

# A Framework for Paper Submission Recommendation System

Dinh V. Cuong\*  
Dac H. Nguyen\*  
Son Huynh\*  
AISIA Research Lab  
Ho Chi Minh, Vietnam

Phong Huynh  
AISIA Research Lab  
VNU HCM - University of Science  
Ho Chi Minh, Vietnam

Cathal Gurrin  
Dublin City University  
Dublin, Ireland

Minh-Son Dao  
National Institute of Information and  
Communications Technology  
Tokyo, Japan

Duc-Tien Dang-Nguyen<sup>†</sup>  
Department of Information Science  
and Media Studies  
University of Bergen  
Bergen, Norway

Binh T. Nguyen<sup>‡</sup>  
AISIA Research Lab  
VNU HCM - University of Science  
Ho Chi Minh, Vietnam

## ABSTRACT

Nowadays, recommendation systems play an indispensable role in many fields, including e-commerce, finance, economy, and gaming. There is emerging research on publication venue recommendation systems to support researchers when submitting their scientific work. Several publishers such as IEEE, Springer, and Elsevier have implemented their submission recommendation systems only to help researchers choose appropriate conferences or journals for submission. In this work, we present a demo framework to construct an effective recommendation system for paper submission. With the input data (the title, the abstract, and the list of possible keywords) of a given manuscript, the system recommends the list of top relevant journals or conferences to authors. By using state-of-the-art techniques in natural language understanding, we combine the features extracted with other useful handcrafted features. We utilize deep learning models to build an efficient recommendation engine for the proposed system. Finally, we present the User Interface (UI) and the architecture of our paper submission recommendation system for later usage by researchers.

## CCS CONCEPTS

• **Computing methodologies** → **Machine learning algorithms; Neural networks; Feature selection;** • **Information systems** → **Recommender systems.**

## KEYWORDS

recommendation system, paper submission, deep learning

\*All authors contributed equally to this research.

<sup>†</sup>Senior author

<sup>‡</sup>Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

ICMR '20, June 8–11, 2020, Dublin, Ireland

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7087-5/20/06...\$15.00

<https://doi.org/10.1145/3372278.3391929>

## ACM Reference Format:

Dinh V. Cuong, Dac H. Nguyen, Son Huynh, Phong Huynh, Cathal Gurrin, Minh-Son Dao, Duc-Tien Dang-Nguyen, and Binh T. Nguyen. 2020. A Framework for Paper Submission Recommendation System. In *Proceedings of the 2020 International Conference on Multimedia Retrieval (ICMR '20)*, June 8–11, 2020, Dublin, Ireland. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3372278.3391929>

## 1 INTRODUCTION

One of the most critical responsibilities of researchers is publishing their latest findings to the scientific community. During the last decades, the number of journals and international conferences is increasing dramatically due to the rapid growth of both science and technology. Up to now, there have been over 13,000 publication venues in computer science [1]. Each place has its specific topics and scopes, and sometimes, organizers or editors update its research topics to attract more submissions from various scientists. For this reason, researchers have a vast array of options for their paper submission and must try to select the most suitable one that is appropriate to the quality of their scientific work as well as increasing the chance of acceptance. This scenario may even occur for researchers who want to explore new domains, work on interdisciplinary projects, or simply don't have much experience. Subsequently, an efficient paper submission recommendation system is essential to assist authors by providing useful information and suggesting the most relevant journals or conferences for a given manuscript. Recently, there exists several paper submission recommendation systems from different publishers. Publishers such as IEEE<sup>1</sup>, Springer<sup>2</sup>, and Elsevier<sup>3</sup> created their systems to help authors quickly determine conferences and/ or journals for submissions. Similarly, domain-specific systems PubMender<sup>4</sup> [2] and Wang et al.<sup>5</sup> [1] focus on recommending within a specific field of research.

In this work, we present a new framework for building a paper submission recommendation system for two research domains; mathematics and computer science. We propose different types of features by applying Chi-square statistics, Term Frequency-Inverse

<sup>1</sup><https://publication-recommender.ieee.org/home>

<sup>2</sup><https://journalsuggester.springer.com/>

<sup>3</sup><https://journalfinder.elsevier.com/>

<sup>4</sup><https://www.keaml.cn:8081/>

<sup>5</sup><http://www.keaml.cn/prs/>

Document Frequency (TF-IDF), and neural networks to select the most influential vocabularies. Subsequently, we train appropriate learning models for developing the recommendation engine. Our current model [3] achieves quite good results and outperforms the state-of-the-art methods in the dataset used in [1]. We aim at applying other contemporary deep learning techniques for continuously improving the recommendation system for paper submission. Subsequently, we describe the architecture of our recommendation system in details to develop the corresponding demo for a paper submission recommendation platform.

