Interactive Visual Analytics for Personalized Online Learning

Meng Xia

07-01-2022
Research Background

Data Visualization

Human Computer Interaction

AI

Education Technology

Prof. Huamin Qu (HKUST)

Prof. Xiaojuan Ma (HKUST)

Prof. Juho Kim (KAIST)

Prof. Vincent Aleven (CMU)
What is Visual Analytics?

Visual analytics provides **visual representations** of datasets and interactive technologies to **augment** human’s ability in finding **insights** in **data**

Input: data

Output: interactive visualizations

Goal: **augmenting** human’s ability in finding **insights** in data
### Table 1: Anscombe’s quartet: four different datasets

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th></th>
<th>II</th>
<th></th>
<th>III</th>
<th></th>
<th>IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
<td>10.0</td>
<td>7.46</td>
<td>8.0</td>
<td>6.58</td>
<td></td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
<td>8.0</td>
<td>6.77</td>
<td>8.0</td>
<td>5.76</td>
<td></td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>8.74</td>
<td>13.0</td>
<td>12.74</td>
<td>8.0</td>
<td>7.71</td>
<td></td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
<td>9.0</td>
<td>7.11</td>
<td>8.0</td>
<td>8.84</td>
<td></td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
<td>11.0</td>
<td>7.81</td>
<td>8.0</td>
<td>8.47</td>
<td></td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.10</td>
<td>14.0</td>
<td>8.84</td>
<td>8.0</td>
<td>7.04</td>
<td></td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.13</td>
<td>6.0</td>
<td>6.08</td>
<td>8.0</td>
<td>5.25</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
<td>4.0</td>
<td>5.39</td>
<td>19.0</td>
<td>12.50</td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>9.13</td>
<td>12.0</td>
<td>8.15</td>
<td>8.0</td>
<td>5.56</td>
<td></td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
<td>7.0</td>
<td>6.42</td>
<td>8.0</td>
<td>7.91</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
<td>5.0</td>
<td>5.73</td>
<td>8.0</td>
<td>6.89</td>
<td></td>
</tr>
</tbody>
</table>
Why Visual Representation?

Table 1: Anscombe’s quartet: four different datasets

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th></th>
<th>II</th>
<th></th>
<th>III</th>
<th></th>
<th>IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
<td>10.0</td>
<td>7.46</td>
<td>8.0</td>
<td>6.58</td>
</tr>
<tr>
<td>y</td>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
<td>8.0</td>
<td>6.77</td>
<td>8.0</td>
<td>5.76</td>
</tr>
<tr>
<td></td>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>8.74</td>
<td>13.0</td>
<td>12.74</td>
<td>8.0</td>
<td>7.71</td>
</tr>
<tr>
<td></td>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
<td>9.0</td>
<td>7.11</td>
<td>8.0</td>
<td>8.84</td>
</tr>
<tr>
<td></td>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
<td>11.0</td>
<td>7.81</td>
<td>8.0</td>
<td>8.47</td>
</tr>
<tr>
<td></td>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.10</td>
<td>14.0</td>
<td>8.84</td>
<td>8.0</td>
<td>7.04</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.13</td>
<td>6.0</td>
<td>6.08</td>
<td>8.0</td>
<td>5.25</td>
</tr>
<tr>
<td></td>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
<td>4.0</td>
<td>5.39</td>
<td>19.0</td>
<td>12.50</td>
</tr>
<tr>
<td></td>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>9.13</td>
<td>12.0</td>
<td>8.15</td>
<td>8.0</td>
<td>5.56</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
<td>7.0</td>
<td>6.42</td>
<td>8.0</td>
<td>7.91</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
<td>5.0</td>
<td>5.73</td>
<td>8.0</td>
<td>6.89</td>
</tr>
</tbody>
</table>

Table 2: Same statistics in Anscombe’s quartet

<table>
<thead>
<tr>
<th>Property (in each set)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of x</td>
<td>9.0</td>
</tr>
<tr>
<td>Variance of x</td>
<td>10.0</td>
</tr>
<tr>
<td>Mean of y</td>
<td>7.50</td>
</tr>
<tr>
<td>Variance of y</td>
<td>3.75</td>
</tr>
<tr>
<td>Correlation between x and y</td>
<td>0.898</td>
</tr>
<tr>
<td>Linear regression line</td>
<td>$y = 0.5x + 3.0$</td>
</tr>
</tbody>
</table>
Why Visual Representation?

- Complement cognition with perception
Don’t need vis when fully automatic solution exists and is trusted. However, when there isn’t, visual analytics can help.
Applications of Visual Analytics

Urban Informatics/Smart Cities, Social Media, Text Analytics, Explainable AI, E-Learning, Social Network, AR/VR
Applications of Visual Analytics

Urban Informatics/Smart Cities, Social Media, Text Analytics, Explainable AI, E-Learning, Social Network, AR/VR
E-learning/Online Learning is Important

=> Flexible learning location.
=> Learning at Scale.

A Framework to Guide and Education Response to the COVID-19 Pandemic by OECD (Organization for Economic Co-operation and Development), 2020
Challenges in Online Learning

=> Flexible learning location.
=> Learning at Scale.
=> Personalized Learning?
# Powered by Learning data

<table>
<thead>
<tr>
<th>Type</th>
<th>Online learning platforms</th>
<th>Examples</th>
<th>Learning Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Learning</td>
<td>Problem-based: Intelligent tutoring system/Test and quiz systems</td>
<td>Algebra Tutor, SmartTutor/LeetCode, Uva</td>
<td>Problem-solving data</td>
</tr>
<tr>
<td></td>
<td>ill-defined tasks: Online forums, Q/A systems</td>
<td>Reddit/StackOverflow</td>
<td>Online forum data</td>
</tr>
<tr>
<td></td>
<td>Video-based: Learning management system</td>
<td>Canvas, Moodle, Coursera, EdX, Udacity</td>
<td>Video watching data</td>
</tr>
<tr>
<td>Live Lessons</td>
<td>Online tutoring platforms</td>
<td>Cambly, Preply, italki, Zoom</td>
<td>Video and audio communications</td>
</tr>
</tbody>
</table>
Personalized Online Learning

Educators (design loop)

Learners (learning loop)

Planning
Customizing learning goals

Data

Reflecting
Regulating learning behaviors

Understanding and Analyzing
Different learners cognitive and non-cognitive behaviors

Designing
Personlized learning materials and instructions

Learning
Personalizing learning activities
Personalized Online Learning

Data

Planning
Customizing learning goals

Performing
Personalizing learning activities

Reflecting
Regulating learning behaviors

Learners (learning loop)

Educators (design loop)

Understanding and Analyzing
Different learners cognitive and non-cognitive behaviors

Redesigning
Personalized learning materials and instructions
Why Visual Analytics in Personalized Online Learning?

