Ph.D. Qualifying Exam

A Survey on Visual Analytics of Online Learning Data

Presenter: Meng XIA
Supervised by: Prof. Huamin Qu & Prof. Xiaojuan Ma
Department of Computer Science & Engineering

Outline

- Introduction
- Data types in online learning
- Visual analytics for different stakeholders
- Conclusion and future work

What is online learning?

Online learning includes learning with the assistance of the Internet and computing technologies. *(Cristobal Romero &Sebastian Ventura, 2010)*

Common online learning platforms

Online learning platforms	Function	Examples
Learning Management System	a software application for administration, documentation, tracking, reporting, and delivery of educational courses.	Canvas, Moodle, Coursera, EdX, Udacity
Intelligent tutoring system	computer system that aims to provide immediate and customized instruction or feedback to learners.	Algebra Tutor, SmartTutor
Test and quiz systems	evaluate and practice students' level of knowledge by providing a series of questions.	Uva, LeetCode
Learning Objects repositories, wikis, forums, educational games, Q/A systems	sometimes used independently or together with other learning platforms.	StakeOverflow

(Cristobal Romero &Sebastian Ventura, 2010)

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Motivation of Learning Data Analytics

It is powered by data:

Online learning platforms	Data collected
Learning Management Systems	student activities involved, covering reading, writing, taking tests, performing tasks, and discussing with peers
Intelligent Tutoring System	all student–tutor interactions, such as mouse clicks, typing, and even speeches or videos.
Test and Quiz Systems	a great deal of information, such as students' answers, calculated scores, and statistics.
Learning Objects repositories, wikis, forums, educational games, Q/A systems	meta knowledge, questions and answers

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Motivation of Learning Data Analytics

It is required by different stakeholders:

- Learner: reflect on and improve learning performance
- Instructor: improve instruction and assist teaching
- Researcher:
  - propose and develop learning analytics
  - compare different pedagogical methods
- Administrator:
  - allocate educational resources in institutions
  - manage the quality and legitimacy of learning programs

What is visual analytics?

Visual analytics provide visual representations of datasets and interactive technologies to help people carry out tasks more effectively.

Keep human in the loop

Data Mining	+ Visualization	→ Visual Analytics
Clearly defined tasks	High-bandwidth channel	Exploratory analysis
Automated process	External perceptual system	Involve human knowledge
Objective results	Interaction	Deeper understanding

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Challenges in visual analytics

- The **data** in online learning is **large** and **heterogeneous**.
- The **needs** of different user groups are quite **diverse**.
- The **end users** usually have **different** knowledge levels on **visual literacy** and data mining techniques.

Outline

- Introduction
- **Data types in online learning**
- Visual analytics for different stakeholders
- Conclusion and future work

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Descriptions of different data types

Meta Data	Profile information	Learners' profile (demographic, student knowledge)
	Learning materials	Course-like learning materials, non-course-like learning materials
Behavior Data	Interactive behavior	Click-stream, problem solving sequence, submission stream, typing stream
	Communicative activity	Forum discussion, chatting room, Blog
Personal status	Mental/physical status	Mental status(engagement), physical status(gaze behavior)
Performance & others	Students' performance	Assignments, quizzes, and exams
	Others	Family factors, etc.

Profile information

Learners' profile & instructors' profile:

Demographic information	age, gender, education level, income level, marital status, occupation, religion
Administrative information	what platform they study, which class he/she belongs to, and who his/her instructor is
Learners' background knowledge	prior knowledge on the learning material

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Learning materials

Course-like learning materials:

- Videos
- Lectures

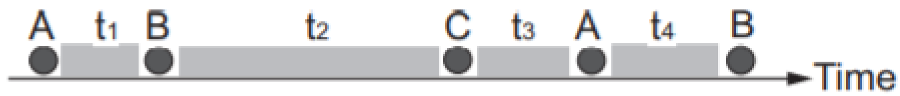
Non-course-like learning materials:

- Questions
- Answers
- Etc.

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Interactive behaviors

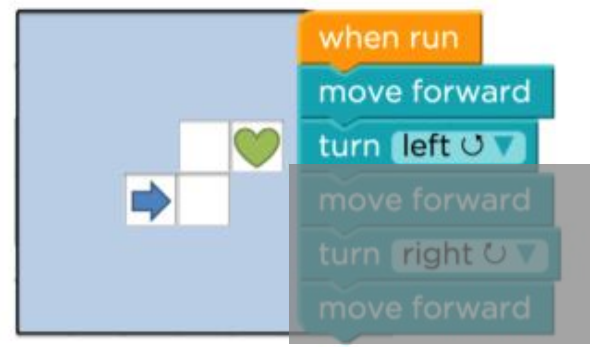
- Clickstream:



$\{User, ClickEvent, Timestamp\}$

- Problem solving sequence:

$\{User, Partial\ Solution, Timestamp\}$



- Submission stream:

$\{User, Problem, Feedback, Timestamp\}$

Run ID	Submit Time	Judge Status	Pro.ID	Exe.Time	Exe.Memory	Code Len.	Language	Author
27657478	2018-12-17 01:19:22	Accepted	1285	31MS	1432K	1063B	G++	赵子龙
27657370	2018-12-17 00:31:18	Wrong Answer	1232	31MS	1376K	636B	G++	赵子龙
27657324	2018-12-17 00:18:13	Wrong Answer	1255	202MS	1776K	3137B	G++	赵子龙

- Typing stream:

Interactive behaviors

Existing datasets:

Datashop	https://pslcdatashop.web.cmu.edu/
ASSISTmentsData	https://sites.google.com/site/assistmentsdata/home/assistment-2009-2010-data
LAK 2018 Challenge	https://sites.google.com/view/lak19datachallenge

Pavlik - Knowledge Tracking PI: Phil Pavlik

Dataset	Domain/LearnLab	Dates	Status	Transactions	
Chinese Vocabulary Spring 2006	Language/Chinese	May 9, 2006 - May 9, 2006	complete	71,928	
Spanish Vocabulary Spring 2006	Other	Feb 4, 2006 - May 23, 2006	complete	17,063	
Chinese Vocabulary Transfer Lab Study Spring 2006	Language/Chinese	May 24, 2006 - Jul 13, 2006	complete	19,008	
Chinese Vocabulary Fall 2006	Language/Chinese	Oct 18, 2002 - Dec 19, 2006	complete	108,206	
Chinese Vocabulary Spring 2007	Language/Chinese	Jan 22, 2007 - May 11, 2007	complete	95,508	
Chinese Vocabulary Schedule Preference Fall 2007	Language/Chinese	Sep 10, 2007 - Dec 16, 2007	complete	127,083	
Chinese Radical Transfer Fall 2007	Language/Chinese	Sep 10, 2007 - Oct 10, 2007	complete	61,323	
Chinese Radical Transfer Fall 2007 - October	Language/Chinese	Oct 1, 2007 - Oct 10, 2007		5,076	

Communicative activities

Forum discussion

Chatting room

Blog

All Threads

Top threads | Last updated | Last created

Thread	Points	Posts	Views
The Role of Goverment in Mexico for Entrepreneurship Started by Héctor Manuel Aceves Ortega · Last post by Héctor Manuel Aceves Ortega (5 days ago)	0 points	1 post	2 views
Course evaluation STAFF REPLIED · Started by anyway samutanda · Last post by Lénárd HORGOS (11 days ago)	0 points	23 posts	125 views
How long will this course site be open and accesible? Started by Amando Sosa · Last post by Amando Sosa (17 days ago)	0 points	5 posts	37 views
Certificate Received. Thanks All & Keep in touch ! STAFF REPLIED · Started by Sandeep Asolkar · Last post by Suzanne Healy STAFF (18 days ago)	0 points	6 posts	28 views
Certification STAFF REPLIED · Started by Wesley Tam · Last post by Suzanne Healy STAFF (19 days ago)	0 points	2 posts	19 views
When will grading be finished?? STAFF REPLIED · Started by Lydia Victor · Last post by Suzanne Healy STAFF (19 days ago)	0 points	7 posts	80 views
Signature Track query STAFF REPLIED · Started by Ekta Navani · Last post by Suzanne Healy STAFF (19 days ago)	0 points	2 posts	9 views
Who are the angel investors in your community? STAFF REPLIED · Started by Anna Perlmutter STAFF · Last post by Lydia Victor (20 days ago)	0 points	64 posts	196 views
Which of the three would you choose to be? STAFF REPLIED · Started by George Pligor · Last post by SERGE ISAAC KOIDJO KOUAME (a month ago)	0 points	20 posts	70 views
The "Role of Anchor Institutions" in Economic Growth Started by Tina Balani · Last post by SERGE ISAAC KOIDJO KOUAME (a month ago)	0 points	18 posts	76 views
The Role of Intermediaries in Entrepreneurship Started by Yong Sebastian Nyam · Last post by SERGE ISAAC KOIDJO KOUAME (a month ago)	0 points	5 posts	19 views
How many organizations we need? Started by DIMTSAS VASILEOS · Last post by SERGE ISAAC KOIDJO KOUAME (a month ago)	0 points	9 posts	17 views

