

# Hybrid Personalized Book Recommender System Based on Big Data Framework

Fan Liu, Suriya Priya R. Asaithambi, Ramanathan Venkatraman

NUS-ISS, National University of Singapore, Singapore

[liufan@nus.edu.sg](mailto:liufan@nus.edu.sg), [suria@nus.edu.sg](mailto:suria@nus.edu.sg), [rvenkat@nus.edu.sg](mailto:rvenkat@nus.edu.sg)

**Abstract**—With the rapid growth of e-books and e-libraries, it becomes difficult for people to find right books from the flood of information. Personalized recommender system has been developed to help users to find appropriate books based on user’ ratings and interests. Most of the existing book recommender systems are based on content-based or collaborative filtering algorithms. However, those recommender systems provide book recommendations only based on books’ contents and users’ personal interests. This paper proposes a hybrid personalized book recommender system which combines both book popularity and personalized interests to recommend books for users. Book popularity is identified by the collected tweets to indicate current reading trends. The book popularity is represented by Lucene score of each book based on the collected tweets. The proposed book recommender system combines model-based recommender algorithms such as item-based and Alternating Least Squares (ALS) with book popularity. A scalable architecture using big data framework is developed to ingest, store, process and analyze both batch and streaming data. Open-source big data tools such as Apache Solr, Apache Kafka, and Apache Spark MLlib are used to implement the proposed hybrid book recommender system. PySpark is used to implement data pre-processing and big data analysis. The ALS recommender algorithm is implemented by using Apache Spark MLlib. The results show that the accuracy of the proposed hybrid book recommender system is acceptable.

**Keyword** — Hybrid recommender algorithm, Big data framework, Collaborative filtering, Alternating Least Squares (ALS), PySpark



**Fan Liu** received the B. S. degree in engineering and the M.S. degree in automation from Northwestern Polytechnical University, Xi’An, China, in 2003 and 2006 respectively. She received the Ph. D degree in Electrical & Electronic Engineering from Nanyang Technological University, Singapore in 2013. In 2015, she worked as a data scientist in Government Technology, Singapore. She is a lecturer at ISS, National University of Singapore since 2019. Her current research interests include big data analytics, machine learning and artificial intelligence. She has authored/co-authored more than a dozen publications in the area of artificial intelligence and data analytics. Dr. Liu serves as a member of Singapore Computer Society (SCS). Dr. Liu is a certified NVIDIA deep learning institute instructor and ambassador.



**Suriya Priya Asaithambi** is working as Member at Institute of System Science, National University of Singapore. She designed & customised courses for industry professionals and MTech students in emerging topics. She consults in the areas of Data Engineering and Software Engineering. Her twenty-four yearlong teaching and consulting portfolio includes (1) Evaluating various reference architecture and choosing between hybrid design alternatives (2) Building customised technical stacks that satisfy given business needs) using combination of cloud native managed services, full-stack frameworks, libraries, programming languages, API end points and related testing tools). And (3) Building analytics solutions to provide operational and strategic business insights {using mathematical, data ETL pipelines, machine learning algorithms and data visualisation libraries}. She holds a PhD from School of Computing, National University of Singapore. She is a certified professional in TOGAF, GCCC and IBM products. Her current research emphasis is on cloud computing and data engineering.



**Dr. Ramanathan Venkatraman** is working as Member, Advanced Technology Application Practice at National University of Singapore. He has served industry and academia for more than 38 years and has a wide spectrum of experience in the fields of IT, business process engineering and optimization techniques. His current research focuses in evolving decision models for business problems and more recently, he has been contributing to frontiers of knowledge by devising innovative architectural models for ICT in domains such as Distributed Systems Architecture, Big Data and Enterprise Cloud Computing. Dr Venkatraman has a strong practice approach having worked in large scale IT projects for organisations across Asia, US, Europe and NZ. He has published more than 30 research papers in leading journals and reputed technical conference proceedings. Apart from research and consulting, he also teaches advanced technical courses for Master’s program at NUS and has been a key architect in setting up innovative software engineering and business analytics curriculum in the fast changing IT education scenario.