Guess Your Favorite or Your Privacy? Privacy Protection of Multimodal Data in E-commerce

Weiqiang Li*

Zhanjiang University of Science and Technology, Zhanjiang, 524086, China

*liyuze@zjkju.edu.cn

Abstract. The rise of e-commerce has seen a surge in personalized recommendation systems, greatly enhancing user experience and boosting sales. This paper explores the integration of multimodal data in e-commerce recommendations and the privacy challenges it brings. Multimodal data, combining different sensory inputs, offers a rich source for recommendations but also raises concerns about data fusion and privacy. To address these challenges, the paper suggests strategies like differential privacy, federated learning, encryption, access control, and anonymization. These approaches aim to maintain effective recommendations while safeguarding user privacy. Looking ahead, the paper discusses future trends in multimodal data privacy, including emerging technologies, standardization, and user education.

Keywords: Multimodal data; Privacy protection; Personalized recommendation

1 Introduction

E-commerce has become a cornerstone of the global economy, driven by technologies like big data and artificial intelligence that power personalized recommendation systems, enhancing user experience and boosting sales [1][2]. However, the increasing volume of user data requires robust privacy measures to prevent data breaches, a pressing concern in e-commerce. Multimodal data, integrating information from text, images, audio, and video [3], enriches recommendation systems by providing a deeper understanding of user interests and needs, leading to more accurate personalized recommendations. However, processing multimodal data introduces new privacy challenges. Balancing effective recommendations with robust privacy protection is critical. This paper explores the application of multimodal data in e-commerce recommendation systems and the related privacy protection challenges, proposing improvement strategies to address these challenges.

© The Author(s) 2024
C. W. K. Chen et al. (eds.), *Proceedings of the 9th International Conference on Engineering Management and the 2nd Forum on Modern Logistics and Supply Chain Management (ICEM-MLSCM 2024)*,
Advances in Engineering Research 243,
https://doi.org/10.2991/978-94-6463-531-7_14

2 Overview of Multimodal Recommendation Systems

2.1 Definition and Types of Multimodal Data

Multimodal data integrating information from various sensory modalities to enhance expressiveness and understanding [4]. Text data includes user reviews, product descriptions, and social media posts, forming the basis for information exchange. Image data adds visual intuitiveness, while audio data provides an additional sensory experience. Video data merges visual and auditory elements, offering dynamic information through product demonstrations and user review videos. User behavior data, including browsing history and purchase records, reflects user interests and preferences, providing insights for personalized recommendations [5].

2.2 Working Principles of Multimodal Recommendation Systems

Multimodal recommendation systems rely on feature extraction, fusion, and analysis to understand user interests deeply [6]. These systems preprocess, clean, and format collected data to ensure quality, using techniques like tokenization and resizing. Feature extraction employs algorithms such as BERT and deep learning models like VGG-16 to capture key features from various data types [7][8]. Feature fusion integrates these features into a unified representation, reflecting comprehensive user interests. Machine learning techniques, including supervised, unsupervised, and reinforcement learning, are used for model training to generate personalized recommendations [9][10]. The system continuously refines recommendation algorithms based on user feedback to enhance accuracy and satisfaction.

2.3 Advantages of Multimodal Data in E-commerce Recommendations

Multimodal recommendation systems are essential in e-commerce for improving user experience and recommendation efficiency. By integrating text, images, audio, and video, these systems provide a comprehensive understanding of products, enhancing recommendation accuracy. Users can quickly grasp a product's appearance through images, understand specifications via text, and experience usage scenarios through audio and video, simulating offline shopping online. These systems analyze user interactions with different data modalities to build interest models, resulting in more accurate recommendations aligned with user needs. For example, frequent viewing of high-resolution images leads to recommendations with high-quality visuals. Overall, multimodal systems enhance the shopping experience by offering diverse sensory channels, making it more engaging and aiding decision-making by providing comprehensive product information, reducing decision time, and minimizing post-purchase regret.

3 Challenges in Privacy Protection for Multimodal Data

Despite significant progress in multimodal data research in e-commerce, several shortcomings remain. Firstly, fusing and integrating multimodal data is challenging due to

the difficulty in leveraging complementary information and managing data heterogeneity and scale differences. Secondly, high-quality annotated data is costly and time-consuming to obtain, and quality discrepancies can affect model performance. Additionally, the complexity and high computational cost of multimodal data models limit practical application. Effective cross-modal information utilization is also difficult, with current methods lacking sufficient fusion and transmission capabilities. Privacy and security concerns are critical, as the collection and use of multimodal data involve user privacy protection. This field faces significant challenges.

