Decision Support for Hydrological Research Funding Agencies Using AI

Mashrekur Rahman[a]

[a]*Department of Land, Air and Water Resources, University of California, Davis, One Shields Avenue, Davis, 95616, CA, USA*

Abstract

Global dependence on citation-based measures of scientific research impact is perpetuating long-existing academic biases. Stakeholders of hydrological research (e.g., grant funders and program managers) often make crucial decisions about directing their resources towards certain research endeavors based on popularity, including citation records, of previous research. I propose a framework for assessing the future impact of hydrological research that takes into account multiple (additional) factors of impact including probabilistic distributions of subtopics within research papers, interdisciplinarity of topics in publication, venues of publication, direct social (research-into-use) implications, and author credibility. I use Natural Language Processing (NLP) to leverage large amounts of textual data in combination with quantitative data and an ensemble of models (Latent Dirichlet Allocation, Attention and Random Forest Regression) to generate "democratic" normalized impact scores. Preliminary results suggest our framework is able to capture early-warning signals for impactful hydrological research. My long-term goal with this project is to help funding decision makers of hydrological science make more

informed decisions based on modern data science.

Interdisciplinarity research has been identified as one of the ways to solve the world's biggest problems [1]. However, the academia continues to be strangled by traditional stereotypes: interdisciplinary proposals are less likely to receive funding [2] and institutions continue to enable this discrimination [1]. Brown et al. [3] suggested multiple solutions to catalyse cross-discipline collaboration, such as cultivating disciplinary researchers who look beyond their training, providing institutional support, and bridging research, policy and practice. Interdisciplinarity has been shown to have positive influence on both academic and broader (socio-economic) impact [4]. In this literature review, we look at the state of interdisciplinarity in hydrological sciences and justify the need for a model framework which leverages big (textual) data to capture early signals of impactful research.

On the question of interdisciplinarity's impact on citation count, Larivière and Gingras [5] found that it varies between disciplines and that highly disciplinary and highly interdisciplinary articles have a low scientific impact. Long-distance interdisciplinarity i.e. papers citing references from subdisciplines positioned far apart attract highest relative citation counts. Wang et al. [6] suggested variety and disparity in scientific articles lead to positive effects on long-term citations and argued that a long citation time window is required to avoid systematically underestimating the impact of interdisciplinary research articles.

Rahman et al. [7] found that although interdisciplinarity in hydrology is a bottom-up effect, and more needs to be done from the top institutions and stakeholders to ensure increasing cross-discipline collaboration. Publication and citation data have traditionally served as valuable parameters in science and technology evaluation by policy-makers [8]. Learning the early signals of impactful research may prove valuable to the hydrology community. As of now, no such tools exist to guide policy makers, funding agencies, program managers, practitioners, and other stakeholders at all levels.

1. Objectives

In this study, I aim to develop a multi-step Artificial Intelligence (AI) based framework for *(i) learning the early signals of impactful articles in hydrological sciences, and in the long-term (ii) predicting the citation counts for articles.*

2. Methodology

The corpus that I will use consists of abstracts from all peer-reviewed articles published in eighteen water science journals between 1991 and 2019 - this is all water science journals with a 2018 Impact Factor (IF) of greater than 0.9 (Scimago Journal and Country Rank) (Figure 1). In total, 74,479 article-abstracts were acquired from the Web of Science core collection in the form of bib files. All data will be preprocessed using multiple cleaning routines and converted to canonical format for model training purposes.

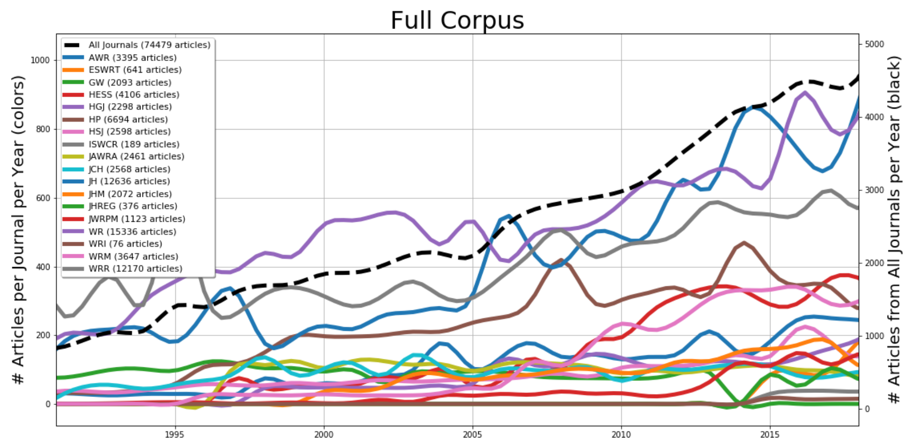

Figure 1: Number of articles published per year between 1991 and 2019 in 18 major water research journals (Source: Web of Science)

The model framework (Figure 2) is an ensemble of a dynamic topic model [9] and a feed forward neural network [10]. In the first step, I will train the dynamic topic model on the corpus and extract the posterior expectations of the model, which includes document-topic distributions and word-topic associations. Shannon entropy is one of the most widely used indicators of interdisciplinarity of journals and articles. Carusi and Bianchi [11] used Shannon entropy as one of the measures of interdisciplinarity in 1258 journals in the field of information and communication technology.