MOOC forum discussion

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Mental/physical status

Mental status
- Emotion
- Engagement

Physical status
- Gaze
- Heart rate
- Body temperature

Data Types in Online Learning

Performance data

Quizzes, Assignments, and Exams

Others

Family factors

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Discussion

Characteristics of different data types

Meta Data	Profile information	easy to acquire, privacy issue
	Learning materials	easy to acquire, widely used
Behavior Data	Interactive behavior	easy to acquire, widely used, need huge storage
	Communicative activity	easy to acquire, widely used, need huge storage
Personal status	Mental/physical status	valuable, need additional sensor
Performance & others	Students' performance	easy to acquire, widely used
	Others	valuable, difficult to acquire

Outline

Visual Analytics

Requirements	Data type	Visual Analytics
Obtaining personalized recommendations	learning materials interactive behavior students performance	(Zhao et al., 2018)
Getting adaptive feedback	learning material interactive behavior students performance	None
Achieving self-regulation	profile information interactive behavior students performance	(Bull et al., 2016) (Azevedo et al., 2017) (Ruiz et al., 2016)
Understanding the learning material utilization	learning material interactive behavior	(Shi et al., 2015) (Chen et al., 2016) (He et al., 2018)
Identifying learning behavior patterns	profile information interactive behavior students performance	(Coffrin et al., 2014) (Chen et al., 2016) (Du et al., 2016) (Chen et al., 2018)
Monitoring learning progress, engagement, collaboration	interactive behavior mental/physical status	(Sharma et al., 2013) (Sharma et al., 2016) (Wang et al., 2017)
Understanding motivation	profile information	None
Behavior modeling	learning material interactive behavior students performance	None
Predicting learners performance	interactive behavior communicative activity	(Piech et al., 2015a)
Social network analysis	interactive behavior communicative behavior	(Fu et al., 2017) (Fu et al., 2018)
Retaining learners or avoiding dropout	interactive behavior others	(Chen et al., 2016)
Anomaly detection and volume control	profile information learning material interactive behavior	None

Learner | Instructor | Researcher | Administrator

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Facilitating study for learners

Requirements	Data Mining	Visual Analytics
Obtaining personalized recommendations	*(Chu et al., 2011)* *(Toledo et al., 2012)* *(Salehi et al., 2013)* *(Huang et al., 2009)* *(Piech et al., 2015)*	*(Zhao et al., 2018)*
Getting adaptive feedback	*(Hieu et al., 2017)* *(R. Singh, 2015)* *(K. Zimmerman and C. R. Rupakheti, 2015)* *(P. Freeman et al., 2016)* *(Thomas et al., 2016)*	
Achieving self-regulation	*(Daniele et al., 2017)*	*(Cicchinelli et al., 2018)* *(Bull et al., 2016)* *(Azevedo et al., 2017)*

Obtaining personalized recommendations

- Content-Based (CB) recommends items based on relationships between learning materials, *e.g., (Chu et al., 2011)*.
  No consideration on user-item interaction.
- Collaborative Filtering (CF) recommends items that were used by other similar learners based on user information such as user ratings, *e.g., (Toledo et al., 2012)*.
  No consideration on content relationship.
- Hybrid techniques consider both learning material and user-related information *(Salehi et al., 2013)*
- *(Huang et al., 2009)* proposed a Markov Chain Model to help learners achieve effective web-based learning transfer based on group-learning paths.
  Performance depends on the majority learners' decisions.
- *(Piech et al., 2015)* applied RNN to modeling and predicting learner performance in solving a series of questions and then achieve recommendation by choosing the questions with high predicted performance.
  Prediction is accurate, but difficult to interpret.

Obtaining personalized recommendations

Problems of these methods: **ignore the explainability and doesn't keep human in the loop**. *(Jingyue Guo et al., 2018)*.

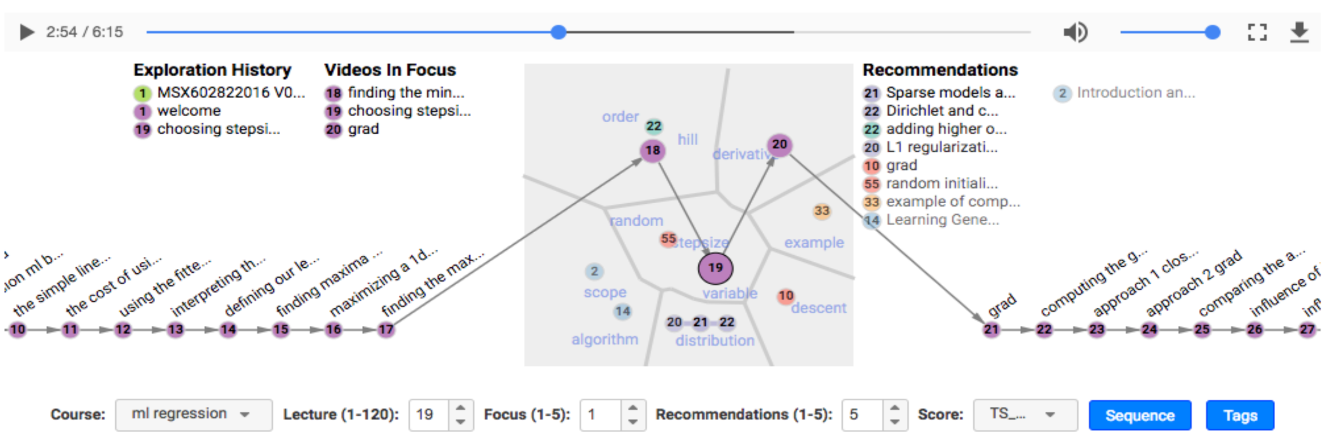
Learning is a different process compared with shopping or watching movies online. Learners can not tell whether the recommendation is good or not in a short time unless they know why they get the recommendation results.

However, few systems inform well learner about why they give such recommendations or give the user the ability to modify the recommendation.

Obtaining personalized recommendations

(Zhao et al., 2018) designed a visual system to assist flexible learning with semantic visual exploration and sequence-based recommendation of MOOC Videos.

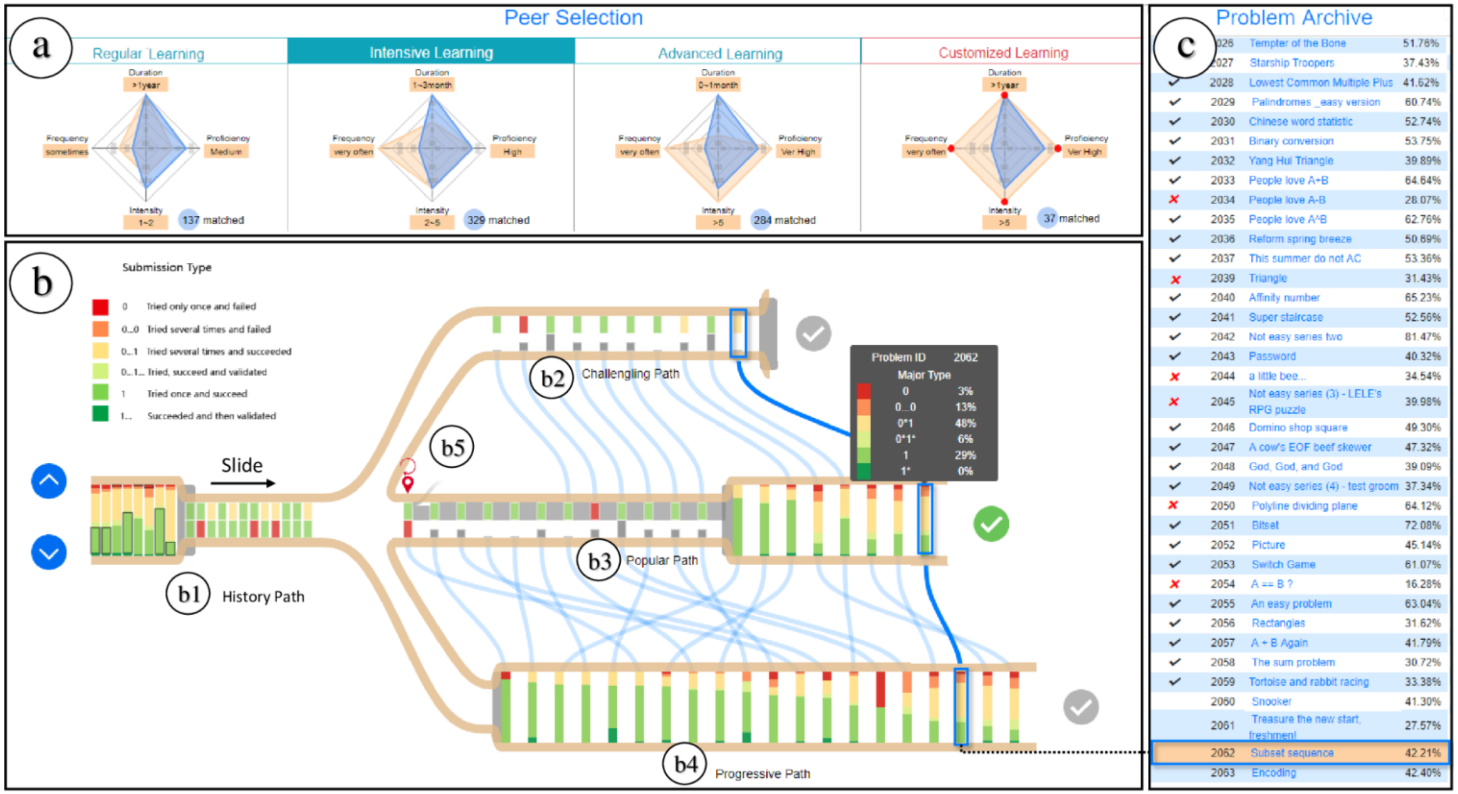
Only for MOOC videos which have predefined sequence and the user's selection will not contribute to the recommendation next time.