Learners (learning loop)

Customizing
learning goals and personalize activities

Reflecting
self-regulations on learning behaviors

Data

Educators (design loop)

Analyzing
learners behaviors and improve learning design

- Learning motivation, status, behaviors can hardly be defined and modelled using algorithms.
- Learning is a high-risk task that needs careful decision making.
Challenges

**Learners (learning loop)**

- Customizing learning goals and personalize activities

**Educators (design loop)**

- Reflecting self-regulations on learning behaviors
- Analyzing learners behaviors and improve learning design

---

**Data**

- Large heterogeneous data
- Limited expertise and time in data analysis
- Not enough guidance and explanations
- No guarantee of data quality

---

- Learning motivation, status, behaviors can hardly be defined and modelled using algorithms.
- Learning is a high-risk task that needs careful decision making.
Challenges

Learners (learning loop)

- Customizing learning goals and personalize activities
- Limited expertise and time in data analysis
- Not enough guidance and explanations
- No guarantee of data quality

Educators (design loop)

- Reflecting self-regulations on learning behaviors
- Large heterogeneous data
- Limited expertise and time in data analysis
- Lack of motivation, consistent mental model, and actionable plans

- Analyzing learners behaviors and improve learning design

- Learning motivation, status, behaviors can hardly be defined and modelled using algorithms.
- Learning is a high-risk task that needs careful decision making.
Challenges

Customizing
learning goals and personalize activities

- Large heterogeneous data
- Limited expertise and time in data analysis
- Not enough guidance and explanations
- No guarantee of data quality

Reflecting
self-regulations on learning behaviors

- Large heterogeneous data
- Limited expertise and time in data analysis
- Lack of motivation, consistent mental model, and actionable plans

Analyzing
learners behaviors and improve learning design

- Large heterogeneous data
- Limited expertise and time in data analysis
- No predefined model

- Learning motivation, status, behaviors can hardly be defined and modelled using algorithms.
- Learning is a high-risk task that needs careful decision making.
My works

**Learners (learning loop)**

**Customizing**
- Learning goals and personalize activities

**Data**
- Large heterogeneous data
- Limited expertise and time in data analysis
- Not enough guidance and explanations
- No guarantee of data quality

**AlgoSolve** (CHI 2022)
**Persua** (CSCW 2021)
**Peerlens** (CHI 2019)

**Reflecting**
- Self-regulations on learning behaviors

**Data**
- Large heterogeneous data
- Limited expertise and time in data analysis
- Lack of motivation, consistent mental model, and actionable plans

**RLens** (L@S 2022)
**“Game the system”** (L@S 2020)

**Educators (design loop)**

**Analyzing**
- Learner behaviors and improve learning design

**Data**
- Large heterogeneous data
- Limited expertise and time in data analysis
- No predefined model

- **BlockLens** (L@S WIP 2022)
- **QLens** (TVCG 2021)
- **Predicting** (LAK 2020)
- **SeqDynamics** (EuroVis 2020)
- **K-12 Mathematics** (VIS 2019, Best Poster Award)

- Large heterogeneous data
- Limited expertise and time in data analysis
- No guarantee of data quality

- AlgoSolve (CHI 2022)
- Persua (CSCW 2021)
- Peerlens (CHI 2019)
- **RLens** (L@S 2022)
- “Game the system” (L@S 2020)
My works

Learners (learning loop)

**Customizing**
learning goals and personalize activities

- Large heterogeneous data
- Limited expertise and time in data analysis
- Not enough guidance and explanations
- No guarantee of data quality

**Reflecting**
self-regulations on learning behaviors

- Large heterogeneous data
- Limited expertise and time in data analysis
- Lack of motivation, consistent mental model, and actionable plans

Educators (design loop)

**Analyzing**
learners behaviors and improve learning design

- Large heterogeneous data
- Limited expertise and time in data analysis
- No predefined model

---

Data

Peerlens (CHI 2019)
RLens (L@S 2022)

QLens (TVCG 2021)
SeqDynamics (EuroVIS 2020)
My works

Learners (learning loop)

Customizing learning goals and personalize activities
- Large heterogeneous data
- Limited expertise and time in data analysis
- Not enough guidance and explanations
- No guarantee of data quality

Reflecting self-regulations on learning behaviors
- Large heterogeneous data
- Limited expertise and time in data analysis
- Lack of motivation, consistent mental model, and actionable plans

Educators (design loop)

Analyzing learners behaviors and improve learning design
- Large heterogeneous data
- Limited expertise and time in data analysis
- No predefined model

Peerlens (CHI 2019)
RLens (L@S 2022)

QLens (TVCG 2021)
SeqDynamics (EuroVIS 2020)
PeerLens: Peer-inspired Interactive Learning Path Planning in Online Question Pool

Meng Xia, Mingfei Sun, Huan Wei, Qing Chen, Yong Wang, Lei Shi, Huamin Qu, Xiaojuan Ma

CHI 2019
What is an online question pool?
Features of question pools

- No pre-determined syllabus
- A lengthy list indexed by their problem IDs
- Hidden intents

<table>
<thead>
<tr>
<th>No.</th>
<th>Prob. ID</th>
<th>Problem Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1069</td>
<td>1069</td>
<td>A + B Problem</td>
</tr>
<tr>
<td>1070</td>
<td>1070</td>
<td>A + B Problem II</td>
</tr>
<tr>
<td>1071</td>
<td>1071</td>
<td>Max Sum</td>
</tr>
<tr>
<td>1072</td>
<td>1072</td>
<td>List the Balloon nose</td>
</tr>
<tr>
<td>1073</td>
<td>1073</td>
<td>Number Sequence</td>
</tr>
<tr>
<td>1074</td>
<td>1074</td>
<td>Tick and Tick</td>
</tr>
<tr>
<td>1075</td>
<td>1075</td>
<td>Quiet Design</td>
</tr>
<tr>
<td>1076</td>
<td>1076</td>
<td>Elevator</td>
</tr>
<tr>
<td>1077</td>
<td>1077</td>
<td>FatHouse / Trade</td>
</tr>
<tr>
<td>1078</td>
<td>1078</td>
<td>Templer of the Bone</td>
</tr>
<tr>
<td>1079</td>
<td>1079</td>
<td>Starship Troopers</td>
</tr>
</tbody>
</table>

- Different learning scenarios
- One learner’s learning scenario may be changing

**Difficulty:** Determine an appropriate order in taking these online questions for their particular learning scenarios
**Current situation**

