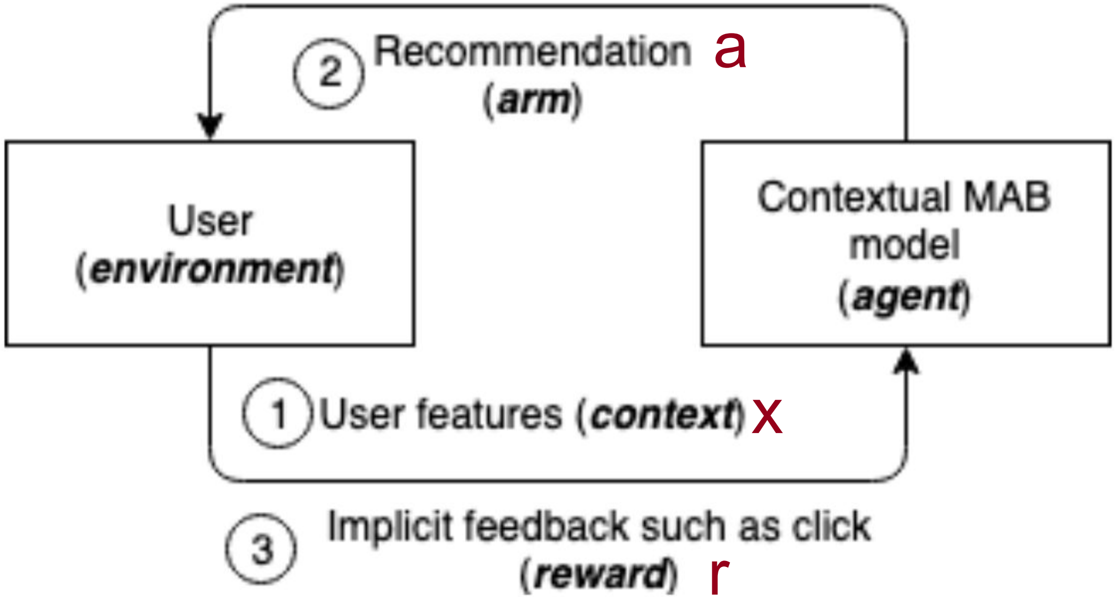
Automated Decision Making

Learn policy $a = f(X)$ to maximize reward

- **Offline: Imitation learning**
  - One step: behavior cloning
  - Sequential: Inverse-RL, soft Q-Imitation learning…
- **Online**
  - One step: Contextual Bandits
  - Sequential: Reinforcement learning

Contextual Bandit (amazon)

recommendation system



$a = f(x)$

Image: amazon

AI & MACHINE LEARNING

Better bandit building: Advanced personalization the easy way with AutoML Tables

Zi Yang
Cloud Technical Resident; Google Cloud Bandits Solutions team

Contextual Bandit (google)

- **Learn Q-function**
  - $Q(s,a) = E[r|s,a]$
  - But there is no history here!!!
- Uses linearly annealed epsilon greedy exploration

Deep Q-Learning (DQN)

Inspired by

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left(R + \gamma \max_{a'} Q(s',a') - Q(s,a)\right)$$

Represent $Q(s,a)$ *by a neural net*

Estimate using gradient descent with loss function:

$$\left(R + \gamma \max_{a'} Q(s',a') - Q(s,a)\right)^2$$

Contextual bandit: γ = 0 gives loss function

$$\left(R - Q(s,a)\right)^2$$

The policy, π*(a),* is then given by maximizing the predicted Q-value

How big a model do we want?

