Judges' Commentary: The Coach Papers

Robert Burks
Defense Analysis Dept.
Naval Postgraduate School
1 University Circle
Monterey, CA 93943–5000
reburks@nps.edu

The Problem

Sports Illustrated, a magazine for sports enthusiasts, is looking for the "best all-time college coach," male or female, for the previous century. Build a mathematical model to choose the best college coach or coaches (past or present) from among either male or female coaches in such sports as college hockey or field hockey, football, baseball or softball, basketball, or soccer. Does it make a difference which time line horizon that you use in your analysis, that is, does coaching in 1913 differ from coaching in 2013? Clearly articulate your metrics for assessment. Discuss how your model can be applied in general across both genders and all possible sports. Present your model's top 5 coaches in each of 3 different sports.

In addition to the MCM format and requirements, prepare a 1–2-page article for *Sports Illustrated* that explains your results and includes a non-technical explanation of your mathematical model that sports fans will understand.

Introduction and Overview

The Coach Problem focused on identifying the factors or metrics for success as a college coach. The problem required students to develop a modeling approach based on these metrics to determine the best coach

The UMAP Journal 35 (2–3) (2014) 181–187.

across all sports, gender, and time. In addition, there was the traditional required nontechnical (*Sports Illustrated* article) paper.

I start this commentary with a short review of the mechanics of this year's judging process. I follow the mechanics with a discussion and observations from the judging on various elements of the problem. I then discuss the importance of sensitivity analysis, assumptions and identifying the strengths and weaknesses of a developed model. I finish by addressing some points concerning communication and conclude with a summary.

The Process

Dr. Kelly Black provided an excellent overview of the judging process in his commentary for the Ultimate Brownie Pan Problem of 2013 [2013]. However, I believe it is beneficial to once again review several elements of the process for this year's problem. In general, it is important to understand that the criteria used to identify good papers gradually change as the judging progresses through the triage and final rounds, with the final papers standing out as the best under a wide variety of criteria.

Triage

The primary objective of the triage is to identify the papers that should be given more detailed consideration from the judges.

Every paper is read by at least two judges seeking to determine if the paper contains all of the necessary elements that make it a candidate for more-detailed readings. If a paper addresses all of the issues and appears to have a reasonable model, then judges are likely to identify it as a paper that deserves more attention.

A paper must be clear and concise to do well in the triage, and the paper's summary is critical at this point in the judging. A good summary provides a brief overview of the problem, the paper's structure, and specific results stated in a clear and concise manner. Small things that make a paper stand out include having a table of contents and ensuring that all required questions are addressed in the paper.

Many papers do not do well in the triage because the summary fails to address all of the questions, and the judge decides that a team's efforts will not compare well with the better papers. For example, one critical question overlooked by many papers this year was how their model could be applied across both genders and all possible sports. Fully developing all of the required elements is a critical area often overlooked in papers.

The sensitivity analysis remains one of the weakest elements in many papers, and these papers do not do well during the triage.

In addition, it is vital that the team express their general approach and results as clearly and concisely as possible in the nontechnical position pa-

per. This means providing a broad overview of the problem, the approach, and specific results in clear and concise nontechnical terms. In other words: Can the article be read and understood by someone without an education in mathematics?

These small things make it much easier for a judge to identify the team's effort and for the paper to do well in the triage round. However, the best models and the best effort is not effective if the results are not adequately communicated. It is important to remember that this is a modeling competition and that effective communication is a critical part of the modeling process.

Final

The final consists of multiple rounds of judging over several days. As the rounds progress, the judging criteria shift from identifying papers that warrant further consideration to a process to identify the very best papers.

The first round of the final begins with each judge reading a set of papers and then all judges meeting to discuss the key aspects of the question and what should be included in a "good" paper. This year these aspects included, in addition to all of the required elements:

- a clear discussion of the assessment metrics,
- how and why these metrics were weighted,
- the incorporation of time in the analysis, and
- emphasis placed on the sensitivity analysis portion of the paper.

As the final progresses, each paper is read multiple times, with the final set of papers being read by all judges. In these last rounds, the modeling process and the mathematical integrity of a paper begin to identify the Outstanding papers in the competition.

The Questions

This year's Coach Problem consisted of three major components:

- The first component required teams to determine the mechanism for selecting the best college coach.
- The second component required teams to address the impact of time on their analysis.
- The last component focused on gender as a factor in the best coach selection process and how their models could be applied across all sports.

The Best College Coach

One major aspect a team must address is what is meant by the term "best." Was it number of wins? Was it number of years coaching? Was it popularity? Was it some combination of a set of factors? Many teams did not take the time to clearly develop the purpose of their model or to define "best" but immediately began modeling this aspect of the problem.

In general, teams did not take an approach of developing a generic definition of coaching success. It appears that many teams started their modeling process by first selecting a sport then developing "successful" coaching metrics based on that sport. This approach had a tendency to result in some sport-specific metrics. For example, the number of Bowl Games was a popular metric for college football. This metric becomes a clear problem when attempting to apply the model across all sports, and such a problem should be addressed in the paper.

Better papers first considered carefully the definition of "best" in terms of generic coaching success, then discussed how it could be measured.

The judges were not looking for a specific set of assessment metrics but were looking for those papers that clearly identified and developed their metrics. Most teams developed a set of metrics (anywhere from 5 to 15) that they collectively modeled to develop rankings in each of the three required sports. The better papers tended to develop a set of global metrics that could be applied across all sports and genders and then applied them individually to a set of three different sports.

