- Personalised web browsing experience is hard
- Especially with a rigorous and respectful privacy policy
- Ultimately, many of the approaches in UMAP strive to find innovative ways to extract a meaningful signal from very noisy data.

Martin Lopatka
mlopatka@mozilla.com
Data Scientist/Applied Statistician

moz://a

Martin Lopatka
mlopatka@mozilla.com
Data Scientist/Applied Statistician

- Martin Lopatka
- Time we have through... Mozilla's approach to recommending browser extensions
- T.A.A.R.
- Curiosity vs. builders -> move **spatial anchor**
- given the time, focus on a very brief overview, and two specific design choices
- Privacy by design and CLLR

Telemetry

Firefox Telemetry (optionally) measures and collects non-personal, performance and usage information.[1]

[1] https://wiki.mozilla.org/Telemetry

moz://a

Telemetry

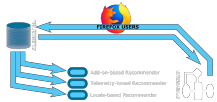
Firefox Telemetry (optionally) measures and collects non-personal, performance and usage information.

- application localization identifier: (ch-de, br-pt)
- operating system
- subsession length
- bookmark count
- open tab count
- unique TLDs
- add-ons installed

Telemetry-Aware Add-on Recommender

moz://a

└**T**elemetry-**A**ware **A**dd-on



- Full system Spec
- Three modules each leveraging different subsets of client information based on availability.
- Individual recommendations combined via linear stacked ensemble
- These are domain specific and specific to our telemetry infra, so lets treat them like black boxes
- more interesting is the comparison of functions for determining individual weighting of the recommendations.

Differential Privacy

Differential Privacy

- Due diligence
- method for releasing a table of frequency counts while preserving a differential privacy guarentee.
- We adapt this technique to generate add-on installation frequency tables for each locale according to the following procedure
- Guards against Overfitting

Log Likelihood Ratio Cost (cLLR)

```
'zh-CN': [('guid_01', 0.75),...,('guid_02', 0.05)],
'fr-FR': [('guid_03', 0.24),...,('guid_04', 0.01)],
...,
'en-US': [('guid_04', 0.18),...,(guid_05', 0.02)]
```

include boxplots vs other metrics.

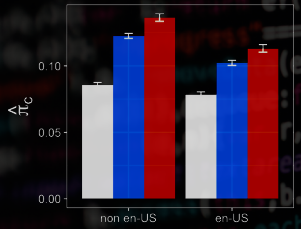
Log Likelihood Ratio Cost (cLLR)

```
'zh-CN': [('guid_01', 0.75),...,('guid_02', 0.05)],
'fr-FR': [('guid_03', 0.24),...,('guid_04', 0.01)],
...,
'en-US': [('guid_04', 0.18),...,(guid_05', 0.02)]
```

include boxplots vs other metrics.

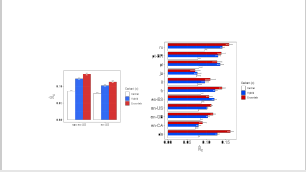
- better usage of full signal if component modules a probabilistic (flavoured)
- better ensembles
- generalizable
- Versus Discounted Cumulative Gain (DCG), and versus Mean Average Precision (MAP) for including in the recommendation list at all

Performance



moz://a

Performance



- variable availability of our data, not only in terms of quantity but in terms of fields
- offer users granular choice in Telemetry
- Performs well and scales with data availability
- And 100% open source
- so if you are just curious or building something -> step back **spatial anchor**

Acknowledgements

moz://a

Acknowledgements

- Thank you all for choosing to come engage with me here
- I'll be happy... questions
- but first... acknowledgements