

Байесовский выбор субоптимальной структуры модели глубокого обучения

О. Ю. Бахтеев

Диссертация на соискание ученой степени
кандидата физико-математических наук
05.13.17 — Теоретические основы информатики
Научный руководитель: д.ф.-м.н. В.В. Стрижов

Московский физико-технический институт
16 июня 2019 г.

Выбор структуры модели глубокого обучения

Цель работы

Предложить метод выбора структуры модели глубокого обучения.

Задачи

- 1 Предложить критерии оптимальной и субоптимальной сложности модели глубокого обучения.
- 2 Предложить алгоритм построения модели субоптимальной сложности и оптимизации параметров.

Исследуемые проблемы

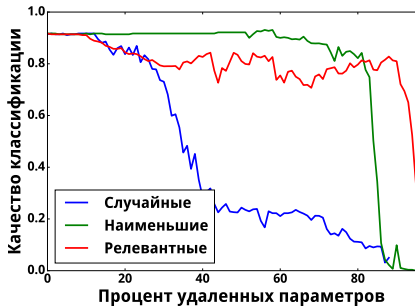
- 1 Большое число параметров и гиперпараметров модели.
- 2 Многоэкстремальность и невыпуклость задачи оптимизации параметров и гиперпараметров модели.
- 3 Высокая вычислительная сложность оптимизации.

Методы исследования

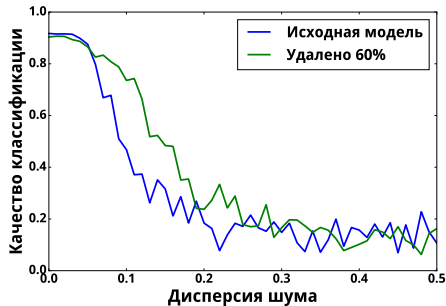
Рассматриваются графовое представление нейронной сети. Используются методы вариационного байесовского вывода. Для получения модели субоптимальной сложности используется метод автоматического определения релевантности параметров с использованием градиентных методов оптимизации гиперпараметров и структурных параметров модели.

Проблема выбора оптимальной структуры модели глубокого обучения

Правдоподобие моделей с избыточным числом параметров не меняется при их удалении.



Избыточность параметров модели



Устойчивость модели

Глубокое обучение предполагает оптимизацию моделей с заведомо избыточной сложностью.

Глубокое обучение

Определение

Моделью $f(\mathbf{w}, \mathbf{x})$ назовем дифференцируемую по параметрам \mathbf{w} функцию из множества признаков описаний объекта во множество меток:

$$f : \mathbb{X} \times \mathbb{W} \rightarrow \mathbb{Y},$$

где \mathbb{W} — пространство параметров функции f .

Особенность задачи выбора модели *глубокого обучения* — значительное число параметров моделей приводит к неприменимости стандартных методов оптимизации и выбора модели.

Эмпирические оценки сложности модели:

- ① число параметров;
- ② число суперпозиций внутри модели.

Выбор структуры: двуслойная нейросеть

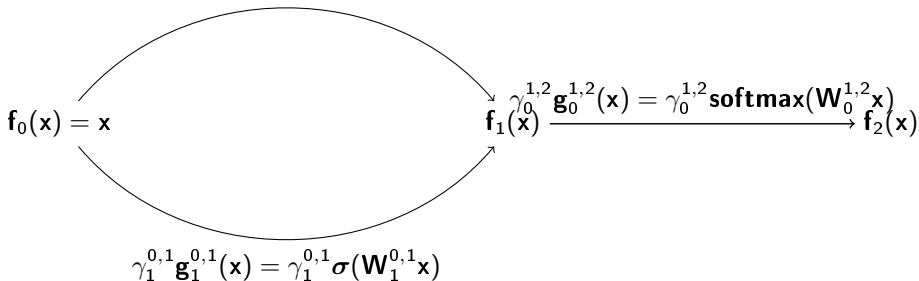
Структурные параметры: $\Gamma = [\gamma^{0,1}, \gamma^{1,2}]$.

