Opportunities for infusing physics in AI/ML algorithms

Animashree Anandkumar

Director of ML Research,
NVIDIA
Bren Professor, Caltech

1.

Neural Programming

Combining Symbolic Expressions & Black-box Function Evaluations in Neural Programs

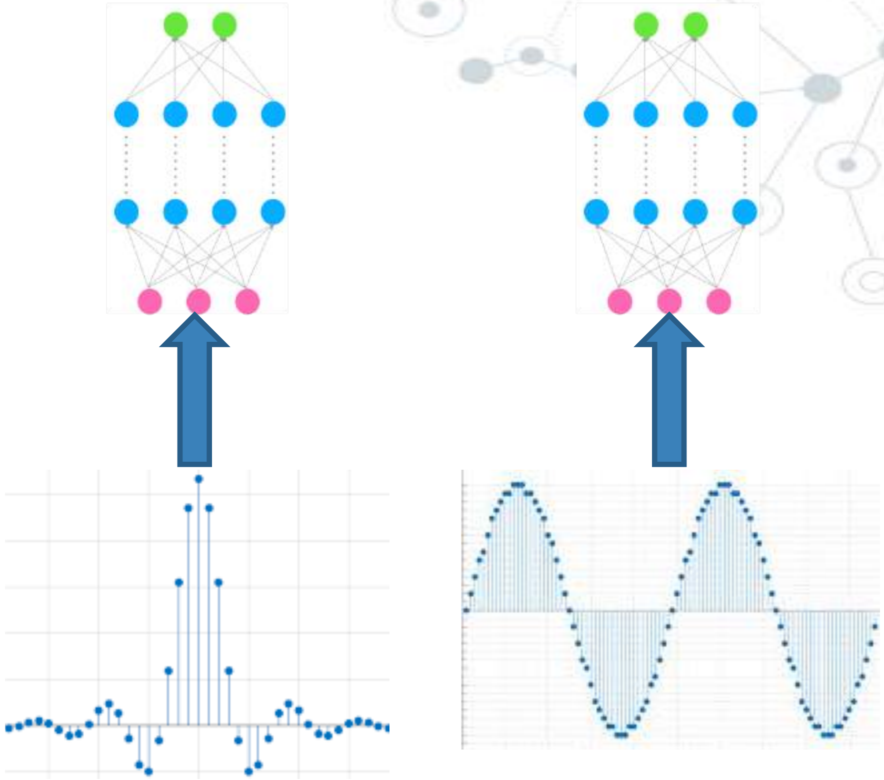
FOROUGH ARABSHAHI, SAMEER SINGH, ANIMA ANANDKUMAR

Symbolic + Numerical Input

Goal: Learn a domain of functions (sin, cos, log…)
- Training on numerical input-output does not generalize.

Data Augmentation with Symbolic Expressions
- Efficiently encode relationships between functions.

Solution:
- Design networks to use both:
symbolic + numeric
- Leverage the observed structure of the data
Hierarchical expressions



Neural Programming

- Data-driven mathematical and symbolic reasoning

- Leverage the observed structure of the data
  - Hierarchical expressions

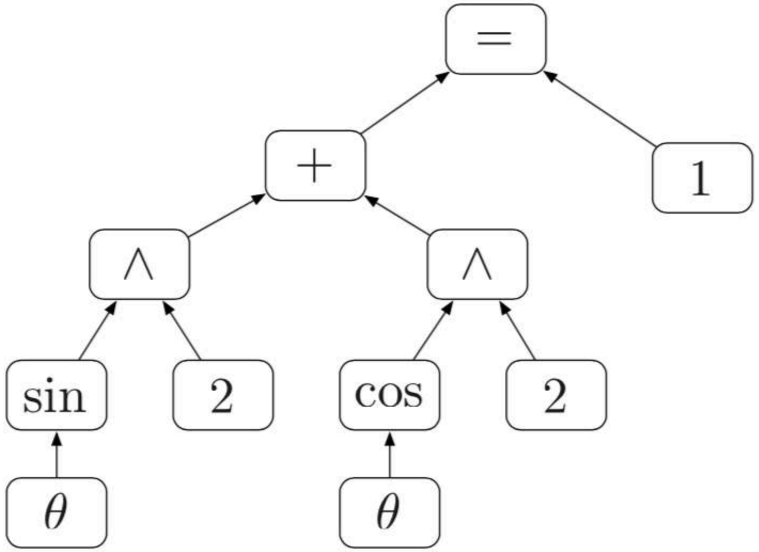
Applications

- Mathematical equation verification
  - $\sin^2\theta + \cos^2\theta = 1$???
- Mathematical question answering
  - $\sin^2\theta + \square^2 = 1$
- Solving differential equations

$$\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} + 4f(x) = \sin(2x)$$

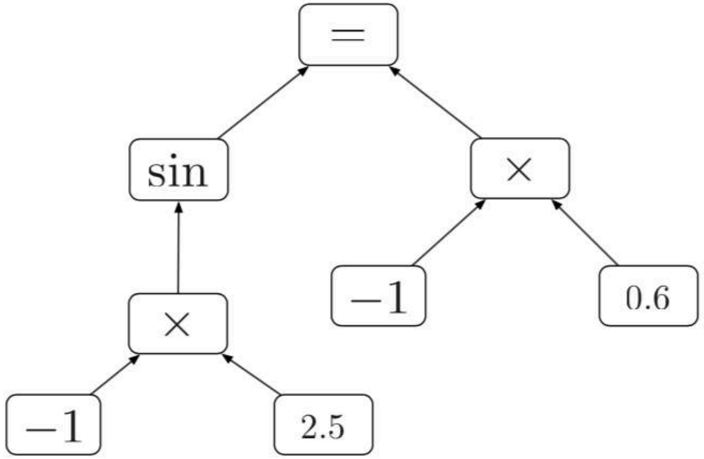
$$f(x) : \frac{1}{8}\sin(2x) - \frac{x}{4}\cos(2x)$$

Examples



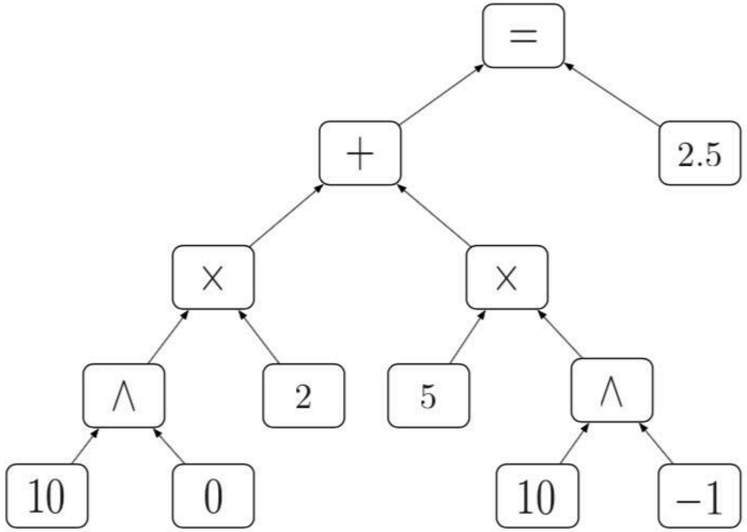
$$\sin^2(\theta) + \cos^2(\theta) = 1$$

Symbolic Data Point



$$\sin(-2.5) = -0.6$$

Function Evaluation Data Point



$$2.5 = 2 \times 10^0 + 5 \times 10^{-1}$$

Number Encoding Data Point

Representing Mathematical Equations

- Grammar rules

$$
\begin{aligned}
I &\rightarrow \; =(E,E),\ \neq(E,E) \\
E &\rightarrow T, F_1(E), F_2(E,E) \\
F_1 &\rightarrow \; \text{sin}, \text{cos}, \text{tan}, \ldots \\
F_2 &\rightarrow +, \wedge, \times, \text{diff}, \ldots \\
T &\rightarrow \; -1, 0, 1, 2, \pi, x, y, \ldots, \\
&\qquad \text{floating point numbers of precision 2}
\end{aligned}
$$

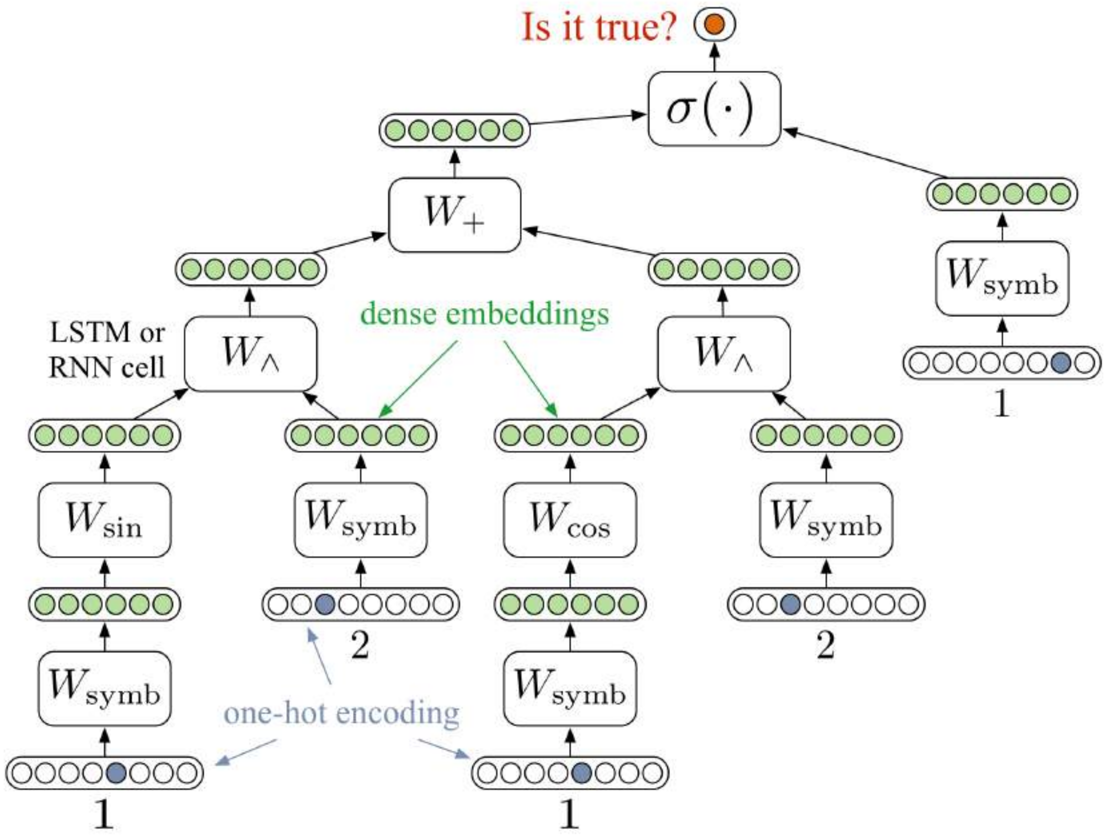
Domain

Unary functions, F_1				
sin	cos	csc	sec	tan
cot	arcsin	arccos	arccsc	arcsec
arctan	arccot	sinh	cosh	csch
sech	tanh	coth	arsinh	arcosh
arcsch	arsech	artanh	arcoth	exp