[Submitted on 29 Mar 2022]

Training Compute-Optimal Large Language Models

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W. Rae, Oriol Vinyals, Laurent Sifre

https://www.lesswrong.com/posts/midXmMb2Xg37F2Kgn/new-scaling-laws-for-large-language-models

Model size vs. Training size

- **Old style (2020-2022):**

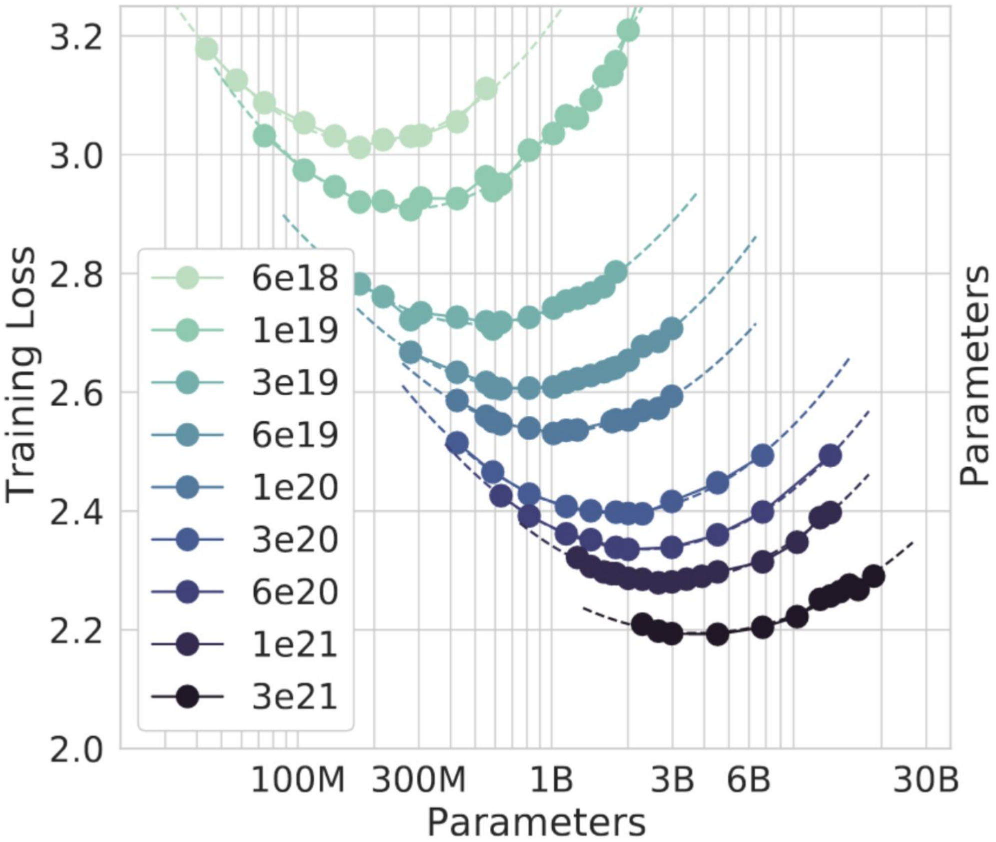
Model	Size (# Parameters)	Training Tokens
LaMDA (Thoppilan et al., 2022)	137 Billion	168 Billion
GPT-3 (Brown et al., 2020)	175 Billion	300 Billion
Jurassic (Lieber et al., 2021)	178 Billion	300 Billion
Gopher (Rae et al., 2021)	280 Billion	300 Billion
MT-NLG 530B (Smith et al., 2022)	530 Billion	270 Billion

We are compute limited

isoflop lines show:

For each 100x in compute, use 10x parameters and 10x data

And it shows the right ratio of parameters/data



Need to get the right ratio

Deepmind: Same compute used in

- **Gopher:** 280-billion parameters, 0.3 trillion tokens
- **Chinchilla:** 70-billion parameters, 1.4 trillion tokens

Chinchilla is much better!

How did OpenAI get it wrong?