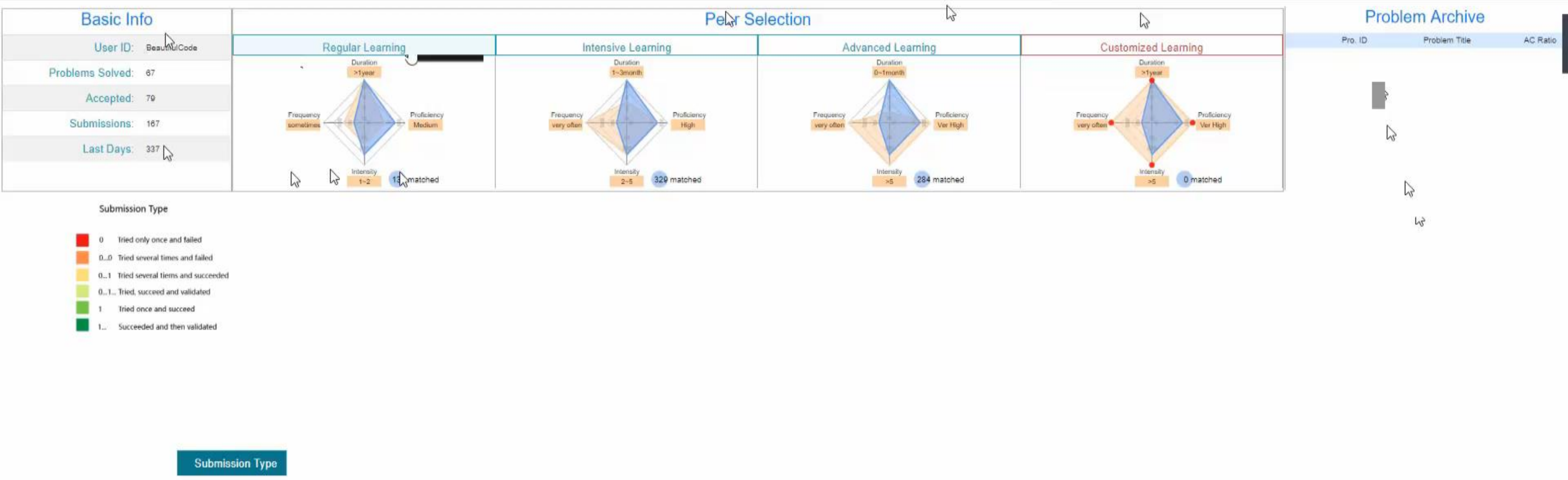
DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Obtaining personalized recommendations

PeerLens: Peer-inspired Interactive Learning Path Planning in Online Question Pool *(conditionally accepted by CHI 2019)*



DEPARTMENT OF



Getting adaptive feedback

(Hieu et al., 2017) gave a classification of data-driven hint generation.

Hint generated method	Problems
Cluster Based Techniques (S. Gross et al., 2014) select the most similar sample solution	no derivation of solution steps from sample solutions
Recommendation Approach (K. Zimmerman and C. R. Rupakheti, 2015) recommends specific code edits according to its closest counterpart in a database of correct solutions	Abstract Syntax Tree based program analysis, semantic similarity of programs and usability testing.
Hint Factory Based Approaches(Thomas et al., 2016) uses a Markov decision process of student problem-solving strategies to serve as a domain model to automatic hint generation	Alternative solutions may not be recognized.

Achieving self-regulation

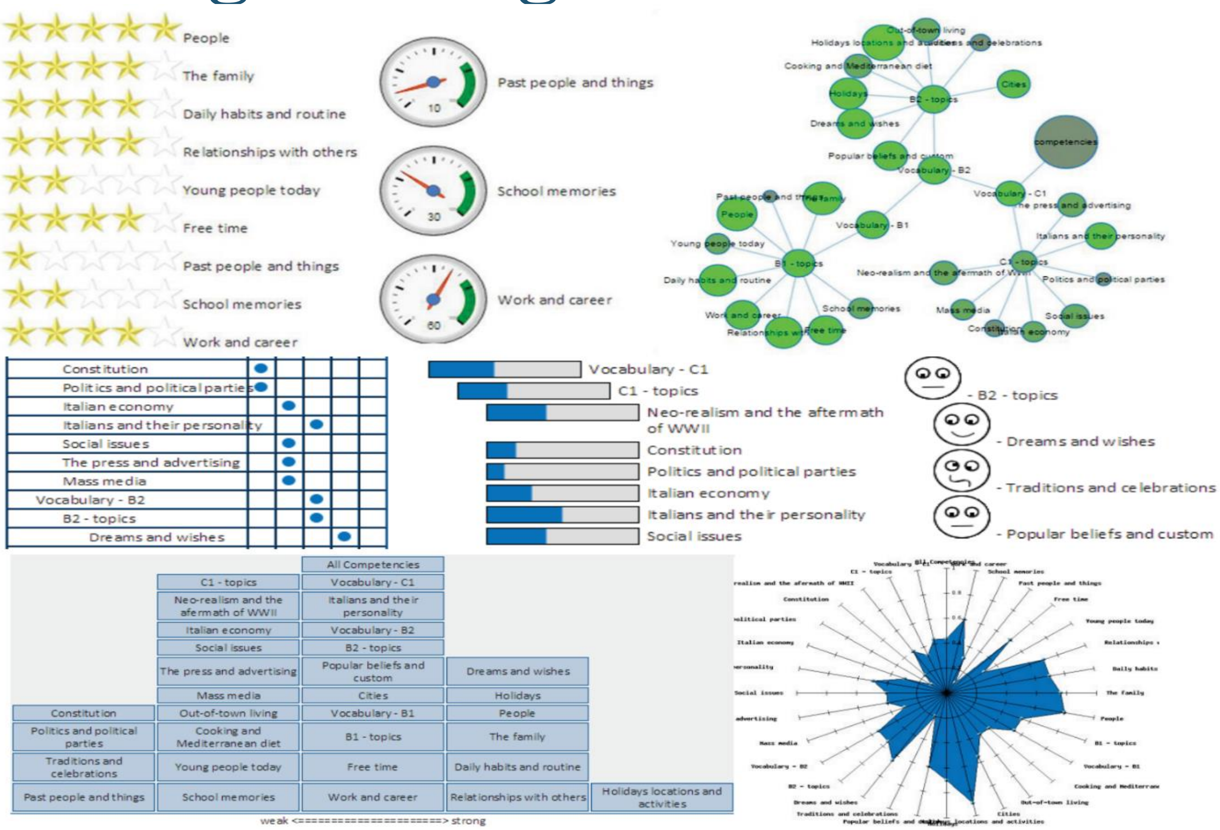
Self-regulated learning refers to the strategies used to **manage learning and regulate mental status** *(Pintrich et al., 1999)*. Generally, three types of strategies are usually used for self-regulation *(Cicchinelli et al., 2018)*: **planning, monitoring, regulating**.

