Judge's Commentary: The Outstanding Contest Judging Papers

Veena B. Mendiratta
Bell Labs
Lucent Technologies
2000 N. Naperville Road
Naperville, IL 60566
veena@lucent.com

The Contest Judging Problem provided the contestants with a challenging real-world problem that lent itself to a range of analysis and modeling methods. In coming up with a "best" selection scheme, the contestants used methods such as rank-ordering, numerical scoring, and bias estimation.

What made the problem interesting and challenging to model, and also to judge, were the less well-defined aspects of the problem. For example, how was bias of judges with respect to ranking and scoring papers handled? Also, how was the issue of ensuring that the best $2W$ papers were not screened out in the early rounds of reading addressed? Consequently, we could not select the winning papers for MCM 1996 based solely on the criterion of how many paper-readings a team's algorithm required; we also considered how the issues were addressed.

Almost all of the successful papers were able to develop a basic model for the ideal case to select the top W papers for the specific parameter values specified in the problem statement. A key assumption for the ideal case is that every judge rank-orders or scores all the the papers as the absolute rank-ordering, that is, there is no judge bias. The so-called ideal model, though unrealistic, sets a lower bound for the total number of reads. The stronger papers went significantly beyond the ideal case.

Characteristics of the best papers included the following:

- An explicit modeling of judges' bias, addressing the issues of systematic bias and the variance of accidental errors in scoring.
- Estimating statistical bounds on the probability of failure to pick the best W papers out of the top $2W$ papers.
- In addition, some of the papers realized the importance of more judges reading the papers remaining in the later screening rounds and included this factor in their models.

- The better papers provided a clear statement of results in terms of the total number of paper readings, the confidence level of the results, and sensitivity of the results to the model parameters.

Various different approaches were selected to address the above issues and some of these are summarized below.

The Gettysburg College team minimizes the probability of eliminating the W best papers in the first round by having two judges read each paper in that round. This same paper models judge error through a functional relationship between the probability of judge error in ranking (with respect to the absolute ranking) and the distance between two compared papers on the absolute scale. The team from the University of Science and Technology of China uses Bayesian estimation to address systematic bias in scoring. They also model the error probability as a function of the percentage of papers eliminated in each round. The Fudan University team shows statistically the conditions under which the "ideal" model can work. The Washington University team models the judge bias and error and, after each round of judging, uses new bias estimates to calculate confidence intervals that determine the number of papers rejected. The St. Bonaventure team implements a novel distribution scheme, which they illustrated very effectively with matrices, to ensure that judges do not receive the same paper more than once and also that the same judge does not receive the top $2W$ papers.

Lastly, the best papers were characterized by clear and logical presentations that brought forth the team's underlying analytical thought process. These papers were well organized and well written, with appropriate tables and graphics for presentation of results, and included a comprehensive summary, all of which made it easier to understand the material presented.

The Contest Judging problem was challenging and many excellent solutions were offered. Finally, however, five papers stood out from the others, and the members of those teams should feel proud of their accomplishments.

About the Author

Dr. Mendiratta (Ph.D., Northwestern University, 1981) has been at Bell Labs (now part of Lucent Technologies) since 1984, working on a wide range of systems. She currently works in the Architecture and Performance area. Her work at Bell Labs has focused on reliability modeling and performance analysis of switching systems, as well as mathematical programming models for switch configurations. Prior to 1984, she worked for three years as Manager of Operations Research for the Illinois Central Gulf Railroad, where she directed the implementation of an empty-freight-car distribution optimization model that was developed as part of her Ph.D. dissertation. Her professional activities include serving on the MCM Advisory Board as well as an MCM judge, being co-President of the INFORMS Chicago Chapter, and being a SIAM Visiting Lecturer.