**Handling Malicious and Non-malicious Noise in Collaborative Recommender Systems**

*Report submitted to   
Indian Institute of technology Kharagpur  
for the partial fulfilment of the degree*

*of*

**Bachelor in technology**

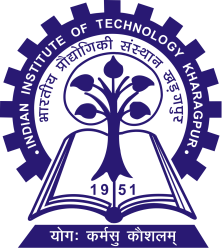
*by*

**PRINCE AGARWAL (16IM30034)**

*under the guidance of*

**Prof. SUJOY BHATTACHARYA**

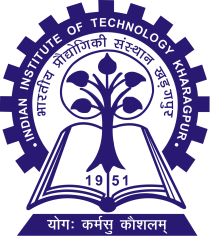
(Vinod Gupta School of Management)



**VINOD GUPTA SCHOOL OF MANAGEMENT**

**INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR**

**NOVEMBER 2019**



**Vinod Gupta School of Management  
Indian Institute of Technology Kharagpur  
Kharagpur, India-721302**

**Certificate**

This is to certify that this project entitled **Handling Malicious and Non-malicious Noise in Collaborative Recommender Systems** submitted by **PRINCE AGARWAL (16IM30034)** to Indian Institute of Technology Kharagpur, is a record of the bona fide project work carried out under my supervision and is worthy of consideration for award of Bachelor of Technology Degree of the Institute.

**Prof. Sujoy Bhattacharya**Vinod Gupta School of Management  
Indian Institute of Technology Kharagpur  
India - 721302

**Declaration**

I certify that,

1. The work contained in the project is original and has been done by me under the guidance of my supervisor;
2. The work has not been submitted to any other institute for any other degree or diploma;
3. I have followed all the guidelines laid by the Institute while doing the project;
4. I have conformed to ethical norms and guidelines while writing the report;
5. Whenever I have used Materials (data, models, figures and text) from other sources, I have given due credit to them by citing them in the text of the report, and giving their details in the references, and taken permission from the copyright owners of the sources, whenever necessary.

**NAME**: PRINCE AGARWAL  
**ROLL NO**: 16IM30034

**Acknowledgement**

I express my deep sense of gratitude to my guide Prof. Sujoy Bhattacharya for his valuable guidance and inspiration throughout the course of this work. I am thankful to him for his kind help, assistance and support, and his motivation toward independent thinking. It has been a great experience working under him in the cordial environment.

**Abstract**

Recommender Systems have become an important part of various web applications by providing higher customer satisfaction through personalised recommendations. Most of the research in this domain focus on improving the recommendation methods for higher accuracy values but the inconsistencies or the noise is often ignored. There exist mainly two types of noise in recommender systems: malicious and non-malicious or natural noise. The main reason behind natural noise existence is imperfect user behaviour i.e. careless/erroneous preference selection. Malicious noise arises when deliberate attempts are made to tamper the output results in some manner. We believe that both classes of noise are important and can adversely effect recommendations. Therefore, we propose to simultaneously handle both malicious and natural noises.

For the detection of natural noise, we characterize users and items based on their ratings into different classes and identify a rating as noise if it contravenes user or item tendencies. We correct these identified naturally noisy ratings by prediction using collaborative filtering approaches. For the malicious noise, we use a value-based neighbor selection which selects neighbors for active users in user-based collaborative filtering recommender systems under shilling attack. We provide an empirical evaluation of our approach for validation.

**Introduction**

Recommender systems play an important role in most top-ranked commercial websites such as Amazon, Netflix, Last.fm or IMDb. The goal of these recommender systems is to increase revenue and present personalized user experiences by providing suggestions for previously unknown items that are potentially interesting for a user. With the growing amount of data on the Internet, the importance of recommender systems increases even more to guide users through the mass of data.

The most popular version of Recommender Systems (RS), learns from user preferences about a predefined set of known items, and predicts the preference degree about unknown items. With this goal, many applications have been built to recommend different types of items like movies, books, TV shows, jokes, news ,scientific papers, web pages, and so on. They have covered diverse areas like e-commerce, e-learning, e-services, tourism, and software engineering.

The two most used techniques in the development of RS are the content-based and the collaborative filtering ones. Content-based recommender systems suggest items with similar features to those that the user chose in the past. On the other hand, collaborative filtering recommender systems recommend items that other similar users liked in the past. Collaborative filtering systems can generate recommendations just using information about users’ preferences regarding a set of items, and due to their simplicity, they have become popular nowadays.

Most of research done in collaborative filtering has been focused on tuning basic algorithms to improve their accuracy. It includes neighbourhood-based user-user and item-item approaches, dimensionality reduction approaches, probabilistic, MDP-based, graph-based, rule-based methods, and their hybridization. In past years, advances in collaborative filtering have overcome important challenges like the Netflix Prize, showing that the rating prediction can be significantly improved.

Unlike classical data mining processes in which the data is usually considered to contain a degree of inconsistency, and there-fore data pre-processing is an important step in the mining pro-cesses, in RS it is assumed that the ratings in the datasets are free of irregularities. Nevertheless, recently Amatriain et al. have shown that users could be inconsistent when they elicit ratings for items, exposing recommender systems data to inconsistencies. These inconsistencies, so called natural noise, are inserted without malicious intent, and its manipulation represents an important open problem to improve the recommender systems performance.