Advancing mathematics by guiding human intuition with AI

Presented by Ye Yuan

Catalog

- Paper background
- Motivation
- The framework proposed in this paper
- Example of successful usage of the framework
- Conclusion and Contribution

Paper Background

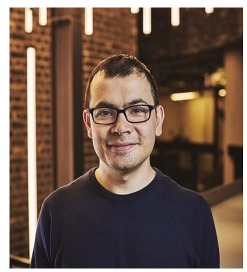
- Received: 10 July 2021
- Accepted: 30 September 2021
- Published online: 1 December 2021
- Featured front cover of Nature on Volume 600 Issue 7887, 2 December 2021

- This article is open accessible, and you don’t have to subscribe Nature journal.
- https://www.nature.com/articles/s41586-021-04086-x

Authors of Paper

- Alex Davies[1], Petar Veličković[1] , Lars Buesing[1] , Sam Blackwell[1] , Daniel Zheng[1] , Nenad Tomašev[1] , Richard Tanburn[1] , Peter Battaglia[1] , Charles Blundell[1] , András Juhász[2] , Marc Lackenby[2] , Geordie Williamson[3] , Demis Hassabis[1] & Pushmeet Kohli[1]

Alex Davies
Founding lead of AI for Maths of DeepMind



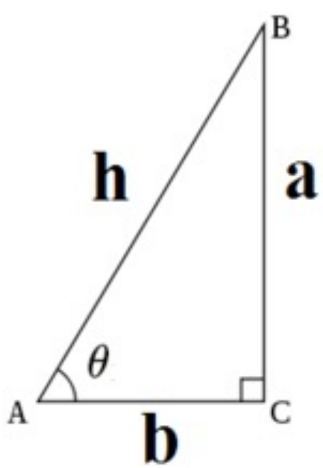
Demis Hassabis
Founder and CEO of DeepMind

Pushmeet Kohli
Head of Research of DeepMind

- [1]DeepMind, [2]University of Oxford, [3]University of Sydney

Motivation

- Conjectures: statements that are suspected to be true but have not been proven to hold in all cases
- Theorem: a theorem is a statement that has been proved or can be proved.
- For a right triangle, we easily find the relation between hypotenuse and right-angle sides.
- Pythagorean theorem



Given		Found
a = 3	b = 4	h = 5
a = 5	b = 12	h = 13
a = 6	b = 8	h = 10
a = 1	b = 1	h = $\sqrt{2}$
a = 2	b = 3	h = $\sqrt{13}$
a = 12709	b = 13500	h = 18541
a = …	b = …	$h^2 = a^2 + b^2$

Motivation

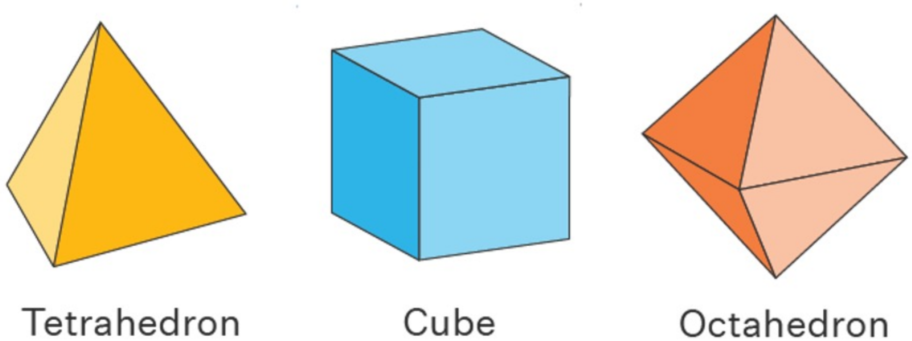
- Sometimes we may think there's no relation between two objects, but there is.

x	$\pi(x)$
10	4
10^2	25
…	…
10^{25}	176,846,309,399,143,769,411,680

- x is a positive integer, π(x) is the number of prime numbers smaller than or equal to x.
- $\pi(x) \sim \frac{x}{\log x}$. This is the famous prime number theorem.
- Sometimes there's no relations between two objects at all. If there's no relation, we should not waste time on trying to find it.
- Machine learning is good at finding relations.
- Contribution: help mathematicians to confirm the existence of the relation and determine which quantities are useful for formulating conjectures.

