ML for Education NIPS Workshop
12/10/2016

ML Approaches for Learning Analytics: Collaborative Filtering Or Regression With Experts?

Kangwook Lee

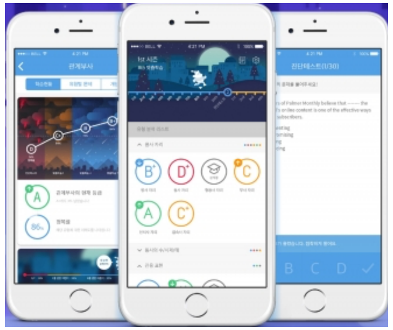


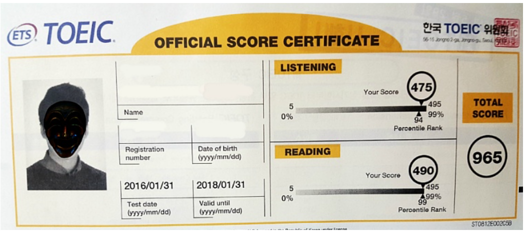
Joint work w/ Jichan Chung, Yeongmin Cha, and Changho Suh

KAIST Riiid!

Learning Analytics

Data Collection → Data Analysis → Optimize Learning



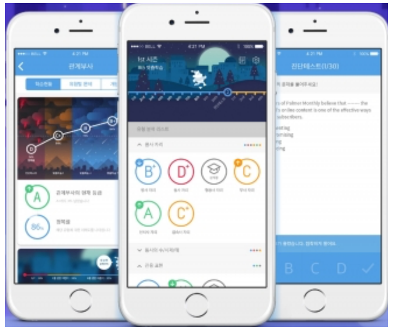


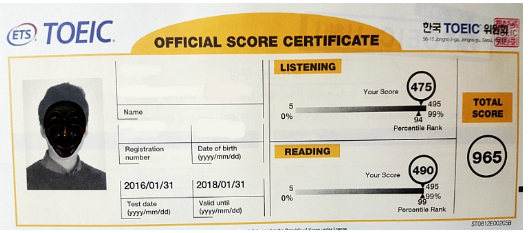
- Rule-based
- Machine Learning

- Prediction
- Recommendation
- Personalization
- Content Design
- …

Learning Analytics

Data Collection → Data Analysis → Optimize Learning

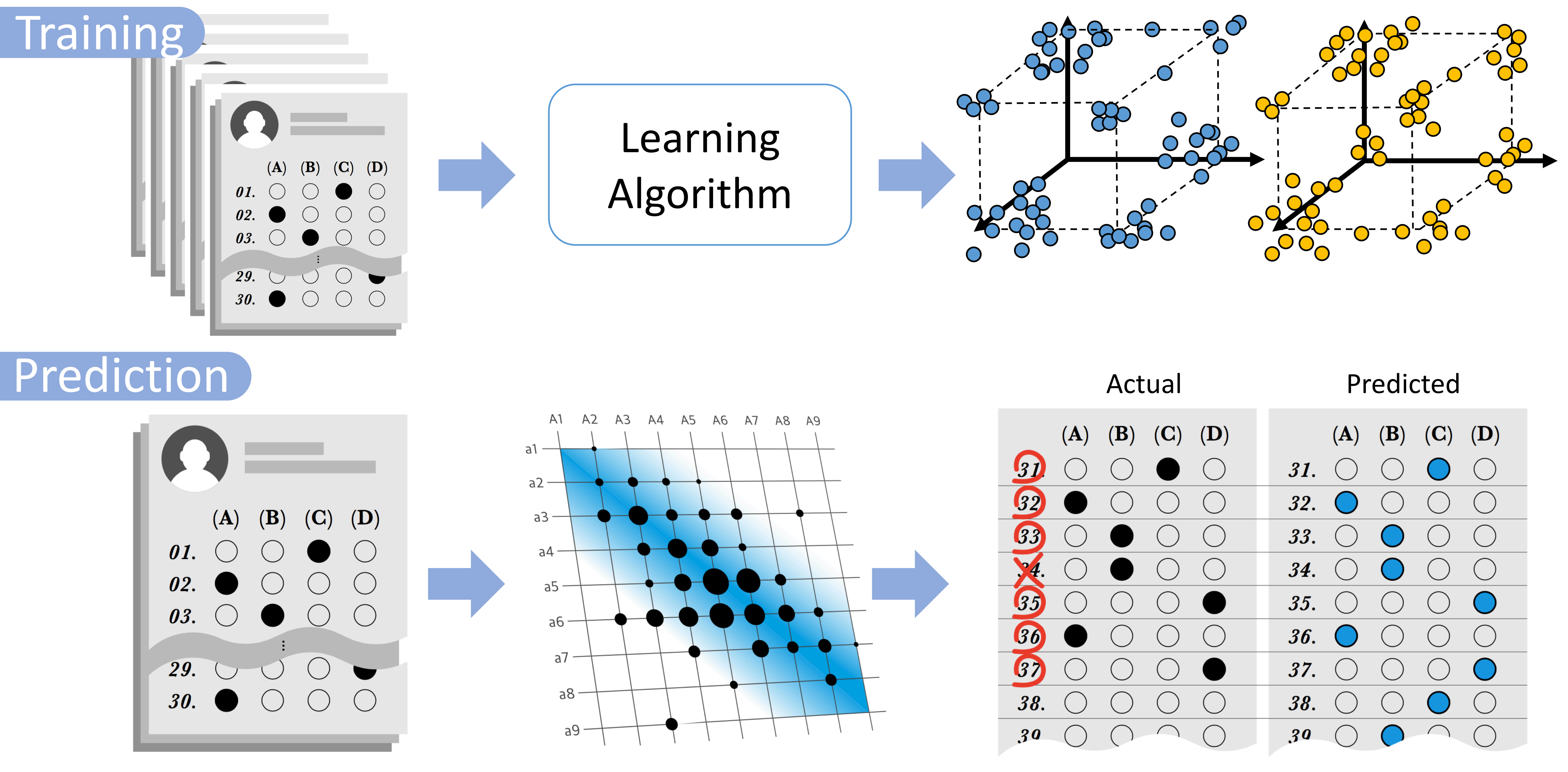



TOEIC
OFFICIAL SCORE CERTIFICATE
LISTENING
Your Score 475
495
99%
Percentile Rank
TOTAL SCORE
965
READING
Your Score 490
495
99%
Percentile Rank
Name
Registration number
Date of birth (yyyy/mm/dd)
2016/01/31
Test date (yyyy/mm/dd)
2018/01/31
Valid until (yyyy/mm/dd)

