ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.5

Cloud-Assisted Video Recommendation System

PoornapragnaVadiraj[1], Sohan M.V[2], Chinmay .S. Watwe[3], Thejas B.S [4], Sanjoy Das[5]
BE Student[1, 2, 3, 4], Assistant Professor[5]
Department of Computer Science
K. S. Institute of Technology, Bangalore, India

Abstract:
With the quick development in multimedia services and the massive offers of video contents in online social networks, users have difficulty in obtaining their interests. Therefore, various personalized recommendation systems have been proposed. However, they ignore that the accelerated proliferation of social media data has led to the big data era, which has greatly impeded the process of video recommendation. In addition, none of them has considered both the privacy of users' contexts (e,g., social status, ages and hobbies) and video service vendors' repositories, which are extremely sensitive and of significant commercial value. To handle these problems, we propose a cloud assisted differentially private video recommendation system based on distributed online learning. In our framework, service vendors are modeled as distributed cooperative learners, recommending videos according to user's context, while simultaneously adapting the video-selection strategy based on user-click feedback to maximize total user clicks (reward). Considering the scarcity and heterogeneity of big social media data, we also propose a novel geometric differentially private model, which can greatly reduce the performance (recommendation accuracy) loss. Our simulation shows the proposed algorithms outperform other existing methods and keep a delicate balance between computing accuracy and privacy preserving level.

Index Terms: Online social networks, multimedia big data, video recommendation, distributed online learning, differential privacy, and media cloud.

I. INTRODUCTION

In recent years, online social networks (OSNs) have been massively growing, where users can share and consume all kinds of multimedia contents. As a result, given the numerous different genres of videos in social media, how to discover the videos of personal interest and recommend them to individual users are of great significance. However, there exist two major challenges in this scenario. The first challenge comes from the big data's role in the personalized recommendation. In detail, OSNs have accelerated the popularity of applications and services, resulting in the explosive increase of social multimedia data. In this case, multimedia big data puts companies in a favorite position to have access to much more contextual information. However, how to harness and actually use big data to effectively personalize recommendation is a monumental task. Traditional stand-alone multimedia systems cannot handle the storage and processing of this large-scale datasets. Besides that, complex and various user-generated multimedia big data in the OSNs results in the sacristy and heterogeneity of users' context data. Hence, it is extremely challenging to implement recommendation with the multimedia big data. Furthermore, the privacy in recommendation has raised widely concern. On the one hand, as declared in, user's sensitive context information may be exposed by the recommendation results. Intuitively, the more detailed the information related to the user is, the more accurate the recommendations for the user are. But once the recommendation records are accessed by a malicious third party, individual features can be inferred by them merely based on the outcome of the recommendation. For example, advertising video of luxury goods recommended to a particular person indicate the income level of this user. Also, basketball video recommendation for the same user exposed its hobby. Then with additional side information, the malicious party may identify the person in real life. On the other hand, the inventory of videos is an important commercial secret for the service vendor. As for the service vendors' incentives, they rely on stored video source files to gain popularity among users. Intuitively, video service vendors are selfish and they refuse the inference of what they have in the inventories by the revenue gain of each video. Consequently, avoiding the divulge of video contents of each service vendor is desirable. The output of a differentially private mechanism is going to be almost exactly the same whether it includes the real me or the pretend me. Differential privacy proposed recently is a heuristic method to solve this problem. Informally, differential privacy means that the output is going to be almost exactly the same whether it includes a single user's data in the input datasets. Therefore, hardly can one make an accurate inference on signal user's feature based on the recommendation results. Besides, adding Laplace noise into the recommendation rewards can hide small changes that arise from a single video's contribution. Thus, the revenue gain of one signal video cannot be deduced. Several studies have incorporated it into recommendation systems, but their works only focus on small-scale media datasets, yet executing differential privacy in a large dataset often impacts little on accuracy, which works extremely efficiently under the big data context. In conclusion, it is necessary to design a privacy-preserving video recommendation that can handle the multimedia big data and achieve high accurate recommendations. In this paper, we introduce differential privacy into distributed online learning to design an efficient and high-accurate timely recommendation system based on multimedia cloud computing. As illustrated in Fig. 1, user-generated multimedia big data (e.g, images, audio clips and videos) is first translated to remote media cloud and stored in decentralized data centers (DCs). Then use technologies such like Bag-of- Features Tagging (BoFT) to extract user's context vectors and convert the results to distributed video service vendors (servers). Finally recommended video contents are pushed to multimedia applications in OSNs. Our main theme in this media cloud based scenario is that video service vendors are modeled as decentralized online learners, who try to learn from user's high-dimensional context data and match it to the optimal video. The service vendors are connected together via a fixed network over the media cloud, each of whom experience inflows of users' context vectors to them. If service vendors cannot find suitable videos in their repositories for the coming user, they can forward the use's context data to neighbor service vendors, who will find out the suitable video in his repository to recommend to this user. At the end of each time slot, the reward of the recommended video is observed. Service vendors can learn from the result and adjust their selection strategy next time. Since the extracted context vectors from multimedia big data are high dimensional and omnifarious, the context space with d dimensions (d is the number of user features) can be extremely huge and heterogeneous. Then, learning the most match able video for each individual can be extremely slow.

