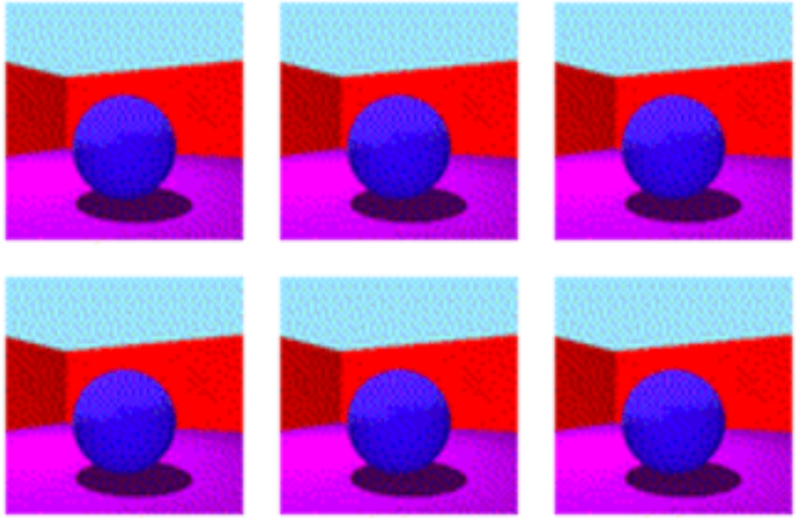
Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations

Кудрявцева Софья

Распутывание представлений

- Метод, при котором путем обучения модели строится вектор независимых параметров, где каждый из них означает отдельный фактор (положение, размер, угол вращения, цвет и т.д.)



disentanglement_lib

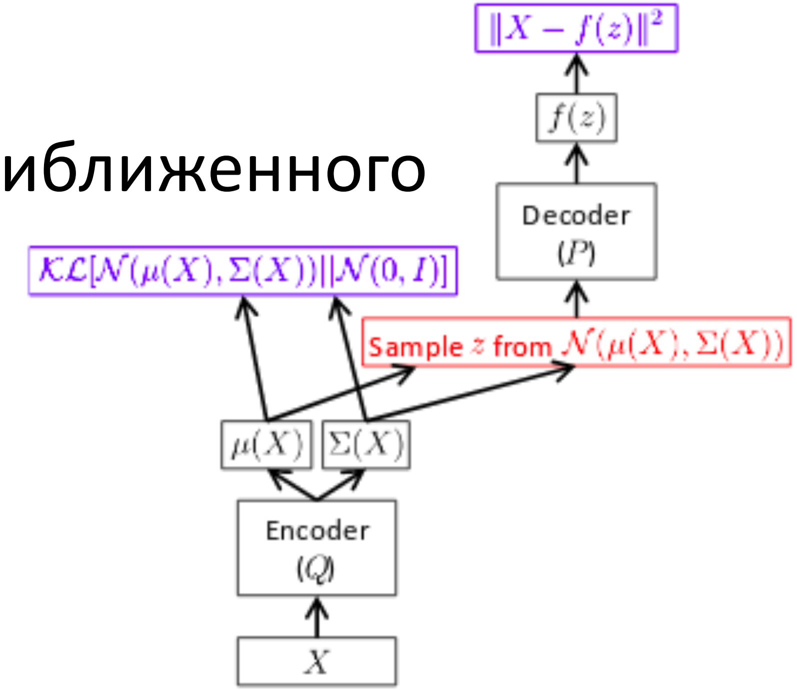
Задачи

- Исследовать возможность unsupervised распутывания представлений теоретически и экспериментально
- Проверить, на сколько полезно распутывание представлений для других задач

State-of-the-art неконтролируемого распутывания

- VAE

r(x) обычно принимается за среднее значение приближенного распределения Q(z|x)



$$\mathcal{L} = \mathbb{E}_{q(z|X)}[\log p(X|z)] - D_{KL}[q(z|X)||p(z)]$$

Теоретическое исследование возможности неконтролируемого распутывания представлений. Теорема 1.

- Для d > 1, пусть z ~ P обозначает любое распределение которое удовлетворяет условию $p(\mathbf{z}) = \prod_{i=1}^{d} p(\mathrm{z}_i)$ Тогда существует бесконечное множество функций f : supp(z) → supp(z) для которых $\frac{\partial f_i(\boldsymbol{u})}{\partial u_j} \neq 0$ почти для всех i и j. (то есть, z и f(z) полностью запутаны относительно друг друга) и P(z ≤ u) = P(f(z) ≤ u) для всех u ∈ supp(z) (то есть они имеют одинаковое предельное распределение).