<table>
<thead>
<tr>
<th>Programming question pools</th>
<th>Has recommendation?</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtCoder</td>
<td>NO</td>
</tr>
<tr>
<td>CodeChef</td>
<td>NO</td>
</tr>
<tr>
<td>CodeFights</td>
<td>NO</td>
</tr>
<tr>
<td>Codeforces</td>
<td>NO</td>
</tr>
<tr>
<td>Codewars</td>
<td>YES (Similar questions)</td>
</tr>
<tr>
<td>LeetCode</td>
<td>YES (Similar questions)</td>
</tr>
<tr>
<td>CodinGame</td>
<td>NO</td>
</tr>
<tr>
<td>Coderbyte</td>
<td>NO</td>
</tr>
<tr>
<td>CSAcademy</td>
<td>NO</td>
</tr>
<tr>
<td>HackerEarth</td>
<td>NO</td>
</tr>
<tr>
<td>HackerRank</td>
<td>NO</td>
</tr>
<tr>
<td>Kattis</td>
<td>NO</td>
</tr>
<tr>
<td>uDebug</td>
<td>NO</td>
</tr>
<tr>
<td>OmegaUp</td>
<td>NO</td>
</tr>
<tr>
<td>Sphere Online Judge</td>
<td>NO</td>
</tr>
<tr>
<td>Topcoder</td>
<td>NO</td>
</tr>
<tr>
<td>Toph</td>
<td>NO</td>
</tr>
<tr>
<td>URI Online Judge</td>
<td>NO</td>
</tr>
<tr>
<td>UVa Online Judge</td>
<td>NO</td>
</tr>
</tbody>
</table>

**Demand:** planning personalized learning path in the context of existing list-based question pools
Related work: Educational Recommendation Techniques

Memory-based techniques

Continuously analyze current data (Drachsler et al., 2008)

- Content-based (e.g., Chu et al., 2011), Collaborative Filtering (e.g., Toledo et al., 2018), Hybrid approach (e.g., Salehi et al., 2013)

Model-based techniques

Utilize a large amount of historical data to model the learning process over time

- Deep learning models (e.g., Piech et al., 2015), other models, such as Markov Chain (e.g., Rajapakse and Ho, 2005; Sarukkai 2000; Huang et al., 2009)
A user-centered design process

- Four domain experts
  - Instruction designers who designed online question pools (E1, E2)
  - Online question pool users (S1, S2)

- Requirements gathering iteratively for three months
  - **R1:** Find peers for a specific learning scenario.
  - **R2:** Compare with peers’ performance.
  - **R3:** Offer flexible learning path suggestions with explanations.
  - **R4:** Provide convenient interaction and intuitive visual designs for learning path planning.
System overflow
Path Planning Engine: Learning Path Modeling

Submission type: the way a user interacts with a problem.

- Captures learners’ knowledge proficiency
- Enables the inference of question difficulty level
Path Planning Engine: Path Suggestion

A given peer path \([ (X_{i_0}, E_{i_0}, t_{i_0}), ..., (X_{i_n}, E_{i_n}, t_{i_n}) ] \) corresponds to a state \( s = \{ X_{i_0}, X_{i_1}, ..., X_{i_n} \} \).

Markov Chain:

Popular path: \( X_1 \rightarrow X_3 \rightarrow X_4 \)

Challenging path: \( X_1 \rightarrow X_4 \)

Progressive path: \( X_3 \rightarrow X_4 \rightarrow X_1 \)
Visual Design: Peer Selection View

- **Regular Learning**: regularly for a long time and solve 1-2 problems per day.
- **Intensive Learning**: 1-3 months, solve 2-5 questions per day with high proficiency.
- **Advanced Learning**: solve many problems per day in short time with high proficiency.

**Yellow diamond plot: selected peers**

**Blue diamond plot: learner himself**
Visual Design: Learning Path View
Visual Design: Learning Path View
Visual Design: Learning Path View
Visual Design: Learning Path View
Visual Design: Learning Path View
Visual Design: Learning Path View

Submission Type:
- Red: Tried only once and failed
- Orange: Tried several times and failed
- Yellow: Tried several times and succeeded
- Green: Tried, succeeded and validated
- Light Green: Succeeded and then validated

 Paths:
- **b1**: History Path
- **b2**: Challenging Path
- **b3**: Popular Path
- **b4**: Progressive Path

Problem ID 2602
- Type: 2%
- 6: 3%
- 9: 13%
- 10: 44%
- 11*: 6%
- 1: 29%
- 1*: 0%

Controls:
- Slide
- OK

Source: [Image](image-url)
Visual Design: Learning Path View
Visual Design: Learning Path View
Evaluation: Experiment Design

**Dataset:**
A popular programming question pool
~4.6M submission records
~54K learners
~5K programming questions

**Participants:**
18 (7 females, 11 males, age:24±2.85), from a local computer science department

**Systems:**
S1. Full PeerLens
S2. Baseline system
S3. Primitive PeerLens
Evaluation: Experiment Design

Learning scenarios:
L1. Basic programming practice
L2. Coding qualification test for IT company interviews
L3. International Programming Contest

Within-subject:
Counter balance the three learning scenarios and three systems

Tasks:
1. Determine the starting question under a specific learning scenario
2. Find the next question to solve given an existing historical learning path
## Evaluation: Questionnaires

<table>
<thead>
<tr>
<th>Informativeness</th>
<th>Decision making</th>
<th>Visual design</th>
<th>System Usability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Q1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The information needed to plan a learning path is easy to access.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Q2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The information needed to plan a learning path is rich.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Q3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The information is sufficient to plan a learning path.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Q4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The system was helpful for me to find a proper learning path for a specific learning scenario.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Q5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am confident that I find a suitable learning path for the learning scenario.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Q6</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The system helps make adjustment according to previous performance.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Q7</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The learning path design is intuitive.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Q8</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The learning path design helps me understand the suggested path.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Q9</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was easy to learn the system.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Q10</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was easy to use the system.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Q11</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I would like to recommend this system to others.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

Informativeness and decision-making efficacy

- Primitive and Full PeerLens > Baseline

- Information richness & sufficiency: Full PeerLens > Primitive

- Information accessibility: No significant differences between Full and Primitive

- Decision-making metrics: Full PeerLens > Primitive
Results

Visual designs and system usability

- Intuitiveness & comprehension: Full PeerLens > Primitive
- Easy to learn & use: No significant difference between Full and Primitive
- Recommendation: Full PeerLens > Primitive
Conclusion

- A novel visual analytics system for comparing event sequence data and providing explanation for recommendation

- A novel zipper-like visualization for showing information step by step to reduce cognitive load

- A within-subject user experiment to show the system usefulness and usability
**My works**

