UNIVERSITY OF PITTSBURGH www.ischool.pitt.edu
School of Information Sciences

SIMULATING USER SELECTIONS OF QUERY SUGGESTIONS

Jiepu Jiang, Daqing He

School of Information Sciences,
University of Pittsburgh

MOTIVATION

- **Lots of works on query suggestion/recommendation**

- **The performance of a query suggestion list depends on**
  - Whether or not the user will take one suggestion
  - Which query suggestion is taken

- **Some facts** (Wu et al. SIGIR'12; Hauff et al. CIKM'10)
  - User judgments of query quality can be different from those evaluated by metrics (e.g. nDCG@10)
    - Absence of search results
    - User judge criteria
  - Users may not be able to identify and adopt the best query suggestion

RELATED WORKS

- **Measuring the quality of a query suggestion**
  - Search performance of the results
  - User ratings, user clicks on query suggestions

- **Measuring the quality of a list of query suggestions (at least two dimension of modeling the problem)**
  - Position or rank related
    e.g. discounting lower ranked query suggestions
  - ***User judgments and selections*** **(our focus)**

SOME EXISTING METHODS

- **Evaluate a list of query suggestions by**
  - The performance of the "best" query
    (Wang & Zhai CIKM'08; Dang & Croft WSDM'10)

  Assumption:

  Users can make perfect judgments and always identify and adopt the best query suggestion.

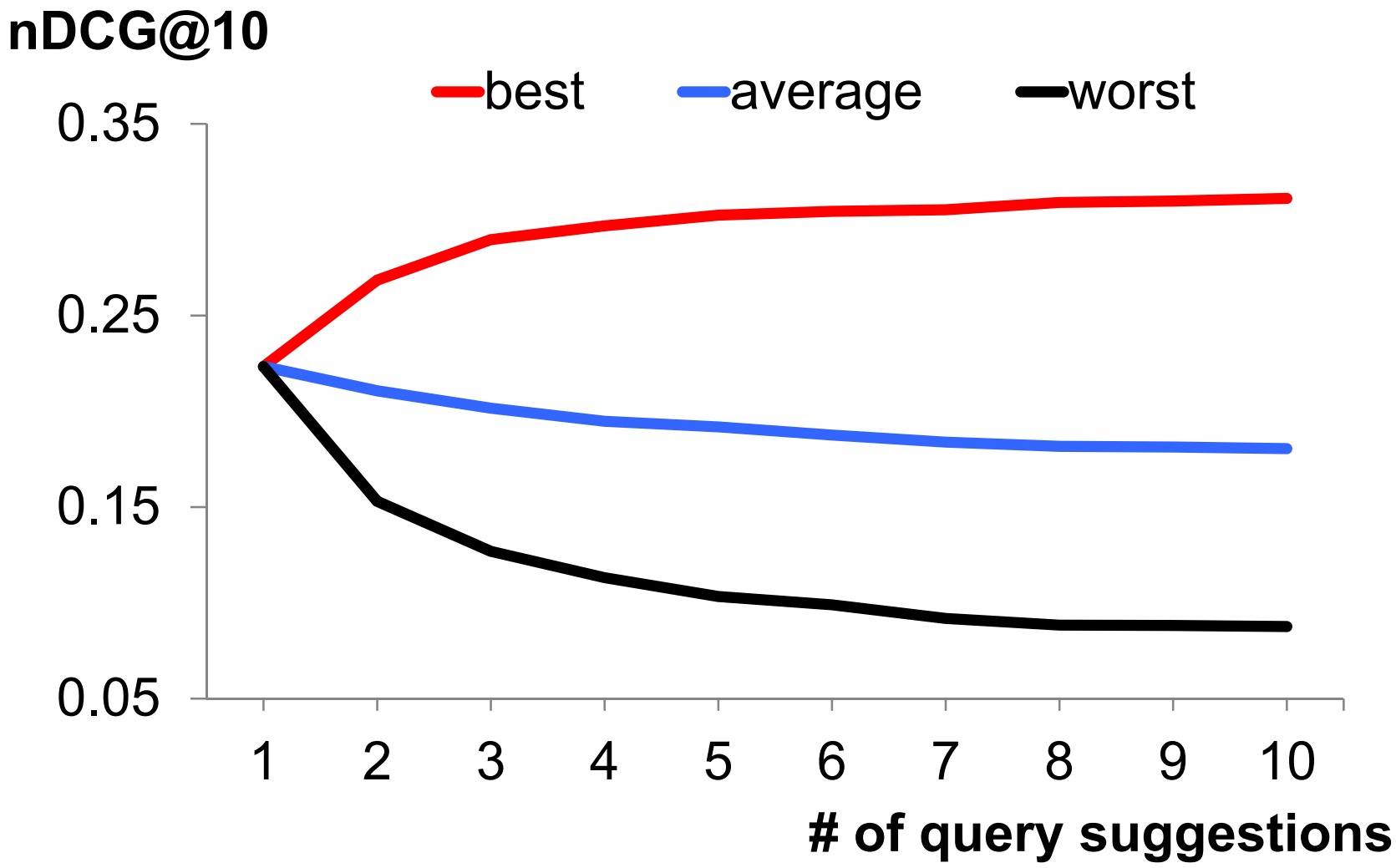
  - The average performance of the queries
    (Sheldon et al. WSDM'11)