Модель: $\mathbf{f}(\mathbf{x}) = \text{softmax}(\mathbf{W}_0^{1,2} \mathbf{f}_1(\mathbf{x}))$, $\mathbf{f}(\mathbf{x}) : \mathbb{R}^n \rightarrow [0, 1]^{|\mathbb{Y}|}$, $\mathbf{x} \in \mathbb{R}^n$.

$$\mathbf{f}_1(\mathbf{x}) = \gamma_0^{0,1} \mathbf{g}_0^{0,1}(\mathbf{x}) + \gamma_1^{0,1} \mathbf{g}_1^{0,1}(\mathbf{x}),$$

где $\mathbf{W} = [\mathbf{W}_0^{0,1}, \mathbf{W}_1^{0,1}, \mathbf{W}_0^{1,2}]^T$ — матрицы параметров, $\{\mathbf{g}_{0,1}^0, \mathbf{g}_{0,1}^1, \mathbf{g}_{1,2}^0\}$ — обобщенно-линейные функции скрытых слоев нейросети.

$$\gamma_0^{0,1} \mathbf{g}_0^{0,1}(\mathbf{x}) = \gamma_0^{0,1} \sigma(\mathbf{W}_0^{0,1} \mathbf{x})$$



Графовое представление модели глубокого обучения

Определение

Пусть

- ① задан ациклический граф (V, E) ;
- ② для каждого ребра $(j, k) \in E$ определен вектор базовых функций $\mathbf{g}^{j,k} = [\mathbf{g}_0^{j,k}, \dots, \mathbf{g}_{K^{j,k}}^{j,k}]$ мощности $K^{j,k}$;
- ③ для каждой вершины $v \in V$ определена функция агрегации \mathbf{agg}_v .

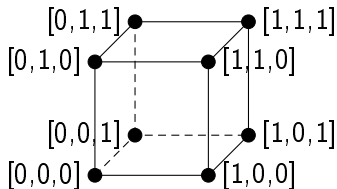
Граф (V, E) со множеством векторов базовых функций $\{\mathbf{g}^{j,k}, (j, k) \in E\}$ и функций агрегаций $\{\mathbf{agg}_v, v \in V\}$ задает *параметрическое семейство моделей* \mathfrak{F} тогда и только тогда, когда функция $\mathbf{f} = \mathbf{f}_{|V|-1}$, задаваемая по правилу

$$\mathbf{f}_v(\mathbf{w}, \mathbf{x}) = \mathbf{agg}_v \left(\{ \langle \gamma^{j,k}, \mathbf{g}^{j,k} \rangle \circ \mathbf{f}_j(\mathbf{x}) \mid j \in \text{Adj}(v_k) \} \right), v \in \{1, \dots, |V| - 1\}, \quad \mathbf{f}_0(\mathbf{x}) = \mathbf{x} \quad (1)$$

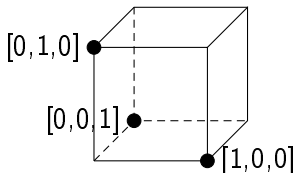
является дифференцируемой по параметрам \mathbf{w} функцией из признакового пространства \mathbb{X} в пространство меток \mathbb{Y} при значениях векторов, $\gamma^{j,k} \in [0, 1]^{K^{j,k}}$.

Ограничения на структурные параметры

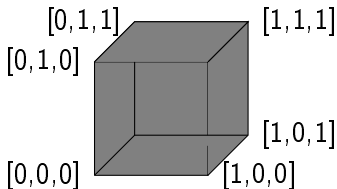
Примеры ограничений для одного структурного параметра γ , $|\gamma| = 3$.



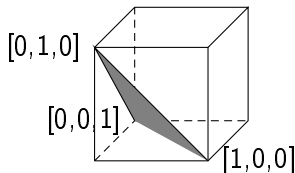
На вершинах куба



На вершинах симплекса



Внутри куба

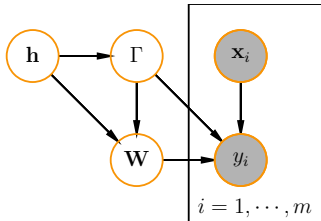


Внутри симплекса

Априорное распределение параметров

Определение

Априорным распределением параметров \mathbf{W} и структуры $\mathbf{\Gamma}$ модели \mathbf{f} назовем вероятностное распределение $p(\mathbf{W}, \mathbf{\Gamma} | \mathbf{h}) : \mathbb{W} \times \mathbb{\Gamma} \times \mathbb{H} \rightarrow \mathbb{R}^+$, где \mathbb{W} — множество значений параметров модели, $\mathbb{\Gamma}$ — множество значений структуры модели.



