

Bayesian Ordinal Peer Grading

Peer Grading Toolkit

[Home](#) [Software and Instructions](#) [Technical Details](#) [Datasets](#) [Contact Us](#)

This tool is designed for the problem of peer-grading/peer-reviewing. Given a set of assignments that need to be graded, we *aggregate* the grades provided by the peer graders/reviewers. The peer-grading toolkit takes as input a set of orderings provided by the reviewers indicating their preferences over the different assignments. For instance, in the example provided below reviewer 1 rates assignment 1 as being better than assignment 2 which in turn is better than assignment 3. Given these orderings you can use the tool to produce an overall ranking of all assignments as well as an estimate of how reliable each of the different reviewers were.

DATA USAGE POLICY: We do not store any of the data uploaded. The output rankings produced by the toolkit are deleted daily.

To learn more about the machine learning techniques we use please check out our [papers](#). You can also download the [code](#) and run it offline.

• Text Input

[Questions about the PGF format?](#)

```
task1 rvwrid_1 asegnid_1 > asegnid_2 > asegnid_3
task1 rvwrid_2 asegnid_1 > asegnid_2 > asegnid_3
task1 rvwrid_3 asegnid_1 > asegnid_3 > asegnid_2
```

[Peer Grade Format File](#)

[CMT Export-Format File \(XLS: XML Spreadsheet\)](#)

[Submit](#)

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.

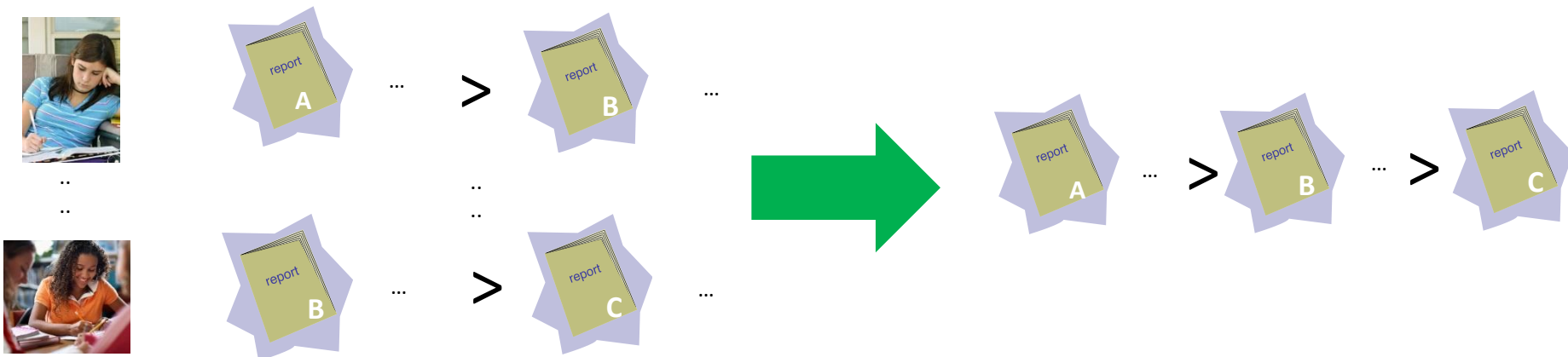


- PEER GRADING: Students grade each other.
- Overcomes scaling limitation of TA grading:
 - Number of “graders” scales with number of students!

Ordinal Peer Grading

[KDD 2014]

