Big Data Infrastructure for Educational Data

SLATE Kick-Off Meeting, 2016-06-09

Klaus Johannsen
Center for Big Data Analysis, Uni Research AS

uni Research

Overview

- Center for Big Data Analysis
- SLATE Cluster 2
- Big Data Infrastructure
- Machine Learning

uni Research

Center for Big Data Analysis

- Vision

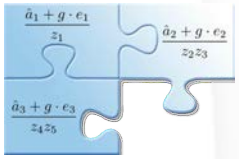
  Create value from data

- Overall objectives
  - *Promote big data in sciences*
  - *Become the leading scientific big data center in Norway*
  - *Realize science based services using big data*

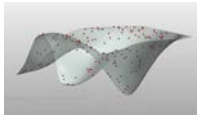
- Key
  - *Master the technical and scientific challenges lying in big data*
  - *Here:*
    (1) Design the big data infrastructure for educational data
    (2) Develop an appropriate machine learning methodology

uni Research

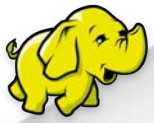
Center for Big Data Analysis – Structure



- Staffing and Competences
  - The staff currently corresponds to 9.3MNOK/a, to increase
  - Expertise in HPC, grid-, cloud computing, e-IS development, middelware development, software engineering, web-programming, data modeling, artificial intelligence, optimization



- IT-infrastructure
  - Hadoop HDP-2.4, Spark, Kafka, HBase, Mahout, Weka, …
  - Linux cluster 200 cores, to increase ...
  - Storage 100TB, to increase ...



uniResearch

SLATE – Cluster 2

- Analyze educational data (big and small) to
  - Understand students: student modeling, performance assessment & prediction, courseware recommendation, grouping of students
  - Assist instructors: analysis of student's progress, automated feedback to improve teaching effectiveness
  - Assist institutions: courseware development, course planning, student cohort comparison, instructor assessment
- Big Data methodologies can help traditional analysis of educational data by
  - Automatically discover complex pattern in large data sets
  - Carry out real-time analysis for immediate feedback

uni Research

SLATE – Cluster 2

- Project 1: Big Data e-Infrastructure for Educational Data
  - Host of raw primary educational data
  - Support for analysis, comparison & manipulation of data
  - Interfaces to external data sources *(MittUiB, Mi Side, MOOC)*
- Project 2: Case study at the University of Bergen
  - Discussion/decision about research agenda
  - Ingestion of data from UiB *(MittUiB, Mi Side)*
  - Learning analytics
- Project 3: National Dugnad: Goals, Guidelines & Policy
  - Broad national discussion about Big Data, Learning Analytics, Educational Data Mining discussing
  - Pedagogical, philosophical, ethical, juridical & cultural issues

uni Research

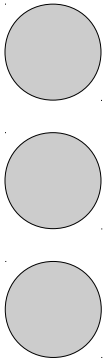
Big Data Infrastructure

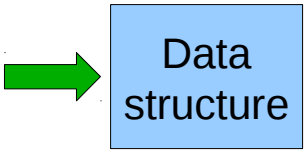
- Big Data compute and storage engine
  - Enables storage, analysis, data science and visualizations
  - Real-time and batch
  - Scalable, no limitation on data sizes or performance
  - Capable of handling a wide range of data types and sources
  - Structured and unstructured data
- Close collaboration with researchers in SLATE and data holders
  - Enables meaningful data structures, analysis and visualizations

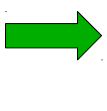
uni Research

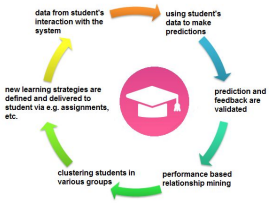
Big Data Infrastructure

Data sources

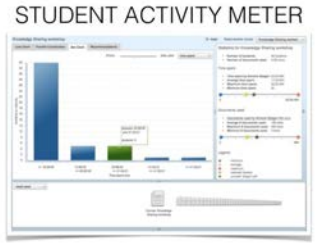








Machine Learning



Dashboards

- Export from Canvas "*MittUiB*" and other sources
- Prepare and import data into Hadoop cluster
- Realize a suitable data structure and interfaces
- Machine learning
- Dashboards, feedback to sources and researchers

uni Research

Machine Learning

- Predicting: a model can infer a single aspect of the data (predicted variable) from some combination of other aspects of the data (predictor variables).

Examples: detecting off-task behavior, failure to answer a question correctly despite having a skill.

- Clustering: finding data points that group together and can be used to split a full dataset into categories.

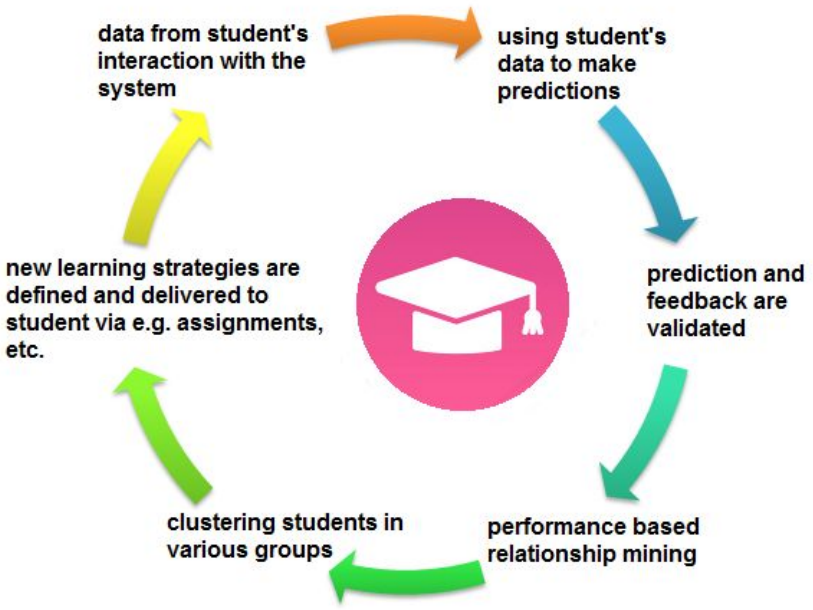
Examples: grouping students based on their learning difficulties and interaction patterns.

- Relationship mining: discovering relationships between variables and encoding them as rules for later use.

Examples: discover associations between student performance and course sequences, discovering most effective pedagogical strategies.

uni Research

Machine Learning



Helps to answer the questions:

- What sequence of topics is most effective for a specific student?
- What student actions are associated with more learning (e.g., higher grades)?
- What student actions indicate satisfaction, engagement, learning progress, etc.?
- What features of an online learning environment lead to better learning?
- What will predict student success?

uniResearch

Thank you

for your attention

uni Research