- Rule-based
- **Machine Learning**

- **Prediction**
- Recommendation
- Personalization
- Content Design
- …

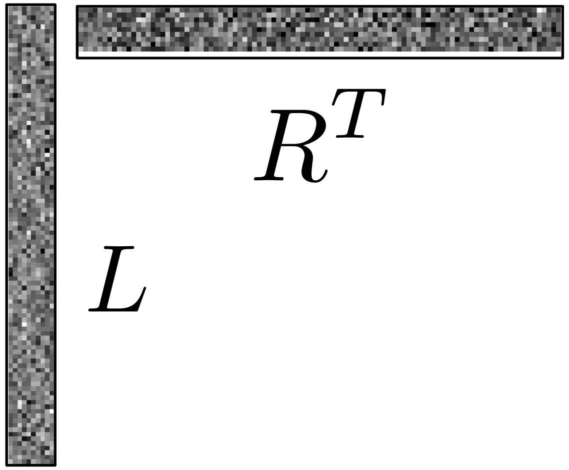
Prediction

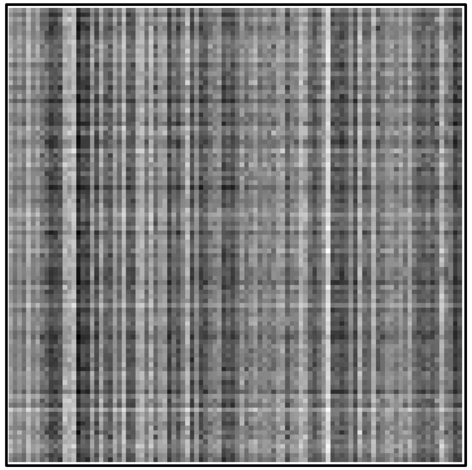


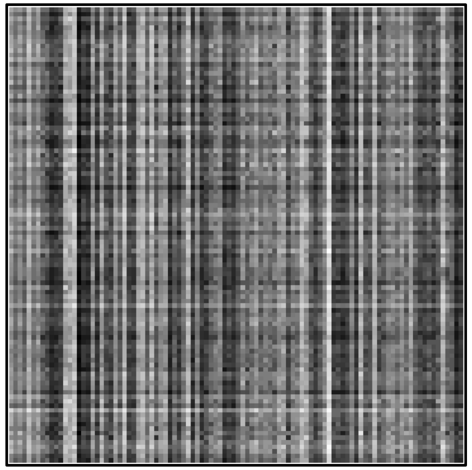
Training
Learning Algorithm
Student Features
Question Features
Prediction
A1 A2 A3 A4 A5 A6 A7 A8 A9
a1 a2 a3 a4 a5 a6 a7 a8 a9
Actual
Predicted
(A) (B) (C) (D)
01. 02. 03. 29. 30.
31. 32. 33. 34. 35. 36. 37. 38.

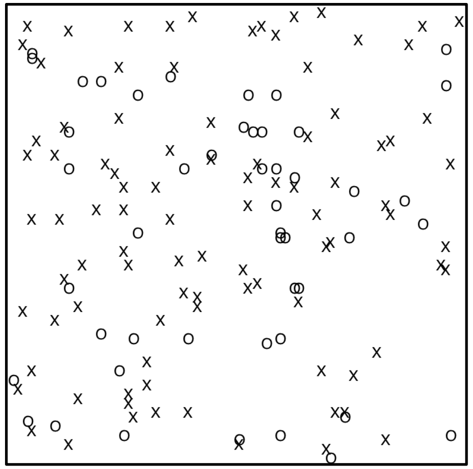
Response Model & Learning

[Bergner et al., 2012], [Lan, Studer, Baraniuk, 2014]









$$L \in [0,1]^{n \times r}$$
$$R \in [0,1]^{m \times r}$$

Student & Question features

$$X = L \times R^T$$

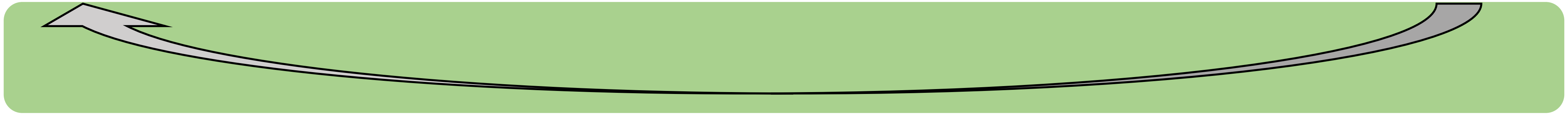
Level of understanding

$$P = \phi(X)$$

Probability of correct guess

$$Y$$

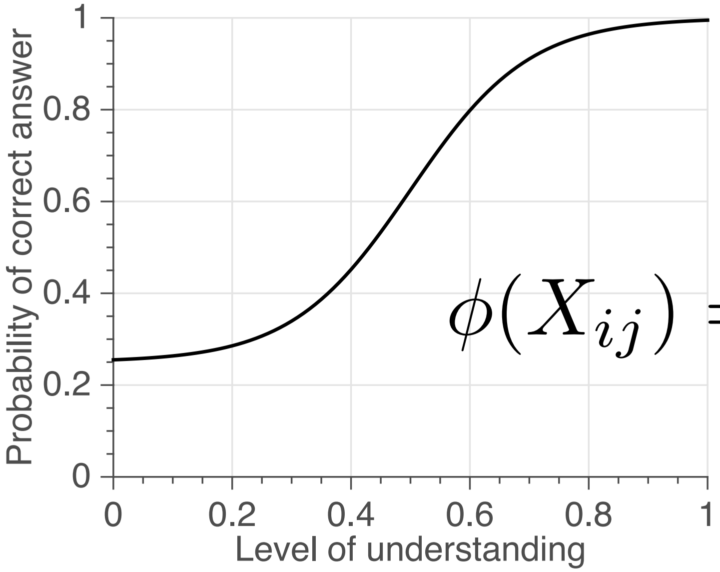
Responses



Learning Algorithm

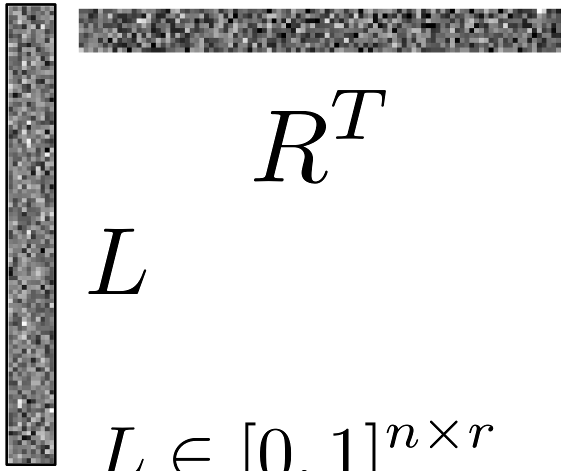
Response Model

- A variation of M2PL (Multidimensional Two-Parameter Logistic) model
- $L_{i,j}$: the level of student i's understanding of the j^{th} hidden concept.
- $R_{i,j}$: the contribution of the j^{th} hidden concept to question i
- R is normalized to sum up 1 so that $X_{i,j} = L_i R_j^T$ is in [0,1]
- Two additional concepts for difficulty & outliers:
  - $(r+1)^{th}$ concept for what is known to everyone
  - $(r+2)^{th}$ concept for what is not known to everyone (e.g., difficult vocab)
- P = Logistic(X)
- Y = Bernoulli(P)

