Human-AI Interaction Design

Grace Hu • 10.08.2020

Standard ML vs Interactive Machine Learning

Standard

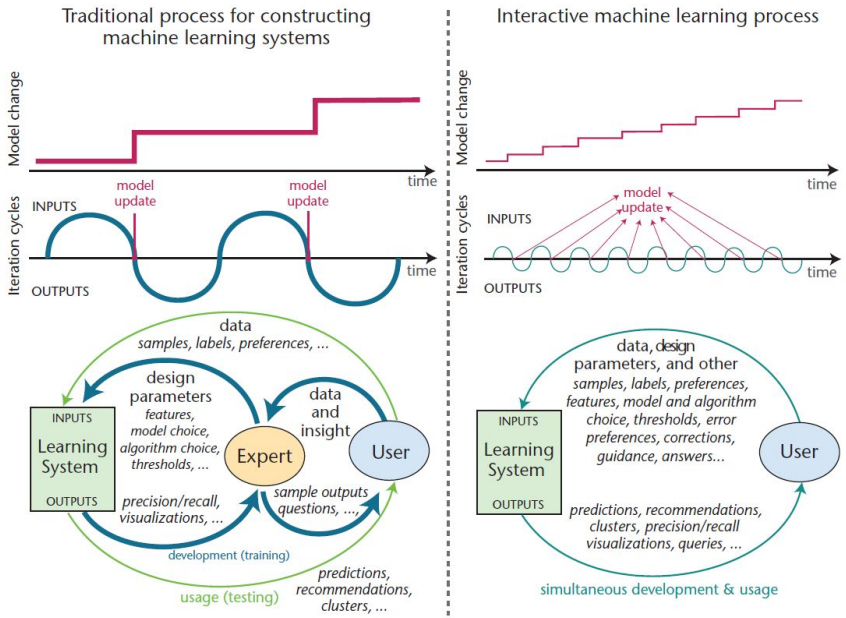
- End users (domain experts or everyday users) only provide data, answer domain-specific questions, and give feedback on learned model
- Design process is lengthy, end users don't affect model that much
- ML practitioners drive model exploration

Interactive

- Learning cycles involve more rapid, incremental model updates
- End users can interact with & influence model
- Democratization of machine learning

More on Interactive Machine Learning

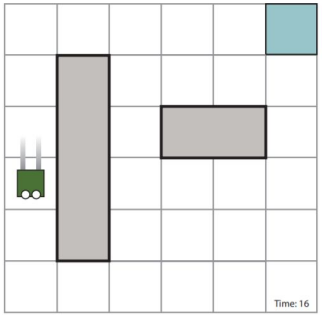
- Model updates are:
  - **Rapid**
  - **Focused**
  - **Incremental**
- End users can immediately see impact of their actions
- Users with little ML expertise can steer model behaviour through trial & error



https://www.aaai.org/ojs/index.php/aimagazine/article/view/2513

Q: Can anyone think of an example where interactive machine learning might be useful?

Downsides to Interactive Learning



Active Learning

- Learner queries "oracle" to request label for unlabelled example
- People get annoyed with having to repeatedly tell the computer if it's right or wrong
    - People can give other types of feedback

Reinforcement Learning

- Agent is given rewards after each action & learns behaviour
- People give more positive than negative rewards (bias)
    - Bypass this by improving algorithms that seek long-term rewards

Summary

- Users can provide a lot more feedback than just the accuracy of labels
- Users want to know more about how the model works
    - Ex. Users who were told the value of their movie rating contribution to the community provided more ratings
- Users like to know the context of what their labelling is used for (active learning)
- Users make mistakes while labelling

