



PURDUE
UNIVERSITY®

Department of Computer Science

Interactive Debugging and Testing Support for Deep Learning

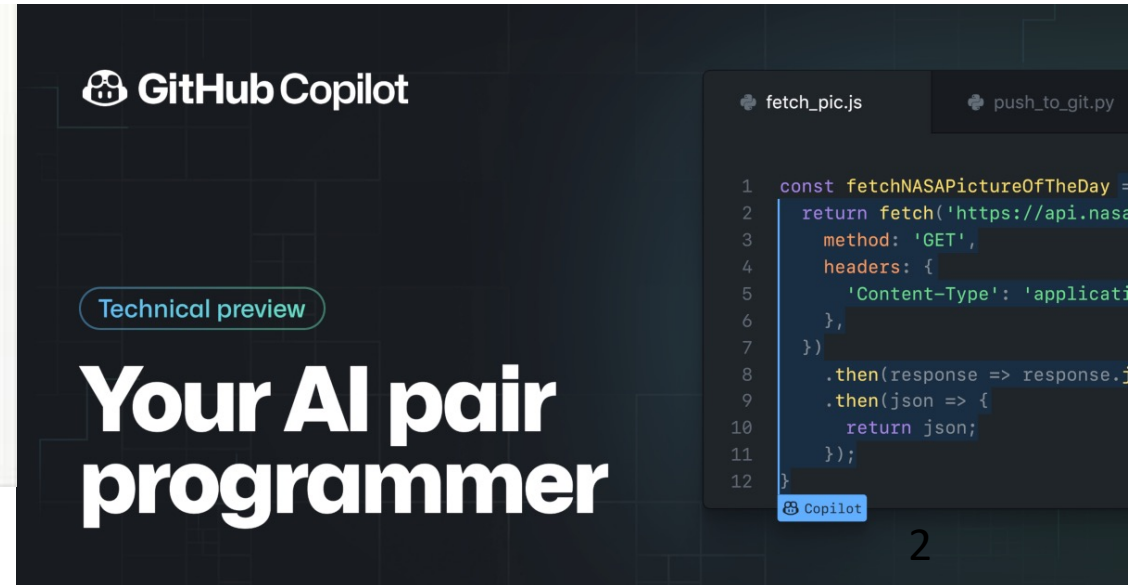
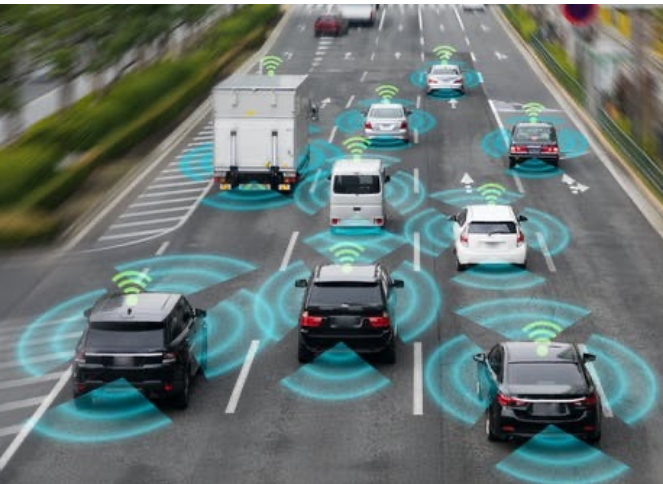
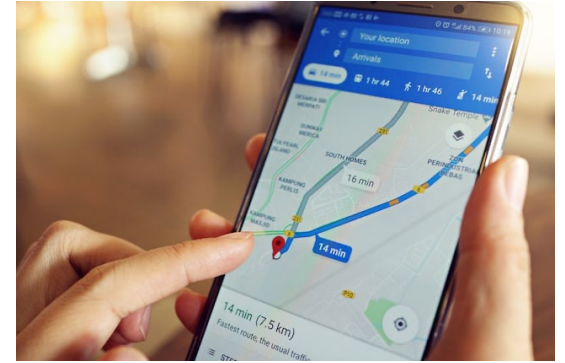
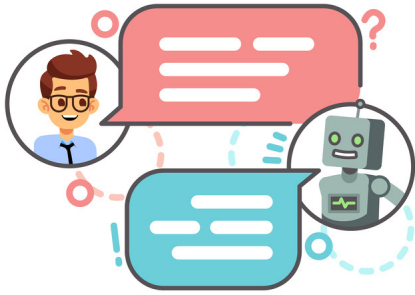
Tianyi Zhang

