International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

A Novel Model for Video Recommendation with Multimedia Big Data

Mr. Venkanna. B & Dr. Ravisankar Malladi
Department of CSE, Vaagdevi Engineering College,Bollikunta, Warangal-506 005, Telangana,India
, Department of CSE, Vaagdevi College of Engineering(Autonomous), Bollikunta,Warangal - 506 005, Telangana,India

Abstract—*With the rapid growth in multimedia services and the enormous 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 the 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 sparsity 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, media cloud.*

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 s different genres of videos in social media, how to discover the videos of personal interest and recommend them to individual l users are of great significance. Recommendation is foreseen to be one of the most important services that can provide such personalized multimedia contents to users. Several l companies have demonstrated initial successes in multimedia recommendation system design. Reported that YouTube won its first Emmy for video recommendations. Actually, most OSNs recommend video content to their users based on the user's rich context information (e.g., social status, ages, professions, health conditions and hobbies) contained in their released multimedia data. Regarding this way, several recommendation systems have been proposed. 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.

Two difficulties into consideration, establishing a privacy-preserving video recommendation system with multimedia bid data can be extremely challenging. Traditional recommender systems for multimedia, including collaborative filtering (CF) and content-based (CB) recommendation can provide meaningful multimedia recommendations at an individual level. However, their stand-alone systems have difficulties in dealing with tremendous highdimensional multimedia big data. As for the privacy concern in recommendation, previously, anonymity was the main tool in recommendation. But the fact that the information can only be partially removed will allow for re-identification. 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

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

accurate inference on signal user's feature based on the recommendation results. Besides, adding lap lace 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 datasets 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-ofFeatures 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.

To goal of each service vendor is to maximize its long term expected total recommendation reward and do not want to reveal their repositories to other service vendors. However, in the cooperation, each service vendor will share some information such as the user's context vectors and the videos' revenue gains with neighbor service vendors. Then, service vendors can infer the repositories of other service vendors from the shared information. To solve this privacy leakage, we adopt Laplace mechanism, adding noise to shared revenue gains. As for the users' privacy, to prevent the exposure of their feature by the recommendation videos, adding noise to the revenue gains is not non-effective. Because the gain is produced after the recommended video is revealed and disturbing the accurate estimation of gains of their own videos with this noise is not necessary. Thus, we employ exponential mechanism to protect the users' privacy, where the service vendors randomly select the video according a computed exponential probabilities. Faced with the fact that user's contexts (d-dimensional point in the context space) are sparse distributed over the context space.

Aim and Objective

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 Omnifurious, 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. Therefore, each service vendor initially groups users (partition the context space) with similar context into rough crowds, and then they dynamically refine the partition strategies over time.

GOALS AND REQUIREMENTS

Our main goal is to design a system that allows consumers to aggregate their data from multiple sources, control how that data is accessed and shared, and to allow them to quickly and easily access that data from any device, at anytime, from anywhere. These top-level goals translate into the following sub goals.

Consolidation: To allow a single view into multiple data streams and cross-correlation between different time series, the system should automatically consolidate energy usage data from multiple sources.

Durability: To allow analysis of usage history, a consumer's energy data should be always available, irrespective of its time of origin.

Portability: To prevent lock-in to a single provider, data and computation should be portable to different cloud providers.

Privacy: To preserve privacy, the system should allow a consumer to determine which other entities can access the data and at what level of granularity.

Flexibility: The system should allow consumers a free choice of analytic algorithms.

Integrity: The system should ensure that a consumer's energy data has not been tampered with by a third party.

Scalability: The system should scale to large numbers of consumers and large quantities of time series data.

Extensibility: It should be possible to add more data sources and analytic algorithms to the system. Good

Performance: Data analysis times and access latencies should be minimized.

MAJOR CHALLENGES

- 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 sparsity and heterogeneity of users' context data. Hence, it is extremely challenging to implement recommendation with the multimedia big data.
- The second challenge, In our problem setting, in order to protect the privacy of neighbor service vendors, we face a big challenge that traditional differential privacy only apply to static database.

