

КЛАССИФИКАЦИЯ КОНТЕКСТНО-ЗАВИСИМЫХ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ ДЛЯ РЕКОМЕНДАЦИИ ТОЧЕК ИНТЕРЕСА

Жданов Павел Сергеевич

Аспирант, университет ИТМО, г. Санкт-Петербург,
Россия
pszhdanov@itmo.ru

CLASSIFICATION OF CONTEXT-AWARE RECOMMENDER SYSTEMS FOR RECOMMENDING POINTS OF INTEREST

P. Zhdanov

Summary. To date, context-aware recommender systems for recommending points of interest are effective tools for analyzing user behavior: they not only reveal user preferences, but also allow you to generate relevant recommendations for locations that will be interesting for the user to visit under given conditions. Initially, for such recommendations, various modifications of models based on matrix factorization with the addition of contextual features were used, however, with the development and mass distribution of deep learning models, researchers began to actively use neural networks in context-aware recommender systems to recommend points of interest, which made it possible to deduce the quality of the algorithms to a whole new level. This study offers a comprehensive review and analysis of context-aware recommender approaches for point of interest recommendations based on matrix factorization and deep learning to form a common understanding among researchers interested in the subject. The paper describes various modifications used in the creation of matrix factorization models and changes in neural network architectures to work with various context factors. Also in the study, in addition to the classification of models, the features of each class of models, their advantages and disadvantages are presented.

Keywords: context-aware recommender systems; recommendation of points of interest; matrix factorization; neural networks.

Аннотация. На сегодняшний день контекстно-зависимые рекомендательные системы для рекомендации точек интереса являются эффективными инструментами анализа поведения пользователей: они не только выявляют пользовательские предпочтения, но и позволяют сформировать релевантные рекомендации локаций, которые будут интересны для посещения пользователю в заданных условиях. Изначально для таких рекомендаций использовались различные модификации моделей на основе матричной факторизации с добавлением контекстных признаков, однако с развитием и массовым распространением моделей глубокого обучения, исследователи стали активнее использовать нейронные сети в контекстно-зависимых рекомендательных системах для рекомендации точек интереса, что позволило вывести качество работы алгоритмов на принципиально новый уровень. В этом исследовании предлагается всесторонний обзор и анализ контекстно-зависимых рекомендательных подходов для рекомендаций точек интереса, основанных на матричной факторизации и глубоком обучении, чтобы сформировать общее представление у исследователей, заинтересованных в предмете. В работе описываются различные модификации, используемые при создании моделей матричной факторизации и изменения в архитектуре нейронных сетей для работы с различными контекстными факторами. Также в исследовании, помимо классификации моделей, представлены особенности каждого класса моделей, их достоинства и недостатки.

Ключевые слова: контекстно-зависимые рекомендательные системы; рекомендация точек интереса; матричная факторизация; нейронные сети.

Введение

Для решения проблемы актуальных для пользователя рекомендаций на сегодняшний день существует множество подходов. Большинство из них используют различные данные: геометки, предпочитаемый контент в интернете, социальные связи, данные о банковских транзакциях и т.д. Некоторые подходы пытаются учитывать контекст, в котором находился пользователь при принятии того или иного решения. На основе всех этих данных могут выдаваться абсолютно разные рекомендательные предложения:

от рекомендации одного конкретного объекта, до полноценного плана действий, включающего различные активности, который максимально учитывает пользовательские интересы и заданные им ограничения.

В [1] контекст определяется, как набор различных факторов, которые имеют определенную структуру и которые характеризуют текущее состояние пользователя в момент предоставления ему рекомендации. К контекстным факторам могут относиться время, локация, эмоциональное состояние и т.д. Авторы приводят следующие классификации контекста: по тому,

насколько системе доступна информация о контексте (полностью наблюдаемый, частично наблюдаемый и ненаблюдаемый) и по тому, как контекст меняется с течением времени (статический, динамический). В [2] дополнительно рассматривается классификация контекста по способам его представления: в виде иерархической структуры и в виде матрицы.

Постановка задачи для контекстно-зависимых рекомендательных систем для рекомендации точек интереса выглядит следующим образом: для заданного набора пользователей $U = \{u_1, u_2, \dots, u_n\}$, точек интереса $I = \{i_1, i_2, \dots, i_m\}$, контекстов $C = \{c_1, c_2, \dots, c_d\}$ и рейтингов $R = \{r_1, r_2, \dots, r_d\}$ (пользовательских оценок для каждого предмета при наличии определенного контекста) необходимо построить модель, которая сможет предсказывать рейтинг точки интереса для пользователя.