(Daniele et al., 2017) used machine learning models to predict the learners' state of self-regulation.

(Kalle et al., 2018) studied how visualizations could be used to support students' self-regulation in online learning. They found that visualizations lacking appropriate **comparison**s between student performance and others may even be harmful to performance-oriented students.

Analytic requirements

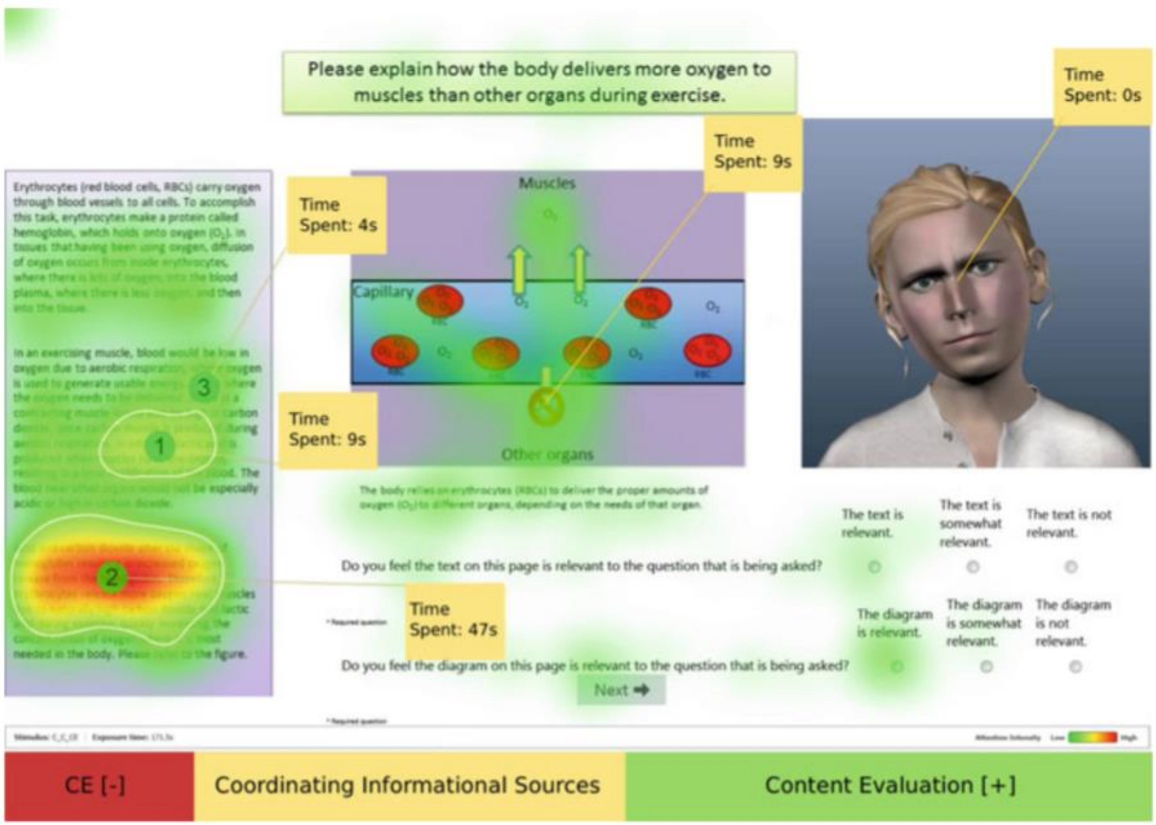
Achieving self-regulation



Open Learner Model (OLM) includes stars, gauges, networks, table, skill meters, smileys, histogram, and radar plot to show the vocabulary learning. *(Bull et al., 2016)*

Analytic requirements

Achieving self-regulation



(Azevedo et al., 2017) visualized the gaze behavior using heatmap, together with text and diagram to show the relationship between gaze and learning content.

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Facilitating study for learners

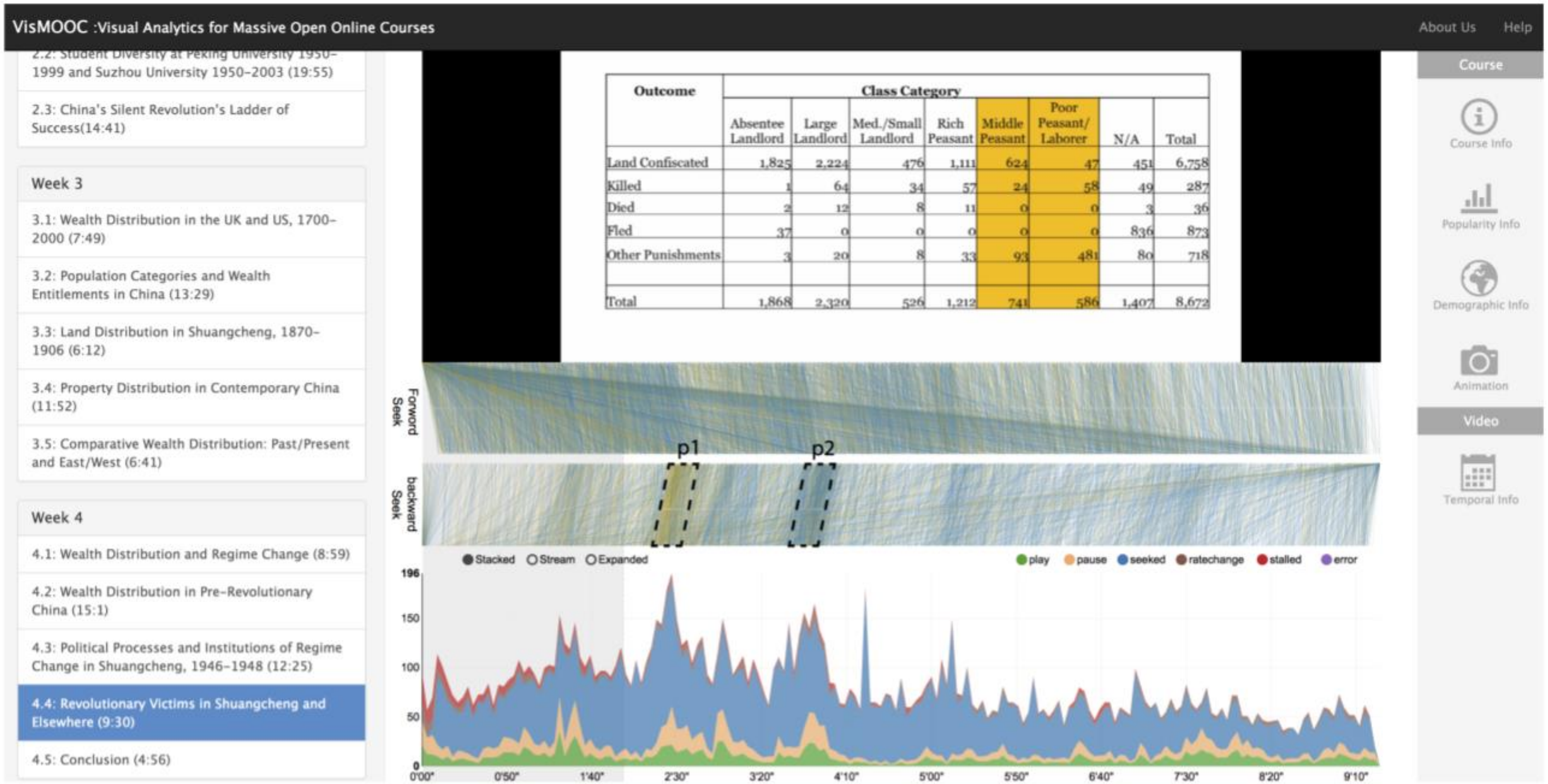
Requirements	Data Mining	Visual Analytics
Obtaining personalized recommendations	*(Chu et al., 2011)* *(Toledo et al., 2012)* *(Salehi et al., 2013)* *(Huang et al., 2009)* *(Piech et al., 2015)*	*(Zhao et al., 2018)* ✓
Getting adaptive feedback	*(Hieu et al., 2017)* *(R. Singh, 2015)* *(K. Zimmerman and C. R. Rupakheti, 2015)* *(P. Freeman et al., 2016)* *(Thomas et al., 2016)*	?
Achieving self-regulation	*(Daniele et al., 2017)*	*(Cicchinelli et al., 2018)* *(Bull et al., 2016)* *(Azevedo et al., 2017)*

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Assisting teaching for instructors