Formulation

- We have a mathematical object z, X(z) is some mathematical features of z and Y(z) is some other mathematical features of z
- A candidate conjecture can be a relation between X(z) and Y(z).
- In other word, we may hypothesize that there exists a function f such that $f(X(z)) \approx Y(z)$.

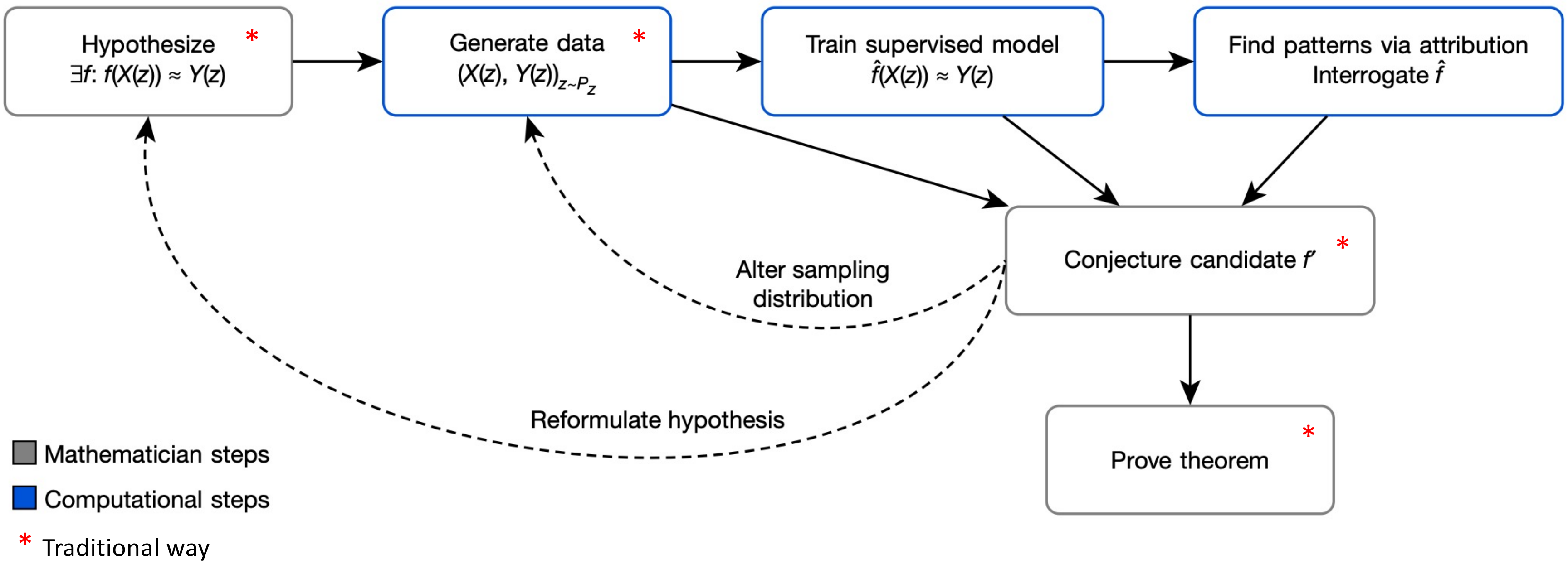


z: convex polyhedrons

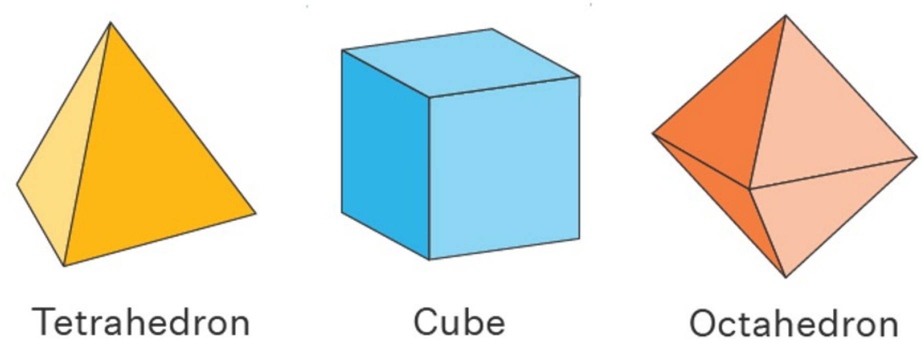
V: number of vertices
E: number of edges
Vol: volume
Sur: surface area
F: number of faces
X(z): $\langle V, E, Vol, Sur \rangle \in Z^2 \times R^2$
Y(z): $\langle F \rangle \in Z$

Flowchart of the proposed framework

- Sometimes there is no relation between X(z) and Y(z). Then you don't have to insist on it. Don't waste your time.
- If there's some relations indeed, the result of machine learning model can give you confidence that you may find and prove some theorem in this direction. That is the intuition.