- Need different annealing rates for different sizes

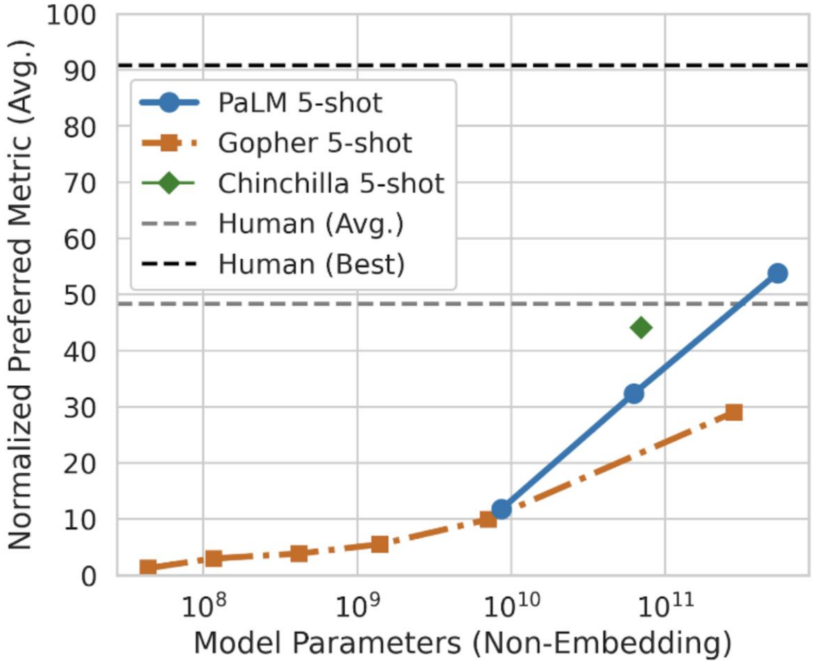
Implications

Better performance without massively larger language models

The latest: Palm

Trained on 6,144 chips

English and multilingual datasets include high-quality web documents, books, Wikipedia, conversations, and GitHub code.