Assistant Professor of Computer Science

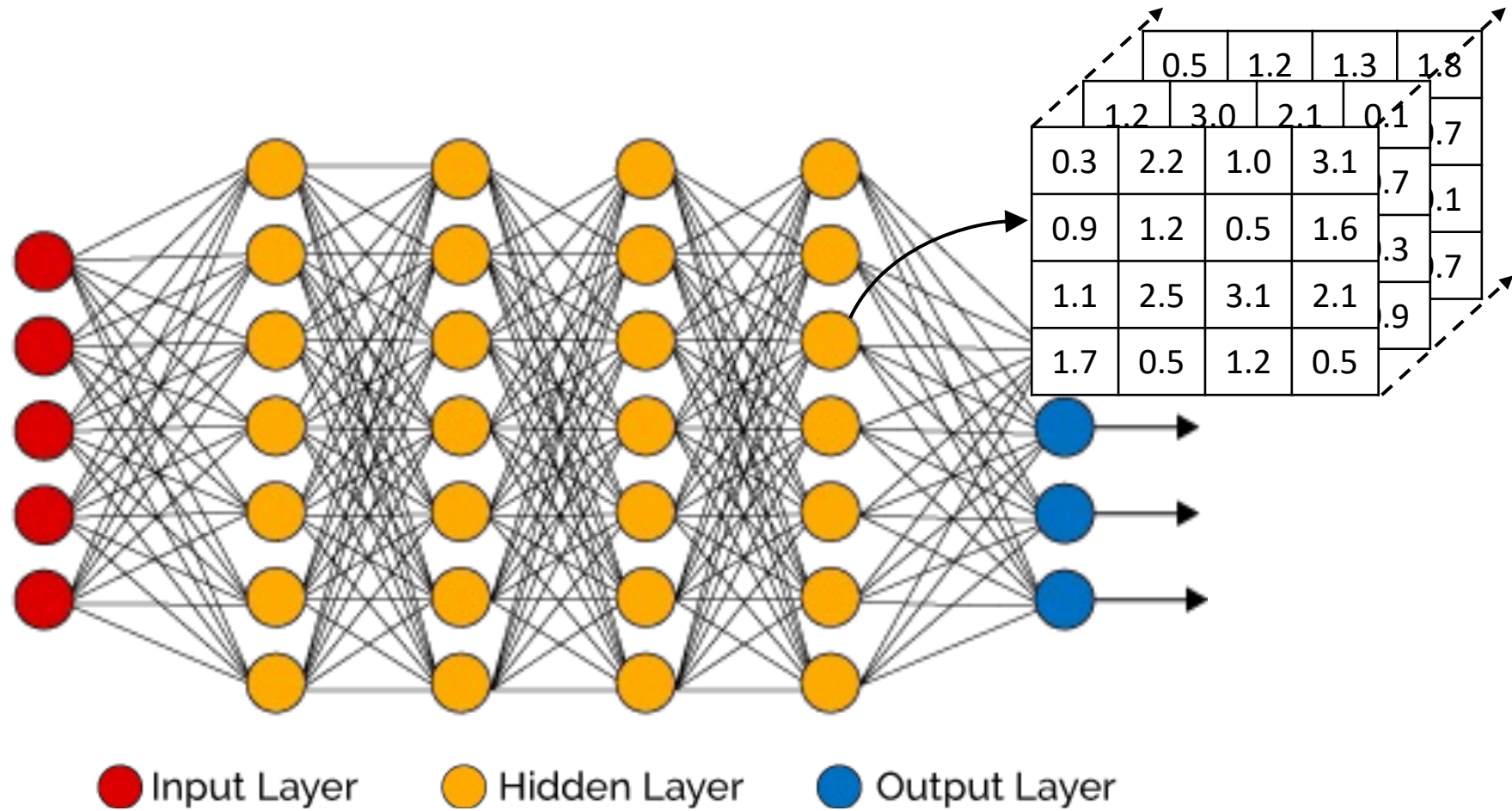
Purdue University



AI is eating software

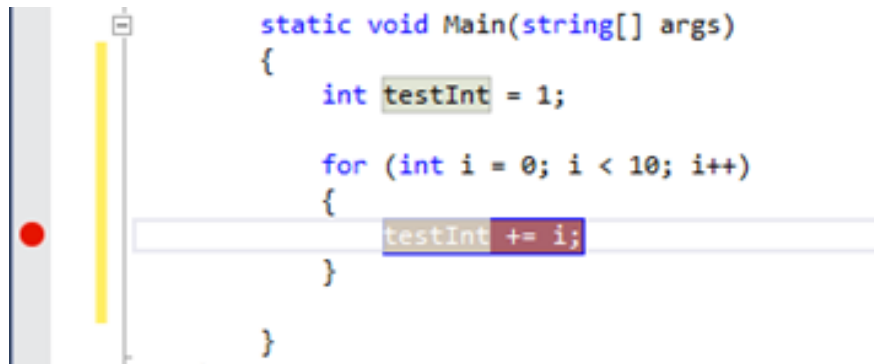


The Key Enabler: Deep Learning



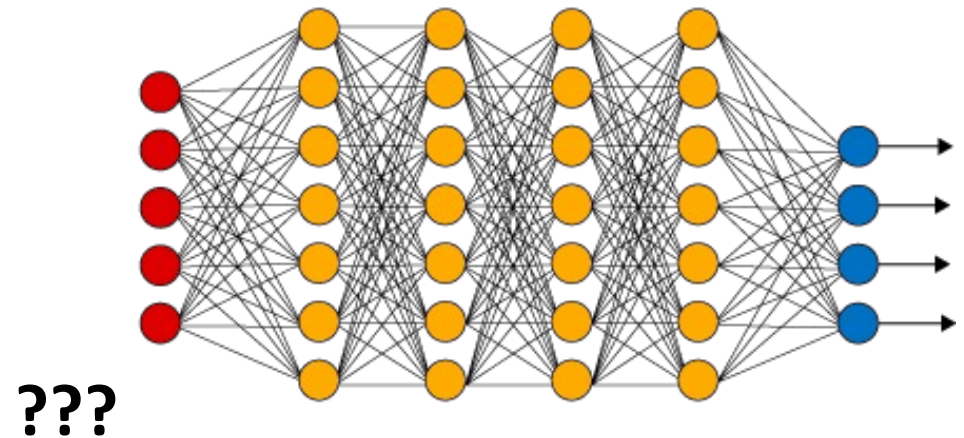
DNNs are hard to debug

Set a breakpoint & step through



traditional software

How to set a break point?



