

# A Fast Clustering-based Recommender System for Big Data

Hong-Quan Do\*<sup>1</sup>, T.H.-An Nguyen\*\*, Quoc-Anh Nguyen\*\*, Trung-Hieu Nguyen\*\*, Viet-Vu Vu\*,  
Cuong Le\*

\*VNU Information Technology Institute, Vietnam National University, Hanoi, Vietnam

\*\*FPT University, Hanoi, Vietnam

<sup>1</sup>Corresponding Author: [quandh@vnu.edu.vn](mailto:quandh@vnu.edu.vn)

**Abstract**—For years, recommender systems (RS) have emerged as a powerful tool to enable users to find appropriate information according to their needs. Different recommendation methods have been proposed and can be categorized as collaborative filter, content-based, and Hybrid/Ensemble approach. However, the exponential growth of digital information in the recent decades often referred to Big Data, poses new challenges for the current RS. Following this spirit, our work proposes a novel fast clustering-based Recommendation method (denoted as FCR) designed on top of Apache Spark. Comprehensive experiments on a real-world dataset have verified the advantages of our proposed method. It is effective in alleviating the problem of data sparsity and item cold-start. The training and inference time is quick while the slight increase of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) is acceptable.

**Keyword**—Recommender System, Big Data, Clustering-based Recommendation, Item cold-start, Data sparsity, Apache Spark



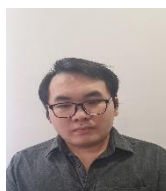
**Hong-Quan Do** received a double M.S. degree in Information and Communication Technology from University of Science and Technology of Hanoi, Vietnam and The University of Rennes 1, France in 2015. He is a researcher at Information Technology Institute, Vietnam National University, Hanoi. His research concentrates primarily on Clustering, Semi-supervised clustering, and Image processing. At the present, he has been involved in many projects related to E-government, and E-Commerce Recommendation applications.



**T.H.-An Nguyen** is studying at the University of Greenwich. She was in the top 3 of the school during the Summer 2021. Her research interests include Recommender Systems and Regression.



**Quoc-Anh Nguyen** is a student in the major of Information Technology at the University of Greenwich Vietnam. He was honored as the best student in the 4<sup>th</sup> semester. He has experiences in developing and implementing real-time, API, Microservices systems. Currently, he has been involved in different website projects including financial, healthcare, and HR management systems.



**Trung-Hieu Nguyen** is a 6<sup>th</sup> semester student at the University of Greenwich (Vietnam). He was an honored student in the 5<sup>th</sup> semester for being the best at Business Intelligent subject. His current study goal is to focus on working with Machine Learning with Python. His future target is working in the field of Clustering and Regression.



Assoc. Prof. **Viet-Vu Vu** received the B.S. degree in Computer Science from Hanoi University of Education in 2000, a M.S. degree in Computer Science from Hanoi University of Technology in 2004, and a Doctor Degree in Computer Science from Paris 6 University in 2011. He is a researcher at Information Technology Institute, Vietnam National University, Hanoi. His research interests include clustering, active learning, semi-supervised clustering, and E-government applications.



Dr. **Cuong Le** is a lecturer in School of Applied Mathematics and Informatics (SAMI), Hanoi University of Sciences and Technology (HUST) from 1998 to 2016. Since 2016 until now he is in Information Technology Institute (ITI), Vietnam National University, Hanoi (VNU, Hanoi) since 2016. He got his PhD at HUST. His research interests lie in the area of information security, mathematics computation and quaternion and Clifford analysis as well as PDE.