



Построение эмбедингов пользователей с помощью нейросети

стажёр Маша Жарова

ментор Женя Михайлова

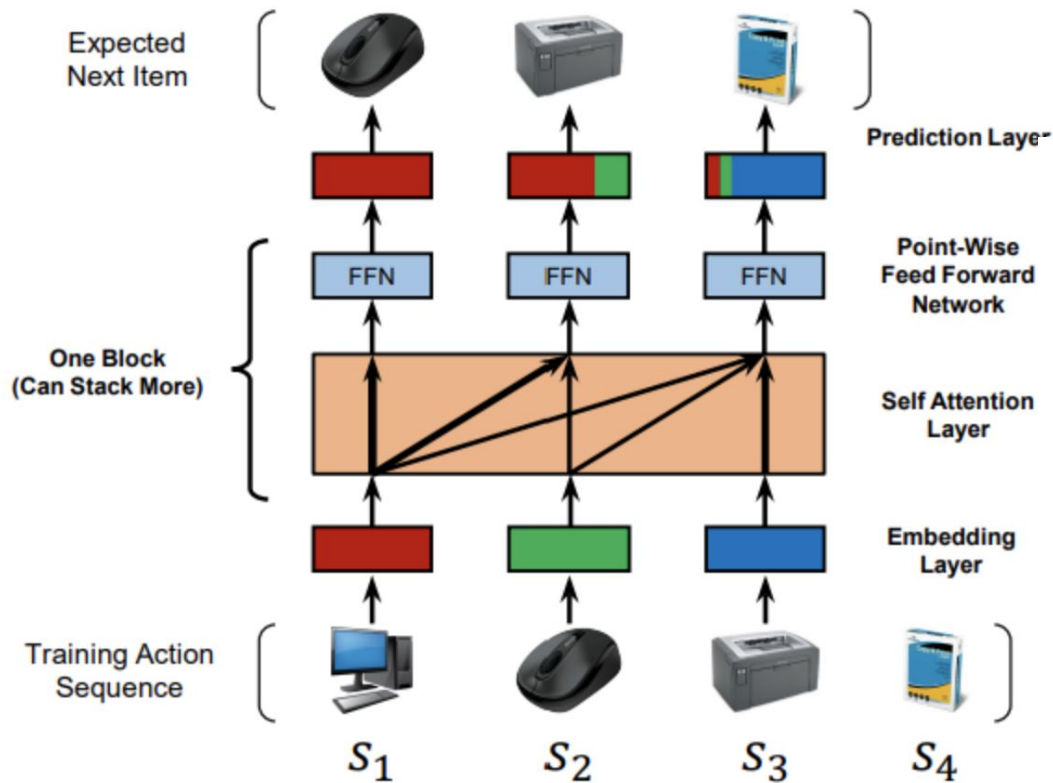
Лето 2022



План:

- Модели-трансформеры
- Сравнение трансформера SASRec с LightFM
- Что не получилось: (
- Что получилось:)
- Где можно использовать

SASRec - модель-трансформер





Сравнение с LightFM



- LightFM (параметры с прода):

hit rate 0.047

(30 эпох, time ~10 мин)

- SASRec:

hit rate 0.140

(60 эпох, time ~30 мин)