Целью данной работы является обзор существующих контекстно-зависимых рекомендательных систем и составление классификации данных методов. Поскольку наиболее популярными рекомендательными алгоритмами являются алгоритмы, которые используют матричную факторизацию и алгоритмы, использующие нейронные сети, то в этой работе будет представлена классификация рекомендательных систем на основе используемых моделей, а также выявлены их достоинства и недостатки.

В работе используются сокращения для названий используемых моделей:

1. MF — matrix factorization, матричная факторизация;
2. AE — autoencoder, автокодировщик;
3. MLP — multi-layer perceptron, полносвязная нейронная сеть;
4. RNN — recurrent neural network, рекуррентная нейронная сеть;
5. LSTM — long short-term memory, сеть долгой краткосрочной памяти;
6. GRU — gated recurrent units, управляемый рекуррентный блок;
7. GNN — graph neural network, графовая нейронная сеть;
8. CNN — convolutional neural network, сверточная нейронная сеть.

Также в работе используются сокращения для обозначения метрик:

1. NDCG — normalized discounted cumulative gain;
2. MAE — mean absolute error;
3. RMSE — root mean squared error;
4. HR — hit ratio;
5. MAP — mean average precision;

6. MRR — mean reciprocal rank.

В первой части работы представлены модели на основе матричной факторизации. Во второй — модели, использующие нейронные сети. В заключении представлены выводы по работе, а также определены достоинства и недостатки каждого класса моделей.

1. Рекомендательные системы на основе матричной факторизации.

Одними из самых первых моделей, которые были использованы для контекстно-зависимых рекомендательных систем для рекомендации точек интереса, были модели, основанные на матричной факторизации. В [3], одной из первых работ, авторы разработали контекстно-зависимую рекомендательную систему для рекомендации точек интереса пользователей. В работе предлагается следующая модель для представления рейтингов:

$$r_{ui c_1 \dots c_k} = v_u q_i^T + \bar{i} + b_u + \sum_{j=1}^k b_{i c_j}, \quad (1.1)$$

где $r_{ui c_1 \dots c_k}$ — рейтинг предмета i для пользователя u при наборе контекстных факторов $c_1 \dots c_k$;
 v_u — вектор, представляющий пользователя u ;
 q_i — вектор, представляющий предмет i ;
 \bar{i} — среднее значение рейтинга предмета i ;
 b_u — базовый параметр для пользователя u ;
 $b_{i c_j}$ — базовый параметр для контекста c_j и предмета i .

Базовые параметры — это параметры модели, которые введены для того, чтобы отразить отклонения рейтингов от среднего значения. Для формулы (1.1) строится соответствующая оптимизационная задача минимизации, которая решается с помощью градиентного спуска.

В [4] авторы предлагают использовать некоторые модификации подхода, описанного в [3]. Рассматриваются три модели:

1. Модель, которая предполагает, что контекст влияет на все рейтинги одинаково, независимо от объекта (CAMF-C);
2. Модель, которая предполагает, что контекст влияет на каждый объект уникальным образом (CAMF-CI);
3. Модель, которая предполагает, что контекст влияет на различные группы объектов (CAMF-CC).

Сравнивая перечисленные подходы с другими базовыми моделями, авторы отмечают, что модель, которая рассматривает пары контекст — категории объектов показывает самый лучший результат.

Таблица 1. Контекстно-зависимые рекомендательные системы на основе матричной факторизации.

Название модели	Контекст	Целевая переменная	Метрики
MF + context [3]	Геоданные, время, данные из опроса пользователя	Рейтинг объекта	MAE, precision, recall
CAMF-CC [4]	Геоданные, время, данные из опроса пользователя	Рейтинг объекта	MAE
HCMF [5]	Информация с сенсоров мобильного устройства и статистики по этим данным.	Рейтинг объекта	RMSE, NDCG, Hit@k
FMFMGM [6]	Геоданные	Вероятность посещения места	Precision@k, Recall@k
GEO-MF [7]	Геоданные	Рейтинг объекта	Precision@k, Recall@k
GEO-MF++ [8]	Геоданные	Рейтинг объекта	NDCG@k, Recall@k
STELLAR [9]	Геоданные и время	Вероятность посещения места	Precision@k, Recall@k

Модель HCMF, предложенная в [5] является модификацией модели из [3], которая дополнительно использует скрытый контекст. Для получения скрытого представления контекста авторы дополнительно используют такие методы снижения размерности, как метод главных компонент и полносвязный автокодировщик. Итоговая модель выглядит следующим образом:

$$r_{uic_1 \dots c_k l_d \dots l_d} = v_u q_i^T + \bar{i} + b_u + \sum_{j=1}^k b_{ic_j} + \sum_{j=1}^d b_{ij} l_j, \quad (1.2)$$