0.3	2.2	1.0	3.1
0.9	1.2	0.5	1.6
1.1	2.5	3.1	2.1
1.7	0.5	1.2	0.5

neural network

Motivating Example: An RNN model for sentiment analysis



100,000 comments
from an online forum

*His ruling makes **perfect** sense. Especially
when you read the citations of law.*

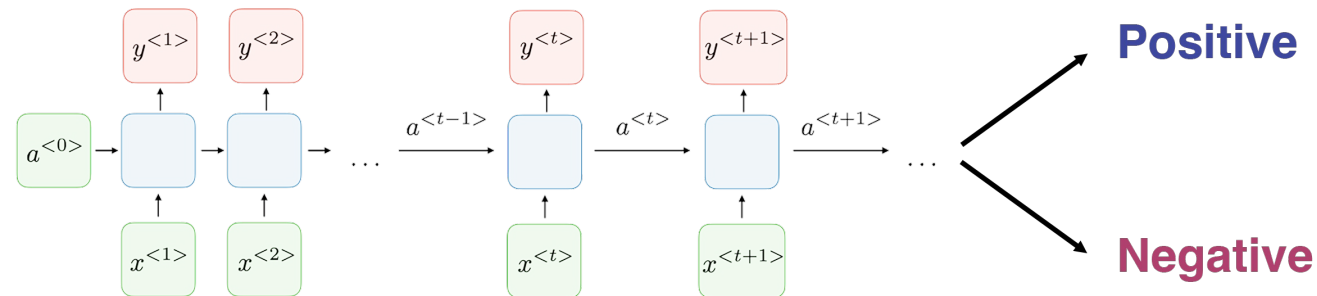


Positive

*So John Kelly is just a more useful **idiot**
than the Mooch.*



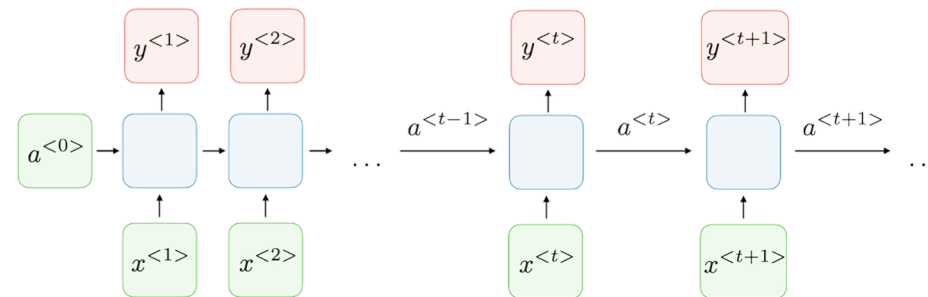
Negative



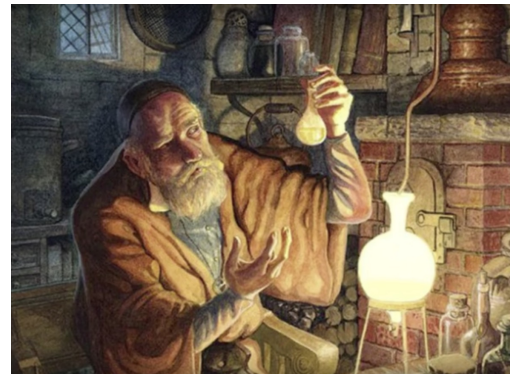
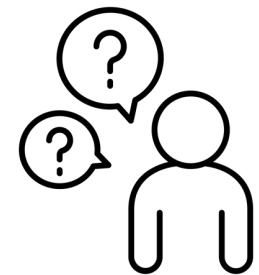
Motivating Example: An RNN model for sentiment analysis

How naïve is Justin. How disingenuous of himself. It's not the money silly. It's being able to plant the seeds of ideas, concepts in the Minister's so that he or she will think favourably on your behalf the next time a related issue raises its ugly head.

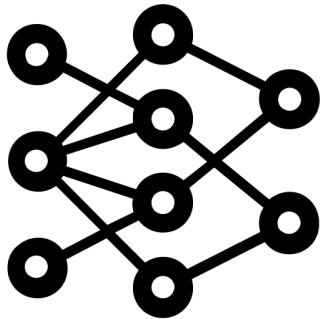
- A comment from Reddit



Positive



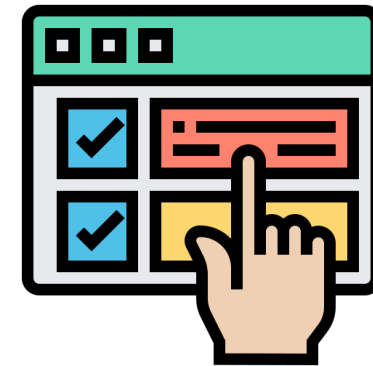
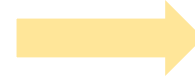
Our Solution: Abstracting DNNs to Finite State Machines



Complex DNN



Finite State Machine



Interaction Support

DeepSeer: An Interactive System for Debugging and Testing RNN Models

The screenshot displays the DeepSeer web interface, which is used for debugging and testing RNN models. The interface is divided into several sections:

- Header:** "DeepSeer" with a menu icon.
- State Diagram:** A network graph with nodes (0-39) and edges. A tooltip for Node 39 shows: "Node: 39, Positive: 5046, Negative: 39948".
- Pattern Summary:** A list of influential and possible buggy patterns. The "INFLUENTIAL PATTERN" section includes:
 - 19 26 13 9 yp oc rites
 - 20 3 11 13 what a stupid
 - 21 3 11 22 the catholic church
 - 22 9 9 9 and inc est
 - 23 11 11 15 lock him up
 - 24 3 25 23 she is a
 - 25 13 13 22 white su pre
 - 26 32 13 1 big ot ry
 - 27 38 13 9 the clown prince
- Instance:** A footer section with filters and statistics: "FILTERS DENSITY TRAIN/TEST POSITIVE: 92739 / NEGATIVE: 7261".