## 2 RELATED WORK

There have been recent works related to the research of building the recommendation system for paper submission. Wang and co-workers extract useful information from the abstract of a given manuscript for suggesting top relevant conferences or journals in computer science. Feng and co-authors [2] utilize deep learning methods to learn suitable recommendation models for the paper submission recommendation problem. Typically, they use pre-trained Word2Vec embeddings and convolutional neural networks to extract deep features and then make predictions. They specialize in biomedical publication venues with a dataset of 880,165 papers from 1,130 journals in PubMed Central. Pradhan and Pal [4] propose a more comprehensive scholarly venue recommendation system that integrates social network analysis and contextual similarity. They consider the following features: centrality measure calculation, citation and co-citation analysis, topic modeling based contextual similarity, and key-route identification based primary path analysis of a bibliographic citation network. The experimental results on the Microsoft Academic Graph (MAG) dataset show that their proposed method outperforms state-of-the-art recommendation techniques. It is worth noting that the paper submission recommendation system is interesting, and there is still room for improvement. Context-based systems prioritize simpleness and speed, but with recent advances in computational engines, more complex and robust methods are practical. Network-based analysis, although exploiting more handy features, requires more personal information or more complex inputs.

On the other hand, there are recommendation systems for paper submission that are available for public usage. IEEE Publication Recommender accepts keywords, key phrases, or article titles as input data of its search engine. This system also supports automated keyword extraction from articles and has an optional filter by date. Springer Nature Journal suggester requires users to insert the manuscript title, the manuscript text, and select the subject areas to find the top relevant Springer and BMC journals related to that manuscript. Similarly, Elsevier Journal Finder needs the manuscript title, the corresponding abstract, optional keywords, and the field of research to return the most relevant Elsevier journals for a given submission.

## 3 OUR PROPOSED SYSTEM

In this section, we introduce our proposed method to construct a paper submission recommendation system in different fields of research.

Our system primarily aims to support researchers to get more useful information and then make the best decision when choosing an appropriate conference or journal to submit their work. To this aim, there are several tasks during the implementation of such a system. For each research topic (such as, e.g., computer science or mathematics), we find the list of existing journals and all papers that are published in these journals to create the corresponding dataset for training and testing the performance of the recommendation engine. After that, we investigate the most appropriate recommendation model by using all data collected and then integrate it into a recommendation server. This recommendation server can listen to APIs (application program interfaces) from a web application, which has a convenient user interface (UI) such that users (researchers) can easily insert the necessary information of their submission and get the computed results. Figure 1 depicts the proposed framework of out paper submission recommendation system.

### 3.1 Data Collection

We consolidate data from available online resources, directly from publication venues or via third parties. Regarding papers, we collect all information about the title, authors, the abstract, keywords, publication dates, DOI, the publication venue, and others. We implement crawling tools for gathering data from each resource. Concerning each publication venue, we also accumulate the corresponding name, scope, indexes, the impact factor, the venue ranking, the venue location, important dates, and others manually. In the end, we can get two datasets for two research fields, including computer science and mathematics. Typically, the computer science dataset is provided by the work of Wang et al. [1]. This dataset contains 14,012 research papers, which are split into 65 different categories (journals or conferences). The remaining dataset consists of 392,894 articles related to Mathematics/Quantitative dataset from the publisher Springer.

### 3.2 Modeling

In the modeling procedure, we investigate different recommendation models by conducting multiple experiments and comparing the performance of various algorithms to choose the best model for the implementation. Generally, there are two popular approaches to construct an appropriate learning model, which are handcrafted-feature-based and deep-learning-based.

*3.2.1 Handcrafted-feature-based modeling.* The recent work of our investigation into paper submission recommendation systems has achieved excellent performance, which outperforms the state-of-the-art algorithms on the computer science dataset [3].

Firstly, all input data is pre-processed in the system using the following technique. For each title and abstract, we remove numbers, punctuation marks, and stop-words, which may carry little information or even create more noise for the problem. It is important to emphasize that English words often have several morphemes, such as prefixes, suffixes, and base words. Typically, words that share base words almost have the same meaning. To get most out of this, we use Porter Stemmer library (in NLTK package) to break down words into their word-stems. It is worth noting that for keywords, authors often come up with phrases that already exist in various conferences or journals.

