

Efficient models to identify groups in group recommender systems



Thesis submitted in partial fulfillment
for the award of degree

Doctor of Philosophy

by

Chintoo Kumar

**DEPARTMENT OF COMPUTER SCIENCE &
ENGINEERING**

**Indian Institute of Technology
(Banaras Hindu University)
Varanasi**

Roll No: 17071017

2022

To
The
Almighty God, My Family, Friends, Guruji..

CERTIFICATE

It is certified that the work contained in the thesis titled **Efficient models to identify groups in group recommender systems** by **Chintoo Kumar** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree. It is further certified that the student has fulfilled all the requirements of Comprehensive Examination, Candidacy and SOTA for the award of Ph.D. Degree.



Supervisor

Dr. Ravindranath Chowdary C

Dept. of Computer Science & Engineering

IIT(BHU)

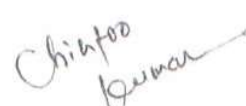
Varanasi - 221005

DECLARATION

I, **Chintoo Kumar**, certify that the work embodied in this thesis is my own bonafide work and carried out by me under the supervision of **Dr Ravindranath Chowdary C** from July-2017 to November-2022, at the Departement of Computer Science & Engineering, Indian Institute of Technology (BHU), Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, theses, etc., or available at websites and have not included them in this thesis and have not cited as my own work.


Date: 26-04-2023

Place: Bengaluru

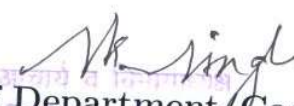

(Chintoo Kumar)

CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of my/our knowledge.


Dr. Ravindranath Chowdary C

पर्यवेक्षक (Supervisor)
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
IIT(BHU), Varanasi
Indian Institute of Technology
(बनारस हिन्दू विश्वविद्यालय)
(Banaras Hindu University)
वाराणसी/Varanasi-221005


Signature of Head of Department, Coordinator of School

समयक विज्ञान एवं अभियान्त्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(बनारस हिन्दू विश्वविद्यालय)
(Banaras Hindu University)
वाराणसी-221005/Varanasi-221005

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: **Efficient models to identify groups in group recommender systems**

Name of Student: **Chintoo Kumar**

Copyright Transfer

The undersigned hereby assigns to the Institute of Technology (Banaras Hindu University) Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the Doctor of Philosophy.

Date: 26-04-2023

Place: Bengaluru


(Chintoo Kumar)

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

Acknowledgments

Though only my name appears on this dissertation's cover, many great people have contributed to accomplishing this job. I owe my gratitude to all who have made this thesis possible and because of whom my post-graduate experience has been one that I will cherish forever. It took a lot of passion, drive and perseverance to make it to the finish line.

I take this opportunity to express my profound gratitude and deep regards to my supervisor Dr. Ravindranath Chowdary C, Associate Professor, Computer Science and Engineering Department, IIT(BHU), Varanasi, for his moral guidance, encouragement and constant support throughout this dissertation. His excellent management skills, strong work ethic, and kind heart set a perfect example for me to follow, which indeed pushed me to nurture similar characters in my life. I am indebted and grateful for his invaluable guidance, advice, and mental and academic support throughout my candidature.

I am obliged and thankful to all other faculty members and staff members of the Computer Science and Engineering Department, IIT(BHU), Varanasi. I would like to express my gratitude to the Research Program Evaluation Committee members Dr. Ravi Shankar Singh and Dr. Lavanya Selvaganesh for their insightful comments, encouragement, and suggestions, which helped me in improving my research from various perspectives. I would like to convey my sincere gratitude to the other faculty members of the Department of Computer Science and Engineering, Prof. Sanjay Kumar Singh, Prof. Kaushal Kumar Shukla, Prof. Anil Kumar Tripathi, Prof. Rajeev Shrivastava, Dr. Anil Kumar Singh, Dr. Bhaskar Biswas, Dr. Sukomal Pal, Dr. Bidyut Kumar Patra, Dr. Pratik Chattopadhyay, Dr. Hari Prabhat Gupta, Dr. Amrita Chaturvedi, Dr. Tanim Dutta, Dr. Lakshamanan Keilasham, Dr. Ajay Pratap, Dr. Prasenjit Chanak, Dr. Mayank Swarnkar, Dr. Obattu Sai Lakshmi Bhavana, Dr. Vignesh Sivaraman and Dr. Vinayak Shrivastava for their guidance and support throughout this tenure. When I look beyond the work en-