Silva et al. [12] assessed the interdisciplinarity of scientific journals using entropy, and found that entropy-based measurement of interdisciplinarity correlates well with impact factors and citation counts. A previous study [13] conducted an interdisciplinarity assessment for Informatics journals using

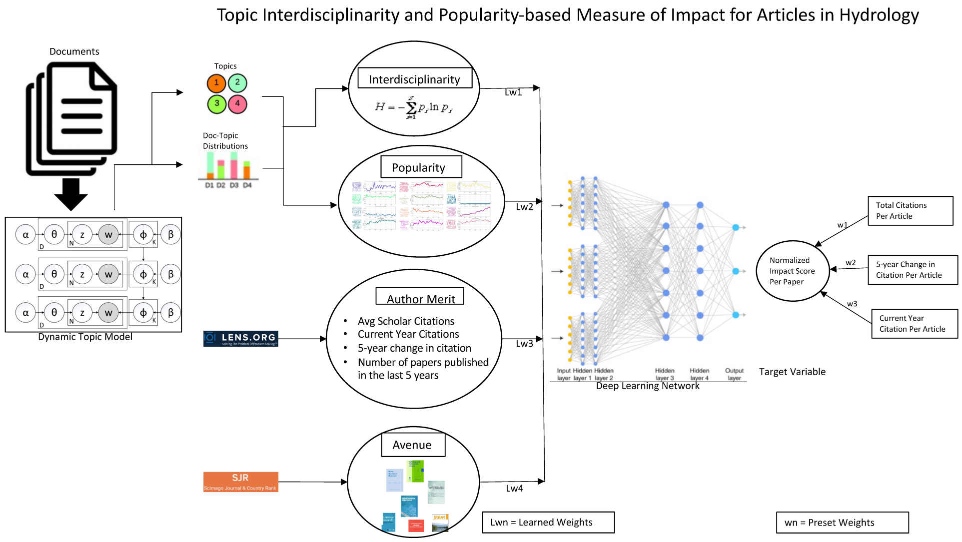

Figure 2: Conceptual model

Topic Modeling with Shannon entropy as a diversity metric. Entropy has been used to measure interdisciplinarity of researchers and research topics [14], research proposals [15], and collaborations [16]. I will therefore use the entropy based diversity metric applied to topic distributions as a primary measure of interdisciplinarity at corpus and article levels.

2.1. Measuring Interdisciplinarity at the Article Level

I used Shannon Diversity to measure the interdisciplinarity per article H_d for each article in our corpus as:

$$H_d = -\sum_{k=1}^{K}(\mu log(\mu)), \tag{1}$$

Where μ is the distribution of topics over document d.

Interdisciplinarity will be one of the four groups of input features for this framework. The other three are relative popularity of topics, merit of the avenue of publication, and merit of the author(s) of the articles. Popularity will be calculated based on the topic trends, while data on avenue merit will be acquired from Scimago Journal Country Ranking, and data on author merit will be downloaded from Lens.org. The potential target variables for training the neural network will be the total citation count for each paper, and n-year change in citations for those papers. After the model is trained and satisfactory performance is achieved, the learned weights can be used to identify early signals of impactful hydrological research.

3. Preliminary Results

As an example of how interdisciplinarity and academic impact can be connected, I investigated the effect of funding diverse research topics on academic impact. Figure 3 shows the major hydrology research funding agencies and the mean scholarly citation per paper resulting from the projects they funded.

Specifically, I looked at the relationship between topic diversity and academic impact of funded research articles by major U.S., Canadian, European, and Australian agencies (Figure 4).

It can be generally observed that more topically diverse research led to higher average scholarly citations per article. This finding serves as one of the preliminary results motivating the development of this framework.

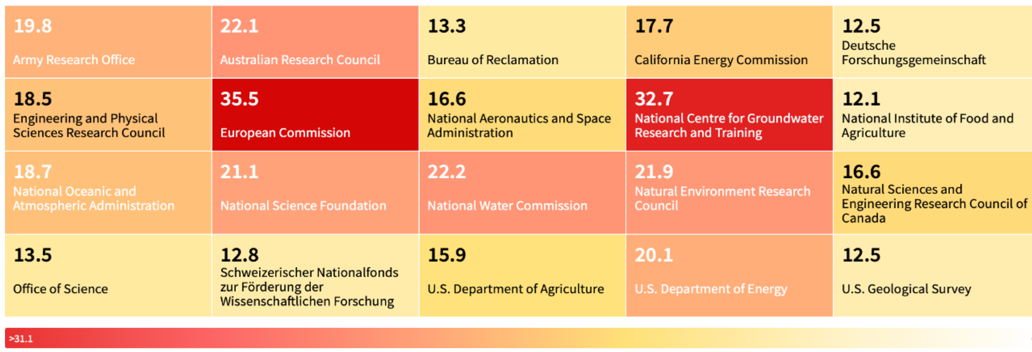

Figure 3: Major hydrology research funding agencies and the mean scholarly citation per paper resulting from the projects they funded

Further analysis is beyond the scope of this term project in terms of the time required. I expect the results from the long-term objectives of this project to be of importance to the stakeholders of hydrological science.

