

Discussion of “Statistical modelling of citation exchange between statistics journals” by Varin, Cattelan and Firth

Piotr Fryzlewicz
Department of Statistics
London School of Economics
Houghton Street
London WC2A 2AE
UK

May 1, 2015

I would like to congratulate the authors on an interesting and thought-provoking paper. My interest in the topic is partly related to my current role as Joint Editor of Series B of the Journal; however, I have never done academic research in this area and therefore my comments below are written from a non-expert’s perspective.

With the recent advances in text analysis, I believe we have the technology to analyse the ‘quality’ (‘weight’, ‘temperature’) of citations, as opposed to their mere number. For example, arguably, the citation in “other recent contributions include Anon (2015)” carries less weight, or has ‘lower temperature’ than that in “this work is motivated by Anon (2015)”. In the same vein, a citation to a paper made once in a manuscript may be ‘less hot’ than a citation to another paper made twice.

I like the concept of ‘exporting intellectual influence’, but I do not think that analysing academic citations only is an adequate way of measuring its strength. Many statistics papers are read by data scientists outside academia, which does not lead to citations. My belief is that one way to capture part of these missing data on the intellectual influence of papers is to equip papers posted online with discussion forums, permitting non-academic users to discuss these pieces of work. I am particularly encouraged to make this comment in light of the conjecture made by the authors of the “read paper effect” - if it is true that read papers “export more intellectual influence”, then why not ‘make every paper a discussion paper’ by enabling an online conversation about it?

In addition to citations within statistics, I believe analysing citations between statistics and other journals could be an informative way of evaluating statistics’ impact on other fields. Besides, I wonder to what extent the “health” of the discipline of statistics can be measured by comparing citations to statistics journals with citations to journals in neighbouring fields such as computer science, EEE or machine learning, and what lessons can be derived from such a comparison.

I will end with a brief comment regarding the methodology used. Fused lasso is known not to be the best tool for sequence segmentation (see e.g. Cho and Fryzlewicz (2011) and Rojas and Wahlberg (2014)), which makes me wonder if it is optimal or appropriate to use it to group rankings, as it is done in Section 5.5 of the paper.

References

- H. Cho and P. Fryzlewicz. Multiscale interpretation of taut string estimation and its connection to Unbalanced Haar wavelets. *Statistics and Computing*, 21:671–681, 2011.
- C. Rojas and B. Wahlberg. On change point detection using the fused lasso method. *Preprint*, 2014.