Следствие теоремы

Неконтролируемое обучение распутыванию невозможно для произвольных генеративных моделей:

Предположим, что у нас есть p(z) и некоторые P(x|z), определяющие генеративную модель. Рассмотрим любой неконтролируемый метод распутывания и предположим, что он находит представление r(x) - совершенно распутанное относительно z в генеративной модели. Тогда по теореме 1 существует эквивалентная генеративная модель со скрытой переменной z’=f(z), где z’ полностью запутана относительно z и, следовательно, также r(x). Так как p(z) = p(z’) почти везде, обе генеративные модели имеют одинаковое предельное распределение наблюдений x по построению, т.е.

$$P(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z} = \int p(\mathbf{x}|\hat{\mathbf{z}})p(\hat{\mathbf{z}})d\hat{\mathbf{z}}.$$

Следствие теоремы

- Поскольку (неконтролируемый) метод распутывания имеет доступ только к наблюдениям x, он, следовательно, не может различать две эквивалентные генеративные модели. После наблюдения x мы можем построить бесконечно много генеративных моделей, которые имеют одно и то же конечное распределение x.

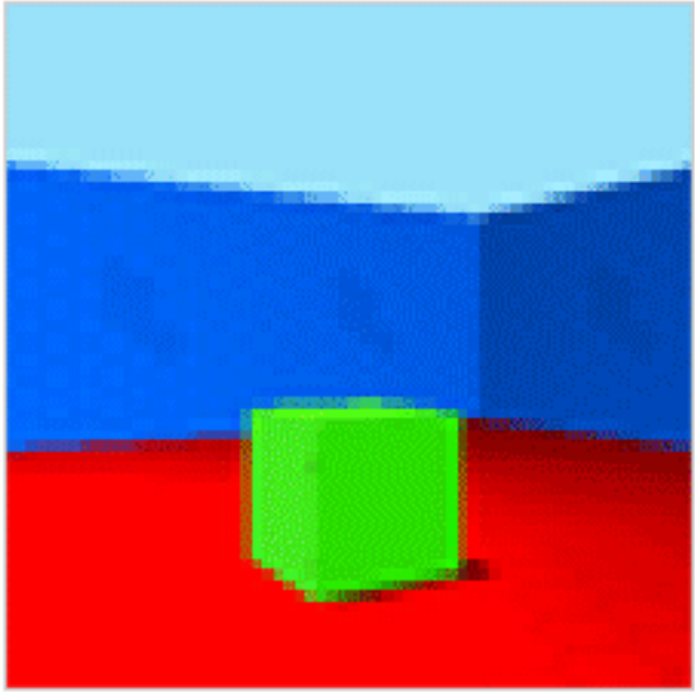
Экспериментальное исследование возможности неконтролируемого распутывания представлений

Все рассмотренные методы увеличивают потери VAE с помощью регуляризатора:

- β-VAE $$\mathcal{L} = \mathbb{E}_{q(z|X)}[\log p(X|z)] - \beta D_{KL}[q(z|X)||p(z)]$$
- AnnealedVAE $$\mathbb{E}_{p(\mathbf{x})}[\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \gamma|D_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z})) - C||]$$
- β-TCVAE $$D_{KL}(q(z)||p(z)) = D_{KL}(q(z)||\prod_j q(z_j)) + \sum_j D_{KL}(q(z_j)||p(z_j))$$

Наборы данных

- 4 набора данных, в которых изображение генерируется из независимого скрытого представления : dSprites, Cars3D, SmallNORB, Shapes3D.



Наборы данных

- 3 набора данных, в которых наблюдения x являются стохастическими с учетом фактора вариаций z: Color-DSPR, Noisy-dSprites и Scream-dSprites.
- В Color-dSprites фигуры окрашиваются случайным цветом. В Noisy-dSprites - фигуры белого цвета на шумном фоне. В Scream-dSprites фон заменяется случайным пятном в случайном цветовом оттенке из картины “Крик”.

