

Исследование механизмов разрежения нейронных сетей на основе важностей весов НИР

А. В. Ребриков

Научный руководитель: к.ф.-м.н., Безносовых А. Н.

Московский физико-технический институт

21 декабря 2024 г.

Литература



Yarin Gal, Jiri Hron, and Alex Kendall.

Concrete dropout.

Advances in neural information processing systems, 30, 2017.



Rohit Keshari, Richa Singh, and Mayank Vatsa.

Guided dropout.

In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4065–4072, 2019.



Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov.

Dropout: a simple way to prevent neural networks from overfitting.

The journal of machine learning research, 15(1):1929–1958, 2014.



Li Wan, Matthew Zeiler, Sixin Zhang, Yann Le Cun, and Rob Fergus.

Regularization of neural networks using dropconnect.

In *International conference on machine learning*, pages 1058–1066. PMLR, 2013.

Слайд об исследованиях

Исследуется проблема регуляризации нейронных сетей за счёт разрежения, например с помощью dropout, DropConnect.

Цель исследования —

предложить метод разрежения нейронных сетей на основе *важности* весов.

Предлагается

- 1) определить *важность* весов,
- 2) разрежение с вероятностями, зависящими от важности весов,
- 3) комбинация с известными методами разрежения для получения лучших результатов.

Решение

Для определения важности весов предлагается исследовать влияние изменения каждого отдельного веса в рамках одного слоя (параметра) сети на функцию ошибки.

Постановка задачи регуляризации

Заданы

- 1) признаки $a_{\text{train}}, a_{\text{test}} \in A^{n+m}$, метки $b_{\text{train}}, b_{\text{test}} \in B^{n+m} = \mathbb{R}^{r \times (n+m)}$,
- 2) веса модели: $x \in X = X_1 \times X_2 \times \dots \times X_L$,
- 3) модель $f : X \times A \rightarrow B$,
- 4) функция ошибки $\mathcal{L} : X \rightarrow \mathbb{R}$.

При классической постановке задачи:

$$\frac{1}{n} \sum_{i=1}^n \mathcal{L}(f(x, a_{\text{train}}^i), b_{\text{train}}^i) \rightarrow \min_{x \in X}$$

Где решение ищется с помощью градиентного спуска.

Дополнительно хотим минимизировать разность с тестовой выборкой:

$$\text{GAP} = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(f(x, a_{\text{test}}^i), b_{\text{test}}^i) - \frac{1}{m} \sum_{i=1}^m \mathcal{L}(f(x, a_{\text{train}}^i), b_{\text{train}}^i)$$

Предложенный метод важности весов

Заданы

- 1) признаки $a_{\text{train}}, a_{\text{test}} \in A^{n+m}$, метки $b_{\text{train}}, b_{\text{test}} \in B^{n+m} = \mathbb{R}^{r \times (n+m)}$,
- 2) веса модели: $x \in X = X_1 \times X_2 \times \dots \times X_L$,
- 3) модель $f : X \times A \rightarrow B$,
- 4) функция ошибки $\mathcal{L} : X \rightarrow \mathbb{R}$,
- 5) важность весов: $w \in W = W_1 \times W_2 \times \dots \times W_L$, $W_i = |\dim X_i| \Delta_{\dim X_i}$.

Поиск важности весов

$$\mathcal{L}(f(x - w \odot (\gamma \nabla f), a_{\text{train}}^i), b_{\text{train}}^i) \rightarrow \min_{w \in W}$$

где γ – шаг градиентного спуска.

Далее проводится DropConnect с вероятностями, зависящими от важности весов.

[2]

Описание предложенного метода

- ▶ Исследовался процесс, при котором полученные важности означали вероятность сохранения веса при DropConnect. Можно рассмотреть и другие варианты, например, наоборот, вероятность удаления веса (с надлежащим нормированием).
- ▶ Для решения задачи поиска важности весов используется зеркальный спуск на симплексе. В качестве регуляризации используется KL-дивергенция с равномерным распределением.
- ▶ Эксперименты проводились на задачах классификации изображений CIFAR-10 с помощью RESNET-18.

Эксперименты

С применением регуляризации мы получаем более информативную гистограмму. Иначе зеркальный спуск сходится к вырожденному решению (выбор одного веса). Для понимания гистограммы: в случае равномерного распределения все value равны единице, в случае вырожденного – все value равны нулю, кроме одного. Мы же стремимся к тому, чтобы выявить группу более важных весов.

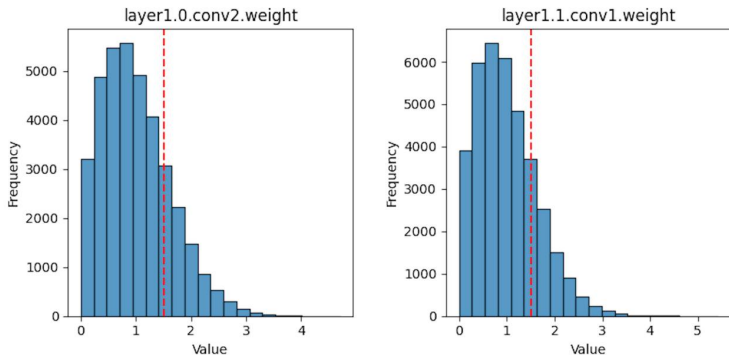


Рис.: Гистограмма важности веса (value)

Эксперименты

Сравнивались

- ▶ классическое обучение модели (baseline),
- ▶ обучение с DropConnect на основе важности весов (impacts),
- ▶ обучение с DropConnect на основе важности весов, но используя регуляризацию с равномерным распределением (impacts+regularization),
- ▶ обучение с классическим DropConnect (dropconnect).

В последних трёх пунктах количество весов, которые не использовались, было одинаковым.

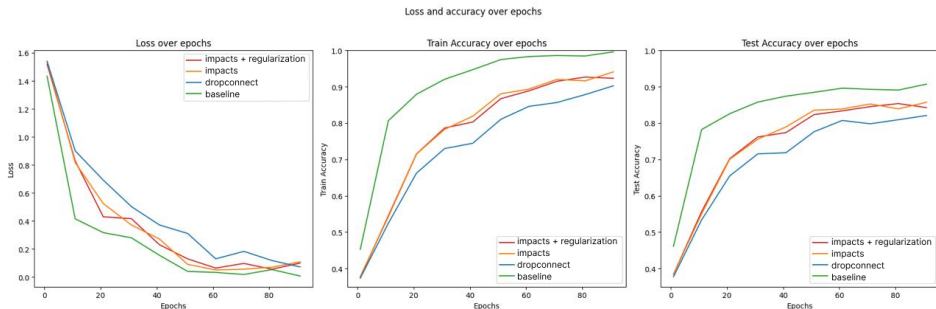


Рис.: Графики сходимости

Выводы

На текущий момент:

1. Предложен метод определения важности весов, основанный на зеркальном спуске на симплексе.
2. Предложен метод регуляризации нейронных сетей: DropConnect на основе важности весов.
3. Экспериментально показано, что предложенный метод позволяет получить лучшие результаты по сравнению с классическим DropConnect (на ранних этапах обучения).

В будущем:

1. Исследовать метод на поздних этапах обучения.
2. Исследовать другие подходы к разрежению на основе важности.