

# Регуляризация траектории параметров модели глубокого обучения на основе дистилляции знаний

М. Горпинич, О. Ю. Бахтеев, В. В. Стрижов

gorpinich4@gmail.com; bakhteev@phystech.edu; strijov@ccas.ru

Исследуется задача оптимизации параметров модели глубокого обучения. Во время оптимизации учитывается информация, содержащаяся в модели с более сложной структурой, то есть применяется дистилляция. Предлагается обобщение методов дистилляции, заключающееся в градиентной оптимизации метапараметров. Под метапараметрами модели понимаются параметры оптимизационной задачи дистилляции, а именно, коэффициенты перед слагаемыми в функции ошибки и температура. Функция ошибки состоит из двух слагаемых: правдоподобия исходной выборки и правдоподобия выборки дистилляции. Исследуются свойства оптимизационной задачи и методы прогнозирования траектории оптимизации метапараметров модели. Предложенное обобщение позволяет получить модель с лучшими эксплуатационными характеристиками и за меньшее число итераций оптимизации. Проиллюстрирован данный подход с помощью вычислительного эксперимента на выборке CIFAR-10 и на синтетической выборке.

**Ключевые слова:** машинное обучение; дистилляция знаний; оптимизация метапараметров; градиентные методы оптимизации; прогнозирование метапараметров

DOI:

## 1 Введение

В работе исследуется задача оптимизации моделей глубоких нейросетей. Проведение оптимизации требует значительных вычислительных мощностей и является затратной по времени. В данной работе предлагается метод оптимизации, позволяющий улучшить точность предсказаний модели, а также ускорить сходимость траектории оптимизации параметров к точке оптимума.

Предлагается обобщение метода оптимизации на основе дистилляции знаний. Назовем *дистилляцией знаний* задачу оптимизации параметров модели прогнозирования, при которой учитывается не только информация, содержащаяся в выборке, но также и информация, содержащаяся в сторонней модели (модели-учителе). Рассматривается *модель-учитель* более сложной структуры, которая была обучена на выборке. Модель более простой структуры предлагается оптимизировать путем переноса знаний модели учителя на более простую модель, называемую *моделью-учеником*. При этом ее качество будет выше по сравнению с качеством, полученным после оптимизации на той же выборке. Данный подход описан в [1]. В [2] предложен подход к дистилляции знаний, переносящий знания на модель с архитектурой, значительно отличающейся от архитектуры модели-учителя.

Предлагается формулировка задачи в виде двухуровневой оптимизации. На первом уровне оптимизируются параметры модели, на втором уровне — ее метапараметры. Данный подход описан в [3–5]. В [3] рассматривается жадный градиентный метод оптимизации метапараметров. В [4] сравниваются различные градиентные методы оптимизации метапараметров, а также метод случайного поиска.

В работе рассматривается подход к прогнозированию значений метапараметров, полученных методом градиентной оптимизации. Под метапараметрами понимаются параметры

задачи оптимизации. Сложность градиентной оптимизации для метапараметров является квадратичной по числу параметров, и потому вычислительно затратна. Предлагается аппроксимация траектории оптимизации метапараметров на основе приближения траектории линейной моделью. Вычислительный эксперимент проводится на выборке изображений CIFAR-10 [6], а также синтетической выборке.

## 2 Постановка задачи

Решается задача классификации вида:

$$\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^m, \quad \mathbf{x}_i \in \mathbb{R}^n, \quad y_i \in \mathbb{Y} = \{1, \dots, K\}, \quad (1)$$

где  $y_i$  — это класс объекта, также будем обозначать  $\mathbf{y}_i$  вектором вероятности для класса  $y_i$ . Разобьем выборку следующим образом:

$$\mathfrak{D} = \mathfrak{D}_{\text{train}} \sqcup \mathfrak{D}_{\text{val}}. \quad (2)$$

Подвыборку  $\mathfrak{D}_{\text{train}}$  будем использовать для оптимизации параметров модели, а подвыборку  $\mathfrak{D}_{\text{val}}$  — для оптимизации метапараметров.

Внешним критерием качества назначена доля правильных ответов:

$$\text{accuracy} = \frac{1}{m} \sum_{i=1}^m [\mathbf{g}(\mathbf{x}_i, \mathbf{w}) = y_i], \quad (3)$$

где  $\mathbf{g}$  — параметрическая модель классификации с параметрами  $\mathbf{w}$ .

**Определение 1.** Траекторией параметров  $\boldsymbol{\nu}$  назовем последовательность  $\mathbf{w}_1, \dots, \mathbf{w}_t, \dots$  обновления параметров в ходе оптимизации, где  $t$  — количество шагов оптимизации.

**Определение 2.** Пусть задана функция  $D : \mathbb{R}^s \rightarrow \mathbb{R}_+$ , задающая схожесть модели-ученика  $\mathbf{g}$  и фиксированной модели-учителя  $\mathbf{f}$ .  $D$ -дистилляцией модели-ученика назовем оптимизацию параметров модели-ученика с траекторией  $\boldsymbol{\nu}$ , такую что  $\lim_{t \rightarrow \infty} D(\mathbf{w}_t) = \min_{\mathbf{w}'_t \in \mathbb{R}^s} D(\mathbf{w}'_t)$ .