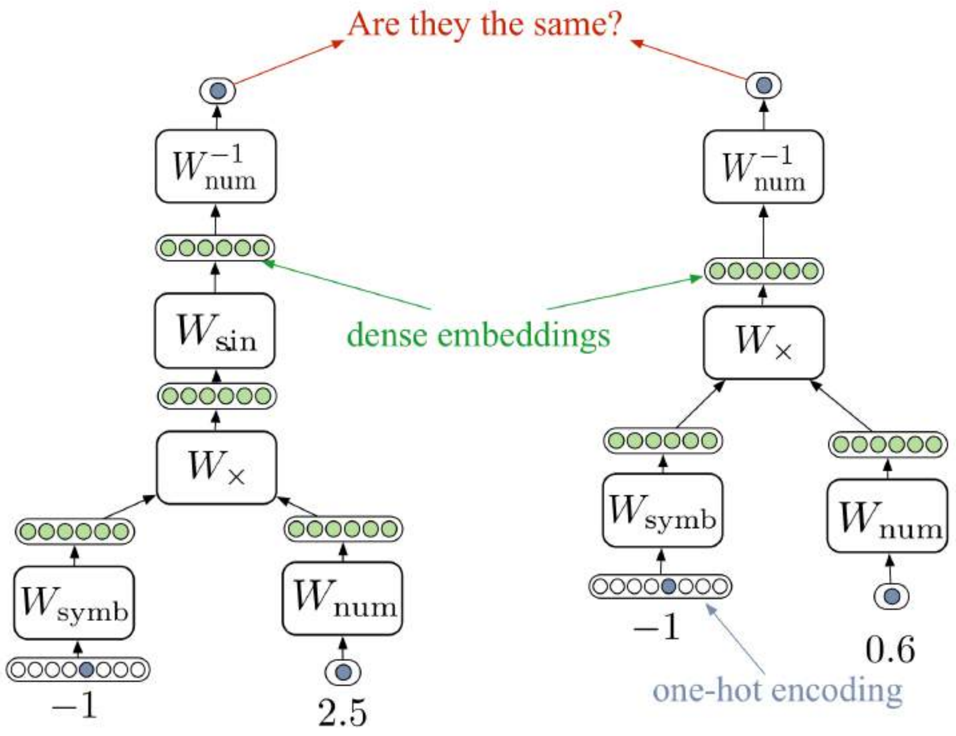
Terminal, T	
0	1
2	3
4	10
0.5	-1
0.4	0.7
π	x

Binary, F_2
$+$
$\times$
$\wedge$
diff

Tree-LSTM for capturing hierarchies



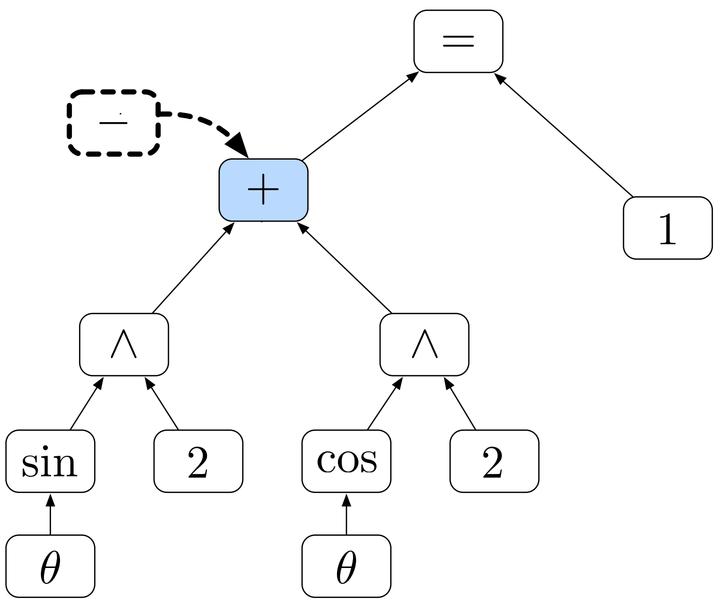
$$\sin^2(\theta) + \cos^2(\theta) = 1$$



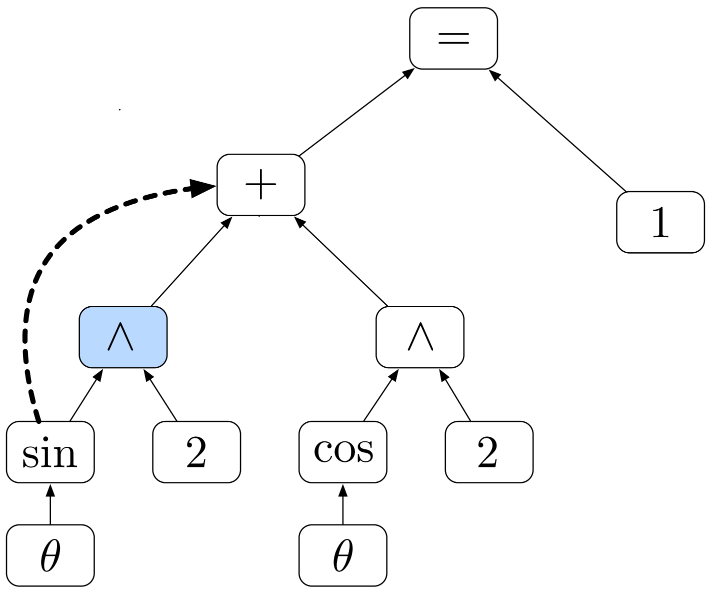
$$\sin(-2.5) = -0.6$$

Dataset Generation

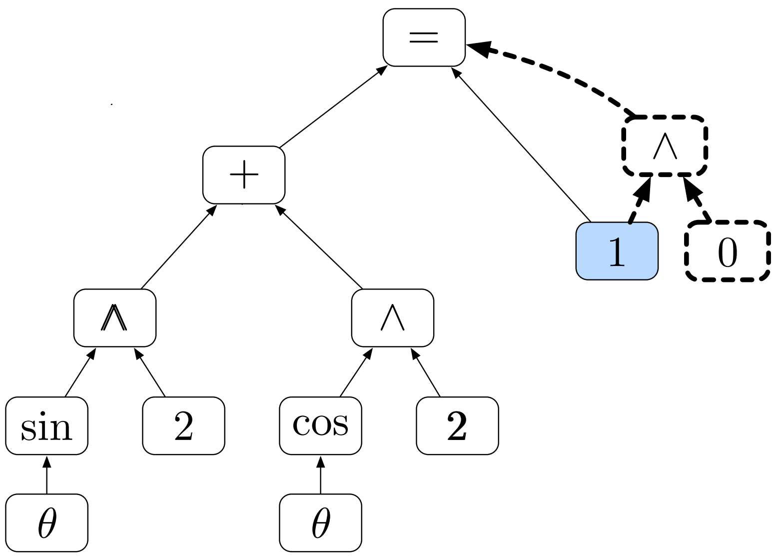
- Random local changes



Replace Node



Shrink Node



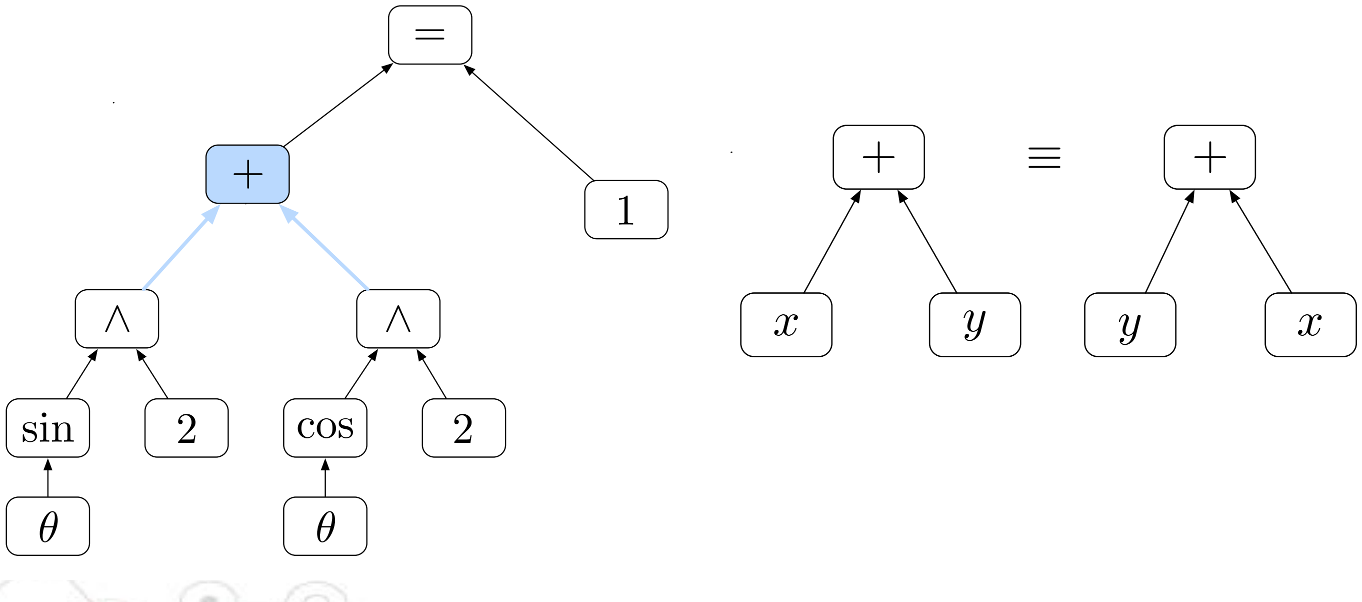
Expand Node

Summary of contributions

- Combine symbolic expressions and function evaluation
- New tasks
  - Equation verification
  - equation completion
  - Solving differential equations
- Balanced dataset generation method
- Generalizable representation of numbers