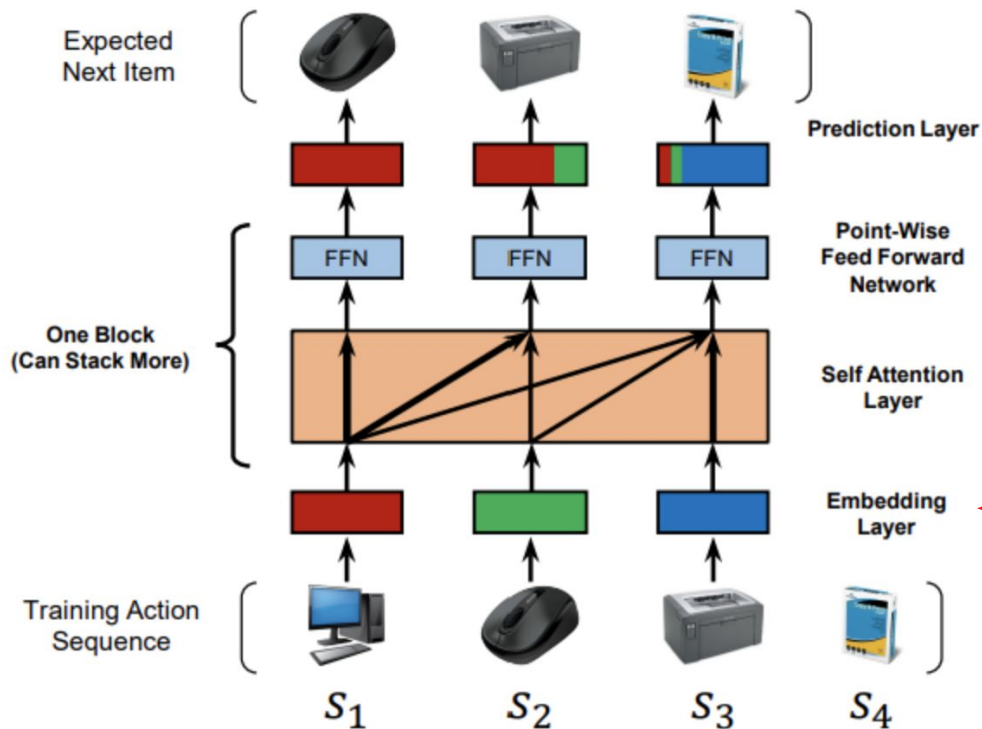
Данные взяты за неделю



1s



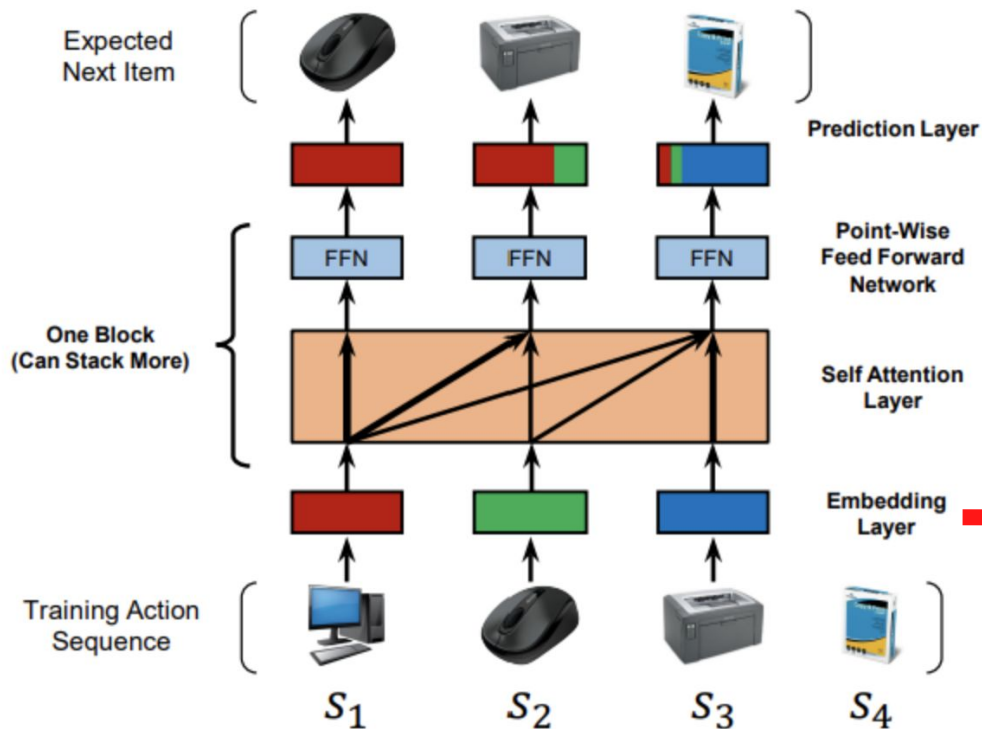
Подстановка эмбедингов из lst



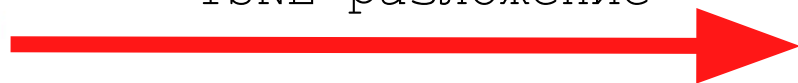
Замена trainable-
люя эмбедингов на
nontrainable,
собранных из lst-
фичей



