
SSLRec

**A Self-Supervised Framework for
Recommendation**



Importancia y Desafíos de Sistemas Recomendadores

- **Rol Clave:** Ayudan a los usuarios a encontrar contenido relevante.
- **Avances Recientes:** Uso de Aprendizaje Profundo mejora Recomendaciones.
- **Desafíos:**
 - **Escasez de datos:** Pocas interacciones en ítems o usuarios de baja actividad.
 - **Ruido en datos:** Interacciones no siempre reflejan intereses reales.
- **Dependencia de etiquetas:** Sistemas tradicionales necesitan datos etiquetados de alta calidad, difíciles de obtener.

Aprendizaje Auto-supervisado

- **¿Qué es?:** Técnica de aprendizaje que **no requiere etiquetas explícitas**.
- **Ventajas:** Ayuda a mejorar la robustez frente a datos escasos o ruidosos.

Conceptos Clave en SSL

Aumentación de Datos (Data Augmentation):

- Proceso de modificar o transformar los datos originales para crear variaciones adicionales.
- Objetivo: Generar más datos a partir de datos limitados, mejorando la capacidad del modelo para generalizar.

Vistas Aumentadas:

- Diferentes versiones de los datos originales obtenidas a través de aumentación.
-

Problemática

- **Falta de enfoque en SSL:** La mayoría de frameworks no cubre en profundidad modelos auto-supervisados.
- **Cobertura limitada de escenarios:** Solo unos pocos abordan recomendaciones colaborativas y secuenciales con SSL.
- **Ausencia de herramientas unificadas:** Falta de módulos estandarizados para data augmentation, funciones de pérdida de SSL, y evaluación consistente.

Estado del Arte

- **ReChorus:** framework diseñado para capturar y modelar las demandas dinámicas de los usuarios.
- Limitaciones:
 - Solo incluye 2 modelos de SSL.
 - Solo abarca 2 escenarios de recomendación.
 - Solo incluye 4 datasets.
 - No tiene ajuste de parámetros automático.

Chenyang Wang, Min Zhang, Weizhi Ma, Yiqun Liu, and Shaoping Ma. 2020. Make it a chorus: knowledge-and time-aware item modeling for sequential recommendation. In SIGIR. 109–118.

Estado del Arte

- **Beta-recsys:** Framework diseñado para construir, evaluar y optimizar sistemas de recomendación.
- Limitaciones:
 - Solo incluye 3 modelos de SSL.
 - Solo abarca 4 escenarios de recomendación.

Zaiqiao Meng, Richard McCreadie, Craig Macdonald, Iadh Ounis, Siwei Liu, Yaxiong Wu, Xi Wang, Shangsong Liang, et al. 2020. Beta-rec: Build, evaluate and tune automated recommender systems. In Recsys. 588–590.

Estado del Arte

- **RecBole:** Framework de código abierto para sistemas de recomendación, diseñado para unificar y simplificar la implementación y evaluación de algoritmos de recomendación.
- Limitaciones:
 - Solo incluye 11 modelos de SSL.

WayneXinZhao, Shanlei Mu, Yupeng Hou, Zihan Lin, Yushuo Chen, Xingyu Pan, Kaiyuan Li, Yujie Lu, Hui Wang, Changxin Tian, et al. 2021. Recbole: Towards a unified, comprehensive and efficient framework for recommendation algorithms. In CIKM. 4653–4664.

Estado del Arte

- **SELFRec:** Framework de sistemas de recomendación enfocado específicamente en modelos de Aprendizaje Auto-supervisado.
- Limitaciones:
 - Solo incluye 11 modelos de SSL.
 - Solo abarca 2 escenarios de recomendación.
 - Solo incluye 4 datasets.
 - No tiene ajuste de parámetros automático.

Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Jundong Li, and Zi Huang. 2022. Self-supervised learning for recommender systems: A survey. arXiv preprint arXiv:2203.15876 (2022).

¿Por qué SSLRec?

- Limitaciones de los enfoques actuales:
 - Inconsistencias en implementaciones.
 - Comparación poco justa entre modelos.
 - Limitada cantidad de herramientas disponibles.
- Propuesta:
 - Crear un marco de trabajo **estándar, flexible y modular** para comparar de manera justa y consistente modelos de recomendación basados en SSL en distintos escenarios.