Dataset Generation

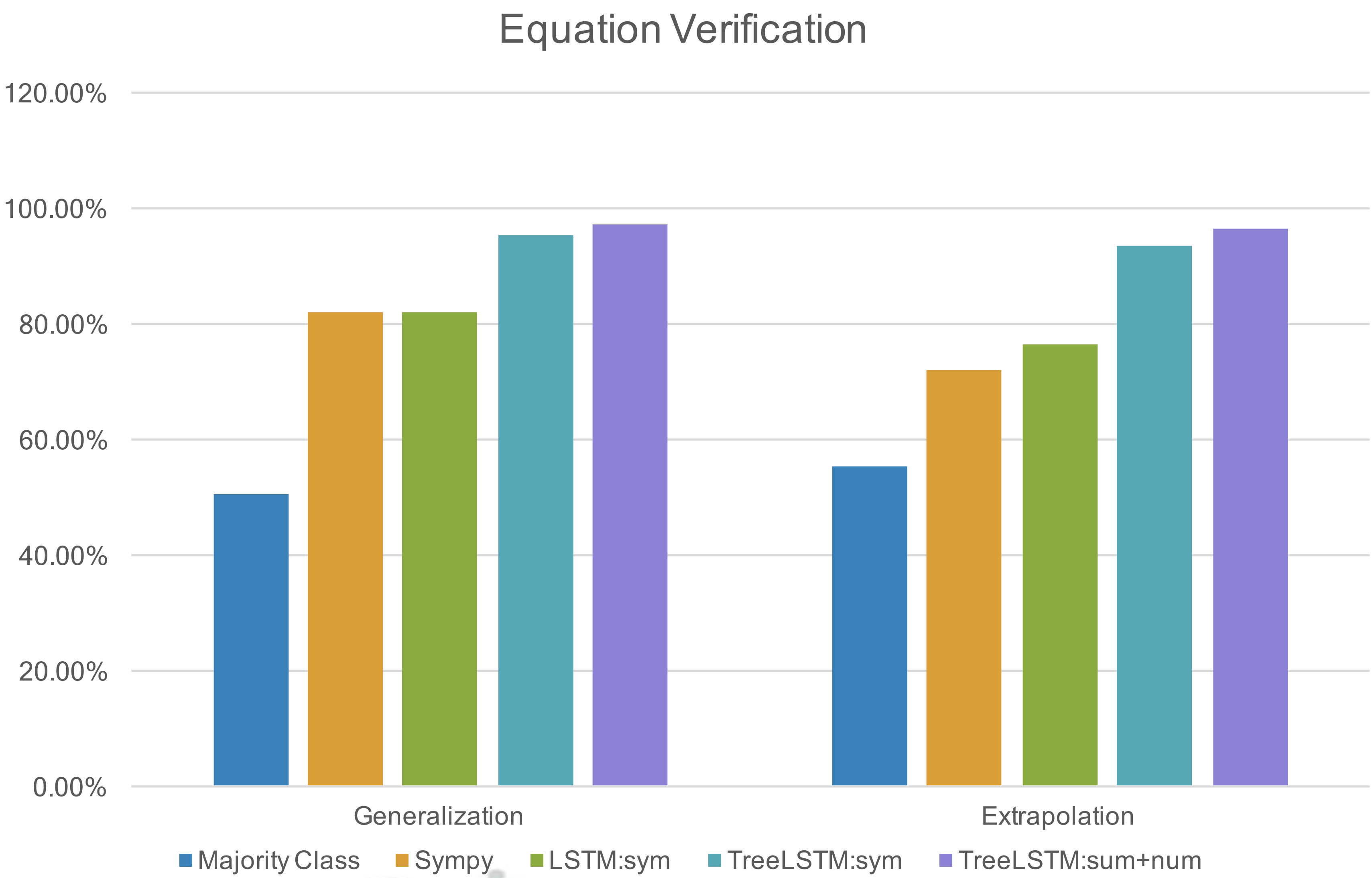
◎ Sub-tree matching



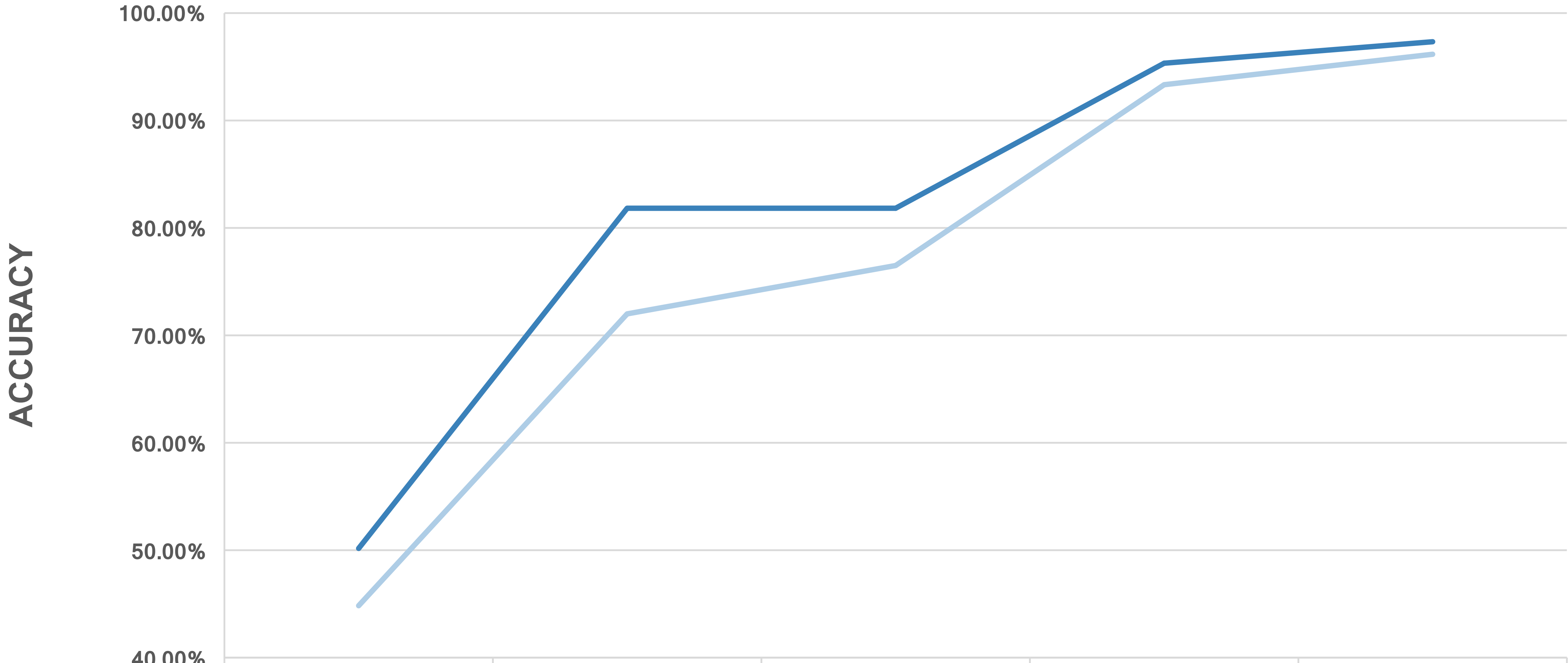
Choose Node　　Dictionary key-value pair　　Replace with value's pattern

Examples of Generated Equations

Examples of correct identities	Examples of incorrect identities
$1^2 = x^{-1\times 0}$	$0.5^{x+2} = sin(0.5)^{x+2}$
$(\arctan 10)^{2^2} = (\arctan 10)^{3+1}$	$\pi \times \csc(x) = -\csc(x)$
$x \times (-1 + x) = x \times (x - 1)$	$-4 = -4^x$
$x^1 = x + 0$	$\frac{\sqrt{2}}{2} \times \sqrt{x} = \sqrt{x}$


Equation Verification
120.00%
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
Generalization
Extrapolation
Majority Class
Sympy
LSTM:sym
TreeLSTM:sym
TreeLSTM:sum+num

EQUATION VERIFICATION

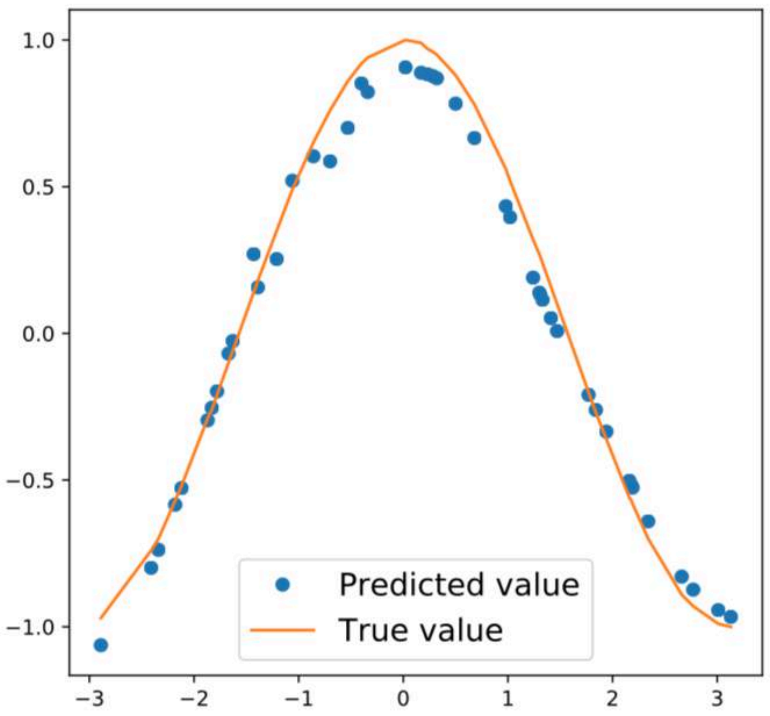


	Majority Class	Sympy	LSTM : sym	TreeLSTM : sym	TreeLSTM:sym+num
Generalization	50.24%	81.74%	81.71%	95.18%	97.20%
Extrapolation	44.78%	71.93%	76.40%	93.27%	96.17%

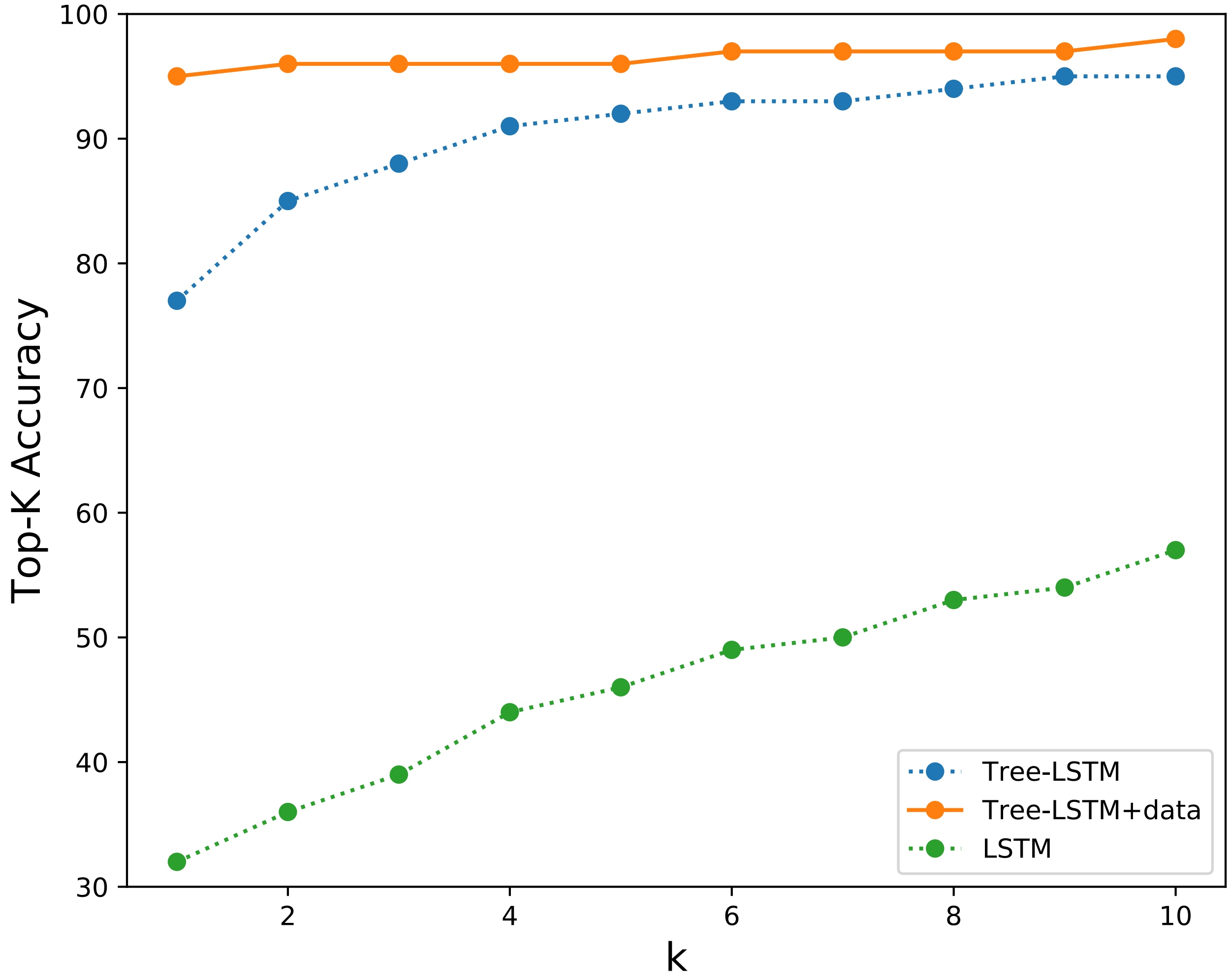
Equation Completion

$4^{\tanh(0)} = \blacksquare^x$	
pred	prob
-2^0	0.9999
1^0	0.9999
7^0	0.9999
-3^0	0.9999

