

IMPROVED METRICS FOR MEASURING THE RESEARCH CONTRIBUTION BASED ON ACADEMIC SOCIAL NETWORK

¹G.Ayyappan, ²C.Nalini

¹Research Scholar, Department of Computer Science and Engineering, Bharath University, Chennai

²Professor, Department of Computer Science and Engineering, Bharath University, Chennai

¹ayyappangmca@gmail.com, ²drnalinichidambaram@gmail.com

Abstract: An Important use of metrics for assessing the research contribution of research work in conferences. Contribution for research we focus quantity which can be measured by size of the outputs in each researcher or author, This paper proposes the individual authors contributes their research work based on the quantity of management research work in researcher society. It is challenging to trace out the imprints of evolution of research in general and also to filter the quantified results for measuring the individual's contribution in their research career.

Keywords: Author, Researcher, Frequency, Metrics, Cartesian Product.

1. Introduction

We measure the individual authors contributions based on quantity frequency in the dataset. We find the solution based on these methods. Tasleem Arif[1] proposed method use a token-based similarity score in this first stage of comparison and based on the results of the first stage it uses a character-based similarity score in the second stage. Experimental results obtained on standard datasets indicate that the proposed technique shows a lot of improvements over the existing methods. Thi Thanh Sang Nguyen et al.[2] proposed and developed the domain knowledge. It used an ontology to represent the domain knowledge. Then It used one automatically generated semantic network to represent domain terms, Web-pages, and the relations between them. Another new model, the conceptual prediction model, have proposed to automatically generate a semantic network of the semantic Web usage knowledge is the integration of domain knowledge and Web usage knowledge.

Jie Tang et al.[3] proposed to deal with the name ambiguity problem (several users with the same name) when integrating the profile information extracted from different sources. the extracted user profiles, and construct the user interest model.

Arantxa Duque Barrachina et al.[4]proposed a Proof of Concept (PoC) end to end solution that utilises the Hadoop programming model, extended ecosystem and the Mahout Big Data Analytics library for categorising similar support calls for large technical support data sets and size of the outputs .

Anand Kumar et al.[5]have developed a prototype of DCMS based on the PostgreSQL system and experiments using real MS data and workload show that DCMS significantly outperforms existing MS softwares. They used it as a platform to test other data management issues such as security and compression.

N K Nagwani [6]proposed and developed technique using semantic similarity based clustering and topic modeling using Latent Dirichlet Allocation (LDA) for summarizing the large text collection over MapReduce framework system.

Kunal Goswami et al.[7]proposed and developed capturing the user social interaction behavior to identify fraud. The problem being solved is one of the characteristics that lead to fraud rather than detecting fraud.

Pekka Pääkkönen et al.[8] proposed and developed an evaluation of data quality management architecture for social media data. Data quality evaluation of Twitter-based data sets. quality management in social media data has been extended and evaluated based on the implementation architecture. The results pointed out that 150–800 tweets/s can be evaluated with two cloud nodes depending on the configuration.

Haoyang Wu et al.[9]proposed and developed data of several types of agricultural commodities and error analysis is made. The result shows that the mixed model is more accurate in forecasting agricultural commodity prices than each single model does, and has better accuracy in warning values. The mixed model, to some extent, forecasts the daily price changes of agricultural commodities.

Here, the dataset and contribution of dataset in section 2, we proposed method for finding individual frequency in section 3, the information about the result and discussion of various classifiers approaches in section 4 represents. We present the conclusion in section 5.

2. Dataset Description

We collect the data from Aminer Academic social network dataset i.e., Topic-paper-author dataset Nodes:33739 authors of 5 topics, Edges:139278 coauthor relationships, Description: Created for cross domain recommendation, DataType: Numeric and Text. We find the measure based on quantity of individual authors contribution on conference from 1990 to 2005 and various domain like as data mining, database, visualization, Theory, Medical Informatics.

3. Materials And Methods

Here we used Microsoft Excel MS Excel: frequency methods or count options using FREQUENCY Function (WS)

The syntax for the FREQUENCY function in Microsoft Excel is:

FREQUENCY(data, intervals)

We apply Cartesian product concepts in all over the dataset and find the individual authors **frequency**.

4. Experiments And Results

We apply Cartesian product or cross product concepts in Topic-Paper-Author dataset we split in to each and every author. The length of authors' position from first author position to thirty second authors position. We avoid redundancy and find there are 22875 individual authors contributions and find top 20 individual authors contributions on conferences from the period of 1990 to 2005 and five topics.

Table 1. Threshold 50 to 30 Individual frequency of authors

Authors	Individual Frequency
Gerhard Weikum	49
Piotr Indyk	48
Beng Chin Ooi	47
Noga Alon	47
Avi Wigderson	46
Hongjun Lu	46
Jon M. Kleinberg	46
Nick Koudas	46
Joseph M. Hellerstein	45
Kian-Lee Tan	45

