Author cocitation analysis of articles from "NeuroImage"

F. Å. Nielsen and L. K. Hansen

Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark

Introduction

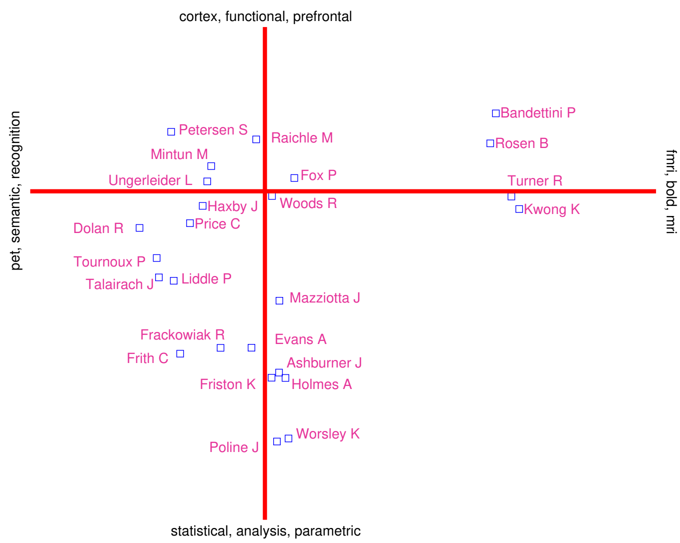
Author cocitation analysis (ACA) [1] is a method to describe a scientific field by analysis and visualization of the cocitation pattern in a corpus of scientific articles. Usual ACA works with data from the Science Citation Index of the Institute of Scientific Information (ISI) using specialized services to form AND-queries between two authors: An author/author-cocitation matrix (similarity matrix) is constructed by counting the number of articles that reference the two (first) authors. To restrict the number of queries the authors to be examined must be determined in advance and usually this involves asking experts to identify key authors. Typical analyses include multidimensional scaling and cluster analysis. Here we present a more *automated analysis* (authors do not have to be preselected and axes can be automatically labelled) working on a *indicator datamatrix* (rather than a cocitation matrix) with data from a *single journal* within the *functional neuroimaging* field.

Method

We downloaded the "References" webpages from of the "NeuroImage" journal (the website "Idealibrary"). Each of these webpages contain bibliographic information (author, abstract, title, ...) for a single article as well as the "out-links" (references). We included 325 articles in the analysis (1997 February - 2000 June) extracting the title (converting to lower case characters) as well as the all authors for each reference identified by the first initials and surname. This data was used to form an indicator datamatrix $\mathbf{X}(n \times p)$ with articles as rows and cited authors as columns with 1 in x_{ij} if the ith article (row) cited the jth author (column), 0 otherwise. We performed simple singular value decomposition on the datamatrix (corresponding to principal component analysis, metric multidimensional scaling or correspondence analysis): $\mathbf{USV}^{\mathsf{T}} = \mathsf{svd}[\mathbf{X}]$. We call the column vectors (principal components) of $\mathbf{U}(n \times r)$ and $\mathbf{V}(p \times r)$ *eigendocuments* and *eigenauthors*, respectively. To automatically label the eigenauthor axes an other indicator datamatrix $\mathbf{Y}(n \times q)$ was formed from the title of the articles, rows corresponding to articles and columns corresponding to title words. The title words were projected onto the eigenauthor space $\mathbf{Z}(q \times r)$ with stopwords ("of", "in", ...) excluded: $\mathbf{Z} = \mathbf{Y}^{\mathsf{T}}\mathbf{U}$. The axis labels for the ith eigenauthor is the elements with the highest score in the ith column of $\mathbf{Z}$.

Result

The $n = 325$ articles contained $p = 11558$ names. Some of these names referenced the same author, e.g. 11 versions refered to Richard S. J. Frackowiak: "Fracknowiak R", "Frackoviak R", ... We did not extract the mean of the matrix, thus the first principal component will (usually) describe those authors with the most references as is the case here with "Friston K" and "Frackowiak R" referenced the most. The component corresponds to "authoritative sources" as defined by [2]. The second principal component segregates between PET and fMRI with a clear fMRI-methodology cluster consisting of Kwong, Turner, Bandettini and Rosen. The axis labels are "fmri", "bold" and "mri" while "pet", "semantic" and "memory" are associated with the opposite direction. The third principal component has labels "mem-



ory", "prefrontal" and "working" in one end with authors such as Haxby and Glover and in the opposite end "motor", "human" and "activation" with, e.g., Mazziotta and Roland. The fourth principal component denotes a statistical/non-statistical dimension with the authors Poline and Worsley scoring high on the axis labeled "statistical", "analysis" and "parametric". Higher principal components are harder to interprete. The figure shows the 25 most cited authors projected on the 2nd and 4th eigenauthors ($\mathbf{v}_2$ and $\mathbf{v}_4$).

Conclusion

We have described the bibliometric method of author cocitation analysis of a single functional neuroimaging journal: "NeuroImage". The method allows us to find "eigenauthors" and "eigendocuments". We also devise a method for automatic labeling of the principal axes from the titles of the citing articles. We find that the main variation in the citation patterns are to be explained by articles describing fMRI versus PET. Other major variations were memory versus motor studies and statistical versus "non-statistical".

Acknowledgement

Danish Research Councils and Human Brain Project (P20 MH57180), NEC Research institute, Katherine McCain.

References

1. White, HD, Journal of the American Society for Information Science, 1981, 32:16–22.
2. Kleinberg, JM, Authoritative sources in a hyperlinked environment, Tech. Rep. 10076, IBM, 1997.