

# Обучение взаимосвязанных информативных представлений в задаче генерации образов

Охотников Никита Владимирович

Научный руководитель: к.ф.-м.н. Исаченко Р.В.

Кафедра интеллектуальных систем ФПМИ МФТИ

Специализация Интеллектуальный анализ данных

Направление: 03.03.01 Прикладные математика и физика

2024

# Введение

Исследуется задача поиска наилучшего дополнения образа — множества взаимосвязанных элементов — элементами конечной коллекции.

## Проблемы

- ▶ Взаимосвязь элементов в образе имеет неизвестную структуру.
- ▶ Точное решение задачи дополнения требует полного перебора.

## Задача

Предложить эффективный приближенный алгоритм дополнения образа несколькими элементами.

## Предлагается

На основе известной функции оценки образа построить функцию для генерации зависимых скрытых представлений элементов, использующихся далее для выбора элементов дополнения

# Постановка задачи

## Основные понятия и обозначения

- ▶ Основная единица данных – *объект* или *элемент*, множество всех объектов –  $\mathcal{X}$
- ▶ Каждый объект  $X \in \mathcal{X}$  есть пара  $X = (I, T)$  из соответственно изображения и текстового описания.
- ▶ Далее под объектом  $X \in \mathcal{X}$  будем понимать его векторное представление  $X \in \mathbb{R}^d$  в общем для элементов пространстве
- ▶ Непустые подмножества множества элементов  $O = \{X_i\}_{i=1}^k \subset \mathcal{X}$ ,  $O \neq \{\emptyset\}$  будем называть *образами*. Множество образов обозначим  $\mathcal{O}$ .
- ▶ Для оценки образов введем функцию *оценки* или *совместимости* его элементов:

$$S : 2^{\mathcal{X}} \longrightarrow [0, 1]$$

$$\forall O \in \mathcal{O} : S(O) > 0$$

*Совместимостью* или *оценкой* образа  $O$  будем называть  $S(O)$

# Постановка задачи

## Задача дополнения образа

### Дано:

$O_n \in \mathcal{O}$ ,  $|O| = n$  — исходный образ

$k \in \mathbb{N}$ ,  $k$  — количество элементов дополнения

### Требуется:

Найти наилучшее в смысле максимизации функции оценки  $\mathcal{S}$  дополнение образа  $O_n$   $k$  элементами  $\{\hat{X}_i\}_{i=1}^k \subset \mathcal{X}$  т.е. решить следующую оптимизационную задачу

$$\{\hat{X}_i\}_{i=1}^k = \operatorname{argmax}_{\{X_i\}_{i=1}^k \subset \mathcal{X}} \mathcal{S}(O_n \cup \{X_i\}_{i=1}^k)$$

Точное решение для известной  $\mathcal{S}$ : полный перебор всех подмножеств  $\mathcal{X}$  размера  $k$ .

Асимптотика:  $|\mathcal{X}|^k$  вызовов функции  $\mathcal{S}$

# Теоретическая часть

- ▶ В качестве аппроксимации функции оценки  $\mathcal{S}$  далее будем рассматривать предобученную модель OutfitTransformer<sup>1</sup>.
- ▶ Для задачи дополнения

$$\{\hat{X}_i\}_{i=1}^k = \operatorname{argmax}_{\{X_i\}_{i=1}^k \subset \mathcal{X}} \mathcal{S}(O_n \cup \{X_i\}_{i=1}^k)$$

существует 2 глобальных подхода

- ▶ Дискретный – оптимизация полного перебора
- ▶ Непрерывный – решение некоторой связанной задачи в непрерывном множестве и выбор ближайших к решению элементов  $\mathcal{X}$

---

<sup>1</sup><https://doi.org/10.48550/arXiv.2204.04812>

# Дополнение образа

## Дискретный подход

- ▶ Решение задачи приближенным перебором
- ▶ Бейзлайн: жадные алгоритмы

«1-step»  $X_1 = \operatorname{argmax}_{X \in \mathcal{X}} \mathcal{S}(O_n \cup X), \dots, X_k = \operatorname{argmax}_{X \in \mathcal{X} \setminus \bigcup_{i=1}^{k-1} X_i} \mathcal{S}(O_n \cup X)$

