Abstract

Research ideation i.e. process of generating research ideas, is an integral part of the researcher's career. Typically, researchers have to access multiple resources to get information about recent research trends. Other sources for research trends identification include scientific literature, online databases and repositories, conferences and workshops, and industry and clinical collaborations. Most of these sources, where the physical participation of the researchers is mandatory, are not feasible for researchers due to time and cost constraints. However, literature databases and data repositories subjected to topic modeling are particularly suitable for this purpose due to their ability to uncover the main themes within the large corpus of data. Presently there are many methods used for topic modeling but Non-Negative Matrix Factorization (NMF) and Latent Dirichlet Allocation (LDA) are most efficient and popular. As field of bioinformatics is continuously evolving through interdisciplinary research and has resulted in millions of research articles available in bioinformatics repositories, so, in this study, the articles related to field bioinformatics have been retrieved from PubMed and have been analyzed using LDA and NMF algorithms to uncover the research trends. This study will help biologists with research interests in biology and bioinformatics to identify the field's current and future research direction.

Keywords: Topic modeling, Non-Negative Matrix Factorization, NMF, Latent Dirichlet Allocation, LDA, PubMed, Bioinformatics