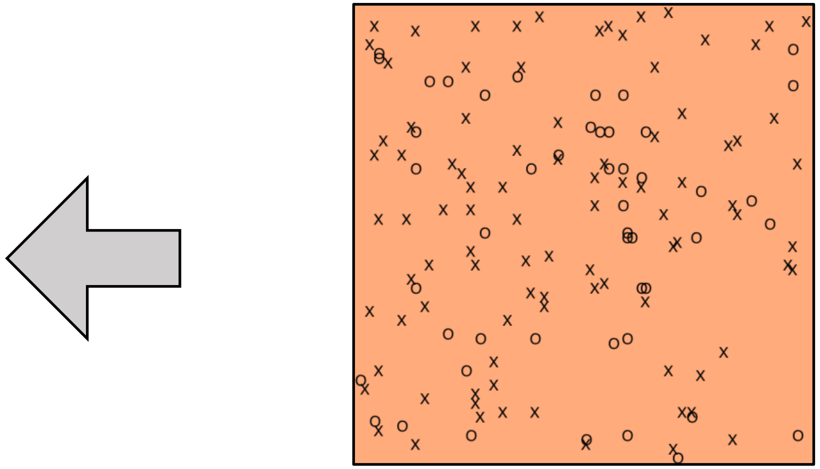


$$\phi(X_{ij}) = \phi_a + \frac{1-\phi_a}{1+e^{-\phi_c(X_{ij}-\phi_b)}}$$

Logistic Regression w/ Experts



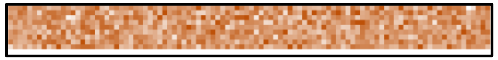
$$L \in [0,1]^{n \times r}$$
$$R \in [0,1]^{m \times r}$$



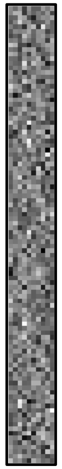
$$Y = Bern(\phi(LR^T))$$

Logistic Regression w/ Experts

If experts can provide us w/ R,

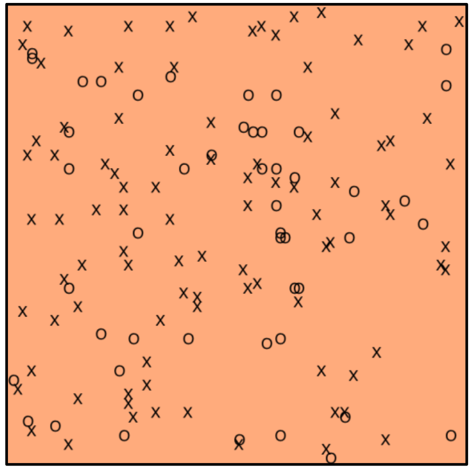


R^T



L

$L \in [0,1]^{n \times r}$
$R \in [0,1]^{m \times r}$



$Y = Bern(\phi(LR^T))$

The MLE of L is

$$\min_{L_i} \sum_{j \in \Omega_{i\star}} [-Y_{ij} \log(P_{ij}) - (1 - Y_{ij}) \log(1 - P_{ij})]$$

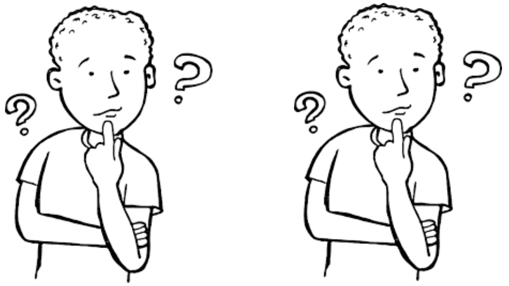
$$\text{s.t.} \quad 0 \le L_{ij} \le 1, \; \sum_j L_{ij} = 1, \; P_{ij} = L_i R_j^T.$$

Logistic Regression w/ Experts

Question j

_____ who want to apply for this position are requested to submit their performance.

(A) You
(B) Those
(C) Another
(D) Some



- Experts
- Crowdsourcing

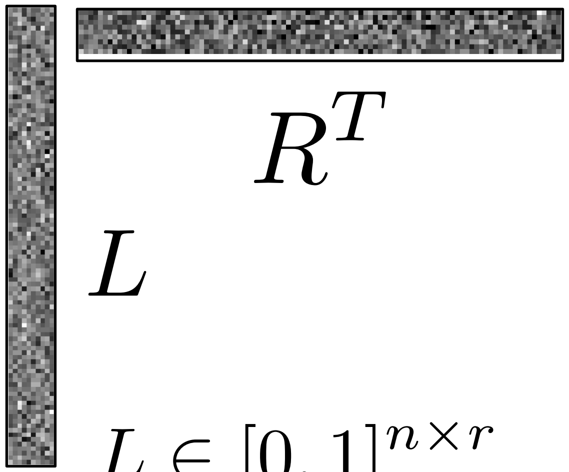
"Pronoun" "Personal pronoun" "Long subjects (?)" "Prepositional phrase"

R_j

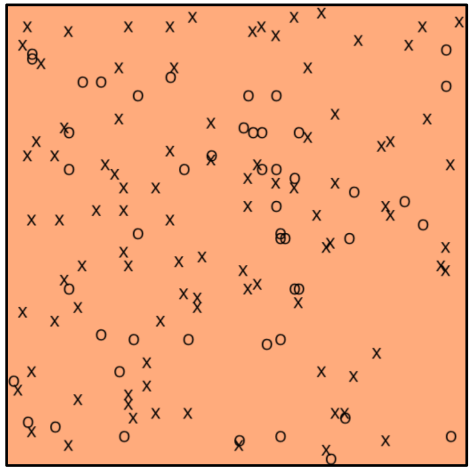
0	.5	.3	0	0	…	0	.1	.1	0

- Noisy, subjective, …
- (Observation) # of concepts is usually very large (prone to overfitting)
- Depends on human knowledge

Binary Matrix Completion (BMC)



$L \in [0,1]^{n\times r}$
$R \in [0,1]^{m\times r}$



$Y = Bern(\phi(LR^T))$

Estimate L and R by solving

$$\min_{L,R} \sum_{(i,j)\in\Omega} [-Y_{ij}\log(P_{ij}) - (1-Y_{ij})\log(1-P_{ij})] + \mu\|LR^T\|_*$$

$$\text{s.t.}\quad 0 \le L_{ij} \le 1,\ 0 \le R_{ij} \le 1,\ P = LR^T,\ \sum_j L_{ij} = 1,\ \forall i.$$

Algorithm for BMC

[Recht, Re´, 2013]

$$\min_{L,R} \sum_{(i,j)\in\Omega} [-Y_{ij}\log(P_{ij}) - (1-Y_{ij})\log(1-P_{ij})] + \mu\|LR^T\|_*$$

$$\text{s.t.} \quad 0 \le L_{ij} \le 1,\ 0 \le R_{ij} \le 1,\ P = LR^T,\ \sum_j L_{ij} = 1,\ \forall i.$$

