Bayesian Ordinal Peer Grading

Karthik Raman, Thorsten Joachims
Cornell University
Evaluation at Scale: Peer Grading

• Conventional Evaluation:
  • Medium-scale classes (20-200 students): TAs perform grading.
  • Scaling to MOOCs (10000+ students)??

• MCQs & Auto-graded questions: Not a good test of understanding.
  • Limits kinds of courses offered.

• **PEER GRADING:** Students grade each other.
  • Overcomes scaling limitation of TA grading:
    • Number of “graders” scales with number of students!
Ordinal Peer Grading

• Students are not trained graders: Need to make feedback process simple!

• *Ordinal feedback* easier to provide and more reliable than *cardinal feedback*:
  • Project X is **better** than Project Y  **vs.**  Project X is a B+.

• **Ordinal Peer Grading**: Graders provide ordering of assignments
  • Need to infer overall ordering and grader reliabilities.
Mallows Ordinal Peer Grading

• GENERATIVE MODEL  
  \[ P(\sigma^{(g)}|\sigma^*) = \frac{e^{-\delta_K(\sigma^*, \sigma^{(g)})}}{\sum_{\sigma'} e^{-\delta_K(\sigma^*, \sigma')}} \]

  \[ \delta_K(\sigma^*, \sigma^{(g)}) \]  is the Kendall-Tau distance between orderings:
  • # of pairs ordered differently between the two rankings.

• Greedy algorithm to find MLE (Maximum-Likelihood estimator).

• Easy to extend: Grader Reliabilities, other aggregation models (Bradley-Terry ..)

• Viable alternate to conventional TA evaluation \([\text{KDD '14}]\)
  • As good (if not better) than cardinal peer grading.
Instructors Want More Details!

- Need finer-grained details to fully trust algorithm.
- Uncertainty information
  - Helps identify most confusing assignments.
- Allows instructors to understand grading output.
Solution: Bayesian Mallows Peer Grading

• Sample orderings using MCMC (Metropolis-Hastings)

\[ J_{MAL}(\sigma' | \sigma_{t-1}) \propto e^{-\delta_K(\sigma', \sigma_{t-1})} \]

Accept/Reject?

\[ \frac{P(\sigma' | \{\sigma^{(g)}; \forall g \})}{P(\sigma_{t-1} | \{\sigma^{(g)}; \forall g \})} \]

\[ \prod_{g \in G} e^{\delta_K(\sigma^{(g)}, \sigma_{t-1})} - \delta_K(\sigma^{(g)}, \sigma') \]

\( O(|D| \log |D|) \) time

\( O(|D|^2) \) time
Experiments

- Meaningful posteriors at cost of minimal grading error