DeepSeer: An Interactive System for Debugging and Testing RNN Models

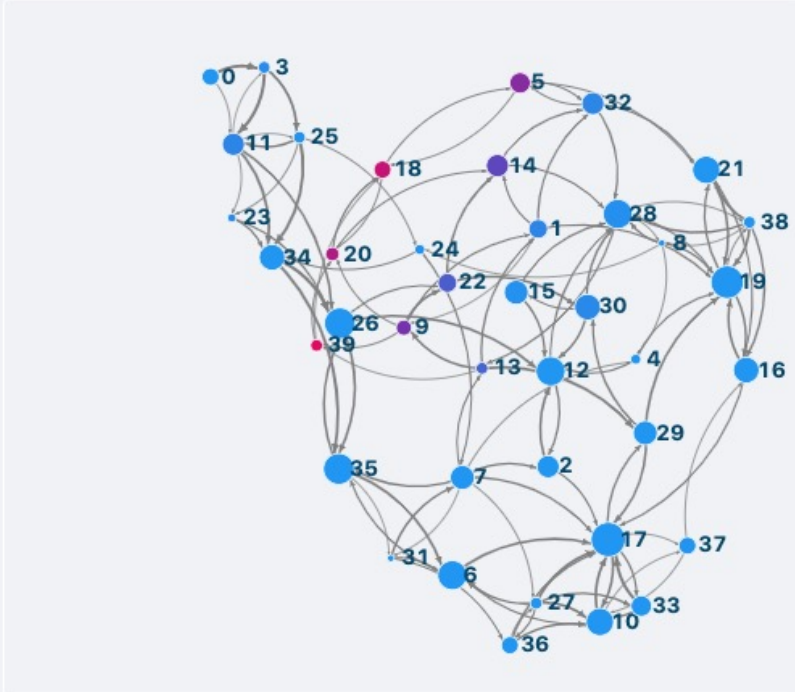
DeepSeer

State Diagram

CLEAR X

Positive Negative

Type to inference a new sentence...



Pattern Summary

INFLUENTIAL PATTERN POSSIBLE BUGGY PATTERN

- 19 26 13 9 yp oc rites
- 20 3 11 13 what a stupid
- 21 3 11 22 the catholic church
- 22 9 9 9 and inc est
- 23 11 11 15 lock him up
- 24 3 25 23 she is a
- 25 13 13 22 white su pre
- 26 32 13 1 big ot ry
- 27 38 13 9 the clown prince

Instance

FILTERS DENSITY TRAIN/TEST POSITIVE: 92739 / NEGATIVE: 7261

Search...

DeepSeer: An Interactive System for Debugging and Testing RNN Models

How naive of Justin. How disingenuous of himself. It's not the money silly. It's being able to plant seeds of ideas, concepts, etc. in the Minister's head so that he or she will think favourably on your behalf the next time a related issue raises its ugly head. Close quarters lobbying I would call it.



Input text: how naive of justin . how di sing en uous of himself . it ' s not the money silly . it ' s being able to plant seeds of ideas , concepts , etc . in the minister ' s head so that he or she will think favour ably on your behalf the next time a related issue raises its ugly head . close quarters lobbying i would call it .

POSITIVE 😊

Trace:

3→13→9→22→22→22→22→30→30→14→14→14→14→1
4→14→14→14→14→14→13→9→9→9→9→9→22→22→1
4→14→32→32→32→32→32→32→32→32→5→5→5→5→
32→32→32→32→5→32→32→32→32→32→32→32→5→5
→5→5→5→5→5→5→5→13→9→9→9→22→1→1→32→32
→32→32

DeepSeer: An Interactive System for Debugging and Testing RNN Models

Further investigation reveals imbalanced training data with the word, “quarters”.

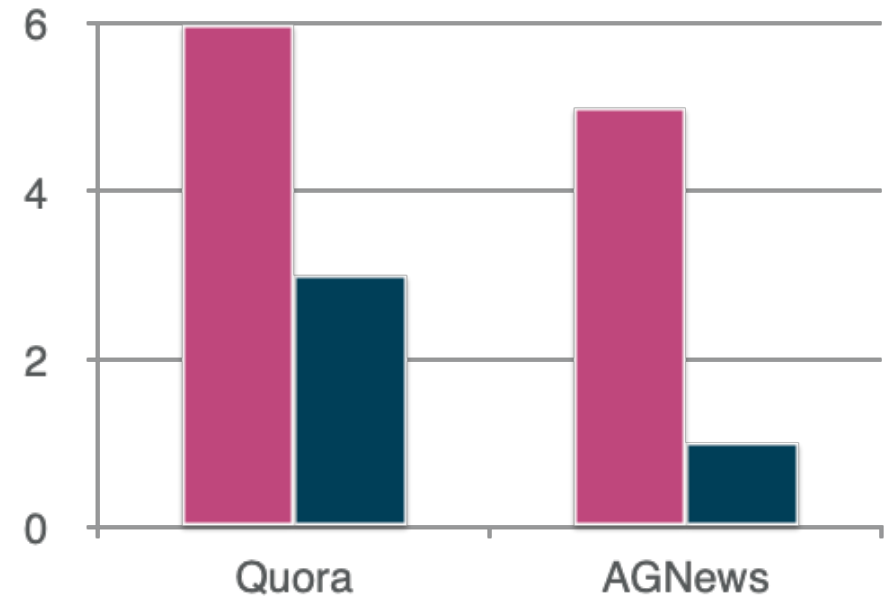
The screenshot displays the DeepSeer interface. At the top, there is a search bar with the query "re' quarters" and a search icon. Below the search bar, a filter bar shows "POSITIVE: 27 / NEGATIVE: 1" highlighted with a red box. The main content area is a table with columns for "Index", "Trace", and "Text". The table lists several instances with their corresponding traces and text snippets.