**Learners (learning loop)**

- Customizing learning goals and personalize activities

  - Large heterogeneous data
  - Limited expertise and time in data analysis
  - Not enough guidance and explanations
  - No guarantee of data quality

- Reflecting self-regulations on learning behaviors

  - Large heterogeneous data
  - Limited expertise and time in data analysis
  - Lack of motivation, consistent mental model, and actionable plans

**Educators (design loop)**

- Analyzing learners behaviors and improve learning design

  - Large heterogeneous data
  - Limited expertise and time in data analysis
  - No predefined model

---

**Data**

- Peerlens (CHI 2019)
- QLens (TVCG 2021)
- RLens (L@S 2022)
- SeqDynamics (EuroVIS 2020)
**RLens:** A Computer-aided Visualization System for Reflecting Language Learning Progress Under Distributed Tutorship

*Meng Xia, Yankun Zhao*, Jihyeong Hong*, Mehmet Hamza Erol*, Taewook Kim, Juho Kim

*L@S 2022*
Background

- **Gig economy** gains popularity
  - Temporary, flexible jobs are commonplace for efficient resource allocation
- New modes of teaching and learning spring up
Background

In particular, online language tutoring platforms (e.g., Cambly,) are becoming increasingly popular.
Background

These online language tutoring platforms:

- Provide temporary jobs for native speakers to work as part-time tutors
- Enable language learners to have 1-1 speaking sessions with native speakers anytime and anywhere
Distributed Tutorship

Our previous work [1] analyzed 15,959 learners’ data on one of these platforms and identified that learners actively distribute their learning time with different tutors during the learning process, which was defined as distributed tutorship.

Distributed Tutorship

There is suggestive evidence that more distributed tutorship might introduce lower learning improvement [1].
An online English tutoring platform. On Ringle, learners can choose tutors and class time for 1:1 online speaking sessions.

16 learners, who have learnt from more than one tutor.

For example, how do you calibrate your progress when you have taken multiple sessions with different tutors? Have you encountered any difficulties?
Challenge #1: Grading Inconsistency

Learners have a hard time knowing their improvement through checking the scores of different tutors, since each tutor might have different grading standards.
Challenge #2: Feedback Discontinuity

Learners are unaware of their common language issues (e.g. tense errors) and they are not sure whether they have corrected the issues or not, since previous corrections are not tracked by different tutors.
Challenge #3: Unorganized Feedback

Learners are uncertain about what to do next, since suggestions given by different tutors are from diverse perspectives.
Challenge 4#: Lack of Context for Feedback Understanding

It would take a long time to find the corresponding place in the audio recording where the feedback was given.
We want to design **intuitive** learning dashboard.

Learners are not data scientists. Exhaustive visual analysis may not be their best option.

We need to build an **intuitive** visual analysis system to reduce the cognitive load.
Challenges

C1: Grading Inconsistency
C2: Feedback Discontinuity
C3: Unorganized Feedback
C4: Lacking context for feedback understanding

Design Requirements

R1: Provide a data-driven assessment and compare with tutors' scores.
R2: Identify common language issues and track feedback uptake behavior.
R3: Organize tutor feedback automatically into different categories.
R4: Map tutor feedback to transcripts.
R5: Provide intuitive visualizations to present learning progress.

Design Elements

Overview to compare computer scores and tutor scores.
Correction View to demonstrate the trend of feedback uptake behaviors.
Suggestion View to display the distribution of suggestion categories.
Transcript View to link corrections with transcript.
NLP +
The correction view shows the common language issues.
Data Driven Evaluation

Feedback Categorization

Grammar

Identify and Track Common Language Issues

Feedback-Context Mapping
**Data Driven Evaluation**

**Vocabulary Complexity**: measure of textual lexical diversity (MTLD), the average length of sequential words a speaker can produce that keep the type-token ratio (TTR) higher than 0.72.

**Grammar Accuracy**: ratio of error-free C-Units to the total number of C-Units, where C-Unit is defined as the minimal communication unit (e.g., ```Yes.```)

**Fluency**: Mean Length of Run (MLR), the average number of syllables per utterance without any pause, where the threshold for pause identification is set to 250 ms.
Data Driven Evaluation

Feedback Categorization

Grammar

Identify and Track Common Language Issues

Feedback-Context Mapping
Corrective feedback: (e.g., two apple -> two apples, think positively -> optimistic)

Feedback uptake behaviors

Grammar: whether still have the grammar issue mentioned by previous tutors

Vocabulary: for each session, check using masked language modeling
- whether the suggested expressions have been used correctly
- whether the original expressions were still used incorrectly
Data Driven Evaluation

Identify and Track Common Language Issues

Feedback Categorization

Feedback-Context Mapping
(1) **Manually selected six categories** with the help of three tutors
(2) **Sentence classification** for each feedback sentence using **natural language inference techniques**
Data Driven Evaluation

Identify and Track Common Language Issues

Feedback Categorization

Feedback-Context Mapping
Map tutors' feedback to the transcripts based on the sentence similarity.
User Study

A between-subjects study on a Baseline system and RLens.

40 learners from Ringle, who have learnt from more than one tutor and 25 sessions.
User Study

Baseline: 20 learners

RLens: 20 learners
User Study

T1: Please describe your overall learning progress.
T2: Please identify your common language issues in the learning process.
T3: Please describe whether you have corrected your common language issues in the learning process.
T4: Please describe the common aspects in tutors' overall feedback.
T5: Please describe how you check the transcript using the system for learning.
T6: Please describe the reasons for ups and downs in scores showing in Overview.
T7: Please describe how you will use this system in learning reflection if it is deployed.
Contributions

- **A computer-aided visualization system** for analyzing audio/text learning data to facilitate learners’ reflection on the learning process under **distributed tutorship**

- **A user study** showing the effectiveness of reflecting learning progress with RLens

- **A set of design considerations** for computer-aided learning systems under distributed tutorship, e.g., surfacing actionable information
My works

Learners (learning loop)

Customizing
learning goals and personalize activities

- Large heterogeneous data
- Limited expertise and time in data analysis
- Not enough guidance and explanations
- No guarantee of data quality

Data

Reflecting
self-regulations on learning behaviors

- Large heterogeneous data
- Limited expertise and time in data analysis
- Lack of motivation, consistent mental model, and actionable plans

Educators (design loop)

Analyzing
learners behaviors and improve learning design

- Large heterogeneous data
- Limited expertise and time in data analysis
- No predefined model

Peerlens (CHI 2019)

RLens (L@S 2022)

QLens (TVCG 2021)

SeqDynamics (EuroVIS 2020)
QLens: Visual Analytics of Multi-step Problem-solving Behaviors for Improving Question Design

Meng Xia, Reshika Palaniyappan Velumani, Panpan Xu, Yong Wang, Huamin Qu, Xiaojuan Ma
TVCG 2021
A Multi-step Problem

Five people stand in a line.
Mark stands ahead of Paul.
Helen stands ahead of Jane.
Paul stands behind Helen but ahead of Luke.
No boy is next to another boy in the line.