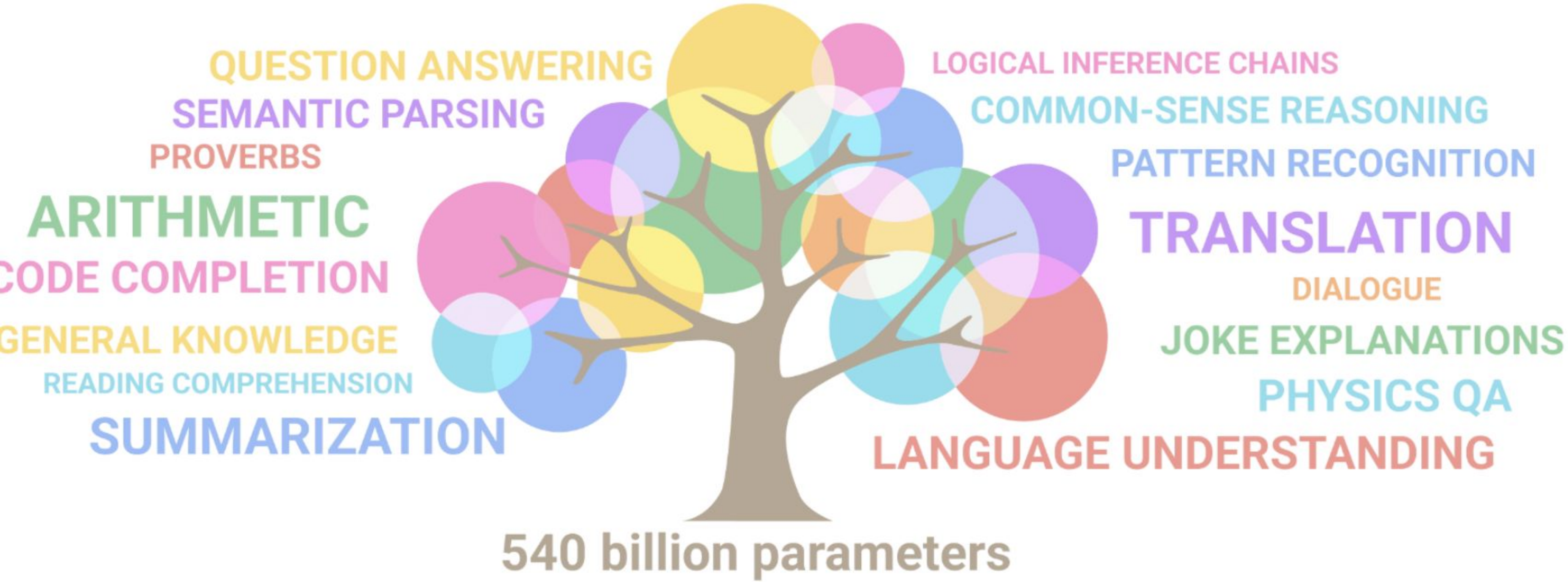


Pathways Language Model (PaLM): Scaling to 540 Billion Parameters for Breakthrough Performance

Monday, April 4, 2022

Posted by Sharan Narang and Aakanksha Chowdhery, Software Engineers, Google Research