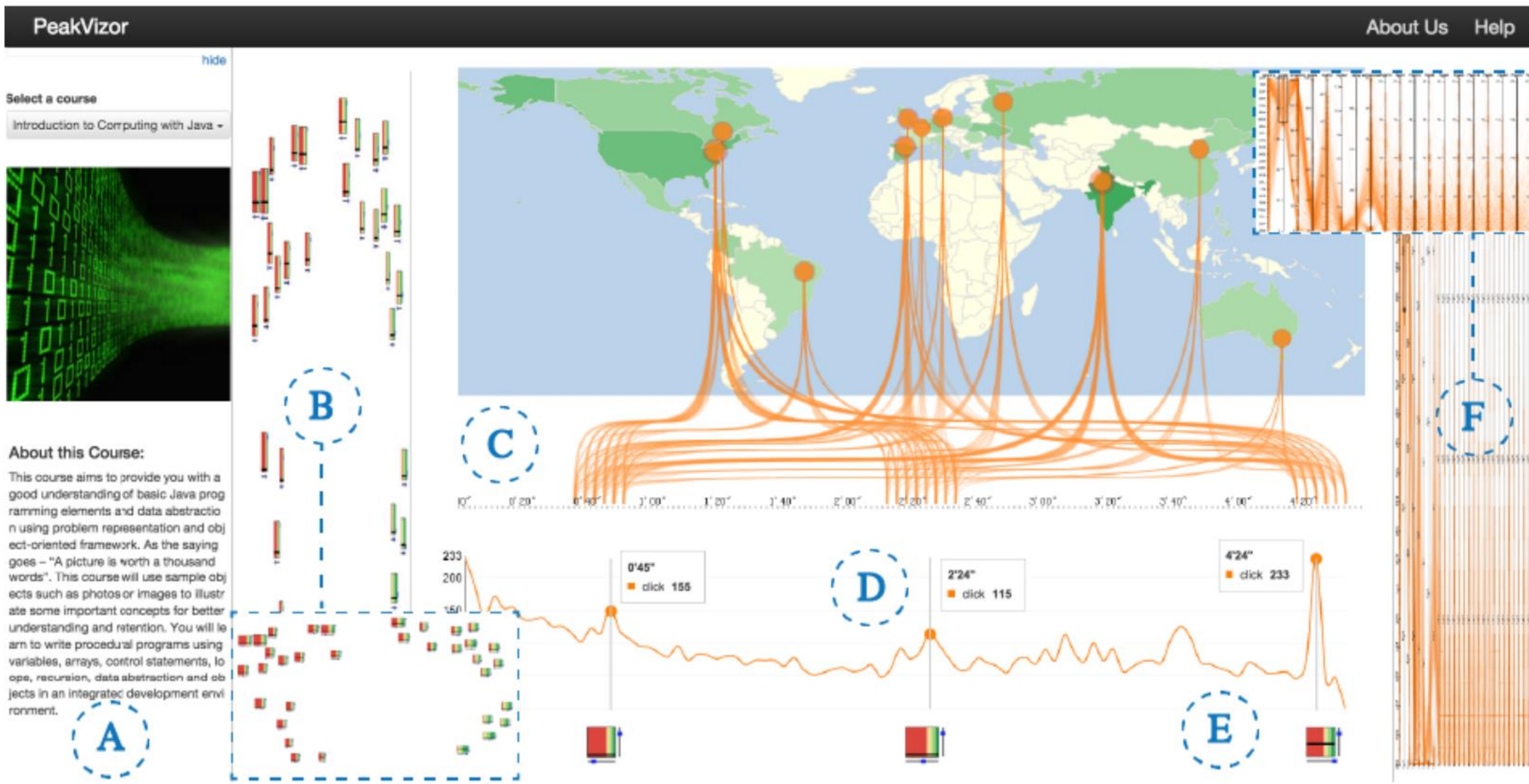
Requirements	Data Mining	Visual Analytics
Understanding learning material usage		*(Shi et al., 2015)* *(Chen et al., 2016)* *(He et al., 2018)*
Identifying learning behavior patterns	*(Hassan et al., 2017)* *(Chui et al., 2017)* *(Wang et al., 2016)*	*(Coffrin et al., 2014)* *(Chen et al., 2018)* *(Du et al, 2016)*
Monitoring learning progress, engagement, collaboration		*(Fatima et al., 2016)* *(Fonseca et al., 2016)* *(Sharma et al., 2016)* *(Wang et al., 2017)*

Understanding learning material usage



A screenshot of VisMOOC *(Shi et al., 2015)*

Understanding learning material usage



PeakVizor system. *(Chen et al., 2016)*

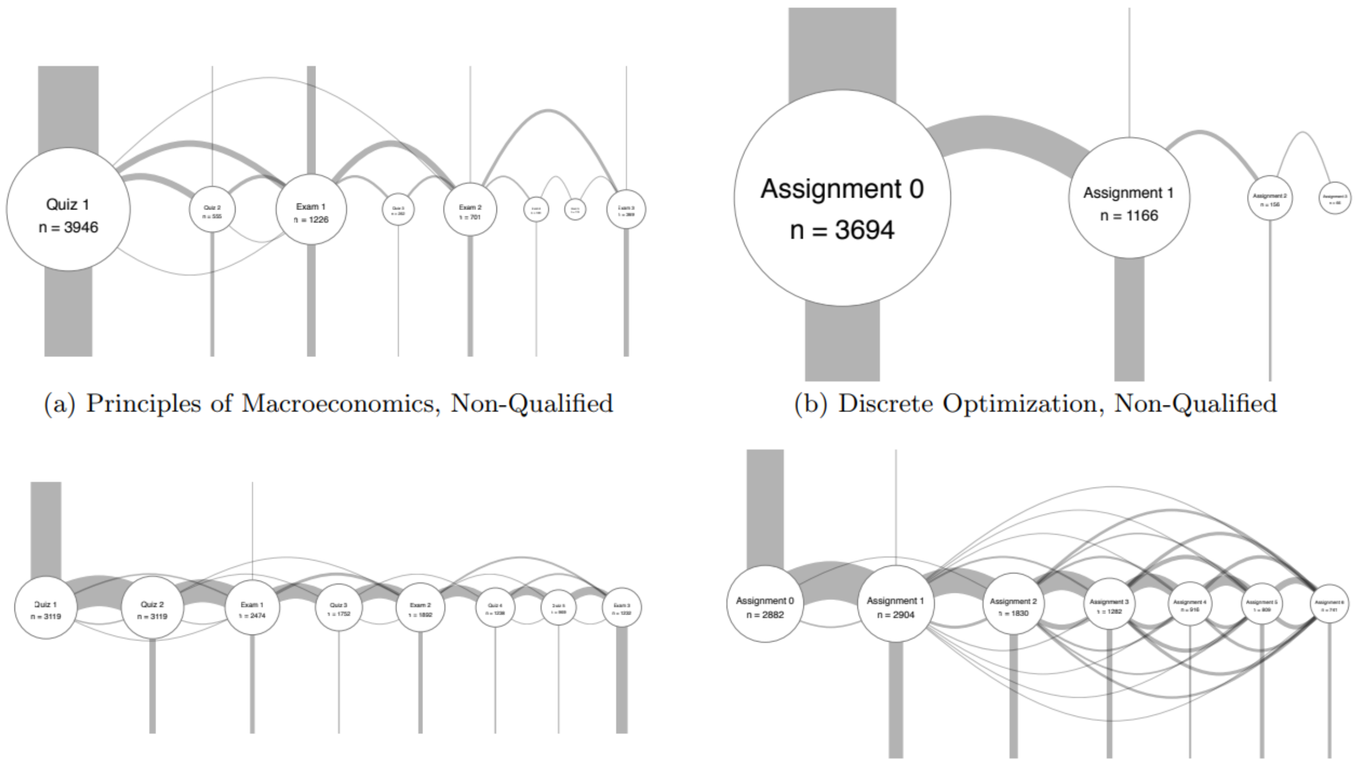
DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Identifying learning behavior patterns

Grouping students only from their performance is inappropriate.

- *(Hassan et al., 2017)* grouped students using k-means and found different clusters of **engagement and performance**.
- *(Chui et al., 2017)* proposed that the impact of tracking on student achievement differs across levels (across schools vs. within a school), across types of classmate resources (**past achievements, reading attitudes, family socioeconomic status**) and depends on a student's academic ability.
- *(Wang et al., 2016)* argued for a combination of features that better define the learning process: **initial accuracy, feedback usage, and attempts required for success**.