Метрики

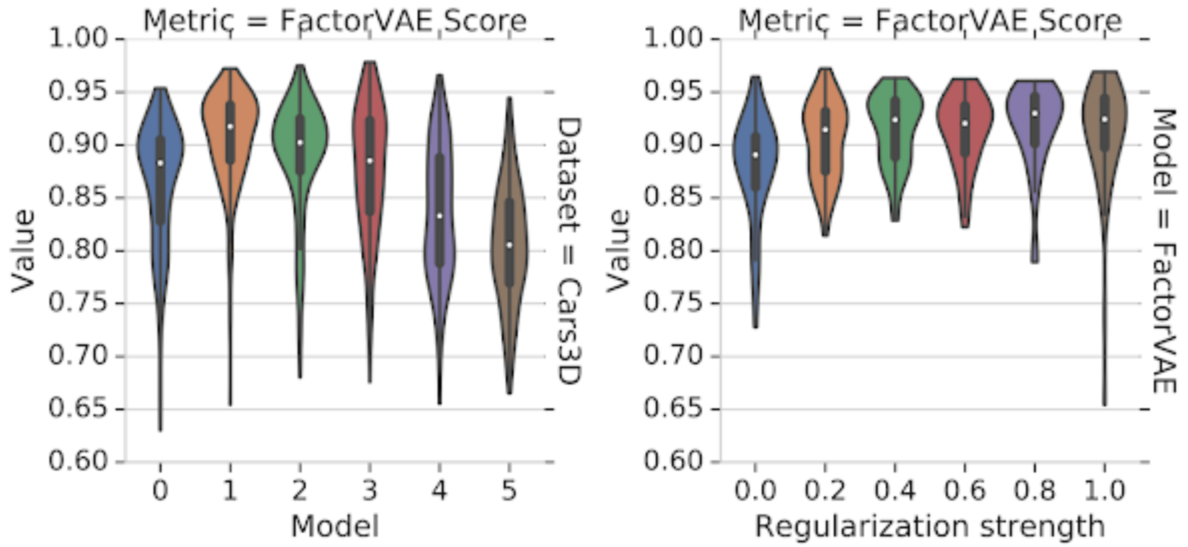
- Метрика BetaVAE - точность линейного классификатора, который предсказывает индекс фиксированного коэффициента вариации.
- FactorVAE - классификатор большинства голосов BetaVAE.
- MIG - измеряет для каждого фактора вариации нормированный разрыв во взаимной информации между первой и второй по величине координатами в r(x).
- Модульность измеряет, зависит ли каждое измерение r(x) не более чем от одного фактора вариации, используя их взаимную информацию.
- DCI Disentanglement - энтропия распределения, полученная путем нормализации важности каждого измерения изученного представления для предсказания значения фактора вариации.
- SAP - среднюю разность ошибок прогнозирования двух наиболее прогностических латентных измерений для каждого фактора.

Эксперименты

- Каждый метод использовал одну и ту же convolutional архитектуру, оптимизатор, размер батча, гаусовский энкодер, Декодер Бернулли и скрытые переменные размером 10.

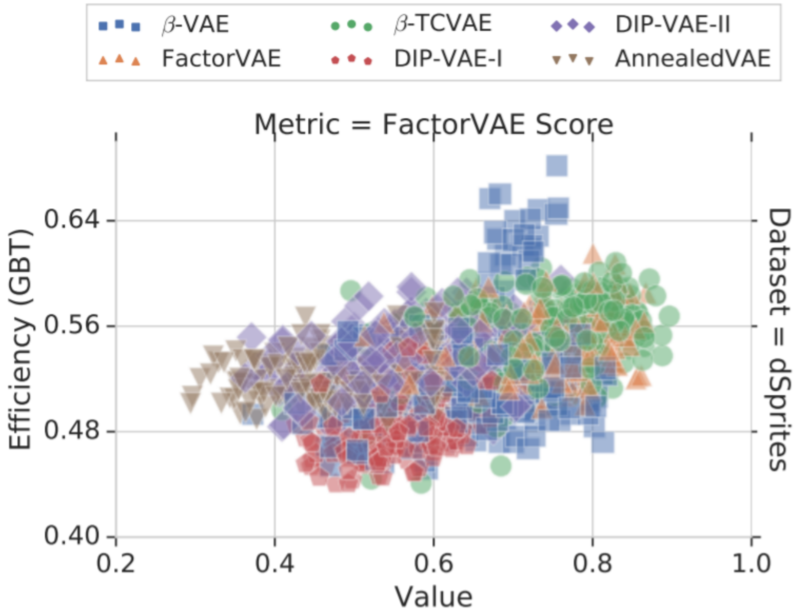
Выводы

- Случайные начальные числа и гиперпараметры имеют большее значение, чем выбор модели

