

Аннотация

Исследуется проблема понижения сложности аппроксимирующей модели при переходе к данным домена меньшей мощности. Вводятся понятия учителя, ученика, слабого и сильного доменов. Признаковые описания моделей ученика и учителя принадлежат разным доменам. Мощность одного домена больше мощности другого. Рассматриваются методы, основанные на дистилляции моделей машинного обучения. Вводится предположение, что решение оптимизационной задачи от параметров обеих моделей и доменов повышает качество модели ученика.

Содержание

1	Введение	4
2	Анализ литературы	4
3	Постановка задачи	5
3.1	Базовая постановка задачи дистилляции Хинтона . . .	5
3.2	Постановка задачи дистилляции для многодоменной выборки	6
4	Вычислительный эксперимент	7
4.1	Базовый эксперимент	8

1 Введение

2 Анализ литературы

В [1] рассматривается метод учета меток учителя, используя функцию softmax с параметром температуры.

В [2] рассматривается метод дистилляции в случае несовпадения признаков описаний ученика и учителя

3 Постановка задачи

3.1 Базовая постановка задачи дистилляции Хинтона

Задано множество объектов Ω и множество целевых переменных \mathbb{Y} . Множество $\mathbb{Y} = \{1, \dots, R\}$ для задачи классификации, где R - число классов, множество $\mathbb{Y} = \mathbb{R}$ для задачи регрессии.

В постановке Хинтона в качестве модели ученика \mathbf{g} рассматривается функция из множества:

$$\mathfrak{G} = \{\mathbf{g} | \mathbf{g} = \text{softmax}(\mathbf{z}(\mathbf{x})/T), \mathbf{z} : \mathbb{R}^n \rightarrow \mathbb{R}^R\}$$

В качестве модели учителя \mathbf{f} рассматривается функция из множества:

$$\mathfrak{U} = \{\mathbf{g} | \mathbf{g} = \text{softmax}(\mathbf{v}(\mathbf{x})/T), \mathbf{v} : \mathbb{R}^n \rightarrow \mathbb{R}^R\}$$

\mathbf{v}, \mathbf{z} - дифференцируемые параметрические функции заданной структуры, T - параметр температуры со свойствами:

- 1) при $T \rightarrow 0$ получаем вектор, в котором один из классов имеет единичную вероятность;
- 2) при $T \rightarrow \infty$ получаем равновероятные классы.

Функция потерь \mathcal{L} учитывает перенос информации от модели учителя \mathbf{f} к модели ученика \mathbf{g} имеет вид

$$\mathcal{L} = - \sum_{i=1}^m \sum_{r=1}^R y_i^r \log g(x_i)|_{T=1} - \sum_{i=1}^m \sum_{r=1}^R f(x_i)|_{T=T_0} \log g(x_i)|_{T=T_0},$$

где $\cdot|_{T=t}$ означает, что параметр температуры T в предыдущей функции равен t .

3.2 Постановка задачи дистилляции для много-доменной выборки

Заданы множества объектов \mathbb{X}, \mathbb{X}' - данные первого и второго доменов, и множество целевых переменных \mathbb{Y} . Множество $\mathbb{Y} = \{1, \dots, R\}$ для задачи классификации, где R - число классов, множество $\mathbb{Y} = \mathbb{R}$ для задачи регрессии. \mathbf{f}, \mathbf{g} - модели учителя и ученика соответственно. Рассматриваются отображения

$$\varphi : \mathbb{X} \rightarrow \mathbb{X}', |\mathbb{X}'| \gg |\mathbb{X}|$$

$$\mathbf{f} : \mathbb{X}' \rightarrow \mathbb{Y}$$

Требуется получить отображение

$$\mathbf{g} : \mathbb{X} \rightarrow \mathbb{Y}$$

4 Вычислительный эксперимент

Выборка FashionMNIST. Эксперимент проводился для задачи классификации для выборки FashionMNIST [3]. В качестве модели учителя f рассматривается четырёхслойная нейросеть, в качестве функции активации рассматривается ReLu. В качестве модели ученика рассматривается однослойная нейросеть.

На рисунках 1, 2 показаны графики зависимостей Аккуратности и кросс-энтропии на тестовой выборке между истинными метками объектов и вероятностями, предсказанными моделью ученика. На графиках видно, что модель, использующая метки учителя, показывает лучшее значение Аккуратности, но большее значение ошибки.

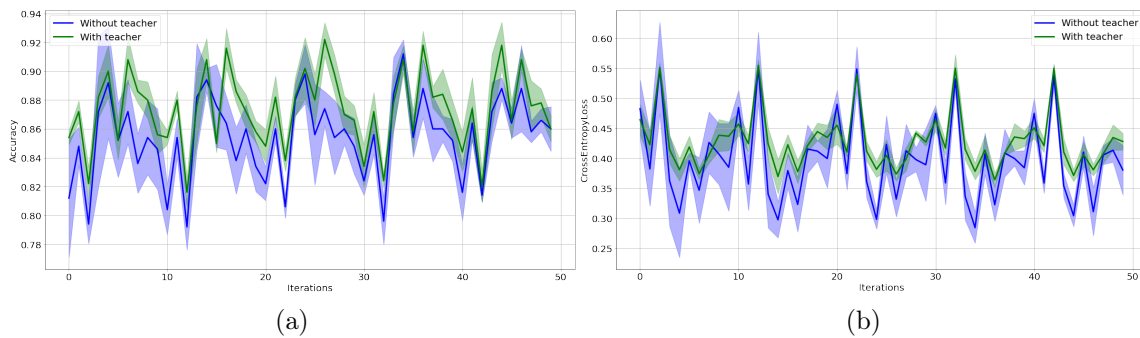


Рис. 1: Зависимость а) Accuracy; б) CrossEntropyLoss между истинными и предсказанными учеником метками от числа итераций на тестовой выборке

4.1 Базовый эксперимент

Список литературы

- [1] *Hinton G., Vinyals O., Dean J* Distilling the Knowledge in a Neural Network // NIPS Deep Learning and Representation Learning Workshop. — 2015.
- [2] *D. Lopez-Paz, L. Bottou, B. Schölkopf, V. Vapnik* Unifying distillation and privileged information // ICLR. — 2016.
- [3] *Xiao H., Rasul K., Vollgraf R.* Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. — 2017. <https://arxiv.org/abs/1708.07747>.