Index	Trace	Text
2259	3→13→9→22→22→22→22→22→14→14→14→14→14	the arrogant self - righteous ness of the media is
10349	3→25→11→11→11→34→26→26→26→26→2→2→2→2→2→2	here ' s the real reason behind dow 21 , 000 (no
13879	3→11→11→11→35→35→35→35→35→35→35→35→35→3	accepting the national constitution center ' s liber
15103	3→11→11→11→34→26→35→35→35→35→35→35→35→3!	so the plant has lost some money in most years -
17269	3→11→11→11→11→34→34→35→35→4→7→26→26→35→35→	one of the things , margaret , that makes you a fa

The medium number of reasonable explanations per user



The medium number of error-inducing keywords found by users



■ DeepSeer

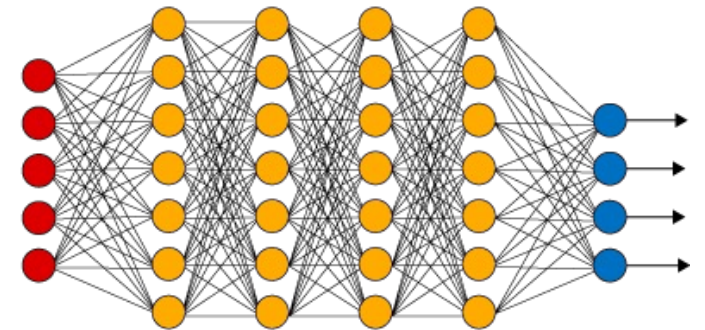
■ LIME

Key Takeaways

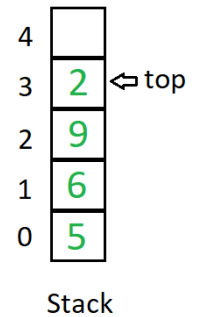
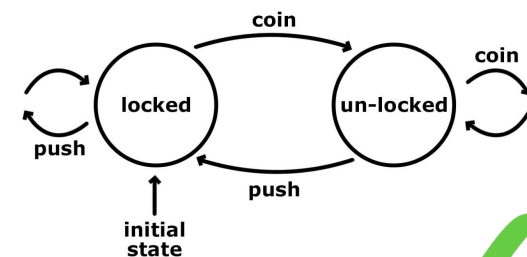
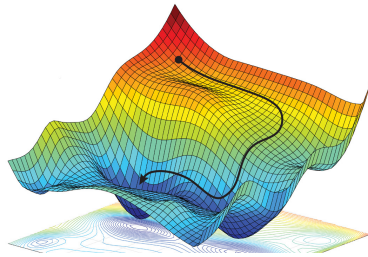


New debugging tools

Responsible AI



		0.5	1.2	1.3	1.8
	1.2	3.0	2.1	0.1	.7
0.3	2.2	1.0	3.1	.7	.1
0.9	1.2	0.5	1.6	.3	.7
1.1	2.5	3.1	2.1	.9	
1.7	0.5	1.2	0.5		



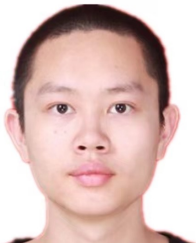
Collaborators:



