
Learning Pulse

A machine learning approach for predicting performance in self-regulated learning using multimodal data

Paper presentation at LAK17
15th March 2017, Vancouver, Canada
Outline

1. Background, context, vision
2. Our approach
3. Data collection
4. Data analysis
5. Conclusions
Data deluge in education

Big Data in Education

In five years, the classroom will learn you.

Welten Institute
Research Centre for Learning, Teaching and Technology

Open Universiteit
welten-institute.org
Collecting learning experiences

Learning is happening everywhere. People learn in many places, doing many things.

The Learner

The Tools

The Activities

Collect the experiences that matter.

This API records activities and delivers data that is:
- Quantifiable
- Sharable
- Trackable

Picture from tincanapi.com

Welten Institute
Research Centre for Learning, Teaching and Technology
Learning happening across spaces

- Physical activity
- Digital activity
- Physical interactions
- Digital interactions

Welten Institute
Research Centre for Learning, Teaching and Technology
Context: Self Regulated Learning

Self-Regulated Learning → no guidance → no feedback → no support
Vision: machine learning approach

\[ y = f(X) \]

- Learning Performance (output space)
- Predictive Model
- Multimodal Data (input space)
Our approach
Research questions

(RQ-MAIN) How can we store, model and analyse multimodal data to predict performance in human learning?

(RQ1) Which architecture allows the collection and storage of multimodal data in a scalable and efficient way?

(RQ2) What is the best way to model multimodal data to apply supervise machine learning techniques?

(RQ3) Which machine learning model is able to produce learner specific predictions on multimodal data?
Participants

- 9 PhD students at Welten institute
- Different disciplines
- Different working setups:
  - Time
  - Tasks
  - Operating systems
# Experimental timeline

<table>
<thead>
<tr>
<th>Phase 0</th>
<th>Pre-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>System architecture tested</td>
</tr>
<tr>
<td>Phase 1</td>
<td>Training</td>
</tr>
<tr>
<td></td>
<td>3 weeks of data collection</td>
</tr>
<tr>
<td>Phase 2</td>
<td>Validation</td>
</tr>
<tr>
<td></td>
<td>2 weeks of data collection and prediction</td>
</tr>
</tbody>
</table>
Input space – multimodal data

**Body:** physiological (heart-rate) and physical responses (steps) - from Fitbit HR

**Activities:** applications used during learning from RescueTime

**Context:** weather data from OpenWeatherMap
Output space – Flow

Csikszentmihalyi, 1972

Flow diagram
source: wikipedia.org/wiki/Flow_(psychology)

Theoretical

Empirical

Welten Institute
Research Centre for Learning, Teaching and Technology

Pagina 13
Activity Rating Tool

Participants rate hourly, from 7AM to 7PM

**Challenge**
How challenging was last activity?

**Abilities**
How prepared did you feel for the activity?

**Productivity**
How productive was last activity?

**Stress**
How stressful was last activity?

FLOW

Welten Institute
Research Centre for Learning, Teaching and Technology

“A very easy to use!”

A scalable web app!
Client: Bootstrap + Jquery
Server: GoogleApp + Python
Data collection
Experience API
Data storing format for the Learning Record Store

Ingredients:
1 SUBJECT      the user
1 VERB          the action
1 OBJECT        the item
1 DATETIME      the timestamp

Directions: MIX WELL AND SERVE ACCORDING TO THE CHOSEN SCENARIO

Tips: add 1 CONTEXT to spice things up

ECO LACE
UvA Inform

Welten Institute
Research Centre for Learning, Teaching and Technology
The data journey

Welten Institute
Research Centre for Learning, Teaching and Technology
Complex architecture
Data collection

- **PULL** data from the 3rd party APIs
- Make the xAPI triples
- **PUSH** data in the LRS

- It’s scalable!
- No collisions
- It’s fast
- It’s Interoperable

Learning Pulse Server + Learning Record Store
Data Processing Application

Script in Python running on a VM which processes data in real time
Data Analysis
Transformed dataset

- **Time Series**: tabular representation
- 5 minutes intervals
- Enough samples now!
- Easier view for Machine Learning
- Signal resampling needed

9410 observations x 29 attributes
Issue 1) Feature extraction from Time Series

Heart Rate Variability and Heart Rate Entropy didn’t work

SOLUTION

- Mean of the signal
- Maximum
- Minimum
- Standard Deviation
- Average change

Welten Institute
Research Centre for Learning, Teaching and Technology
Issue 2) Activity data very sparse

Rule based grouping of applications

Applications used are too sparse

SOLUTION
Let’s create application categories

Learners’ activity can be compared!

Welten Institute
Research Centre for Learning, Teaching and Technology
Issue 3) Number of labels available

Trade-off:

- Number of labels
- Seamlessness of the data collections

NO SOLUTION
Issue 5) Random vs continuous data

\[ y_t = \alpha + \beta X_t + e_t \]

Independence constraint

Knowing one value of \( e_t \) for one observation does not help us to guess value of \( e_{t+1} \)

\[ \text{cov}(e_t, e_{t+1}) = 0 \]

SOLUTION follows...
## Mixed Effect Linear Model

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Random Effects</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_0$</td>
<td>$x_1$</td>
<td>$x_2$</td>
</tr>
<tr>
<td>$t_0$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td>$t_1$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td>$t_2$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td>$t_p$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td>$t_p$</td>
<td>$?$</td>
<td>$?$</td>
</tr>
</tbody>
</table>

Used R-squared for goodness-test

**LIMITATIONS**

- Convergence time
- Mono-output

**Welten Institute**
Research Centre for Learning, Teaching and Technology
Issue 6) Inter-subject variability

i.e. Participants have rated very differently

SOLUTION

Predictions are normalised wrt each learner

\[ x_{\text{new}} = (x_{\text{max}} - x_{\text{min}}) \times \frac{x_i}{100} + x_{\text{min}} \]
Conclusions
RQ1) Architecture

The architecture developed was able of:

1. Importing great number of sensor data in xAPI format;
2. combining sensor data with self-reports
3. programmatically transform xAPI data
4. train predictive models & reuse them
5. save the predictions to compare with actual values
RQ2) Represent multimodal data

- *Multiple Instance Representation*
- Each learning sample is a 5 minute interval
- It’s suitable for machine learning
RQ3) Machine learning model

- *Linear Mixed Effect Models* allow
  1. taking into account data specific to each learner
  2. distinguish between fixed and random effects
  3. Take categorical data into account.
Limitations

• **Low accuracy of predictions**
  R-Square tests Stress: 0.32, Challenge: 0.22, Flow score: 0.16, Abilities: 0.08, Productivity: 0.05.

• **Real-time issues**
  Fitbit synchronisation, Virtual Machine performance

• **3rd party API constraints**

• **No great solution for grouping activity data** (manual grouping)
Opportunities

• Data driven
• Real Time feedback
• Visualisations can show feedback
• Seamless data collection
• Multimodal dataset for research
• Reusable architecture

Example visualisation: The Feedback Cube*

*Börner, Tabuenca, Storm, Happe, and Specht. 2015
Thanks for listening!

Daniele Di Mitri
ddm@ou.nl
@dimstudi0