Функция потерь  $\mathcal{L}_{\text{train}}$ , в которой учитывается перенос информации от модели учителя  $\mathbf{f}$  к модели ученика  $\mathbf{g}$ , имеет вид:

$$\begin{aligned} \mathcal{L}_{\text{train}}(\mathbf{w}, \boldsymbol{\lambda}) = & -\lambda_1 \sum_{(\mathbf{x}, y) \in \mathfrak{D}_{\text{train}}} \underbrace{\sum_{k=1}^K y^k \log \frac{e^{\mathbf{g}(\mathbf{x}, \mathbf{w})_i/T}}{\sum_j e^{\mathbf{g}(\mathbf{x}, \mathbf{w})_j/T}} \Big|_{T=1}}_{\text{исходная функция потерь}} \\ & - \lambda_2 \sum_{(\mathbf{x}, y) \in \mathfrak{D}_{\text{train}}} \underbrace{\sum_{k=1}^K \frac{e^{\mathbf{f}(\mathbf{x})_i/T}}{\sum_j e^{\mathbf{f}(\mathbf{x})_j/T}} \Big|_{T=T_0} \log \frac{e^{\mathbf{g}(\mathbf{x}, \mathbf{w})_i/T}}{\sum_j e^{\mathbf{g}(\mathbf{x}, \mathbf{w})_j/T}} \Big|_{T=T_0}}_{\text{слагаемое дистилляции}}, \quad (4) \end{aligned}$$

где  $y^k$  —  $k$ -я компонента целевого вектора (ответ на  $k$ -м классе),  $T$  — параметр температуры. Параметр температуры  $T$  имеет следующие свойства:

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

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

Зададим множество метапараметров  $\lambda$  как вектор, состоящий из температуры и коэффициента перед слагаемым дистилляции:

$$\lambda = [\lambda_1, \lambda_2, T].$$

**Утверждение 1.** Функция потерь при  $\lambda_1 = 0$  является KL-дистилляцией модели.

*Доказательство* При  $\lambda_1 = 0$ :

$$\mathcal{L}_{\text{train}} = \sum_{(\mathbf{x}, y) \in \mathcal{D}_{\text{train}}} \sum_{k=1}^K \mathbf{f}(\mathbf{x})|_{T=T_0} \log \mathbf{g}(\mathbf{x}, \mathbf{w})|_{T=T_0} = D_{KL}(\mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x}, \mathbf{w}))$$

По определению  $D$ -дистилляции получим, что должно выполняться условие:

$$\lim_{t \rightarrow \infty} D_{KL}(\mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x}, \mathbf{w}_t)) = \min_{\mathbf{w}'_t \in \mathbb{R}^s} D_{KL}(\mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x}, \mathbf{w}'_t))$$

Так как для оптимизации используется градиентный спуск, то

$$D_{KL}(\mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x}, \mathbf{w}'_{i+1})) \leq D_{KL}(\mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x}, \mathbf{w}'_i))$$

Также выполняется  $D_{KL}(\mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x}, \mathbf{w})) \geq 0$

Тогда получаем, что

$$\lim_{t \rightarrow \infty} D_{KL}(\mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x}, \mathbf{w}_t)) = \min_{\mathbf{w}'_t \in \mathbb{R}^s} D_{KL}(\mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x}, \mathbf{w}'_t))$$

□

Итоговая оптимизационная задача выглядит следующим образом:

$$\hat{\lambda} = \arg \max_{\lambda \in \mathbb{R}^3} \mathcal{L}_{\text{val}}(\hat{\mathbf{w}}, \lambda), \quad (5)$$

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{R}^s} \mathcal{L}_{\text{train}}(\mathbf{w}, \lambda), \quad (6)$$

где функция  $\mathcal{L}_{\text{val}}$  определяется как:

$$\mathcal{L}_{\text{val}}(\mathbf{w}, \lambda) = \sum_{(\mathbf{x}, y) \in \mathcal{D}_{\text{val}}} \sum_{k=1}^K y^k \log \mathbf{g}(\mathbf{x}, \mathbf{w})|_{T=1}. \quad (7)$$

### 3 Градиентные методы оптимизации

В данном разделе рассматриваются детали оптимизации метапараметров градиентными методами.

**Определение 3.** Назовем *оператором оптимизации* алгоритм  $U$  выбора вектора параметров  $\mathbf{w}'$  по параметрам предыдущего шага  $\mathbf{w}$ :

$$\mathbf{w}' = U(\mathbf{w}).$$

Оптимизируем параметры  $\mathbf{w}$  при помощи  $\eta$  шагов оптимизации:

$$\hat{\mathbf{w}} = U \circ U \circ \dots \circ U(\mathbf{w}_0, \boldsymbol{\lambda}) = U^\eta(\mathbf{w}_0, \boldsymbol{\lambda}), \quad (8)$$

где  $\mathbf{w}_0$  — начальное значение вектора параметров  $\mathbf{w}$ ,  $\boldsymbol{\lambda}$  — совокупность метопараметров модели.

Переопределим задачу минимизации согласно определению оператора  $U$ :

$$\hat{\boldsymbol{\lambda}} = \arg \max_{\boldsymbol{\lambda} \in \mathbb{R}^3} \mathcal{L}_{\text{val}}(U^\eta(\mathbf{w}_0, \boldsymbol{\lambda})). \quad (9)$$

Схема оптимизации метопараметров:

1. Для каждого  $i = \overline{0, l}$ , где  $l$  — число итераций, используемых для оптимизации метопараметров.
2. Решим задачу (9) и получим новое значение метопараметров  $\boldsymbol{\lambda}'$ .
3. Положим  $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$ .