vironment, I would not be who I am now without the help of many senior people. I want to express my deepest gratitude to a few of my seniors, Dr. Shivang Agarwal, Dr. Shushant Kumar Pandey, for their kind acts and moral support. Besides, my colleagues Mrs. Deepika Shukla, Dr. Amit Biswas and Mr. Paras Tiwari, whose valuable comments and discussions are also important in accomplishing my thesis. I also want to offer my sincere thanks to my other colleagues Mr. Siba Shankar Sahu, Mr. Rahul Mishra, Ms. Dipty Tripathi, Ms. Shruti Bajpai, Mr. Vishal Srivastava, Mr. Randheer Bagi, Mr. Ankit Jaiswal, Mr. Amit Kumar, Mr. Agrya Halder and Social Projects' club's family members Mr. Yash Jain, Mr. Ravi Kumar, Mr. Shishir Gupta, Mr. Harshit Krishna Chowdhary, Ms. Shrishti Sharma, Mr. Piyush Singh, Mr. Prajwal M, Mr. Aditya Patodia, Mr. Shivam, Ms. Yashika Chaudhary, Mr. Ritwik Ghorui, Ms. Sai Prashritha, Ms. Yashvi Bansal, Ms. Bhavika Guiney, Mr. Kumar Nachiketa who provided me with technical, moral and emotional support during my research program. I feel humbled and grateful to everyone for all the help and support I have received along the way.

Last but not least, I thank my parents and my younger brother and sisters for their constant support, prayer and encouragement, without which this thesis would not have been completed. I am grateful to my other family members who have supported me along the way.

Date: 26/04/2023

Chintoo Kumar

Abstract

The use of E-commerce has increased due to the affordability and availability of network bandwidth. The user base has grown multi-folds for these online markets, that has led to the data explosion on the web. It is a challenging task to filter this large volumes of data. In this scenario, information filtering systems are pivotal in making content/item search more accessible and efficient to provide personalised recommendations. Recommender systems are information filtering systems that provide items/contents/services from a plethora of choices available on the web by mining users' past preferences and feedback. Recommendation helps to overcome the information overload problem and as there are many options (for a specific requirement) and more will be added at regular intervals. Recommender systems aim to help users by offering a choice from a large set of items.

Man is a social animal. Therefore, majority of his activities involve a group, like watching a movie with friends, planning a tour itinerary with family, etc. In these scenarios, there is a requirement to recommend items to a group based on preferences of users. Now, the research has gained visibility in discovering user groups with similar preferences and generating recommendations for each group. The prevalence of social group activities led to the development of a group recommender system. The task of a group recommender system is to generate a group recommendation vector by integrating individual user's diversified interest. In general, a group activity increases social bonding. In the group recommendation, individual member activities show his/her satisfaction or dissatisfaction score for a particular item, and all members are equally important in the recommendation process. The entire group must be satisfied with the recommended products in group recommendations.

The decision-making process of a group is more complex, as different group members have diverse interests and contribute unevenly to the final decision. The composition of

a group is a vital problem in group recommendation. A majority of the previous works have developed group recommender systems with prior knowledge about group members. However, in many realistic scenarios, this information is generally not available. The existing models auto-detect groups by computing correlations among all individual users and are thus computationally very inefficient. Traditional clustering approaches reduce time complexity but do not solve dimensionality reduction, sparsity problems, etc. So, this thesis aims to provide efficient models for auto-detecting groups in group recommendations to mitigate these issues. Exploring a particular group from the available information is a challenging task. In many cases, the group members who share similar characteristics are unfamiliar with each other. This motivated us to explore a group before generating the recommendation, as a group composition is essential in an automatically identified group in group recommendations.

As group formation is an unsupervised task, it is inevitable to auto-detect groups from the available information where each group member shares some common characteristics. This thesis focuses on presenting efficient models for identifying user groups automatically. Groups are formed using probabilistic hashing techniques, e.g., locality-sensitive hashing and one permutation hashing, dimensionality reduction technique, i.e., autoencoder, etc. Automatically identified groups are also formed by incorporating swarm intelligence in K-means technique; an optimization technique (Gaussian mixture model) to cluster similar users and a community detection-based algorithm (Louvain algorithm). This thesis also studies the performance of benchmark clustering approaches in group formation. We conducted extensive experiments on real-world datasets and found that the proposed models to identify groups in group recommendations maximize consensus among the users in a group.