The latest: Palm

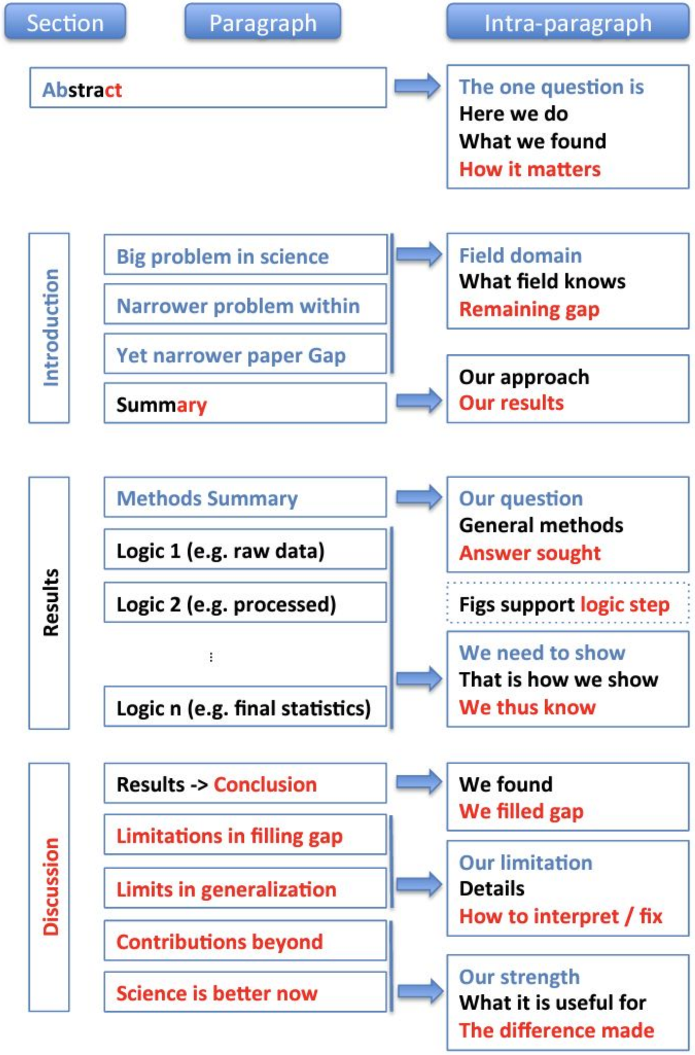


https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html

Projects

Ten simple rules for structuring papers

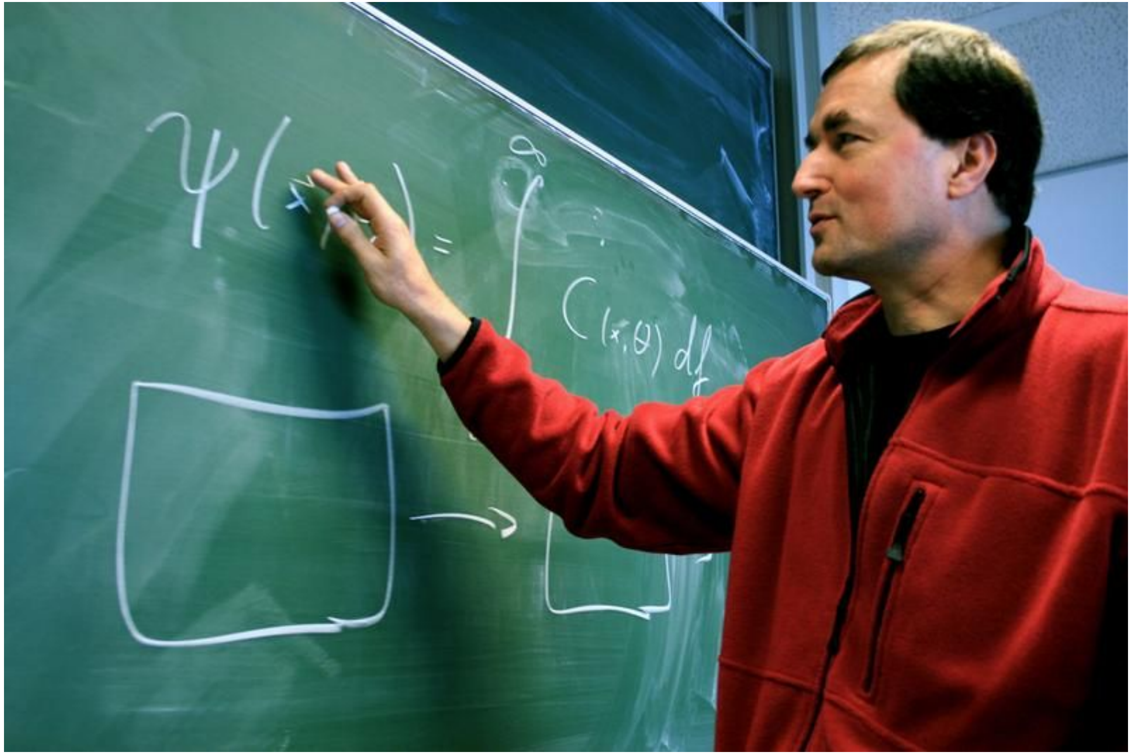
Brett Mensh, Konrad Kording



https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005619

Ten simple rules for structuring papers

@spiralmensh, @kordinglab



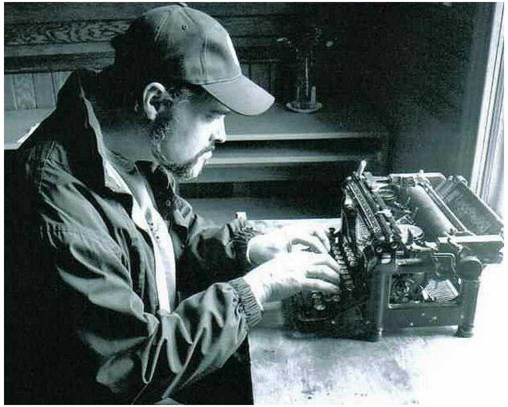
Rule 1: One central idea

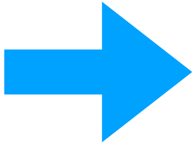
- Attention is all you need
- Resnet
- Xavier/ He initialization