Оптимизационную задачу (5) и (6) решает оператор градиентного спуска:

$$U(\mathbf{w}, \boldsymbol{\lambda}) = \mathbf{w} - \gamma \nabla \mathcal{L}_{\text{train}}(\mathbf{w}, \boldsymbol{\lambda}), \quad (10)$$

где  $\gamma$  — длина шага градиентного спуска.

Используем метод градиентного спуска, который зависит только от значений параметров  $\mathbf{w}$  на предыдущем шаге. На каждой итерации получим следующее значение метопараметров:

$$\boldsymbol{\lambda}' = \boldsymbol{\lambda} - \gamma_{\boldsymbol{\lambda}} \nabla_{\boldsymbol{\lambda}} \mathcal{L}_{\text{val}}(U(\mathbf{w}, \boldsymbol{\lambda}), \boldsymbol{\lambda}) = \boldsymbol{\lambda} - \gamma_{\boldsymbol{\lambda}} \nabla_{\boldsymbol{\lambda}} \mathcal{L}_{\text{val}}(\mathbf{w} - \gamma \nabla \mathcal{L}_{\text{train}}(\mathbf{w}, \boldsymbol{\lambda}), \boldsymbol{\lambda}). \quad (11)$$

Градиентная оптимизация является вычислительно затратной, поэтому предлагается аппроксимировать траекторию оптимизации модели. Предлагается предсказывать траекторию изменения метопараметров модели (а конкретно, их градиенты) с помощью линейных моделей через определенное число итераций:

$$\boldsymbol{\lambda}' = \boldsymbol{\lambda} + \mathbf{c}^\top \begin{pmatrix} z \\ 1 \end{pmatrix} \quad (12)$$

где  $\sigma$  — сигмоида,  $z$  — номер итерации по модулю периодичности обучения линейной модели,  $\mathbf{c}$  — коэффициенты линейного многочлена. В остальное время используются градиентные методы.

## 4 Вычислительный эксперимент

Целью эксперимента является проверка работоспособности предложенного метода дистиляции моделей, а также анализ полученных моделей и их метопараметров. Эксперимент проводится на двух выборках: синтетической модели и выборке CIFAR-10. Результаты данной работы и исходный код эксперимента опубликованы в [7] и могут быть проверены или использованы в дальнейшей работе.

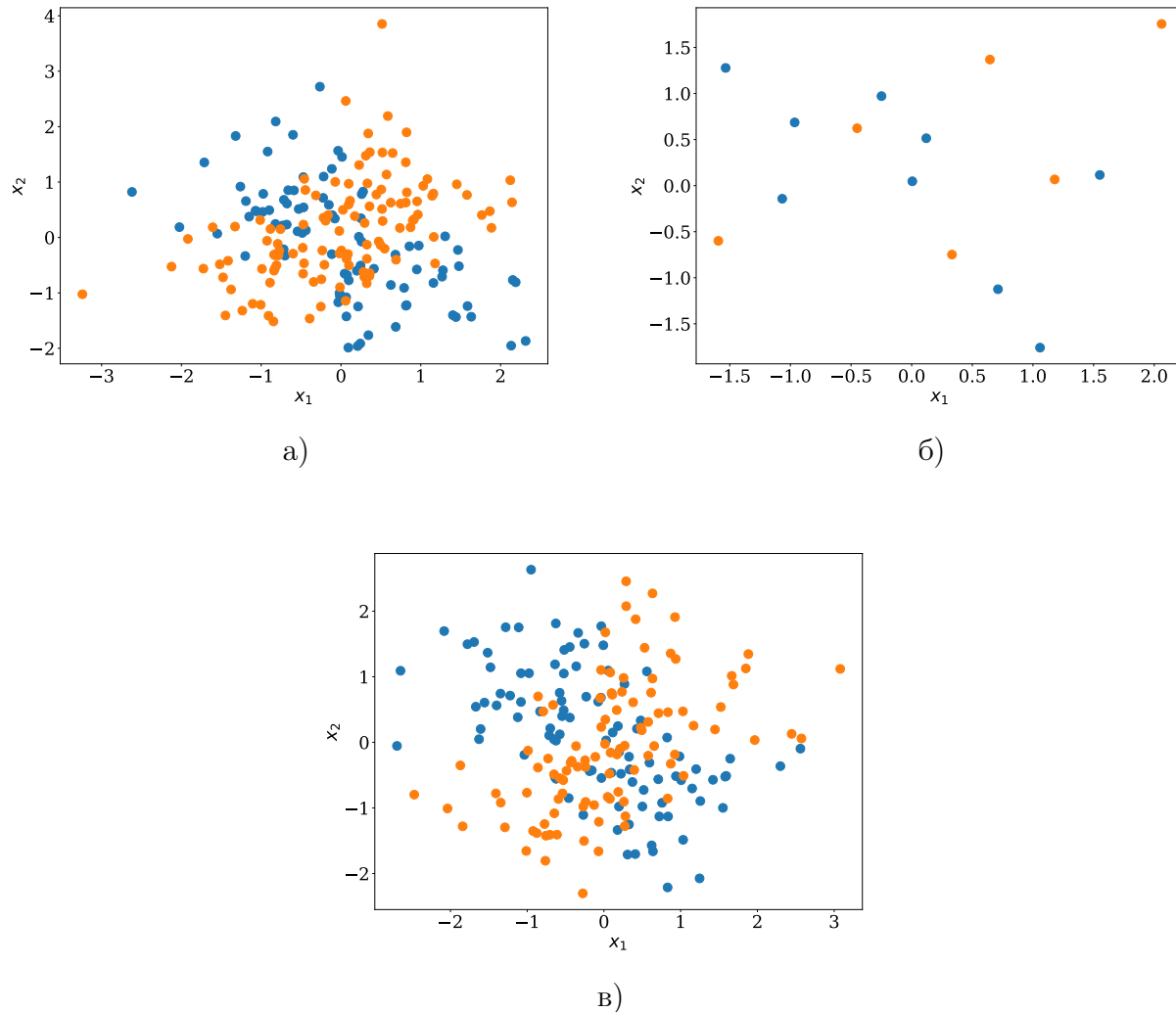
### 4.1 Эксперимент на синтетической выборке

