

# Поиск согласованных нейросетевых моделей в задаче мультидоменного обучения

К. Д. Яковлев<sup>1</sup>      О. Ю. Бахтеев<sup>1,2</sup>    В. В. Стрижов<sup>1,2</sup>  
{iakovlev.kd, bakhteev, strijov}@phystech.edu

<sup>1</sup>Москва, Московский физико-технический институт

<sup>2</sup>Москва, Вычислительный центр им. А.А. Дородницына ФИЦ ИУ РАН

2022

# Цель исследования

## Цель

Предложить метод поиска архитектуры модели глубокого обучения в задаче мультидоменного обучения.

## Проблема

Модели, не учитывающие разделение выборки на домены, имеют низкую обобщающую способность.

## Метод решения

Предлагаемый метод основан на построении мультимодели. Для каждого домена оптимизируется отдельная структура. Также предлагаются два метода регуляризации: структурная и регуляризация пространства скрытых представлений модели.

-  Hanxiao Liu and Karen Simonyan and Yiming Yang. *DARTS: Differentiable Architecture Search*. CoRR, 2018.
-  Wang, Q., Ke, J., Greaves, J., Chu, G., Bender, G., Sbaiz, L., Go, A., Howard, A., Yang, M., Gilbert, J. & Others *Multi-path neural networks for on-device multi-domain visual classification*. CoRR, 2021.
-  Yakovlev, K., Grebenkova, O., Bakhteev, O. & Strijov, V. *Neural Architecture Search with Structure Complexity Control*. CoRR, 2022.

## Постановка задачи поиска архитектуры

- Архитектура модели представляет собой ориентированный ациклический граф. Каждому ребру ставится в соответствие отображение  $\mathbf{g}^{(i,j)}$ , причем

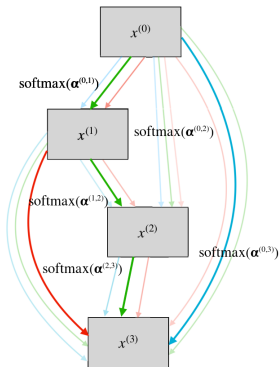
$$\mathbf{x}^{(j)} = \sum_{i < j} \mathbf{g}^{(i,j)}(\mathbf{x}^{(i)}).$$

- Пусть вектор  $\vec{\mathbf{g}}^{(i,j)}$  – вектор, составленный из доступных для ребра  $(i,j)$  отображений. Пусть вектор  $\alpha^{(i,j)}$  – вектор структурных параметров. Смешанная операция

$$\hat{\mathbf{g}}^{(i,j)}(\mathbf{x}^{(i)}) = \langle \mathbf{softmax}(\alpha^{(i,j)}), \vec{\mathbf{g}}^{(i,j)}(\mathbf{x}^{(i)}) \rangle.$$

- Задана выборка  $\mathcal{D} = \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{val}}$ . Задана функция потерь  $\mathcal{L}_{\text{train}}$ ,  $\mathcal{L}_{\text{val}}$ . Пусть  $\alpha = [\alpha^{(i,j)}]$ . Пусть  $\mathbf{w}$  – параметры модели. Двухуровневая задача оптимизации

$$\begin{aligned} \min_{\alpha} \mathcal{L}_{\text{val}}(\mathbf{w}^*, \alpha), \\ \text{s.t. } \mathbf{w}^* = \arg \min_{\mathbf{w}} \mathcal{L}_{\text{train}}(\mathbf{w}, \alpha) \end{aligned}$$



► Смешанная операция:

$$\hat{\mathbf{g}}^{(i,j)} = \text{softmax}(\alpha^{(i,j)})_1 \mathbf{g}_1^{(i,j)}(\mathbf{x}^{(i)}) + \\ \text{softmax}(\alpha^{(i,j)})_2 \mathbf{g}_2^{(i,j)}(\mathbf{x}^{(i)}) + \\ \text{softmax}(\alpha^{(i,j)})_3 \mathbf{g}_3^{(i,j)}(\mathbf{x}^{(i)})$$