где $r_{uic_1 \dots c_k}$ — рейтинг предмета i для пользователя u при наборе контекстных факторов $c_1 \dots c_k$ и при наборе скрытых контекстных факторов $l_1 \dots l_d$;

- v_u — вектор, представляющий пользователя u ;
- q_i — вектор, представляющий предмет i ;
- \bar{i} — среднее значение рейтинга предмета i ;
- b_u — базовый параметр для пользователя u ;
- b_{ic_j} — базовый параметр для контекста c_j и предмета i ;
- d — размерность скрытого пространства;
- b_{ij} — базовый параметр, отражающий отклонения рейтинга предметов от среднего значения для скрытого контекста l_j .

Аналогично [3] для (1.2) формируется оптимизационная задача, которая решается с помощью метода градиентного спуска.

В [6] авторы предлагают модель FMFMGM, которая для рекомендации точек интереса использует социальные и географические факторы. Для моделирования геопространственных признаков исследователи используют модель смеси Гауссовых распределений, а данные о социальных взаимодействиях учитывают матричную факторизацию. В основе этой модели лежат следующие предположения:

- ♦ У каждого пользователя есть свои «центры интересов», в которых точки интереса распределены по нормальному закону.
- ♦ Действия пользователя подчиняются следующей закономерности — чем дальше от центра интереса пользователя находится точка интереса, тем меньше вероятность, что пользователь ее посетит.
- ♦ Социальное взаимодействие также оказывает влияние на выбор пользователя, но не такое значительное, как сами точки интереса.

В [7] авторы предлагают GeoMF — алгоритм, в котором расширяются пользовательские латентные факторы и латентные факторы точек интереса в модели факторизации с помощью векторов активных участков пользователей (значений, характеризующих вероятность появления пользователя в определенной локации), а также векторов влияния точек интереса на локацию (количественная характеристика, описывающая влияние одной точки интереса на все остальные точки интереса). Также в рассматриваемой модели используется следующее предположение: если пользователь не посещал точки интереса, расположенные рядом с теми точками интереса, которые он часто посещает, то, скорее всего, предлагать для посещения эти ближайшие точки интереса ему не релевантно.

В [8] представлена GeoMF++ — модификация GeoMF, главной отличительной особенностью которой является то, что векторы, характеризующие влияние локаций, отображаются в низкоуровневое векторное представление и прибавляются к векторам, характеризующим локацию.

В [9] авторы предлагают модель под названием STELLAR, где дополнительно учитывается временной фактор в решении задачи последовательных рекомендаций точек интереса. Помимо обычной матрицы, кото-

рая описывает взаимодействие пользователя и точки интереса, дополнительно вводятся матрицы, характеризующие взаимодействие точки интереса и времени, а также самих точек интереса друг с другом. Поскольку в работе решается задача последовательного предсказания точек, то авторы ставят перед собой цель ранжирования наиболее подходящих мест для пользователя в заданный момент времени, а не обычную задачу предсказания точек интереса.

Итоговая таблица с описанием особенностей контекстно-зависимых рекомендательных систем на основе матричной факторизации представлена в таблице 1.

Главным достоинством описанных выше моделей является возможность интерпретации полученных результатов. К недостаткам этих моделей можно отнести низкое качество работы этих алгоритмов и необходимость самостоятельно определять набор данных, который используется в качестве пользовательского контекста, что может привести к снижению качества работы алгоритма, в особенности при неправильном выборе контекстных данных.

2. Рекомендательные системы на основе нейронных сетей

В последние несколько лет исследователи стали активно применять модели на основе нейронных сетей при решении задач, связанных с контекстно-зависимыми рекомендательными системами для рекомендации точек интереса. Например, в [10] предлагается новый вариант LSTM, названный ST-CLSTM, который использует фильтры времени и расстояний, чтобы фиксировать пространственно-временные отношения между последовательными отметками о посещениях точек интереса. В частности, один из таких фильтров предназначен для управления обновлением краткосрочного интереса, а другая пара фильтров для управления обновлением долгосрочного интереса.

В [11] авторы разработали модель SLCM, которая использует контекстную информацию в явном и скрытом виде, пытаясь выучить функцию взаимодействия между пользователями, предметами и контекстами. Для извлечения скрытого контекста авторы используют два подхода:

1. Получение вектора скрытого текущего контекста с помощью полносвязного автокодировщика.
2. Получение последовательного скрытого контекста на основе нескольких предшествующих контекстов с помощью LSTM автокодировщика.