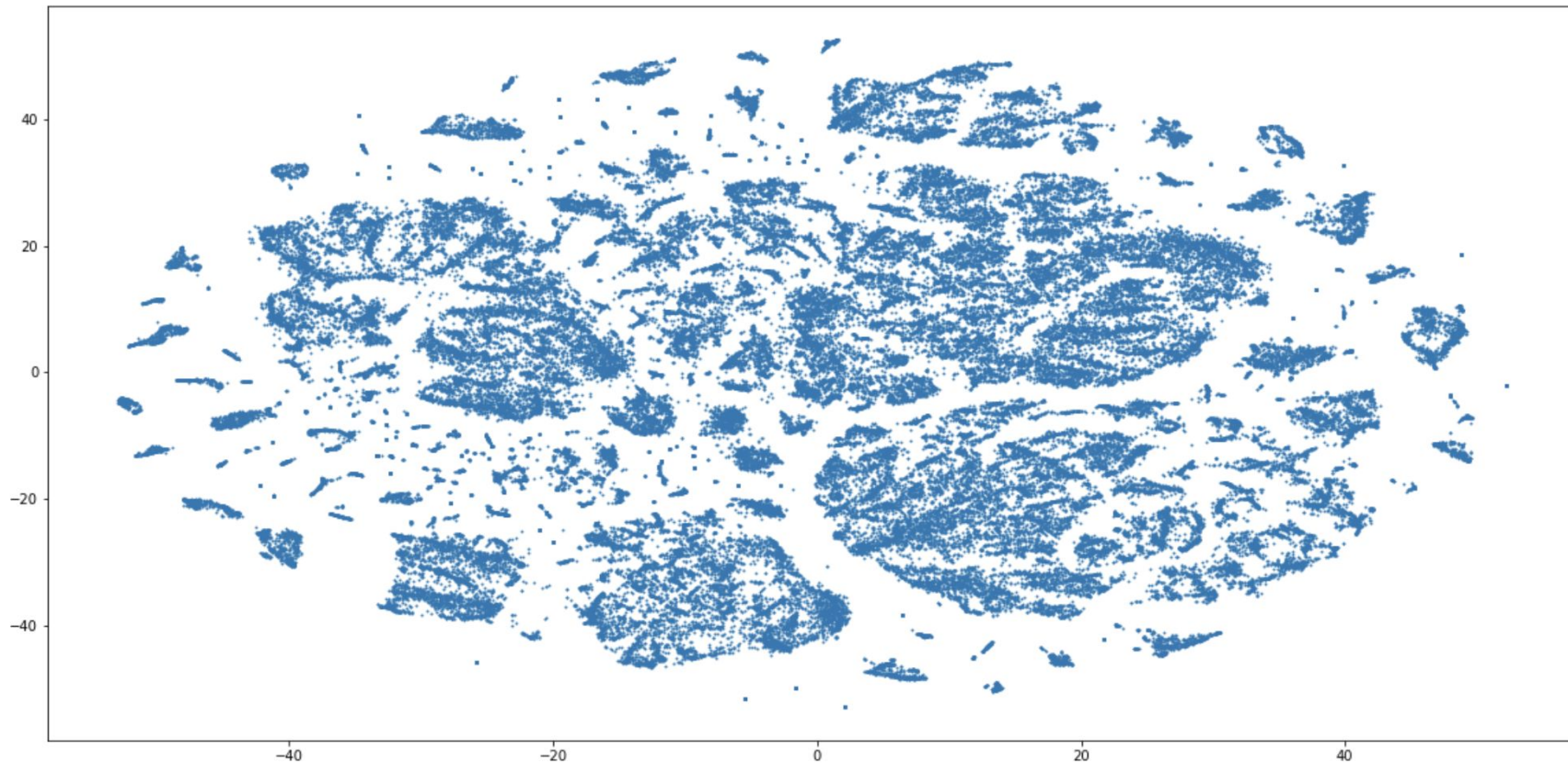
Извлечение эмбедингов объектов



В стандартной модели достали эмбединги объектов и построили по ним TSNE-разложение