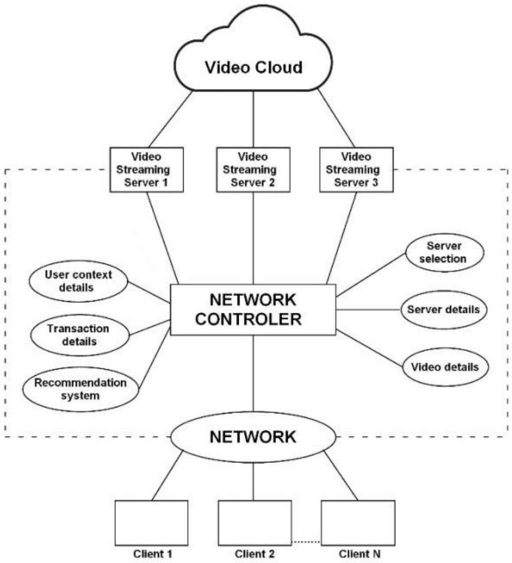

Figure.1. Mockup of the system architecture

II. BACKGROUND

A. Differential Privacy

The concept of *differential privacy* is originally introduced by Dwork*et al.* [12], which gives us a riorous definition of privacy.

Definition 1: (Differential Privacy*]*). A randomized algorithm
A has ε differential privacy if for any two input sets
A and B differing on a single entry, and for any set of outcomes
$R \in \mathrm{Range}(A) P[A(A) \in R] \leq \exp(\varepsilon)$
$\times P[A(B) \in R]$.

Informally, differential privacy means that the outcome of two nearly identical input datasets (different for a single component) should also be nearly identical. Thus, individual information can hardly be inferred by comparing the query result of A and B.

In our model, the input datasets are users' context vectors. The privacy ε is the parameter to measure the privacy level of the algorithm. The choice of ε is a tradeoff between the privacy and the accuracy of the output.

B. Online Learning

Our proposed distributed learning method derives from contextual bandits. This algorithm learns from the context information available at each time, which, in this case, are the users' context vectors. Then, it keeps an index that weights the *estimated performance and uncertainty* of each action (recommended video or neighbor service vendor in this case) and chooses the action with highest index at each time. Furthermore, the indices for the next time slot for all actions are updated based on the feedback received from the chosen action (users click feedback). There exists some work studying the contextual bandit, where the best action given the context is learned online. C. Tekin*et al.* first proposed a distributed contextual bandit framework for big data classification and social recommendations. But the uniform partition method proposed in their work does not fit into the sparse big data is a heuristic work. Nonetheless, the single-learner framework cannot satisfy the need of the massive multimedia big data.