## Постановка задачи поиска архитектуры на мультидоменных данных

- ▶ Задано  $D$  доменов,  $\{\mathfrak{D}_d\}_{d=1}^D$ ,  $\mathfrak{D}_d = (\mathbf{X}_d, \mathbf{y}_d)$ ,  $\mathbf{y}_d \subset \mathcal{Y}$ . Пусть заданы функции потерь  $\mathcal{L}_{\text{train}}^{(d)}$  и  $\mathcal{L}_{\text{val}}^{(d)}$ . Модель задается разделяемыми между доменами параметрами  $\mathbf{w}$ . Пусть  $\alpha_d$  – вектор структурных параметров домена  $d$ ,  $\alpha = [\alpha_d]_{d=1}^D$ .
- ▶ Задано распределение на доменах  $d \sim \text{Categorical}(\mathbf{p})$ . Двухуровневая задача оптимизации

$$\begin{aligned} \alpha^* &= \arg \min_{\alpha} \sum_{d=1}^D p_d \mathcal{L}_{\text{valid}}^{(d)}(\mathbf{w}^*, \alpha_d), \\ \text{s.t. } \mathbf{w}^* &= \arg \min_{\mathbf{w}} \sum_{d=1}^D p_d \mathcal{L}_{\text{train}}^{(d)}(\mathbf{w}, \alpha_d). \end{aligned}$$

- ▶ Получение итоговой архитектуры для каждого из доменов:

$$\mathbf{g}^{(i,j)}(.) = \vec{\mathbf{g}}_{k^*}^{(i,j)}, \quad k^* = \arg \max_k (\alpha_d^{(i,j)})_k.$$

## Предлагаемые регуляризаторы

- Заданы распределения на ребрах графа  $P_d^{(i,j)} = \text{Categorical}(\text{softmax}(\alpha_d^{(i,j)}))$ . Структурный регуляризатор:

$$\mathcal{L}_{\text{struct}}(\alpha) = \frac{1}{D(D-1)} \sum_{(i,j)} \sum_{d=1}^D \sum_{d'=1, d' \neq d}^D \text{JS}(P_d^{(i,j)} || P_{d'}^{(i,j)}),$$

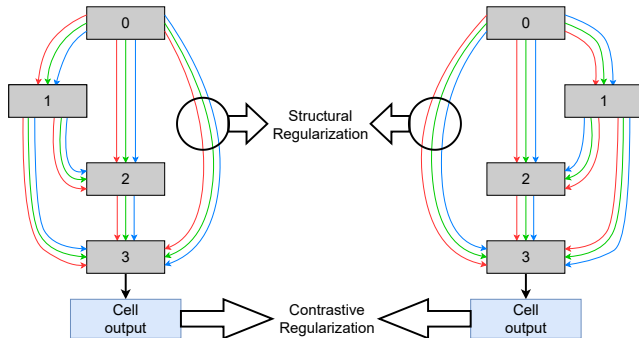
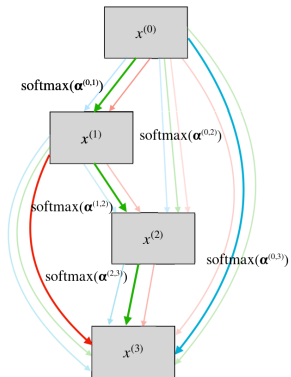
где  $\text{JS}(.||.)$  – дивергенция Йенсена—Шеннона.