Identifying learning behavior patterns



(a) Principles of Macroeconomics, Non-Qualified

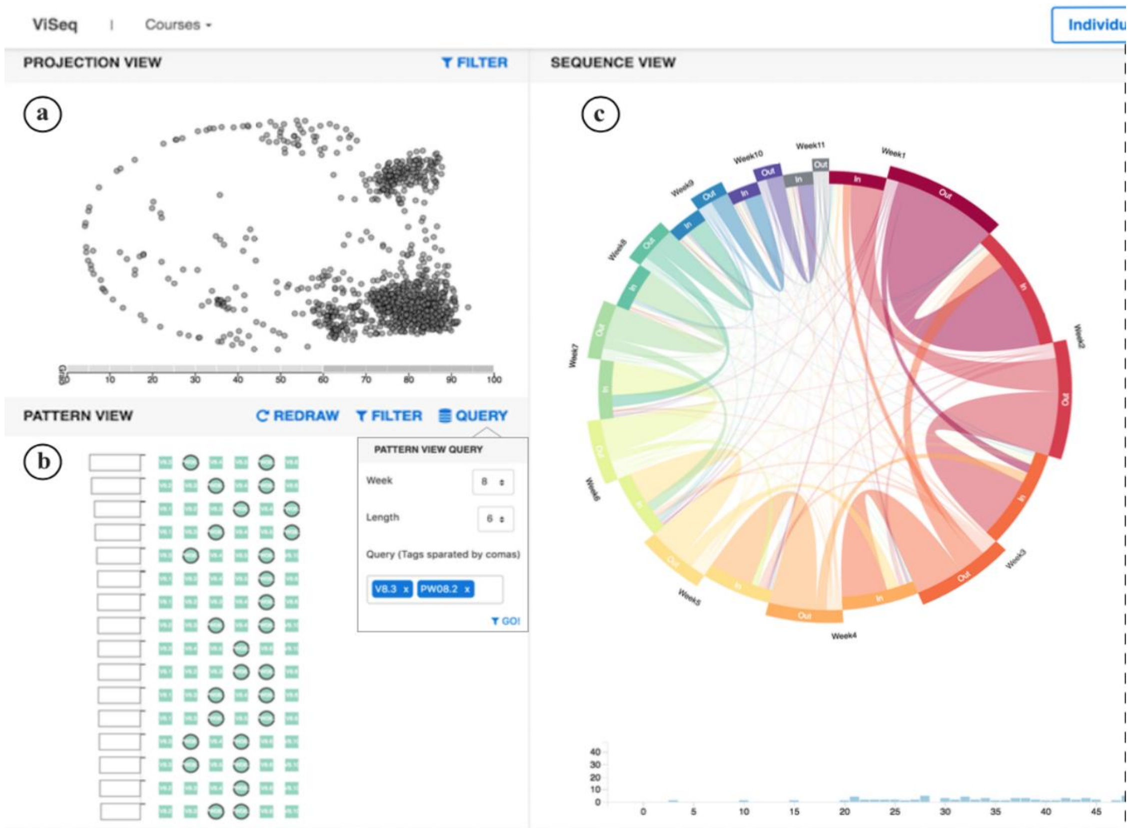
(b) Discrete Optimization, Non-Qualified

(c) Principles of Macroeconomics, Qualified

(d) Discrete Optimization, Qualified

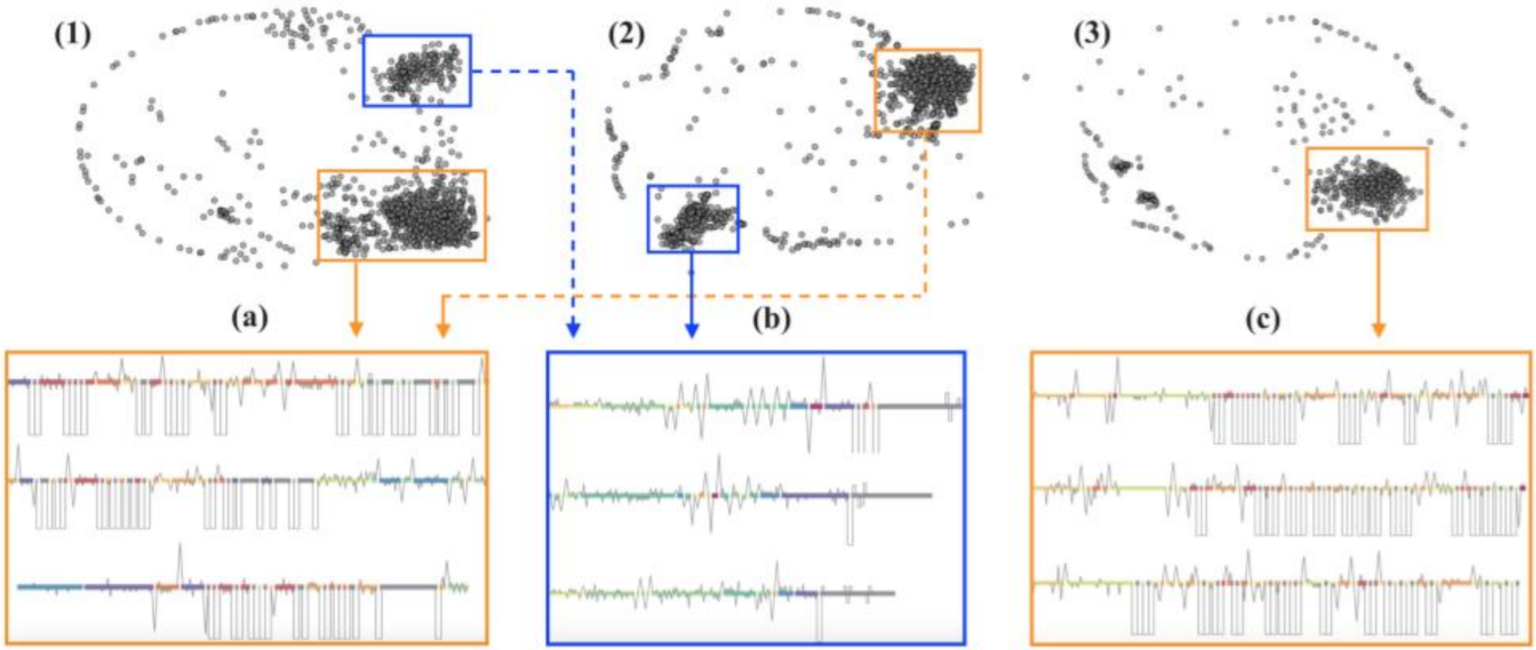
Student Assignment Submissions Transitions.*(Coffrin et al., 2014)*

Identifying learning behavior patterns



ViSeq system. *(Chen et al., 2018)*

Identifying learning behavior patterns



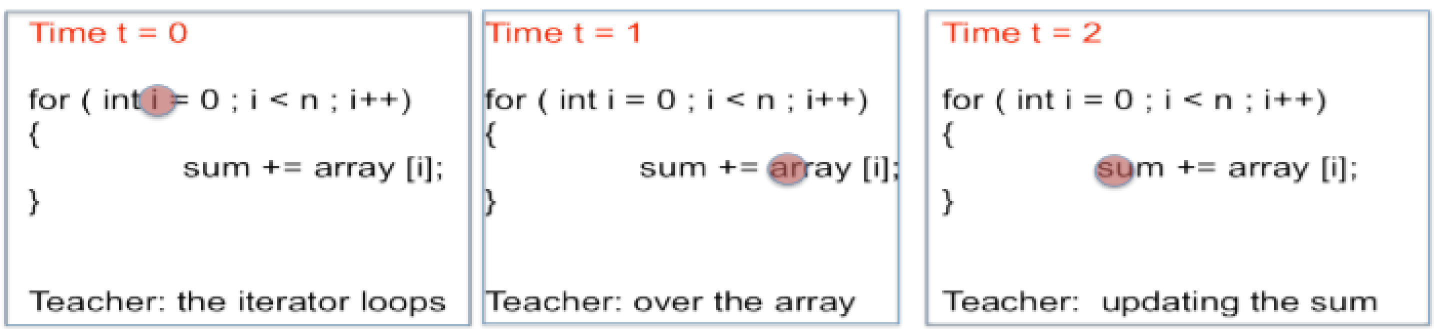
Three projection views (1) (3) from the three MOOCs are presented with different groups. *(Chen et al., 2018)*

Analytic requirements

Monitoring learning progress, engagement, collaboration

(Fatima et al., 2016) showed how to create the Problem Solving Markov Model (PSMM), **a graphical representation of all attempted solutions** proposed by the class.

(Fonseca et al., 2016) designed a dashboard for teachers to monitor students programming process by **code snapshots**.



