Human-Centered Computing
RESEARCH GROUP

Freie Universität Berlin

WIKIMEDIA FOUNDATION

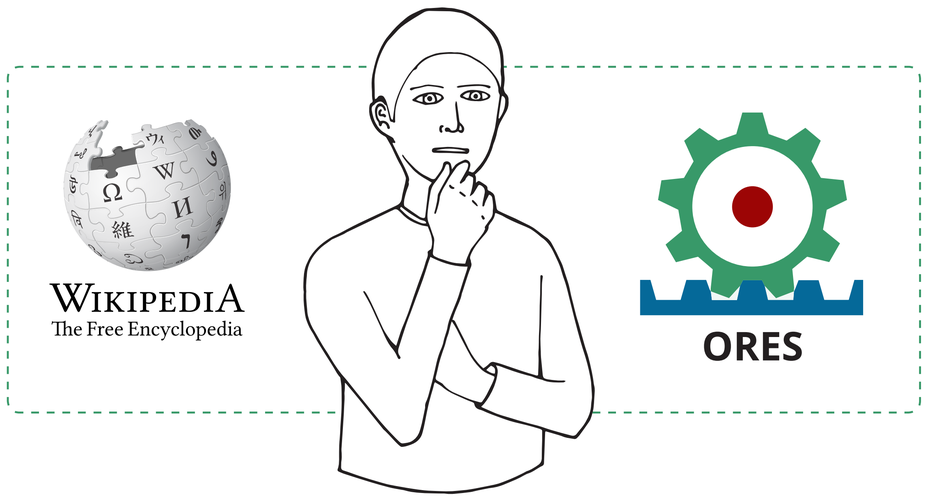
PreCall: A Visual Interface for Threshold Optimization in ML Model Selection

Christoph Kinkeldey[1], Claudia Müller-Birn[1], Tom Gülenman[1], Jesse Josua Benjamin[1], Aaron Halfaker[2]

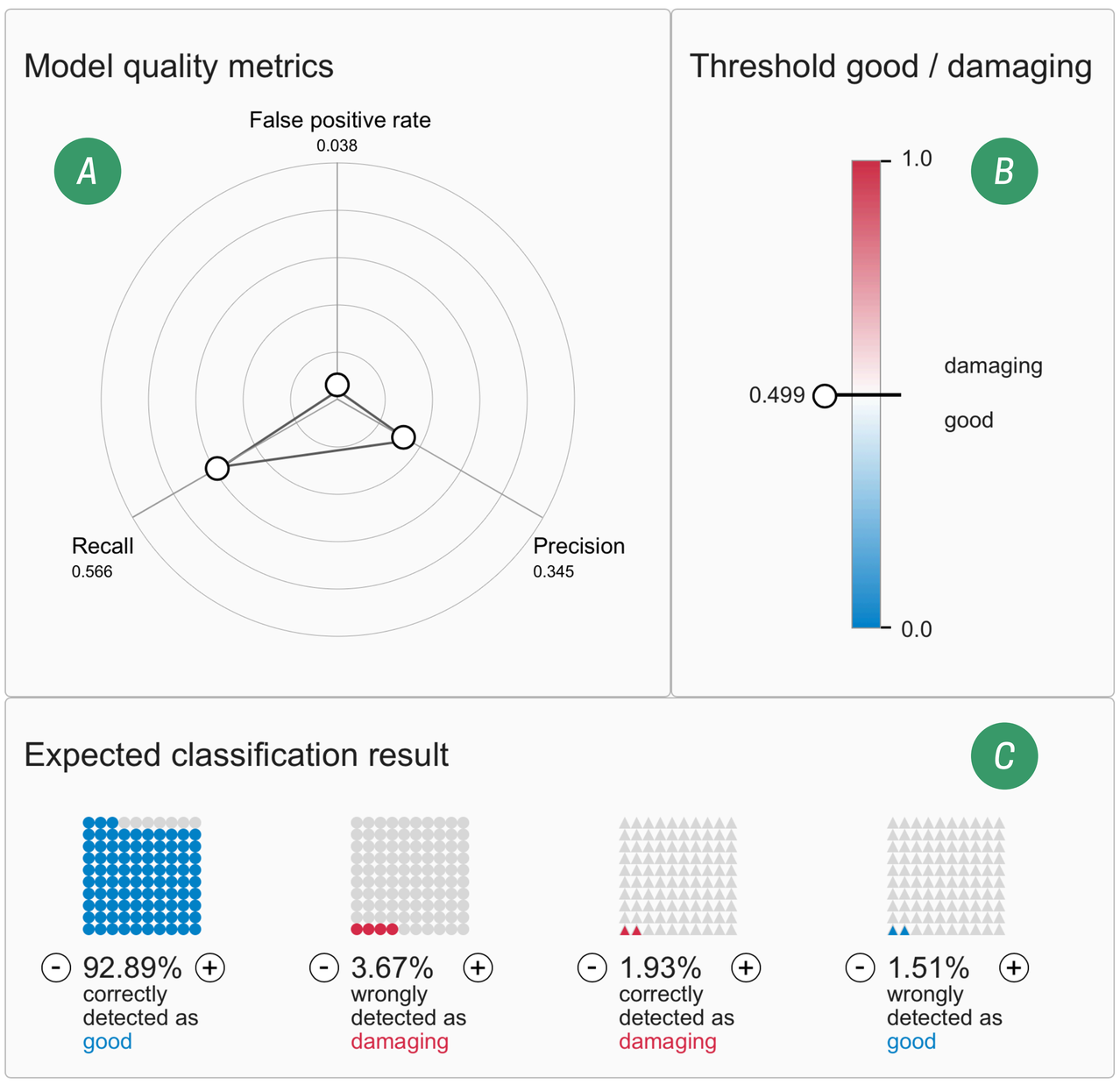
[1] Human-Centered Computing, Freie Universität Berlin
[2] Wikimedia Foundation