3.1 Data Fusion and Privacy Leakage Risks

Data fusion is vital for multimodal recommendation systems, combining diverse data to understand user needs comprehensively. However, this process can pose privacy risks, particularly with personally identifiable information (PII). Different data sources may have varying privacy standards, requiring strict adherence to relevant policies during fusion. Data processing, such as transformation and encoding, can inadvertently expose personal information, necessitating careful workflow design to minimize breaches. Sensitive user behavior data, like clicks and purchase histories, require additional caution. Moreover, the complexity of algorithms used in model training can obscure transparency, making it challenging for users to understand data processing and increasing the risk of privacy breaches. Developing transparent, interpretable algorithms and ensuring user understanding of data use are essential for privacy protection.

3.2 Diversity and Complexity of Multimodal Data

Handling multimodal data presents a significant challenge in modern recommendation systems due to its richness and complexity. Multimodal data, including text, images, audio, and video, varies in format, structure, and relationships, complicating processing and integration to accurately meet user needs. This complexity manifests in three aspects: the variety of data types, requiring specific techniques for feature extraction; the diversity of data sources, necessitating management of data from different channels; and the diversity of user interactions, where behaviors such as browsing and sharing become important data sources. The challenge lies in associating, processing, and fusing these data types. Correlations exist between modalities, demanding deep learning, pattern recognition, and signal processing techniques. Effective data fusion involves integrating these modalities to consider all relevant information, requiring efficient algorithms to handle large-scale datasets and intelligent strategies to maintain data consistency and integrity.

3.3 Sensitivity of User Behavior Data

User behavior data in multimodal recommendation systems poses significant privacy protection issues, including browsing history, purchase records, and social media interactions. This data reflects personal preferences and potentially reveals identity, making

privacy and security crucial. Unauthorized access could lead to privacy violations, especially as collection and analysis processes can be channels for breaches. Large data volumes make systems vulnerable to cyberattacks or misuse. Furthermore, recommendation systems create user profiles based on this data, influencing the information and services users receive, potentially limiting choices and influencing decision-making.

4 Privacy Protection Techniques for Multimodal Data

4.1 Differential Privacy

Differential Privacy is a statistical technique that protects individual privacy by ensuring that the probability distribution of query results does not significantly change when a single record is added or removed from the database. This approach prevents attackers from determining if a specific record exists, even if they have access to all other data. A randomized algorithm A satisfies ε-differential privacy if, for all neighboring datasets D and D' differing by at most one element, and for all S in the output domain of A, the inequality holds (see Equation 1):

$$\Pr[A(D) \in S] \leq e^{\varepsilon} \cdot \Pr[A(D') \in S] \tag{1}$$

Differential privacy is typically implemented by introducing random noise during data publication or queries, with the noise level controlled by the parameter ε. A smaller ε implies stronger privacy protection but may affect data accuracy (see Fig. 1). The principle is that if an algorithm's output behavior remains nearly unchanged for any two datasets differing by one element, it can protect individual data privacy. Suppose a hospital wants to publish the average age of its patients without revealing any individual's age. Using differential privacy, the hospital can add some random noise to the calculated average age. This way, even if an individual's data is added or removed, the published average age will only be slightly affected, thus protecting personal privacy.

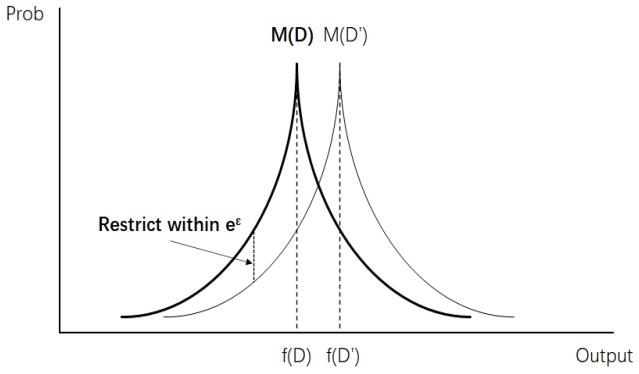

Fig. 1. Probability of Random Algorithms on Neighboring Datasets

4.2 Federated Learning

Federated Learning (FL) is a distributed machine learning framework that enables collaborative model building without sharing raw data, ensuring data privacy and security. In FL, a central server sends a global model to clients, who train locally and send model updates back to the server for aggregation. This process allows FL to be applied in privacy-sensitive domains like healthcare and finance. FL's key strength lies in its data privacy protection, as data remains local, reducing the risk of breaches. Moreover, FL is communication-efficient, transmitting only model updates instead of raw data. Training on diverse local datasets also improves model generalization (see Equation 2). For instance, several banks want to jointly develop a credit scoring model. Each bank has its own customer data but does not want to share this data directly due to privacy and security concerns. Through federated learning, each bank can train part of the model on its own data and then send only the model updates to a central server for aggregation. This way, all banks share the knowledge of the model without sharing specific customer data.