В эксперименте используется синтетическая выборка:

$$\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^m, \quad x_{ij} \in (0, 1), \quad j = 1, 2, \quad x_{i3} = [\text{sign}(x_{i1}) + \text{sign}(x_{i2}) > 0]$$

$$y_i = \text{sign}(x_{i1} \cdot x_{i2} + \delta) \in \mathbb{Y},$$

106 где  $\delta$  — это шум. При этом размер выборки модели-ученика намного меньше размера  
 107 выборки модели-учителя.

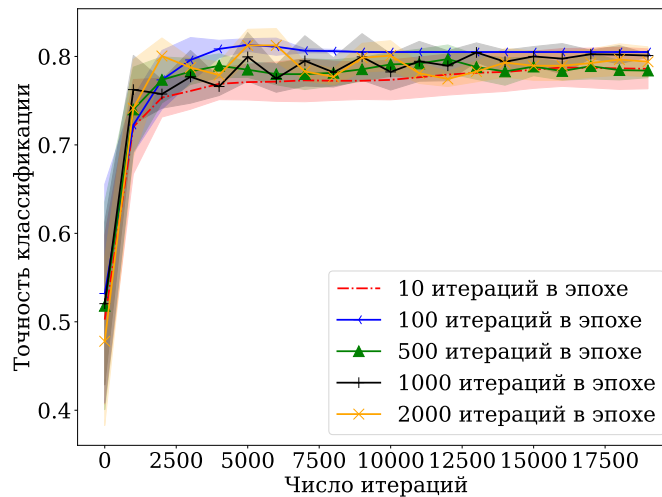


**Рис. 1** Визуализация выборки а) для обучения учителя; б) для обучения ученика; в) тестовой выборки

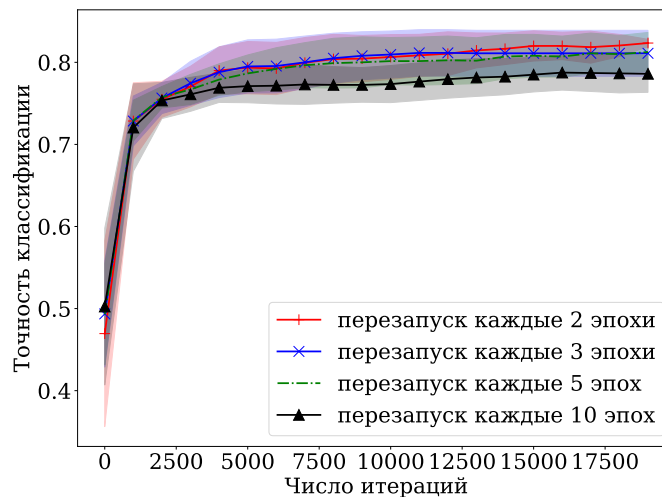
108 Обучение модели-ученика проводилось несколькими методами: с использованием ди-  
 109 стилляции и оптимизации метапараметров градиентными методами, дистилляции с пред-  
 110 сказанием траектории оптимизации модели, дистилляции со случайными метапараметра-  
 111 ми. При этом для обучения модели с использованием линейной модели дополнительно  
 112 проводились серии экспериментов для определения наилучшего размера эпохи и наилуч-  
 113 шего числа эпох между предсказаниями траектории с помощью линейной модели.

114 На рис. 2 показан график зависимости точности от номера итерации при различных  
 115 размерах эпохи. Согласно данному графику размер эпохи был выбран равным 100.

116 Пусть  $n$  — число эпох между использованием линейной модели. На рис. 3 показан  
 117 график зависимости точности от номера итерации различных  $n$ . Наилучшие результаты  
 118 достигнуты при  $n = 2$ .