Определение

Гиперпараметрами $\mathbf{h} \in \mathbb{H}$ модели назовем параметры распределения $p(\mathbf{w}, \mathbf{\Gamma} | \mathbf{h})$ (параметры распределения параметров модели \mathbf{f}).

Правдоподобие как статистическая сложность

Статистическая сложность модели f :

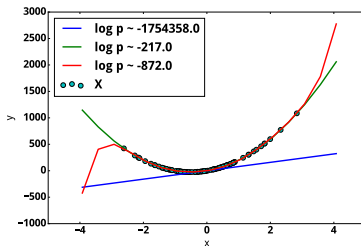
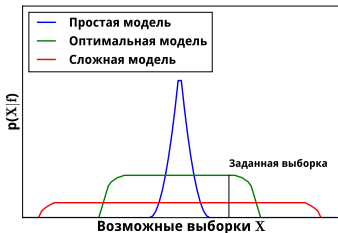
$$\text{MDL}(\mathbf{y}, \mathbf{f}) = -\log p(\mathbf{h}) - \log (p(\mathbf{y}|\mathbf{X}, \mathbf{h})\delta\mathfrak{D}),$$

где $\delta\mathfrak{D}$ — допустимая точность передачи информации о выборке \mathfrak{D} .

Правдоподобие модели:

$$Q(\mathbf{h}|\theta^*, \mathbf{X}, \mathbf{y}) = \log p(\mathbf{h}|\mathbf{X}, \mathbf{y}) = \log p(\mathbf{h}) + \log \int_{\mathbf{W}, \Gamma} p(\mathbf{y}|\mathbf{X}, \mathbf{W}, \Gamma) p(\mathbf{W}, \Gamma|\mathbf{h}) d\mathbf{W} d\Gamma,$$

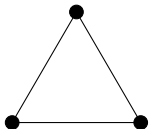
$$L(\theta|\mathbf{h}, \mathbf{X}, \mathbf{y}) = \log p(\mathbf{W}, \Gamma|\mathbf{X}, \mathbf{y}, \mathbf{h}) \propto \log p(\mathbf{y}|\mathbf{X}, \mathbf{W}, \Gamma, \mathbf{h}) + \log p(\mathbf{W}, \Gamma|\mathbf{h}).$$



Выбор модели по правдоподобию Аппроксимация выборки полиномами

Априорное распределение на структуре модели

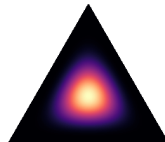
Распределение Дирихле: $\Gamma \sim \text{Dir}(\mathbf{s}, c_{\text{temp}})$



$c_{\text{temp}} \rightarrow 0$

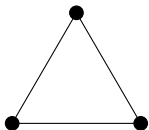


$c_{\text{temp}} = 0.995$



$c_{\text{temp}} = 5.0$

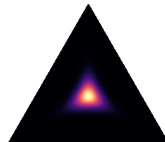
Распределение Гумбель-софтмакс: $\Gamma \sim \text{GS}(\mathbf{s}, c_{\text{temp}})$



$c_{\text{temp}} \rightarrow 0$



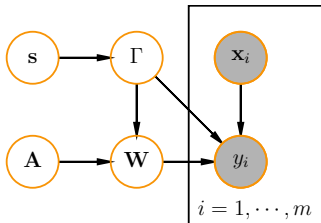
$c_{\text{temp}} = 0.995$



$c_{\text{temp}} = 5.0$

Байесовский выбор модели

- Параметры модели:
 $w_i \sim \mathcal{N}(0, \gamma_{(i)} \alpha_i),$
 $\alpha \sim \text{inv-gamma}(c_1, c_2).$
- Структурные параметры модели:
 $\Gamma = \{\gamma_i\}. \gamma \sim \text{GS}(\mathbf{s}, c_{\text{temp}}).$
- Гиперпараметры модели: $\mathbf{h} = \{\mathbf{A}, \mathbf{s}\}.$
- Метапараметры: $c_1, c_2, c_{\text{temp}}.$



Основные проблемы выбора оптимальной модели

- Интеграл правдоподобия $p(\mathbf{y}|\mathbf{X}, \mathbf{h})$ невычислим аналитически.
- Задача его оптимизации многоэкстремальна и невыпукла.