$\cos(-\blacksquare) = -0.57$		
pred	modelErr	trueErr
3	1.8e−5	1.7e−1
2.17	1.9e−5	9.9e−5
2.16	2.6e−5	3.9e−4
2.18	**1.9e−4**	**0**



Equation Completion



Take-aways

- Vastly Improved numerical evaluation: 90% over function-fitting baseline.
- Generalization to verifying symbolic equations of higher depth

LSTM: Symbolic	TreeLSTM: Symbolic	TreeLSTM: symbolic + numeric
76.40 %	93.27 %	**96.17 %**

- **Combining symbolic + numerical data helps in better generalization for both tasks: symbolic and numerical evaluation.**

Solving Differential Equations

- Traditional methods:
  - Gather numerical data from a differential equation
  - Design a neural network for training
- Drawback:
  - Trained model can be used only for that differential equation
  - Train a new model for differential equation
  - Not generalizable

Solving Differential Equations

- Steps:
  - Find a set of candidate solutions
  - Accept the correct candidate using the neural programmer
- Advantage:
  - Jointly train for many functions
  - Generalizable
  - Can be used for solving any differential equation

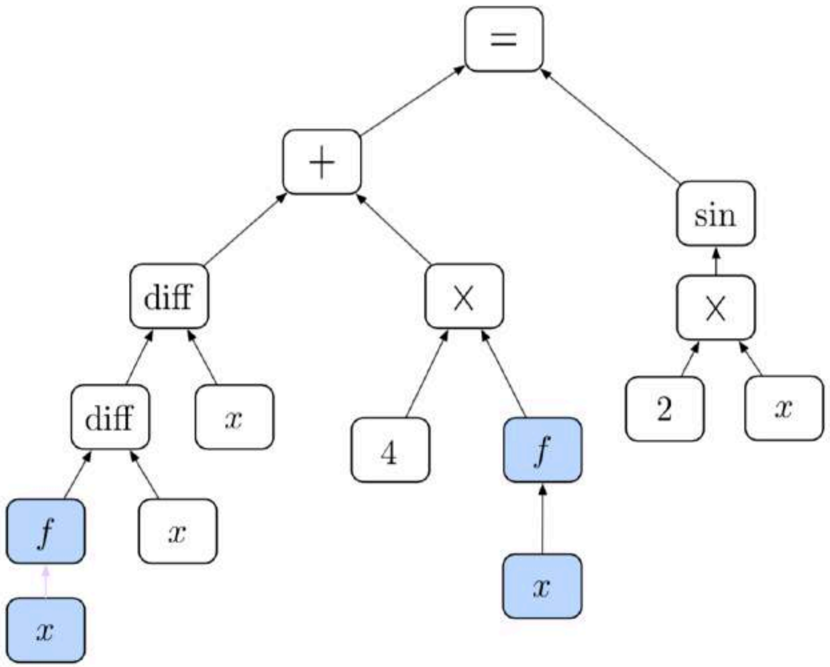
Ordinary Differential Equations

- n^{th} order ODE

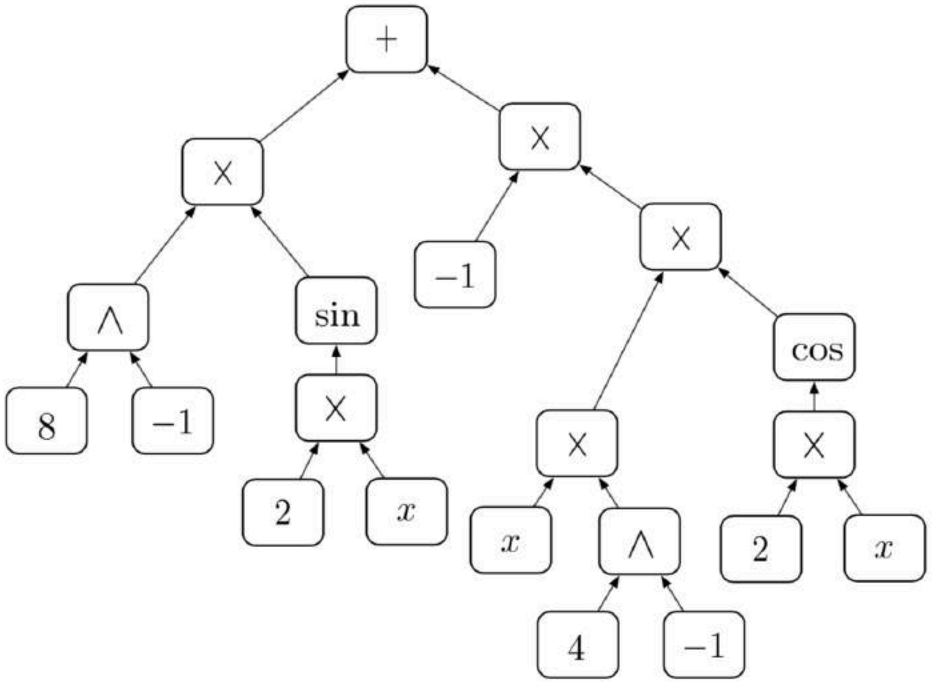
$$a_0(x)f(x) + a_1(x)\frac{\mathrm{d}f(x)}{\mathrm{d}x} + \cdots + a_n(x)\frac{\mathrm{d}^n f(x)}{\mathrm{d}x^n} = b(x)$$

- Find $f(x)$ that satisfies it

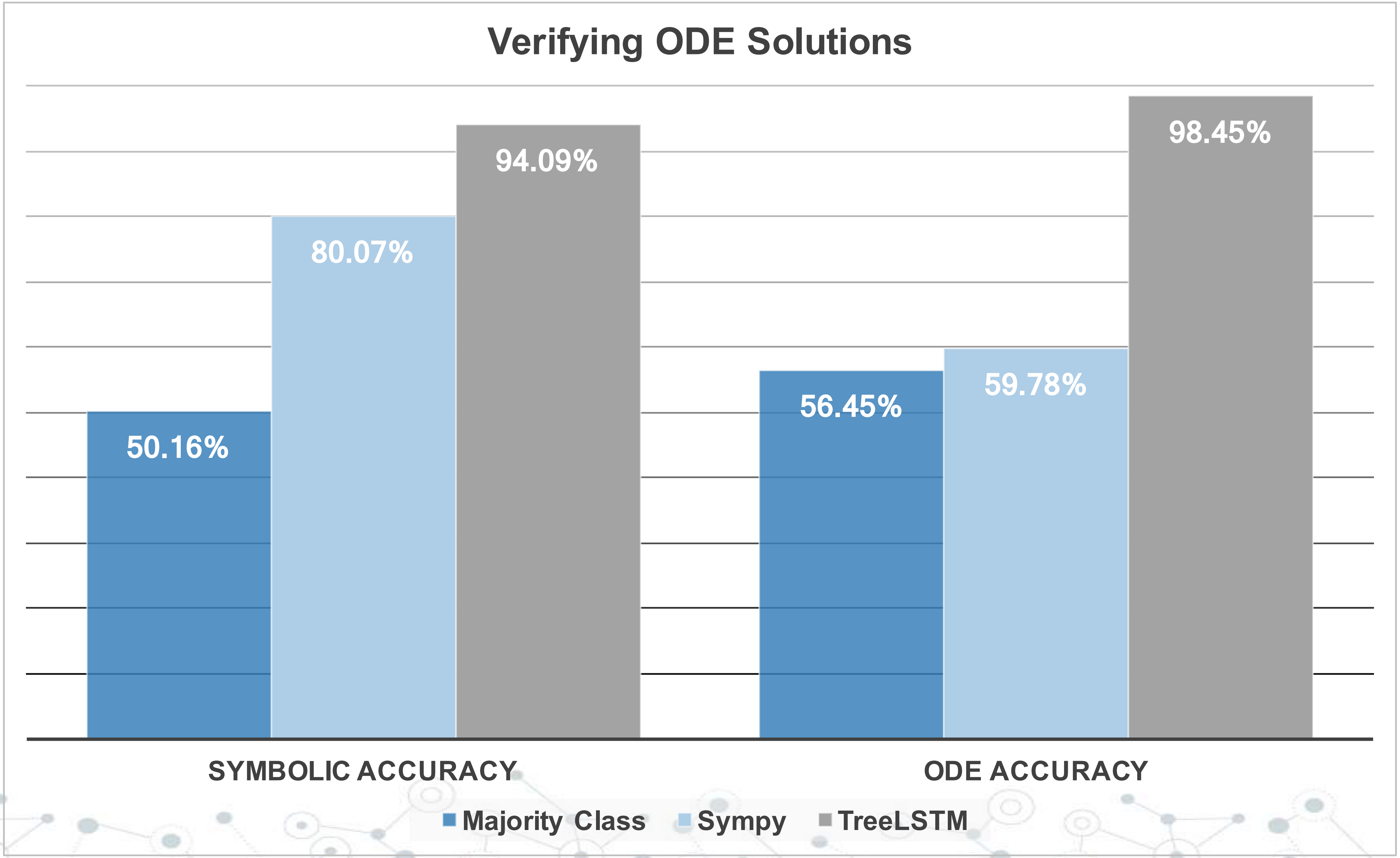
Extension to solving differential equations



$$\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} + 4f(x) = \sin(2x)$$



$$f(x) : \frac{1}{8}\sin(2x) - \frac{x}{4}\cos(2x)$$


Verifying ODE Solutions
50.16%
80.07%
94.09%
56.45%
59.78%
98.45%
SYMBOLIC ACCURACY
ODE ACCURACY
Majority Class
Sympy
TreeLSTM

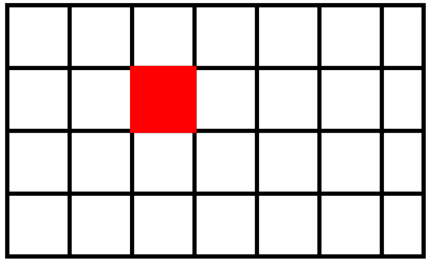
2.