Move each person to their place in the line.
Motivation

- Problem-solving logic
- Engagement level
- Difficulties

Question Designer
Related work

Feature-based projections for effective playtrace analysis (Liu et al., 2011)

Visual Analytics of Student Learning Behaviors on K-12 Mathematics E-learning Platforms (Xia et al., VIS 2019) Best Poster Award

States cannot reflect students’ thinking logic
QLens for question designers

Inspect

Analyze

Compare
A user-centered design process

● Four domain experts
  ○ Question designers (E1, E2)
  ○ System developer (E3)
  ○ Project manager (E4)

● Requirements gathering iteratively >= one year

  R1: Show students’ overall problem-solving performance.
  R2: Summarize and present the multi-step problem-solving behaviors.
  R3: Enable the comparison of students from different groups.
  R4: Evaluate the feasibility of providing feedback based on existing data.
System overview

Dataset
- Mouse Movements Collection

Preprocessing
- Intermediate Steps Mapping

Analysis
- Hybrid State Transition Modelling
- Data-driven Path Recommendation

Visualization

Interaction

Query
1. Data Preprocessing

<table>
<thead>
<tr>
<th>Source URL</th>
<th><a href="http://mad9.learnlex.com/storage/mad/questions/2xbbe2fdb4aec4e218/">http://mad9.learnlex.com/storage/mad/questions/2xbbe2fdb4aec4e218/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Element Path</td>
<td>HTML#,.BODY#.en,DIV#question_content.singlepage,DIV#std_wrapper...</td>
</tr>
<tr>
<td>Question ID</td>
<td>geometry23567</td>
</tr>
<tr>
<td>Time Stamp</td>
<td>20190122T1022</td>
</tr>
<tr>
<td>Client Width</td>
<td>1920</td>
</tr>
<tr>
<td>X</td>
<td>567</td>
</tr>
<tr>
<td>Touch Screen</td>
<td>True/False</td>
</tr>
<tr>
<td>Platform</td>
<td>Windows/MacOS/iOS</td>
</tr>
</tbody>
</table>

April 2019 to January 2020, 2,30,644 records from 5,266 students and 1,718 mathematical questions.
1. Data Preprocessing

For each question:

1. Data Preprocessing

For each student:

Step1: """", Mark
Step2: Paul,"""", Mark
Step3: Paul, Helen,"""", Mark
Step4: Paul,"""", Helen,"""", Mark

2 11 4 7 3 8 8 9 ...

...
2. Data Analysis - State Transition Model

**Step:** the smallest user interface interaction that changes the intermediate answers

**Stage:** the number of conditions the current answer fulfills

**Condition:** one criteria that students need to fulfill to get the partial score

<table>
<thead>
<tr>
<th>Condition</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark &gt; Paul</td>
<td>149/233</td>
</tr>
<tr>
<td>Helen &gt; Jane</td>
<td>140/233</td>
</tr>
<tr>
<td>Luke &gt; Paul &gt; Helen</td>
<td>78/233</td>
</tr>
<tr>
<td>No boys near each other</td>
<td>0/233</td>
</tr>
</tbody>
</table>

Step1: ,,„,Mark  
Step2: Paul,„„,Mark  
Step3: Paul,Helen,„„,Mark  
Step4: Paul„„,Helen„„,Mark

...
3. Visualization - State Transition Visualization

Cannot show event sequence data in which each step has multiple events.

Sequence Synopsis: Optimize Visual Summary of Temporal Event Data (Chen et al., 2017)

Pathviewer: Visualizing pathways through student data (Wang et al., 2017)
3. Visualization - State Transition Visualization

State:

Level 1: \{Step, Stage\} + \{Condition array, Time elapse, Trajectory length\}

Level 2: \{Intermediate answer\}
3. Visualization - State Transition Visualization

State:

Level1: \{Step, \text{Stage}\} + \{Condition array, Time elapse, Trajectory length\}

Level2: \{Intermediate answer\}
3. Visualization - State Transition Visualization

State:

Level 1: \{Step, Stage\} + \{Condition array, Time elapse, Trajectory length\}

Level 2: \{Intermediate answer\}
3. Visualization - State Transition Visualization

State:

Level 1: \{Step, Stage\} + \{Condition array, Time elapse, Trajectory length\}

Level 2: \{Intermediate answer\}
3. Visualization - State Transition Visualization

State:

Level1: {Step, Stage} + {Condition array, Time elapse, Trajectory length}

Level2: {Intermediate answer}
3. Visualization - State Transition Visualization

State:

Level 1: \{\text{Step, Stage}\} + \{\text{Condition array, Time elapsed, Trajectory length}\}

Level 2: \{\text{Intermediate answer}\}
Case studies
Use all the digits to make two 3-digit numbers whose product is the largest possible.

1 2 3 4 5 6

Correct answer: 
6 3 1 \times 5 4 2

The larger the multiplicand and the multiplier are, the larger the product is.

To make the largest product, the two largest numbers should be used as the hundreds digit.

6 \times 5

Then the next two largest numbers (3 and 4) should be used for the tens digit.

Consider 2 cases:

Case 1. 
637 \times 40, must be larger than 24000

Case 2. 
647 \times 30, must be smaller than 24000
Five people stand in a line.

Mark stands ahead of Paul.
Helen stands ahead of Jane.
Paul stands behind Helen but ahead of Luke.
No boy is next to another boy in the line.

Move each person to their place in the line.
Consider the condition with the most restrictions: “No boy is next to another boy in the line.” There are only 3 boys and 2 girls, so we have

```
Back  Boy  Girl  Boy  Girl  Boy  Front
```

“Mark stands ahead of Paul”, “Paul stands ahead of Luke”. Therefore,

```
Back  Luke  Girl  Paul  Girl  Mark  Front
```

“Helen stands ahead of Jane”. Therefore,

```
Back  Luke  Jane  Paul  Helen  Mark  Front
```
Evaluation

- **Cases studies** with four domain experts during the development
- **Semi-structured interviews** with another three domain experts (two questions designers form a different education company, one senior manager); each interview lasts about 1.5 hours
Evaluation

System usefulness

“The insights from Transition View will be very useful for the question designer (for example to decide which question is more suitable for which grade students) and the system developer.”

--- E6

“As more and more learning activities conducted are online, it was also very useful to compare students from different schools (e.g., international and local ones) or regions.”

--- E5

“The on-the-fly guidance is what we expected but needs more considerations.”