Вариационная нижняя оценка правдоподобия

Интеграл правдоподобия невычислим аналитически.

Правдоподобие модели:

$$p(y|\mathbf{X}) = \int_{\mathbf{W}, \Gamma} p(y|\mathbf{X}, \mathbf{W}, \Gamma) p(\mathbf{W}, \Gamma) d\mathbf{W} d\Gamma.$$

Определение

Вариационными параметрами модели $\theta \in \mathbb{R}^u$ назовем параметры распределения q , приближающие апостериорное распределение параметров и структур $p(\mathbf{W}, \Gamma | \mathbf{X}, y, \mathbf{h})$:

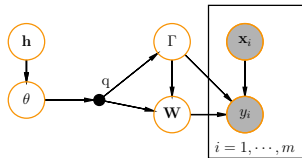
$$q \approx \frac{p(y|\mathbf{X}, \mathbf{W}, \Gamma, \mathbf{h}) p(\mathbf{W}, \Gamma | \mathbf{h})}{\iint_{\mathbf{W}', \Gamma'} p(y|\mathbf{X}, \mathbf{W}', \Gamma', \mathbf{h}) p(\mathbf{W}', \Gamma' | \mathbf{h}) d\mathbf{W}' d\Gamma'}.$$

Получим нижнюю оценку интеграла правдоподобия.

$$\log p(y|\mathbf{X}) \geq \mathbb{E}_q \log p(y|\mathbf{X}, \mathbf{W}, \Gamma) - D_{\text{KL}}(p(\mathbf{w}, \Gamma) \| q(\mathbf{W}, \Gamma)) = \log \hat{p}_{q_{\mathbf{W}} q_{\Gamma}}(y|\mathbf{X}).$$

Полученная оценка совпадает с интегралом правдоподобия при

$$D_{\text{KL}}(q(\mathbf{W}, \Gamma) \| (p(\mathbf{W}, \Gamma | y, \mathbf{X}))) = 0.$$



Задача выбора модели

Зададим вариационное распределение q с параметрами θ , приближающие апостериорное распределение $p(\mathbf{W}, \Gamma | \mathbf{X}, \mathbf{y}, \mathbf{h})$ параметров и структуры.

Определение

Функцией потерь $L(\theta | \mathbf{h}, \mathbf{X}, \mathbf{y})$ назовем дифференцируемую функцию, качество модели на обучающей выборке при параметрах θ распределения q .

Функцией валидации $Q(\mathbf{h} | \theta, \mathbf{X}, \mathbf{y})$ назовем дифференцируемую функцию, качество модели при векторе θ , заданном неявно.

Задачей выбора модели \mathbf{f} назовем решение двухуровневой задачи оптимизации:

$$\mathbf{h}^* = \arg \min_{\mathbf{h} \in \mathbb{H}} Q(\mathbf{h} | \theta^*, \mathbf{X}, \mathbf{y}),$$

где θ^* — решение задачи оптимизации

$$\theta^* = \arg \min_{\theta \in \mathbb{R}^u} L(\theta | \mathbf{h}, \mathbf{X}, \mathbf{y}).$$

Выбор оптимальной модели

Требуется

Предложить метод поиска субоптимального решения задачи оптимизации, обобщающего различные алгоритмы оптимизации:

- Оптимизация правдоподобия.
- Последовательное увеличение и снижение сложности модели.
- Полный перебор вариантов структуры модели.

Общая задача оптимизации

Теорема (будет)

Следующая оптимизационная задача обобщает алгоритмы оптимизации: оптимизация правдоподобия, последовательное увеличение и снижение сложности модели, полный перебор вариантов структуры модели:

$$\begin{aligned} \mathbf{h}^* &= \arg \max_{\mathbf{h}} Q = & (Q^*) \\ &= c_{\text{train}} E_{q^*} \log p(\mathbf{y} | \mathbf{X}, \mathbf{W}, \mathbf{\Gamma}, \mathbf{h}, c_{\text{prior}}) - \\ &- c_{\text{prior}} D_{KL}(p(\mathbf{w}, \mathbf{\Gamma} | \mathbf{h}, c_{\text{temp}}) || q^*(\mathbf{W}, \mathbf{\Gamma})) - \\ &- c_{\text{comb}} \sum_{p' \in \mathcal{P}} D_{KL}(\mathbf{\Gamma} | p'), \end{aligned}$$