PROBLEM DEFINTION

We propose a novel geometric differentially private method to promote the total reward. This paper makes the following contributions:

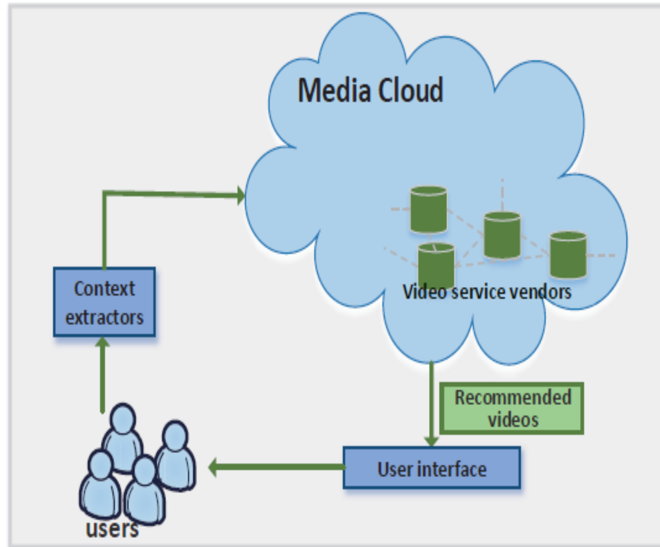

Fig :- A general illustration of multimedia cloud based videorecommendation system.

- We propose a media cloud based video recommendation system and rigorously formulate it as a distributed online learning problem. In our model, decentralized service vendors work cooperatively to deal with large-scale contextual data.
- To handle the dimensionality and sparsity of the multimedia big data, our method adaptively partitions the context space for each service vendor. Our evaluation results show this method has lower performance loss and converges fast to optimal strategy.
- To the best of our knowledge, we are the first to deal with the privacy issue of both the social media users and video service vendors in recommendation. We integrate exponential mechanism and Laplace mechanism simultaneously into distributed learning systems. We guarantee ε-differential privacy while not coming at substantial expense in total reward.
- We propose a "geometric differentially private model" to deal with the sparse contextual data, which can reduce the performance loss extensively.

PRIVATE DISTRIBUTED ONLINE LEARNING ALGORITHM

Since the reward of each recommended video for different users have unknown stochastic distributions, the natural way to learn a video's performance is to

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

record and update its sample mean reward for the same context vector. Using such an empirical value to evaluate the expected reward is the basic approach to help the service vendors to learn. However, the context space X can be very large, recording and updating the sample mean reward for each context are scarcely possible. The memory capacity of the server cannot meet the need of keeping a sample mean reward for all contexts. To overcome the difficulty, we dynamically partition the entire context space into multiple smaller context subspaces (according to the number of arriving users). Then, we maintain and update the sample mean reward estimates for each subspace. This is due to the fact that the expected rewards of a video are likely to be similar for similar contexts. In our distributed framework, each service vendor i∈ M dynamically partitions the context space X when context $x_i(t)$ arrives to them.

IMPLEMENTATION

Exploration and Reward Estimation:-

Upon each context data arrival, service vendor i first checks to which subspace C in the set P^t the context belongs and the level of C. To get accurate performance estimation of each arm k $\in M_i$, service vendor i needs to judge whether k has been fully explored . Since service vendor i does not know the performance service vendor k's videos, it needs to send neighbor service vendor k some context samples to train it and make sure it will mostly select optimal video. The $N^i_{1,k,c}(t)$ denotes the times when k ∈ $M_{-i}$is selected for training. In the training process, service vendor idose not need to communicate with service vendor k to observe the reward $f^i_{k,x}(t)$. If each service vendor k ∈ M_{-i}has been fully trained, service vendor i start to explore the performance of leaner k ∈ $M_{-i}$and observe the reward of each k . The control function G1(t), G2(t) and G3(t) ensure that video is selected sufficiently many number of times so that the sample mean estimates $r^{-i}_{k,C}$ (t) are accurate enough. And we set different control function for k ∈ M_{-i} and k ∈ M_i, i.e., G2(t) is larger than G1(t). Because for k∈ M_{-i}, the reward $r^{-i}_{k,C}(t)$ is added with noise, we need more times to evaluate performance of k ∈ M_{-i} .

Decision with Privacy Protection:-

For subspace C, when all arms have been fully explored, there are accurate sample mean estimations for each arm. In traditional bandit algorithms, the learners (service vendor in this case) usually select the arm with the highest sample mean reward. However, the optimal arm will expose the individual feature. Thus, to protect the user's privacy, service vendor I first randomly choose one arm k_i∈ M_i according to the computed probability distribution, where Δu is the sensitivity of exponential mechanism. Then, it select another arm k_j∈ $M_{-i}$with the highest estimated reward. Finally service vendor i compare the estimated reward of k_j and k_i , then it select the one with higher estimated reward for context $x_i(t)$. We will prove this randomly selection scenario guarantee ε-differential privacy in next our analysis section.