Approximation

$$\|X\|_* = \min_{X=LR^T} \frac{1}{2}\left(\|L\|_F^2 + \|R\|_F^2\right)$$

$$\min_{L,R} \sum_{(i,j)\in\Omega} [-Y_{ij}\log(P_{ij}) - (1-Y_{ij})\log(1-P_{ij})] + \frac{\mu}{2}\left(\|L\|_F^2 + \|R\|_F^2\right)$$

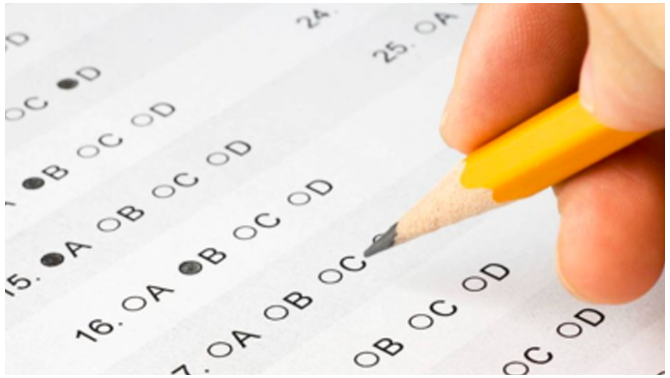
$$\text{s.t.} \quad 0 \le L_{ij} \le 1,\ 0 \le R_{ij} \le 1,\ P = LR^T,\ \sum_j L_{ij} = 1,\ \forall i.$$

Projected SGD

$$L_{i_k}^{(k+1)} = \Pi_{P_L}\left(\left(1 - \frac{\mu_1\alpha_k}{|\Omega_{i_k\star}|}\right)L_{i_k}^{(t)} - \alpha_k \frac{\phi_c\left(Y_{i_kj_k} - \phi(L_{i_k}R_{j_k}^T)\right)}{\phi(L_{i_k}R_{j_k}^T)\left(1+e^{-\phi_c(L_{i_k}R_{j_k}^T - \phi_b)}\right)} R_{j_k}^{(t)}\right),$$

$$R_{j_k}^{(k+1)} = \Pi_{P_R}\left(\left(1 - \frac{\mu_1\alpha_k}{|\Omega_{\star j_k}|}\right)R_{j_k}^{(t)} - \alpha_k \frac{\phi_c\left(Y_{i_kj_k} - \phi(L_{i_k}R_{j_k}^T)\right)}{\phi(L_{i_k}R_{j_k}^T)\left(1+e^{-\phi_c(L_{i_k}R_{j_k}^T - \phi_b)}\right)} L_{i_k}^{(t)}\right)$$

Experiments: Data Set



TOEIC (Test Of English for International Communication)

-A test with 150 multiple-choice questions

-7 parts

- Part 5, Part 6

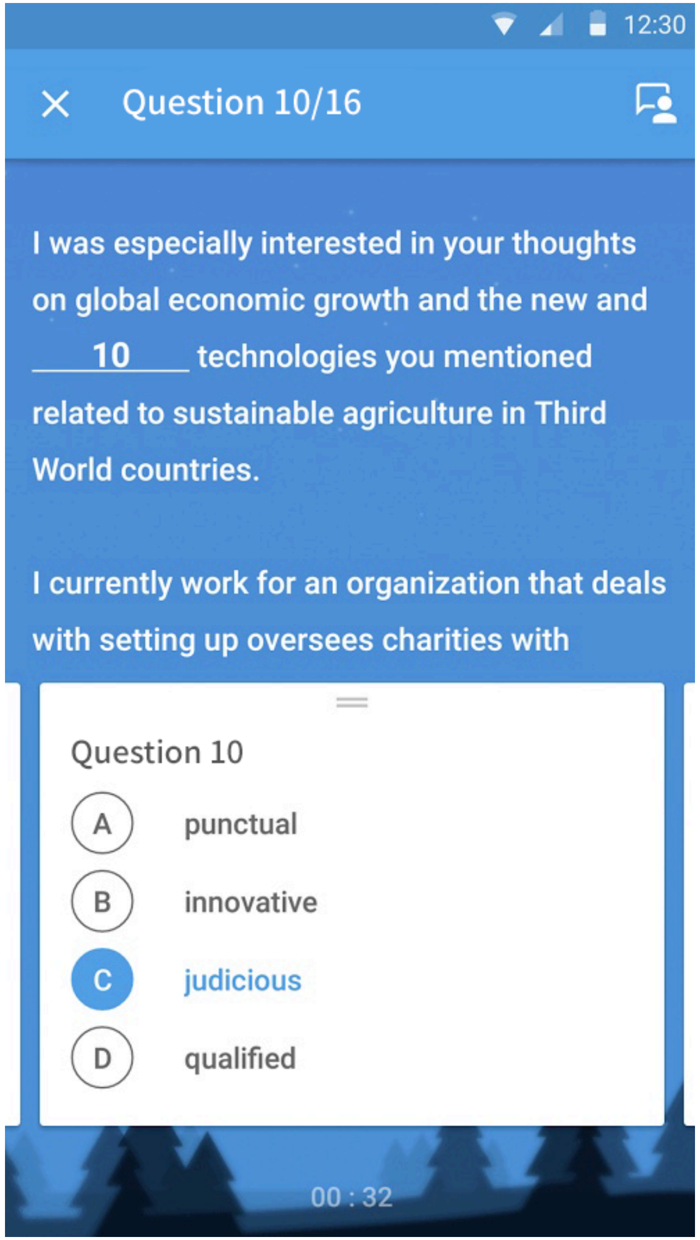
Our office security door is scheduled to _____ this week so all staff members are required to return their security cards to the front desk.

(A) replace
(B) replaced
(C) being replaced
(D) be replaced

Seasons Greetings. As a _____ customer, we wanted you to be among the first to know about our upcoming holiday sale. All craft paper, specialty printer paper, and decorative envelopes will be reduced by 50% for the month of December.