где

$$\begin{aligned} q^* &= \arg \max_q L = E_q \log p(\mathbf{y} | \mathbf{X}, \mathbf{W}, \mathbf{\Gamma}, \mathbf{A}^{-1}, c_{\text{temp}}) & (L^*) \\ &- c_{\text{reg}} D_{KL}(p(\mathbf{w}, \mathbf{\Gamma} | \mathbf{A}^{-1}, \mathbf{m}, c_{\text{temp}}) || q(\mathbf{W}), q(\mathbf{\Gamma})). \end{aligned}$$

$c_{\text{train}}, c_{\text{prior}}, c_{\text{temp}}, c_{\text{comb}}$ — метапараметры оптимизации.

Адекватность задачи оптимизации

Теорема

Пусть задано параметрическое множество вариационных распределений: $q \in \Omega$. Пусть $c_{\text{train}} = c_{\text{prior}} = c_{\text{reg}} > 1$, $c_{\text{comb}} = 0$. Тогда:

- 1 Задача оптимизации (??) доставляет максимум вариационной оценки правдоподобия при фиксированном параметре c_{temp} :
$$\log p(y|X, h, c_{\text{temp}}) \rightarrow \max_h .$$
- 2 Вариационное распределение q приближает апостериорное распределение $p(W, \Gamma|y, X, h, c_{\text{temp}})$ наилучшим образом среди множества распределений Ω :

$$D_{\text{KL}}(q||p(W, \Gamma|y, X, h), c_{\text{temp}}) \rightarrow \min_{q \in \Omega} .$$

Пусть также распределение q декомпозируется на два независимых распределения для параметров W и структуры Γ модели f :

$$q = q_W q_\Gamma, q_\Gamma \approx p(\Gamma|y, X, h), q_W \approx p(W|\Gamma, y, X, h).$$

Тогда вариационные распределения q_W, q_Γ приближают апостериорные распределения $p(\Gamma|y, X, h, c_{\text{temp}}), p(W|\Gamma, y, X, h, c_{\text{temp}})$ наилучшим образом:

$$D_{\text{KL}}(q_\Gamma||p(\Gamma|y, X, h, c_{\text{temp}})) \rightarrow \min, \quad D_{\text{KL}}(q_W||p(W|\Gamma, y, X, h, c_{\text{temp}})) \rightarrow \min .$$

Оператор оптимизации

Определение

Назовем *оператором оптимизации* алгоритм T выбора вектора параметров θ' по параметрам предыдущего шага θ .

Оператор стохастического градиентного спуска:

$$\begin{aligned}\hat{\theta} &= T \circ T \circ \dots \circ T(\theta_0, \mathbf{A}^{-1}, \mathbf{m}) = T^\eta(\theta_0, \mathbf{A}^{-1}, \mathbf{m}), \quad \text{где } T(\theta, \mathbf{A}^{-1}, \mathbf{m}) = \\ &= \theta - \beta \nabla L(\theta, \mathbf{A}^{-1}, \mathbf{m})|_{\hat{\mathcal{D}}},\end{aligned}$$

γ — длина шага градиентного спуска, θ_0 — начальное значение параметров θ , $\hat{\mathcal{D}}$ — случайная подвыборка исходной выборки \mathcal{D} .

Перепишем итоговую задачу оптимизации:

$$\mathbf{h}^* = \arg \max_{\mathbf{h}} Q \left(T^\eta(\theta_0, \mathbf{A}^{-1}, \mathbf{m}) \right),$$

где θ_0 — начальное значение θ .

Нижняя вариационная оценка правдоподобия на основе мультистарта

$$\log p(\mathbf{y}|\mathbf{X}, \mathbf{h}) \geq \mathbb{E}_{q(\mathbf{w})} \log p(\mathbf{y}, \mathbf{W}|\mathbf{X}, \mathbf{h}) - \mathbb{E}_{q(\mathbf{w})} (-\log(q_{\mathbf{w}})).$$

Теорема [Бахтеев, 2016]

Пусть L — функция потерь, градиент которой — непрерывно-дифференцируемая функция с константой Липшица C .