Tensorized deep learning

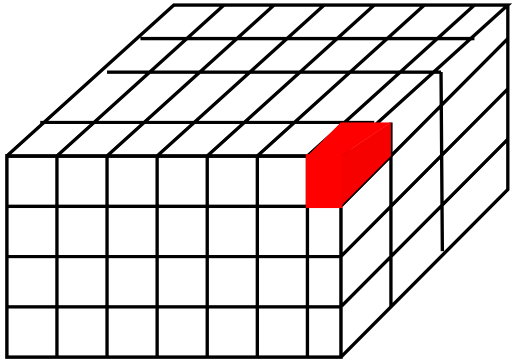
Tensors for multi-dimensional data and higher order moments

Images: 3 dimensions

Videos: 4 dimensions



Pairwise correlations



Triplet correlations

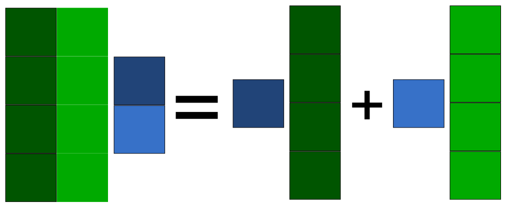
Operations on Tensors: Tensor Contraction

Tensor Contraction

Extends the notion of matrix product

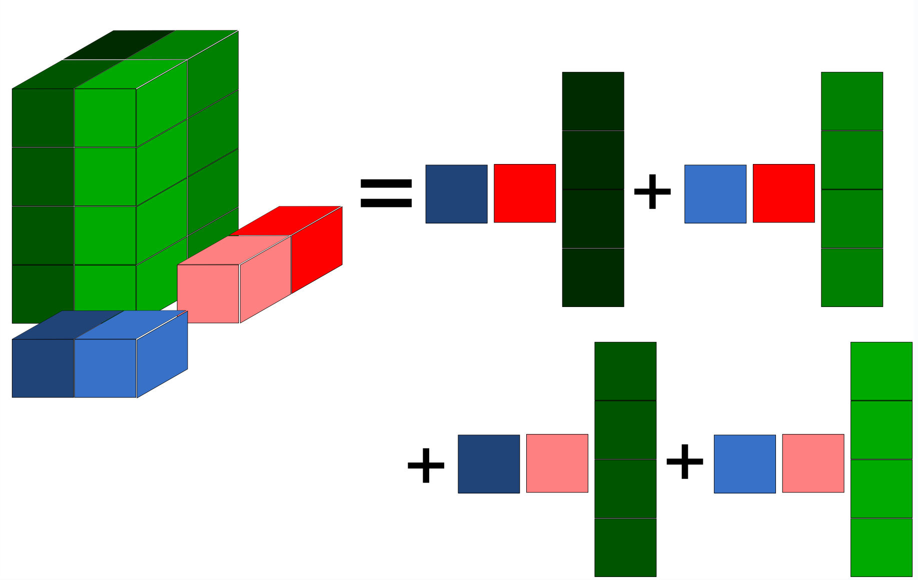
Matrix product

$$Mv = \sum_j v_j M_j$$

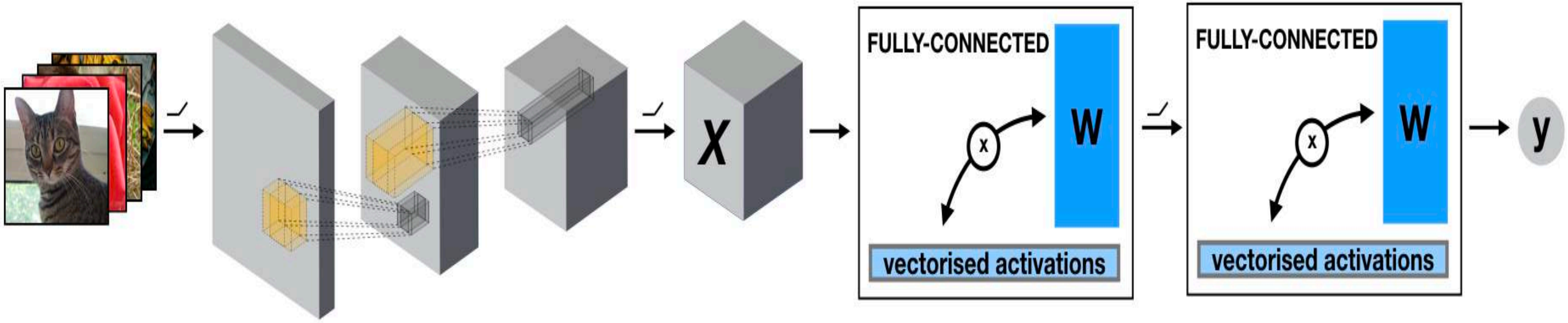


Tensor Contraction

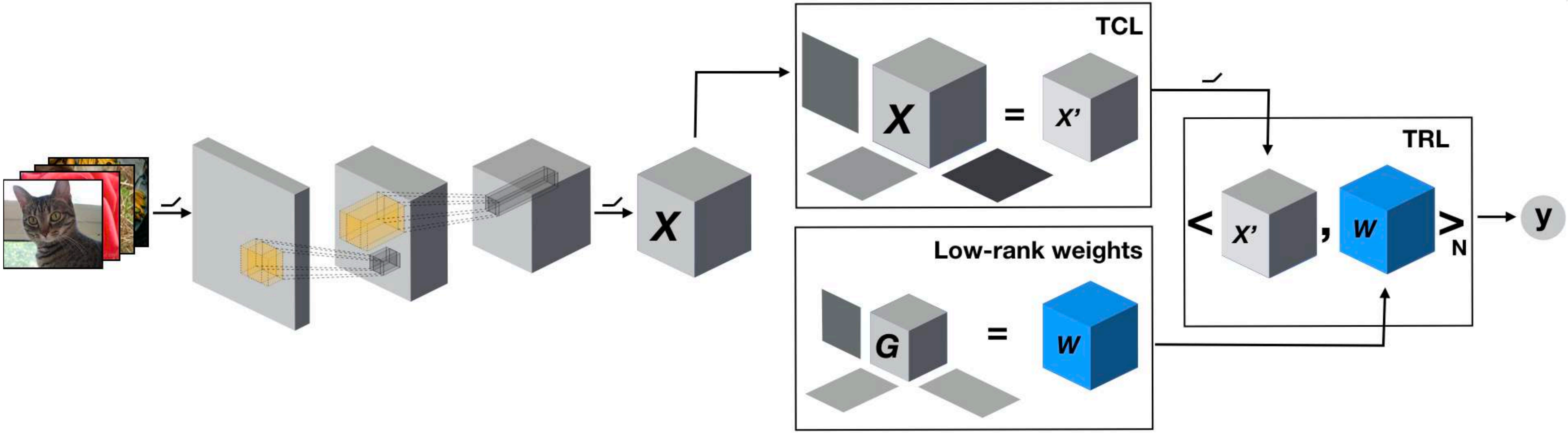
$$T(u, v, \cdot) = \sum_{i,j} u_i v_j T_{i,j,:}$$