An example



z: convex polyhedrons

- V: number of vertices, E: number of edges
- Vol: volume, Sur: surface area, F: number of faces
- X(z): <V, E, Vol, Sur> $\in Z^2 \times R^2$ and Y(z): <F> $\in Z$
- We hypothesized that there is a relation between X(z) and Y(z)
- i.e., find a f such that $f(X(z)) \approx Y(z)$
- X(z) are the features, and Y(z) are the labels. Use a supervised learning.
- Now, we need to collect data. That is, we collect many convex polyhedrons and record their number of vertices, edges, and faces, as well as volume and surface area.
- Fit the model. We can simply use a linear regression model.
- We found that $X(z) \cdot (-1, 1, 0, 0) + 2 = Y(z)$, that is $-V + E + 2 = F$.
- If we rewrite the formula as $2 = F + V - E$, you got the Euler's polyhedron formula.

Problems

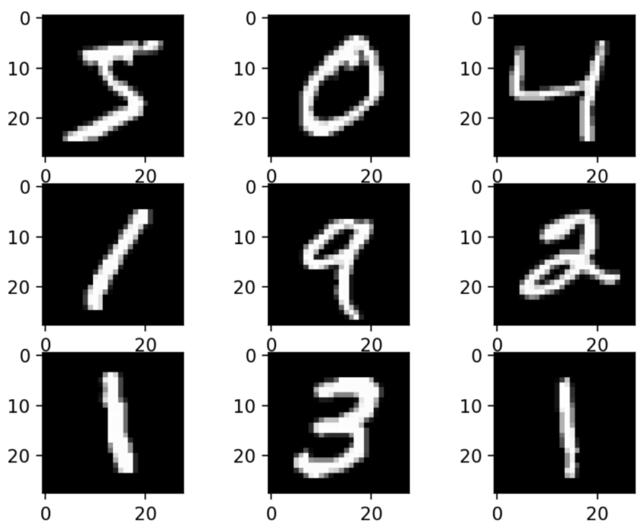
- $X(z) \cdot (-1, 1, 0, 0) + 2 = Y(z)$. The volume and the surface area are not predictors of the number of faces.

- This statement is obvious when we use a linear model. However, when $X(z)$ and $Y(z)$ have high dimensions, we may use a deep model, e.g., MLP. We don't want lose any interpretability.

- We still want to know which features are useful for predicting the results, because we need them to formulate a conjecture. (Attribution Interrogate in the diagram on slide 8.)

Attribution Technique

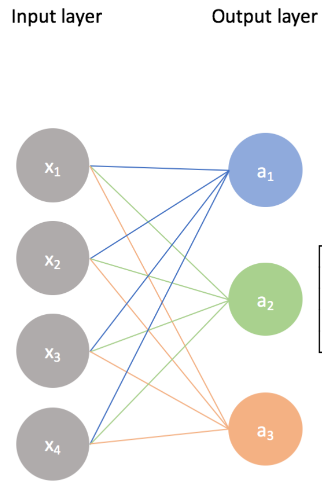
- Problem: which features have more influence on the result?
- Which pixels can tell you that this is number one?
- Why this is a picture of cat?



