
Исследование механизмов разрежения нейронных сетей на основе важностей весов

A Preprint

А. В. Ребриков*

Физтех-школа прикладной математики и информатики
Московский физико-технический институт (МФТИ)
rebryakov.av@phystech.edu

А. Н. Безноси́ков†

Московский физико-технический институт (МФТИ)
beznosikov.an@phystech.edu

Abstract

В данной работе рассматривается задача регуляризации нейронных сетей путём их разрежения. Традиционные методы, такие как Dropout и DropConnect, «выключают» часть весов случайно, что помогает бороться с переобучением. Мы предлагаем иной подход к определению вероятности выключения весов, основываясь на их важности — оценке вклада конкретных параметров в итоговую функцию ошибки. Эксперименты с ResNet-18 на задаче классификации изображений CIFAR-10 показывают преимущество предлагаемых идей на ранних этапах обучения.

Keywords Разрежение · Нейронные сети · DropConnect · Важность весов · Регуляризация

1 Введение

В последние годы глубокие нейронные сети достигли значительного прогресса в решении сложных задач компьютерного зрения, обработки естественного языка и многих других областей. Однако большим практическим вызовом остаётся проблема переобучения, а также избыточное количество параметров, которое делает модели ресурсоёмкими и сложными для интерпретации.

Одним из наиболее известных подходов к сдерживанию переобучения является метод Dropout [Srivastava et al., 2014], при котором случайным образом «обнуляются» некоторые нейроны на стадии обучения. Аналогичным образом в DropConnect [Wan et al., 2013] «обнуляются» отдельные веса. Как правило, вероятность зануления задаётся глобально и не учитывает особенности и вклад отдельных весов.

С другой стороны, в последнее время появилось множество работ, исследующих механизмы разрежения нейронных сетей, основанные на важности весов. Например, в работе [Keshari et al., 2019] предлагается использовать «силу нейронов», т. е. оценку их вклада в функцию ошибки, и обучать только наименее сильные.

В данной работе мы предлагаем метод разрежения, использующий важность весов, т. е. оценку их вклада в функцию ошибки. Для определения важности предлагается зеркальный спуск по соответствующему симплексу для каждого параметра, с дополнительной регуляризацией, препятствующей вырождению в пользу одного веса.

* Данная работа выполнена в рамках научно-исследовательской работы (НИР) в Московском физико-техническом институте.

† Научный руководитель.

2 Постановка задачи

Рассмотрим задачу обучения нейронной сети на выборке:

1. Дано n обучающих примеров $(a_{\text{train}}^i, b_{\text{train}}^i)$, а также m тестовых примеров $(a_{\text{test}}^i, b_{\text{test}}^i)$.
2. Пусть $x \in X = X_1 \times \dots \times X_L$ — совокупность весов сети (все слои).
3. Модель $f(x, a): X \times A \rightarrow B$ возвращает результат (например, вектор вероятностей классов).
4. Функция ошибки $\mathcal{L}(f(x, a^i), b^i)$ суммируется по всем объектам обучающей выборки.

Традиционная постановка задачи обучения нейронной сети:

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

При этом нас дополнительно интересует «обобщающая способность», измеряемая разностью качества между тестовой и обучающей выборками (GAP):

$$\text{GAP} = \frac{1}{m} \sum_{i=1}^m \mathcal{L}(f(x, a_{\text{test}}^i), b_{\text{test}}^i) - \frac{1}{n} \sum_{i=1}^n \mathcal{L}(f(x, a_{\text{train}}^i), b_{\text{train}}^i). \quad (2)$$

Целью регуляризации служит снижение риска переобучения, то есть уменьшение величины GAP.

3 Определение важности весов и разрежение

Для построения механизма разрежения на основе важности весов необходимо:

1. Определить меру важности для каждого веса x_i .
2. На основе этой меры задавать вероятности «включения» (или «выключения») весов при DropConnect.
3. В случае необходимости сочетать полученную схему с классическим DropConnect или другими известными техниками.

3.1 Зеркальный спуск на симплексе

В работе предлагается найти распределение важности w размерности, совпадающей с количеством весов. Для этого решается задача:

$$\mathcal{L}(f(x - w \odot (\gamma \nabla f(x)), a_{\text{train}}), b_{\text{train}}) \rightarrow \min_w, \quad (3)$$

где w лежит на соответствующем симплексе, умноженном на количество весов, γ — параметр «шага». Решение реализуется методом зеркального спуска с регуляризацией в виде KL-дивергенции:

$$R(w) = \text{KL}(w \parallel \mathbf{u}), \quad (4)$$

где \mathbf{u} — равномерное распределение (в данном случае, все компоненты равны 1). Это препятствует «вырождению» в пользу одного веса, одновременно позволяя различать важные и менее важные веса.

4 Применение DropConnect на основе важности

DropConnect [Wan et al., 2013] предполагает, что в процессе обучения каждая компонента x_i (вес сети) «выключается» с некоторой вероятностью p . Мы же можем задать это p как функцию от найденной важности w_i . Например,

$$p_i = 1 - \alpha w_i \quad (\text{или наоборот } p_i = \alpha w_i),$$

где α — параметр, задающий уровень разрежения.