**Рис. 2** График зависимости точности классификации от номера итерации при различных значениях размера эпохи

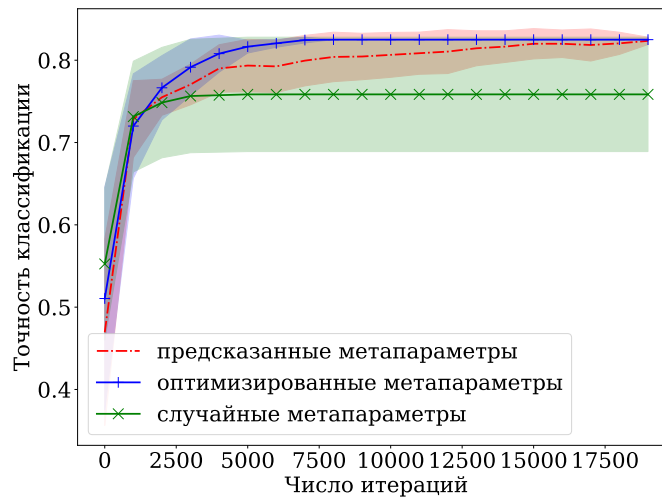


**Рис. 3** График зависимости точности классификации от номера итерации при различных  $n$

На рис. 10 показан график зависимости точности от номера итерации при различных подходах к обучению модели. Наилучшие результаты достигнуты при использовании оптимизированных метаметров, но предсказание траектории с помощью линейной модели показало результат не намного хуже предыдущего, причем с увеличением числа итераций точность этих двух методов становилась одинаковой.

## 4.2 Эксперимент на выборке CIFAR-10

В эксперименте используется выборка CIFAR-10, которая состоит из 60000 цветных изображений размера  $32 \times 32$  пикселя, разделенных на 10 непересекающихся классов. К каждому классу относится 6000 изображений. Выборка делится на обучающую (50000



**Рис. 4** График зависимости точности классификации от номера итерации

128 изображений) и тестовую (10000 изображений) подвыборки. В тестовой выборке содер-  
 129 жится 1000 изображений каждого класса.

130 Внешним критерием качества модели является точность (3). В качестве моделей-  
 131 учителей рассматриваются модели из [2], а именно, ResNet-18 и сверточная нейросеть  
 132 с тремя сверточными слоями и двумя слоями полносвязной нейросети.

133 Проведено сравнение среднего качества обучения модели-ученика без дистилляции по-  
 134 сле 5 запусков, с дистилляцией с моделью-учителем ResNet и сверточной нейросетью после  
 135 20 запусков. Значение коэффициента  $\lambda_1$  лежит в пределах от 0 до 1, значение температу-  
 136 ры — от 0.1 до 10.

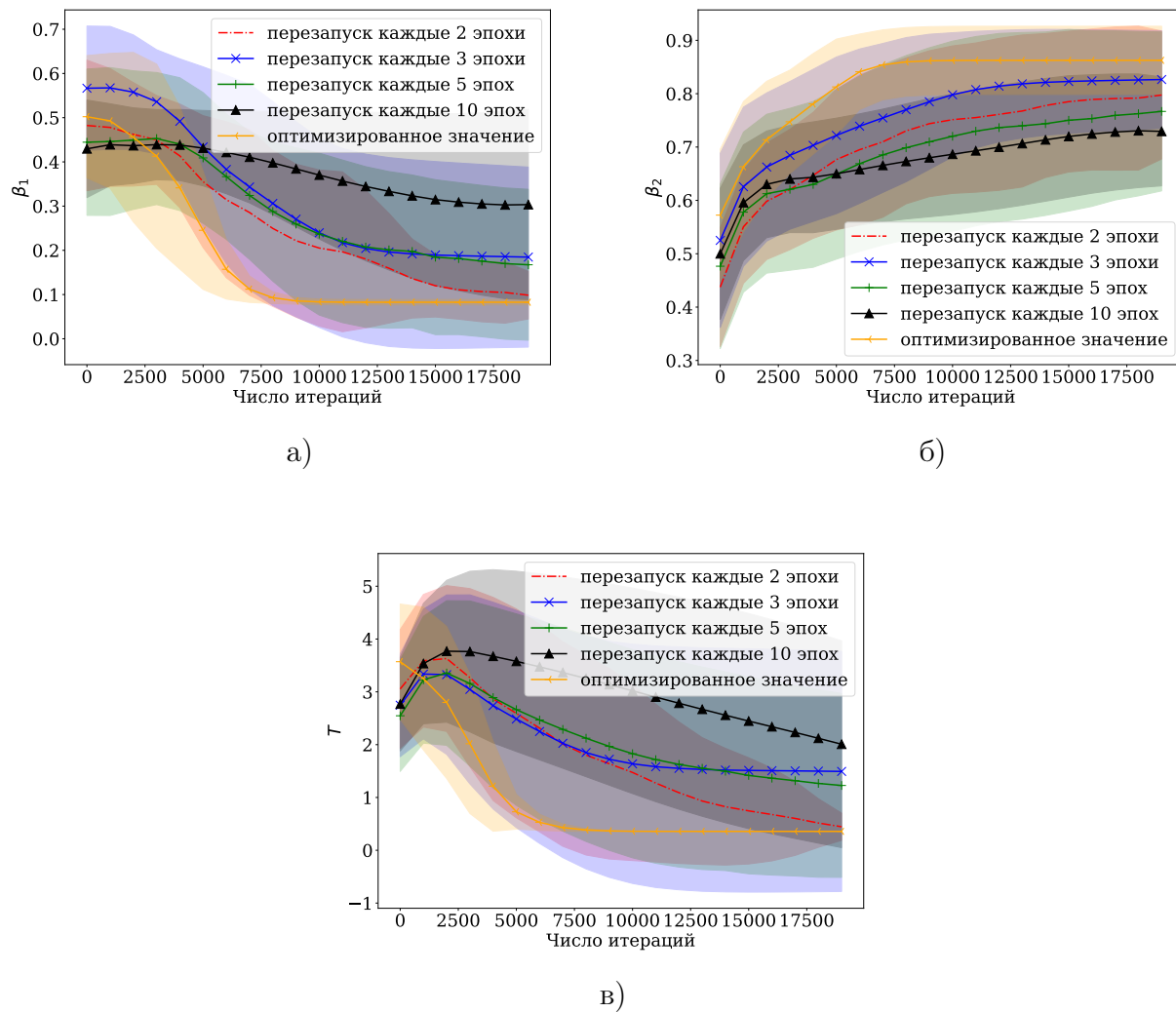
137 На рис. 6 изображена зависимость точности от величины коэффициента  $\lambda_1$ . Различ-  
 138 ные точки отвечают за точность модели без дистилляции, с дистилляцией ResNet и CNN.  
 139 Можно заметить, что с уменьшением значения коэффициента  $\lambda_1$  значение точности уве-  
 140 личивается.