Пусть $\theta = [\mathbf{W}^1, \dots, \mathbf{W}^k]$ — начальные приближения оптимизации модели, β — шаг градиентного спуска.

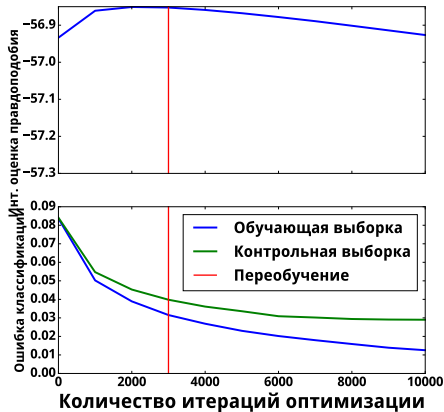
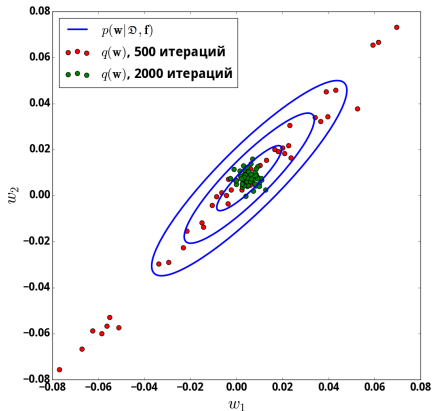
Тогда разность энтропий на смежных шагах оптимизации приближается следующим образом:

$$\mathbb{E}_{q_{\mathbf{W}}^{\tau}} (-\log(q_{\mathbf{W}}^{\tau})) - \mathbb{E}_{q_{\mathbf{W}}^{\tau-1}} (-\log(q_{\mathbf{W}}^{\tau-1})) \approx \frac{1}{k} \sum_{r=1}^k (\beta \text{Tr}[\mathbf{H}(\mathbf{W}^r)] - \beta^2 \text{Tr}[\mathbf{H}(\mathbf{W}^r) \mathbf{H}(\mathbf{w}^r)]),$$

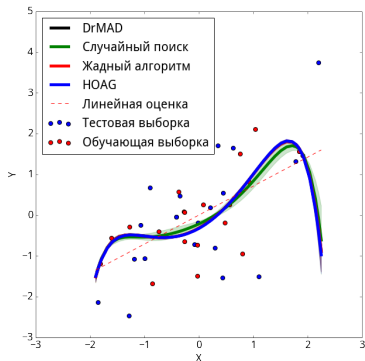
где \mathbf{H} — гессиан функции потерь L , $q_{\mathbf{W}}^{\tau}$ — распределение $q_{\mathbf{W}}^{\tau}$ в момент оптимизации τ .

Градиентный спуск как вариационная оценка правдоподобия модели

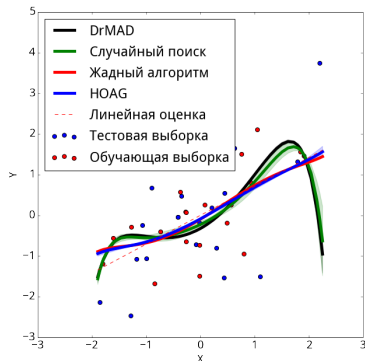
Для вычисления правдоподобия был предложен ряд алгоритмов, основанных на стохастическом градиентном спуске.



Оптимизация гиперпараметров: пример



Кросс-Валидация



Вариационная оценка

Оптимизация правдоподобия модели

Теорема.

TODO

Пусть $c_{\text{reg}} > 0$. Тогда оптимизация функции

$$L = E_q \log p(y|\mathbf{X}, \mathbf{W}, \mathbf{\Gamma}, \mathbf{A}^{-1}, c_{\text{temp}}) - \\ - c_{\text{reg}} D_{KL}(p(\mathbf{w}, \mathbf{\Gamma}|\mathbf{A}^{-1}, \mathbf{m}, c_{\text{temp}}) || q(\mathbf{W}), q(\mathbf{\Gamma}))$$

эквивалентна минимизации ожидаемой дивергенции $E_{\hat{\mathbf{X}}, \hat{\mathbf{y}} \sim p(\mathbf{X}, y)} D_{KL}(q || p(\mathbf{w}, \mathbf{\Gamma}|\hat{\mathbf{X}}, \hat{\mathbf{y}}))$, где $\hat{\mathbf{X}}, \hat{\mathbf{y}}$ — случайные подвыборки мощностью $\frac{m}{c_{\text{reg}}}$ из генеральной совокупности при $n \rightarrow \infty$.