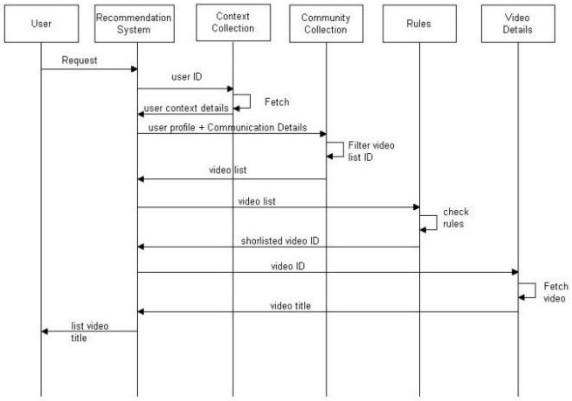

Figure.2. Sequence Diagram of the Recommendation Module

III.ADVERSARY MODEL AND DESIGN GOALS

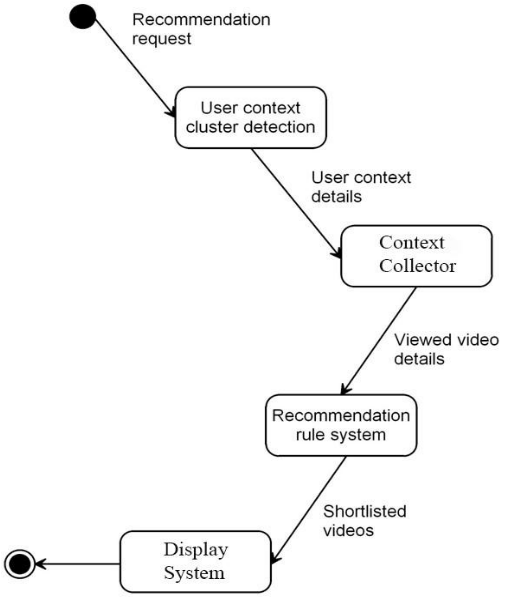

Figure.3. Flow Diagram of the proposed architecture

Similar to the privacy concerns proposed in, we consider a adversary model as follows: 1) *Malicious third party* who can gain access to the recommendation outputs and owns some side information such as location about some users. The goal of this malicious third part is to deduce a particular user's features by observing the recommendation outputs. Then, they can identify the media user in the real world with deduced features and additional side information. 2) *Selfish and curious service vendors* who want to infer neighbors' repositories from shared information. For example, the curious service vendor forwards a sports fan's context vector to a neighbor service vendor, who receives a high reward from its recommendation. Since the reward is shared with the curious service vendor for estimation, the curious service vendors infer that this neighbor service vendor may own a video about sport. To address the adversary models above, we proposed a differentially private learning algorithm. Our scheme achieves privacy protection and performance guarantees as follows.

1) ***Users' privacy guarantee*:** Even if the malicious party can gain access to the recommendation outputs, it is less likely for it to infer the user's feature from the recommended result. And we prove that our proposed algorithm can preserve ε-differential privacy for user's privacy.
2) ***Service vendors' privacy guarantee*:** The video of neighbor service vendors cannot be inferred by shared information. The proposed algorithm can preserve ε-differential privacy for service vendors. 3) *Performance guarantee*: The proposed algorithm can guarantee the regret in equation (6) is sublinear converged, i.e., $R(T) = O(T_\gamma)$ such that $\gamma < 1$. A smaller γ will result in faster convergent rate.
4) *Privacy-reward tradeoff*: Our analysis shows that the higher the privacy level is preserved, the lower the total reward is received. By varying the value of the privacy parameter ε, we can keep a tradeoff between the total recommendation reward and the privacy preservation level.

IV. EXPERIMENTAL GOALS AND ANALYSIS

In this section, we demonstrate the theoretical regret bounds for our algorithms with empirical results based on very large real-world datasets, which includes massive multimedia data and social media users-generated big data. We show that:
1) Regret bounds are sub linear converged over time.
2) Our differentially private methods work well and do not come at the expense of recommendation reward.
3) *Geometric differentially private* method has a lower regret bound and higher accuracy. Finally, we use users' context vectors refined from real datasets to test the recommendation accuracy (reward) of our algorithms.

V. CONCLUSION

In this paper, we have presented a differentially private learning framework for video recommendation for OSNs. To tackle with the large volume and heterogeneity of big data, we employee a adaptive space partition method to improve the convergence rate. Concerned with the privacy of social network users and video service vendors, we use Exponential Mechanism and Laplace Mechanism in our model simultaneously. Furthermore, to alleviate the performance loss due to introducing differential privacy, we refine our framework to novel *geometric differentially private* model. We have theoretically analyzed our algorithms in terms of performance loss (regret) and privacy preserving level.

VI. ACKNOWLEDGMENT

Authors would also like to thank VGST(Vision Group on Science and Technology),Government of Karnataka, India for providing infrastructure facilities through K-FIST level I project at KSIT,CSE R&D department, Bangalore. We would also like to thank our project guide, Mr. Sanjoy Das for his help and support provided to carry out the project and complete it successfully.

VII. REFERENCES

[1]. Z. Wang, W. Zhu, and P. Cui, "Social media recommendation," in Social Media Retrieval. London, U.K.: Springer, 2013, pp. 23–42.

[2]. A. Jeckmans et al., "Privacy in recommender systems," in Social Media Retrieval, London, U.K.: Springer, 2013, pp. 263–281.

[3]. A. Samuel, M. I. Sarfraz, and H. Haseeb, "A framework for composition and enforcement of privacy-aware and context-driven authorization mechanism for multimedia big data," IEEE Trans. Multimedia, vol. 17,no. 9, pp. 1484–1494, Sep. 2015.

[4]. M. J. Pazzani and D. Billsus, "Content-based recommendation systems," in The Adaptive Web. Berlin, Germany: Springer, 2007, pp. 325–341.

[5]. W. Zhu, C. Luo, and J. Wang, "Multimedia cloud computing," IEEE Signal Process. Mag., vol. 28, no. 3, pp. 59–69, May 2011.

[6]. P. Zhou, Y. Zhou, D. Wu, and H. Jin. "Differentially private online learning for cloud-based video recommendation with multimedia big data in social networks," Sep. 2015.