Gaze visualization. *(Sharma et al., 2016)*

Assisting teaching for instructors

Requirements	Data Mining	Visual Analytics
Understanding learning material usage		*(Shi et al., 2015)* *(Chen et al., 2016)* *(He et al., 2018)*
Identifying learning behavior patterns	*(Hassan et al., 2017)* *(Chui et al., 2017)* *(Wang et al., 2016)*	*(Coffrin et al., 2014)* *(Chen et al., 2018)* *(Du et al, 2016)* ?
Monitoring learning progress, engagement, collaboration		*(Fatima et al., 2016)* *(Fonseca et al., 2016)* *(Sharma et al., 2016)* *(Wang et al., 2017)*

Behavior analysis for Researchers

Requirements	Data Mining	Visual Analytics
Understanding motivation	*(José et al., 2016)* *(Chen et al., 2017)*	
Learning behavior modeling	*(Corbett et al., 1994)* *(Cen et al., 2006)* *(Pavlik et al., 2009)* *(Piech et al., 2015a)*	
Predicting learners' performance	*(Alireza et al., 2015)* *(Okubo et al., 2017)* *(Christopher et al, 2015)*	
Social network analysis	*(Cui and Wise, 2015)* *(Gillani et al., 2014)*	*(Fu et al., 2017)* *(Fu et al., 2018)*

Understanding motivation

(Zheng et al., 2015)	fulfilling current needs, preparing for the future, satisfying curiosity, and connecting with people	clickstream
(José et al., 2016)	earning different badges	badges
(Chen et al., 2017)	improving grades, refreshing theoretical understanding, and solving practical problems	clickstream

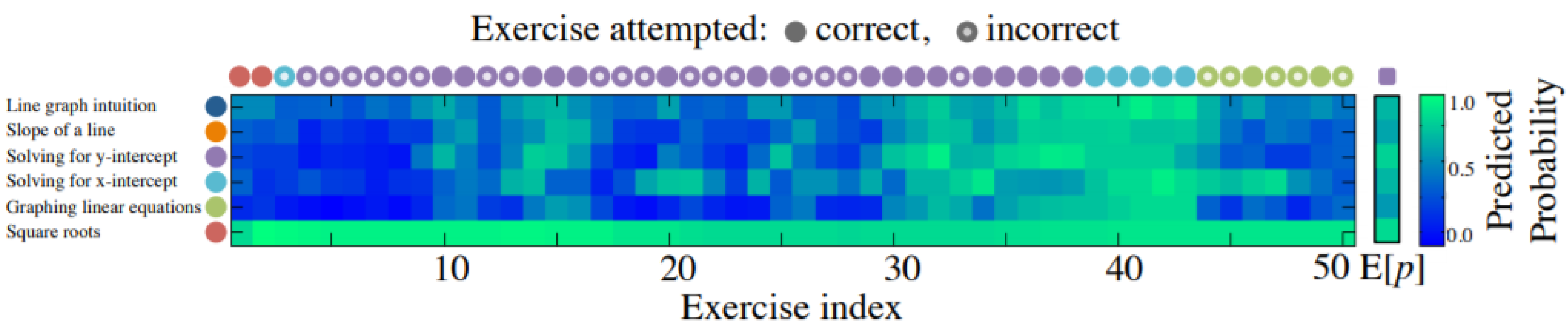
Learning behavior modeling

Predicting learners' performance

(Corbett et al., 1994)	Bayesian knowledge tracing	assumes that the concepts are independent
(Cen et al., 2006) (Pavlik et al., 2009)	Dynamic Probabilistic Models(Learning Factors Analysis, Performance Factors Analysis)	require accurate concept labelling
(Piech et al., 2015a)	Deep knowledge tracing model(RNN)	difficult to interpret

Predicting learners' performance

An interesting case to show the role of visualization:



A single student and her predicted responses as she solves 50 Khan Academy exercises. She seems to master finding x and y intercepts and then has trouble transferring knowledge to graphing linear equations. *(Piech et al., 2015a)*

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

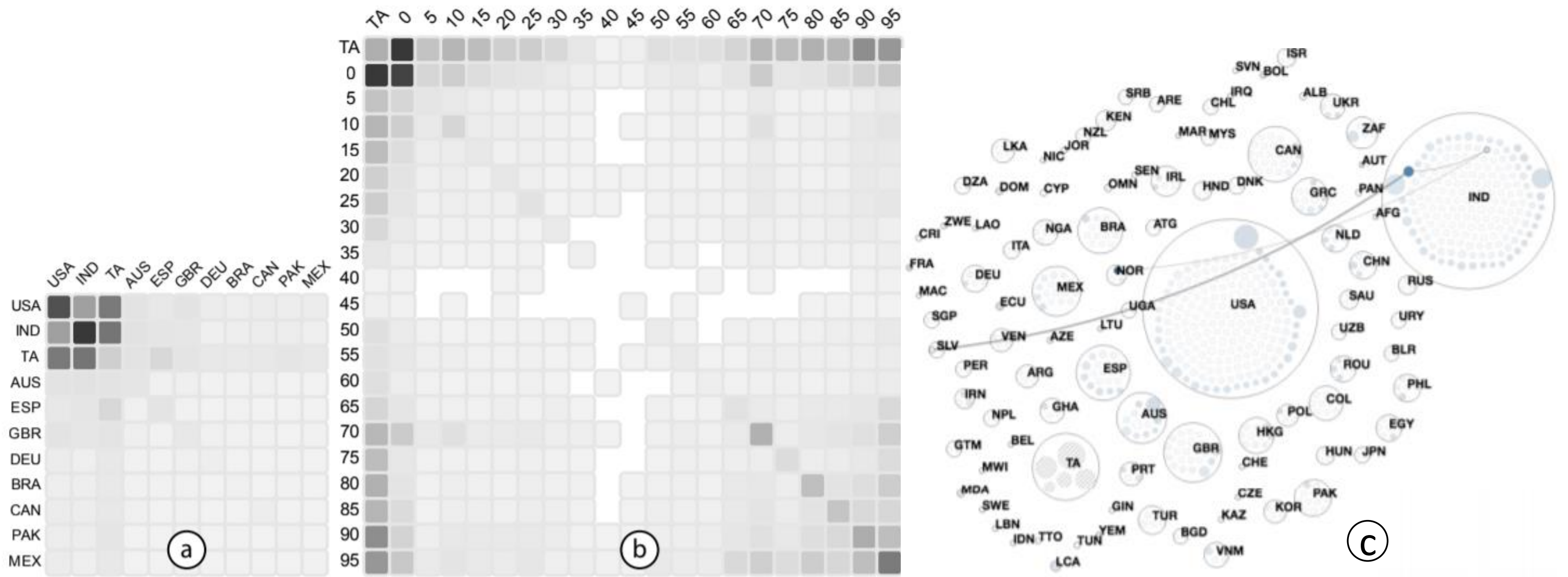
Social network analysis

Researches analyzed the forum data from two perspectives: the **communication related** and **the content related**.

- *(Cui and Wise, 2015)* apply content analysis and machine learning to identify forum threads where participants discuss the course content, which is essential for the investigation of information exchange.
- *(Gillani et al., 2014)* argued that the coherence of the social structure mainly depends on a small set of central users and the forum users can be considered as a loosely connected crowd rather than a strongly connected learning community.

However, the analysis process is difficult to understand without figure illustrations and detailed information is difficult to describe.

Social network analysis



Using iForum to explore the MOOC forum of a JAVA programming course.