(A) value
(B) valued
(C) valid
(D) validate

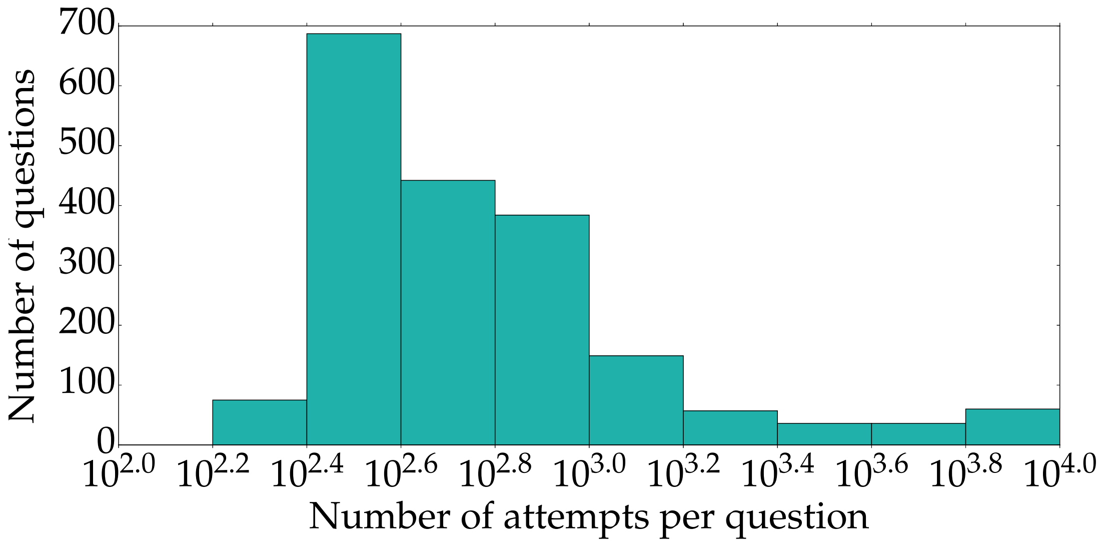
Experiments: Data Set

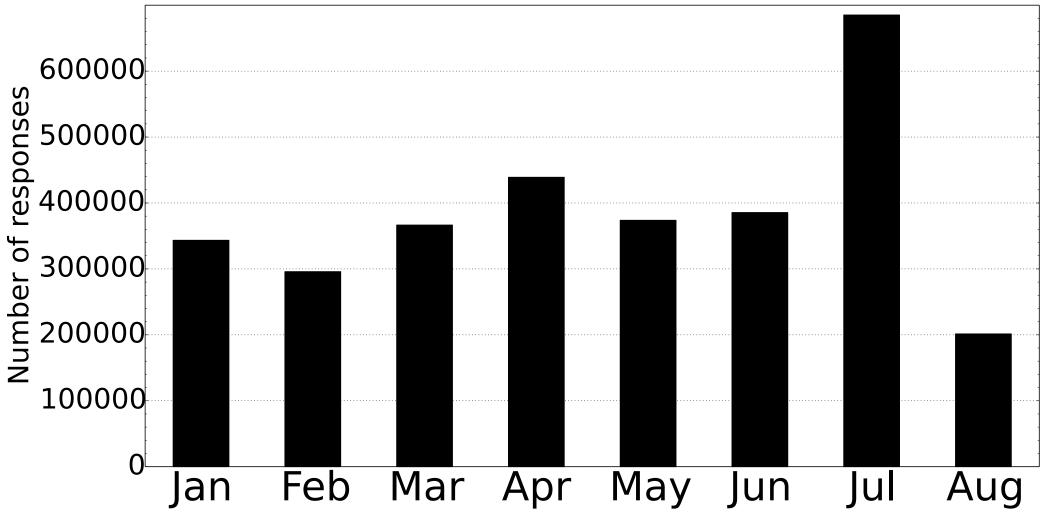


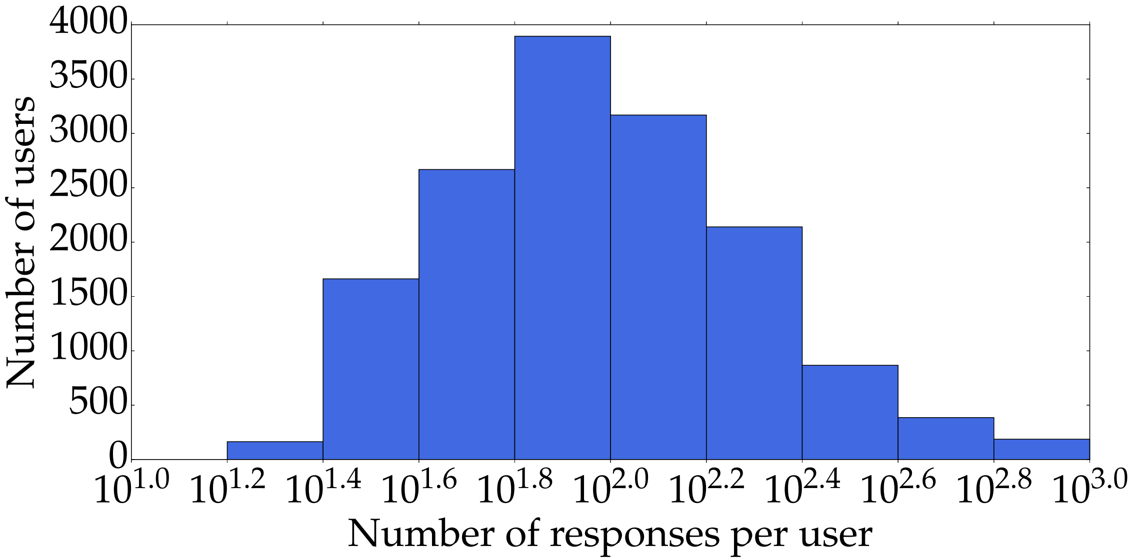
- Mobile applications (iOS/Android) launched in Korea
- Equipped w/ 4,202 TOEIC questions
- Data was collected from 1/1/2016 to 8/10/2016
- As a result,
  - 106k students signed up, 13m responses collected
  - => On average 130 questions per student
  - Many many outliers
    - Our app became so popular that a lot of people signed up just for checking out
    - Needed to preprocess the data

Experiments: Data Set

- Data Filtering
  - > 30 questions per student
  - > 3 seconds per question (on average)
  - > 400 students per question
- After filtering
  - n ~= 15k students
  - m ~= 2k questions
  - # of observed entries
    ~= 1.9m questions (6.5%)