Escenarios Implementados

Se implementan 5 escenarios de recomendación tradicionales que pueden ser mejorados a través de técnicas de SSL:

- **General Collaborative Filtering (CF)**
 - **Multi-Behavior Recommendation**
 - **Sequential Recommendation**
 - **Social Recommendation**
 - **Knowledge Graph-enhanced Recommendation**
-

Herramientas de SSL en SSLRec

Se abordan tres paradigmas de SSL:

- **Contrastive SSL:** Comparación entre vistas aumentadas, acercando ejemplos similares y alejando los disimilares.
 - **Generative SSL:** Reconstrucción de datos de entrada para mejorar la representación.
 - **Predictive SSL:** Predicción de atributos específicos como etiquetas auto-generadas.
-

Framework SSLRec



Modularización

Ofrece módulos para data augmentation y objetivos (loss functions)



Interfaz

Todos los modelos siguen la misma interfaz, asegurando métodos públicos



Ejecución unificada

El entrenamiento y validación siguen un flujo de ejecución agnóstico del modelo

Módulos

Data Augmentation

- Data-based
 - Basada en Grafos
 - Recomendación social
 - Recomendación secuencial
- Feature-based
 - Basadas en embeddings aprendidos

Objetivos (loss functions)

- Contrastive/Predictive SSL
 - InfoNCE
 - DirectAU
 - Generative SSL
 - Cross Entropy
 - Scaled Cosine Error (SCE)
 - Divergencias
 - Kullback-Leibler (KL)
 - Jensen-Shannon (JS)
-

Interfaz estándar



forward

Genera el embedding para la predicción



cal_loss

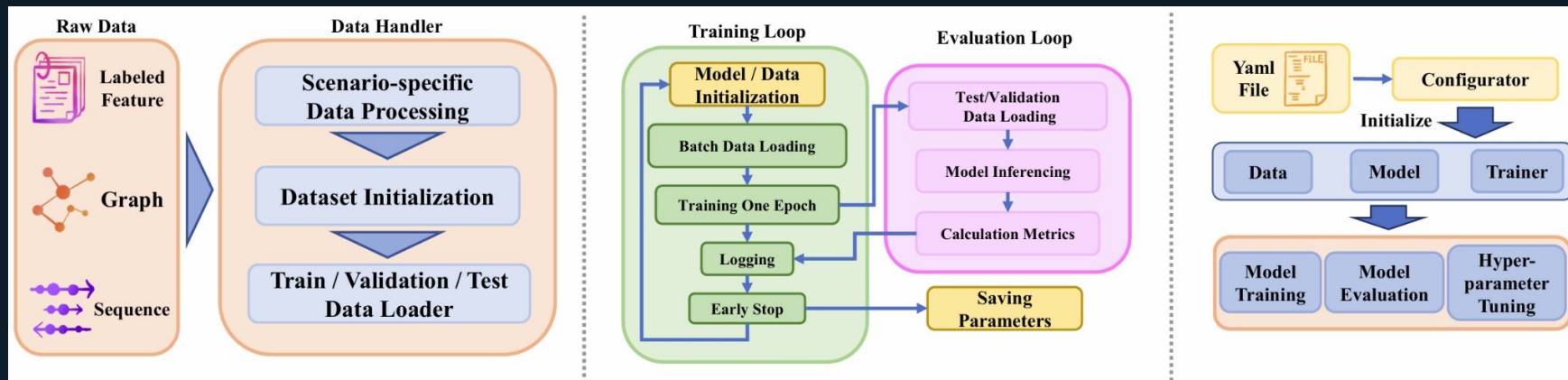
Hace la predicción en base al embedding y calcular la loss function



