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 дополнения образа

- Основная единица данных – *объект* или *элемент*, множество векторных представлений всех объектов – $\mathcal{X} \subset \mathbb{R}^d$.
- 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 образов введем функцию *оценки* или *совместимости* элементов:

$$\mathcal{S} : \ 2^{\mathcal{X}} \longrightarrow [0, 1], \quad \forall O \in \mathcal{O} : \ \mathcal{S}(O) > 0.$$

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

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

Теоретическая часть

- Примем функцию оценки $\mathcal{S}$ известной. (В эксперименте 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

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

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

Дискретный подход

- Решение задачи приближенным перебором
- Бейзлайн: жадные алгоритмы

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

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

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

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: $\geqslant k \cdot |\mathcal{X}|$ вызовов функции $\mathcal{S}$
- Крайне неэффективно

Непрерывный подход

Будем рассматривать 2 непрерывных алгоритма:

1. Релаксация и градиентный спуск
2. Генерация скрытых представлений

Градиентный спуск

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

Градиентный спуск, необходимое условие

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

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

$$\Updownarrow$$

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

- Назовем последнее «условием плотности» множества элементов $\mathcal{X}$

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

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

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

- Свели задачу к генерации представлений недостающих элементов $\{\tilde{X}_i\} \subset \mathbb{R}^d$, наиболее близких в смысле функции ρ к $\{\hat{X}_i\}_{i=1}^k$:

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

  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 $\mathcal{X}_n = \{\{\hat{X}_j^i\}_{j=1}^k\}_{i=1}^n \subset \mathcal{X}^k$
- Тогда на параметры θ получаем следующую задачу:

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

Аппроксимация функции F_θ^k

- 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

Генерация скрытых представлений, итоги

1. Аппроксимация напрямую решений дискретной, а не релаксированной задачи
2. Асимптотика: один вызов функции, аппроксимирующей F_k^θ
3. Произвольное количество элементов дополнения за одну итерацию
4. Моделирование зависимостей между элементами дополнения с помощью GNN
5. В качестве $\mathcal{X}_n$ можно рассмотреть набор решений задачи многошаговым жадным алгоритмом

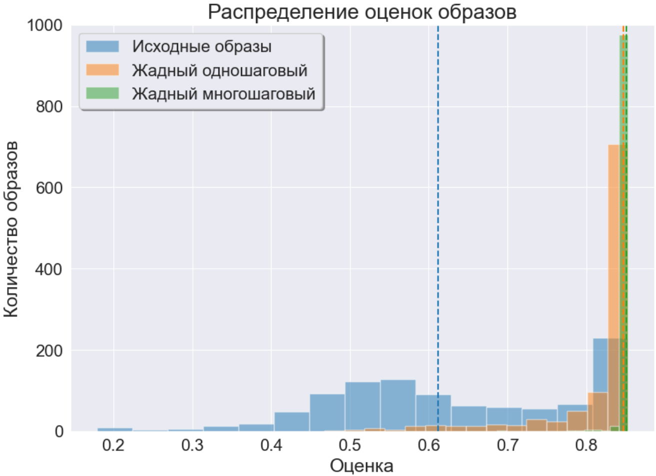
Условия вычислительного эксперимента

- 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

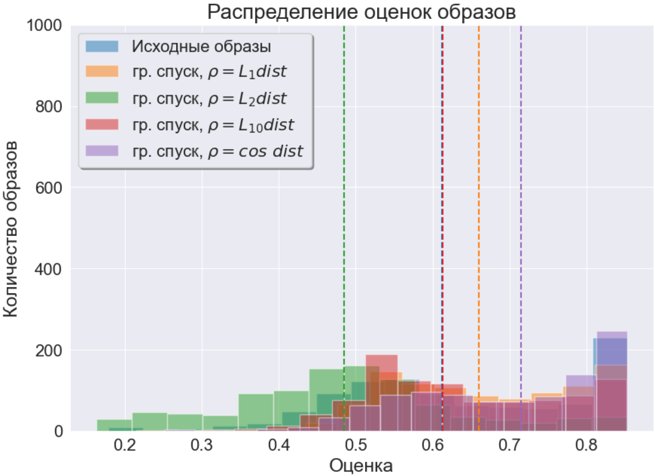
Дискретный подход (жадные алгоритмы)



Показывают хороший результат, но требуют слишком много времени

Вычислительный эксперимент

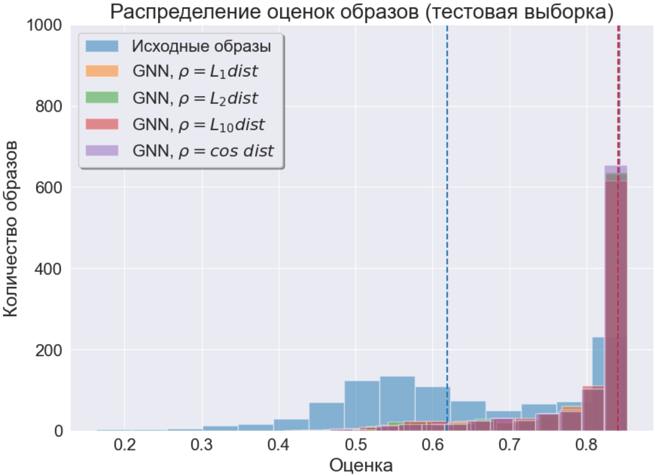
Непрерывный подход (градиентный спуск)



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

Вычислительный эксперимент

Непрерывный подход (генераций представлений)



Качество гораздо выше градиентного спуска, достижимое за десятки миллисекунд. Интересно, что разницы между различными ρ почти нет.

Сравнение рассмотренных 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

Выводы

1. Жадные алгоритмы показывают хороший результат, но не применимы на практике.
2. Подход с решением релаксированной задачи себя не оправдал — для данных не выполнено «условие плотности».
3. Генерация представлений существенно быстрее и почти не уступает в качестве.

Выносится на защиту

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