141 На рис. 7 изображена зависимость точности от  $T$ . Для изображения значений темпера-  
 142 туры используется логарифмическая шкала. По графику видно, что значение температу-  
 143 ры уменьшается при увеличении логарифма температуры, но при значениях логарифма  
 144 от 0.5 до 1 наблюдается резкое уменьшение точности.

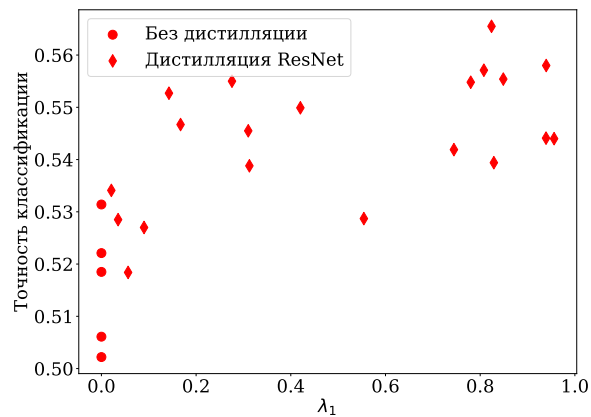
145 На рис. 8 изображена зависимость  $\lambda_1$  от величины  $T$  с выделенной цветом ассигасу.  
 146 Заметим, что точки с большим значением точности в основном расположены в правом  
 147 нижнем углу графика, а именно, при значениях  $\lambda_1$  от 0 до 0.5 и значениях  $\log(T)$  от -1  
 148 до 0. Наоборот, точки с низким значением точности, расположены в правом верхнем углу  
 149 графика.

150 На рис. 9 изображена зависимость метапараметров от числа итераций.

151 На рис. 10 изображена зависимость точности от числа эпох для обучения модели-  
 152 ученика без дистилляции, обучения с дистилляцией и случайными метапараметрами, обу-  
 153 чения с дистилляцией и оптимизацией метапараметров, а также обучения с дистилляцией  
 154 и оптимальными метапараметрами, полученными в ходе их оптимизации. Можно заме-  
 155 тить, что точность обучения с дистилляцией гораздо выше, чем без дистилляции. Также



**Рис. 5** График зависимости а)  $\lambda_1$ ; б)  $\lambda_2$ ; в) температуры от номера итерации



**Рис. 6** График зависимости точности от  $\lambda_1$



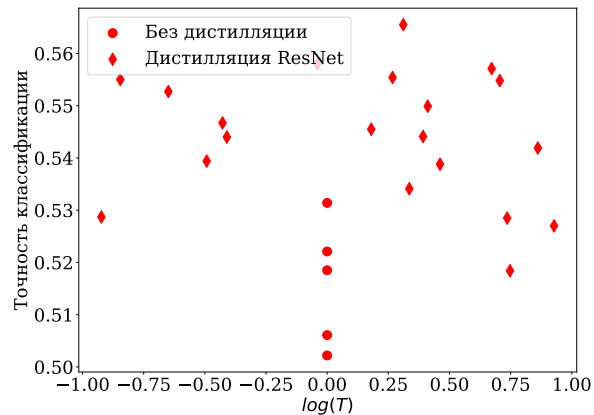
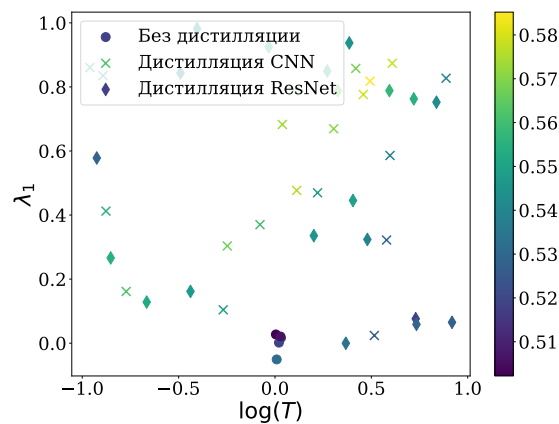