full_predict

Utilizado para evaluar, calcula los scores de preferencia para todos los items

Ejecución unificada



1

Flujo de datos

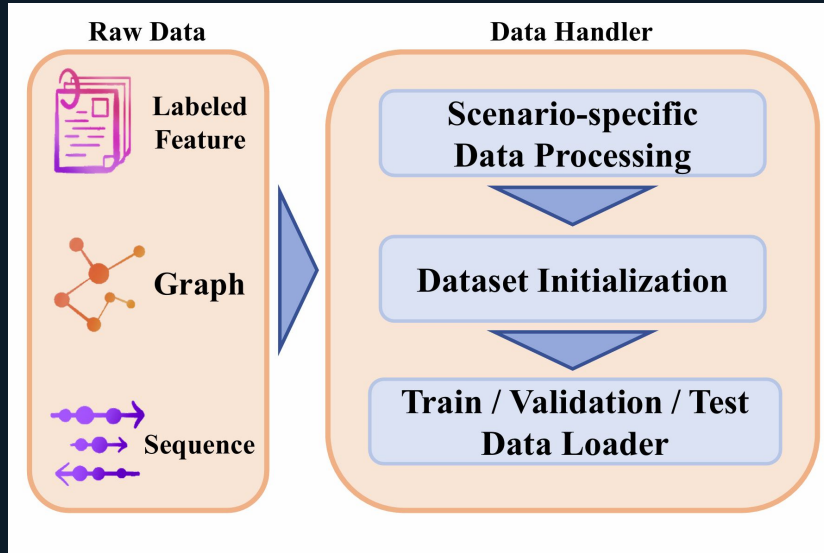
2

Entrenamiento y
validación

3

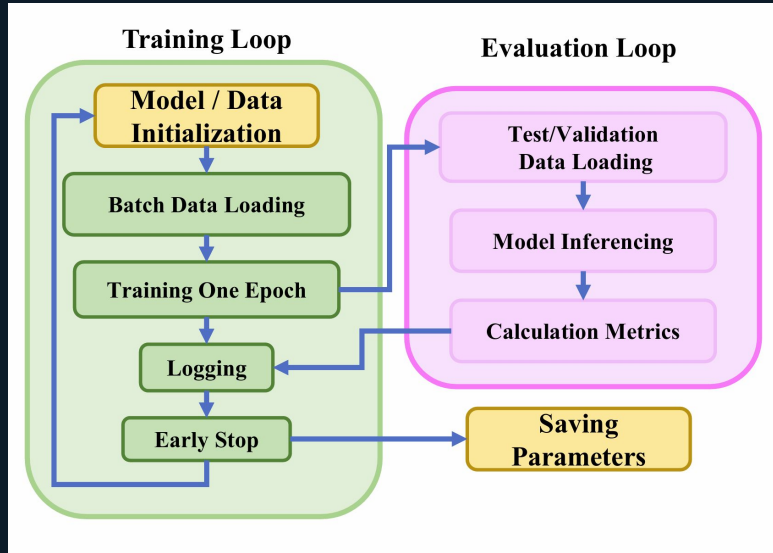
Ajuste de hiper
parámetros

1. Flujo de datos



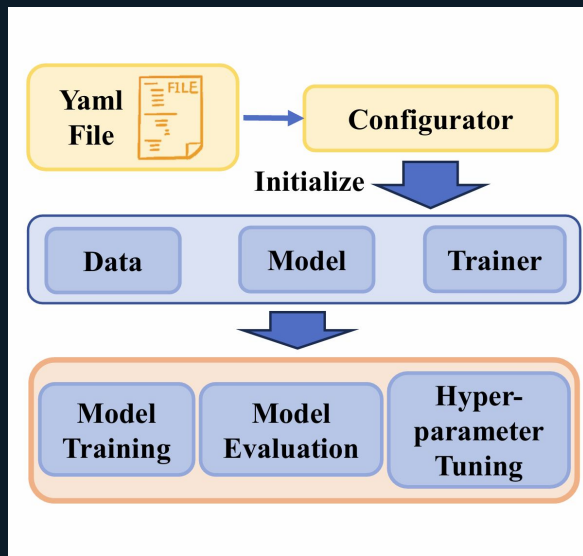
- Se aplican técnicas de preprocesamiento estándares, para asegurar calidad.
- DataHandler carga los datos y los transforma en estructuras de datos
- Dataset crea objetos DataLoader de pytorch para entrenar y evaluar los modelos

