International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

Institutional association commendation: An expertise-based framework using NLP and Network examination

Rajeev Ranjan, Lokendra singh Songare

PG Scholar, CSED, Dr. APJ Abdul Kalam University Indore, M.P., India

Assistant Professor, CSED, Dr. APJ Abdul Kalam University Indore, M.P., India

ABSTRACT

The shift from 'trust-based funding' to 'performance-based funding' is one of the factors that has forced institutions to strive for continuous improvement of performance. Several studies have established the importance of collaboration in enhancing the performance of paired institutions. However, identification of suitable institutions for collaboration is sometimes difficult and therefore institutional collaboration recommendation systems can be vital. Currently, there are no well-developed institutional collaboration recommendation systems. In order to bridge this gap, we design a framework that recognizes thematic strengths and core competencies of institutions, which can in turn be used for collaboration recommendations. The framework, based on NLP and network analysis techniques, is capable of determining the strengths of an institution in different thematic areas within a field and thereby determining the core competency and potential core competency areas of that institution. It makes use of recently proposed expertise indices such as x and x(g) indices for determination of core and potential core competency areas and can toss two kinds of recommendations: (i) for enhancement of strength of strong areas or core competency areas of an institution and (ii) for complementing the potentially strong areas or potential core competency areas of an institution. A major advantage of the system is that it can help to determine and improve the research portfolio of an institution within a field through suitable collaboration, which may lead to the overall improvement of the performance of the institution in that field. The framework is demonstrated by analyzing the performance of 195 Indian institutions in the field of 'Computer Science'. Upon validation using standard metrics for novelty, coverage and diversity of recommendation systems, the framework is found to be of sufficient coverage and capable of tossing novel and diverse recommendations. The article thus presents an institutional collaboration recommendation system which can be used by institutions to identify potential collaborators.

Key Words: Institutional Collaboration, Recommendation System, NLP, Network Analysis, Research Expertise, Expertise indices

INTRODUCTION

Institutional organization, being one of the three major organizations of science (the other two are intellectual and social organizations) has played a pivotal role in the progress of science. Funding is one of the major fuels of R&D activities. An early focus of funding based on 'trust' outlined the importance of the role of institutions in scientific progress. However, in the past few decades, a shift from 'trust-based' funding to 'performance-based' funding (Sörlin, 2007) forced funding agencies to adopt sharp performance assessment methods. At the same time, this shift also forced institutions to strive for continuous improvement of performance. The rise of many international ranking systems such as QS, THE, ARWU, CWTS, etc., is a natural consequence of the shift to 'performance-based funding' and several funding agencies rely on these rankings. Some funding agencies prefer 'thrust area performance' as a yardstick for fund allocation. National agencies in many countries are entrusted to formulate national strategies for nurturing institutions of excellence in thrust/priority areas. For instance, back in 2006, an Indian working group on thrust areashand-picked cyber security, multi-scale modelling, Quantum theory and applications, etc., as some of the thrust areas in engineering sciences. The establishment of the Interdisciplinary Cyber Physical Systems (ICPS) division by the Department of Science and Technology (DST), Govt. of India is another evidence for the increasing emphasis on 'thrust area performance-based' funding. When it comes to 'performance-based funding' or 'thrust area performance-based funding', institutions are always required to remain innovative and relevant.

In the case of an institution, its scholarly contributions might span over many fields. A research field comprises of many thematic areas. For example, for the subject 'Computer Science', one can consider 'Software Engineering', 'Data Science', 'NLP', 'Data Mining' etc. as some important thematic areas. The level of contribution of an institution with respect to a thematic area in a given field might vary, or in other words the institution may have varied 'thematic research strengths'

OBJECTIVE OF THE WORK

Determination of research strengths of an institution in different thematic areas (and thereby determining the core competency and potential core competency areas) is the key task of the framework. Retrieval of suitable recommendations based on previously mentioned diligently designed strategies 1 and 2 is the second major task. Thus, our recommendation system framework has two sections-(i) the expertise determination section and (ii) the recommendation retrieval section. A schematic diagram of the framework showing both the sections is given in

Determination of the thematic strengths of an institution is the first step. For that, an institution's scientific publications have to be mapped to respective thematic areas and then metrics that reflect the performance of an institution with respect to thematic areas have to be decided and method(s) to compute these scores are to be designed. The foremost concern regarding this is how to determine the thematic areas of research? As scientific literature can be treated as a body of knowledge and several levels of representation of knowledge is possible for it viz., level of thematic areas/subfields, fields of research, major disciplines or broad subjects, there is no proper way or hard and fast rules to define the confinements of each level.

Result:-

The indicative performance of our recommendation system can be summarized as shown in table 5. As a whole, our system can be treated as capable of tossing novel and diverse recommendations that are also of satisfactory intra-set and inter-set coverage. The ability of the recommendation system to toss three sets of recommendations ensures the high performance of the system with respect to evaluation metrics except for Inter-set coverage. However, as discussed earlier, from the computed inter-set coverage scores between strategy-1 retrieval, strategy-2 (high priority) retrieval and strategy-2 (low priority) retrieval, the exhibition of moderate inter-set dissimilarity scores between Strategy-1 and Strategy-2 (high priority) recommendations alone will not be a vital indication of compromised performance in terms of coverage.