A clustering approach on a utility matrix mainly considers users' ratings for items. It does not consider meta-data to form a group. The review texts are auxiliary information available in most of the datasets. This thesis also considers the textual similarities among the review texts to auto-detect a group. In a group recommendation scenario, group members will have different tastes and achieving consensus is a non-trivial task. A common set of interested items taken from users' preferences is the most vital information to provide recommendations. In reality, recommendations are generated by considering only the set of common interest items instead of considering uninterested items. We hy-

pothesise that these uninterested items may influence the overall recommendations by interpreting users' preferences in a better way.

In this thesis, we first discuss some major issues and challenges through a detailed literature survey on existing group recommendations and auto-detecting groups in group recommender systems. Then, we propose effective solutions to those issues. We also include the importance of uninterested items of the group members while generating the recommendation list. We have performed experiments over automatically identified groups and random groups to check the efficacy of the proposed models. We conducted experiments on publicly available real-world datasets. We found that incorporating uninterested items in group recommendations play a positive role in group recommendations while considering order and flexibility in user preferences. We observed that the overall group satisfaction scores are better in an automatically identified group.

Contents

Abstract	iii
List of Tables	ix
List of Figures	xi
1 Introduction	1
1.1 Research Gap and Research Goal	5
1.2 Motivation	6
1.3 Contributions	7
1.4 Thesis outline	8
2 Preliminaries and Literature Survey	10
2.1 Preliminaries	10
2.2 Literature survey	12
2.2.1 Order preference model	12
2.2.2 Flexible size preference model	13
2.2.3 Group Recommender Systems	18
2.2.4 Memory-based approaches in group recommendation	19
2.2.5 Model-based approaches in group recommendation	20
2.2.6 Group formation in group recommendations	22
3 Automatically detecting groups using locality-sensitive hashing in group recommendations	27
3.1 Problem definition and model overview	28
3.2 Proposed Model	28
3.2.1 MinHash technique	28

3.2.2	Locality-sensitive hashing for finding the candidate pairs of users . . .	32
3.3	Experimental setup and Result analysis	38
3.3.1	Experimental setup and data preparation	38
3.3.2	Results and analysis	40
3.4	Summary	46
4	OPHAencoder: An unsupervised approach to identify groups in group recommendations	48
4.1	Proposed work	48
4.1.1	One permutation hashing	50
4.1.2	Autoencoder	53
4.1.3	OPHAencoder	54
4.2	Experimental setup and result analysis	56
4.2.1	Pre-processing step:	57
4.2.2	Group formation step:	57
4.2.3	Recommendation step:	58
4.2.4	Baseline and state-of-the-art models:	59
4.2.5	Results and analysis	60
4.3	Summary	66
5	Auto-detecting groups based on textual similarity for group recommendations	67
5.1	Methodology	68
5.1.1	Model 1	68
5.1.2	Model 2	69
5.1.3	Model 3	70
5.1.4	Cluster validation	73
5.2	Experimental setup and result analysis	76
5.2.1	Pre-processing step:	77
5.2.2	Group formation step:	77
5.2.3	Recommendation step:	78
5.2.4	Results and discussions	79
5.2.5	Baseline and state-of-the-art models:	81

5.3	Summary	83
6	A study on the role of uninterested items in group recommendations	85
6.1	Models	88
6.1.1	Proposed models	88
6.1.2	Automatically identified group	93
6.2	Results and discussions	98
6.2.1	Experimental setup and data-preprocessing	98
6.2.2	Results and analysis	99
6.3	Summary	109
7	Conclusion and Future directions	110
A	List of Publications	112
A.1	Journal Papers	112
A.1.1	Published Papers	112
A.1.2	Communicated papers	113