2. Entrenamiento y Validación



- Clase Trainer se encarga de todo el proceso de entrenamiento y evaluación
- Para modelos que requieren un entrenamiento específico, se pueden sobrescribir métodos del entrenador

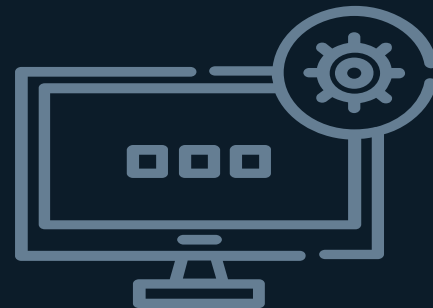
3. Ajuste de hiper parámetros



- Modelos se configuran con archivo YAML
- El módulo configurator inicializa instancias necesarias según archivo YAML
- Se pueden especificar rangos para los hiper parámetros, automatizando la búsqueda de valores óptimos

Benchmarking

Evaluación y comparación de modelos



Evaluación de modelos

- Se presentan los resultados de la evaluación de modelos según los distintos escenarios de recomendación.
- Los modelos son evaluados según su ranking-based performance, utilizando las métricas Recall@K y nDCG@K.

$$Recall@K = \frac{|FirstKRecommendations \cap RelevantItems|}{|RelevantItems|}$$

$$DCG@K = \sum_{i=1}^k \frac{2^{rel_i} - 1}{\log_2(1 + i)} \quad nDCG@K = \frac{DCG@K}{iDCG@K}$$

rel_i : Relevancia asignada al ítem i

$iDCG@K$: DCG@K óptimo

General Collaborative Filtering

Table 3: Model performance reproduced by SSLRec for General Collaborative Filtering on Gowalla dataset.

Method	R@10	R@20	R@40	N@10	N@20	N@40
LightGCN	0.1526	0.2258	0.3199	0.1230	0.1451	0.1716
SGL	0.1640	0.2369	0.3268	0.1321	0.1540	0.1795
HCCF	0.1634	0.2370	0.3239	0.1305	0.1529	0.1778
SimGCL	0.1605	0.2358	0.3335	0.1302	0.1529	0.1802
NCL	0.1666	0.2441	0.3401	0.1342	0.1574	0.1844
DirectAU	0.1533	0.2276	0.3221	0.1211	0.1439	0.1702
LightGCL	0.1607	0.2387	0.3395	0.1285	0.1522	0.1807
AutoCF	0.1747	0.2506	0.3482	0.1407	0.1635	0.1913
DCCF	0.1672	0.2464	0.3476	0.1352	0.1589	0.1873
GFormer	0.1730	0.2487	0.3456	0.1426	0.1651	0.1924

Gowalla Dataset - #users: 25557, #items: 19747, #interacciones: 294983

General Collaborative Filtering

Table 3: Model performance reproduced by SSLRec for General Collaborative Filtering on Gowalla dataset.

Method	R@10	R@20	R@40	N@10	N@20	N@40
LightGCN	0.1526	0.2258	0.3199	0.1230	0.1451	0.1716
SGL	0.1640	0.2369	0.3268	0.1321	0.1540	0.1795
HCCF	0.1634	0.2370	0.3239	0.1305	0.1529	0.1778
SimGCL	0.1605	0.2358	0.3335	0.1302	0.1529	0.1802
NCL	0.1666	0.2441	0.3401	0.1342	0.1574	0.1844
DirectAU	0.1533	0.2276	0.3221	0.1211	0.1439	0.1702
LightGCL	0.1607	0.2387	0.3395	0.1285	0.1522	0.1807
AutoCF	0.1747	0.2506	0.3482	0.1407	0.1635	0.1913
DCCF	0.1672	0.2464	0.3476	0.1352	0.1589	0.1873
GFormer	0.1730	0.2487	0.3456	0.1426	0.1651	0.1924

Gowalla Dataset - #users: 25557, #items: 19747, #interacciones: 294983

General Collaborative Filtering

Table 3: Model performance reproduced by SSLRec for General Collaborative Filtering on Gowalla dataset.