TSNE-разложение эмбедингов объектов



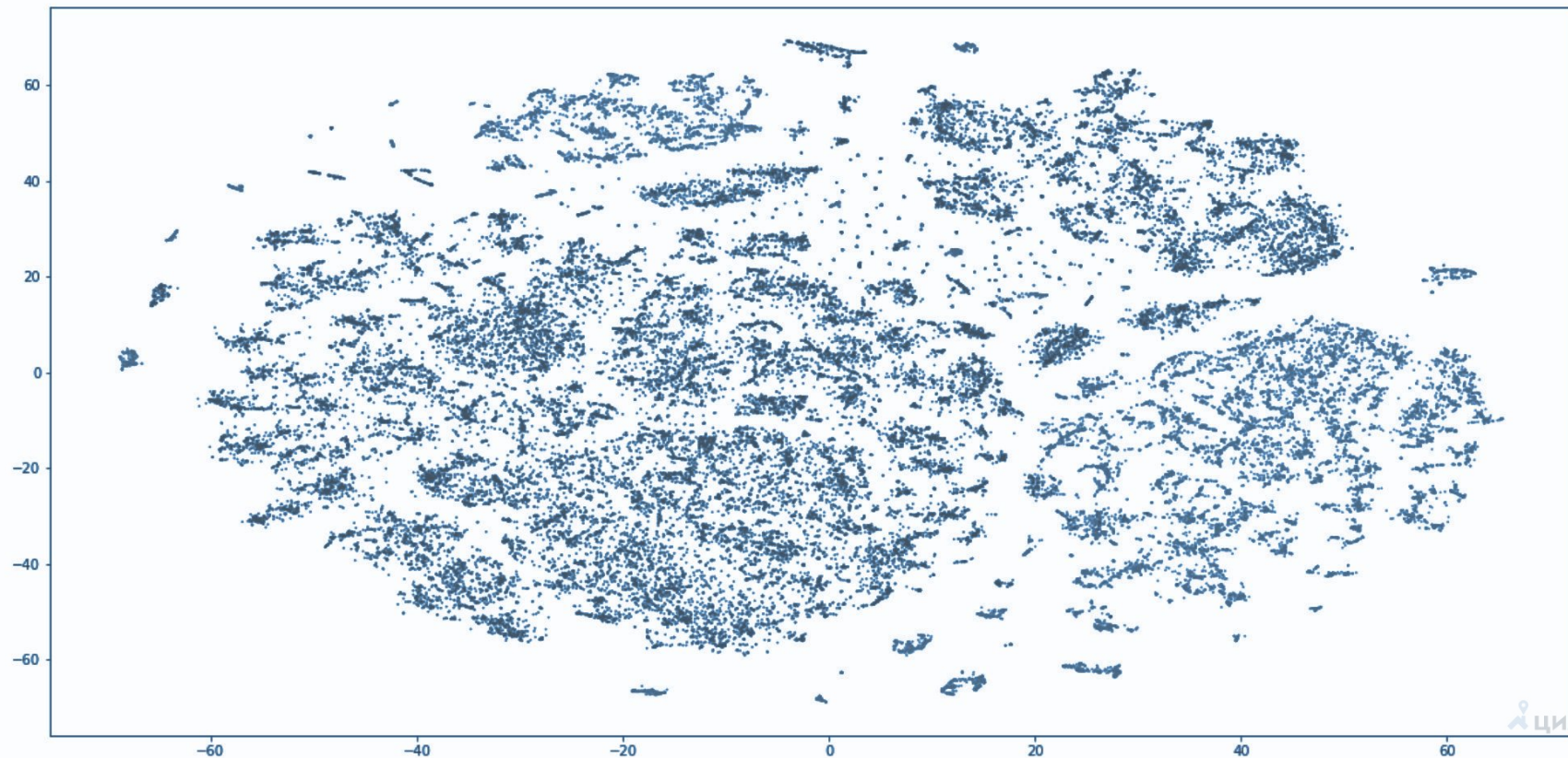


Эмбединги пользователей

Каждый пользователь характеризуется последовательностью объявлений, на которые он ткнул.

Таким образом, можем узнать группу его предпочтений.