Prof. Lei Ma



Zhijie Wang



Yuheng Huang



Da Song



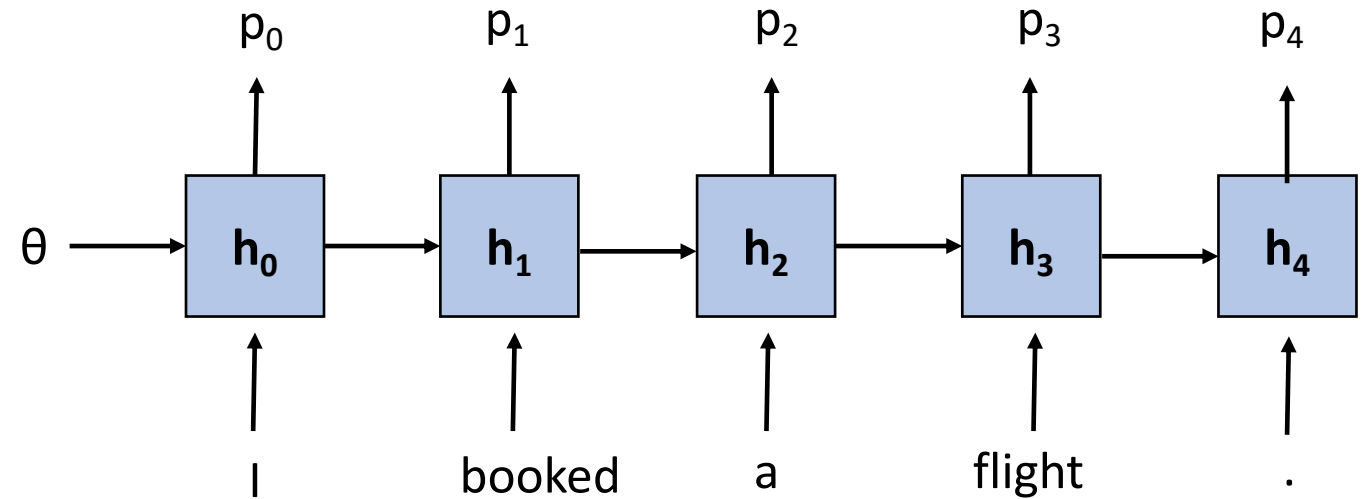
**UNIVERSITY
OF ALBERTA**

tianyi@purdue.edu

<https://tianyi-zhang.github.io/>



Approach Overview

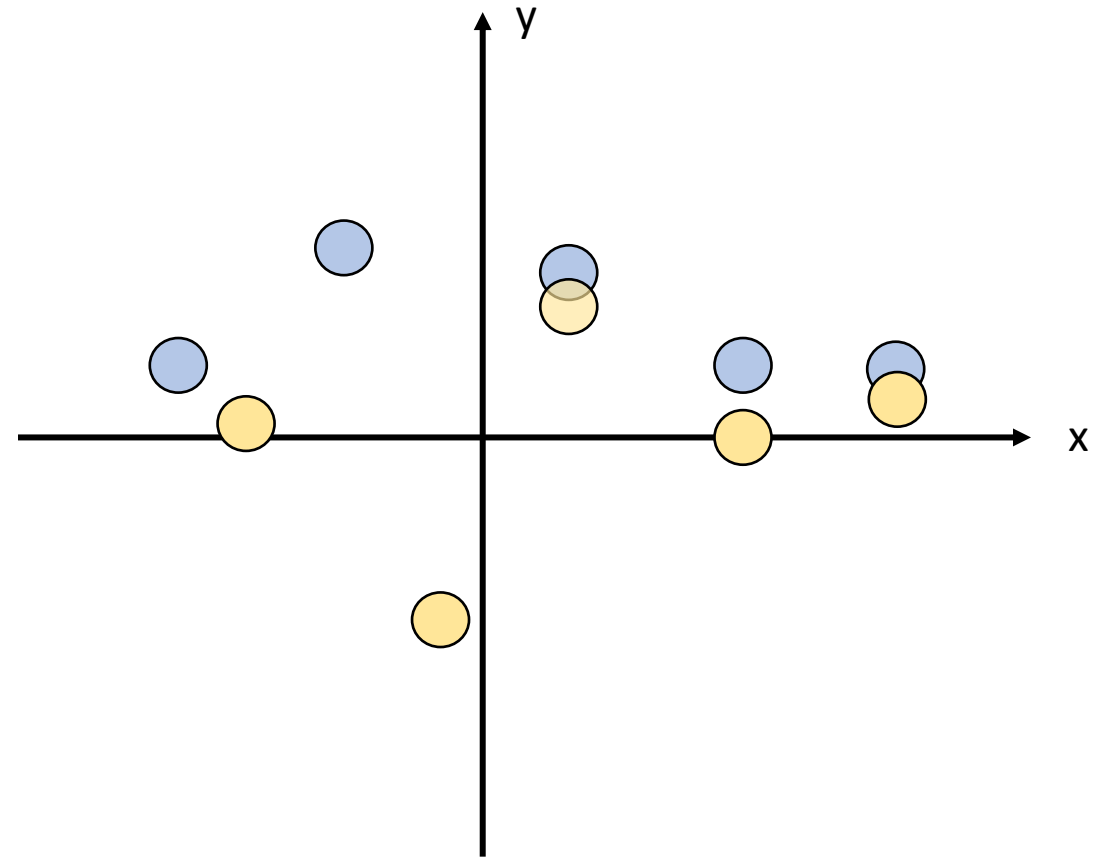
- State profiling



Approach Overview

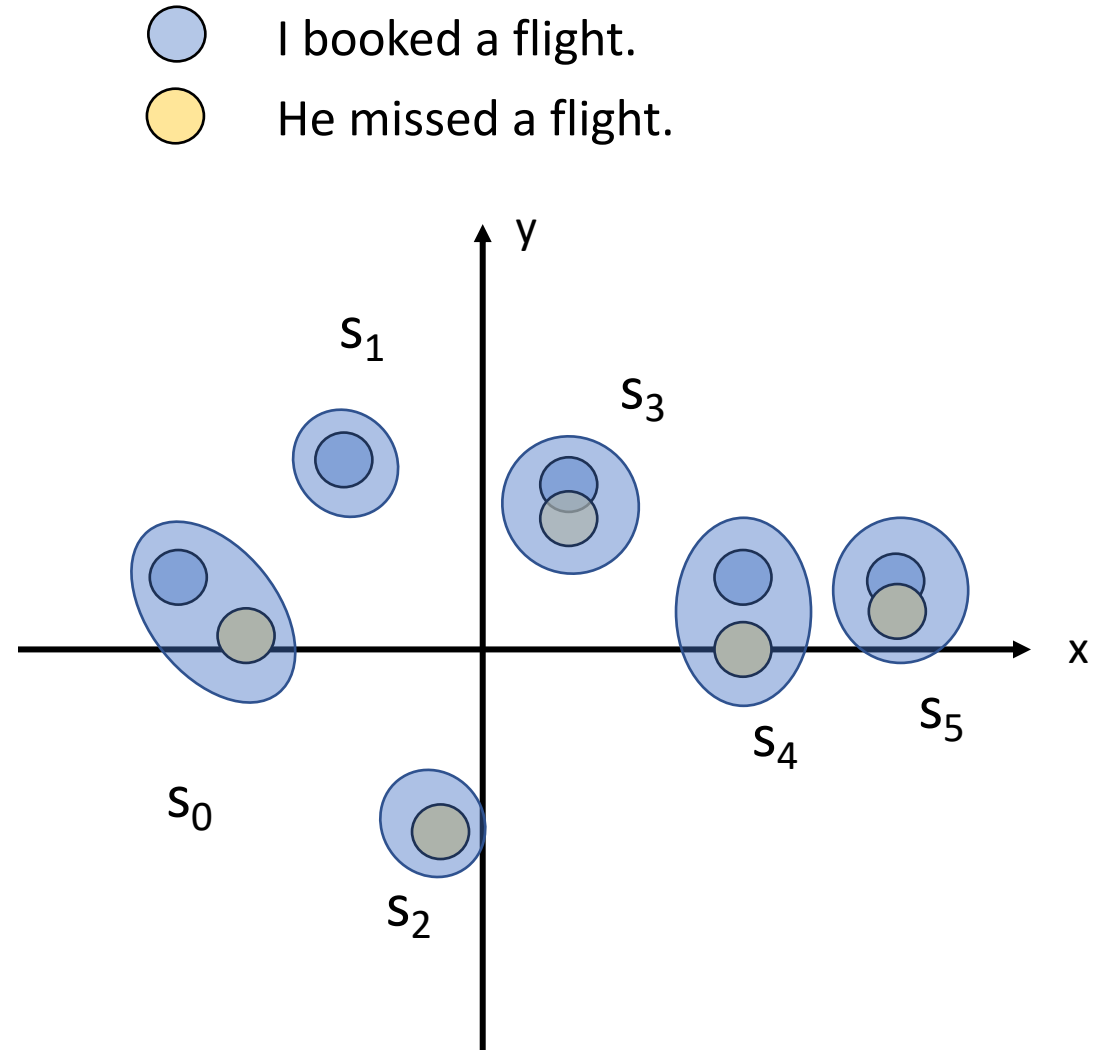
- State profiling
- Dimension reduction
 - Principal Component Analysis