Асимптотика:  $|\mathcal{X}|$  вызовов функции  $\mathcal{S}$

«k-step»  $X_1 = \operatorname{argmax}_{X \in \mathcal{X}} \mathcal{S}(O_n \cup X), \dots, X_k = \operatorname{argmax}_{X \in \mathcal{X} \setminus \bigcup_{i=1}^{k-1} X_i} \mathcal{S}(O_n \cup X_1 \dots X_{k-1} \cup X)$

Асимптотика:  $k \cdot |\mathcal{X}|$  вызовов функции  $\mathcal{S}$

- ▶ Альтернатива: алгоритм beam-search. В граничных случаях вырождается либо в полный перебор, либо в k-step алгоритм выше.  
Асимптотика:  $\geq k \cdot |\mathcal{X}|$  вызовов функции  $\mathcal{S}$
- ▶ Очень неэффективно

# Дополнение образа

## Непрерывный подход (градиентный спуск)

- ▶ Функция  $\mathcal{S}$  липшицева с некоторой константой  $M$
- ▶ Есть доступ не только к  $\mathcal{S}$ , но и к  $\nabla \mathcal{S}$
- ▶ Идея: перейдем к релаксированной задаче:

$$\{\tilde{X}_i\}_{i=1}^k = \operatorname{argmax}_{\{X_i\}_{i=1}^k \subset \mathbb{R}^d} \mathcal{S}(O_n \cup \{X_i\}_{i=1}^k)$$

- ▶ Далее выберем  $\{\hat{X}_i\} \subset \mathcal{X}$  как ближайшие к решениям в смысле функции близости  $\rho$ :

$$\hat{X}_i = \operatorname{argmin}_{X \in \mathcal{X}} \rho(\tilde{X}_i, X)$$

- ▶ Полученная задача разрешима с помощью градиентного спуска.
- ▶ Асимптотика:  $n$  вызовов  $\mathcal{S}$  и  $\nabla \mathcal{S}$ , где  $n$  – количество шагов градиентного спуска (не зависит от  $|\mathcal{X}|$ )

# Дополнение образа

## Непрерывный подход (градиентный спуск)

- ▶  $S$  –  $M$ -липшицева
- ▶ рассмотрим  $L_\rho$  метрику в качестве  $\rho$ , тогда

$$\sum_{i=1}^k \rho(\hat{X}_i, \tilde{X}_i) < \varepsilon \longrightarrow \left| S(O_n \cup \{\tilde{X}_i\}_{i=1}^k) - S(O_n \cup \{\hat{X}_i\}_{i=1}^k) \right| < M \cdot \varepsilon$$

- ▶ Проблема подхода:  $\exists \{\hat{X}_i\} \subset \mathcal{X} : \sum_{i=1}^k \rho(\hat{X}_i, \tilde{X}_i) < \varepsilon$  — очень сильное условие и требует по крайней мере

$$\exists \{\hat{X}_i\}_{i=1}^k \subset \mathcal{X} : S(O_n \cup \{\hat{X}_i\}_{i=1}^k) \geq \max_{\{X_i\}_{i=1}^k \subset \mathbb{R}^d} S(O_n \cup \{X_i\}_{i=1}^k) - M\varepsilon$$



$$\max_{\{X_i\}_{i=1}^k \subset \mathcal{X}} S(O_n \cup \{X_i\}_{i=1}^k) \geq \max_{\{X_i\}_{i=1}^k \subset \mathbb{R}^d} S(O_n \cup \{X_i\}_{i=1}^k) - M\varepsilon$$



# Дополнение образа

## Непрерывный подход (генерация скрытых представлений)

- ▶ Предлагается *полностью* отказаться от вызовов функции  $\mathcal{S}$
- ▶ Переформулируем задачу как поиск аппроксимации функции

$$\mathcal{F}_k : \mathcal{O} \longrightarrow \mathcal{X}^k, \quad O_n \in \mathcal{O}, \quad \mathcal{F}_k(O_n) = \operatorname{argmax}_{\{X_i\}_{i=1}^k \subset \mathcal{X}} \mathcal{S}(O_n \cup \{X_i\}_{i=1}^k)$$