An interesting observation is that number of scores termed as 'Very High' seems to decline (Strategy-2 (Low Priority) > Strategy-2 (High Priority)> Strategy-1) as the goal

of recommendation retrieval turns more ambitious. This can be treated as a reflection of the gradient of difficulty that might occur in the real-world (a kind of indirect issuance of a caution note) to engineer more ambitious collaborations/ties. Thus, an advantage of our system over systems that tosses ambitious recommendations is that our system can toss ambitious recommendations and relatively less ambitious recommendations by appropriately cautioning institutional level decision-makers about the level of difficulty that may incur while proceeding to engineer the recommended collaborations, without compromising the level of performance or maintaining asubstantial level of performance in terms of novelty, coverage and diversity.

Conclusion-

Institutional organization of science and its influence and dependency on other organizations of science such as social organization and intellectual organization of science is vital for the advancement of science. As funding of research happens largely at an institutional level than at any other level, the role and responsibility of institutions to scaleup their contribution in quantity and quality towards scientific progress is steadily increasing. Together with this, the shift witnessed in funding patterns from ‘trust-based funding’ to ‘performance-based funding’ forced institutions to devise effective strategies to enhance their performance. Collaboration is known to be one of the effective measures to enhance the performance of partnering institutions. However, the determination of a suitable partner for collaboration is a key factor that determines the success of collaborative ventures. As it is not always an easy task, the importance of recommendation systems lies there. Though scientific literature related to collaboration studies is quite rich and includes many studies related to the development of collaboration recommendation systems for academia using a multitude of approaches including network approach, machine learning, etc., and their combinations, most of these are concentrated on individual/author level collaborations. There are a few studies on co-institutional relationship patterns using networks. But the development of institutional collaboration recommendation systems is almost unexplored. We attempt to bridge this gap by

an attempt to design an institutional collaboration recommendation system framework using NLP and a network approach that can be fully developed and implemented as a functional recommendation system. Our framework is not free from certain limitations and that serve as an opportunity for further improvement. The multi-layered NLP module has ensured the reduction of ambiguities to a great extent but still, some author keywords that cannot represent thematic areas are successfully evading the NLP processing pipeline. Usage of more advanced NLP techniques when they materialize might improve the performance of the NLP module by ensuring better processing of such keywords. As the framework is capable of utilizing metrics other than citations such as altmetric scores for determining thematic strengths, such endeavours can also be attempted. These are some of the possible improvements of the computational framework that forms the backbone of the full-fledged recommendation system. When it comes to the (software) development and implementation of the system, advanced 'software architecture concepts' such as the one given by Elammari & Elfrjany (2012) that ensures reduced complexity will be considered.

REFERENCES

1. Abramo, G., D'Angelo, C. A., & Di Costa, F. (2009). Research collaboration and productivity: is there correlation? Higher education, 57(2), 155-171. Adams, J. (2012). The rise of research networks. Nature, 490(7420), 335-336.
2. Afolabi, I. T., Ayo, A., & Odetunmibi, O. A. (2021). Academic Collaboration Recommendation for Computer Science Researchers Using Social Network Analysis. Wireless Personal Communications, 121(1), 487–501. Springer US. Retrieved from https://doi.org/10.1007/s11277-021-08646-2
3. Ali, Z., Qi, G., Muhammad, K., Ali, B., & Abro, W. A. (2020). Paper recommendation based on heterogeneous network embedding. Knowledge-Based Systems, 210, 106438. Elsevier B.V. Retrieved from https://doi.org/10.1016/j.knosys.2020.106438
4. Alshareef, A. M., Alhamid, M. F., & El Saddik, A. (2019). Academic Venue Recommendations Based on Similarity Learning of an Extended Nearby Citation Network. IEEE Access, 7, 38813–38825. IEEE.
5. Ammar, W., Groeneveld, D., Bhagavatula, C., Beltagy, I., Crawford, M., Downey, D.,

Dunkelberger, J., Elgohary, A., Feldman, S., Ha, V., Kinney, R., Kohlmeier, S., Lo, K., Murray, T., Ooi, H. H., Peters, M., Power, J., Skjonsberg, S., Wang, L. L., … Etzioni, O. (2018). Construction of the literature graph in semantic scholar. NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 3, 84–91. https://doi.org/10.18653/v1/n18-3011 Ashraf, I., Hur, S., & Park, Y. (2020). An Investigative Analysis on Finding Patterns in Co-Author and Co-Institution Networks for LIDAR Research. INTERNATIONAL ARAB JOURNAL OF INFORMATION TECHNOLOGY, 17(6), 875-884. Batagelj, V. (2012). Social network analysis, large-scale. In A. Robert Meyers (Ed.), Computational complexity: Theory, techniques, and applications (pp. 2878–2897). New York: Springer. Bouraga, S., Jureta, I., Faulkner, S., & Herssens, C. (2014). Knowledge-based recommendation systems: A survey. International Journal of Intelligent Information Technologies (IJIIT), 10(2), 1-19.

6. Brandão, M.A., Moro, M.M., Lopes, G.R., Oliveira, J.P. (2013). Using link semantics to recommend collaborations in academic social networks, in Proceedings of the 22nd International Conference on World Wide Web, pp. 833-840.

7. Canto, I., & Hannah, J. (2001). A partnership of equals? Academic collaboration between the United Kingdom and Brazil. Journal of Studies in International Education, 5(1), 26-41.

8. Caselles-Dupré, H., Lesaint, F., & Royo-Letelier, J. (2018). Word2vec applied to Recommendation: Hyperparameters matter. RecSys 2018 - 12th ACM Conference on Recommender Systems, 352–356. https://doi.org/10.1145/3240323.3240377