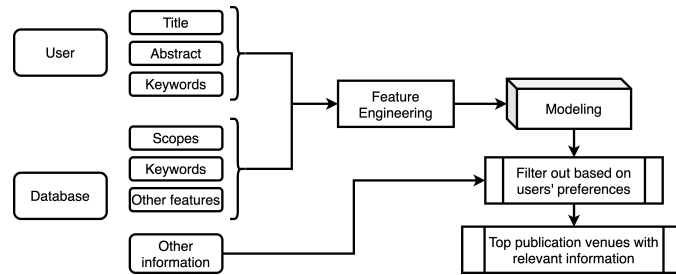


Figure 1: Our proposed workflow of a user’s query in the paper submission recommendation system.

Secondly, we aim to convert texts into suitable weighted vector representations. TF-IDF is an appropriate choice since it can capture the importance of a token according to its frequency. However, since abstracts are often lengthy, there is an abundant amount of unique tokens. As a result, we apply Chi-square statistics to score those tokens and select top vocabularies, which are most valuable for each category. This procedure generates a vocabulary  $W_A$  for abstract tokens. At the same time, we do not apply chi-square and create a vocabulary  $W_T$  containing all title tokens. Using only tokens from  $W_T$  and  $W_A$ , both title and abstract of the manuscript are converted into vectors by TF-IDF. On the contrary, keywords are entirely meaningful and well processed so we encode keywords by one-hot vectors.

Finally, all vectors computed from different inputs are concatenated and fed into a machine learning model. In our experiments, a small neural network is used. This network has one single hidden layer of 200 units. The output layer contains a number of neurons equal to publication venues. Importantly, this network is well trained by using the stochastic gradient descent algorithms.

**3.2.2 Deep-learning-based modeling.** Deep learning models have been proven to be ubiquitously useful in many natural language processing tasks. Pre-training techniques, such as Word2Vec, Glove, Fasttext, ELMO, BERT, or XLNet [5–10], are playing a crucial role in solving many problems. Consequently, deep-learning architectures can make a significant contribution to the robustness and effectiveness of a system.

In this system, we implement various techniques that are most suitable for each part. In specific, pre-training word-level embedding may be used to vectorize keywords of the manuscript, keywords of the venue’s domains or words in a full sentence for better word presentations. Generally, word2vec or Glove are designed to capture linguistic contexts of words but ignore the morphology of the words. For this reason, they rarely handle unseen words. In contrast, Fasttext takes the advantage of sub-word information to compute word representations that may not appear in the training data. On the other hand, deeper pre-training bidirectional models make use of a very large corpus on generalizable tasks to encode the whole context and that can be reusable on downstream tasks like ours. We use these models to deal with the context of documents, like titles or abstracts. Each of these methods has its own input normalization, some appropriate pre-processing to the inputs may be applied to improve the performance of the system. Overall, our end-to-end model architecture has different components that are

responsible for different inputs. Then, a classifier can utilize all the components to make a final prediction. The classifier architecture may be a simple multi-layer feed-forward neural network, stacked recurrent layers, DeepFM, Wide&Deep, or stacked transformers [11–13].

### 3.3 System Architecture

Figure 2 depicts the architecture of our paper submission recommendation platform. As mentioned previously, we collect various datasets, including journal articles and conference papers, and train the most suitable recommendation engine for the platform. After this step, we integrate this recommendation engine into a recommendation server. This server can be deployed in a cloud server providing necessary APIs such that the web/mobile application can easily get the top relevant journals or conferences by giving enough information related to the title, the abstract, and the list of keywords from a given submission. A user can access the paper submission recommendation platform via a web/mobile app for necessary information.

### 3.4 User Interface

The User Interface (UI) of our paper submission recommendation system is shown in Figure 3. In this UI, one user can easily insert necessary information into the search bar, run the query, and get the results for our recommendation model which is a list of target venues. There are two main components in the recommendation system for paper recommendation. The web application can collect necessary information of each manuscript and then run the query to get the corresponding results. The recommendation server is integrated with our proposed algorithm and returns useful results via APIs.

## 4 CONCLUSION

We have presented a new recommendation system for paper submission target venues. We have also shown how to build this system by collecting data, processing data, investigating the most suitable algorithm, and integrating the recommendation algorithm into a recommendation server with an appropriate architecture and UI. In the future, we plan to enhance the performance of this platform and add more useful features for users.