$$w_{t+1} = w_t + \eta \sum_{k=1}^{K} \frac{n_k}{n} (w_{t+1,k} - w_t) \tag{2}$$

4.3 Anonymization and Obfuscation Techniques

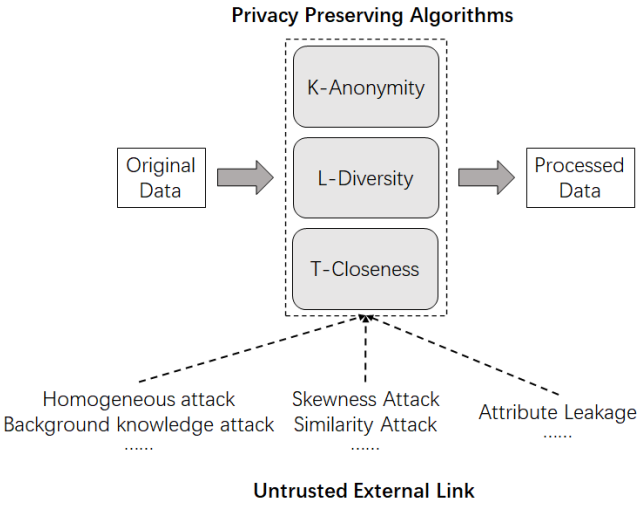

Fig. 2. Data Desensitization Method

Balancing privacy protection with data usability in multimodal environments is challenging. Anonymization and obfuscation techniques are vital for minimizing the risk of identifying individuals while maintaining data utility. Anonymization methods like data desensitization, pseudonymization, aggregation, generalization, and randomization transform personal information into an untraceable state. Obfuscation techniques,

such as traffic obfuscation, protect users' communication behaviors. Desensitization algorithms like K-Anonymity, L-Diversity, and T-Closeness quantify identifiability levels in datasets (see Fig. 2). K-Anonymity ensures each record is indistinguishable from at least k-1 others, preventing attackers from determining complete information. However, it may be vulnerable to homogeneity attacks. L-Diversity requires each equivalence class to have at least l different sensitive attribute values, mitigating this risk. T-Closeness enhances privacy by ensuring the distribution of sensitive attributes within an equivalence class is close to the overall dataset distribution, reducing the risk of inference attacks.

5 Conclusion

The integration of multimodal data in e-commerce has undoubtedly transformed personalized recommendation systems, providing a richer and more engaging user experience. However, this integration will undoubtedly affect user privacy. This article highlights the importance of maintaining a delicate balance between leveraging the power of multimodal data to enhance recommendations through multiple channels and protecting user privacy. The development of technology, the establishment of sound standards, and user education will be key to shaping the future landscape of multimodal data privacy. Emerging technologies such as blockchain and homomorphic encryption may further enhance data privacy, while standardization efforts will simplify practices across the industry. User education will enable consumers to make informed decisions about their data sharing preferences.

Acknowledgments

The research is supported by “Zhanjiang University of Science and Technology 2022 National College Student Innovation and Entrepreneurship Training Program” (202212622007S)

References

1. Gupta, U., Wu, C. J., Wang, X., Naumov, M., Reagen, B., Brooks, D., ... & Zhang, X. (2020, February). The architectural implications of facebook's dnn-based personalized recommendation. In 2020 IEEE International Symposium on High Performance Computer Architecture (HPCA) (pp. 488-501). IEEE.
2. Werner, D., Adam, M., & Benlian, A. (2022). Empowering users to control ads and its effects on website stickiness. Electronic Markets, 32(3), 1373-1397.
3. Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. International Journal of Information Management, 48, 108-119.
4. Zhang, Y. D., Dong, Z., Wang, S. H., Yu, X., Yao, X., Zhou, Q., ... & Gorriz, J. M. (2020). Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation. Information Fusion, 64, 149-187.

5. Salah, A., Truong, Q. T., & Lauw, H. W. (2020). Cornac: A comparative framework for multimodal recommender systems. Journal of Machine Learning Research, 21(95), 1-5.
6. Cai, W., Song, Y., & Wei, Z. (2021). Multimodal Data Guided Spatial Feature Fusion and Grouping Strategy for E-Commerce Commodity Demand Forecasting. Mobile Information Systems, 2021(1), 5568208.
7. Kim, K., & Park, S. (2023). AOBERT: All-modalities-in-One BERT for multimodal sentiment analysis. Information Fusion, 92, 37-45.
8. Ghosh, T., & Jayanthi, N. (2024). An efficient Dense-Resnet for multimodal image fusion using medical image. Multimedia Tools and Applications, 1-28.
9. Gomez, R., Gomez, L., Gibert, J., & Karatzas, D. (2019). Self-supervised learning from web data for multimodal retrieval. In Multimodal Scene Understanding (pp. 279-306). Academic Press.
10. Rudovic, O., Zhang, M., Schuller, B., & Picard, R. (2019, October). Multi-modal active learning from human data: A deep reinforcement learning approach. In 2019 international conference on multimodal interaction (pp. 6-15).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.