Update and Partition the Context Subspace:-

At the end of each time slot, the algorithm first updates $M^i_C(t)$, $r^{-i}_{k,C}$ (t) and $N^i_{k,C}(t)$, where M^i_C (t) = M^i_C (t) + 1, $N^i_{k,C}(t)$ = $N^i_{k,C}(t)$ + 1 and $r^{-i}_{k,C}P$ (t) = $\sum_{x(t)\varepsilon C} f^i_{k,x(t)}$ /$N^i_{k,C}$ (t). Then the algorithm decides whether to further partition the current subspace C, depending on whether we have sufficient context vectors arrivals in C. Specifically, if $M^i_C(t) \geq Am^{pl}$ at time t, C will be further partitioned, where p and m are positive numbers. When partitioning is needed, C is uniformly partitioned into md smaller hypercubes. Each hypercube is a level-(L+1) subspace with side-length 1/m of that of C. Then C is removed from the current context set P t . New subspaces are added into P t . Fig. 3 provides us an illustration of this partition process when m = 2, d = 2. Then, we describe Algorithm 2 as follows. In our problem setting, in order to protect the privacy of neighbor service vendors, we face a big challenge that traditional differential privacy only apply to static database.

Tree based aggregation:-

Assume for simplicity that T = $2^\alpha$for some positive integer α. We create a binary tree, i.e., $Tree_k$ for each video k ∈ M_i with its leaf nodes being f_1, ..., f_T . As illustrated…. at each time slot, when new reward is produced, we insert the value of the reward into the leaf node. Over the entire time sequence [T], the rewards are inserted sequentially. Each internal node x in $Tree_k$stores the sum of all the leaf nodes in the tree rooted at x. First notice that one can compute any v_t using at most log(T) nodes of $Tree_k$. Second, notice that for any two neighboring datasets D and D′ different in leaf node fi and fi′ at most log(T) nodes in $Tree_k$gets modified. So, if we flatten the complete tree as a vector then for any neighboring datasets D and D′ one can easily show that $\|Tree(D) - Tree(D')\|_1 \leq \log(T)$. We will further bound the amount of the noise added to each tree in section V when evaluating the performance of our algorithm.

CONCLUSION

In this paper, we have presented a differential private distributed learning framework for video

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

recommendation for online social networks. To tackle with the large value and heterogeneity of big data, we adopt dynamic space partition to distributed contextual bandit. Concerned with the privacy of social network users and that of video service vendors, we use exponential mechanism and Laplace mechanism 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. We have also evaluated our algorithms, demonstrating their sub-linear converged regrets, delicate trade-off between performance loss and privacy preserving level and extensively reduction.

FURTHER WORK

The privacy in recommendation has raised widely concern. On the one hand, as declared , 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 recommendation s for the user are. But once the recommendation records are accessed by a malicious third party, individual features ca n 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 it's 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.

REFERENCES

[1] Z. Wang, W. Zhu and P. Cui, "Social media recommendation," Social Media Retrieval. Springer London, pp. 23-42, 2013.
[2] C. Ming, S. James and Z. Jie, "Connection discovery using big data of user-shared images in social media," Multimedia, IEEE Transactions on. vol. 17, no. 9, pp. 1417-1428, 2015.
[3] "YouTube wins its first Emmy for video recommendations you can't resist," The Verge. [Online]. Available:http://www.theverge.com/2013/ 8/1/4578544/youtube-wins-first-emmy-for-video-recommendations-youcant-resist
[4] "Privacy by the Numbers: A New Approach to Safeguarding Data," Scientific American. [Online]. Available:http://www.scientificamerican.com/ article/privacy-by-the-numbers-a-new-approach-to-safeguarding-data/
[5] A. Jeckmans, M. Beye and Z. Erkin et al, " Privacy in recommender systems," in Social media retrieval, Springer London, pp. 263-281, 2013.
[6] M. Cheung and J. She, "Bag-of-features tagging approach for a better recommendation with social big data," in Proceedings of the 4th International Conference on Advances in Information Mining and Management, pp. 83-88, 2014.
[7] 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," Multimedia, IEEE Transactions on. vol. 17, no. 9, pp. 1484-1494, 2015.
[8] A. Machanavajjhala, A. Korolova and A. D. Sarma, "Personalized social recommendations accurate or Private?," in Proceedings of the VLDB Endowment, vol. 4, no. 7, pp. 440-450, 2011.
[9] M. Tkalcic, A. Odic and A. Kosir, "Affective labeling in a content-based recommender system for images," Multimedia, IEEE Transactions on. vol. 15, no. 2, pp. 391-400, 2013.