AI Competencies for Undergraduate Data Science Curricula

Andrea Danyluk

Williams College & Northeastern University

Scott Buck

Intel Corporation

EAAI

January 2019

Why we're here

- Not to make a direct contribution to AI education
- To ask for your help in articulating what every undergraduate major in Data Science should know about AI

ACM Data Science Task Force Charter

To add to the broad, interdisciplinary conversation on data science, with an articulation of the role of computing discipline-specific contributions to this emerging field. The task force should seek to define what the computing contributions are to this new field, and should provide guidance for undergraduate data science programs of study.

To create a report, which may then be used to invite collaboration and coordination with other (non-computing) professional societies.

Other Data Science Efforts

- EDISON Project (2017)
  - A competency-based framework to be used as guidance for educators, employers, etc.
  - Most similar to ACM effort; Europe focus.
- Park City Report (2017)
  - Topics and learning outcomes for undergraduate data science curricula
  - Sample course outline
  - Statistics leaning?
- National Academies report (2018)
  - Higher level articulation of the importance of data science education

Organized by Knowledge Area with articulated competencies

- CS Fundamentals
  - Programming, Data Structures, Algorithms, Software Engineering
- Data Acquisition & Governance
- Data Management, Storage, Retrieval
- Data Privacy, Security, Integrity
- Machine Learning
- Data Mining
- Big Data
  - Complexity, Distributed Systems, Parallel Computing, HPC
- Analysis & Presentation
  - HCI, Visualization
- Professionalism

Competencies:
too much? too little? missing?

- Professionalism:
  - Analyze the impact on design of the requirement to provide insights into decisions made autonomously by machines
- Machine Learning:
  - Derive a (current) learning algorithm from first principles and/or justify a (current) learning algorithm from a mathematical or statistical perspective.
- Analysis & Presentation:
  - Identify mechanisms for providing explanations of decisions
- Data Mining:
  - Tokenizing, tagging, etc.?
- CS "Fundamentals"
  - Probabilistic programming?
  - Combinatorial explosion of search spaces and its consequences?

Next Steps

- See the first draft of the report
  - http://www.cs.williams.edu/~andrea/ DSReportInitialFull.pdf
- Visit the poster
  - Tuesday, 6:30-8:30 PM
  - Coral Lounge, Hilton Hawaiian Village
  - EAAI16