- Заданы наборы объектов для каждого из доменов  $\{\mathcal{S}_b^{(d)}\}_{d=1}^D$ . Регуляризатор пространства скрытых представлений модели:

$$\mathcal{L}_{\text{contr}}(\mathbf{w}, \alpha) = \mathbb{E}_{d,d' \sim \text{Categorical}(\mathbf{p})} \mathbb{E}_{\mathcal{S}_b^{(d)}, \mathcal{S}_b^{(d')}} \mathcal{L}_{\text{triplet}}(\mathbf{w}, \alpha, \mathcal{S}_b^{(d)}, \mathcal{S}_b^{(d')}),$$

где  $\mathcal{L}_{\text{triplet}}$  – триплетная функция потерь для скрытых представлений  $\mathcal{S}_b^{(d)}, \mathcal{S}_b^{(d')}$ .

# Построение мультимодели





## Задача оптимизации

- Заданы коэффициенты регуляризации  $\beta_{\text{trip}} \geq 0$ ,  $\beta_{\text{struct}} \geq 0$ . Оптимальный вектор структурных параметров в задаче выбора архитектуры на мультидоменных данных находится из следующей задачи оптимизации:

$$\begin{aligned} \alpha^* &= \arg \min_{\alpha} \sum_{d=1}^D p_d \mathcal{L}_{\text{val}}^{(d)}(\mathbf{w}^*, \alpha) + \beta_{\text{trip}} \mathcal{L}_{\text{contr}}(\mathbf{w}, \alpha) + \beta_{\text{struct}} \mathcal{L}_{\text{struct}}(\alpha), \\ \text{s.t. } \mathbf{w}^* &= \arg \min_{\mathbf{w}} \sum_{d=1}^D p_d \mathcal{L}_{\text{train}}^{(d)}(\mathbf{w}, \alpha) + \beta_{\text{trip}} \mathcal{L}_{\text{contr}}(\mathbf{w}, \alpha). \end{aligned}$$

## Постановка вычислительного эксперимента

- ▶ Цель – получение зависимости качества работы мультимодели и количества ее параметров в зависимости от используемого регуляризатора.
- ▶ Эксперимент проводится на подвыборке MNIST. В качестве доменов рассматриваются изображения, повернутые на угол, кратный  $\pi/2$ . Число доменов меняется от 1 до 4. Сравниваются следующие модели: мультимодель со структурной регуляризацией, мультимодель с регуляризацией скрытых представлений, а также модель, не учитывающая разбиение выборки на домены.
- ▶ Оценивается средняя точность (accuracy) на тестовой выборке для каждого из доменов. Также приводится количество параметров для каждой модели.

## Результаты вычислительного эксперимента

model	accuracy	num. of params
1 domain		
single	60.59	5029
2 domains		
single, union	66.95	6560
multimodel, struct	62.86	<b>5248</b>
multimodel, contr	<b>69.64</b>	9328
3 domains		
single, union	63.02	5685
multimodel, struct	64.85	<b>6826</b>
multimodel, contr	<b>65.01</b>	12096
4 domains		
single, union	67.16	6560
multimodel, struct	63.15	<b>7872</b>
multimodel, contr	<b>67.98</b>	13685

## Вычислительный эксперимент на задаче языкового моделирования

1. Задача языкового моделирования решалась на выборке IWSLT14.
2. Рассматривались одиночные модели, решающие задачу на одном домене и на объединении доменов, а также предлагаемая мультимодель.

Model	En, ppl	De, ppl
Single, En	<b>59.35</b>	-
Single, De	-	-
Single, En+De	72.68	82.15
Multimodel	<b>64.91</b>	<b>74.06</b>

## Заключение

- ▶ Рассмотрена задача поиска архитектуры модели глубокого обучения на мультидоменных данных. Задача рассматривалась как задача мультимоделирования.
- ▶ Предложены два метода регуляризации: регуляризация структуры и регуляризация пространства скрытых представлений модели.
- ▶ Продемонстрирована работоспособность предлагаемого решения. При использовании первого регуляризатора мультимодель имеет меньшее число параметров. При использовании второго регуляризатора модель имеет лучшую точность классификации.
- ▶ В дальнейшем планируется провести вычислительный эксперимент на задаче мультиязычного языкового моделирования.