Motivation

- ORES is a web service that provides machine learning as a service for Wikimedia projects, like Wikipedia and Wikidata (e.g., for vandalism detection).
- But: it is not straightforward for developers to come up with configurations that meet their operational requirements.
- Selection of a model configuration is a trade-off between ***precision*** (i.e., "all detected edits are damaging") and ***recall*** (i.e., "find all damaging edits").



„Recall? Precision?"



The PreCall Interface.

github.com/tguelenman/PreCall

GET THE CODE!

A. Model quality metrics Keep track of the interdependencies between recall, precision and false positive rate. Handle dragging interaction is supported.

B. Threshold good/damaging The model's current threshold for classifying edits as damaging or not.

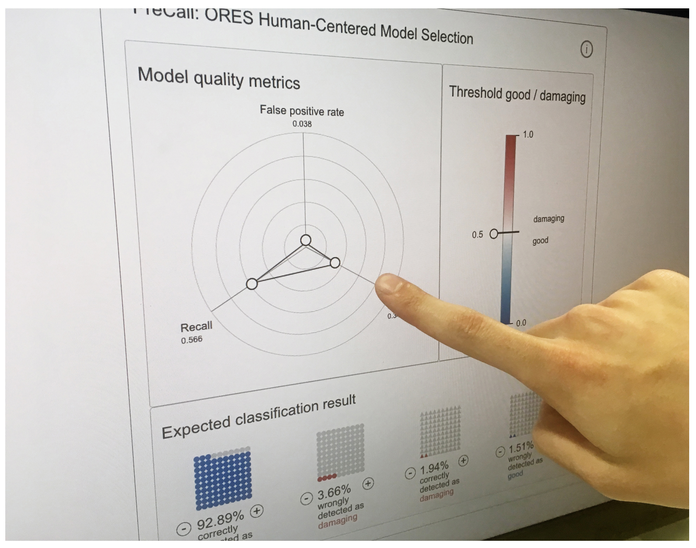
C. Expected classification result Classifier's performance specifying the confusion matrix outputs. Increase and decrease directly by clicking and holding the buttons.

Pilot Study

- Think aloud protocol with three participants and two tasks
- Participants gained new insights (e.g. non-linear relation between parameters)
- Most participants directly manipulated the '**Expected classification result**' view (C)

Outlook

- 'Result-driven' vs. 'Metric-driven' model selection
- Field studies in the Wikipedia Community: ***Does PreCall increase the acceptance of the ORES service?***



Pilot Study