How to Select the Best Paper: Towards Justification (and Possible Enhancement) of Current Semi-Heuristic Procedures

Francisco Zapata[1], Olga Kosheleva[2], and Vladik Kreinovich[3]
[1]Department of Industrial, Manufacturing, and
Systems Engineering
[2]Department of Teacher Education
[3]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, Texas 79968, USA
fazg74@gmail.com, olgak@utep.edu, vladik@utep.edu

Abstract

To select the best paper at a conference or in a journal, people use reasonably standard semi-heuristic procedures like averaging scores. These procedures usually work well, but sometimes, new situations appear for which the existing procedures are not automatically applicable. Since the existing procedures are heuristic, it is often not clear how to extend them to new situations. In this paper, we provide a possible explanation for the existing procedures. This explanations enables us to naturally generalize these procedures to possible new situations.

1 How Best Papers Are Selected Now: Description and Challenges

How best papers are usually selected. Many conferences and journals have a tradition of selecting the best paper; for journals, it is usually the best paper published in a certain time period. To select the best paper, the conference or the journal asks several respected researchers to form a committee. Members of the committee state their opinions. These opinions are then combined into a single score, and the paper with the largest score is proclaimed the winner.

Sometimes, experts provide numerical estimates of the papers' quality. In some cases, committee members are asked to evaluate each paper by a number on a given scale – e.g., on a scale from 0 to 10. As a result, for each expert i and for each expert j, we get a number v_{ij} describing the j-th expert's estimate of the quality of the i-th paper.

How these estimates are usually combined? Usually, for each paper, we simply add the experts' estimates of this paper – or, which is equivalent in terms of the resulting ordering of papers, we take the arithmetic average of different scores.

Sometimes, experts are asked to rank the papers. In some cases, instead of asking the experts to evaluate each paper on a scale, the conference simply asks each expert to rank all the candidate papers. Then, if there are n candidate papers, for each expert, the top paper gets n points, second best gets $n-1$ points, etc., all the way to the worst paper that gets 1 point. These points are then averaged, and the paper with the largest overall score is proclaimed the best.

Challenges. In most cases, the above procedure works well, but sometimes, unusual situations occur. For example, sometimes, one of the experts, instead of ranking all the papers, just selects a paper which is the best according to him/her, and does not provide the ranking of all the other papers. How do we take the opinion of this expert into account?

This is not an easy question to answer because the existing strategy is semi-heuristic, it is not based on any well-defined set of principles, so it is not clear how to extend the existing strategies to such new cases.

As a result, for each such new case, people have to invent new ideas. It would be therefore nice to come up with a theoretical explanation of the existing strategies, an explanation that would enable us to automatically generalize these existing strategies to new situations.

In this paper, we provide such an explanation.

2 How to Justify the Current Semi-Heuristic Methods for Selecting Best Papers

Why arithmetic average. The first procedure that we described above was taking an arithmetic average of estimates provided by several experts. A natural first question is: why arithmetic average? Why not mean squared value? Why not geometric average? To answer this question, let us analyze this situation in detail.

For each paper j, have several estimates v_{1j}, v_{2j}, ..., v_{nj} provided by different experts. We want to combine them into a single estimate. A natural idea is to assume that for each paper, there is the actual objective value e_j, and expert estimates v_{ij} are approximations to this desired value. In other words, we want to find the value e_j for which $v_{1j} \approx e_j$, $v_{2j} \approx e_j$, etc.

This is a typical situation in data processing, when we have several results v_{1j}, v_{2j}, etc., of measuring the desired quantity e_j, and we want to combine these measurement results into a single – more accurate – estimate of this quantity.

There are usually many different (and reasonable independent) factors due to which the measurement result differs from the actual value. The difference $v_{ij} - e_j$ is the result of the joint effect of all these factors. It is known that, under reasonable condition, the probability distribution of the sum of a large number of independent similar-size random variables is close to Gaussian (normal); see, e.g., [10]. Thus, it makes sense to assume that the differences $v_{ij} - e_j$ are normally distributed, with some standard deviation σ.

This means that for each value e_j and for each expert i, the probability of having the estimate v_{ij} is proportional to $\exp\left(-\frac{(v_{ij} - e_j)^2}{2\sigma^2}\right)$. The experts are usually assumed to be independent. So, the probability to have all the given estimates is equal to the product of the product of the corresponding probabilities:

$$\prod_{i=1}^{n} \exp\left(-\frac{(v_{ij} - e_j)^2}{2\sigma^2}\right). \quad (1)$$

Out of all possible values e_j, it is reasonable to select the *most probable* value, i.e., the value for which the probability (1) is the largest. Maximizing the expression (1) is equivalent to minimizing its negative logarithm, i.e., equivalently, minimizing the sum

$$\sum_{i=1}^{n} (v_{ij} - e_j)^2.$$

Differentiating this expression with respect to e_j and equating the derivative to 0, we get

$$e_j = \frac{1}{n} \cdot \sum_{i=1}^{n} v_{ij},$$

i.e., we get a justification of the usual arithmetic average.

Why n points for the top rank, $n - 1$ points for next ranked paper, etc. The above subsection describes how to combine the numerical grades, but what if instead of numerical grades, experts only provide rankings of different papers? How do we convert a ranking into numerical grades?