Далее обучения сети проходит с учётом того, что веса выбираются случайно в соответствии с вероятностями p_i .

5 Эксперименты

5.1 Настройка экспериментов

Эксперименты проводились на датасете CIFAR-10 с помощью архитектуры ResNet-18. Рассматривались следующие методы:

- baseline: обучение без разрежения;
- impacts: обучение с DropConnect, где вероятность задаётся на основе важностей весов;
- impacts+regularization: то же, что impacts, но с дополнительной регуляризацией KL-дивергенции, препятствующей вырождению распределения важности;
- dropconnect: классический DropConnect с фиксированной вероятностью выключения.

Во всех трёх вариантах с разрежением старались «выключать» одинаковое количество весов, чтобы иметь сопоставимые условия.

5.2 Результаты

На Рисунке 1 показаны гистограммы важных весов с регуляризацией. При отсутствии регуляризации наблюдалось «выделение» единственного веса с максимальной важностью, что приводило к вырожденной ситуации. С KL-регуляризацией распределение более плавное, сохраняя «группу» наиболее важных весов.

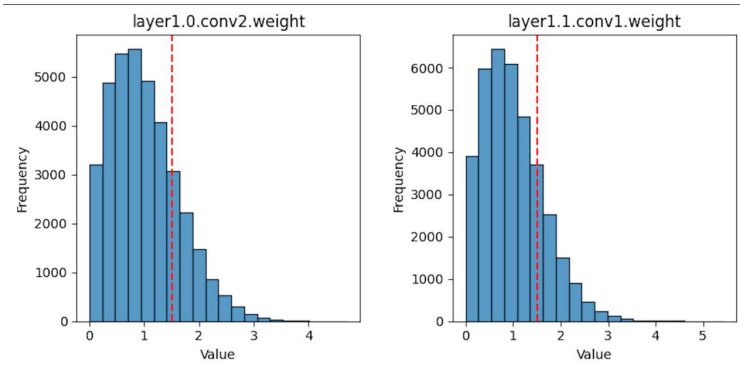


Рис. 1: Гистограммы важностей весов с регуляризацией.

На Рисунке 2 представлены графики обучения (точность на валидации и ошибка), показывающие, что на ранних этапах метод impacts+regularization даёт лучший результат по сравнению с классическим DropConnect.

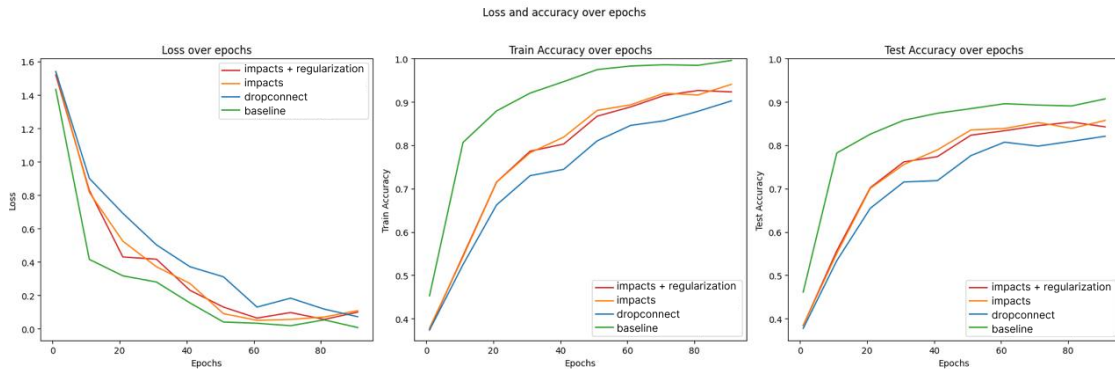


Рис. 2: Сходимость на ранних эпохах обучения ResNet-18 с разрежением в вариантах baseline, dropconnect, impacts и impacts+regularization.

6 Выводы и дальнейшие исследования

В работе представлен новый подход к регуляризации нейронных сетей, основанный на вычислении важности весов и применении DropConnect с вероятностями, зависящими от важности. Эксперименты показали преимущество такого подхода на ранних этапах обучения по сравнению с классическими методами.

Основные результаты:

1. Предложен метод определения важности весов с помощью зеркального спуска на симплексе и KL-регуляризации.
2. Показано, что основанный на важностях метод DropConnect даёт более высокое качество в начальной фазе обучения (и сопоставимое в дальнейших фазах).
3. Сохранение «небольшой группы важных весов» выглядит более разумным, чем равномерное вероятностное выключение всех весов.

Перспективы:

- Исследовать устойчивость и поведение предлагаемого подхода на поздних этапах обучения.
- Рассмотреть комбинирование с другими методами разрежения, включая различные эвристики пороговых значений весов, L1-регуляризацию и т. д.
- Применение на более сложных архитектурах и задачах (ImageNet, NLP).

Благодарности

Авторы благодарят МФТИ за предоставленную вычислительную инфраструктуру и поддержку исследований.

Список литературы

- 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.
- Rohit Keshari, Richa Singh, and Mayank Vatsa. Guided dropout. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4065–4072, 2019.
- Yarin Gal, Jiri Hron, and Alex Kendall. Concrete dropout. *Advances in neural information processing systems*, 30, 2017.