List of Tables

2.1	Group satisfaction with respect to items.	12
2.2	Individual user satisfaction after selecting item 1.	12
2.3	User Preference Matrix	14
2.4	Points Vector(P)	14
2.5	Priority value of users	15
2.6	User Satisfaction Matrix	15
2.7	Priority values for the users in case of modified least misery with priority(MLMMP)	17
3.1	Permuted rows using hash functions for characteristic matrix.	31
3.2	Jaccard similarities.	31
3.3	User Preference Matrix	36
3.4	User Satisfaction Matrix.	36
3.5	score vector (s) when $k = 5$	37
3.6	Weight updation of each member of the group.	38
3.7	Average group satisfaction of an automatically identified group of composed using Louvain algorithm of cluster size ($n=60$, $m=17632$) using LMM and LMMP on Last.fm dataset	43
4.1	one permutation hashing on four different sets	51
4.2	Estimate similarity between users in one permutation hashing	51
4.3	Densification in one permutation hashing	52
4.4	Jaccard similarities.	53
4.5	Pairwise similarities among users from the signature matrix into a similarity matrix	55

4.6	Similarity matrix	55
4.7	Recommendation result on the formed group using NCF model	59
4.8	Comparison with baseline methods on the basis of total group satisfaction scores of the automatically identified group of different group sizes (n) and ($m=3706$) using GRAM on MovieLens (ml-1m) dataset	65
4.9	Comparison with baseline methods on the basis of total group satisfaction scores of the automatically identified group of different group sizes (n) and ($m=3706$) using LMM on MovieLens (ml-1m) dataset	66
5.1	Contingency Table	74
5.2	Comparison with baseline methods on the basis of total group satisfaction scores of the automatically identified group of cluster size ($n=40$, $m=2580$) using GRAM on Amazon-music dataset	82
5.3	Comparison with baseline methods on the basis of total group satisfaction scores of the automatically identified group of different group sizes (n) and ($m=36742$) using GRAM on Goodreads Book Reviews dataset	83
5.4	Comparison with baseline methods on the basis of total group satisfaction scores of the automatically identified group of different group sizes (n) and ($m=36742$) using LMM on Goodreads Book Reviews dataset	83
6.1	Movie-genre matrix (MG)	90
6.2	Uninterested genre vector (urec) for the users	91
6.3	User preference matrix (U)	91
6.4	Score vector (s) when $k = 5$	91
6.5	User satisfaction matrix (S)	92
6.6	Priority values for the users in case of modified least misery with priority(MLMMP)	96
6.7	Total group satisfaction of a random group of cluster size($n=50$, $m=1682$) using GRAM and MGRAM on ml-100k dataset without order preferences .	102

List of Figures

1.1	Utility Matrix	2
1.2	Thesis structure	9
3.1	Signature matrix generation using MinHash.	31
3.2	Applying LSH technique in generated MinHash Signature to find candidate pairs of users.	34
3.3	Average group satisfaction for a cluster ($30 \leq clusterSize \leq 40$) using a) GRAM and b) MGRAM on ml-100k dataset.	42
3.4	Average group satisfaction of cluster ($30 \leq clusterSize \leq 40$) using a) AV and b) LMMP with flexible preferences without order on ml-100k dataset.	42
3.5	Average group satisfaction of cluster size ($55 \leq clusterSize \leq 70$) using a) LMM and b) LMMP with order preferences on ml-100k dataset.	43
3.6	Average group satisfaction of cluster size ($55 \leq clusterSize \leq 70$) using a) MLMM and b) MLMMP on ml-100k dataset.	43
3.7	Average group satisfaction using LMM, LMMP a) with order and b) without order preferences on Last.fm dataset.	44
3.8	Average group satisfaction of cluster size ($50 \leq clusterSize \leq 60$) using a) HAM and b) MHAM on Last.fm dataset.	44
3.9	Average group satisfaction of cluster size ($50 \leq clusterSize \leq 60$) using a) GRAM and b) MGRAM on Last.fm dataset.	44
3.10	Average group satisfaction of cluster size ($55 \leq clusterSize \leq 70$) using a) HAM and b) MHAM on ml-100k dataset.	45
3.11	Average group satisfaction of cluster size=70 and a) m=500 and b) m=1000 using LMM and LMMP on ml-100k dataset.	45

