Обучение взаимосвязанных информативных представлений в задаче генерации образов

Охотников Никита Владимирович

МФТИ

2023-2024

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 на основе близости в латентном пространстве.

Постановка задачи

Основные понятия и обозначения

- Основная единица данных, рассматривающаяся в работе – элемент одежды, далее будем называть его *объектом* или *элементом*, множество всех рассматриваемых объектов – $\mathcal{X}$
- Каждый объект $X \in \mathcal{X}$ есть пара $X = (I, T)$ из соответственно изображения о текстового описания.
- Далее под объектом $X \in \mathcal{X}$ будем понимать его векторное представление $X \in \mathbb{R}^d$ в общем для всех элементов признаковом пространстве.
- Непустые подмножества множества элементов $O = \{X_i\}_{i=1}^{k} \subset \mathcal{X}, O \neq \{\varnothing\}$ 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

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

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

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

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

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 $\mathcal{S}$ непрерывно дифференцируема почти всюду и с ограниченным по норме градиентом, а значит липшицева с некоторой константой M
- Есть доступ не только к значению функции оценки, но и к ее градиенту
- Идея: заменим дискретную задачу непрерывной:

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

- 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 за разумное время с помощью стохастического градиентного спуска.
- Асимптотика n вызовов функции оценки и ее градиента, 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: $\exists\{\hat{X}_i\} \subset \mathcal{X}: \sum_{i=1}^{k} \rho(\hat{X}_i, \tilde{X}_i) < \varepsilon$ — 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

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

Композицией функций

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

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

- $d \gg 1$ поэтому далее, следуя рекомендациям из статьи[2] будем в эксперименте использовать в качестве ρ косинусную близость

[2] https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0144059

Дополнение образа

Непрерывный подход (генерация скрытых представлений)

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

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

  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

Message passing GNN[3]

- 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:

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

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

[3] https://arxiv.org/pdf/1704.01212

Вычислительный эксперимент

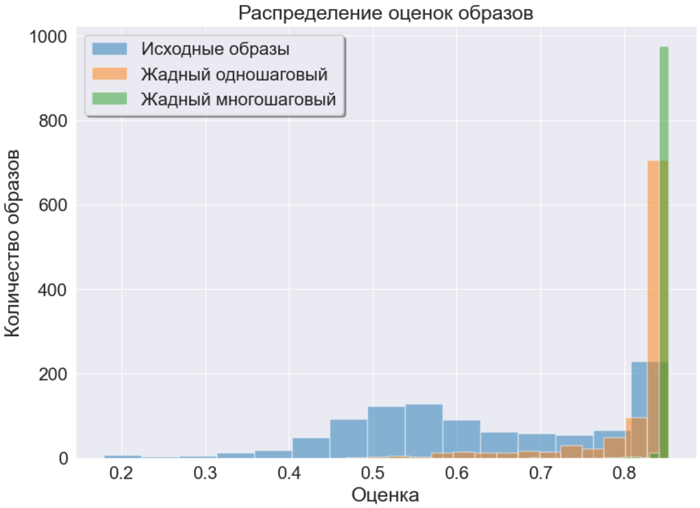
Условия эксперимента

- Данные: датасет Polyvore[4] — 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