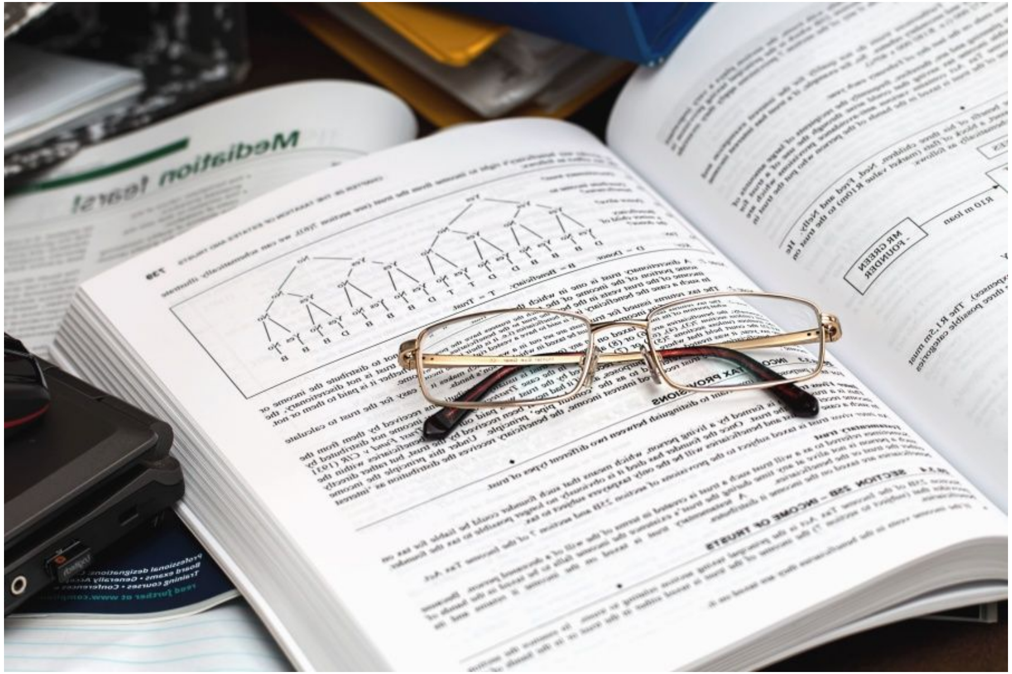
Problems of >1 idea

- 1 title
- 1 abstract
- 1 intro/results/discussion
- brains hate zig-zag
- Modern search patterns

Rule 2: Write for naive humans







Rule 3: Context/ Contents/ Conclusion

To rule out the possibility of central or peripheral fatigue, we compared the maximum voluntary force before, in the middle of, and after data collection in these dual-task experiments. In Experiment 3, MVCs of all effectors do not change significantly (one-way ANOVA on timing, $p = 0.588, 0.919, 0.723$, and 0.868 for the right finger, palm, leg, and left finger, respectively). For the two drug experiments (Experiments 4 and 5), we performed a 2 (drug) $\times$ 3 (timing) repeated-measures ANOVA on MVCs of the finger and the palm. Neither the main effects nor the interaction is significant, indicating that fatigue does not contribute to the observed effects and that MVCs are not affected by cyproheptadine or paroxetine intake.

From: Wei et al 2014 - Serotonin affects movement gain control in the spinal cord