Let us assume that all these grades are from some interval. Without losing generality, we can safely assume that this interval is the interval $[0, 1]$. We want to assign, to each paper, a number so that paper ranked better would get the higher number. If we denote the value assigned to the worst paper by v_1, the values assigned to the second worst paper by v_2, all the way to the best of n paper to which we assign the value v_n, then these n numbers must satisfy the inequality

$$0 \le v_1 < v_2 < \ldots < v_n \le 1. \quad (2)$$

There are many different tuples $(v_1, \ldots, v_n)$ with this property. We have no reason to assume that one of these tuples is preferable, so it makes sense to assume that all these tuples are equally probable. In precise terms, this means that we assume that we have a uniform distribution on the set of all the tuples that satisfy the inequality (2).

We have different possible tuples, we need to combine all these different possible tuples into a single tuple. Similarly to the previous section, we can argue that the most appropriate combination is taking the arithmetic average of all these tuples, i.e., in mathematical terms, taking the mean values $m_1, \ldots, m_n$ of the corresponding components $v_1, \ldots, v_n$ – mean values in the sense of the above uniform distribution.

It turns out that these mean values have a single form $m_i = \dfrac{i}{n+1}$; see, e.g., [1, 2, 3, 4, 5, 6, 7, 8, 9]. Thus:

- the amount of points assigned to the most highly ranked paper, with $i = n$, is proportional to n;
- the next paper gets the number of points proportional to $n - 1$,
- etc.
- until we get to the lowest ranked paper, to which we assign the number of points proportional to 1.

This is exactly how this assignment is usually done, which means that this usual assignment has also been justified.

3 How to Use These Justifications to Handle Possible Challenges

The ideas behind the above justifications can be used to provide a recommendation on what to do in other situations as well. For example, let us consider the above challenge when one of the experts, instead of ranking all n papers, just selects the best one.

In this case, no matter how we rank other papers, the selected paper get the value n. Depending on how we rank all other papers, other papers get values from 1 to $n - 1$. Since we do not have any reason to assume that one of the rankings is preferable, it is reasonable to conclude that all these rankings are equally probable. Thus, similarly to the previous session, the number assigned to each of the remaining $n-1$ papers should be equal to the arithmetic average – i.e., to the mean – of all the values assigned to this paper according to different possible rankings.

To compute these averages, there is no need to consider all $(n-1)!$ possible rankings: it is sufficient to take into account that, first, the number assigned to each of the $n - 1$ papers is the same, and second, since the sum of numbers m assigned to all $n - 1$ papers is always equal to

$$1 + 2 + \ldots + (n-1) = \frac{(n-1) \cdot n}{2},$$

the mean values should also add up to the same sum:

$$m + \ldots + m \ (n-1 \text{ times}) = (n-1) \cdot m = \frac{(n-1) \cdot n}{2}.$$

Thus, $m = n/2$. So, in this situation:

- to the highest ranked paper, we assign the value n, and
- to every other paper, we assign the value $n/2$.

Acknowledgments

This work was supported in part by the National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

References

[1] M. Ahsanullah, V. B. Nevzorov, and M. Shakil, *An Introduction to Order Statistics*, Atlantis Press, Paris, 2013.

[2] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*, Society of Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, 2008.

[3] H. A. David and H. N. Nagaraja, *Order Statistics*, Wiley, New York, 2003.

[4] O. Kosheleva, V. Kreinovich, J. Lorkowski, and M. Osegueda, "How to Transform Partial Order Between Degrees into Numerical Values", *Proceedings of the 2016 IEEE International Conferences on Systems, Man, and Cybernetics SMC'2016*, Budapest, Hungary, October 9–12, 2016.

[5] O. Kosheleva, V. Kreinovich, M. Osegueda Escobar, and K. Kato, "Towards the Most Robust Way of Assigning Numerical Degrees to Ordered Labels, With Possible Applications to Dark Matter and Dark Energy", *Proceedings of the 2016 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2016*, El Paso, Texas, October 31 – November 4, 2016.

[6] J. Lorkowski and V. Kreinovich, "Interval and Symmetry Approaches to Uncertainty – Pioneered by Wiener – Help Explain Seemingly Irrational Human Behavior: A Case Study", *Proceedings of the 2014 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2014*, Boston, Massachusetts, June 24–26, 2014.

[7] J. Lorkowski and V. Kreinovich, "Likert-type fuzzy uncertainty from a traditional decision making viewpoint: how symmetry helps explain human decision making (including seemingly irrational behavior)", *Applied and Computational Mathematics*, 2014, Vol. 13, No. 3, pp. 275–298.

[8] J. Lorkowski and V. Kreinovich, "Granularity Helps Explain Seemingly Irrational Features of Human Decision Making", In: W. Pedrycz and S.-M. Chen (eds.), *Granular Computing and Decision-Making: Interactive and Iterative Approaches*, Springer Verlag, Cham, Switzerland, 2015, pp. 1–31.

[9] J. Lorkowski and V. Kreinovich, "Fuzzy Logic Ideas Can Help in Explaining Kahneman and Tversky's Empirical Decision Weights", In: L. Zadeh et al. (Eds.), *Recent Developments and New Direction in Soft-Computing Foundations and Applications*, Springer Verlag, 2016, pp. 89–98.

[10] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.