Теорема [Бахтеев, 2018].

Пусит $p(\mathbf{W}, \mathbf{\Gamma}|\mathbf{y}, \mathbf{X}, \mathbf{h}, c_{\text{temp}}) \in \mathfrak{Q}$.

Пусть также $c_{\text{reg}} = c_{\text{train}} = c_{\text{prior}} > 0, c_{\text{comb}} = 0$, Тогда двухуровневая задача оптимизация эквивалентна задаче оптимизации правдоподобия модели при фиксированной температуре:

$$\arg \min_{\mathbf{h}} p(\mathbf{y}|\mathbf{X}, c_{\text{temp}}).$$

Параметрическая сложность

Обозначим за $F(c_{\text{reg}}, c_{\text{train}}, c_{\text{prior}}, c_{\text{comb}}, \mathbf{P}, c_{\text{temp}})$ множество экстремумов функции L при решении задачи двухуровневой оптимизации.

Теорема [Бахтеев, 2018].

Пусть $\mathbf{f} \in F(1, 1, c_{\text{prior}}, 0, \emptyset, c_{\text{temp}})$. При устремлении c_{prior} к бесконечности параметрическая сложность модели \mathbf{f} устремляется к нулю.

$$\lim_{c_{\text{prior}} \rightarrow \infty} C_{\text{param}}(\mathbf{f}) = 0.$$

Теорема [Бахтеев, 2018].

Пусть $\mathbf{f}_1 \in F(1, 1, c_{\text{prior}}^1, 0, \emptyset, c_{\text{temp}})$, $\mathbf{f}_2 \in F(1, 1, c_{\text{prior}}^2, 0, \emptyset, c_{\text{temp}})$, $c_{\text{prior}}^1 < c_{\text{prior}}^2$. Пусть вариационные параметры моделей \mathbf{f}_1 и \mathbf{f}_2 лежат в области U , в которой соответствующие функции L и Q являются локально-выпуклыми. Тогда модель \mathbf{f}_1 имеет параметрическую сложность, не меньшую чем у \mathbf{f}_2 .

$$C_{\text{param}}(\mathbf{f}_1) \geq C_{\text{param}}(\mathbf{f}_2).$$

Структурная сложность

Теорема [Бахтеев, 2018].

Пусть для каждого ребра (i, j) семейства моделей \mathcal{F} априорное распределение

$$p(\gamma_{i,j}) = \lim_{c_{\text{temp}} \rightarrow 0} \mathcal{GS}(c_{\text{temp}}).$$

Пусть $c_{\text{reg}} > 0$, $c_{\text{train}} > 0$, $c_{\text{prior}} > 0$. Пусть $\mathbf{f} \in F(c_{\text{reg}}, c_{\text{train}}, c_{\text{prior}}, 0, \emptyset, c_{\text{temp}})$. Тогда структурная сложность модели \mathbf{f} равняется нулю.

$$C_{\text{struct}}(\mathbf{f}) = 0.$$

Теорема [Бахтеев, 2018].

Пусть $\mathbf{f}_1 \in F(c_{\text{reg}}, c_{\text{train}}, c_{\text{prior}}, 0, \emptyset, c_{\text{temp}}^1)$, $\mathbf{f}_2 \in \lim_{c_{\text{temp}}^2 \rightarrow \infty} F(c_{\text{reg}}, c_{\text{train}}, c_{\text{prior}}, 0, \emptyset, c_{\text{temp}}^2)$. Пусть вариационные параметры моделей \mathbf{f}_1 и \mathbf{f}_2 лежат в области U , в которой соответствующие функции L и Q являются локально-выпуклыми. Тогда разница структурных сложностей моделей ограничена выражением:

$$C_{\text{struct}}(\mathbf{f}_1) - C_{\text{struct}}(\mathbf{f}_2) \leq E_q^1 \log p(\mathbf{y}|\mathbf{X}, \mathbf{W}, \mathbf{\Gamma}, \mathbf{A}^{-1}, c_{\text{temp}}^1) - E_q^2 \log p(\mathbf{y}|\mathbf{X}, \mathbf{W}, \mathbf{\Gamma}, \mathbf{A}^{-1}).$$