(Fu et al., 2017)

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Behavior analysis for Researchers

Requirements	Data Mining	Visual Analytics
Understanding motivation	*(José et al., 2016)* *(Chen et al., 2017)*	
Learning behavior modeling	*(Corbett et al., 1994)* *(Cen et al., 2006)* *(Pavlik et al., 2009)* *(Piech et al., 2015a)*	?
Predicting learners' performance	*(Alireza et al., 2015)* *(Okubo et al., 2017)* *(Christopher et al, 2015)*	
Social network analysis	*(Cui and Wise, 2015)* *(Gillani et al., 2014)*	*(Fu et al., 2017)* *(Fu et al., 2018)*

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Platform management for administrators

Requirements	Data Mining	Visual Analytics
Retaining learners or avoiding dropout	*(Sinha et al., 2014)* *(Crossley et al., 2016)* *(Sharkey et al., 2014)* *(Boyer et al., 2015)*	*(Chen et al., 2016)*
Anomaly detection and volume control		

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

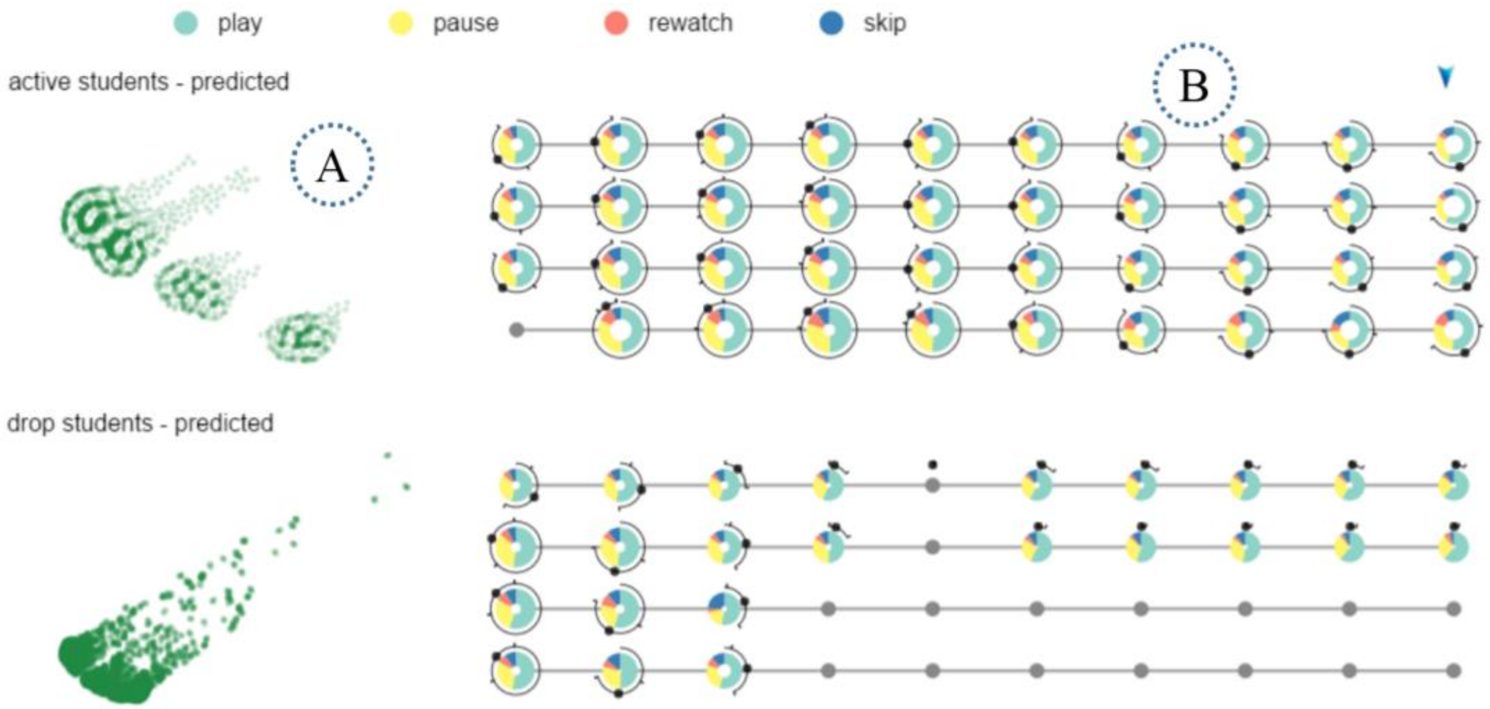
Retaining learners or avoiding dropout

(Sinha et al., 2014) formed a directed graph to represent the sequential structure and extracted features based on several typical short activity sequences rather than single activities.

With the extracted features, many classical machine learning models have been tested, including Support Vector Machine (SVM) *(Crossley et al., 2016)*, Decision Tree *(Sharkey et al., 2014)*, and Logistic Regression *(Boyer et al., 2015)*.

Parameters of the algorithm are different for each course.

Retaining learners or avoiding dropout

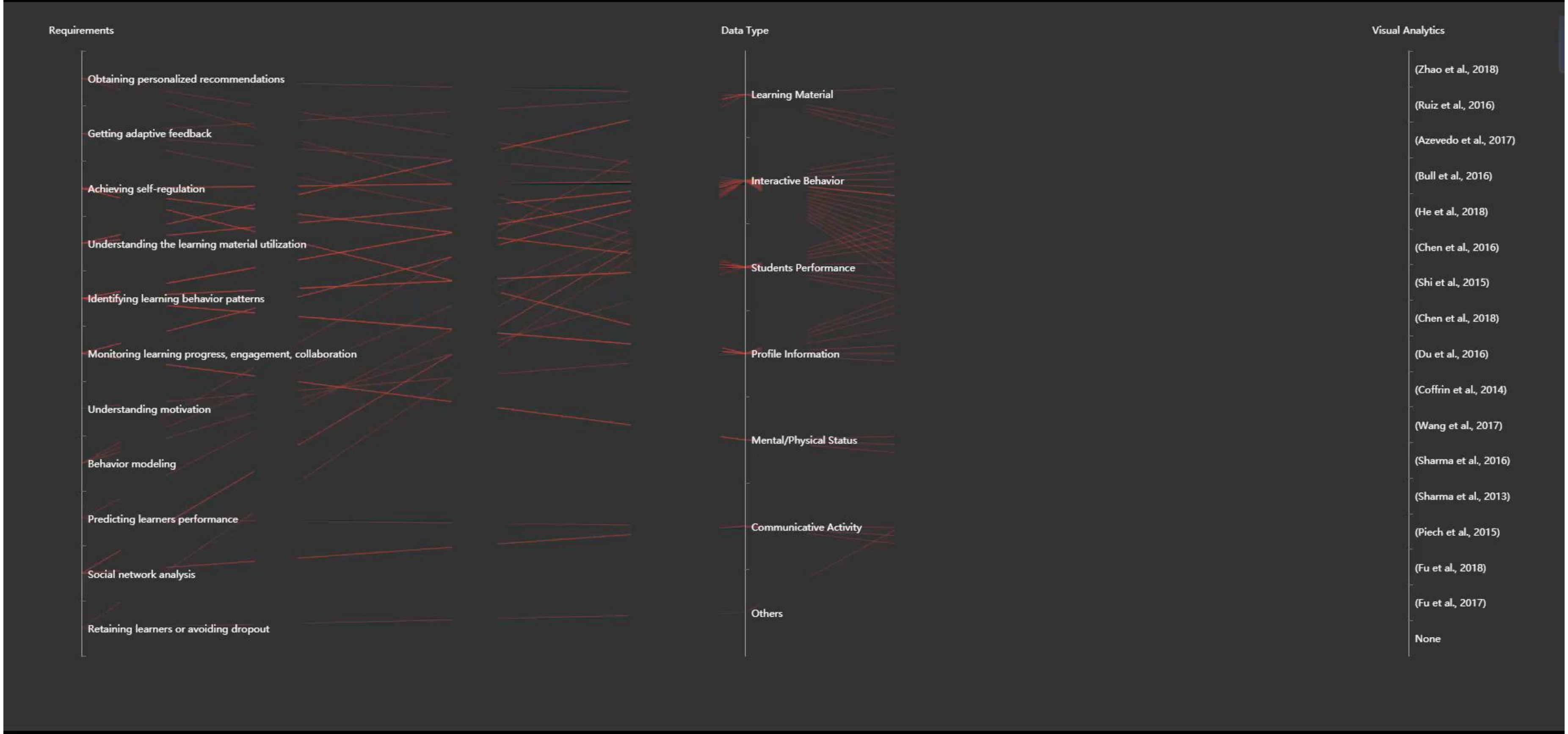


DropoutSeer *(Chen et al., 2016)*

Platform management for administrators

Requirements	Data Mining	Visual Analytics
Retaining learners or avoiding dropout	*(Sinha et al., 2014)* *(Crossley et al., 2016)* *(Sharkey et al., 2014)* *(Boyer et al., 2015)*	*(Chen et al., 2016)* ?
Anomaly detection and volume control		

Analytic requirements



Outline

Introduction

Data types in online learning

Visual analytics for different stakeholders

Conclusion and future work

Conclusion

- Introduce the background of online learning and the motivation of analysis for online learning data and point out the reasons and challenges using visual analytics.
- List all kinds of online learning data types that have been or can be used to do learning analysis.
- Discuss the analytical requirements of online data for four different stakeholders are.
- Introduce the state-of-the-art visual analytic systems that aim to fulfill the user-oriented requirements.

Future Work

- For learners, instructors, and administrators, we should design appropriate visual analytics to help them understand the recommendation, prediction as well as retention/drop out. Especially, we need to close the loop to do more field studies to see which level of complexity and which kinds of visual designs are preferred in the real world.

- For researchers, more visual analytic technologies should be developed to help them build the learning behavior model as well as the prediction model.

Thank you!