Gradient Saliency

- X = input, Y = output
- σ is the activation function
- $Y = \sigma(w^T \cdot X + b)$
- $X = (x_1, x_2, x_3, x_4)$ and $Y = (y_1, y_2, y_3)$

- $y_1 = \sigma\,(w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b_1)$
- calculate: $\frac{\partial y_1}{\partial x_1}, \frac{\partial y_1}{\partial x_2}, \frac{\partial y_1}{\partial x_3}, \frac{\partial y_1}{\partial x_4}$
- y_1 is most sensitive to the change of x_i if $\frac{\partial y_1}{\partial x_i}$ has the largest value



A simple neural network

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

$$w_1 = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}, w_2 = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}, w_3 = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}$$

$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \end{bmatrix} \text{becomes} \begin{bmatrix} \leftarrow w_1^T \rightarrow \\ \leftarrow w_2^T \rightarrow \\ \leftarrow w_3^T \rightarrow \end{bmatrix}$$

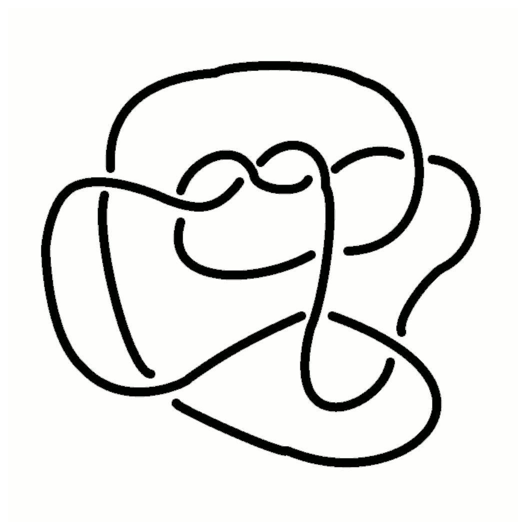
$$\begin{bmatrix} b \\ b \\ b \end{bmatrix} \text{becomes} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

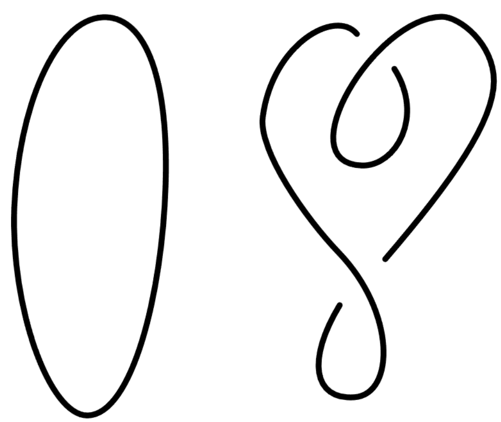
Successful Example

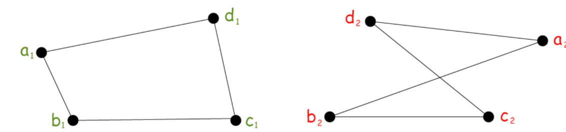
Knots Theory in Topology

Knots Theory

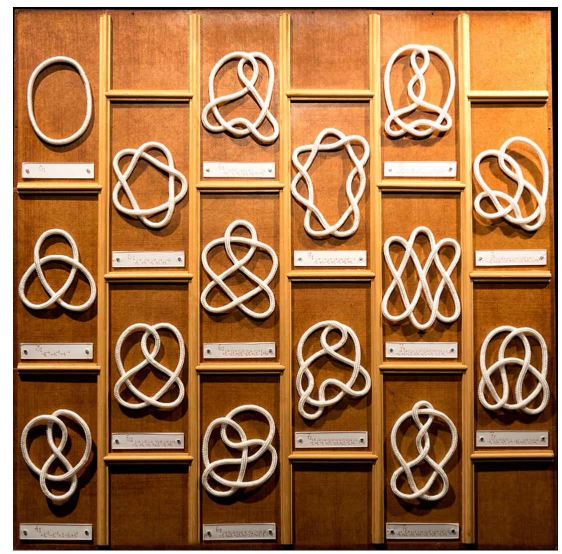
- Knots: ropes in 3D world where two ends are joined
- Knots equivalency: if you can change a knot to other knot without using scissors to cut it off, then they are considered as equivalent.
- Knot Invariants: a "quantity" that is the same for equivalent knots. There are two types of Knot Invariants, Geometric and Algebraic.







example of isomorphism
If two graph are isomorphic,
some quantity must be same.



