Sistemas de Recomendación con Privacidad Preservada: Aplicación de Privacidad Diferencial a Modelos basados en SVD

Diego-Esteban Quintero-Rey,
Ciberseguridad, 2023-I
Ingeniería de Sistemas y Computación,
Universidad Nacional de Colombia

***Abstract*— This study investigates the application of differential privacy in recommendation systems using the SVD matrix factorization model. The methodology involves data preparation, data perturbation using the Laplace mechanism, modification of the SVD model, and evaluation of the modified model's performance. Results show that increasing the privacy budget parameter (ϵ) leads to a convergence of the modified model towards the original model, balancing privacy preservation and model accuracy. However, challenges in generalization were observed, highlighting the need for further research. The conclusions emphasize the importance of considering privacy in recommendation systems, suggesting future directions for improving generalization, exploring alternative privacy mechanisms, and applying the approach to other domains and datasets. Overall, this study contributes to the understanding of differential privacy in sensitive recommendation environments.**

I. INTRODUCCIÓN

En la era actual de grandes cantidades de datos y avances en el aprendizaje automático, los sistemas de recomendaciones se han vuelto indispensables para personalizar y mejorar las experiencias de los usuarios en una variedad de aplicaciones, como comercio electrónico, música, películas y más. Sin embargo, el uso de estos sistemas también ha generado preocupaciones significativas sobre la privacidad y la protección de la información personal sensible de los usuarios.

En este contexto, la privacidad diferencial ha surgido como un enfoque prometedor para abordar las preocupaciones de privacidad en los sistemas de recomendaciones. La privacidad diferencial ofrece una forma de proteger los datos personales al introducir ruido controlado en los datos de entrenamiento, lo que dificulta la identificación de información específica de los individuos.

Este artículo presenta un estudio sobre la aplicación de la privacidad diferencial en sistemas de recomendaciones, con un enfoque específico en el modelo de factorización de matrices SVD (Singular Value Decomposition).

II. METODOLOGÍA

El objetivo principal fue aplicar el modelo SVD (Singular Value Decomposition) de factorización de matrices a un conjunto de datos utilizando la librería Surprise y el dataset MovieLens 100k. Se siguieron las siguientes etapas:

1. *Preparación de los datos:*

Se utilizó el conjunto de datos MovieLens 100k como base para el entrenamiento y evaluación del modelo. Se realizó una división del conjunto de datos en datos de entrenamiento y datos de prueba. El conjunto de prueba representó el 20% del total de datos disponibles.

2. *Perturbación de los datos:*

Se aplicó el mecanismo de Laplace para perturbar el conjunto de datos de entrenamiento. La perturbación se basó en una sensibilidad de 4.0, que se calculó como la diferencia entre la calificación máxima (5.0) y la calificación mínima (1.0) en el conjunto de datos. Además, se aplicó una restricción (clamp) para que los resultados no sobrepasaran la calificación mínima y máxima. Se utilizó un parámetro de privacidad ϵ dado para controlar el nivel de privacidad diferencial en la perturbación de datos. El mecanismo de Laplace consiste en agregar ruido a los datos usando la siguiente función de distribución:

$$f(x \mid \mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} \text{ con } b = \frac{\Delta f}{\epsilon}$$

Donde Δf es la sensibilidad y ϵ es el parámetro de privacidad.

3. *Modificación del modelo SVD:*

Se realizó una modificación en el método de entrenamiento de la clase SVD para perturbar los gradientes durante el proceso de entrenamiento del modelo. Los parámetros del modelo SVD incluyeron la tasa de aprendizaje γ, el parámetro de regularización λ, el número de iteraciones para la perturbación de gradientes y el número de factores para el SVD. Además, se utilizó el valor de ϵ dado.

Tabla 1 - Parámetros utilizados

Número de factores	3
Regularización λ	0.06
Tasa de aprendizaje γ	0.001
Número de iteraciones (perturb. de gradientes)	5

La figura 1 muestra el algoritmo utilizado para la perturbación de gradientes, tomado de [1]. En nuestro caso, las puntuaciones preprocesadas corresponden al conjunto de datos perturbado. Se utilizó un valor de $e_{\max} = 2.0$ según se recomienda en [1].

Input:
$R = \{r_{ui}\}$ – preprocessed user ratings,
d – number of factors,
γ – learning rate parameter,
λ – regularization parameter,
k – number of gradient descent iterations,
$e_{\max}$ – upper bound on per-rating error,
ϵ – privacy parameter
Output:
Latent factor matrices $P_{n\times d}$ and $Q_{m\times d}$

Initialize random factor matrices P and Q.
for *k iterations* **do**
  for *each* $r_{ui} \in R$ **do**
    $e'_{ui} = r_{ui} - p_u q_i^\intercal + \text{Laplace}(k\Delta r/\epsilon)$
    Clamp e'_{ui} to $[-e_{\max}, e_{\max}]$
    $q_i \leftarrow q_i + \gamma(e'_{ui} \cdot p_u^\intercal - \lambda \cdot q_i)$
    $p_u \leftarrow p_u + \gamma(e'_{ui} \cdot q_i^\intercal - \lambda \cdot p_u)$
return P and Q.