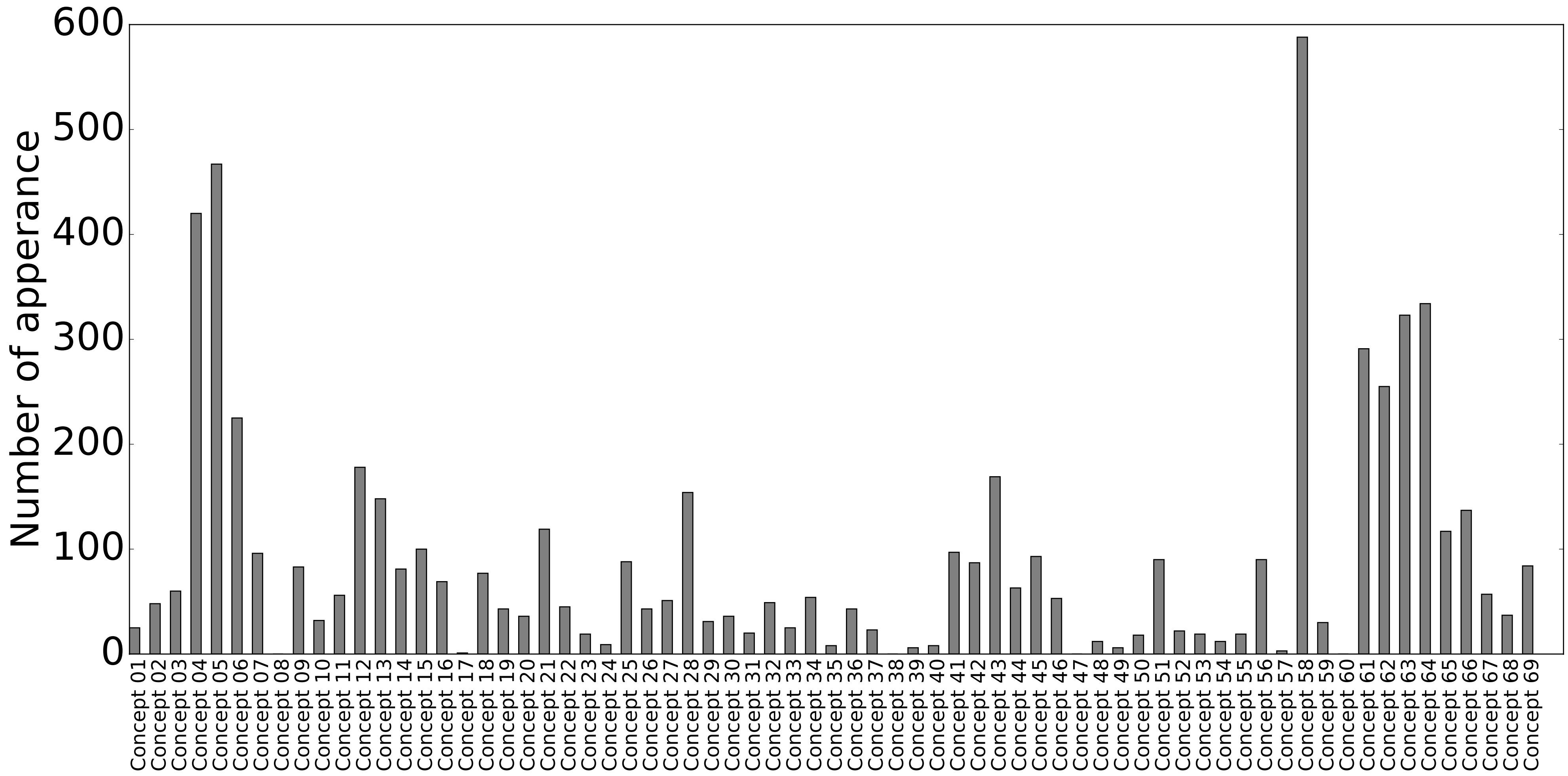




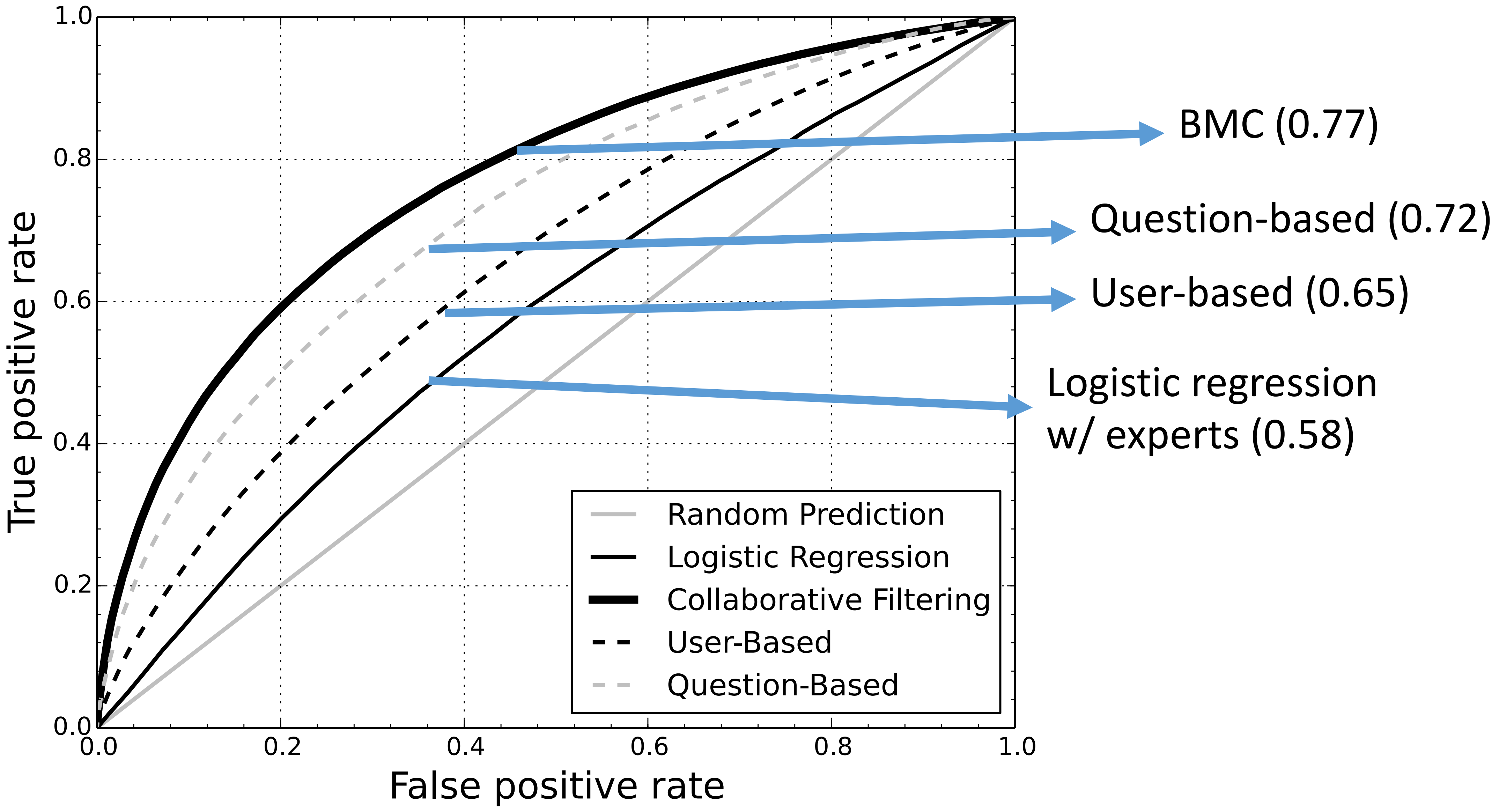


Experiments: Data Set

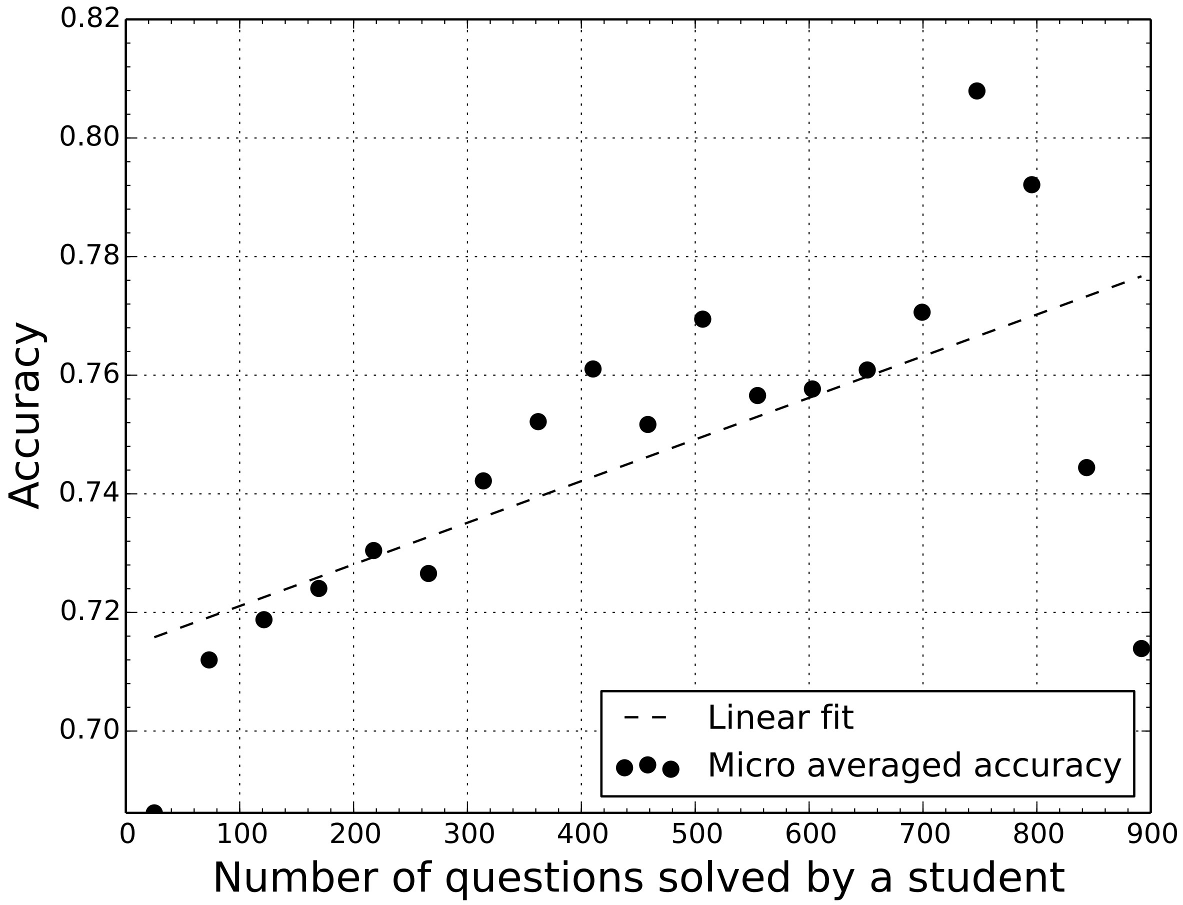
- m ~= 2k questions are manually tagged by experts