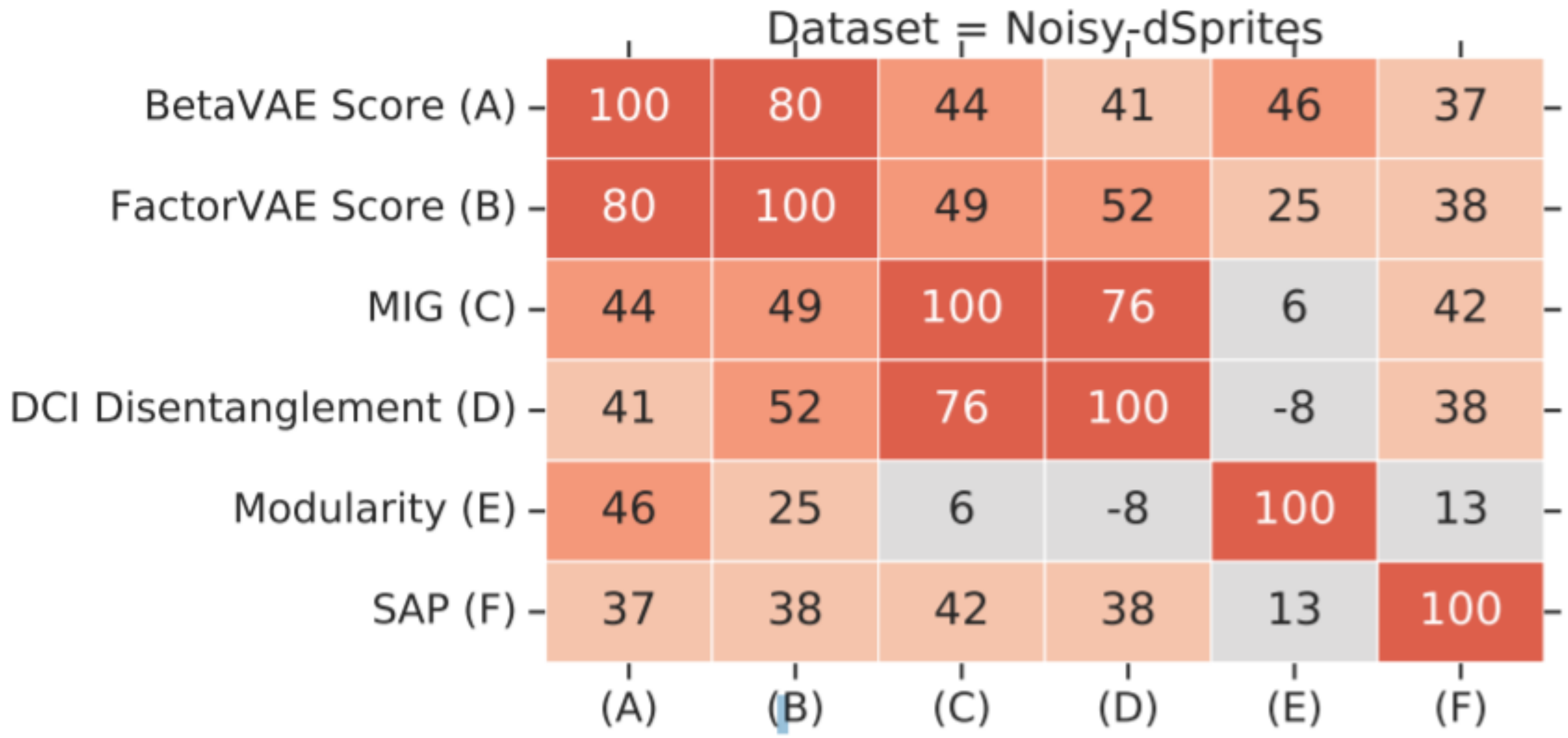


Выводы

- Для рассматриваемых моделей и датасетов неподтверждается, что распутывание полезно для последующих задач. Например, что с распутанными представлениями можно обучаться на меньшем количестве размеченных наблюдений.



Корреляция метрик



Дальнейшие направления исследования

- Исследовать эту задачу с supervised или semi-supervised learning
- Добавление во входные данные дополнительной информации
- Эксперименты на различных данных. Для этого выпущена библиотека с предобученными VAE

Ссылки

- https://github.com/google-research/disentanglement_lib - библиотека от Google AI содержит ***10800 вариационных автоэнкодеров***, обученных на семи датасетах