Mining Scientific Trends Based on Topics in Conference Call for Papers

Abu Bakar[1[0000-0002-8006-8886]], Noor Arshad[1[0000-0002-3915-9153]]
Iqra Safder [1[0000-0001-9818-4693]] and Saeed-Ul Hassan[1[0000-0002-6509-9190]]

[1] Information Technology University, Ferozepur Road, Lahore 54000, Pakistan
saeed-ul-hassan@itu.edu.pk

1 Introduction

Ever since analyzing scientific topics and evolution of technology have become vital for researchers, academics, funding institutes and research administration departments, there is a crucial need to mine scientific trends to fill this appetite more rigorously. In this paper, we procured a novel Call for Papers (CFPs) dataset in order to analyze scientific evolution and prestige of conferences that set scientific trends using scientific publications indexed in DBLP. Using ACM CSS, 1.3 million publications that appear in 146 data mining conferences are mapped into different thematic areas by matching the terms that appear in publication titles with ACM CSS. In recent years, an attempt termed as Topic Detection and Tracking (TDT) [1] is made to find the solution for the problem of "well-awareness" on this dynamic data. As conference ranking has been made by different forums on the basis of mixed indicators[1]. ERA[2] ranks Australia's higher education research institutions. The major contributions of this paper are as follows: (i) compilation of CFPs dataset, (ii) identification of topics and keywords from CFP corpus, and (iii) measure the impact of these extracted hot topics from CFPs.

2 Datasets, Methodology and Results

We compiled a temporal CFPs corpus of 146 data mining conferences (FOR: 804) ranked by (ARC). The CFPs corpus is the main and novel dataset for our proposed research. Firstly, we extracted the publications metadata information such as article title and year for each paper in DBLP. Meanwhile, we extracted topic keywords with year from CFPs. Pre-processing techniques of text mining are applied. Furthermore, to measure the significance of terms, tf-idf weights [2] are used and we adopt following weight schemas: probabilistic, gram, relative, accumulative and hierarchal. For improved matching, we enrich vocabulary of terms by WordNet dictionary and Growbag dataset. Fig. 1 shows year wise frequent keywords from CFPs dataset.

[1] http://portal.core.edu.au/conf-ranks/

[2] http://universityrankings.com.au/research-excellence-rankings.html

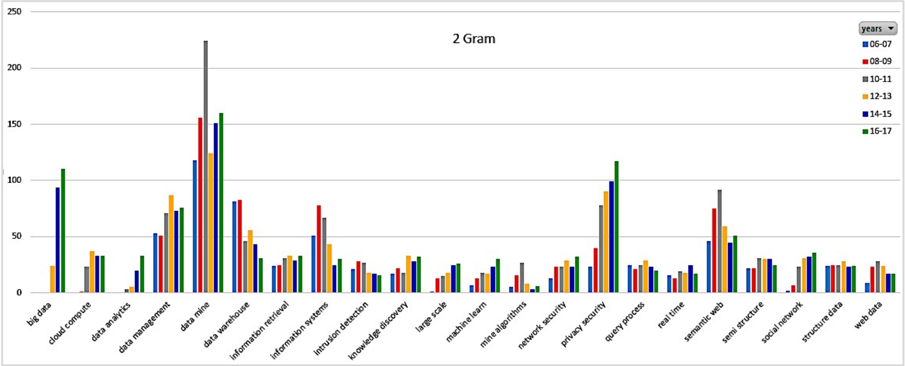

Fig. 1. Frequent keywords from CFPs Dataset - Year wise – 2 gram

Overall we observed that the term “Big Data” becomes frequent after 2011. It is also noticed that the rise of the term “Big Data” in 2012-2013 has pushed the popularity of the term “Data Warehouse” down during 2012 to 2017 along with decline of another hot term “Data Mining” during this era. Note that the term “Big Data” appeared in 2012-13 and immediately this term has got attention from research community. Moreover, the term “Big Data” is appeared as consistent focus of researchers since 2012. Along this, another hot area “Security Privacy” is constantly attracting the researchers since 2006.

3 Concluding Remarks

We mine scientific trends based on conference CFPs topics and publications. Our results indicate the rise of “big data” CFPs topics in past years. While “privacy security” show exponential increase, “semantic web” show a down fall. Moreover, we show that conferences from A* and A tier not merely set publication trends, since similar CFPs are targeted by B or C tier. We believe that analyzing scientific trends using CFPs datasets could help early career researchers. The proposed approach could help to identify trending conferences with respect to contribution in emerging topics. The data and code used can be downloaded from: https://github.com/slab-itu/research_trends/.

References

1. Yeh, J.F., Tan, Y.S. and Lee, C.H.: Topic detection and tracking for conversational content by using conceptual dynamic latent Dirichlet Allocation. Neurocomputing, 216, (2016).
2. Wu, H.C., Luk, R.W.P., Wong, K.F. and Kwok, K.L.: Interpreting tf-idf term weights as making relevance decisions. ACM Transactions on Information Systems (TOIS), 26(3), 13 (2008).