--- E5

Visual design & interactions

“It is so clear to view the problem-solving process using the visualization like this (Transition View).”

--- E7

Overall, all experts confirmed the **usefulness** and the **intuitiveness** of the system.
Our collaborator, TrumpTech, uses QLens to improve questions design. The company now serves for 100,000 students from more than 500 schools in Hong Kong.
Conclusion

- An interactive visual analytics system on multi-step question design by analyzing click stream data

- A novel glyph-embedded Sankey diagram for analyzing event sequence trend and comparison, where each step has multiple events

- Three case studies and interviews with domain experts to show the usefulness and usability
My works

Learners (learning loop)

Customizing
learning goals and personalize activities

- Large heterogeneous data
- Limited expertise and time in data analysis
- Not enough guidance and explanations
- No guarantee of data quality

Data

Reflecting
self-regulations on learning behaviors

- Large heterogeneous data
- Limited expertise and time in data analysis
- Lack of motivation, consistent mental model, and actionable plans

Educators (design loop)

Analyzing
learners behaviors and improve learning design

- Large heterogeneous data
- Limited expertise and time in data analysis
- No predefined model

- Large heterogeneous data
- Limited expertise and time in data analysis
- No guarantee of data quality

Peerlens (CHI 2019)

QLens
(TVCG 2021)

SeqDynamics
(EuroVIS 2020)

RLens (L@S 2022)
SeqDynamics: Visual Analytics for Evaluating Online Problem-solving Dynamics

Meng Xia, Min Xu, Chuan-en Lin, Ta Ying Cheng, Huamin Qu, Xiaojuan Ma

EuroVis 2020
Elite Selection in University

Interview in IT Company
Elite Selection in University

Interview in IT Company

Cognitive skills (think, read, learn, remember, reason, and pay attention)

Noncognitive traits (motivation, conscientiousness, perseverance, self-regulation, and collaboration)
Elite Selection in University

Interview in IT Company

Cognitive skills (think, read, learn, remember, reason, and pay attention)

Noncognitive traits (motivation, conscientiousness, perseverance, self-regulation, and collaboration)

Exams/Technical interviews

Performance and behavior on a long period.

OR
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>23412857</td>
<td>2017-12-28 00:03:33</td>
<td>Accepted</td>
<td>2046</td>
<td>0MS</td>
<td>1700K</td>
<td>310B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23412041</td>
<td>2017-12-27 22:24:33</td>
<td>Accepted</td>
<td>2045</td>
<td>0MS</td>
<td>1696K</td>
<td>309B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23411734</td>
<td>2017-12-27 21:52:45</td>
<td>Wrong Answer</td>
<td>2045</td>
<td>0MS</td>
<td>1700K</td>
<td>388B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23411669</td>
<td>2017-12-27 21:45:25</td>
<td>Wrong Answer</td>
<td>2045</td>
<td>0MS</td>
<td>1696K</td>
<td>382B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23411286</td>
<td>2017-12-27 21:10:04</td>
<td>Accepted</td>
<td>2044</td>
<td>15MS</td>
<td>2052K</td>
<td>410B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23411278</td>
<td>2017-12-27 21:09:05</td>
<td>Wrong Answer</td>
<td>2044</td>
<td>0MS</td>
<td>2048K</td>
<td>404B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23410518</td>
<td>2017-12-27 20:38:17</td>
<td>Wrong Answer</td>
<td>2044</td>
<td>0MS</td>
<td>2052K</td>
<td>404B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23408905</td>
<td>2017-12-27 17:39:49</td>
<td>Accepted</td>
<td>2043</td>
<td>0MS</td>
<td>1700K</td>
<td>899B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23408506</td>
<td>2017-12-27 16:52:47</td>
<td>Accepted</td>
<td>2042</td>
<td>0MS</td>
<td>1696K</td>
<td>254B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23405316</td>
<td>2017-12-27 01:33:38</td>
<td>Accepted</td>
<td>2041</td>
<td>15MS</td>
<td>1708K</td>
<td>360B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23405296</td>
<td>2017-12-27 01:21:02</td>
<td>Accepted</td>
<td>2040</td>
<td>78MS</td>
<td>1688K</td>
<td>433B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23405284</td>
<td>2017-12-27 01:11:27</td>
<td>Accepted</td>
<td>2039</td>
<td>0MS</td>
<td>1740K</td>
<td>280B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23405283</td>
<td>2017-12-27 01:10:20</td>
<td>Wrong Answer</td>
<td>2039</td>
<td>0MS</td>
<td>1744K</td>
<td>283B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23405282</td>
<td>2017-12-27 01:09:41</td>
<td>Wrong Answer</td>
<td>2039</td>
<td>15MS</td>
<td>1692K</td>
<td>280B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
<tr>
<td>23405277</td>
<td>2017-12-27 01:05:09</td>
<td>Accepted</td>
<td>2037</td>
<td>15MS</td>
<td>1708K</td>
<td>729B</td>
<td>G++</td>
<td>xiameng552180</td>
</tr>
</tbody>
</table>
Problem-solving Dynamics

The process and progress of solving a series of problems over time.

<table>
<thead>
<tr>
<th>Problem ID</th>
<th>Results</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Easy</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Medium</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Easy</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Medium</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>Easy</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>Hard</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>Hard</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>Hard</td>
</tr>
</tbody>
</table>
Problem-solving Dynamics

The process and progress of solving a series of problems over time.
Problem-solving Dynamics

The process and progress of solving a series of problems over time.

<table>
<thead>
<tr>
<th>Problem ID</th>
<th>Results</th>
<th>Difficulty</th>
<th>Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Easy</td>
<td>Jan. 1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Medium</td>
<td>Jan. 1</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Easy</td>
<td>Jan. 2</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Medium</td>
<td>Jan. 3</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>Easy</td>
<td>Jan. 3</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>Hard</td>
<td>Jan. 20</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>Hard</td>
<td>Jan. 20</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>Hard</td>
<td>Jan. 20</td>
</tr>
</tbody>
</table>

Non-cognitive Trait
  e.g., self-regulation

Calendary

117
SeqDynamics

Interactive

Multi-dimensional

Time-series
A user-centered design process

Four domain experts
- Recruiters from the competitive programming team (E1, E2)
- Student coaches (E3, E4)

Requirements gathering iteratively for three months
- R1: Show a clear overview of overall students’ problem-solving performance.
- **R2: Understand problem-solving dynamics from different perspectives over time.** (i.e., cognitive and non-cognitive).
- **R3: Compare/Combine the problem-solving performance at different scales.**
- R4: Support an interactive and customized exploration of the evaluation.
System overview
Problem-solving Feature Extraction