Figura 1 - Algoritmo de perturbación de gradientes para SVD

4. *Entrenamiento del modelo SVD modificado:*

Se aplicó la técnica de validación cruzada de K pliegues con $K = 5$ al conjunto de datos de entrenamiento, usando una partición de 20% para validación. En cada pliegue de entrenamiento, se aplicó la función de perturbación de los datos y se entrenó el modelo SVD con la perturbación de gradientes. La perturbación de gradientes se realizó el número de veces dado por el número de iteraciones definido.

5. *Evaluación del modelo:*

Se calculó el promedio del error cuadrático medio (RMSE) obtenido en cada pliegue como medida de evaluación del modelo. También se obtuvo el RMSE sobre los datos de prueba usando el modelo modificado y el modelo sin modificaciones. Al final se graficaron los resultados en función del parámetro de privacidad ϵ.

III. RESULTADOS

La figura 2 muestra los resultados obtenidos.

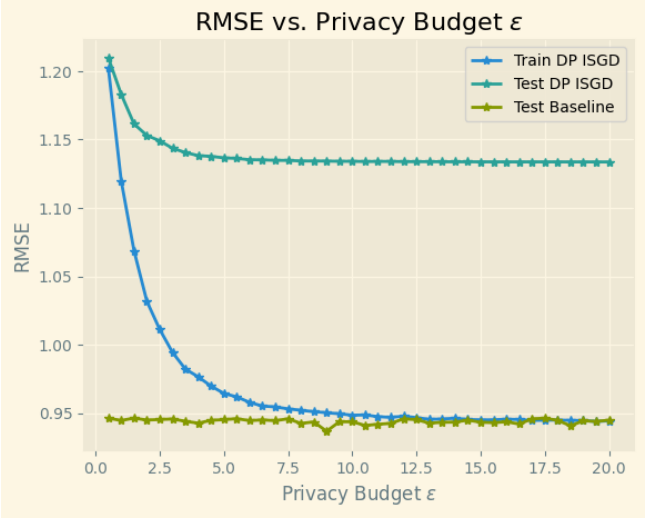

Figura 2 - RMSE vs. Privacy Budget

A partir de los resultados obtenidos, se pueden destacar las siguientes conclusiones:

1. *Impacto de ϵ en el rendimiento del modelo:*

La gráfica de RMSE vs. ϵ muestra que a medida que se incrementa el valor de ϵ (privacy budget), la curva de RMSE sobre los datos de entrenamiento se acerca gradualmente a la curva del modelo sin modificaciones. Esto era esperado, ya que a medida que se reduce la cantidad de ruido agregado con el mecanismo de Laplace, el modelo se aproxima más a su rendimiento original. Sin embargo, es importante tener en cuenta que este acercamiento al modelo sin modificaciones se da a expensas de la privacidad diferencial. A medida que se incrementa ϵ, se reduce la protección de la información privada en los datos de entrenamiento.

2. *Problemas de generalización del modelo modificado:*

La curva de RMSE sobre los datos de prueba del modelo modificado se encuentra considerablemente alejada de las curvas del modelo sin modificaciones y del RMSE sobre los datos de entrenamiento. Esto sugiere que el modelo modificado está teniendo dificultades para generalizar y obtener resultados precisos en datos no vistos previamente. La perturbación de los gradientes durante el entrenamiento del modelo puede estar introduciendo ruido adicional que afecta negativamente la capacidad de generalización del modelo. Es posible que se requieran ajustes en los parámetros o técnicas adicionales para mejorar la capacidad de generalización del modelo modificado.

3. *Beneficios y compensaciones de la privacidad diferencial:*

El análisis realizado permite observar el comportamiento esperado de que a medida que se incrementa ϵ, disminuye el error RMSE. Esto indica que al aumentar el nivel de privacidad diferencial (mayor ϵ), se reduce el impacto negativo del ruido en la calidad de las recomendaciones. Sin embargo, es fundamental tener un equilibrio entre privacidad y calidad de

las recomendaciones. Es necesario considerar las necesidades y requisitos específicos del contexto en el que se implementa el sistema de recomendación, para determinar el valor de ϵ adecuado y encontrar el balance óptimo entre privacidad y precisión en las recomendaciones.