Rule 4: Optimize logical flow

Rule 5: Everything in Abstract

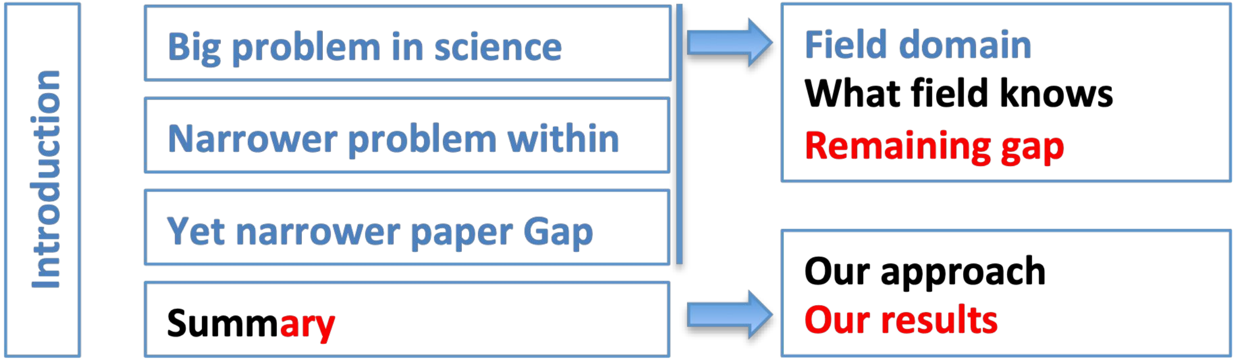
Abstract

The one question is
Here we do
What we found
How it matters

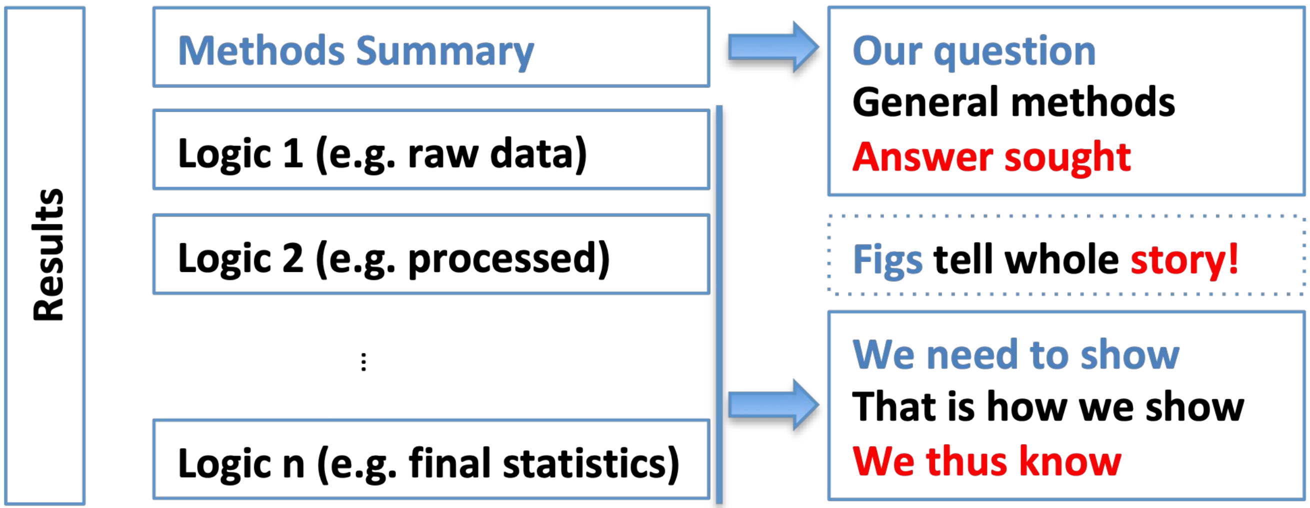
Example

Whereas before 2006 it appears that deep multi-layer neural networks were not successfully trained, since then several algorithms have been shown to successfully train them, with experimental results showing the superiority of deeper vs less deep architectures. All these experimental results were obtained with new initialization or training mechanisms. Our objective here is to understand better why standard gradient descent from random initialization is doing so poorly with deep neural networks, to better understand these recent relative successes and help design better algorithms in the future. We first observe the influence of the non-linear activations functions. We find that the logistic sigmoid activation is unsuited for deep networks with random initialization because of its mean value, which can drive especially the top hidden layer into saturation. Surprisingly, we find that saturated units can move out of saturation by themselves, albeit slowly, and explaining the plateaus sometimes seen when training neural networks. We find that a new non-linearity that saturates less can often be beneficial. Finally, we study how activations and gradients vary across layers and during training, with the idea that training may be more difficult when the singular values of the Jacobian associated with each layer are far from 1. Based on these considerations, we propose a new initialization scheme that brings substantially faster convergence.