Method	R@10	R@20	R@40	N@10	N@20	N@40
LightGCN	0.1526	0.2258	0.3199	0.1230	0.1451	0.1716
SGL	0.1640	0.2369	0.3268	0.1321	0.1540	0.1795
HCCF	0.1634	0.2370	0.3239	0.1305	0.1529	0.1778
SimGCL	0.1605	0.2358	0.3335	0.1302	0.1529	0.1802
NCL	0.1666	0.2441	0.3401	0.1342	0.1574	0.1844
DirectAU	0.1533	0.2276	0.3221	0.1211	0.1439	0.1702
LightGCL	0.1607	0.2387	0.3395	0.1285	0.1522	0.1807
AutoCF	0.1747	0.2506	0.3482	0.1407	0.1635	0.1913
DCCF	0.1672	0.2464	0.3476	0.1352	0.1589	0.1873
GFormer	0.1730	0.2487	0.3456	0.1426	0.1651	0.1924

Gowalla Dataset - #users: 25557, #items: 19747, #interacciones: 294983

General Collaborative Filtering

Table 3: Model performance reproduced by SSLRec for General Collaborative Filtering on Gowalla dataset.

Method	R@10	R@20	R@40	N@10	N@20	N@40
LightGCN	0.1526	0.2258	0.3199	0.1230	0.1451	0.1716
SGL	0.1640	0.2369	0.3268	0.1321	0.1540	0.1795
HCCF	0.1634	0.2370	0.3239	0.1305	0.1529	0.1778
SimGCL	0.1605	0.2358	0.3335	0.1302	0.1529	0.1802
NCL	0.1666	0.2441	0.3401	0.1342	0.1574	0.1844
DirectAU	0.1533	0.2276	0.3221	0.1211	0.1439	0.1702
LightGCL	0.1607	0.2387	0.3395	0.1285	0.1522	0.1807
AutoCF	0.1747	0.2506	0.3482	0.1407	0.1635	0.1913
DCCF	0.1672	0.2464	0.3476	0.1352	0.1589	0.1873
GFormer	0.1730	0.2487	0.3456	0.1426	0.1651	0.1924

Gowalla Dataset - #users: 25557, #items: 19747, #interacciones: 294983

General Collaborative Filtering

Table 3: Model performance reproduced by SSLRec for General Collaborative Filtering on Gowalla dataset.

Method	R@10	R@20	R@40	N@10	N@20	N@40
LightGCN	0.1526	0.2258	0.3199	0.1230	0.1451	0.1716
SGL	0.1640	0.2369	0.3268	0.1321	0.1540	0.1795
HCCF	0.1634	0.2370	0.3239	0.1305	0.1529	0.1778
SimGCL	0.1605	0.2358	0.3335	0.1302	0.1529	0.1802
NCL	0.1666	0.2441	0.3401	0.1342	0.1574	0.1844
DirectAU	0.1533	0.2276	0.3221	0.1211	0.1439	0.1702
LightGCL	0.1607	0.2387	0.3395	0.1285	0.1522	0.1807
AutoCF	0.1747	0.2506	0.3482	0.1407	0.1635	0.1913
DCCF	0.1672	0.2464	0.3476	0.1352	0.1589	0.1873
GFormer	0.1730	0.2487	0.3456	0.1426	0.1651	0.1924

Gowalla Dataset - #users: 25557, #items: 19747, #interacciones: 294983

General Collaborative Filtering

Table 3: Model performance reproduced by SSLRec for General Collaborative Filtering on Gowalla dataset.

Method	R@10	R@20	R@40	N@10	N@20	N@40
LightGCN	0.1526	0.2258	0.3199	0.1230	0.1451	0.1716
SGL	0.1640	0.2369	0.3268	0.1321	0.1540	0.1795
HCCF	0.1634	0.2370	0.3239	0.1305	0.1529	0.1778
SimGCL	0.1605	0.2358	0.3335	0.1302	0.1529	0.1802
NCL	0.1666	0.2441	0.3401	0.1342	0.1574	0.1844
DirectAU	0.1533	0.2276	0.3221	0.1211	0.1439	0.1702
LightGCL	0.1607	0.2387	0.3395	0.1285	0.1522	0.1807
AutoCF	0.1747	0.2506	0.3482	0.1407	0.1635	0.1913
DCCF	0.1672	0.2464	0.3476	0.1352	0.1589	0.1873
GFormer	0.1730	0.2487	0.3456	0.1426	0.1651	0.1924

Gowalla Dataset - #users: 25557, #items: 19747, #interacciones: 294983

Sequential Recommendation

Table 4: Performance Evaluation by SSLRec for Sequential Recommender Systems on MovieLen-20M Dataset.