Changes of these features below over time:

Cognitive ability *(Ausubel et al., 1968)*
• L1: number of problems solved
• L2: ratio of hard problems solved
• L3: diversity of problems solved

Non-cognitive traits *(Farkas, 2003)*
• L4: number of submissions *(diligence level)*
• L5: time starting to trying hard problems *(willingness to take challenge)*
• L6: ratio of active days *(perseverance)*
(a) A learner who has many submissions but solves a few problems.  
(b) A learner who has relatively fewer submissions but solves more problems.
Two design alternatives.
Evaluation

Three usage scenarios
Elite Analysis and Selection
Personal Analysis and Training
Team Formation

Five expert interviews
(Three coaches of competitive programming teams and two instructors teaching programming courses)

☐ System Usability
☐ System Effectiveness
☐ Visual Designs
☐ Interactions

“The hexagon can clearly show the strength and weakness of two candidates”

“The encoding (glyph) is very intuitive and I can tell a learner’s talent at a glance”

Overall, all five experts commented that SeqDynamics was useful and easy to use.
Conclusion

● An interactive visual analytical system to compare and rank objects with multiple temporal variables

● Novel glyphs and bilateral stacked graph for comparison over different levels of detail

● Three usage scenarios and five expert interviews to show the system usefulness and usability
Summary

Learners (learning loop)

Customizing
learning goals and personalize activities

- Large heterogeneous data
- Limited expertise and time in data analysis
- Not enough guidance and explanations
- No guarantee of data quality

Data

Reflecting
self-regulations on learning behaviors

- Large heterogeneous data
- Limited expertise and time in data analysis
- Lack of motivation, consistent mental model, and actionable plans

Educators (design loop)

Analyzing
learners behaviors and improve learning design

- Large heterogeneous data
- Limited expertise and time in data analysis
- No predefined model

Peerlens (CHI 2019)

QLens (TVCG 2021)

RLens (L@S 2022)

SeqDynamics (EuroVIS 2020)
1. **Integration of learning analytics and learning design**
   a. how to lower the barrier of learning design
   b. how to support data-driven learning design
2. Real-time/synchronized personalized learning
   a. Zoom: how to engage both instructors and learners?
   b. Class-room/situated education: can we utilize immersive learning analytics?

Exploring Interactions with Printed Data Visualizations in Augmented Reality, (Tong et al., VIS 2022, conditionally accepted)
Future Work

3. **Empower learners**
   a. how to infer learners' psychological state
   b. how to motivate learners
   c. how to design effective data visualization with low cognitive load
   d. how to guarantee data quality [1]

---


Future Work

Learners (learning loop)

Customizing
learning goals and personalize activities

Reflecting
self-regulations on learning behaviors

Educators (design loop)

Analyzing
learners behaviors and improve learning design

Data

4. Personalization in diverse learning scenarios

✓ Learning with different teachers or platforms [1]
✓ Learning with different hardware (smartphones, tablets, smart glasses)
✗ Learning with different scenarios (collaborative learning)


Future Work

5. Visual analytics for personalization in other domains

Customizing learning goals and personalize activities

Reflecting self-regulations on learning behaviors

Analyzing learners behaviors and improve learning design

Data

5. Visual analytics for personalization in other domains

Ongoing: NFT investment strategy

Ongoing: Surgery Data Analysis

Ongoing: XAI

Healthcare

Finance

AQX: Explaining Air Quality Forecast for Verifying Domain Knowledge using Feature Importance Visualization, Reshika et al., IUI 2022
Persuading people to change their opinions is a common practice in online discussion forums on topics ranging from political campaigns to relationship consultation. Enhancing people’s ability to write persuasive arguments could not only practice their critical thinking and reasoning but also contribute to the effectiveness and civility in online communication. It is, however, not an easy task in online discussion settings where written words are the primary communication channel. In this paper, we derived four design goals for a tool...
Publication List

1. **Persua: A Visual Interactive System to Enhance the Persuasiveness of Arguments in Online Discussion**
   Meng Xia, Qian Zhu, Xingbo Wang, Fei Nie, Huamin Qu, Xiaojuan Ma, CSCW 2022

2. **RLens: A Computer-aided Visualization System for Supporting Reflection on Language Learning under Distributed Tutorship**
   Meng Xia, Yankun Zhao*, Jihyeong Hong*, Mehmet Hamza Erol*, Taewook Kim, Juho Kim, L@S 2022

3. **Understanding Distributed Tutorship in Online Language Tutoring**
   Meng Xia, Yankun Zhao, Mehmet Hamza Erol, Jihyeong Hong, Juho Kim, ACM LAK (Learning Analytics & Knowledge) 2022

4. **Exploring Interactions with Printed Data Visualizations in Augmented Reality**
   Wai Tong, Zhutian Chen, Meng Xia, Linping Yuan, Leo Yu Ho Lo, Benjamin Bach, Huamin Qu, VIS 2022 (conditionally accepted)

5. **Bias-Aware Design for Informed Decisions: Raising Awareness of Self-Selection Bias in User Ratings and Reviews**
   Qian Zhu, Leo Yu Ho Lo, Meng Xia, Zixin Chen, Xiaojuan Ma, CSCW 2022 (Accept with minor revision)

6. **Mobile-Friendly Content Design for MOOCs: Challenges, Requirements, and Design Opportunities**
   Jeongyeon Kim, Yubin Choi, Meng Xia, Juho Kim, CHI 2022, Best Paper Award

   Mingzhe Li*, Francesca Spector*, Meng Xia*, Mina Oh*, Peter Cederberg, Yuqi Gong, Kristen Shinozara, Patrick Carrington, CHI 2022

8. **AlgoSolve: Supporting Subgoal Learning in Algorithmic Problem-Solving with Learnersourced Microtasks**
   Kabdo Choi, Hyungyu Shin, Meng Xia, Juho Kim, CHI 2022

9. **Explaining Air Quality Forecast for Verifying Domain Knowledge using Feature Importance Visualization**
   Reshika Palaniyappan Velumani, Meng Xia, Jun Han, Chaoli Wang, Alexis Lau, Huamin Qu, IUI 2022

10. **BlockLens: Visual Analytics of Student Coding Behaviors in Block-Based Programming Environments**
    Sean Tsung, Huan Wei, Haotian Li, Meng Xia, Yong Wang, Huamin Qu, L@S 2022 (Work In Progress)

11. **QLens: Visual Analytics of Multi-step Problem-solving Behaviors for Improving Question Design**
    Meng Xia, Reshika Palaniyappan Velumani, Yong Wang, Huamin Qu, Xiaojuan Ma, VIS 2020 (TVCG 2021)
12. Investigating the Effects of Robot Engagement Communication on Learning from Demonstration
Mingfei Sun, Zhenhui Peng, Meng Xia, Xiaojuan Ma, International Journal of Social Robotics 2021