Wei Wang	45
Rajeev Rastogi	44
Prabhakar Raghavan	43
Jeffrey Scott Vitter	42
Shree K. Nayar	42
Kenneth A. Ross	41
Madhu Sudan	41
Baruch Awerbuch	40
Laks V. S. Lakshmanan	40
C. Mohan	39
Dennis Shasha	39
Hamid Pirahesh	39
Sudipto Guha	39
Heikki Mannila	38
Abraham Silberschatz	37
Alan M. Frieze	37
Jian Pei	37
Moni Naor	37
Ravi Kumar	36
Sanjeev Khanna	36
Umeshwar Dayal	36
Vimla L. Patel	36
Yuval Rabani	36
Christos H. Papadimitriou	35
David B. Lomet	35
David Eppstein	35
David W. Bates	35
Haixun Wang	35
Max A. Viergever	35
Serge Abiteboul	35
Uri Zwick	35
Yannis E. Ioannidis	35
Éva Tardos	35
Daniel A. Keim	34
David R. Karger	34
Divyakant Agrawal	34
Frank Thomson Leighton	34
Krithi Ramamritham	34
Mikkel Thorup	34
Moses Charikar	34
Raymond T. Ng	34
Takeo Kanade	34
Bernd Hamann	33
Charu C. Aggarwal	33
Dan Suciu	33
Elke A. Rundensteiner	33
Johannes Gehrke	33
Joseph Naor	33
M. Tamer Özsu	33
Marianne Winslett	33
Pankaj K. Agarwal	33
Stanley B. Zdonik	33

Eamonn J. Keogh	32
Oded Goldreich	32
Russell Impagliazzo	32
S. Sudarshan	32
Thomas S. Huang	32
Avrim Blum	31
Kyuseok Shim	31
Ming-Syan Chen	31
Uriel Feige	31
Amos Fiat	30
Christian S. Jensen	30
Ke Wang	30
Ling Liu	30
Philip A. Bernstein	30
Santosh Vempala	30
Thomas Ertl	30

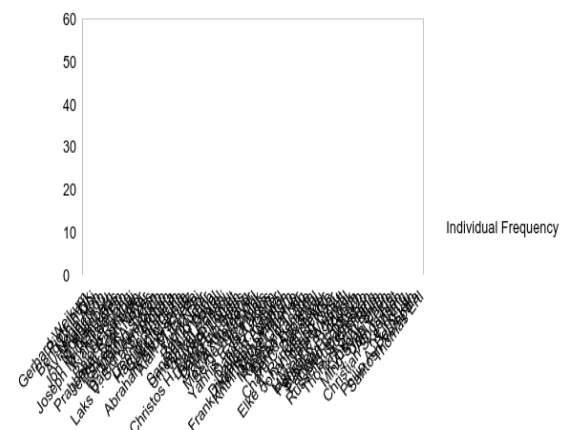


Figure 1. Authors Vs.Individual Frequency of 50-30

We fix the threshold below 50 to 30 for the count of individual frequency in TopicPaperAuthor dataset. We find top 79 authors of individual frequency who are all contribute the conference in between 1990 to 2005 in five topics.

5. Conclusion

Social Networks like ‘AMiner’ helps the miners to go for such direction and they obtain results to throw more results. There are types of researchers initiating (originating, by authoring), collaborating (by coauthoring), and supporting (by reference or citations). In this paper we aim at novel estimation for the absolute frequency of individual contribution and relative weighted frequency with respect to their collaboration also. In future, we estimate and evaluate authors contributions based on H index and I index of each authors using social networks

References

- [1] Tasleem Arif, Exploring , The Use Of Hybrid Similarity Measure For Author Name Disambiguation, INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 4, ISSUE 12, DECEMBER 2015 ISSN 2277-8616, pp:171-175.
- [2] Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu, Web-Page Recommendation Based on Web Usage and Domain Knowledge, IEEE Transactions on Knowledge and Data Engineering (Volume: 26, Issue: 10, Oct. 2014),Page(s): 2574 - 2587,Page(s): 2574 - 2587Print ISSN: 1041-4347.
- [3] Jie Tang, Limin Yao, Duo Zhang, and Jing Zhang, A Combination Approach to Web User Profiling ,ACM Transactions on Knowledge Discovery from Data, Vol. V, No. N, March 2010, Pages 1–38.
- [4] Arantxa Duque Barrachina and Aisling O’Driscoll, A big data methodology for categorising technical support requests using Hadoop and Mahout, Duque Barrachina and O’Driscoll Journal of Big Data 2014, 1:1 <http://www.journalofbigdata.com/content/1/1/1>
- [5] Anand Kumar, Vladimir Grupcev, Meryem Berrada, Joseph C Fogarty, Yi-Cheng Tu, Xingquan Zhu, Sagar A Pandit and Yuni Xia, DCMS: A data analytics and management system for molecular simulation, Journal of Big Data 2014, 1:9, <http://www.journalofbigdata.com/content/1/1/9>.
- [6] N K Nagwani, Summarizing large text collection using topic modeling and clustering based on MapReduce framework, DOI 10.1186/s40537-015-0020-5.
- [7] Kunal Goswami, Younghee Park and Chungsik Song, Impact of reviewer social interaction on online consumer review fraud detection, DOI 10.1186/s40537-017-0075-6.
- [8] Pekka Pääkkönen^{1*} and Juha Jokitalo², Quality management architecture for social media data, DOI 10.1186/s40537-017-0066-7.
- [9] Haoyang Wu, Huaili Wu, Minfeng Zhu, Weifeng Chen and Wei Chen, A new method of large-scale short-term forecasting of agricultural commodity prices: illustrated by the case of agricultural markets in Beijing, DOI 10.1186/s40537-016-0062-3
- [10] <https://www.techonthenet.com/excel/formulas/frequency.php>
- [11] <https://aminer.org/data>