- 15 experts first come up with 69 concepts for describing part 4/5 questions
- Each question is randomly assigned to 2 experts among 15 experts

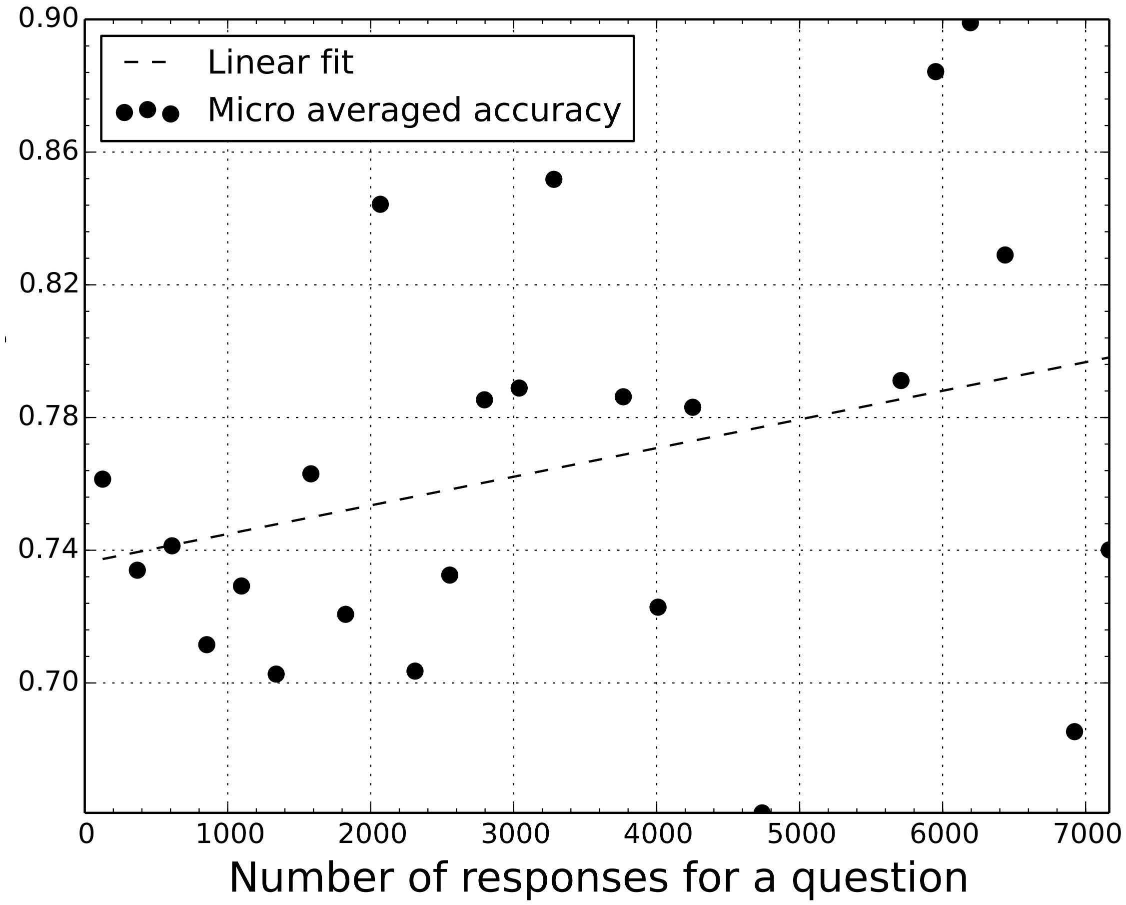


Experiments: Results (AUC)