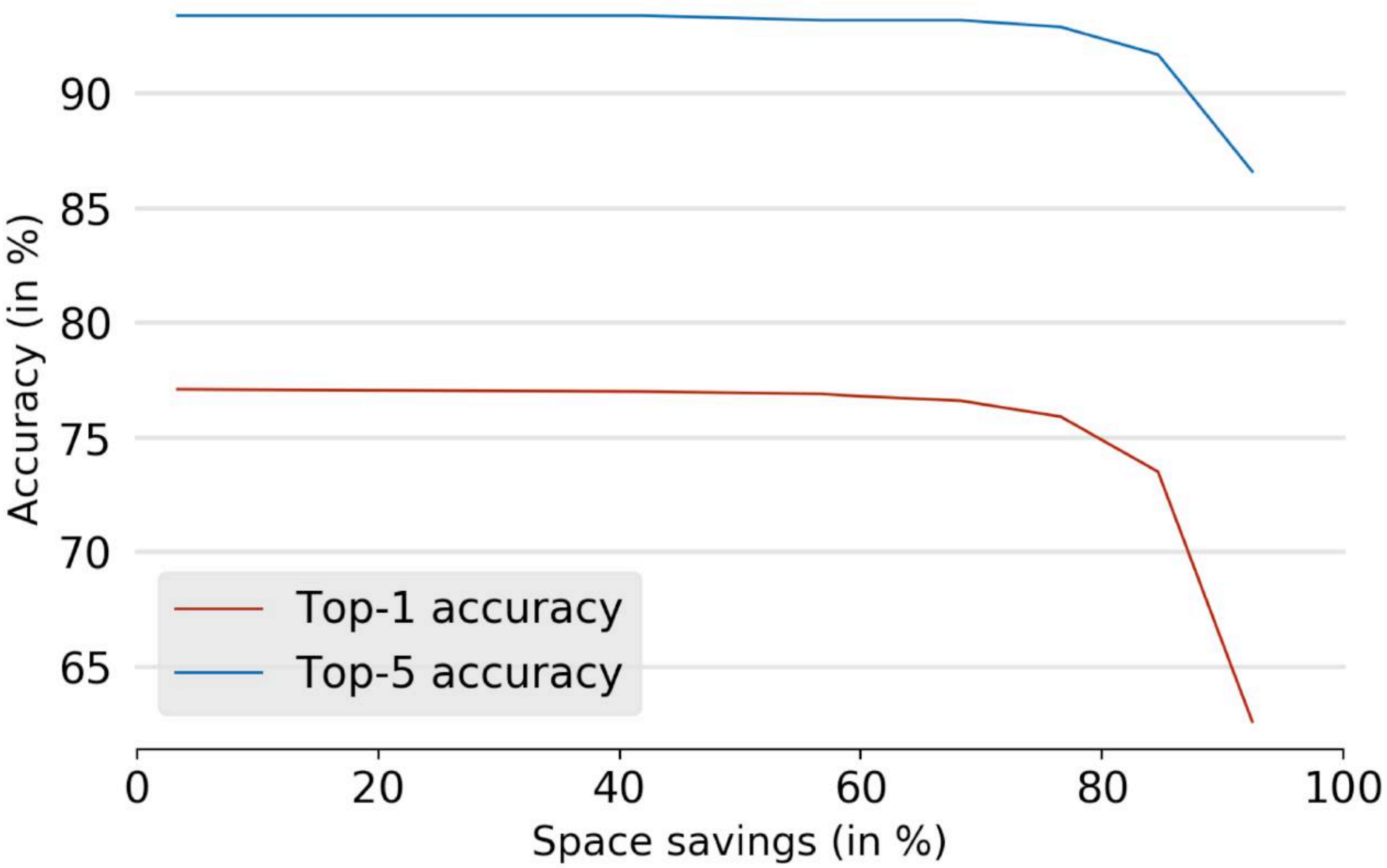
Deep Neural Nets: Transforming Tensors



Deep Tensorized Networks



Space Saving in Deep Tensorized Networks

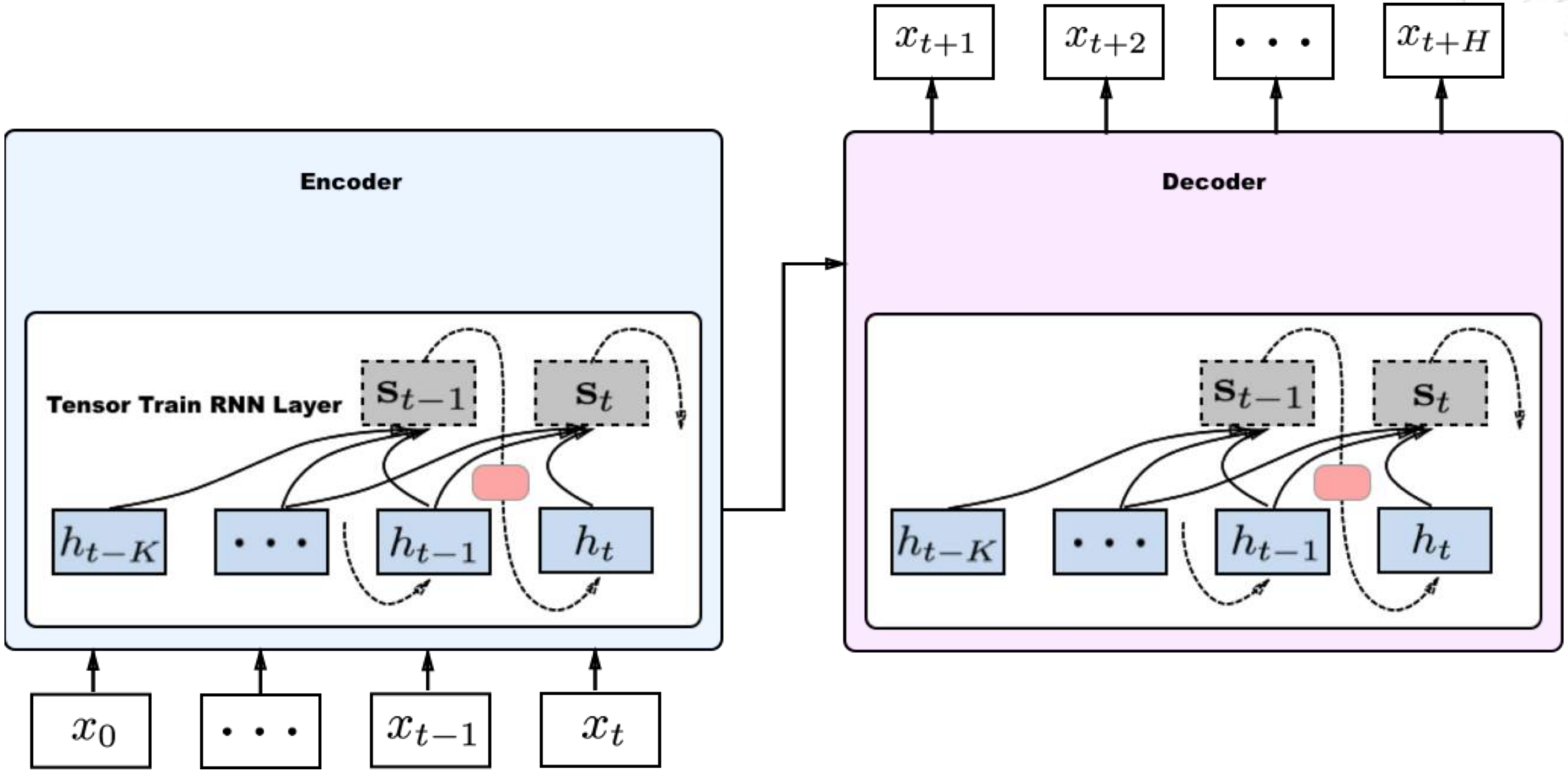


Tensors for long-term forecasting

Tensor Train RNN and LSTMs

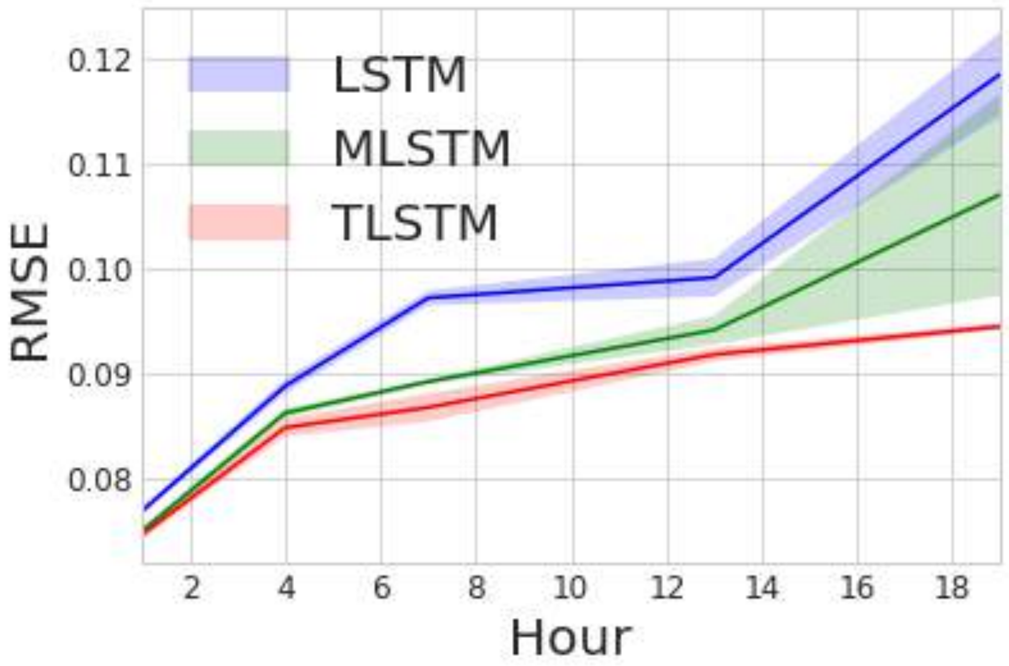
Challenges in forecasting:

- Long-term dependencies
- High-order correlations
- Error propagation

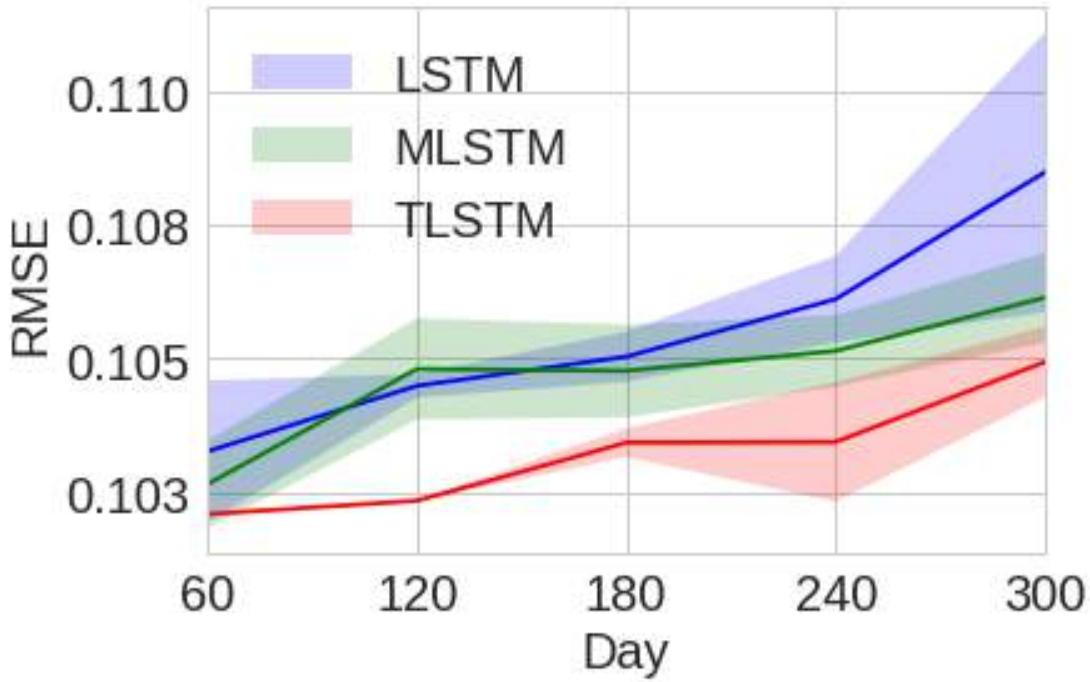


Tensor LSTM for Long-term Forecasting

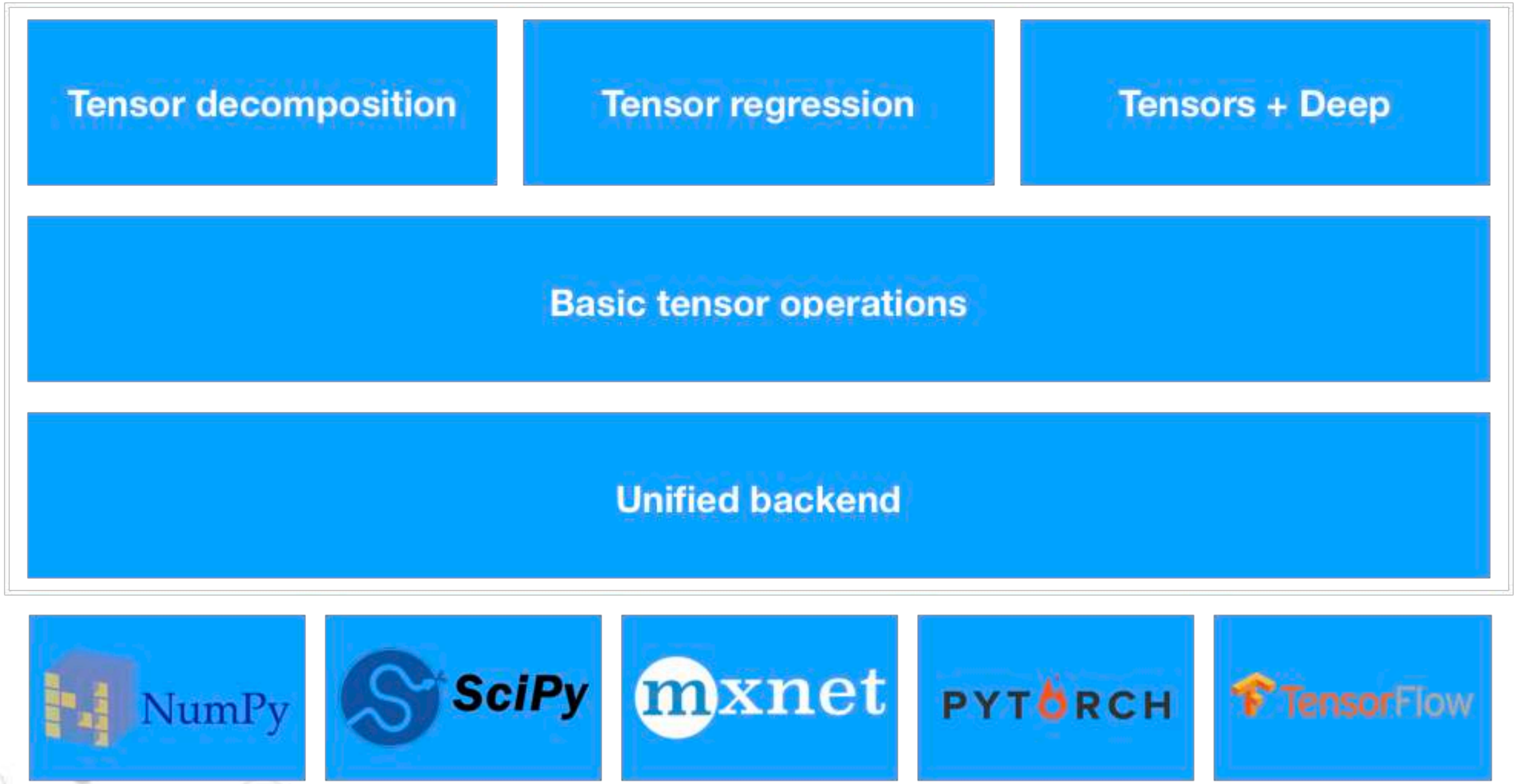
Traffic dataset



Climate dataset



Tensorly: High-level API for Tensor Algebra



- Python programming
- User-friendly API
- Multiple backends: flexible + scalable
- Example notebooks in repository

Unsupervised learning of Topic Models through tensor methods

Topics

Justice

Education

Sports

SECTIONS HOME SEARCH The New York Times

COLLEGE FOOTBALL

At Florida State, Football Clouds Justice

By MIKE McINTIRE and WALT BOGDANICH OCT. 10, 2014

Now, an examination by The New York Times of police and court records, along with interviews with crime witnesses, has found that, far from an aberration, the treatment of the Winston complaint was in keeping with the way the police on numerous occasions have soft-pedaled allegations of wrongdoing by Seminoles football players. From criminal mischief and motor-vehicle theft to domestic violence, arrests have been avoided, investigations have stalled and players have escaped serious consequences.

In a community whose self-image and economic well-being are so tightly bound to the fortunes of the nation's top-ranked college football team, law enforcement officers are finely attuned to a suspect's football connections. Those ties are cited repeatedly in police reports examined by The Times. What's more, dozens of officers work second jobs directing traffic and providing security at home football games, and many express their devotion to the Seminoles on social media.