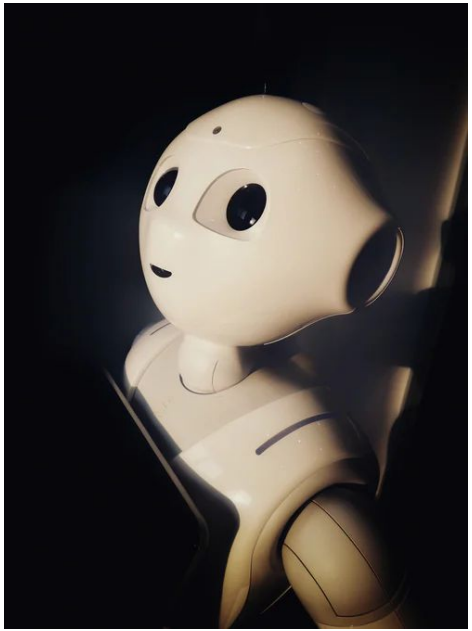
Human Perception of ML Quality Metrics

Q: Why might the "democratization of machine learning" be important?

Why is it important that laymen and everyday users understand what's happening in the model?

Human Perception of ML Models

Do people trust my model?



How good is good enough for users?

Do people trust my model?

1. Does a model's stated accuracy affect people's trust in the model?
2. Does this trust change after people observed the model in action?
3. Does a model's observed accuracy affect people's trust in it?

Study: Amazon Mechanical Turk participants & model to predict outcome of speed dating events.

Yin, Vaughan, Wallach 2019

Findings: Users are affected more by their observations of a ML model in action than by the accuracy metric given to them => People rely on their own interactions with a model when deciding how much to trust it.

What this means for Interactive ML:

- ML practitioners must properly communicate the uncertainty of each prediction.
- Let users make predictions themselves before using model so they can assess how much they should trust the model.

http://www.mjskay.com/papers/chi_2015_accuracy-acceptability-mjskay.pdf

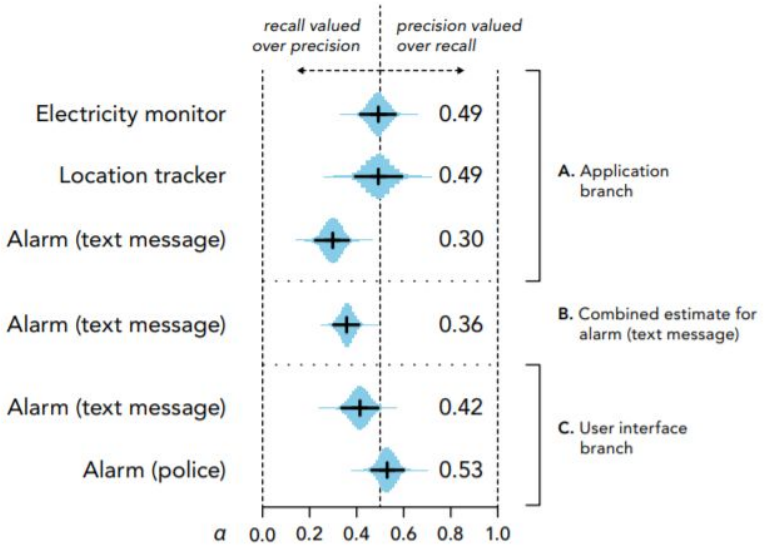
How good is good enough?

- Is 90% accuracy good enough for a image classifier? An autonomous vehicle?
- Do users value recall or precision more? Do users even understand what the metrics mean?

https://www.facebook.com/convolutionalmemes/posts/745586122562557

Is Your Classifier Good Enough?

- Intelligibility of user interfaces: Does the reasoning of the system make sense to users? Will users find it useful?
- Ex. Unweighted F1 score is commonly used to evaluate models, but users may prefer one of recall/precision over the other
- Solution: "Acceptability of Accuracy" metric



https://dl.acm.org/doi/pdf/10.1145/3290605.3300509

18 Guidelines for Human-AI Interaction 2019

https://www.microsoft.com/en-us/research/blog/guidelines-for-human-ai-interaction-design/

Initially (before interaction)

1. **Make clear what the system can do.**
2. **Make clear how well the system can do what it can do.**

During Interaction

4. **Show contextually relevant information.**

When Wrong

9. **Support efficient correction.**

Over Time

13. **Learn from user behavior.**