Композицией функций

$$F_k^\theta : \mathcal{O} \longrightarrow \mathbb{R}^d, \quad F_k^\theta(O_n) = \{\tilde{X}_i\}_{i=1}^k$$

$$\text{и } \rho_{\mathcal{X}} : \mathbb{R}^d \longrightarrow \mathcal{X}, \quad \rho_{\mathcal{X}}(\tilde{X}_i) = \operatorname{argmax}_{\hat{X}_i \in \mathcal{X}} \rho(\tilde{X}_i, \hat{X}_i)$$

# Дополнение образа

## Непрерывный подход (генерация скрытых представлений)

- Свели задачу к генерации скрытых представлений недостающих элементов  $\{\tilde{X}_i\} \subset \mathbb{R}^d$ , наиболее близких в смысле функции  $\rho$  к точным решениям

$$\{\hat{X}_i\}_{i=1}^k = \operatorname{argmax}_{\{X_i\}_{i=1}^k \subset \mathcal{X}} \mathcal{S}(O_n \cup \{X_i\}_{i=1}^k)$$

с помощью функции  $F_k^\theta$  с вектором параметров  $\theta$ .

- Рассмотрим образы  $O_n = \{O^i\}_{i=1}^n \subset \mathcal{O}$  и множество известных точных решений  $\mathcal{X}_n = \{\{\hat{X}_j^i\}_{j=1}^k\}_{i=1}^n \subset \mathcal{X}^k$
- Тогда на параметры  $\theta$  получаем следующую задачу:

$$\theta = \operatorname{argmin}_{\hat{\theta}} \left( \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k \rho(X_j^i, [F_k^{\hat{\theta}}(O^i)]_j) \right)$$

# Дополнение образа

## Непрерывный подход (генерация скрытых представлений)

- ▶ Задача симметрична к перестановке  $\implies$  разумно рассматривать операции эквивариантные относительно группы перестановок.
- ▶ Тогда представим функцию  $F_k^\theta$  с помощью графовой нейронной сети (GNN)
- ▶ Вершины графа — представления элементов образа
- ▶ Общий вид преобразования  $h_i^{(t)}$  скрытого состояния  $i$ -ой вершины на шаге  $t$  в message passing GNN<sup>2</sup>:

$$h_i^{(t)} = U^{(t)} \left( h_i^{(t-1)}, \bigoplus_{j \in \overline{1, n}} M^{(t)} \left( h_i^{(t-1)}, h_j^{(t-1)} \right) \right),$$

где  $U^{(t)}, M^{(t)}$  — дифференцируемые функции,  $\bigoplus$  — дифференцируемая агрегирующая функция, инвариантная к перестановкам (в эксперименте будем использовать сумму)

---

<sup>2</sup><https://arxiv.org/pdf/1704.01212>

# Дополнение образа

## Непрерывный подход (генерация скрытых представлений)

- ▶ Аппроксимация напрямую решений дискретной, а не релаксированной задачи
- ▶ Асимптотика: один вызов функции  $F_k^\theta$
- ▶ Произвольное количество элементов дополнения за одну итерацию
- ▶ GNN моделирует зависимости между элементами дополнения
- ▶ В качестве  $\mathcal{X}_n$  можно рассмотреть набор решений задачи многошаговым жадным алгоритмом

# Вычислительный эксперимент

## Условия эксперимента

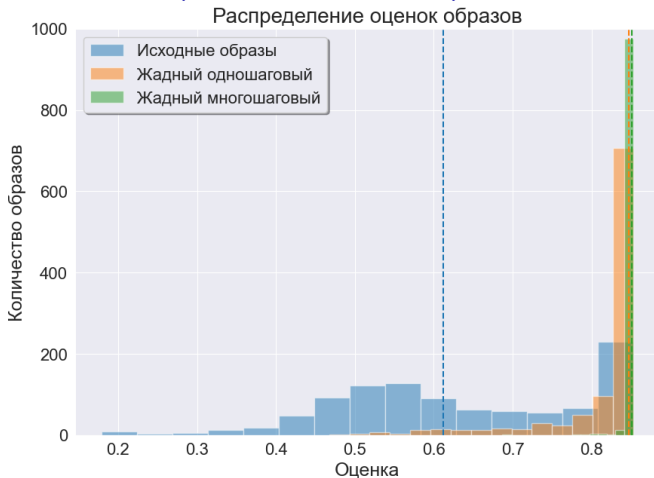
- ▶ Данные: датасет Polyvore<sup>3</sup> — 17000 образов из 65000 объектов
- ▶ Случайно выберем 1000 образов
- ▶ Зафиксируем количество элементов дополнения  $k = 2$
- ▶ Оцениваем алгоритмы на основании распределения оценок дополненных образов
- ▶ Бейзлайн: распределение оценок исходных образов
- ▶ В качестве  $\rho$  рассмотрим метрики  $L_1$ ,  $L_2$  и  $L_{10}$  и косинусное расстояние.