TSNE-разложение эмбедингов пользователей



Дополнение

Собрали датасет:

фичи – эмбединги пользователей

таргет – предпочтения по аренде/продаже

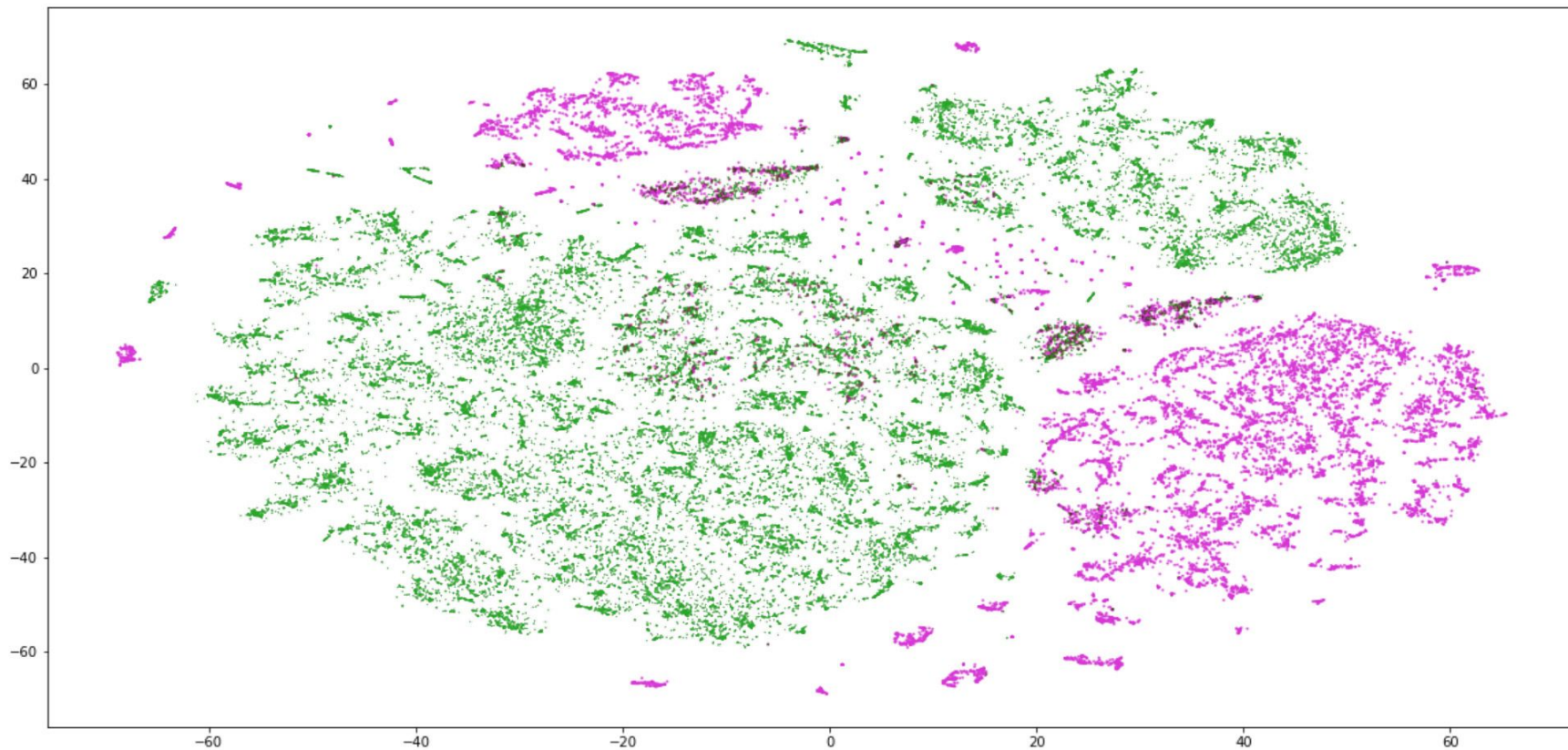
Обучили на нём классификатор CatBoost:

precision 0.97

recall 0.96

roc-auc 0.99

Разделение пользователей по классам



Итоги

- Обучили SASRec и получили высокие метрики
- Достали из неё эмбединги пользователей
- По схеме с обучением классификатора можем предсказывать для новых пользователей их предпочтения на основе их эмбедингов

не подвожу итоги года, потому что планирую в последнюю ночь все доделать



Спасибо за внимание!