  Assumption:

  Users will randomly adopt a query suggestion.

IS IT IMPORTANT?

EXAMPLE 1: HOW MANY SUGGESTIONS SHOULD BE DISPLAYED?



EXAMPLE 2: WHEN SHOULD WE DISPLAY QUERY SUGGESTIONS?

- **Is there a chance that query suggestion could lead to decline of search performance?**

- **Probably yes, e.g. the user may take query suggestions that underperform the queries could be reformulated by the user him/herself**

- **A possible proof (Kelly et al. SIGIR'09)**
  - *"query suggestions seem to have an advantage when subjects face a cold-start problem and when they exhaust their own ideas for searches"*

A SIMULATION PROCESS

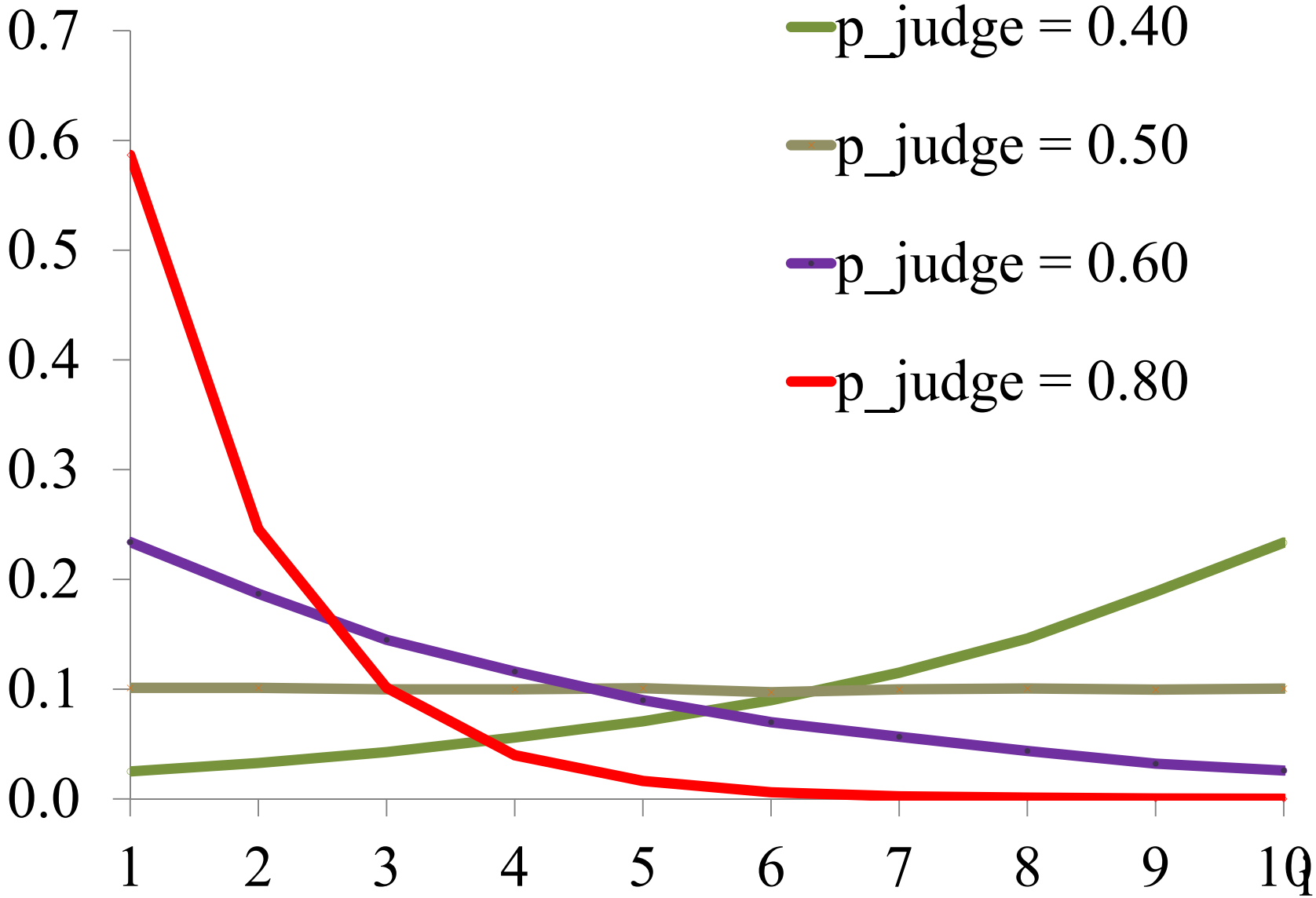
- **$P_{judge}(q_1, q_2)$: the probability that user can make a correct pairwise judgment on the quality of q_1 and q_2.**
  - $P_{judge}(q_1, q_2) = 1$ ➔ perfect judgments
  - $0.5 < P_{judge} < 1$ ➔ positively correlated
  - $P_{judge}(q_1, q_2) = 0.5$ ➔ random selection
  - $0 < P_{judge} < 0.5$ ➔ negatively correlated

- **A query selection "tournament"**
  - q_a vs. q_b: the better query has the probably $P_{judge}(q_a, q_b)$ to win and get 1 point
  - Iteratively compare each pair of queries to come out the "winner" query
  - Run the tournament many times to estimate the probability of selecting each query.

A SIMULATION PROCESS

- **The probability of selecting a query depends on**
  - The performance of the query compared with others
  - The user's judge ability, as characterized by P_{judge}
- **Some effects of the model**
  - The better the user's judging ability is, the more likely that the user can select the best query in C.
  - The better a query's quality is, the more likely that the user will select the query.
- **And (at this point), independent of its rank in the list**
  - But we can model the rank by other approaches (see details in the workshop paper)

THE PROBABILITY OF SELECTING THE ITH BEST QUERY (FIXED P_{JUDGE})



SOME EFFECTS

- **The better the user's judging ability is, the more likely that the user can select the best query from the list.**

- **The better a query's quality is, the more likely that the user will select the query (if the user's selection is positively correlated with ground truth).**

ON-GOING WORK

- **User experiments involving user judgments & use of query suggestions in a search session**

- **Aiming for studying**
  - What are good measures for evaluating a single query suggestion's quality
  - Predictive factors for user judgments and selections of queries

THANK YOU!