ACM Data Science Task Force

Second Draft Report

and Opportunity for Feedback

Paul Leidig
Grand Valley State University
leidig@gvsu.edu

Lillian (Boots) Cassel
Villanova University
lillian.cassel@villanova.edu

Outline

- The Data Science Curriculum Task Force Effort
  - Committee
  - Background
- The Second Draft Curriculum Report
  - Contents
  - Knowledge Areas and Competencies
- Community Engagement
  - Timeline
  - Discussion

ACM Data Science Task Force

- Andrea Danyluk, Williams C. and Northeastern U., USA (co-chair)
- Paul Leidig, Grand Valley State University, USA (co-chair)
- Lillian (Boots) Cassel, Villanova University, USA
- Andrew McGettrick, University of Strathclyde, UK
- Scott Buck, Intel, USA
- Christian Servin, El Paso Community College, USA
- Hongzhi Wang, Harbin Institute of Technology, China
- Weining Qian, East China Normal University, China
- Maureen Doyle, Northern Kentucky, USA
- Karl Schmitt, Valparaiso University, USA
- Suzanne McIntosh, New York University, USA
- Tin Kam Ho, IBM, USA
- Jian Pei, University of South Florida, USA

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.

Background

- ACM Ed. Council summer meeting 2017
  - Build on the efforts of Boots Cassel & Heikki Topi, as well as other groups
  - Articulate importance of computing in the interdisciplinary data science space
  - Identify computing-based competencies for an undergraduate data science curriculum

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

Computing Competencies for Undergraduate Data Science Curricula

Second Draft Report

http://dstf.acm.org/DSReportDraft2Full.pdf

Second Draft Report Contents

The Body of Knowledge: Computing Competencies for Data Science

- Analysis and Presentation (AP)
- Artificial Intelligence (AI)
- Big Data Systems (BDS)
- Computing and Computer Fundamentals (CCF)
- Data Acquisition, Management, and Governance (DG)
- Data Mining (DM)
- Data Privacy, Security, Integrity, and Analysis for Security (DPSIA)
- Machine Learning (ML)
- Professionalism (PR)
- Programming, Data Structures, and Algorithms (PDA)
- Software Development and Maintenance (SDM)

Competency Framework

- Following ACM/IEEE-CS IT 2017; moving in the direction of CC 2020.
- Utilize a working definition of competency that connects knowledge, skills, and dispositions.
- Includes, but moves beyond articulation of topics and learning outcomes. [e.g., CS2013]

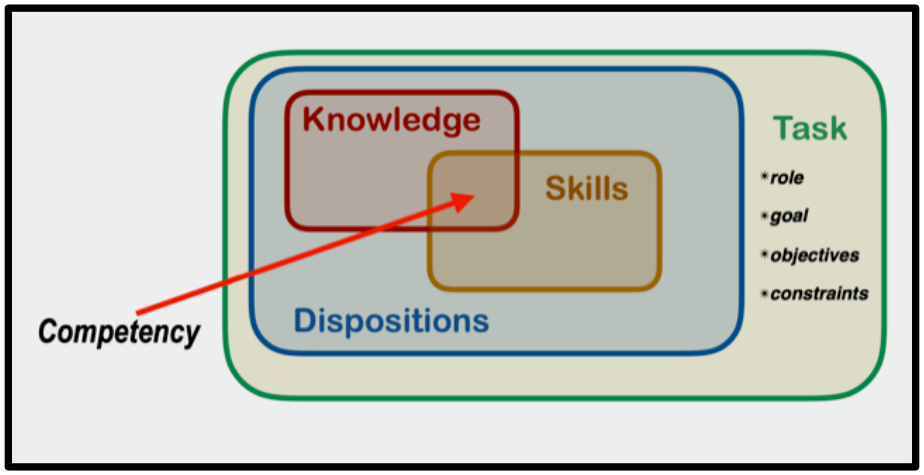
Competency = Knowledge + Skills + Dispositions

- Knowledge
  - Mastery of content
  - Transfer of learning
- Skills
  - Capabilities and strategies for higher-order thinking
  - Interactions with others and world around
- Dispositions
  - Personal qualities (socio-emotional skills, behaviors, attitudes) associated with success in college and career

From IT 2017; adapted from a publication by Council of Chief State School Officers (2013).

Competencies

CC2020



Competencies = Knowledge (K) + Skills (S) + Dispositions (D)

- **Knowledge** - a fact/idea that enables satisfactory performance of relevant tasks
- **Skill** - a degree of mastery in applying a fact/idea to achieve a valued outcome
- **Dispositions** - values and motivation that moderates skilled behavior to influence a quality of professional performance

Current KA Structure

Data Acquirement and Governance

There can be no analysis of date without the data itself. A data scientist must understand the source and quality of their data, as well as understand appropriate processes for acquiring and maintaining high quality data.

Scope	Competencies
Acquiring data from physical world and extracting data to a form suitable for analysis.	Construct and tune the data acquirement and governance process according to the requirements of an application, including the selection of data sources, data acquirement equipment, and data preparation algorithms.

Additional Examples of Competencies

Computing Fundamentals: Algorithms

Scope	Competencies
Comparison of well-known algorithms' complexity, including machine learning and statistics techniques	Provide the big-Oh time and space complexity for a given procedure.

Additional Examples of Competencies

Computing Fundamentals: Software Engineering

Scope	Competencies
Software engineering principles, including design, implementation and testing of programs.	Implement a small software project that uses a defined coding standard.

Additional Examples of Competencies

Data Privacy, Security, Integrity: Privacy

Scope	Competencies
Technologies to safeguard data privacy.	Evaluate common practices and technologies, and identifying the tools that reduce the risk of data breaches while safeguarding data privacy.

Additional Examples of Competencies

Machine Learning

Scope	Competencies
Problems related to model expressivity as well as availability of data, and techniques for mitigating their effects. E.g., problem of overfitting and regularization techniques for mitigating effects of overfitting; curse of dimensionality and feature selection/weighting/ reformulation techniques for mitigating effects.	Exhibit knowledge of methods to mitigate the effects of overfitting and curse of dimensionality in the context of machine learning algorithms.

Additional Examples of Competencies

Professionalism: Teamwork

Scope	Competencies
Team selection, the need to complement abilities and skills of team members on techniques for mitigating effects.	Document and justify the considerations involved is selecting a team to undertake a specific data science investigation

Timeline

- Early 2019:
  - Draft report out for comment
  - Outreach and gathering of feedback
  - Note: Initial comment period ends March 31
- Spring 2019 (f2f @ SIGCSE)
  - Begin work on next phase, including new KAs and competency details
- Spring 2019+: Outreach, presentations and information gathering
- Summer 2019:
  - Next draft to SIGCSE Education Advisory Committee (formerly Education Council)
  - Call for joint task force
- Fall 2019:
  - Draft report out for comment
- Early 2020
  - Release final report

Call for Example Courses

- http://dstf.acm.org/callForExamples.html

Discussion

- http://http://dstf.acm.org
- Comments on the report?
  - Positive and negative reactions
  - Thoughts on the Knowledge Areas and Competencies?
  - Thoughts on the proposed expansion of the report?
- Additional comments?