GLORIA GHENO

Innovative Data Analysis, Italy

AN NEW ALGORITHM FOR CITATION ANALYSIS

Abstract:

The bibliographic coupling and co-citation analysis methodologies were proposed in the early 60s and 70s to study the structure and the production of scientific communities. Bibliographic coupling is fundamental to understand the current state of a particular research area and its possible and potential future direction, while co-citation analysis is used to map the roots of academic works, fundamental to the development of a specific research field. With the first method, papers which have a common reference are paired and the strength of the link is given by the number of the references in common. With the second, instead, the papers co-cited by one or more documents are grouped. Both methodologies assume that papers, citing the same articles or cited from the same article, have similar aspects. Because these two methodologies have been considered separately until now, I propose a new algorithm, based on the bicluster analysis, which applies them together and I create an index to measure the similarity of the elements of the obtained clusters. Therefore, this new method groups together the bibliographically coupled papers and the co-cited references. In the obtained bicluster, the references grouped together represent the roots from which is born the trend to which the citing papers, grouped together, adhere. I apply this new method to economic papers, published between 2011 and 2020, which have "big data" among the keywords, so as to understand in a more exhaustive and rapid way how the current state and the future direction of the study of the big data are in the economic sector.

Keywords:

Bibliographic coupling, bicluster, big data, co-citation analysis

JEL Classification: C19, N01, O10