Topology-Knots Theory

- We want to find relations between the geometric and algebraic invariants
- Prior Conjecture (Volume Conjecture): the hyperbolic volume of a knot (a geometric invariant) should be encoded within the asymptotic behavior of its colored Jones polynomials (which are algebraic invariants)
- But the mathematicians think there might be some undiscovered relations between the geometric and algebraic invariants

z: **Knot**	$X(z)$: **Geometric invariants**				$Y(z)$: **Algebraic invariants**		
	Volume	Chern–Simons	Meridional translation	...	Signature	Jones polynomial	...
	2.0299	0	i	...	0	$t^{-2} - t^{-1} + 1 - t + t^2$	...
	2.8281	–0.1532	$0.7381 + 0.8831i$	...	–2	$t - t^2 + 2t^3 - t^4 + t^5 - t^6$	...
	3.1640	0.1560	$-0.7237 + 1.0160i$	...	0	$t^{-2} - t^{-1} + 2 - 2t + t^2 - t^3 + t^4$	...

Topology-Knots Theory

- z: Knots
- X(z): <Im(Meridional translation), Longitudinal translation, Re(Meridional translation), Im(Short geodesic), Injectivity radius, Cusp volume, Symmetry group, Torsion degree, Re(Short geodesic), Volume, Chern–Simons, Adjoint torsion degree>
- Y(z): <Signature>

- Hypothesis: $\exists f$: $f(X(z)) \approx Y(z)$

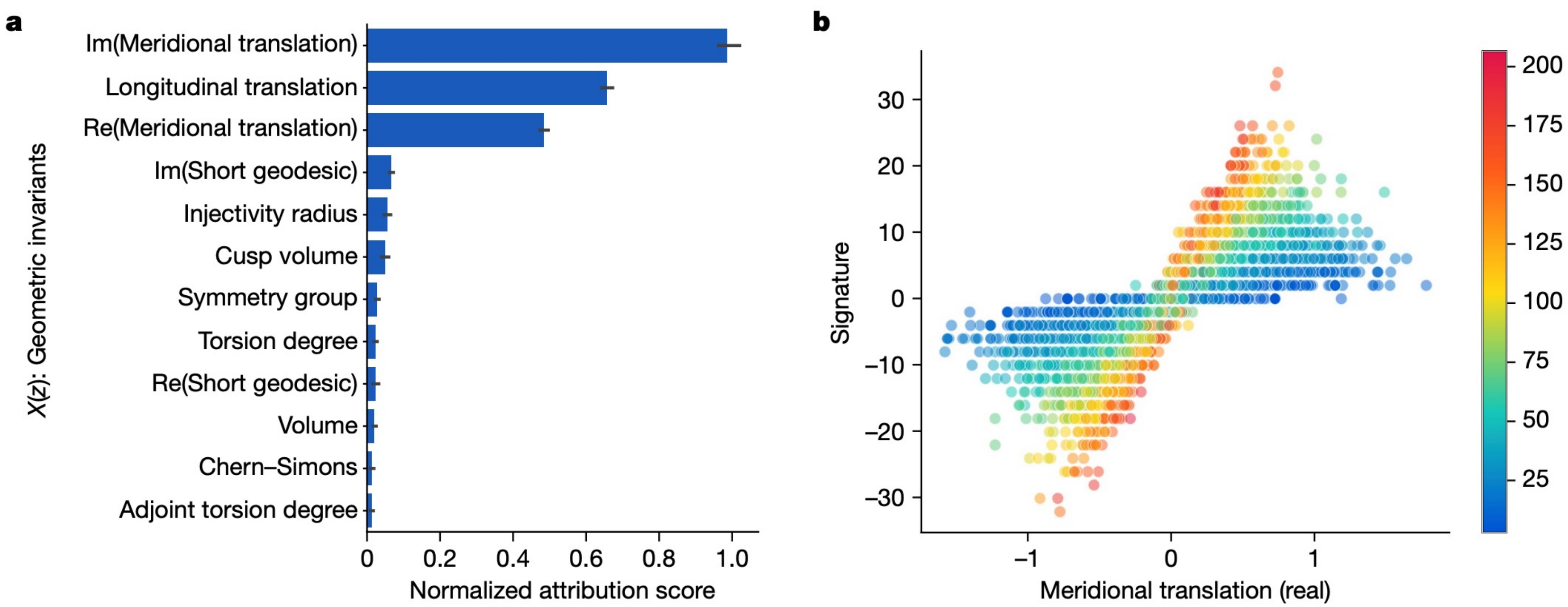
- Collect or create lots of knots to make the dataset.

