Artificial Intelligence Competencies for Data Science Undergraduate Curricula

Andrea Danyluk (Williams College & Northeastern University), Scott Buck (Intel Corporation)
for the ACM Data Science Task Force

Introduction

The ACM Data Science Task Force was created by the ACM Education Council and tasked with articulating the role of computing discipline-specific contributions to this emerging field.

Many areas of AI are directly relevant to Data Science, including machine learning, computer vision, intelligent interfaces, and speech and natural language processing.

We seek the AI Education community's input on AI competencies that should be considered central for students in an undergraduate Data Science program.

Curriculum Report Overview

The Task Force is preparing a report that is both a computing-focused curricular volume in the tradition of other ACM curricula and a white paper calling for a multidisciplinary task force to build on this work. An initial draft is available for review and comment.

Introduction & Background

Includes the task force mandate, committee work and processes, results of a survey of academic and industry professionals, Knowledge Areas (KAs) for Data Science undergraduate curricula, and an introduction to the competency framework.

Competency Framework

Meaning of competency, levels of learning outcomes, and relationship to professional practice.

Knowledge Areas and Competencies

For each computing-focused KA, an articulation of competencies for an undergraduate student completing a Data Science degree.

AI Competencies for Undergraduate Data Science Curricula

Competencies to be achieved by students in an undergraduate Data Science program are organized by Knowledge Area. Though there is no KA titled *Artificial Intelligence*, AI-based competencies appear throughout. Below is a sample of AI competencies in the draft curricular recommendations.

Machine Learning

Scope	Competenies
Categories of machine learning approaches (e.g., supervised, unsupervised)	Compare and contrast classes of learning approaches with a focus on inputs, outputs, and problem types to which they can be applied.
Algorithms and tools in each category	Select and apply a broad range of machine learning tools and implementations to real data.
Machine learning as a set of principled algorithms, rather than a "bag of tricks"	Derive a (current) learning algorithm from first principles and/or justify an algorithm from a mathematical or statistical perspective.
Notion of hypothesis space; relationship to expressive power of learned models	Express formally the representational power of models learned by an algorithm, and relate that to issues such as expressiveness and overfitting.
Problems related to model expressivity and availability of data; techniques for mitigating their effects	Exhibit knowledge of methods to mitigate the effects of overfitting and curse of dimensionality in the context of machine learning algorithms.
Performance objectives	Provide an appropriate performance metric to evaluate a machine learning algorithm/tool for a given problem.
Evaluation methodology	Apply appropriate empirical evaluation methodology to assess performance on a given problem or to compare algorithms/tools to each other.
Interpretability of learned models	Compare differences in interpretability of learned models.
Algorithmic and data bias, data integrity, professional responsibility	Be aware of problems related to algorithmic and data bias, as well as privacy and integrity of data.
Implementation	Implement machine learning programs from their algorithmic specifications.

How you can help

We appreciate your feedback on the competencies identified in the draft report and welcome suggestions for other AI competencies that should be considered central for Data Science.

Knowledge Areas

The Task Force has identified the following computing-focused KAs for Data Science:

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

References

ACM (2018) Curricula Recommendations. https:// www.acm.org/education/curricula-recommendations

Cassel, B. and Topi, H. (2015) Strengthening Data Science Education Through Collaboration. Technical Report.

DeVeaux, R. et al. (2017) Curriculum Guidelines for Undergraduate Programs in Data Science. *Annual Review of Statistics and Its Application* , 4:15-30.

EDISON (2018) The EDISON Data Science Competence Framework. http://edison-project.edu/edison/edison-data-science-framework-edsf

National Academies of Sciences, Engineering, and Medicine (2018) *Data Science for Undergraduates: Opportunities and Options*, Washington, DC: The National Academies Press.

Where to find the draft report