

# Выбор оптимальной сети глубокого обучения в задачах классификации временных рядов

М. С. Попова

Научный руководитель: В. В. Стрижов  
Московский физико-технический институт  
Факультет управления и прикладной математики  
Кафедра интеллектуальных систем

23 июня 2015 г.

**Цель исследования:** предложить стратегию выбора моделей и решить задачу классификации физической активности человека по измерениям акселерометра.

**Мотивация:** построение устойчивой модели для классификации временных рядов.

**Проблема:** выбор устойчивых моделей в условиях мультиколлинеарности выборки затруднен из-за необходимости оценки большого числа параметров.

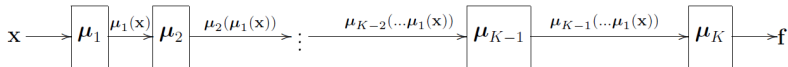
**Метод решения задачи:** процедура получения модели с оптимальным числом параметров, использующая методы последовательного добавления и удаления параметров.

**Выборка:**  $\mathcal{D} = \{(\mathbf{x}_i, t_i), i = 1, \dots, N\}$  –  $N$  пар объект–ответ. Объекты  $\mathbf{x}_i$  – сегменты временного ряда ( $\mathbf{x}_i \in \mathbb{R}^n$ ). Всего  $M$  классов:  $t_i \in \{1, \dots, M\}$ .

**Модель классификации:** суперпозиция функций

$$\mathbf{f}(\mathbf{x}, \mathbf{w}) = \mu_1(\mu_2(\dots \mu_K(\mathbf{x}))) : \mathbb{R}^n \rightarrow [0, 1]^M,$$

где  $\mu_k$ ,  $k \in \{1, \dots, K\}$ , – модели из класса нейронных сетей с соответствующими векторами параметров  $\mathbf{w}_k$ ,  $k \in \{1, \dots, K\}$ ,  $\mathbf{w} = [\mathbf{w}_1, \dots, \mathbf{w}_K]^T$ ,  $\mathbf{f}(\mathbf{x}; \mathbf{w}) = [p(y_1 = 1 | \mathbf{x}; \mathbf{w}), \dots, p(y_M = M | \mathbf{x}; \mathbf{w})]^T$ .



Автокодировщик – суперпозиция блоков

$$\mu = \varphi(\mathbf{g}(\mathbf{x})),$$

где  $\mathbf{g}(\mathbf{x}) = \sigma(\mathbf{W}_g \mathbf{x} + \mathbf{b}_g)$ , – кодирующий блок, или encoder,

$\varphi(\mathbf{g}(\mathbf{x})) = \sigma(\mathbf{W}_h \mathbf{g}(\mathbf{x}) + \mathbf{b}_h)$ , – декодирующий блок, или decoder, а

$$\sigma(\mathbf{t}) = \frac{1}{1 + \exp(-\mathbf{t})} \text{ – сигмоидная функция.}$$

Вектор параметров:  $\Theta = (\mathbf{W}_g, \mathbf{W}_h, \mathbf{b}_g, \mathbf{b}_h)$ ,  $\mathbf{W}_h = \mathbf{W}_g^\top$ .

Функция ошибки:  $S(\Theta, \mathbf{x}) = \|\mathbf{f}(\mathbf{x}|\Theta) - \mathbf{x}\|_2^2$ .

## Постановка задачи оптимизации параметров

$$\hat{\Theta} = \operatorname{argmin}_{\Theta} \frac{1}{2|\mathcal{L}|} \sum_{\mathbf{x} \in \mathcal{L}} S(\Theta, \mathbf{x}).$$

Двухслойная нейронная сеть – это отображение вида

$$\mathbf{a}(\mathbf{x}) = \mathbf{W}_2^T \tanh(\mathbf{W}_1^T \mathbf{x}),$$

$$\mu(\mathbf{x}) = \frac{\exp(\mathbf{a}(\mathbf{x}))}{\sum_{j=1}^n \exp(a_j(\mathbf{x}))}.$$

Вектор параметров:  $\mathbf{w} = \text{vec}(\mathbf{W}_1^T | \mathbf{W}_2^T)$ ,  $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1})$ , где  $\mathbf{A}^{-1}$  – ковариационная матрица параметров,  $\mathbf{w}^T \mathbf{A} \mathbf{w} > \mathbf{0}$  для любого  $\mathbf{w} \in \mathbb{R}^k$ .

## Функция ошибки

$$S(\mathbf{w} | \mathcal{R}) = - \sum_{\mathbf{x} \in \mathcal{R}} \sum_{\xi=1}^M [y_t = 1] \ln(f_{\xi}(\mathbf{x}, \mathbf{w})).$$

- 1 **Сложность.** Число ненулевых параметров модели, мощность структуры.

$$C(\mathbf{w}) = \sum_{i=1}^k [w_i \neq 0] = |\mathcal{A}|$$

где  $\mathcal{A} = \{j : w_j \neq 0\} \subseteq \mathcal{J}$  – структура модели.