Полный перебор

Пусть для каждого ребра (i, j) семейства моделей \mathfrak{F} априорное распределение

$$p(\gamma_{i,j}) = \lim_{c_{\text{temp}} \rightarrow 0} \mathcal{GS}(c_{\text{temp}}).$$

Рассмотрим последовательность \mathbf{P} , состоящую из $N = \prod_{(j,k) \in E} K_{j,k}$ моделей, полученных в ходе оптимизаций вида:

$$f_1 \in F(c_{\text{reg}}, 0, 0, \emptyset, c_{\text{comb}}, c_{\text{temp}}),$$

$$f_2 \in F(c_{\text{reg}}, 0, 0, \{q_1(\Gamma)\}, c_{\text{comb}}, c_{\text{temp}}),$$

$$f_3 \in F(c_{\text{reg}}, 0, 0, \{q_1(\Gamma), q_2(\Gamma)\}, c_{\text{comb}}, c_{\text{temp}}),$$

где $c_{\text{reg}} > 0, c_{\text{comb}} > 0$.

Теорема

Вариационные распределения q_{Γ} структур последовательности \mathbf{P} вырождаются в распределения вида $\delta(\hat{\mathbf{m}})$, где $\hat{\mathbf{m}}$ — точка на декартовом произведении вершин симплексов структуры модели.

Последовательность соответствует полному перебору структуры Γ .

Результаты, выносимые на защиту

- ① Предложен метод выбора модели наиболее правдоподобной структуры, обобщающий ранее описанные алгоритмы оптимизации:
 - ▶ оптимизация правдоподобия;
 - ▶ последовательное увеличение сложности модели;
 - ▶ последовательное снижение сложности модели;
 - ▶ полный перебор вариантов структуры модели.
- ② Предложен алгоритм оптимизации параметров, гиперпараметров и структурных параметров моделей глубокого обучения.
- ③ Проведено исследование свойств алгоритмов выбора модели при различных значениях мета-параметров.
- ④ Проведен вычислительный эксперимент, иллюстрирующий работу предложенного метода.

Список работ автора по теме диссертации

Публикации ВАК

- 1 Бахтеев О.Ю., Попова М.С., Стрижов В.В. Системы и средства глубокого обучения в задачах классификации. // Системы и средства информатики. 2016. № 26.2. С. 4-22.
- 2 Бахтеев О.Ю., Стрижов В.В. Выбор моделей глубокого обучения субоптимальной сложности. // Автоматика и телемеханика. 2018. №8. С. 129-147.
- 3 Огальцов А.В., Бахтеев О.Ю. Автоматическое извлечение метаданных из научных PDF-документов. // Информатика и её применения. 2018.
- 4 Смердов А.Н., Бахтеев О.Ю., Стрижов В.В. Выбор оптимальной модели рекуррентной сети в задачах поиска парафраза. // Информатика и ее применения. 2019.
- 5 Грабовой А.В., Бахтеев О.Ю., Стрижов В.В. Определение релевантности параметров нейросети. // Информатика и её применения. 2019.

Выступления с докладом

- 1 “Восстановление панельной матрицы и ранжирующей модели в разнородных шкалах”, Всероссийская конференция «57-я научная конференция МФТИ», 2014.
- 2 “A monolingual approach to detection of text reuse in Russian-English collection”, Международная конференция «Artificial Intelligence and Natural Language Conference», 2015.
- 3 “Выбор модели глубокого обучения субоптимальной сложности с использованием вариационной оценки правдоподобия”, Международная конференция «Интеллектуализация обработки информации», 2016.
- 4 “Author Masking using Sequence-to-Sequence Models”, Международная конференция «Conference and Labs of the Evaluation Forum», 2017.
- 5 “Градиентные методы оптимизации гиперпараметров моделей глубокого обучения”, Всероссийская конференция «Математические методы распознавания образов ММРО», 2017.
- 6 “Детектирование переводных заимствований в текстах научных статей из журналов, входящих в РИНЦ”, Всероссийская конференция «Математические методы распознавания образов ММРО», 2017.
- 7 “Байесовский выбор наиболее правдоподобной структуры модели глубокого обучения”, Международная конференция «Интеллектуализация обработки информации», 2018.