Method	R@5	R@10	R@20	N@5	N@10	N@20
BERT4Rec	0.1493	0.2223	0.3197	0.1003	0.1237	0.1483
CL4SRec	0.1491	0.2245	0.3221	0.1007	0.1250	0.1495
DuoRec	0.1507	0.2242	0.3182	0.1021	0.1258	0.1494
DCRec	0.1569	0.2327	0.3303	0.1067	0.1311	0.1557

MovieLens 20M Dataset - #users: 96726, #items: 10154, #interacciones: 193452

Sequential Recommendation

Table 4: Performance Evaluation by SSLRec for Sequential Recommender Systems on MovieLen-20M Dataset.

Method	R@5	R@10	R@20	N@5	N@10	N@20
BERT4Rec	0.1493	0.2223	0.3197	0.1003	0.1237	0.1483
CL4SRec	0.1491	0.2245	0.3221	0.1007	0.1250	0.1495
DuoRec	0.1507	0.2242	0.3182	0.1021	0.1258	0.1494
DCRec	0.1569	0.2327	0.3303	0.1067	0.1311	0.1557

MovieLens 20M Dataset - #users: 96726, #items: 10154, #interacciones: 193452

Sequential Recommendation

Table 4: Performance Evaluation by SSLRec for Sequential Recommender Systems on MovieLen-20M Dataset.

Method	R@5	R@10	R@20	N@5	N@10	N@20
BERT4Rec	0.1493	0.2223	0.3197	0.1003	0.1237	0.1483
CL4SRec	0.1491	0.2245	0.3221	0.1007	0.1250	0.1495
DuoRec	0.1507	0.2242	0.3182	0.1021	0.1258	0.1494
DCRec	0.1569	0.2327	0.3303	0.1067	0.1311	0.1557

MovieLens 20M Dataset - #users: 96726, #items: 10154, #interacciones: 193452

Sequential Recommendation

Table 4: Performance Evaluation by SSLRec for Sequential Recommender Systems on MovieLen-20M Dataset.

Method	R@5	R@10	R@20	N@5	N@10	N@20
BERT4Rec	0.1493	0.2223	0.3197	0.1003	0.1237	0.1483
CL4SRec	0.1491	0.2245	0.3221	0.1007	0.1250	0.1495
DuoRec	0.1507	0.2242	0.3182	0.1021	0.1258	0.1494
DCRec	0.1569	0.2327	0.3303	0.1067	0.1311	0.1557

MovieLens 20M Dataset - #users: 96726, #items: 10154, #interacciones: 193452

Sequential Recommendation

Table 4: Performance Evaluation by SSLRec for Sequential Recommender Systems on MovieLen-20M Dataset.

Method	R@5	R@10	R@20	N@5	N@10	N@20
BERT4Rec	0.1493	0.2223	0.3197	0.1003	0.1237	0.1483
CL4SRec	0.1491	0.2245	0.3221	0.1007	0.1250	0.1495
DuoRec	0.1507	0.2242	0.3182	0.1021	0.1258	0.1494
DCRec	0.1569	0.2327	0.3303	0.1067	0.1311	0.1557

MovieLens 20M Dataset - #users: 96726, #items: 10154, #interacciones: 193452

Sequential Recommendation

Table 4: Performance Evaluation by SSLRec for Sequential Recommender Systems on MovieLen-20M Dataset.

Method	R@5	R@10	R@20	N@5	N@10	N@20
BERT4Rec	0.1493	0.2223	0.3197	0.1003	0.1237	0.1483
CL4SRec	0.1491	0.2245	0.3221	0.1007	0.1250	0.1495
DuoRec	0.1507	0.2242	0.3182	0.1021	0.1258	0.1494
DCRec	0.1569	0.2327	0.3303	0.1067	0.1311	0.1557

MovieLens 20M Dataset - #users: 96726, #items: 10154, #interacciones: 193452

Social Recommendation

Table 5: Performance Evaluation by SSLRec for Social Recommendation on Yelp (Social) Dataset.