- Model: fit an MLP -> accuracy: 83% (somebody achieved 97%) -> confirm f exist

- There must be some undiscovered relations between X(z) and Y(z)

Topology- Knots Theory

- Use attribution technique



| **Knot theory attribution. a**, Attribution values for each of the input $X(z)$. The features with high values are those that the learned function is most sensitive to and are probably relevant for further exploration. The 95% confidence interval error bars are across 10 retrainings of the model. **b**, Example visualization of relevant features—the real part of the meridional translation against signature, coloured by the longitudinal translation.

Topology-Knots Theory

- Im(Meridional translation), Longitudinal translation, Re(Meridional translation) are the influential predictors
- We define μ = Meridional translation and λ = Longitudinal translation
- We define slope(K) = Re(λ/μ), where K is a knot.
- We define signature as $\sigma(K)$, volume as vol(K) and the injectivity radius inj(K)
- Conjecture candidate: There exist constants c1 and c2 such that, for every hyperbolic knot K,

$$|2\sigma(K) - \mathrm{slope}(K)| < c_1 \mathrm{vol}(K) + c_2$$

- But found a counterexample in the dataset
- Final Conjecture: **There exists a constant c such that, for any hyperbolic knot K,**

$$|2\sigma(K) - \mathrm{slope}(K)| \leq c\,\mathrm{vol}(K)\,\mathrm{inj}(K)^{-3}$$

- Proved and published on the paper "The signature and cusp geometry of hyperbolic knots"

Contribution

Traditional Way to Prove a Theorem

Hypothesize $\exists f$: $f(X(z)) \approx Y(z)$ (Q1)

Generate data $(X(z),Y(z))z$~Pz

Conjecture candidate f' (Q2)

Prove theorem

Problem:

Q1. Cannot confirm there exists f. Waste time if f doesn't exist.

Q2. Even if f exists indeed, we don't know which parts of X(z) should be considered when formulating the conjecture.

Proposed Framework to Prove a Theorem

Hypothesize $\exists f$: $f(X(z)) \approx Y(z)$

Generate data $(X(z),Y(z))z$~Pz

Train supervised model $f^*(X(z)) \approx Y(z)$ (Solve Q1)

Find patterns via attribution Interrogate f^* (Solve Q2)

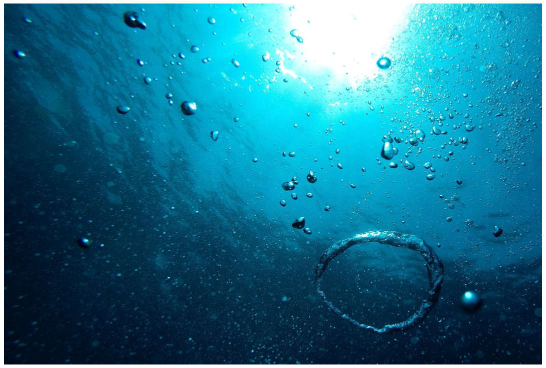
Conjecture candidate f'

Prove theorem

Contribution

Traditional Way to Prove a Theorem

- There might be some treasures under the sea, but we are not sure about this. We don't know how to find it even if there is a treasure.



Proposed Framework to Prove a Theorem

- There must be some treasures in the sea, and you can find it if you search along this direction. (There is no treasure under the sea, and don't waste your time.)

Conclusion of the Paper

- The example of knots theory demonstrate how a foundational connection in a well-studied and mathematically interesting area can go unnoticed, and how the framework allows mathematicians to better understand the behaviour of objects that are too large for them to otherwise observe patterns in.

- Advantages:
  - help mathematicians confirm the hypothesized relation exists
  - give an intuition about formulating a conjecture (determine which features are useful)

- Limitations:
  - must be able to generate large datasets for detectable patterns
  - helpful for proposing a conjecture but not proving a theorem
  - some mathematical domains cannot be benefitted with this framework

My Thoughts

1

Ensure what contribution you want to make.

2

Never be satisfied with current achievements.

3

Even though the machine learning model is simple, it’s still extremely useful in other domains.

References

1. Davies, A., Veličković, P., Buesing, L. *et al.* Advancing mathematics by guiding human intuition with AI. *Nature* **600,** 70–74 (2021).

2. Adams, Colin (2004), The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots, American Mathematical Society, ISBN 978-0-8218-3678-1

Thanks

ANY QUESTIONS ABOUT THE PAPER?

ANY COMMENTS AND FEEDBACK FOR MY PRESENTATION?