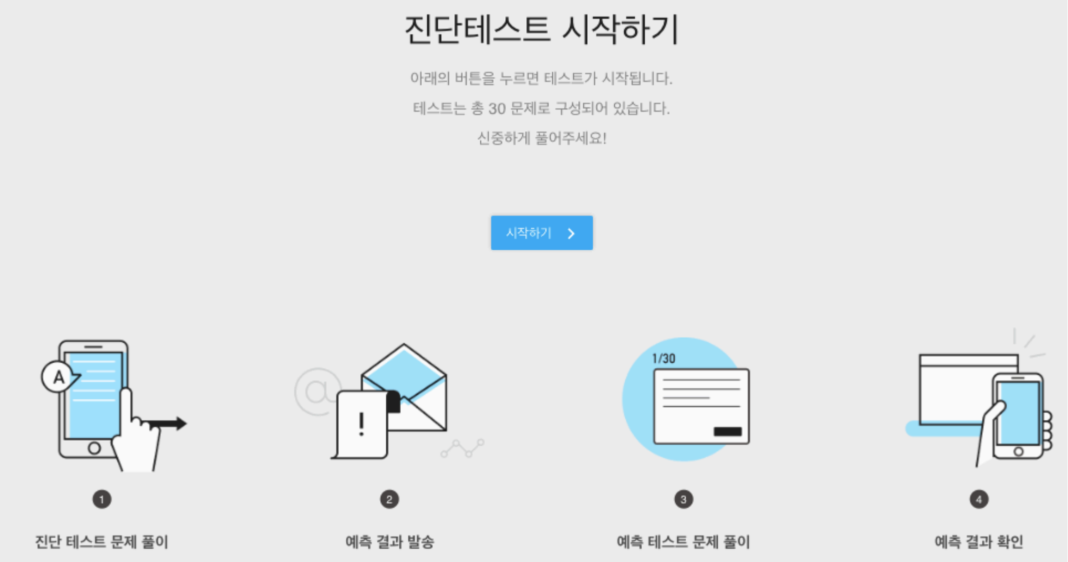


Experiments: Results (Accuracy)

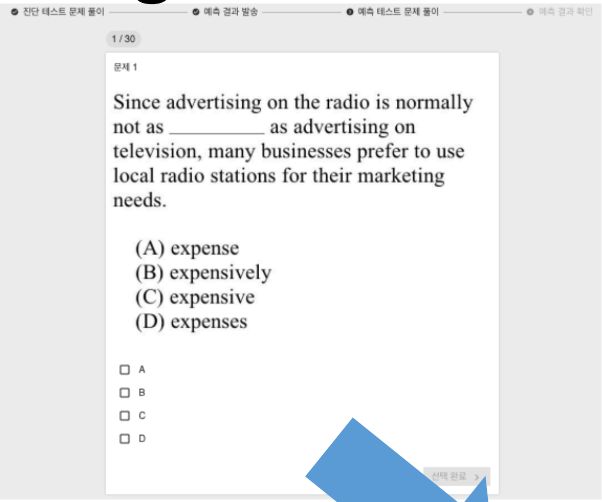


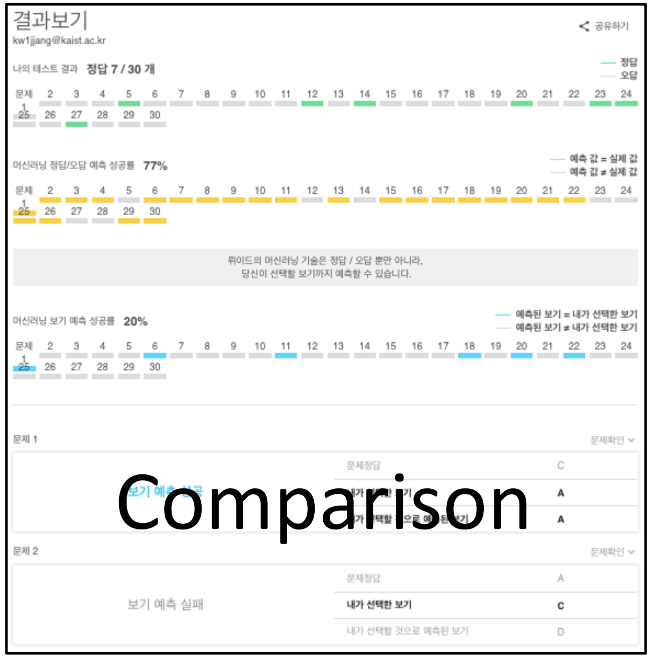


Prediction API in Products

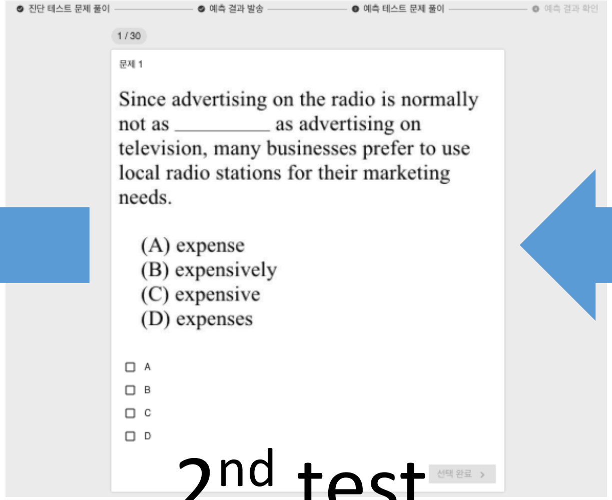


Diagnostic tests

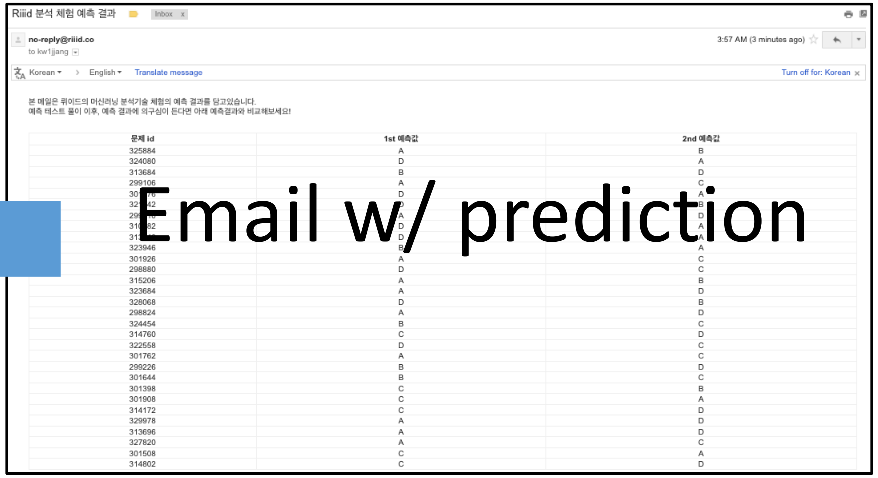




Comparison



2nd test



Email w/ prediction

Conclusion & Discussion

- ML framework for response prediction
  - Based on a variation of M2PL
  - Two algorithms:
    - Logistic regression with manually tagged questions
    - Binary matrix completion
- A large-scale experiment
  - Collected 13m responses from 106k students
  - A filtered data set is used for this work
  - Experimental results show that BMC works the best
- Deployed in products (email me if you want to try it yourself ☺)
- Many open problems & new directions
  - Interpretation of hidden concepts for an efficient design of edu. resources
  - Prediction of choices
  - Time-varying L, Sparse R
  - Convergence, Sample complexity, Biased sampling