



Matteo Paltenghi Contact: mattepalte@live.it https://matteopaltenghi.com/

Understanding Neural Models of Code

Ph.D. advisor and co-author: Michael Pradel

Problem

- current neural models are evaluated end-to-end. It is
- still unclear what they actually learn, and if their reasoning matches that of skilled humans.
- risk of deploying models that are right for the wrong reason
 Background

Neural models of code [1]: any neural network model that aims to solve or ease a software engineering task.

Research Questions and Results [4]

RQ1: Are the attention weights correlated with human attention while performing a code-related task?



Regular Attention Layer [2]:

- specialized nerual network layer that focuses on parts of the input according to their relevance for prediction
- it can be visualized as a heatmap over the input

Copy attention [3]: specialized attention layer that selects which input should be copied verbatim to the output.

Task on Source Code

• Method naming task: predict the name of a method given its content.



Spearman rank correlation (mean)

<pre>synchronousDestination.afterPropertiesSet();</pre>	<pre>synchronousDestination.afterPropertiesSet();</pre>
<pre>synchronousDestination.open();</pre>	<pre>synchronousDestination.open();</pre>
doTestSend(synchronousDestination);	<pre>doTestSend(synchronousDestination);</pre>
Copy attention of neural model	Human attention (5x per method)

Human attention (5x per method)

Our Approach: Human Data Collection [4]

- we introduce the Human Reasoning Recorder (HRR): web interface with blurred code, move to unblur the code tokens.
- HRR more scalable than eye tracking
- We compare it to the attention produced by popular neural models.



	All methods	Methods with $F1 \ge 0.5$
CNN, regular attention	0.08	0.24
CNN, copy attention	0.49	0.55
Transformer, regular attention	-0.20	0.02
Transformer, copy attention	0.47	0.55

Contributions [4]

- Novel and scalable methodology for recording human attention while performing a task on source code
- Dataset with more than 1500 attention human heatmaps.
- In-depth comparison of attention of human vs neural models.

Insights

 Neural models and humans often agree about what tokens to copy verbatim from the input to the output, but less on what other tokens to pay regular attention.



Code and Data



https://github.com/MattePalte/thinking-like-a-developer

Live HRR Demo!



https://demo-thinking-like-a-developer.herokuapp.com/

 A higher human-model correlation coincides with more effective predictions by the neural models

References

2021

- [1] Michael Pradel and Satish Chandra. 2021. "Neural Software Analysis". Commun. ACM 2021
 [2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. "Neural Machine Translation by Jointly Learning to Align and Translate", ICLR 2015
 [3] Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. "Pointer Networks.", NeurIPS 2015
- [4] Matteo Paltenghi and Michael Pradel. 2021. "Thinking Like a Developer? Comparing the Attention of Humans with Neural Models of Code". ASE