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

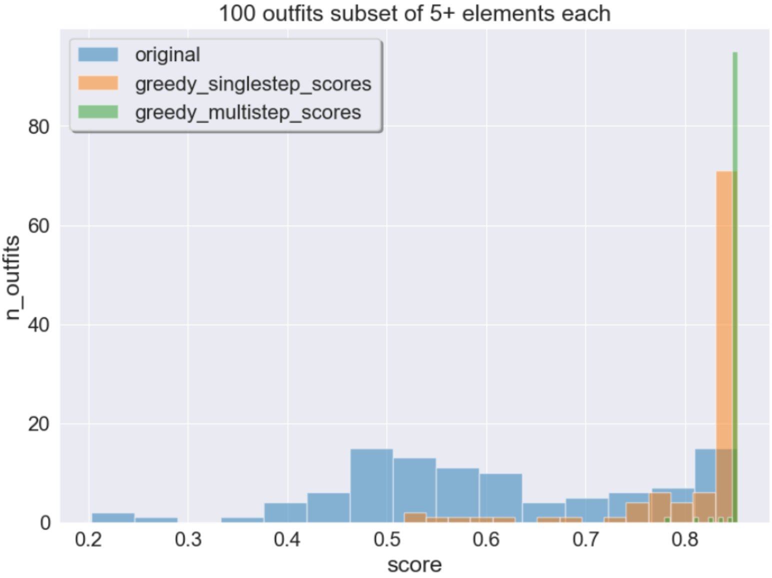
Вычислительный эксперимент

Жадные алгоритмы



Вычислительный эксперимент

Жадные алгоритмы



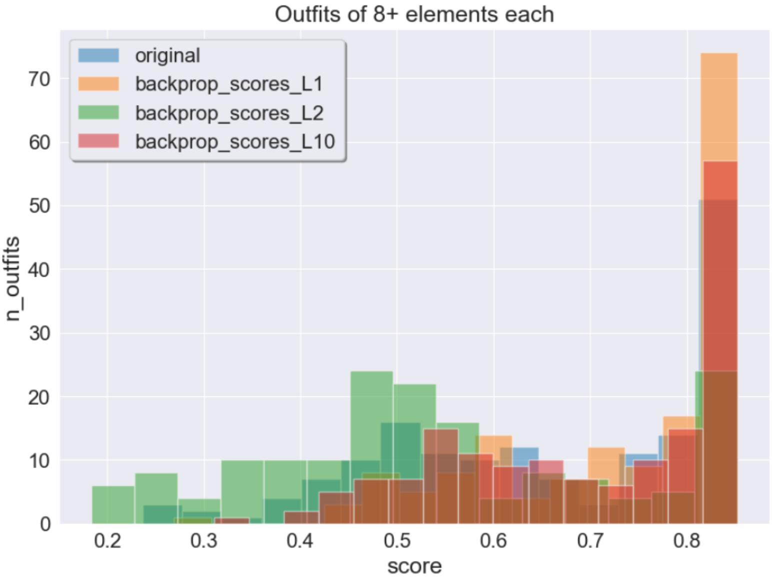
Вычислительный эксперимент

Непрерывная аппроксимация

- Рассматриваем те же самые образы, удаляем из каждого 2 элемента и рассматриваем дополнение получившихся образов
- Замораживаем веса модели оценки
- Добавляем 2 обучаемых эмбединга для недостающих элементов
- С помощью оптимизатора Adam получаем эмбединги, максимизирующие оценку
- Выбираем из всей коллекции элементы, ближайшие к полученным по некоторой метрике, в данном случае взяты L_1, L_2, L_{10}
- Вычисляем оценку получившегося образа

