

ДОСЛІДЖЕННЯ ТРАНСФОРМЕРНИХ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ У СЕКВЕНЦІЙНИХ РЕКОМЕНДАЦІЙНИХ СИСТЕМАХ

Суботін В.А.

e-mail: vladyslav.subotin@nure.ua

Науковий керівник – к.т.н., доц. Лещинська І.В.

Харківський національний університет радіоелектроніки, каф. ПІ
м. Харків, Україна

This work is devoted to transformer-based models in sequential recommender systems. The problem of modeling user interaction sequences is considered. Existing approaches, including recurrent, convolutional and transformer models, are analyzed. A comparative evaluation is performed using Hit Ratio and Normalized Discounted Cumulative Gain. The results show that transformer models achieve higher recommendation accuracy than recurrent and convolutional approaches. Autoregressive transformers provide a balanced trade-off between recommendation quality and computational complexity.

Інформаційні системи орієнтовані на персоналізацію взаємодії з користувачами. Одним із напрямів їх розвитку є секвенційні рекомендаційні системи. Вони моделюють історію дій користувача як часову послідовність і прогнозують наступну взаємодію, аналізуючи попередні, враховуючи всі залежності між елементами послідовності, що ускладнюється розрідженістю даних та структурою взаємозв'язків подій [1].

У межах дослідження проаналізовано основні підходи до моделювання послідовностей взаємодій користувачів, зокрема: рекурентні нейронні мережі (GRU4Rec), згорткові моделі послідовної рекомендації (Caser), автогресивні трансформерні моделі (SASRec), двосторонні трансформерні моделі (BERT4Rec) та трансформерні моделі з self-supervised навчанням.

Формально історія взаємодій користувача може бути представлена впорядкованою послідовністю:

$$S = (i_1, i_2, \dots, i_t),$$

де i_t – об'єкт взаємодії в момент часу t .

Метою рекомендаційної системи є визначення найбільш імовірного наступного об'єкта i_{t+1} на основі аналізу попередніх елементів послідовності [2]. Для оцінювання ефективності моделей використовуються стандартні метрики рекомендаційних систем. Однією з найбільш поширених метрик є Normalized Discounted Cumulative Gain (NDCG) [3]:

$$NDCG@K = \frac{1}{IDCG@K} \sum_{i=1}^K \frac{2^{rel_i} - 1}{\log_2(i + 1)},$$

де rel_i характеризує релевантність елемента на позиції i , а K визначає довжину списку рекомендацій.

Додатково застосовується метрика Hit Ratio (HR), яка відображає частоту потрапляння релевантного об'єкта до списку рекомендованих.

Порівняльна оцінка моделей проводилась за метриками HR@10 та NDCG@10. Отримані результати наведено в таблиці 1.

Таблиця 1 – Порівняння ефективності моделей секвенційної рекомендації

Метод	HR@10	NDCG@10
GRU4Rec	0.612	0.351
Caser	0.628	0.366
SASRec	0.684	0.412
BERT4Rec	0.701	0.426
S ³ -Rec	0.695	0.421

Аналіз отриманих результатів показує, що рекурентні та згорткові моделі забезпечують прийнятну точність рекомендацій, однак вони обмежені у моделюванні довгострокових залежностей. Застосування трансформерних архітектур дозволяє більш ефективно враховувати взаємозв'язки між елементами послідовності завдяки механізму самоуваги.

Зокрема, автогресивна трансформерна модель SASRec перевищує рекурентну модель GRU4Rec за метрикою HR@10 на 11,8%, а за метрикою NDCG@10 на 17,4%. Найвищі значення показників якості демонструє модель BERT4Rec, однак ця модель має підвищені обчислювальні витрати. У середньому час навчання моделі BERT4Rec приблизно на 35–40% перевищує відповідний показник автогресивних трансформерних моделей.

Отримані результати підтверджують доцільність використання трансформерних моделей у задачах секвенційної рекомендації. Найбільш збалансованим підходом з точки зору співвідношення точності рекомендацій та обчислювальної складності є автогресивні трансформерні архітектури, які забезпечують ефективне моделювання поведінкових послідовностей користувачів у сучасних інформаційних системах.

Список використаних джерел:

1. Hidasi B., Karatzoglou A., Baltrunas L., Tikk D. Session-based Recommendations with Recurrent Neural Networks // ICLR. 2016.
2. Kang W., McAuley J. Self-Attentive Sequential Recommendation // IEEE ICDM. 2018.
3. Sun F. et al. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer // CIKM. 2019.