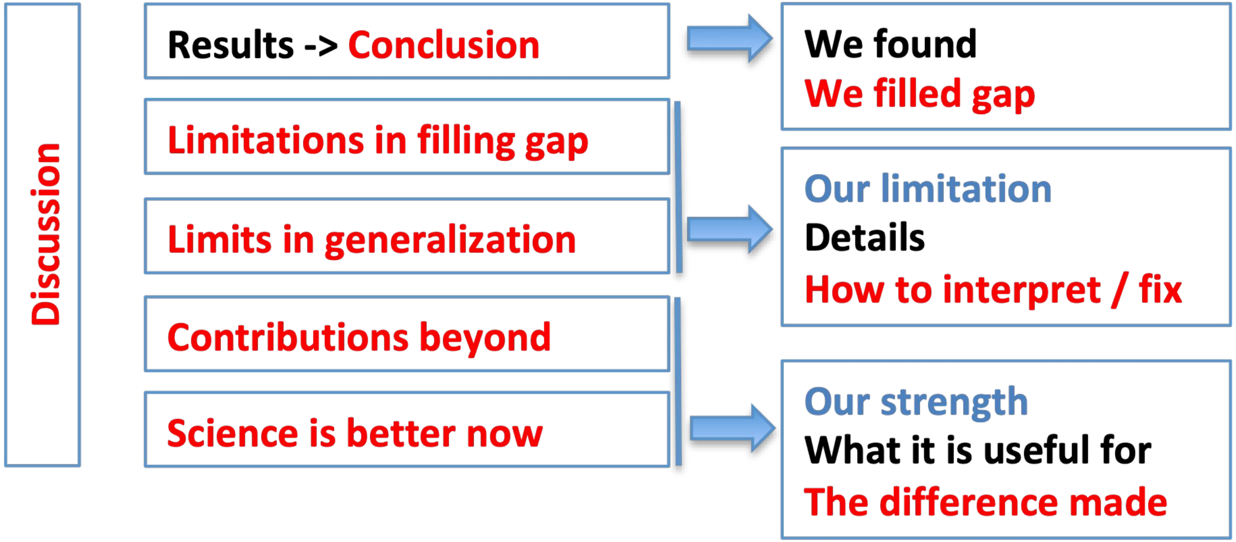
Rule 6: Why it matters in the Introduction


Introduction
Big problem in science
Narrower problem within
Yet narrower paper Gap
Summary
Field domain
What field knows
Remaining gap
Our approach
Our results

Rule 7: A sequence of logically dependent points


Results
Methods Summary
Logic 1 (e.g. raw data)
Logic 2 (e.g. processed)
⋮
Logic n (e.g. final statistics)
Our question
General methods
Answer sought
Figs tell whole story!
We need to show
That is how we show
We thus know

Rule 8: Preempt criticism & future impact



Discussion
Results -> Conclusion
Limitations in filling gap
Limits in generalization
Contributions beyond
Science is better now
We found
We filled gap
Our limitation
Details
How to interpret / fix
Our strength
What it is useful for
The difference made

Rule 9: Allocate writing time wisely


TIMEX
QUARTZ

Rule 10: Reuse, Reduce, Recycle

Signs of violation

Rule	Sign it is violated
1: Focus on one big idea	Readers cannot give 1-sentence summary.
2: Write for naive humans	Readers do not “get” the paper.
3: Use context, content, conclusion structure	Readers ask why something matters or what it means.
4: Optimize logical flow	Readers stumble on a small section of the text.
5: Abstract: Compact summary of paper	Readers cannot give the “elevator pitch” of your work after reading it.
6: Introduction: Why the paper matters	Readers show little interest in the paper.
7: Results: Why the conclusion is justified	Readers do not agree with your conclusion.
8: Discussion: Preempt criticism, give future impact	Readers are left with unanswered criticisms and/or questions on their mind.
9: Allocate time wisely	Readers struggle to understand your central contribution despite your having worked hard.
10: Iterate the story	The paper’s contribution is rejected by test readers, editors, or reviewers.

https://doi.org/10.1371/journal.pcbi.1005619.t001