A Scalable Learning Analytics Platform for Automated Writing Feedback

Nicholas Lewkow
Jacqueline Feild
Neil Zimmerman
Mark Riedesel
Alfred Essa

David Boulanger
Jeremie Seanoksy
Vive Kumar
Kinshuk

Sandhya Kode
Motivation

- MHE data science collaboration with universities
  - Work to develop new models
  - Combine MHE data with university data
  - Run experiments with universities
- Need research environment for collaboration
  - Collect and store heterogeneous student data from LMS or other learning tools via an ingestion API
  - Build and test models on data
  - Store aggregated data from models in a database
  - Expose aggregated data to students/teachers/researchers via API
  - Promote models from research to “production”
- No existing systems fit all of our needs!
Solution: OpenACRE

Open Analytics Collaboration Research Environment -> OpenACRE

- Requirements for OpenACRE
  - Resilient to data loss and fault tolerant
  - Handle the challenges of scalability
  - Easily deployable to Amazon Web Services (AWS) cloud
  - Use all open source technology!
Solution: OpenACRE

System Overview

- Ingestion and storage of data from external systems
  - Input REST API accepts heterogeneous data (JSON)
  - Sends data to distributed queue (Apache Kafka)
  - Collection service pulls data from queue and writes to distributed long-term database (Apache Hadoop HDFS)
Solution: OpenACRE

System Overview

- Develop and test models on student data
  - Models are developed/tested on student data in long-term storage
  - Models are run in parallel (Apache Spark) for scalability
  - Results from models are written to results storage (PostgreSQL)
Solution: OpenACRE

System Overview

- Access the results of the models
  - Output REST API returns model outputs from results storage
  - Output can be used to populate insight visualizations, dashboards, or other learning tools
Solution: OpenACRE

Deploying OpenACRE to Amazon Cloud Services

- Use Terraform by HashiCorp to automatically create cloud instances
Solution: OpenACRE

Deploying OpenACRE to Amazon Cloud Services

- Use Terraform by HashiCorp to automatically create cloud instances

```bash
$ terraform apply
```
Solution: OpenACRE

Deploying OpenACRE to Amazon Cloud Services

- Use Terraform by HashiCorp to automatically create cloud instances

$ terraform apply
Solution: OpenACRE

Open Source OpenACRE

- Release an open source reference implementation to Github in 2016!

Terraform

$ terraform apply

AWS

Collection

Results Store

Output API

Worker 1

Worker 2

Master

kafka

Spark

Hadoop

kafka

Spark

Hadoop

Flask

PostgreSQL

Input API
Writing Competency Feedback

MHE and Athabasca University Collaboration

- Athabasca has a model to generate 17 competencies from essay text
  - Competencies are displayed to students in summary and timeline views

**Traditional Metrics**
- Vocabulary
- Spelling
- Grammatical accuracy
- Lexical diversity

**Advanced Metrics**
- Modifier complexity
- Noun phrase complexity
- Tense agreement

**Sentiment Metrics**
- Negative tone
- Neutral tone
- Positive tone

**Flow Metrics**
- Local cohesion
- Global cohesion
- Connectivity

**Descriptive Metrics**
- Concreteness
- Imagery
- Familiarity
- Conciseness
Writing Competency Feedback

MHE and Athabasca University Collaboration

- Typical student workflow:
  1. Log in to course using LMS
  2. Start a writing assignment in LMS
  3. View feedback visualizations
  4. Edit writing
  5. Repeat 3-4 as needed
  6. Submit assignment

Problem!  Model takes ~12 sec/essay to generate feedback!!

- Example: MOOC with 50k students -> 1 assignment feedback ~ 7 days
- We need a scalable solution to provide large numbers of students with feedback in near real time
Implementing Writing Feedback Model on OpenACRE

- Open source Moodle LMS instrumented to send Caliper events
- Writing feedback model adapted to run in parallel using Apache Spark
- Results API setup to access writing feedback model results

Allows feedback to be calculated much faster!
OpenACRE and Writing Feedback Model Pilot Study

- Efficacy study with 800 students underway at IIIT Hyderabad in India
  - Engineering students taking an English course
  - Propensity score analysis and surveys are being used
  - Study will test how effective and useful the competencies are as well as the visualizations which are used to present them
- More pilot studies planned with college students from different countries
Conclusions

- OpenACRE scalable analytics research environment
  - Easily deploy OpenACRE to AWS cloud
  - Run arbitrary models on data in parallel using Apache Spark
  - Release an open source reference implementation to Github in 2016!

- Collaboration with Athabasca University
  - Writing competency feedback model is the first model implemented on OpenACRE to provide scalable analytics
  - Efficacy study on the writing competency model is underway
Thank you!

Questions?