- 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

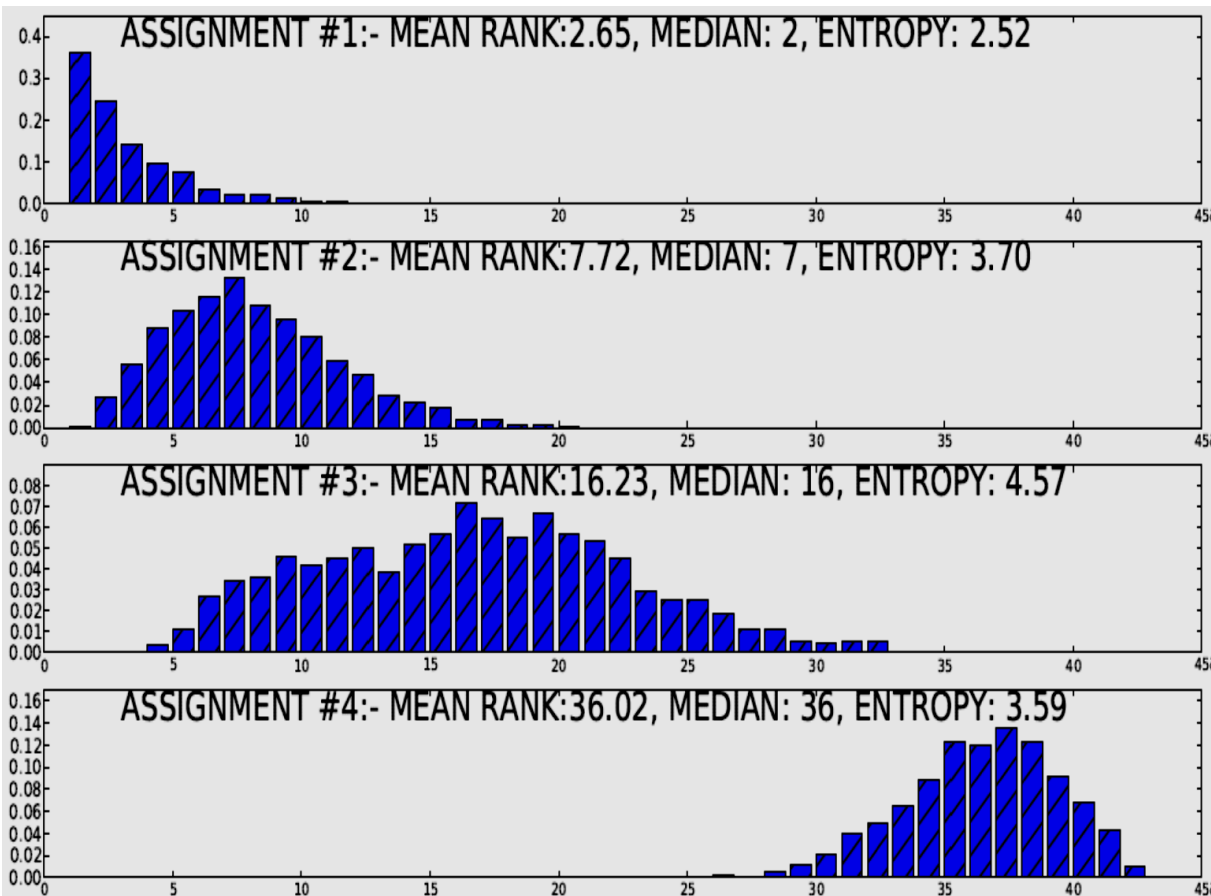
- GENERATIVE MODEL $P(\sigma^{(g)}|\sigma^*) = \frac{e^{-\delta_K(\sigma^*, \sigma^{(g)})}}{\sum_{\sigma'} e^{-\delta_K(\sigma^*, \sigma')}}$

$\delta_K(\sigma^*, \sigma^{(g)})$ 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 [KDD '14]
 - As good (if not better) than cardinal peer grading.

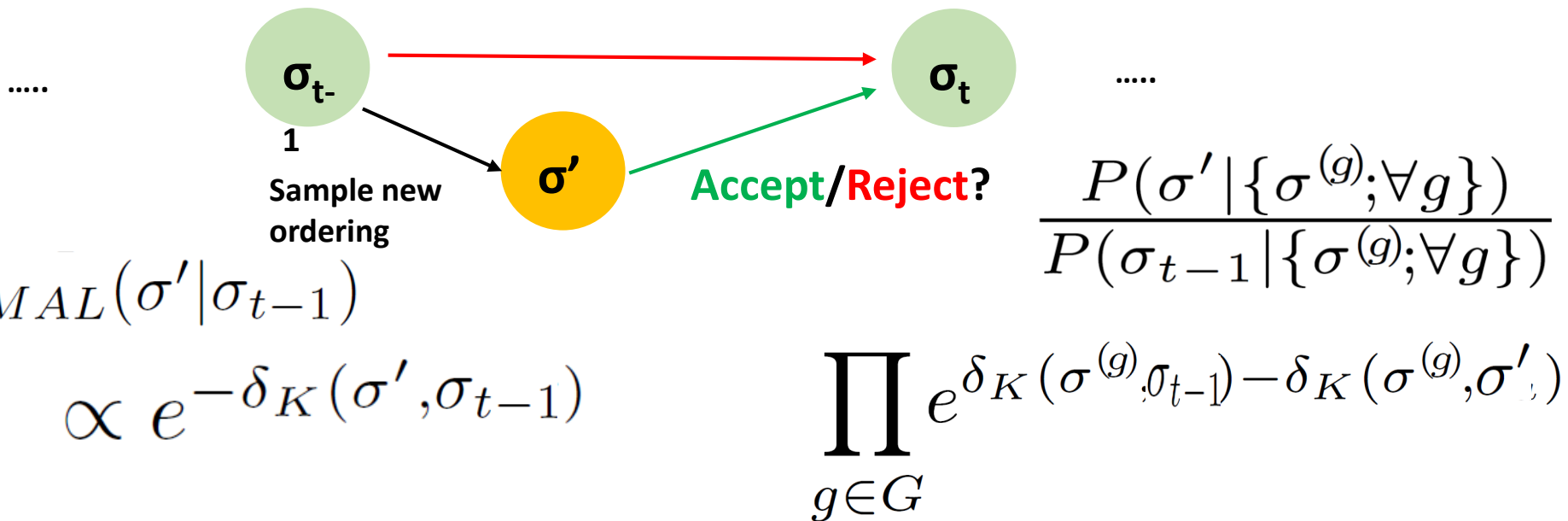
Instructors Want More Details!

- 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)



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

$O(|D|^2)$ time

Experiments

- Meaningful posteriors at cost of minimal grading error