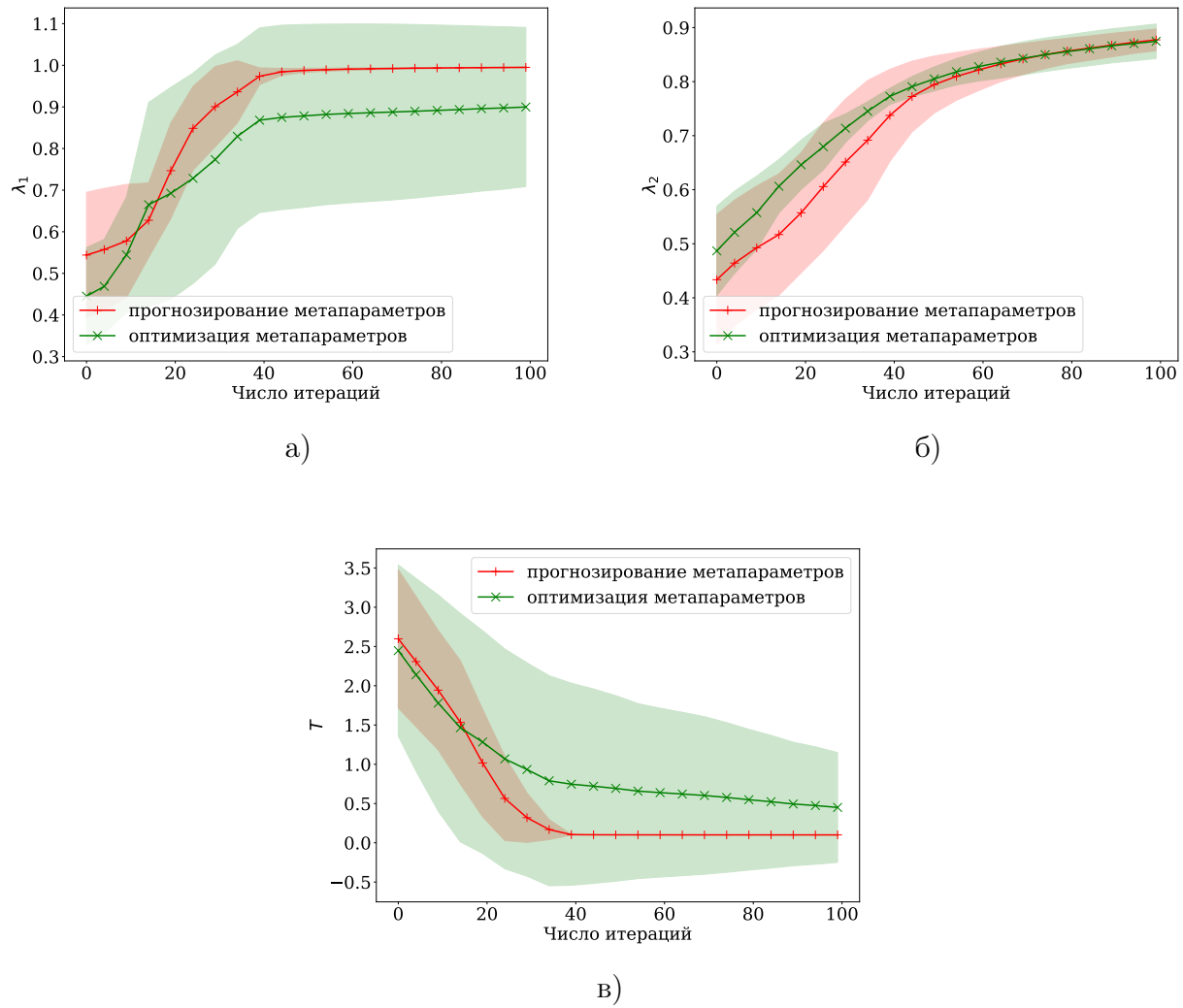
Рис. 7 График зависимости точности от температуры

Рис. 8 График зависимости  $\lambda_1$  от температуры с выделенной цветом ассигасу

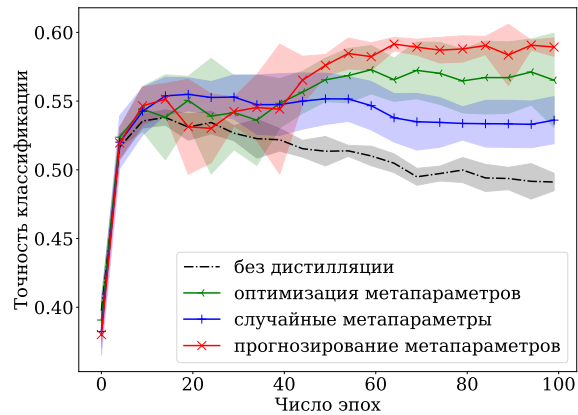
156 наибольшая точность достигается при обучении с дистилляцией и прогнозированными  
 157 метопараметрами.