Many papers treated this requirement as a multi-objective decision-making problem. Popular modeling approaches included the Analytical Hierarchal Process (AHP), Principal Component Analysis (PCA), Fuzzy Comprehensive Evaluation, TOPSIS, Artificial Neural Network, and Dynamic Network Analysis (DNA). However, by far, the most common modeling approach was the AHP. The large volume of papers utilizing the AHP would seem to suggest that this problem was developed for this approach—but that is not the case. There were several very successful papers that did not utilize the AHP approach.

However, many teams, regardless of the modeling approach, failed to recognize the inherit nature of having to assume some sort of weighting mechanism for their metrics in the modeling process. The judges considered the discussion of weighting as a critical criterion for "good" papers. The better papers recognized this, discussed how they developed their weights, and then conducted a sensitivity analysis on their assumed weights.

The Impact of Time

Has the nature of coaching changed over time, or is a coach in 1913 the same as a coach in 2013? An analysis of the impact of time had mixed

results and was one of the weaker modeling components in many papers. It appears that many teams ran short of time and provided little analytical effort to this component. The analysis of time was viewed by the judges as a critical criterion for "good" papers. The judges were looking for some recognition that in 100 years of college sports history, the rules of the game, the number of games, the nature and duration of training programs, the social environment, and a host of additional factors have changed—and these changes may influence the metrics used in their model.

Many teams approached this aspect as a time-series analysis problem. These teams evaluated how their model's metrics may have changed over time, to see if there was a correlation between their metric and year. If a team discovered a correlation, typically it was simply noted in the paper that time was important and influenced their model's ranking; but the team usually made no adjustments to their models. The better papers adjusted their metrics, usually the weighting mechanism, and generated a new coach ranking.

Gender

The last question examined how the model could be applied across both genders and all sports. This requirement consisted of two distinct discussion points, addressing both genders and modeling across all sports. Most teams addressed the gender requirement in their paper but were weaker at analyzing how their model could be applied across all sports.

The judges were not looking for any one particular approach to address the gender requirement. Most papers addressed this requirement by modeling a traditional women's college team and ranking the best set of woman coaches. This was presented as evidence that the model worked for both genders. However, a handful of teams used their model to rank a combined list of men- and women-coached teams, developing a top-5 ranking that contained both men and women coaches. The judges viewed this as a superior approach to model the requirement.

In terms of applying the model across all sports, most teams provided their rankings for three different college sports as evidence that the model applied across all sports. Only a handful of papers used the developed model to produce a single ranking that encompassed all sports and genders as evidence. The judges viewed this as a superior approach that went beyond the basic question and requirement of the problem.

Analysis: Assumptions, Sensitivity

The judges realize the limited time available to the teams to complete their models is a considerable constraint, and they do not expect perfect models. However, the judges do expect teams to analyze their models in

a structured way and to critically assess their models. A vital part of the mathematical modeling process is this critical analysis of the model. This analysis ranges from examining the impact of the basic assumptions on the modeled conclusions to examining the shortcomings of the techniques employed in the model.

As in previous years, the judging criteria placed a large emphasis on assumptions and sensitivity analysis. Many papers neglected to fully consider these issues and were scored lower by the judges.

Assumptions

The basic assumptions that a team makes is the starting point for their modeling efforts. The judges did not place restrictions on the basic assumptions other than that they need to make sense and be necessary. However, simply listing assumptions is not enough; papers should include a discussion of why they are making the assumption and their potential impact/influence on the model.

It is also important to recognize that stating the assumption is not the end of the process, but that examining the impact on the modeled conclusions when a change in the assumption takes place is a vital part of the modeling process. If changing an assumption results in a change in the coach rankings, then the team should indicate that as a potential weakness.

Sensitivity Analysis

Sensitivity analysis was appropriate and necessary for all modeling approaches. For AHP, sensitivity analysis would have involved varying the weights (or pairwise rankings) to explore what conditions would cause the alternative ranking to change. Many papers included a sensitivity analysis section in their paper but only addressed the theoretical aspects of sensitivity analysis as opposed to actually changing the value of an assumption or parameter to understand the impact.

Communication

Papers were judged on the quality of the writing, with special attention to the summary and to the nontechnical (*Sports Illustrated*) article. In general, the quality of writing is continuing to improve. The strongest summaries this year included a definition of what the team meant by "best," a general overview of the modeling process, and an explicit result of the model analysis.

The judges continue to be surprised by the number of papers where the summary only describes what the team will attempt without telling the results.

Similarly, many of the nontechnical articles focused more on the mathematics and modeling process than on the details of the rankings. A nontechnical article does not mean that numbers are not included. It means that the article can be read meaningfully by someone without an education in advanced mathematics.

Conclusions

The Outstanding teams modeled and presented all the aspects of the problem described in the problem statement, including the fully-developed standard elements (assumptions, sensitivity analysis, strengths and weaknesses, etc.), developed an effective model, explained the modeling choices made, and were clearly and concisely written. The judges continue to be impressed with the quality of the submissions, especially considering the time constraints. The growth in the quality and number of submissions is very encouraging to those who work to promote the practice of good mathematical modeling.

Reference

Black, Kelly. Judges' Commentary: The Ultimate Brownie Pan papers. *The UMAP Journal* 34 (2) (2013): 141–149.

About the Author

Robert Burks is a senior analytic consultant in the Dept. of Defense Analysis at the Naval Postgraduate School. He received his undergraduate degree in Aerospace Engineering from the United States Military Academy, his Master's in Operations Research from the Florida Institute of Technology, and his Ph.D. in Operations Research from the Air Force Institute of Technology. He has wide-ranging research interests, including diffusion of information and epidemiology agent-based modeling. Dr. Burks served as both a triage and final judge on the Coach Problem.