В [12] предлагается модель CARA, в основе которой лежит архитектура рекуррентной нейронной сети

с механизмом внимания для предсказания точек интереса. Особенностью архитектуры являются фильтры контекста, а также фильтр времени и пользовательского местоположения. Разработанная модель позволяет контролировать влияние скрытых состояний между рекуррентными ячейками, поддерживает как непрерывный, так и дискретный вход и может рассматривать разнообразные виды контекста по-разному, таким образом, обходя ряд ограничений, с которыми сталкивались исследователи в более ранних моделях.

В [13] авторы предлагают основанную на глубоком обучении модель для контекстных рекомендаций. Модель состоит из архитектуры AE, дополненной контекстно-управляемым механизмом внимания, называемым ACDA. Механизм внимания используется для кодирования контекстных атрибутов в скрытое представление предпочтений пользователя, которое связывает персонализированный контекст с предпочтениями каждого пользователя, чтобы предоставить рекомендации, ориентированные на этого конкретного потребителя.

В [14] исследователи рассматривают использование моделей гетерогенной информационной сети и графовых нейронных архитектур для задач рекомендаций. Авторы предлагают новый метод выстраивания гетерогенной информационной сети, SgWalk, который исследует близость между пользователями и местоположениями, а также генерирует рекомендации по местоположению через выстраивание узлов на основе подграфов. SgWalk выполняет четыре шага: построение пользовательских подграфов в соответствии с контекстом местоположения, генерация случайных последовательностей обхода пользовательских подграфов, изучение встраивания узлов в граф, а также генерация рекомендаций по местоположению с использованием векторного представления узлов. SgWalk отличается от существующих методов, основанных на метапутях или двудольных графах, за счет использования контекстного пользовательского подграфа.

В [15] авторы концентрируются на проблеме разреженности данных при прогнозировании посещения точек интереса. Исследователи предлагают модель на основе глубокого обучения под названием CatDM, которая включает категорию точки интереса и географическое влияние объектов для уменьшения пространства поиска и преодоления разреженности данных. Авторы разрабатывают два глубоких кодировщика: первый кодировщик фиксирует пользовательские предпочтения в категориях точек интереса, тогда как второй использует пользовательские предпочтения в самих точках интереса. Учитывая влияние времени во втором кодировщике, авторы разделяют пользовательскую историю посещений точек интереса на несколько разных

Таблица 2. Контекстно-зависимые рекомендательные системы на основе нейронных сетей

Название модели	Модель	Целевая переменная	Метрики
ST-CLSTM [10]	LSTM	Вероятность посещения места	Accuracy@k, MAP
SLCM [11]	LSTM	Рейтинг объекта	RMSE, MAR, HR@k
CARA [12]	GRU + механизм внимания	Вероятность выбора объекта	HR, NDCG
ACDA [13]	AE + механизм внимания	Вероятность выбора объекта	Precision@k, MAP, Recall@k, NDCG
SgWalk [14]	GNN	Рейтинг объекта	Precision@k, Recall@k, FMeasure@k
CatDM [15]	LSTM + механизм внимания	Рейтинг объекта	Precision@k, Recall@k,
HGMAP [16]	CNN + механизм внимания	Рейтинг объекта	Precision@k, Recall@k,
LSA [17]	LSTM + механизм внимания	Рейтинг объекта	Precision

временных окон и получают итоговый результат с помощью персонализированного механизма внимания.

В [16] исследователи предлагают новую структуру рекомендаций HGMAP, которая представляет собой комбинацию графовых сверточных нейронных сетей и механизма внимания. HGMAP строит пространственный граф на основе географического расстояния между парами точек интереса и использует графовые сверточные сети для поисков закономерностей разных точек интереса. Кроме того, HGMAP использует социальные отношения пользователя с другими пользователями и различает предпочтения пользователя по различным аспектам точек интереса с помощью механизма внимания.

В [17] авторы предлагают модель LSA, которая является комбинацией LSTM и механизма внимания. На первом этапе используется построение векторных представлений для количественной оценки информации о пользователе, его дружеских отношениях, информации о точке интереса. Далее на основе истории пользователя и текущей последовательности регистрации посещений точек интереса строится механизм долгосрочного и краткосрочного внимания, где количественная информация используется в качестве входных данных LSA для лучшего изучения долгосрочных и краткосрочных интересов пользователя. На последнем этапе информация из социальной сети и семантическая информация помещаются в разные входные слои, а информация о времени и географическом местоположении исторического поведения пользователя используется для рекомендации ему следующей точки интереса.