TMZ, the gossip website, also requested the police report and later asked the school's deputy police chief, Jim L. Russell, if the campus police had interviewed Mr. Winston about the rape report. Mr. Russell responded by saying his officers were not investigating the case, omitting any reference to the city police, even though the campus police knew of their involvement. "Thank you for contacting me regarding this rumor — I am glad I can dispel that one!" Mr. Russell told TMZ in an email. The university said Mr. Russell was unaware of any other police investigation at the time of the inquiry. Soon after, the Tallahassee police belatedly sent their files to the news media and to the prosecutor, William N. Meggs. By then critical evidence had been lost and Mr. Meggs, who criticized the police's handling of the case, declined to

lson after the Seminoles' first game; five

am's second-leading receiver.

Rape Accusation

On Jan. 10, 2013, a female student at Florida State spotted the man she believed had raped her the previous month. After learning his name, Jameis Winston, she reported him to the Tallahassee police.

In the 21 months since, Florida State officials have said little about how they handled the case, which is no

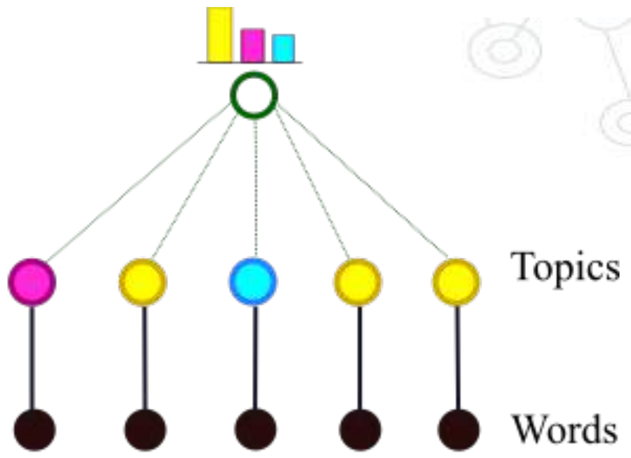
investigated by the federal Depart

Most recently, university officials suspended Mr. Winston for one game after he stood in a public place on campus and, playing off a running Internet gag, shouted a crude reference to a sex act. In a news conference afterward, his coach, Jimbo Fisher, said, "Our hope and belief is Jameis will learn from this and use better judgment and language and decision-making."

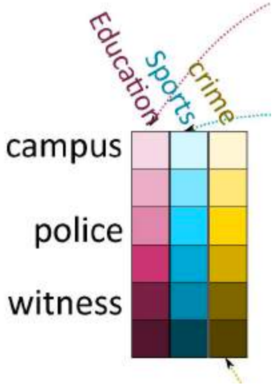
As The Times reported last April, the Tallahassee police also failed to aggressively investigate the rape accusation. It did not become public until November, when a Tampa reporter, Matt Baker, acting on a tip, sought records of the police investigation.

Upon learning of Mr. Baker's inquiry, Florida State, having shown little curiosity about the rape accusation, suddenly took a keen interest in the journalist seeking to report it, according to emails obtained by The Times.

"Can you share any details on the requesting source?" David Perry, the university's police chief, asked the Tallahassee police. Several hours later, Mr.

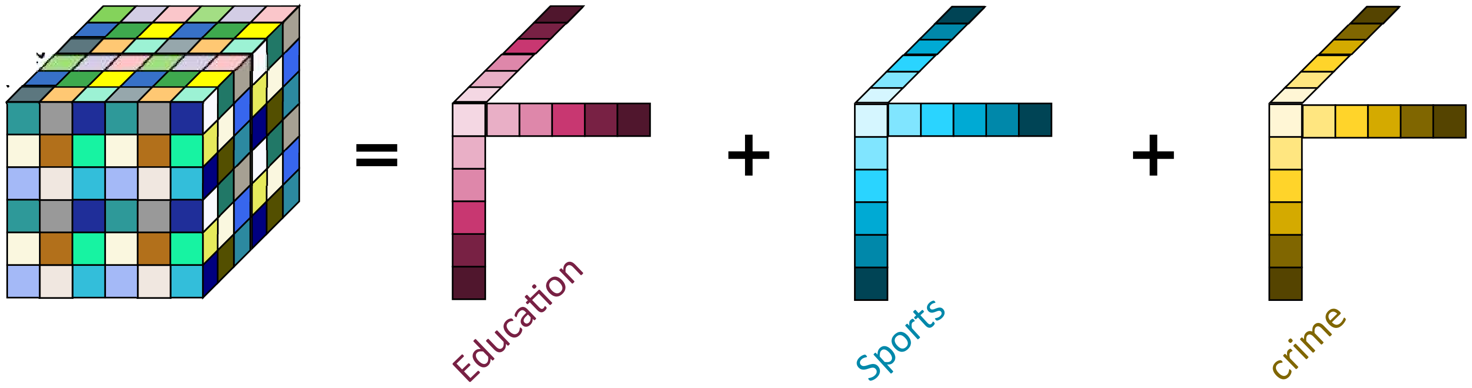


Learning LDA Model through tensors



- Topic-word matrix $\mathrm{P}[\text{word} = i | \text{topic} = j]$
- Topic proportions $\mathrm{P}[\text{topic} = j | \text{document}]$

Moment Tensor: Co-occurrence of Word Triplets



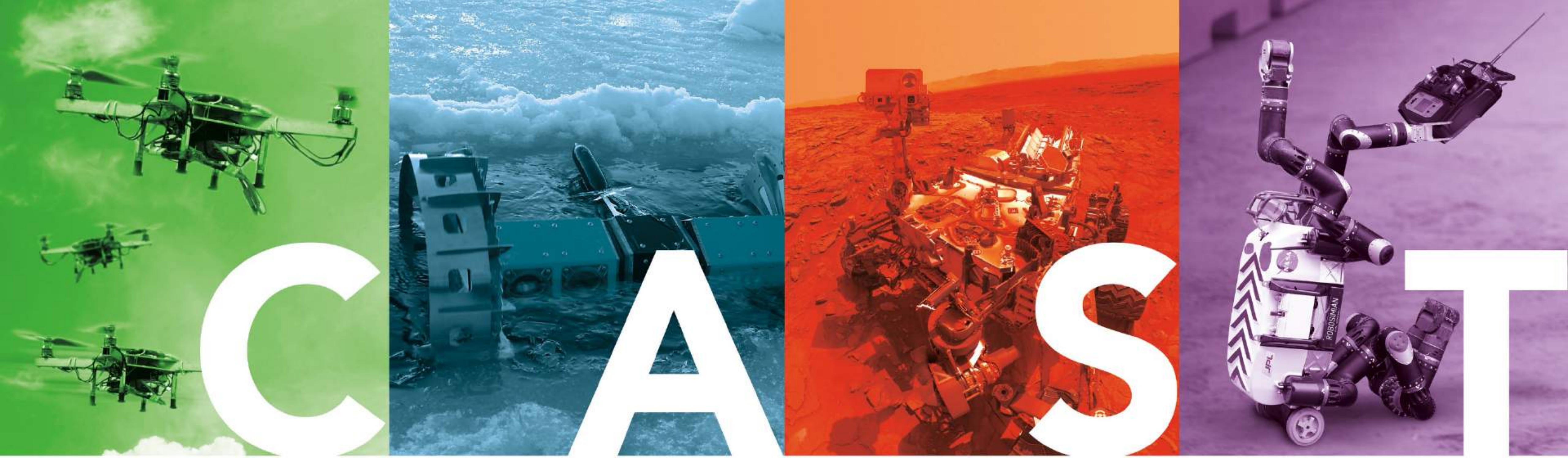
Tensor-based LDA training is faster

- Mallet is an open-source framework for topic modeling
- Benchmarks on **AWS SageMaker Platform**
- Bulit into **AWS Comprehend NLP service.**

3.

Learning to land a drone

Guanya Shi, Xichen Shi, Michael O'Connell, Rose Yu, Kamyar Azizzadenesheli,
A. , Yisong Yue, and Soon-Jo Chung



Center for Autonomous Systems and Technologies

A New Vision for Autonomy

Caltech

Physical Model for a Quadrotor drone

- Dynamics:

$$\begin{cases} \dot{\mathbf{p}} = \mathbf{v}, & m\dot{\mathbf{v}} = m\mathbf{g} + R\mathbf{f}_u + \mathbf{f}_a \\ \dot{R} = RS(\boldsymbol{\omega}), & J\dot{\boldsymbol{\omega}} = J\boldsymbol{\omega} \times \boldsymbol{\omega} + \boldsymbol{\tau}_u + \boldsymbol{\tau}_a \end{cases}$$

- Control:

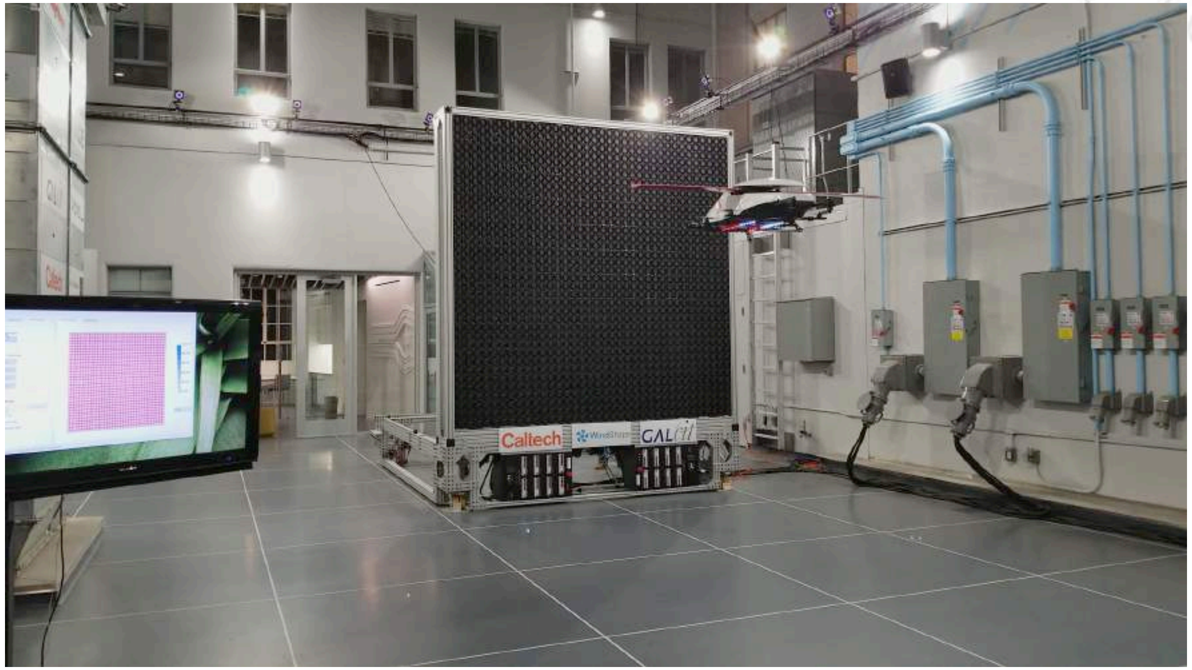
$$\begin{cases} \mathbf{f}_u = [0, 0, T]^\top \\ \boldsymbol{\tau}_u = [\tau_x, \tau_y, \tau_z]^\top \\ \begin{bmatrix} T \\ \tau_x \\ \tau_y \\ \tau_z \end{bmatrix} = \begin{bmatrix} c_T & c_T & c_T & c_T \\ 0 & c_T l_{\text{arm}} & 0 & -c_T l_{\text{arm}} \\ -c_T l_{\text{arm}} & 0 & c_T l_{\text{arm}} & 0 \\ -c_Q & c_Q & -c_Q & c_Q \end{bmatrix} \begin{bmatrix} n_1^2 \\ n_2^2 \\ n_3^2 \\ n_4^2 \end{bmatrix} \end{cases}$$