## 158 5 Заключение

159 Была исследована задача оптимизации параметров модели глубокого обучения. Было  
 160 предложено обобщение методов дистилляции, заключающееся в градиентной оптимиза-  
 161 ции метопараметров. На первом уровне оптимизируются параметры модели, на втором —  
 162 метопараметры, задающие вид оптимизационной задачи. Были исследованы свойства оп-  
 163 тимизационной задачи и методы предсказания траектории оптимизации метопараметров  
 164 модели. Под метопараметрами модели понимаются параметры оптимизационной задачи  
 165 дистилляции. Предложенное обобщение позволило производить дистилляцию модели с  
 166 лучшими эксплуатационными характеристиками и за меньшее число итераций оптимиза-  
 167 ции. Комбинация данных подходов была проиллюстрирована с помощью вычислительного  
 168 эксперимента на выборке CIFAR-10 и на синтетической выборке. Вычислительный экспе-  
 169 римент показал эффективность градиентной оптимизации для задачи выбора метопара-  
 170 метров дистилляционной функции потерь. Проанализирована возможность аппроксими-  
 171 ровать траекторию оптимизации метопараметров локально-линейной моделью. Планиру-



**Рис. 9** График зависимости а)  $\lambda_1$ ; б)  $\lambda_2$ ; в) температуры от номера итерации



**Рис. 10** График зависимости точности от числа эпох

ется дальнейшее исследование оптимизационной задачи и анализ качества аппроксимации траектории оптимизации метапараметров более сложными прогностическими моделями.

## Литература

- [1] *Hinton Geoffrey E., Vinyals Oriol, Dean Jeffrey.* Distilling the knowledge in a neural network // CoRR, 2015. Vol. abs/1503.02531. URL: <http://arxiv.org/abs/1503.02531>.
- [2] *Passalis Nikolaos, Tzelepi Maria, Tefas Anastasios.* Heterogeneous knowledge distillation using information flow modeling // CVPR. — 2020. P. 2336–2345. URL: <https://ieeexplore.ieee.org/xpl/conhome/9142308/proceeding>.
- [3] *Luketina Jelena, Berglund Mathias, Greff Klaus, Raiko Tapani.* Scalable gradient-based tuning of continuous regularization hyperparameters // CoRR, 2015. Vol. abs/1511.06727. URL: <http://arxiv.org/abs/1511.06727>.
- [4] *Bakhteev Oleg Yu., Strijov Vadim V.* Comprehensive analysis of gradient-based hyperparameter optimization algorithms // Ann. Oper. Res, 2020. Vol. 289. No. 1. P. 51–65.
- [5] *Maclaurin Dougal, Duvenaud David, Adams Ryan P.* Gradient-based hyperparameter optimization through reversible learning // CoRR, 2015. Vol. abs/1502.03492. URL: <http://arxiv.org/abs/1502.03492>.
- [6] *Krizhevsky Alex et al.* Learning multiple layers of features from tiny images, 2009.
- [7] URL: <https://github.com/Intelligent-Systems-Phystech/2021-Project-84>.

*Received*

## Regularizing optimization trajectory of deep learning model parameters with knowledge distillation

*M. Gorpinich, O. Yu. Bakhteev, V. V. Strijov*

gorpinich4@gmail.com; bakhteev@phystech.edu; strijov@ccas.ru

The paper investigates parameter optimization problem for deep learning neural networks. The knowledge of a cumbersome model is considered during optimization, i.e. the knowledge distillation is used. The paper proposes generalization of knowledge distillation method to optimize meta-parameters by gradient descent. Meta-parameters are the parameters of knowledge distillation optimization problem, namely, the coefficients before terms in error function and the temperature factor. The error function is a sum of likelihood of the initial dataset and the one of distillation dataset. Temperature is a factor of logits of models in softmax function. The authors investigate the properties of optimization problem and methods to predict the optimization path of meta-parameters. Generalized method produces models with higher performance and uses less number of iterations. The algorithm is evaluated on CIFAR-10 dataset and synthetic data.

**Keywords:** *machine learning; knowledge distillation; metaparameter optimization; gradient-based optimization; metaparameter selection*

**DOI:**

## References

- [1] *Hinton Geoffrey E., Vinyals Oriol, Dean Jeffrey.* Distilling the knowledge in a neural network // CoRR, 2015. Vol. abs/1503.02531. URL: <http://arxiv.org/abs/1503.02531>.

- [2] *Passalis Nikolaos, Tzelepi Maria, Tefas Anastasios*. Heterogeneous knowledge distillation using information flow modeling // CVPR. — 2020. P. 2336–2345. URL: <https://ieeexplore.ieee.org/xpl/conhome/9142308/proceeding>.
- [3] *Luketina Jelena, Berglund Mathias, Greff Klaus, Raiko Tapani*. Scalable gradient-based tuning of continuous regularization hyperparameters // CoRR, 2015. Vol. abs/1511.06727. URL: <http://arxiv.org/abs/1511.06727>.
- [4] *Bakhteev Oleg Yu., Strijov Vadim V.* Comprehensive analysis of gradient-based hyperparameter optimization algorithms // Ann. Oper. Res, 2020. Vol. 289. No. 1. P. 51–65.
- [5] *Maclaurin Dougal, Duvenaud David, Adams Ryan P.* Gradient-based hyperparameter optimization through reversible learning // CoRR, 2015. Vol. abs/1502.03492. URL: <http://arxiv.org/abs/1502.03492>.
- [6] *Krizhevsky Alex et al.* Learning multiple layers of features from tiny images, 2009.
- [7] URL: <https://github.com/Intelligent-Systems-Phystech/2021-Project-84>.

*Received*