Method	R@10	R@20	R@40	N@10	N@20	N@40
KCGN	0.0202	0.0355	0.0578	0.0094	0.0133	0.0178
SMIN	0.0216	0.0368	0.0599	0.0103	0.0141	0.0188
MHCN	0.0314	0.0525	0.0841	0.0163	0.0216	0.0280
DSL	0.0335	0.0536	0.0846	0.0167	0.0218	0.0281

Yelp Dataset - #users: 43043, #items: 66576, #interacciones: 283512

Social Recommendation

Table 5: Performance Evaluation by SSLRec for Social Recommendation on Yelp (Social) Dataset.

Method	R@10	R@20	R@40	N@10	N@20	N@40
KCGN	0.0202	0.0355	0.0578	0.0094	0.0133	0.0178
SMIN	0.0216	0.0368	0.0599	0.0103	0.0141	0.0188
MHCN	0.0314	0.0525	0.0841	0.0163	0.0216	0.0280
DSL	0.0335	0.0536	0.0846	0.0167	0.0218	0.0281

Yelp Dataset - #users: 43043, #items: 66576, #interacciones: 283512

KG-Enhanced Recommendation

Table 6: Recall (R) and NDCG (N) reproduced by SSLRec for KG-enhanced Recommendation on Alibaba dataset.

Method	R@5	R@10	R@20	N@5	N@10	N@20
KGIN	0.0746	0.1139	0.1678	0.0569	0.0711	0.0870
KGCL	0.0749	0.1147	0.1685	0.0577	0.0720	0.0879
KGRec	0.0708	0.1188	0.1735	0.0595	0.0741	0.0903

Alibaba dataset - #users: 114737, #items: 30040, #interactions: 1781093,
#relations: 51, #entities: 59156, #triplets: 686516

KG-Enhanced Recommendation

Table 6: Recall (R) and NDCG (N) reproduced by SSLRec for KG-enhanced Recommendation on Alibaba dataset.

Method	R@5	R@10	R@20	N@5	N@10	N@20
KGIN	0.0746	0.1139	0.1678	0.0569	0.0711	0.0870
KGCL	0.0749	0.1147	0.1685	0.0577	0.0720	0.0879
KGRec	0.0708	0.1188	0.1735	0.0595	0.0741	0.0903

Alibaba dataset - #users: 114737, #items: 30040, #interactions: 1781093,
#relations: 51, #entities: 59156, #triplets: 686516

Multi-Behavior Recommendation

Table 7: Recall (R) and NDCG (N) reproduced by SSLRec for Multi-behavior Recommendation on Retail Rocket dataset.

Method	R@10	R@20	R@40	N@10	N@20	N@40
MBGMN	0.0405	0.0478	0.0529	0.0227	0.0245	0.0255
HMG-CR	0.0363	0.0446	0.0584	0.0163	0.0184	0.0211
S-MBRec	0.0336	0.0391	0.0593	0.0124	0.0138	0.0180
CML	0.0428	0.0492	0.0570	0.0241	0.0257	0.0272
KMCLR	0.0428	0.0501	0.0557	0.0240	0.0258	0.0269

Retail-rocket dataset - #users: 2174, #items: 30113, #buy: 9551, #view: 75374, #cart: 12456

Conclusiones

- **Un Marco Estandarizado y Completo:** SSLRec proporciona una plataforma modular y unificada que facilita la implementación y evaluación de modelos de recomendación auto-supervisados en varios escenarios.
- **Principales Herramientas y Paradigmas:** Incorpora tres enfoques de SSL: contrastivo, generativo y predictivo, además de herramientas específicas al dominio, como distintas funciones de pérdida.
- **Flexibilidad y Estandarización:** Con su diseño flexible, y protocolos de evaluación estandarizados, SSLRec facilita la experimentación, promueve la reproducibilidad y permite una comparación justa entre modelos.
- **Impacto para la investigación:** Establece un estándar en el campo de SSL, proporcionando una infraestructura versátil para futuras investigaciones.