13. Using Information Visualization to Promote Students’ Reflection on “Gaming the system” in Online Learning
Meng Xia, Yuya Asano, Joseph Jay Williams, Huamin Qu, Xiaojuan Ma, L@S 2020

14. SeqDynamics: Visual Analytics for Evaluating Online Problem-solving Dynamics
Meng Xia, Min Xu, Chuan-en Lin, Ta-ying Cheng, Huamin Qu, Xiaojuan Ma, EuroVIS 2020

15. Predicting Student Performance in Interactive Online Question Pools Using Mouse Interactions
Huan Wei, Haotian Li, Meng Xia, Yong Wang, Huamin Qu, ACM LAK (Learning Analytics & Knowledge) 2020

16. Visual Analytics of Student Learning Behaviors on K-12 Mathematics E-learning Platforms
Meng Xia, Huan Wei, Min Xu, Leo Yu Ho Lo, Yong Wang, Rong Zhang, Huamin Qu, IEEE VIS 2019 Poster, Best Poster Award

17. PeerLens: Peer-inspired Interactive Learning Path Planning in Online Question Pool
Meng Xia, Mingfei Sun, Huan Wei, Qing Chen, Yong Wang, Lei Shi, Huamin Qu, Xiaojuan Ma, CHI 2019

18. Generation of Thangka Relief from Line Drawings
Meng Xia, Rong Zhang, Ren Peng, Jinhui Yu, SCIENTIA SINICA Informationis 2018

Ke Xu, Meng Xia, Xing Mu, Yun Wang, Nan Cao, TVCG 2018

20. Exploring How Software Developers Work with Mention Bot in GitHub
Zhenhui Peng, Jeehoon Yoo, Meng Xia, Sunghun Kim, Xiaojuan Ma, in Proc. of Chinese CHI 2018

Mingfei Sun, Yiqing Mou, Hongwen Xie, Meng Xia, Michelle Wong, Xiaojuan Ma, in Proc. of IEEE SMC 2018

22. Deep Spherical Panoramic Representation for 3D Shape Recognition

23. Designing Kinect Game based on Video Tracking
Yinglie Zhang, Meng Xia, Linqiang Chen, Computer Engineering and Applications 2015
Thank you!

Email: mengxia@andrew.cmu.edu
Homepage: https://www.xiameng.org
Projects

01. An Open Learning Design, Data Analytics and Visualization Framework for E-learning
   - HKUST & HKU & MIT
   - Core Member
   - 2018 - 2021

02. Integration of Learning Design and Learning Analytics
   - HKUST & HKU
   - Project Coordinator
   - 2020 - 2021

03. Analysis of Learning Progress and Recommendation of Personalized Learning Paths for English Learners
   - KAIST
   - Project Coordinator
   - 2021 - 2022

04. Cognitive Tutor Authoring Tools (CTAT) for educational researchers
   - Carnegie Mellon University
   - Project Coordinator
   - Since 2022
Advisors/
Mentors

Collaborators

Interns /
Mentees

7 countries

Offers from Stanford,
CMU, Oxford, etc.
Publication List

CCF A (9): CHI (4, including 1 Best Paper), TVCG (3), CSCW (2)

CORE A Conference in Education Technology (5): LAK (2), L@S (3, including 1 Work in Progress)

CCF B (2): IUI (1), EuroVis (1)

CCF C (1): SMC (1)

CCF T1 (2): SCIENTIA SINICA Informationis (1), CAD&CG (1)

Others (4): VIS (1 Best Poster), Computer Engineering and Applications (1), International Journal of Social Robotics (1), Chinese CHI (1)
Scholarships and Awards

KAIST

Best paper award at CHI 2022

HKUST

RGC Postdoctoral Fellowship (PDFS) 2021 (only 50 each year in HK)
SENГ TOP RPg Award, 2018-2019
Best Poster Award at VIS, 2019
Overseas Research Award, 2018-2019

Zhejiang University

National Scholarship, 2015
Chairman of Postgraduate Association of Computer Science Department, 2014-2015
Outstanding graduate student and student cadres, 2014-2015

Hangzhou Dianzi University

National Scholarship, 2011

Best Poster Award at VIS, 2019
Best Paper Award at CHI, 2022
Teaching Plan

1. Data Visualization
2. Human Computer Interaction
3. Computer Organization
5. Personalized Online Learning
Professional Service

- Program Committee member for CHI 2023
- Program Committee member for VIS 2022
- Program Committee member for CHI 2022 LBW
- Program Committee member for VIS 2021
Patent List

1. Apparatus and Method for Evaluating Search Engine Performance, and Dashboard
   ○ KAIST, Jaehoon Lee, Juho Kim, Kabdo Choi, Mehmet Hamza Erol, Hyunwoo Kim, and Meng Xia, 10-2022-0026112

2. English conversation skill analysis using dialogue transcript
   ○ Jihyeong Hong, Meng Xia, Mehmet Hamza Erol, Juho Kim, KAIST, 10-2021-0106202

3. Utilizing tutor feedback for fine-grained learning progress reflection in online English tutoring via interactive visualization
   ○ Meng Xia, Jihyeong Hong, Mehmet Hamza Erol, Juho Kim, KAIST, 10-2021-0106212

4. QLens: Visual Analytics of Multi-step Problem-solving Behaviors for Improving Question Design
   ○ Meng Xia, Reshika Palaniyappan Velumani, Yong Wang, Huamin Qu, Xiaojuan Ma, Hong Kong University of Science and Technology, No.: US 63/102508

5. 一种将唐卡线描图生成浮雕效果的方法
   ○ 于金辉, 夏梦, 浙江大学, ZL 2015 1 1003097.2
Talks and presentations

CHI 2019
IEEE VIS 2019
L@S 2020
Euro VIS 2020
IEEE VIS 2021
LAK 2022
L@S 2022

**CHI 2022: session chair**

Invited Talk at KAIST HCI Course:

**Visual Analytics and Its Application in Education**

Invited Talk at VIS group at HKUST and ShanghaiTech University
Plan on the Research Career

- Pushing forward the research in HCI and data visualization, particularly about personalized learning
- Build a team of undergraduate interns, master students, and PhD students from CS, education, and design
- Teaching courses about HCI, Data Visualization, Personalized Online Learning
- Apply for the Overseas Excellent Youth
Personalized Online Learning

Learners (learning loop)

Customizing
learning goals and personalize activities

Data

Reflecting
self-regulations on learning behaviors

Educators (design loop)

Analyzing
learners behaviors and improve learning design