Итоговая таблица с описанием особенностей контекстно-зависимых рекомендательных систем на основе нейронных сетей представлена в таблице 2.

Среди главных достоинств рассматриваемого класса моделей можно выделить высокое качество работы алгоритмов и отсутствие необходимости в явном виде выделения контекстных признаков для моделей. К недостаткам можно отнести долгое обучение, а также трудности с интерпретацией результатов.

Заключение

В работе представлен обзор существующих контекстно-зависимых рекомендательных систем для рекомендации точек интереса и классификация методов, которые лежат в их основе. Все модели были разделены на два класса — модели, которые используют матричную факторизацию и модели, использующие нейронные сети. Изначально, в качестве контекстно-зависимых рекомендательных систем для рекомендации точек интереса использовались модели на основе матричной факторизации. Главным достоинством этих моделей является возможность интерпретации полученных результатов. Среди недостатков этих моделей можно выделить низкое качество работы этих алгоритмов и необходимость самостоятельно определять набор данных, который используется в качестве пользовательского контекста. После массового распространения нейронных сетей их стали использовать и в решении задач контекстно-зависимых рекомендательных систем для рекомендации точек интереса. Основными достоинствами этих моделей является высокое качество работы по сравнению с моделями, использующими матричную факторизацию, а также отсутствие необходимости в явном виде выделения контекстных признаков для моделей. К недостаткам данных моделей можно отнести долгое обучение и трудности с интерпретацией результатов. Так, каждый из рассмотренных классов моделей имеет свои достоинства и недостатки. Другими словами, лучше использовать модели в зависимости от рассматриваемой задачи.

ЛИТЕРАТУРА

1. Adomavicius G., Tuzhilin A. Context-aware recommender systems //Recommender systems handbook. — Springer, Boston, MA, 2011. — С. 217–253.
2. Seyednezhad S.M. et al. A review on recommendation systems: Context-aware to social-based //arXiv preprint arXiv:1811.11866. — 2018.
3. Baltrunas L. et al. Context relevance assessment and exploitation in mobile recommender systems //Personal and Ubiquitous Computing. — 2011. — С. 1–20.
4. Baltrunas L., Ludwig B., Ricci F. Matrix factorization techniques for context aware recommendation //Proceedings of the fifth ACM conference on Recommender systems. — 2011. — С. 301–304.
5. Unger M. et al. Towards latent context-aware recommendation systems //Knowledge-Based Systems. — 2016. — Т. 104. — С. 165–178.
6. Cheng C. et al. Fused matrix factorization with geographical and social influence in location-based social networks //Aai. — 2012. — Т. 12. — С. 17–23.
7. Lian D. et al. GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation //Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. — 2014. — С. 831–840.
8. Lian D. et al. GeoMF++ Scalable Location Recommendation via Joint Geographical Modeling and Matrix Factorization //ACM Transactions on Information Systems (TOIS) . — 2018. — Т. 36. — № 3. — С. 1–29.
9. Zhao S., Lyu M.R., King I. STELLAR: Spatial-Temporal Latent Ranking Model for Successive POI Recommendation //Point-of-Interest Recommendation in Location-Based Social Networks. — Springer, Singapore, 2018. — С. 79–94.
10. Zhao P. et al. Where to go next: A spatio-temporal lstm model for next poi recommendation //arXiv preprint arXiv:1806.06671. — 2018.
11. Livne A. et al. Deep Context-Aware Recommender System Utilizing Sequential Latent Context //arXiv preprint arXiv:1909.03999. — 2019.
12. Manotumruksa J., Macdonald C., Ounis I. A contextual attention recurrent architecture for context-aware venue recommendation //The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. — 2018. — С. 555–564.
13. Jhamb Y., Ebesu T., Fang Y. Attentive contextual denoising autoencoder for recommendation //Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval. — 2018. — С. 27–34.
14. Canturk D., Karagoz P. SgWalk: Location Recommendation by User Subgraph-Based Graph Embedding //IEEE Access. — 2021. — Т. 9. — С. 134858–134873.
15. Yu F. et al. A category-aware deep model for successive POI recommendation on sparse check-in data //Proceedings of the web conference 2020. — 2020. — С. 1264–1274.
16. Zhong T. et al. Hybrid graph convolutional networks with multi-head attention for location recommendation //World Wide Web. — 2020. — Т. 23. — С. 3125–3151.
17. Wang K., Wang X., Lu X. POI recommendation method using LSTM-attention in LBSN considering privacy protection //Complex & Intelligent Systems. — 2021. — С. 1–12.

© Жданов Павел Сергеевич (pszhdanov@itmo.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»