PURDUE UNIVERSITY | Department of Computer Science

CS57100: Artificial Intelligence
Takeaways from ICDM22

Prof. Chris Clifton

2 December, 2022

PURDUE UNIVERSITY
Department of Computer Science

Announcements

- Final projects due this weekend
  - Writeup
  - Presentation materials
- Presentations next week
  - See course web page for (randomly assigned) schedule
  - Please let me know right away if you have a conflict
- Final exam 15 December, 1-3pm, BRNG 2290

PURDUE UNIVERSITY
Department of Computer Science

Takeaways from ICDM'22

- ICDM'22 held November 28-December 1 in Orlando
  - World's premier research conference in data mining
  - Annual conference running since 2001
  - 870 submissions, ~180 accepted

3

PURDUE UNIVERSITY
Department of Computer Science

Observations

- Problems with graph-structured data are hot
  - "Specialized" graphs (e.g., recommender systems generalized as bipartite graphs)
- Active and Reinforcement Learning seem to be growing
  - Particularly novel models that tie well to real-world problems

4

PURDUE UNIVERSITY
Department of Computer Science

Vipin Kumar (Minnesota) Knowledge-Guided ML

- Long-standing leader in use of data mining for environmental challenges
- Competing approaches to climate modeling
  - Process-based models of dynamical systems
  - Machine learning
- *Combining the two results in significantly better models*
- Example: predicting lake temperature
  - Data collection too sparse for full physics-based models (e.g., energy input/output)
  - ML-based approaches don't generalize well to unseen data
- Hybrid: Physics-based energy loss
  - even for days with no observation
  - See 43:55 in talk

5

PURDUE UNIVERSITY
Department of Computer Science

Cynthia Rudin (Duke) Simple Models

- ML moving towards complicated, uninterpretable models
- Simple models often do (nearly) as well
  - Interpretable
  - More likely to generalize
- Example at 13:00 in talk
- *Have we forgotten a long-established principle?*

6

PURDUE UNIVERSITY
Department of Computer Science

Occam's Razor

- Pluralitas non est ponenda sine necessitate
  - *Plurality should not be posited without necessity*
- Of competing theories, the simpler explanation is to be preferred
- *William of Ockham, 1285-1347*

PURDUE UNIVERSITY
Department of Computer Science

Another Example: "Feature Creep"

- Too few radiologists
  - Can we use AI/Image Analysis?
- Problem: Does patient have pneumonia?
  - Many features – image, also metadata (patient characteristics, machine used, was image taken while patient standing or lying down, etc.)
- "Learned" classification
  - *If patient lying down when x-ray taken, they have pneumonia*

PURDUE UNIVERSITY
Department of Computer Science

Erik Xing, CMU
"Standard Model" for learning

- $\min_{q,\theta} -\alpha H(q) + \beta D\big(q(x), p_{theta}(x)\big) - E_{q(x,y)}[f(x,y)]$
  - E – experiences term (data examples, rules, etc.)
  - D – Divergence (how good the model fits)
  - H – Uncertainty (e.g., Entropy)
- Application to classification obvious
  - But can also apply to learning other things – depending on how the model and fit of model to data is defined

9

PURDUE UNIVERSITY
Department of Computer Science

Panel
Full Stack AI: The Missing Piece

- Raj Acharya, IU; Chris Clifton, Purdue; Lawrence Hall, USF; Cynthia Rudin, Duke
- Key takeaways:
  - HCI and AI – how does AI interact with humans?
    - Particularly explainability
  - Robustness
    - Can we ensure no catastrophic mistakes?
    - Can we have reliable confidence bounds on outcomes?
  - Applications
    - How do we apply AI to important problems?

10

PURDUE UNIVERSITY
Department of Computer Science

What about the future?
(My feelings based on what I heard)

- Leaderboard/benchmark-driven work isn't going to solve real problems
  - Data incorporates hidden assumptions that don't address the real-world problems they are supposed to model
    - Hence, best on the leaderboard doesn't translate to real-world success
  - Optimization criteria too narrow, don't address breadth of real-world constraints
- Generalized frameworks important
  - But must apply generally
  - Incorporate knowledge/reasoning, not just ML

11