-  I booked a flight.
-  He missed a flight.



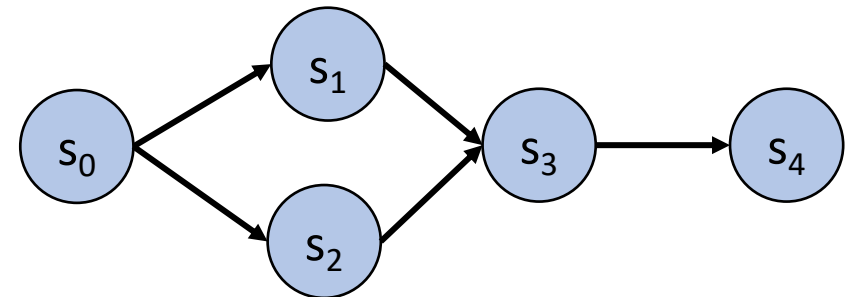
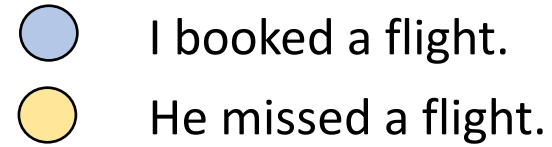
Approach Overview

- State profiling
- Dimension reduction
 - Principal Component Analysis
- State abstraction
 - Gaussian Mixture Models



Approach Overview

- State profiling
- Dimension reduction
 - Principal Component Analysis
- State abstraction
 - Gaussian Mixture Models

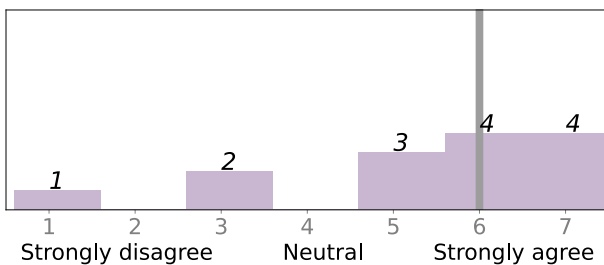


Evaluation

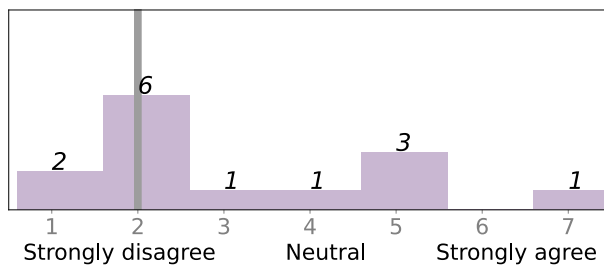
A User Study with 28 Participants

— Median ■ Num of participants

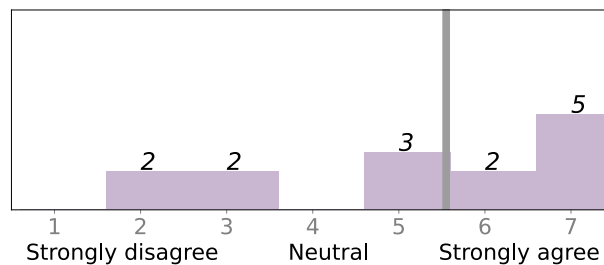
Q1. State diagram is useful.



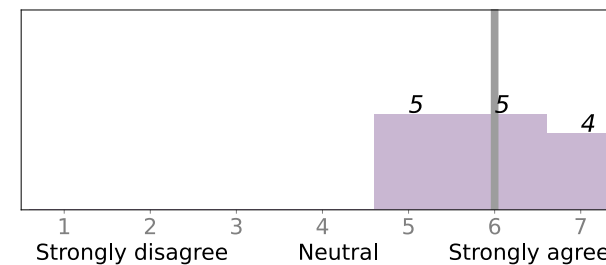
Q2. State diagram is overwhelming or distracting.



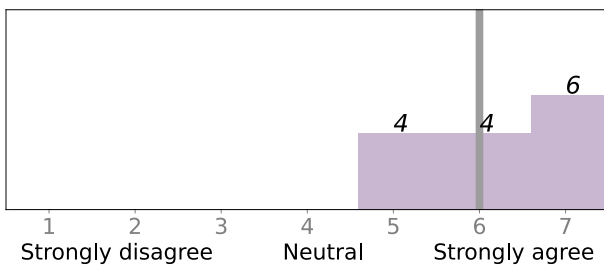
Q3. Interaction with state is useful.