3.12	Average group satisfaction for a cluster ($30 \leq clusterSize \leq 40$) using a) MPP and b) MPMP on ml-100k dataset.	45
3.13	Runtime complexity analysis of MinHash technique.	46
4.1	Signature matrix generation using one permutation hashing.	53
4.2	Encoder Model	54
4.3	NCF model summary on the formed group by OPHAencoder	59
4.4	Average group satisfaction of cluster size ($70 \leq clusterSize \leq 80$) using a) LMM and b) MLMM on ml-100k dataset.	61
4.5	Average group satisfaction of cluster size ($70 \leq clusterSize \leq 80$) using a) GRAM and b) MGRAM on ml-100k dataset.	61
4.6	Average group satisfaction of cluster size ($70 \leq clusterSize \leq 80$) using a) LMMP and b) MLMMP on ml-100k dataset.	62
4.7	Average group satisfaction of cluster size ($70 \leq clusterSize \leq 80$) using a) GRAM and b) MGRAM when item (m)=300 on ml-100k dataset.	62
4.8	Plotted curve (value Vs. batch) for both training and validation sets of NCF model	63
4.9	Average group satisfaction of cluster size ($50 \leq clusterSize \leq 55$) using a) AV with flexible preference without order and b) HAM on Amazon-music dataset.	64
4.10	Average group satisfaction of cluster size ($50 \leq clusterSize \leq 55$) using a) GRAM and b) MGRAM on Amazon-music dataset.	64
4.11	Average group satisfaction of cluster size ($50 \leq clusterSize \leq 55$) using a) LMM and b) MLMM on Amazon-music dataset.	64
4.12	Average group satisfaction of cluster size ($50 \leq clusterSize \leq 55$) using a) LMMP and b) MLMMP on Amazon-music dataset.	65
5.1	Siamese neural network	70
5.2	Assigning single hash value in multiple planes	72
5.3	Two clustering assignments	74
5.4	a.) GRAM results b.) Results using various consensus functions on Amazon-music Dataset	79
5.5	LMM and LMMP results on Amazon-music dataset.	80

5.6	GRAM & LMM results on Clothing Fit Data Dataset	80
5.7	Total group satisfaction of cluster size ($130 \leq clusterSize \leq 140$) on Amazon-music dataset using a.) GRAM b.) MPP	81
6.1	Average group satisfaction of cluster size ($55 \leq clusterSize \leq 60$) of an automatically identified group on ml-100k dataset using GRAMIUI and GRAM . . .	103
6.2	Average group satisfaction of cluster size ($55 \leq clusterSize \leq 60$) of an automatically identified group on ml-100k dataset using MLMMIUI and MLMM . . .	103
6.3	Average group satisfaction of cluster size ($55 \leq clusterSize \leq 60$) of an automatically identified group on ml-100k dataset using LMMIUI and LMM	103
6.4	Average group satisfaction of cluster size ($55 \leq clusterSize \leq 60$) of an automatically identified group on ml-100k dataset using LMMPIUI and LMMP . . .	104
6.5	Average group satisfaction of cluster size ($50 \leq clusterSize \leq 55$) of an automatically identified group on Amazon-music dataset using LMMIUI and LMM .	104
6.6	Average group satisfaction of cluster size ($50 \leq clusterSize \leq 55$) of an automatically identified group on Amazon-music dataset using MLMMIUI and MLMM	104
6.7	Average group satisfaction of cluster size ($50 \leq clusterSize \leq 55$) of an automatically identified group on Amazon-music dataset using LMMPIUI and LMMP	105
6.8	Total group satisfaction of an automatically identified group of cluster size ($n=55$ and $m=1682$) using MPMP and LMMP on ml-100k dataset.	105
6.9	Total group satisfaction of an automatically identified group of cluster size a)($n=55$) and b)($n=130$) using HAM on ml-100k dataset.	105
6.10	Total group satisfaction using different consensus functions on ml-100k dataset for order preferences incorporating uninterested item-based model ($n=55$ and $m=200$) using a) Automatically identified group and b) Random group	106
6.11	Total group satisfaction using different consensus functions on ml-100k dataset for flexible-size preferences incorporating uninterested item-based model ($n=55$ and $m=200$) using a) Automatically identified group and b) Random group . . .	106
6.12	Comparison of a random group and an automatically identified group using GRAM incorporating uninterested item-based model (GRAMIUI)	108