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 (на примере элементов одежды) — элементами конечной коллекции.

Проблемы

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

Задача

Предложить применимый на практике 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* (одежды), 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:

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

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

Совместимостью или *оценкой* образа O будем называть $\mathcal{S}(O)$

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

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

- 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[1].
- Для задачи дополнения

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

существует 2 глобальных подхода

  - Дискретный – оптимизация полного перебора
  - Непрерывный – решение релаксированной задачи в $\mathbb{R}^d$ и поиск ближайших к решению элементов $\mathcal{X}$

[1] https://doi.org/10.48550/arXiv.2204.04812

Дополнение образа

Дискретный подход

- Решение задачи приближенным перебором
- Бейзлайн: жадные алгоритмы

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

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

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

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

- Альтернатива: алгоритм beam-search, активно применяемый в языковых моделях. В граничных случаях вырождается либо в полный перебор, либо в k-step алгоритм выше.
  Асимптотика: $\geqslant k \cdot |\mathcal{X}|$ вызовов функции $\mathcal{S}$

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:

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

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

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

- Полученная задача разрешима с помощью стохастического 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)

- $\mathcal{S}$ – M-липшицева
- рассмотрим L_p метрику в качестве ρ, тогда

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

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

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

$$\Updownarrow$$

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

Дополнение образа

Непрерывный подход (генерация скрытых представлений)

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

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

Композицией функций

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

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

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 ρ к точным решениям

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

  с помощью функции F_k^θ с вектором параметров θ.
- Рассмотрим образы $\mathcal{O}_n = \{O^i\}_{i=1}^n \subset \mathcal{O}$ и множество известных точных решений задачи дополнения для них $\mathcal{X}_n = \{\{\hat{X}_j^i\}_{j=1}^k\}_{i=1}^n \subset \mathcal{X}^k$
- Тогда на параметры θ получаем следующую оптимизационную задачу:

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

Дополнение образа

Непрерывный подход (генерация скрытых представлений)

- Задача симметрична к перестановке $\Longrightarrow$ разумно рассматривать операции эквивариантные относительно группы перестановок.
- Тогда представим функцию F_k^θ 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[2]:

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

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

[2] 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^θ
- 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

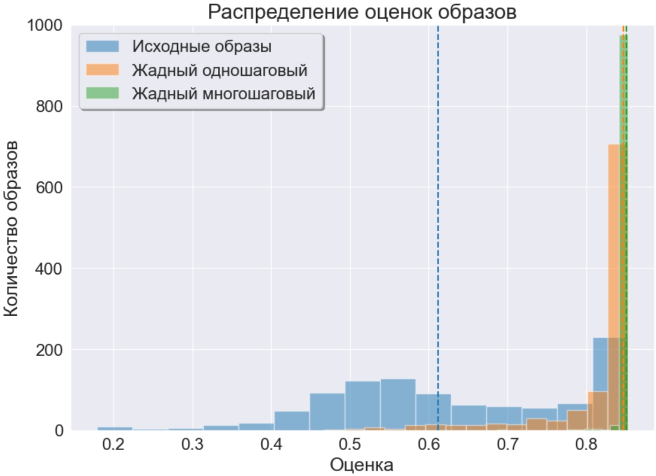
Условия эксперимента

- Данные: датасет Polyvore[3] — 17000 образов из 65000 объектов
- Случайно выберем 1000 образов
- Зафиксируем количество элементов дополнения $k = 2$
- Оцениваем алгоритмы на основании распределения оценок дополненных образов
- Бейзлайн: рапределение оценок исходных образов
- В качестве ρ рассмотрим метрики L_1, L_2 и L_{10} и косинусное расстояние.

[3] http://arxiv.org/abs/1707.05691

Вычислительный эксперимент

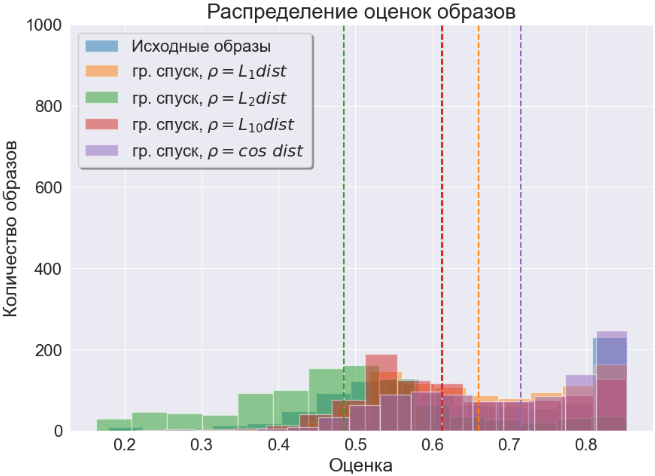
Дискретный подход (жадные алгоритмы)



Показывают хороший результат, но требуют слишком много времени

Вычислительный эксперимент

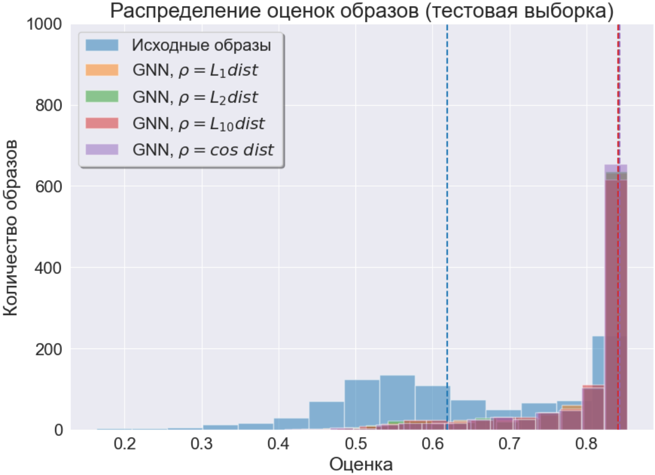
Непрерывный подход (градиентный спуск)



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 ρ почти нет

Вычислительный эксперимент

Сравнение методов

ρ	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

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.

Выносится на защиту

- Предложен 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