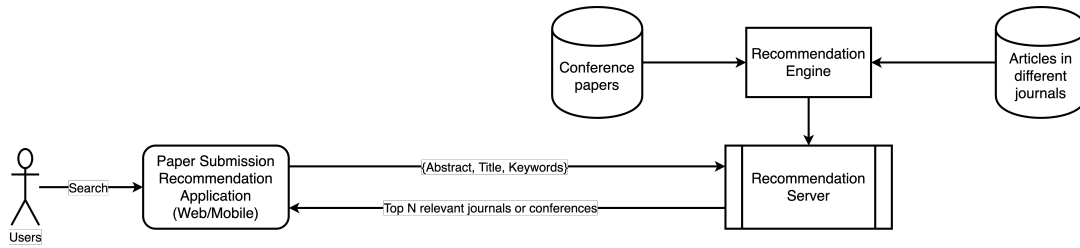


Figure 2: The proposed architecture for our paper submission recommendation system.

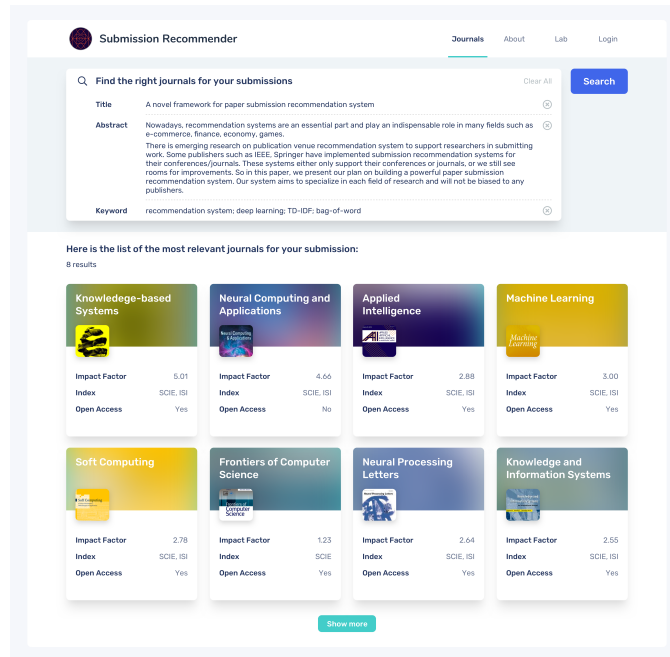


Figure 3: The proposed user interface for our paper submission recommendation system.

## ACKNOWLEDGEMENT

This research is conducted under the Collaborative Research Agreement between National Institute of Information and Communications Technology and University of Science, Vietnam National University at Ho Chi Minh City. We acknowledge the support of Science Foundation Ireland under grant number SFI/13/RC/2106 and L. Meltzers Høyskolefonds, UiB 2019/2259-NILSO.

## REFERENCES

- [1] D. Wang, Y. Liang, D. Xu, X. Feng, and R. Guan, "A content-based recommender system for computer science publications," *Knowledge-Based Systems*, vol. 157, pp. 1 – 9, 2018.
- [2] X. Feng, H. Zhang, Y. Ren, P. Shang, Y. Zhu, Y. Liang, R. Guan, and D. Xu, "Pubmender: A deep learning based recommender system for the biomedical publication venue (preprint)," 11 2018.
- [3] H. Son, H. Phong, N. Dac, D. V. Cuong, and N. T. Binh, "S2rscs: An efficient scientific submission recommendation system for computer science," *The 33th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems*, 2020.
- [4] T. Pradhan and S. Pal, "A hybrid personalized scholarly venue recommender system integrating social network analysis and contextual similarity," *Future Generation Computer Systems*, 11 2019.
- [5] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *CoRR*, vol. abs/1301.3781, 2013.
- [6] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014.
- [7] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2016.
- [8] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," *ArXiv*, vol. abs/1802.05365, 2018.
- [9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *NAACL-HLT*, 2019.
- [10] Z. Yang, Z. Dai, Y. Yang, J. G. Carbonell, R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," in *NeurIPS*, 2019.
- [11] H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, "Deepfm: A factorization-machine based neural network for ctr prediction," in *IJCAI*, 2017.
- [12] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. S. Corrado, W. Chai, M. Ispir, R. Anil, Z. Haque, L. Hong, V. Jain, X. Liu, and H. Shah, "Wide & deep learning for recommender systems," in *DLRS 2016*, 2016.
- [13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *NIPS*, 2017.