- 2 **Точность.** Величина функции ошибки на выборке:

$$S(\mathbf{w}|\mathcal{K}) = - \sum_{i \in \mathcal{K}} \sum_{\xi=1}^z t_{i\xi} \ln(f_{\xi}(\mathbf{x}_i, \mathbf{w})).$$

- 3 **Устойчивость.** Число обусловленности матрицы, обратной к ковариационной:

$$\eta(\hat{\mathbf{w}}) = \frac{\lambda_{\max}}{\lambda_{\min}},$$

где  $\lambda_{\max}$  — максимальное, а  $\lambda_{\min}$  — минимальное собственные числа матрицы  $\mathbf{A}$ .

## Множество допустимых моделей

$$\mathfrak{F} = \bigcup_{A \subseteq \mathcal{J}} \{\mathbf{f}_A\}.$$

## Отношение доминирования

Модель  $\mathbf{f}'$  доминирует модель  $\mathbf{f}$  ( $\mathbf{f}' \succ \mathbf{f}$ ), если

$$C' \leq C, \quad \eta' \leq \eta, \quad S' \leq S,$$

Модель  $\mathbf{f} \in \mathcal{F}$  назовем *оптимальной по Парето*, если не существует  $\mathbf{f}' \in \mathcal{F}$  такой, что  $\mathbf{f}' \succ \mathbf{f}$ .

Множество оптимальных по Парето моделей назовем *Парето оптимальным фронтом*  $\text{POF}_{\mathfrak{F}}$  множества  $\mathfrak{F}$ .

## Задача выбора оптимальной модели

состоит в том, чтобы найти Парето оптимальный фронт  $\text{POF}_{\mathfrak{F}}$  множества допустимых моделей  $\mathfrak{F}$ .

## Определение стратегии

Стратегия задается следующими математическими объектами:

- набором критериев качества  $\{Q\}$ ;
- набором ограничений на структуру и параметры модели  $\mathcal{A} \subseteq \mathcal{J}$ ,  $\mathbf{w} = \hat{\mathbf{w}}_{\mathcal{A}} = \underset{\mathbf{w}_{\mathcal{A}} \in \mathbb{R}^k}{\operatorname{argmin}} S(\mathbf{w}_{\mathcal{A}} | \mathcal{L})$ ;
- критериями прореживания и наращивания модели;
- критериями останова шагов удаления и добавления структурных единиц в модель;
- критерием останова процедуры выбора модели.

## Критерии качества

Сложность  $C$ , точность  $S$ , устойчивость  $\eta$ .



Приращение  $\Delta S$  вблизи точки локального минимума  $\mathbf{w}_0$ :

$$\Delta S = S(\mathbf{w}_0 + \Delta \mathbf{w}) - S(\mathbf{w}_0) = \frac{1}{2} \Delta \mathbf{w}^\top \mathbf{H} \Delta \mathbf{w}.$$

Задача условной минимизации:

$$\Delta S = \frac{1}{2} \Delta \mathbf{w}^\top \mathbf{H} \Delta \mathbf{w} \rightarrow \min, \quad \mathbf{e}_j^\top \Delta \mathbf{w} + w_j = 0.$$

Для решения задачи строим Лагранжиан:

$$L = \frac{1}{2} \Delta \mathbf{w}^\top \mathbf{H} \Delta \mathbf{w} - \lambda_j (\mathbf{e}_j^\top \Delta \mathbf{w} + w_j).$$

Дифференцируем  $L$  по  $\Delta \mathbf{w}$ :

$$L_j = \frac{w_j^2}{2[\mathbf{H}^{-1}]_{j,j}},$$

$$\hat{j} = \underset{j \in \mathcal{A}}{\operatorname{argmin}} L_j.$$

$\mathbf{W}$  — матрица реализаций вектора  $\hat{\mathbf{w}}$ .

Оценка ковариационной матрицы параметров:

$$\mathbf{A}^{-1} = \text{cov}(\mathbf{W}) = \mathbb{E}(\mathbf{W}^T \mathbf{W}) = \frac{1}{r} \mathbf{W} \mathbf{W}^T.$$

Выполнив сингулярное разложение  $\mathbf{W}$ , получим:

$$\mathbf{A}^{-1} = (\mathbf{W} \mathbf{W}^T) = (\mathbf{U} \mathbf{\Lambda} \mathbf{V}^T \mathbf{V} \mathbf{\Lambda}^T \mathbf{U}^T) = (\mathbf{U} \mathbf{\Lambda} \mathbf{\Lambda}^T \mathbf{U}^T) = \mathbf{U} \mathbf{\Lambda}^2 \mathbf{U}^T.$$

$$\hat{l} = \operatorname{argmax}_{l \in \mathcal{A}} \frac{\lambda_{\max}}{\lambda_l}.$$

$$\hat{j} = \operatorname{argmax}_{j \in \mathcal{A}} \frac{u_{ij}^2 \lambda_{jj}^2}{\sigma(w_{ij})},$$

## Критерии последовательного прореживания и наращивания

$$\hat{j} = \operatorname{argmin}_{j \in \mathcal{A}} S(\mathbf{w}_{\mathcal{A}} \setminus w_j | \mathcal{T}).$$

$$\hat{j} = \operatorname{argmin}_{j \in \mathcal{J} \setminus \mathcal{A}} S(\mathbf{w}_{\mathcal{A}} \cup w_j | \mathcal{T}).$$