- Unknown forces & moments:

$$\begin{cases} \mathbf{f}_a = [f_{a,x}, f_{a,y}, f_{a,z}]^\top \\ \boldsymbol{\tau}_a = [\tau_{a,x}, \tau_{a,y}, \tau_{a,z}]^\top \end{cases}$$

Challenges in landing a Quadrotor drone

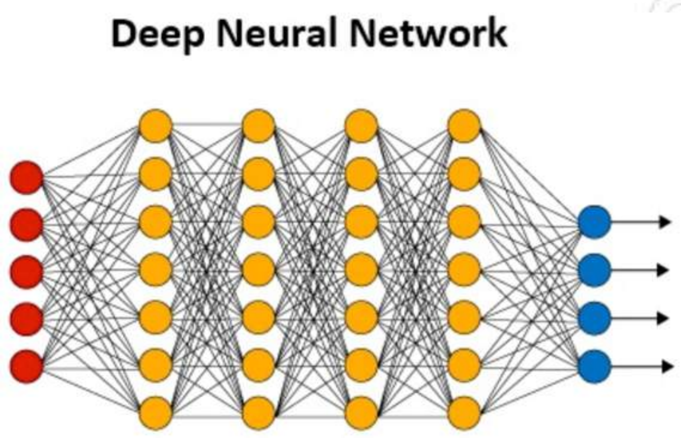
- Unknown aerodynamic forces & moments.
- *Example 1:* when drone is close to ground.
- *Example 2:* as velocity goes up, air drag.
- *Example 3:* external wind conditions.



Wind generation in
CALTECH CAST wind tunnel

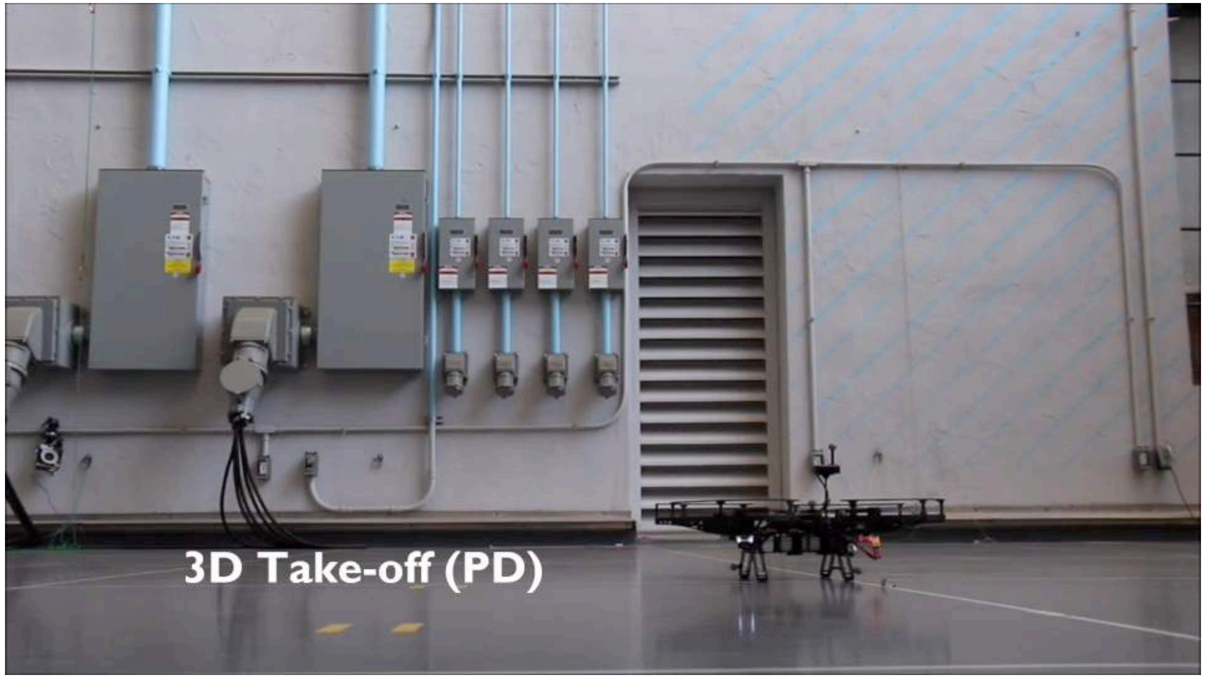
Challenges in using DNNs to Learn Unknown Dynamics

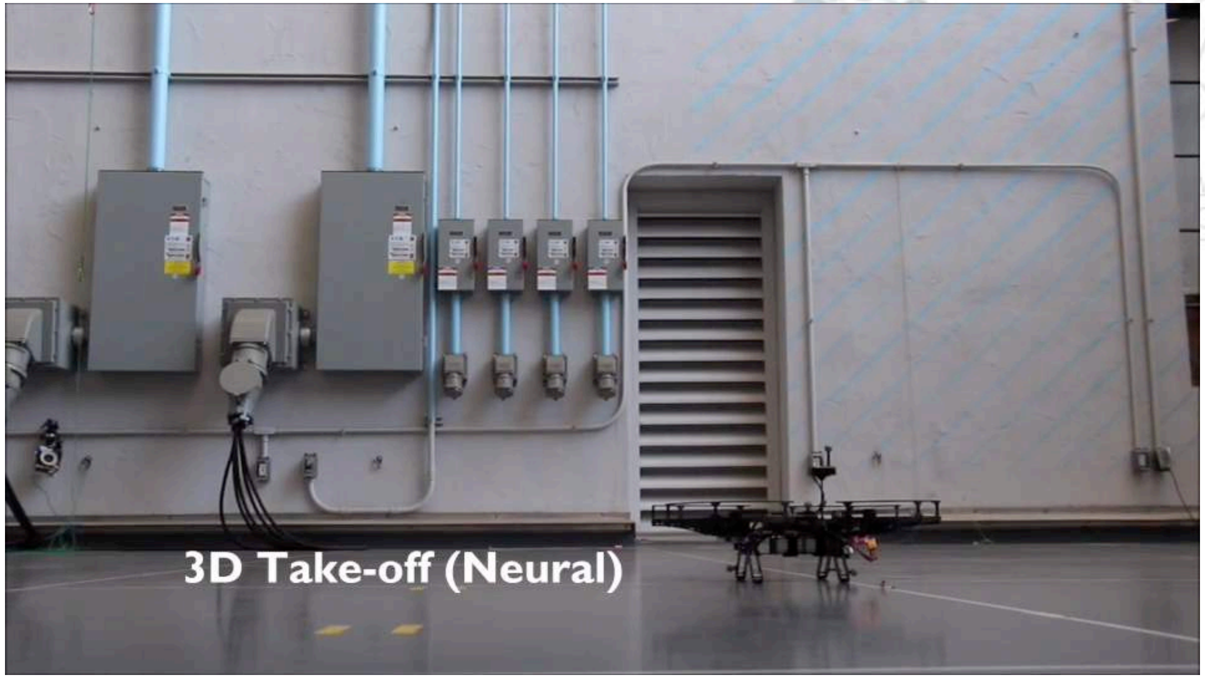
- Our idea: using DNNs to learn unknown aerodynamic forces and then design nonlinear controller to cancel it (unknown moments are very limited in landing)
- *Challenge 1:* DNNs are data-hungry
- *Challenge 2:* DNNs can be unstable and generate unpredictable output
- *Challenge 3:* DNNs are difficult to analyze and design provably stable controller based on them
- **Our approach**: using Spectral Normalization to control Lipschitz property of DNNs and then design stable nonlinear controller (**Neural-Lander**)



Neural Lander Demo 1

Neural Lander Demo 2



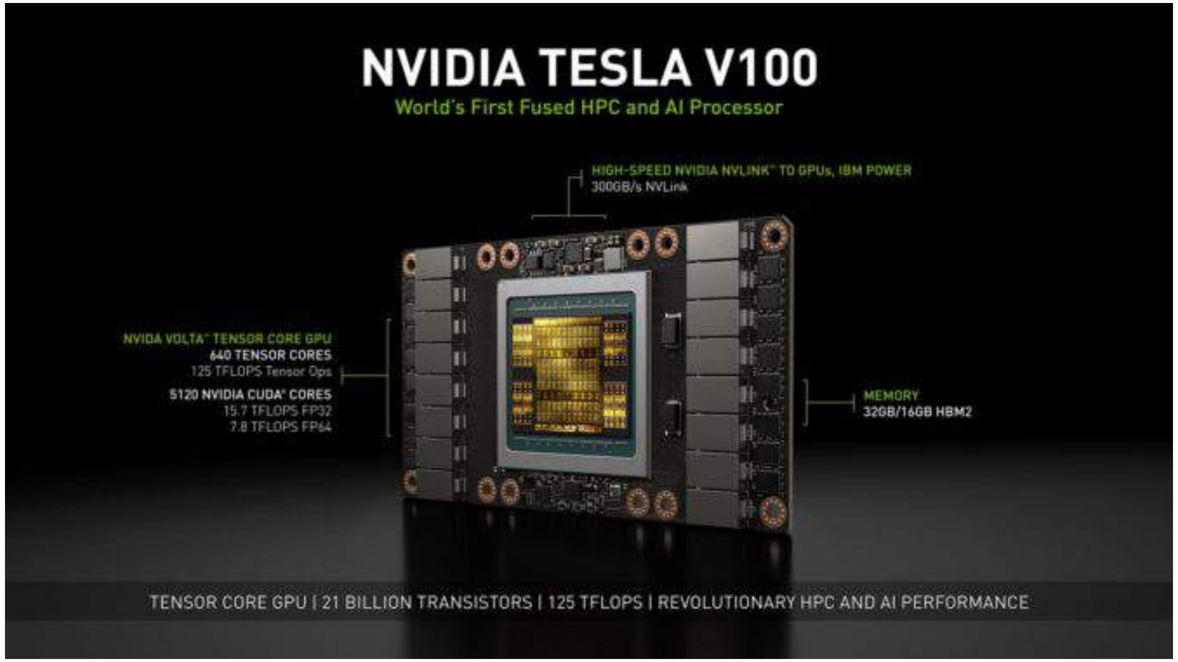


4.

Lots of efforts at NVIDIA..

Exascale Deep Learning for Climate Analytics

- 3 Exaops for AI
- ~27k Volta 100 GPUs



Some research leaders at NVIDIA

Chief Scientist

Bill Dally

Graphics

Dave Luebke

Alex Keller

Aaron Lefohn

Learning & Perception

Jan Kautz

Robotics

Dieter Fox

Computer vision

Sanja Fidler

Core ML

Me !

Applied research

Bryan Catanzaro

Michael Garland

Networks

Larry Dennison

Architecture

Steve Keckler

Dave Nellans

Mike O'Connor

VLSI

Brucek Khailany

Circuits

Tom Gray

Conclusions

- Rich opportunities to infuse physical domain knowledge in AI algorithms.
- Jointly using **symbolic** and numerical data greatly helps neural programming generalization.
- **Tensors** expand learning into any dimension. Tensorized neural networks capture dependencies better.
- Learning unknown **aerodynamics** using spectrally normalized DNNs.
- Many efforts at NVIDIA to scale AI/ML for physics applications.

Thanks!

Any questions?