---

<sup>3</sup><http://arxiv.org/abs/1707.05691>

# Вычислительный эксперимент

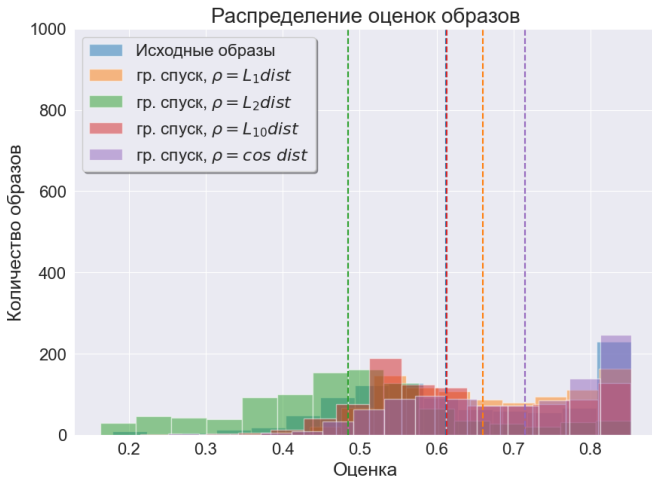
## Дискретный подход (жадные алгоритмы)



Показывают хороший результат, но требуют слишком много времени

# Вычислительный эксперимент

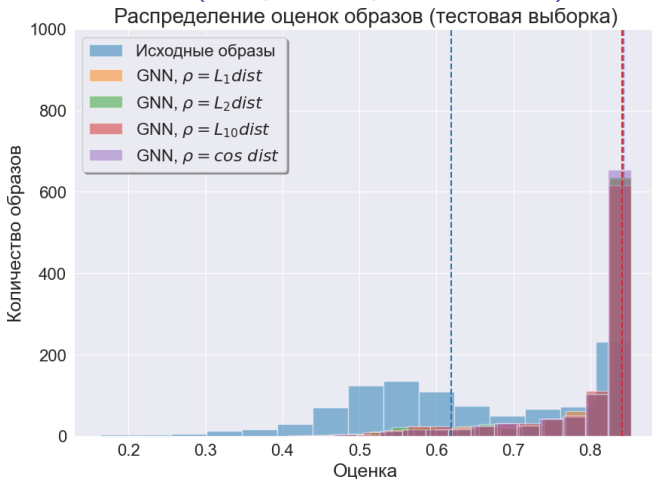
## Непрерывный подход (градиентный спуск)



Результат заметно хуже чем для жадных алгоритмов, а время вычислений еще медленнее, чем ожидалось

# Вычислительный эксперимент

## Непрерывный подход (генераций представлений)



Хорошее качество, достижимое за десятки миллисекунд. Интересно, что разницы между различными  $\rho$  почти нет



# Вычислительный эксперимент

## Сравнение методов

$\rho$	Выборочная медиана оценок образов			
	ж. однош.	ж. многош.	гр. спуск	GNN
$L_1$	0.8511	0.8467	0.6602	0.8417
$L_2$			0.4850	0.8417
$L_{10}$			0.6132	0.8415
$\cos dist$			0.7142	0.8428
Задержка, с	$\sim 4$	$\sim 8$	$\sim 5$	$\sim 0.03$

## Выводы

- ▶ Жадные алгоритмы показывают хороший результат, но не применимы на практике.
- ▶ Подход с решением релаксированной задачи себя не оправдал — для данных не выполнено теоретическое предположение.
- ▶ Генерация представлений существенно быстрее и почти не уступает в качестве.

## Выносятся на защиту

- ▶ Предложен и обоснован эффективный алгоритм дополнения образа произвольным числом взаимосвязанных элементов
- ▶ Предложен способ пополнения обучающих данных для модели в условиях недостатка образов с высокой оценкой
- ▶ Реализован программный код для вычислительного эксперимента и проведена оценка предложенных подходов