4. Future Direction

Once developed, this data-driven framework can serve as a tool for stakeholders at various levels to identify signals which lead to higher academic impact in the future. Such a tool will also serve towards minimizing long-standing academic biases, making it easier for newer, diverse, and more innovative ideas to be funded.

References

[1] H. Ledford, How to solve the world's biggest problems, Nature News 525 (2015) 308.

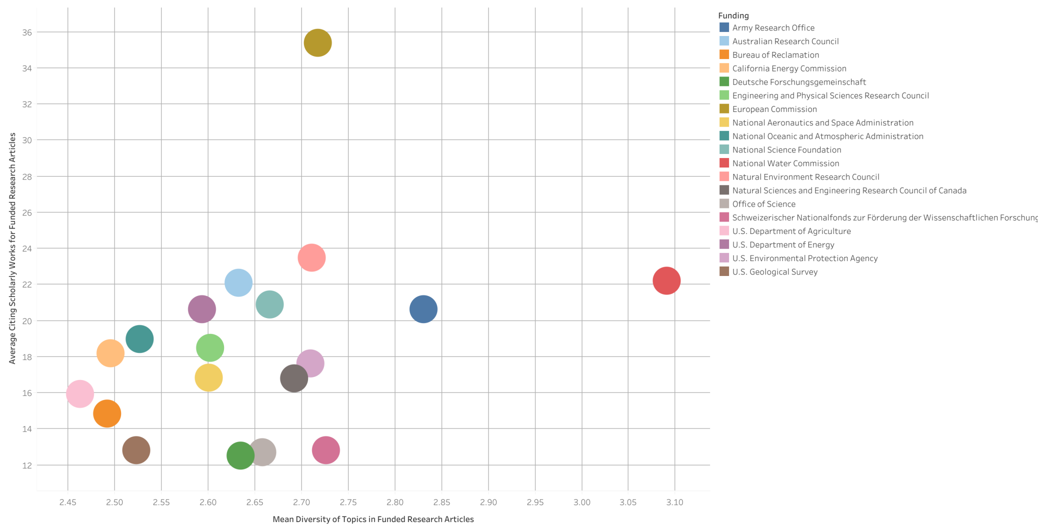

Figure 4: Relationship between topic diversity and academic impact in funded research articles

[2] L. Bromham, R. Dinnage, X. Hua, Interdisciplinary research has consistently lower funding success, Nature 534 (2016) 684–687.

[3] R. R. Brown, A. Deletic, T. H. Wong, Interdisciplinarity: how to catalyse collaboration, Nature News 525 (2015) 315.

[4] L. Zhang, B. Sun, L. Jiang, Y. Huang, On the relationship between interdisciplinarity and impact: Distinct effects on academic and broader impact, Research Evaluation (2021).

[5] V. Larivière, Y. Gingras, On the relationship between interdisciplinarity and scientific impact, Journal of the American Society for Information Science and Technology 61 (2010) 126–131.

[6] J. Wang, B. Thijs, W. Glänzel, Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity, PloS one 10 (2015) e0127298.

[7] M. Rahman, J. Frame, J. Lin, G. Nearing, Water science is becoming more interdisciplinary (2021).

[8] E. Garfield, A. Welljams-Dorof, Citation data: their use as quantitative indicators for science and technology evaluation and policy-making, Science and Public Policy 19 (1992) 321–327.

[9] D. M. Blei, J. D. Lafferty, Dynamic topic models, in: Proceedings of the 23rd international conference on Machine learning, 2006, pp. 113–120.

[10] G. Bebis, M. Georgiopoulos, Feed-forward neural networks, IEEE Potentials 13 (1994) 27–31.

[11] C. Carusi, G. Bianchi, A look at interdisciplinarity using bipartite scholar/journal networks, Scientometrics 122 (2020) 867–894.

[12] F. N. Silva, F. A. Rodrigues, O. N. Oliveira Jr, L. d. F. Costa, Quantifying the interdisciplinarity of scientific journals and fields, Journal of Informetrics 7 (2013) 469–477.

[13] S. A. Jin, M. Song, Topic modeling based interdisciplinarity measurement in the informatics related journals, Journal of the Korean Society for information Management 33 (2016) 7–32.

[14] H. Sayama, J. Akaishi, Characterizing interdisciplinarity of researchers and research topics using web search engines, Plos One 7 (2012) e38747.

[15] M.-G. Seo, S. Jung, K.-m. Kim, S.-H. Myaeng, Computing interdisciplinarity of scholarly objects using an author-citation-text model., in: BIR@ ECIR, 2017, pp. 62–72.

[16] T. Bergmann, R. Dale, N. Sattari, E. Heit, H. S. Bhat, The interdisciplinarity of collaborations in cognitive science, Cognitive Science 41 (2017) 1412–1418.