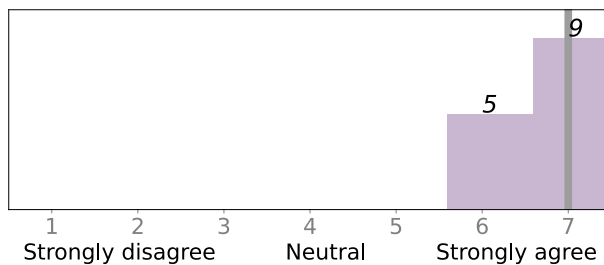
Q4. Pattern summary and filtering is useful.



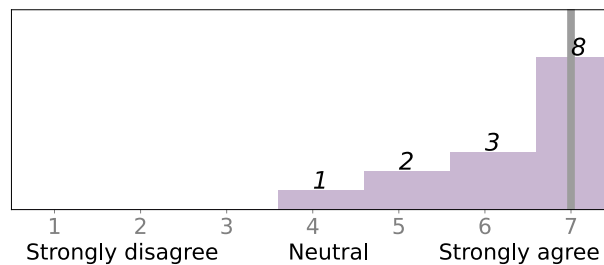
Q5. Instance view is useful.



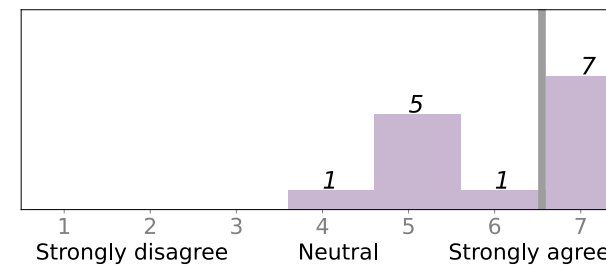
Q6. Intermediate prediction results are useful.



Q7. Search for similar data instances is useful.



Q8. I want to use DeepSeer to debug my own models.



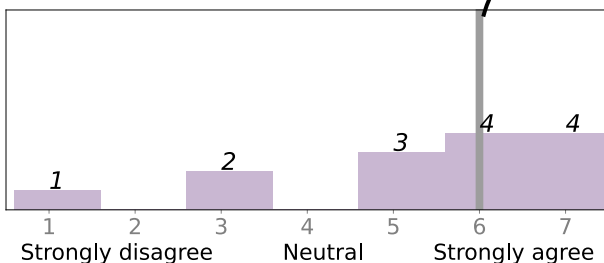
Evaluation

A User Study with 28 Participants

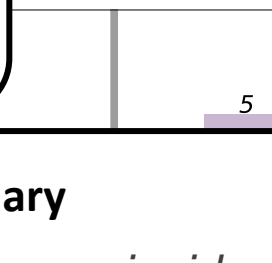
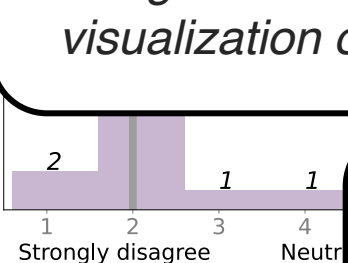
State Diagram

P10 said, "I think DeepSeer is in good design. The best feature I like the visualization of positive/negative states."

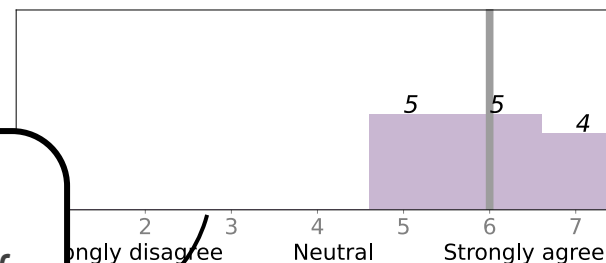
Q1. State diagram is useful.



state is useful.



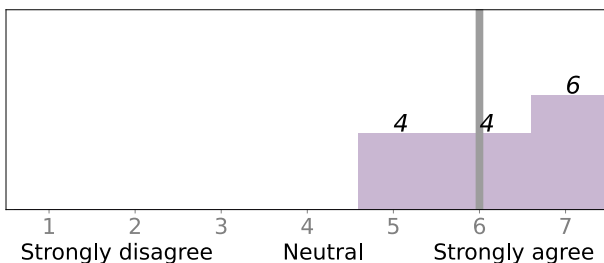
Q4. Pattern summary and filtering is useful.



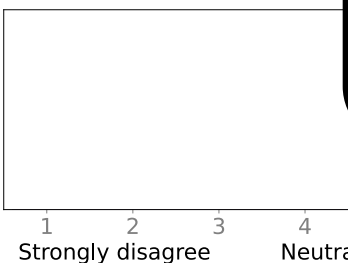
Pattern Summary

P2 said, "it is good for us to see inside of the model and find the bug with possibly buggy pattern."

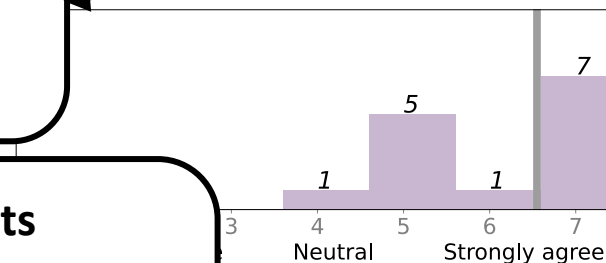
Q5. Instance view is useful.



Q6. Intermediate prediction results are useful.



I want to use DeepSeer to debug my own models.



Intermediate Prediction Results

P5 said, "typing out my own example, and seeing the sequence of internal nodes that the RNN went through (and its prediction) was quite useful."

Design Implications

- Opening a black-box model is critical for debugging.
- Users wish to see the internal decision-making process of a model.
- Explanations should be interpretable to users.