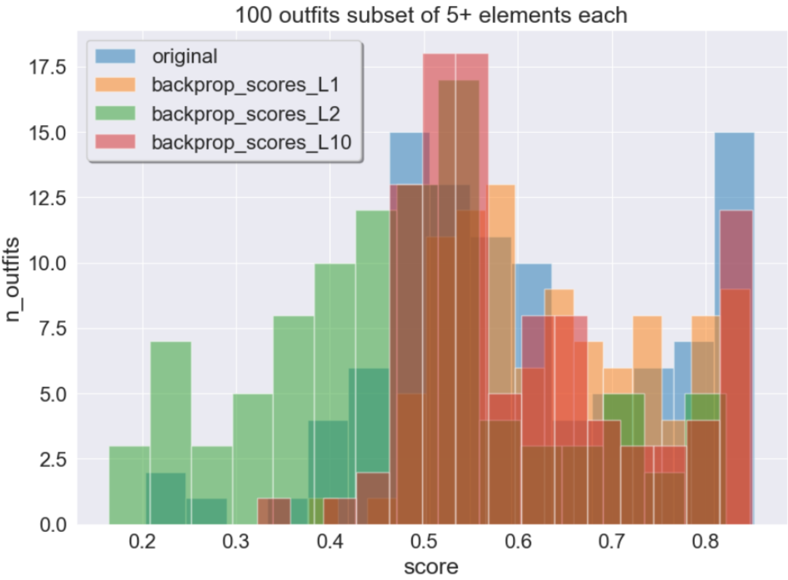
Вычислительный эксперимент

Непрерывная аппроксимация



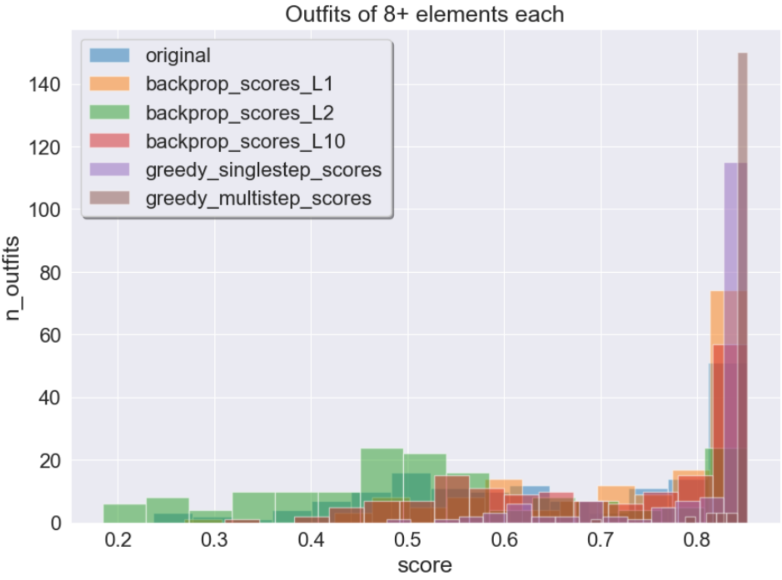
Вычислительный эксперимент

Непрерывная аппроксимация



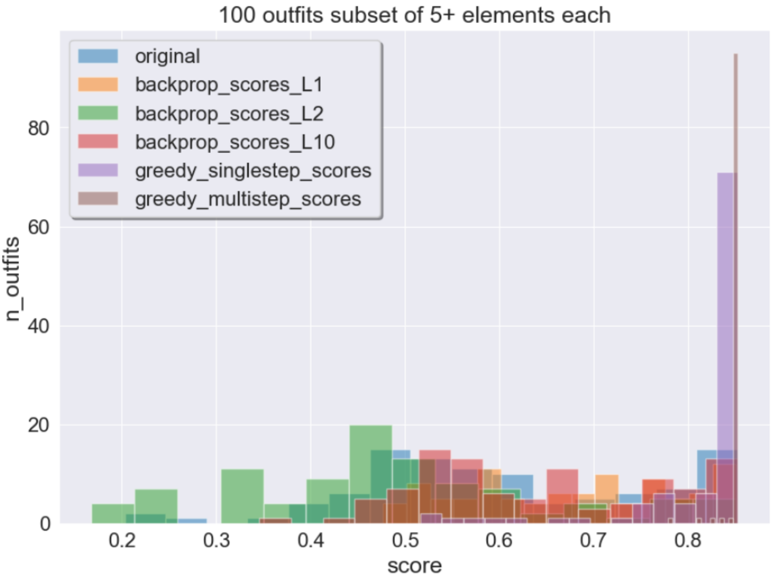
Вычислительный эксперимент

Сравнение



Вычислительный эксперимент

Сравнение



Вычислительный эксперимент

Выводы

- Жадные алгоритмы показывают хороший результат, но вычисления крайне не эффективны и занимают слишком много времени
- Метод непрерывного восстановления векторных представлений недостающих элементов серьезно уступает жадным
- Структура пространства представлений элементов слишком сложна, чтобы простые метрики близости позволяли выбрать лучший элемент коллекции
- Предлагается рассмотреть возможности агрегации представлений всех элементов перед выбором ближайшего для учета структуры пространства и взаимодействия элементов между собой.
- Агрегация может быть обучаемой. С учетом симметрии задачи, предлагается рассмотреть графовую нейронную сеть
- Исходя из постановки задачи, необходимо рассмотреть способы поощрения инвариантности к порядку выбора элементов
- Для обучения в дальнейшем можно применять элементы выбранные жадным образом, поскольку более точное решение задачи вряд ли достижимо за разумное время.