IV. CONCLUSIONES

En general, este proyecto ha resaltado la importancia de considerar la privacidad diferencial al desarrollar sistemas de recomendaciones en entornos sensibles. La privacidad de los datos de los usuarios es una preocupación creciente en la era de la recopilación masiva de información y el aumento de las amenazas cibernéticas. La privacidad diferencial se ha convertido en un enfoque prometedor para abordar esta preocupación al permitir que los datos se utilicen de manera agregada y anónima, sin revelar información personal o sensible.

Aunque la perturbación de datos y la modificación del modelo SVD han presentado desafíos en términos de generalización y precisión, la privacidad diferencial ofrece beneficios significativos al proporcionar una capa adicional de protección de la información sensible. Los resultados obtenidos en este proyecto han demostrado que al aumentar el valor de ϵ (privacy budget), se logra una reducción gradual en el error RMSE y una mayor aproximación al rendimiento del modelo sin modificaciones. Sin embargo, es esencial encontrar el equilibrio adecuado entre la privacidad y la calidad de las recomendaciones, ya que un mayor nivel de privacidad puede comprometer la precisión del modelo.

En cuanto a los posibles próximos pasos en la investigación, se sugieren las siguientes direcciones:

Mejora de la capacidad de generalización: Es necesario investigar enfoques y técnicas adicionales para mejorar la capacidad de generalización del modelo modificado. Esto podría implicar explorar métodos de regularización más eficientes, ajustes finos en los parámetros del modelo o incluso la incorporación de algoritmos de aprendizaje automático más avanzados. El objetivo es lograr un equilibrio entre la privacidad y la precisión, donde el modelo pueda proporcionar recomendaciones útiles y relevantes sin comprometer la protección de la privacidad.

Evaluación de otros mecanismos de privacidad diferencial: Además del mecanismo de Laplace utilizado en este proyecto, existen otros enfoques de privacidad diferencial, como el mecanismo de ruido de Gauss o la privacidad basada en el árbol de decisiones. Investigar la viabilidad y eficacia de estos enfoques en sistemas de recomendaciones podría ser una línea de investigación interesante.

Exploración de otras métricas de evaluación: Además del RMSE, es importante considerar otras métricas de evaluación, como la precisión, el recall y el f1 score. Estas métricas brindan una visión más completa del rendimiento del modelo en términos de la calidad de las recomendaciones realizadas. Investigar cómo estas métricas se ven afectadas por la privacidad diferencial y cómo se pueden optimizar en modelos modificados sería un paso importante para comprender mejor la relación entre privacidad y rendimiento del modelo.

Aplicación a otros dominios y conjuntos de datos: Este proyecto se centró en el conjunto de datos MovieLens 100k y en sistemas de recomendaciones de películas. Sin embargo, sería interesante evaluar el enfoque de privacidad diferencial en otros dominios y conjuntos de datos, como recomendaciones de productos, música, libros, entre otros. Cada dominio puede presentar desafíos y características únicas, lo que ampliaría la comprensión de la aplicabilidad de la privacidad diferencial en diferentes contextos.

Mediante mejoras en la capacidad de generalización, exploración de otros mecanismos de privacidad diferencial, evaluación de diferentes métricas y aplicaciones en diversos dominios, se pueden lograr avances significativos en la protección de la privacidad y la precisión de los sistemas de recomendaciones.

REFERENCES

[1] A. Berlioz, A. Friedman, M. A. Kaafar, R. Boreli, and S. Berkovsky, "Applying Differential Privacy to Matrix Factorization," NICTA, Australia, doi: https://doi.org/10.1145/2792838.2800173.

[2] X. Zhu and Y. Sun, "Differential Privacy for Collaborative Filtering Recommender Algorithm," Department of Software Engineering, Shandong University and Department of Computer Science, The University of Hong Kong, doi: https://doi.org/10.1145/2875475.2875483.

[3] J. Yang, X. Li, Z. Sun, and J. Zhang, "A Differential Privacy Framework for Collaborative Filtering," College of Computer and Control Engineering, Qiqihar University, Qiqihar, China, doi: https://doi.org/10.1155/2019/1460234.

[4] A. Friedman, S. Berkovsky, and M. A. Kaafar, "A differential privacy framework for matrix factorization recommender systems," doi: 10.1007/s11257-016-9177-7 .

[5] L. Fang, B. Du, and C. Wu, "Differentially private recommender system with variational autoencoders," Department of Computer Science, The University of Hong Kong, Hong Kong, doi: https://doi.org/10.1016/j.knosys.2022.109044.

[6] J. Ren, X. Xu, Z. Yao, and H. Yu, "Recommender Systems Based on Autoencoder and Differential Privacy," Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai, China, doi: DOI 10.1109/COMPSAC.2019.00059.