## Критерий смены шагов Add-Del

$$S(\hat{\mathbf{w}}_{\mathcal{A}} | \mathcal{T}) \geq S_{\min} + \delta S,$$

## Критерий останова процедуры

$$H(\mathcal{A}, \mathcal{A}') = -\rho(\mathbf{z}, \mathbf{z}') \ln(\rho(\mathbf{z}, \mathbf{z}')),$$
$$z_j = \begin{cases} 0, & \text{если } w_j + \Delta w_j = 0, \text{ т. е. } j \notin \mathcal{A}; \\ 1, & \text{иначе.} \end{cases}$$

### Цели вычислительного эксперимента

- повышение точности классификации для различных типов суперпозиции моделей и сравнение их со значениями точности из работ зарубежных авторов;
- получение устойчивых моделей с помощью описанной стратегии и сравнение различных критериев прореживания.

## Первый набор данных

Сегменты временных рядов акселерометра, всего 4 типа физической активности человека – ходьба, бег, стояние и сидение. Сегмент – 10 секундный отрезок, состоящий из 600 измерений проекции ускорения.

*Kwapisz J. R., Weiss G. M., Moore S. Activity recognition using cell phone accelerometers // SIGKDD Explorations, 2010. Vol. 12. No 2. P. 74–82.*

Класс	3 уровня	4 уровня	5 уровней	Kwapisz et. al.
Бег	98%	95%	97%	98%
Ходьба	95%	94%	96%	92%
Сидение	100%	100%	100%	95%
Стояние	89%	82%	84%	92%

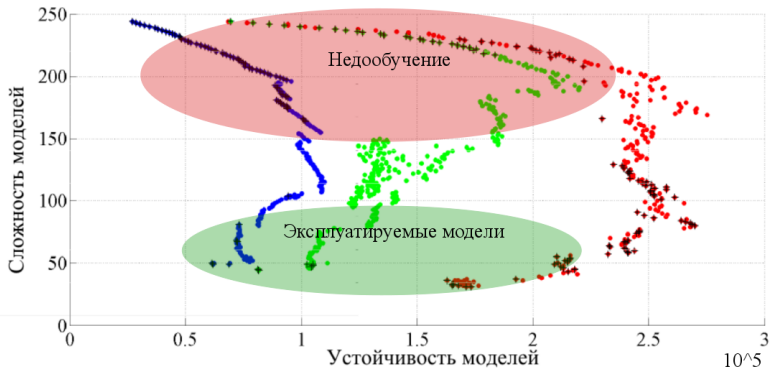
## Второй набор данных

Векторы признаков – сегменты временных рядов с акселерометра и гироскопа, всего 6 типов физической активности – ходьба, бег, стояние, сидение, подъём и спуск. Сегмент – 2,5 секунднй отрезок, состоящий из 561 измерений проекции ускорения.

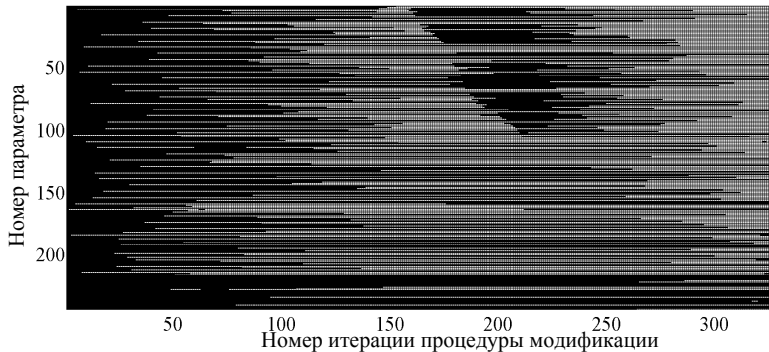
*Anguita D., Ghio A., Oneto L., Parra X., Luis Reyes-Ortiz J.*  
Human Activity Recognition on Smartphones // Proceedings of the 4th International Workshop (IWAAL 2012). – Springer, 2012.

Класс	3 уровня	4 уровня	5 уровней	Anguita et. al.
Ходьба	96%	98%	96%	97%
Подъём	94%	93%	80%	87%
Спуск	98%	92%	96%	72%
Сидение	91%	65%	84%	95%
Стояние	75%	80%	71%	97%
Лежание	99,7%	98%	92%	100%

# Эксперимент: результаты



Стратегия	C	S	$\eta$
Оптимальное прореж-ние	50	877	$2.2 \cdot 10^5$
Последовательное прореж-ние	50	870	$1.2 \cdot 10^5$
Устойчивое прореж-ние	50	866	$6 \cdot 10^4$



- каждый столбец описывает структуру модели на данной итерации;
- белая клетка – параметр неактивен;
- черная клетка – параметр активен.



- Предложена стратегия пошаговой модификации двухслойной нейронной сети для получения точных и устойчивых моделей оптимальной сложности;
- Предложена суперпозиция моделей для выделения информативных признаков;
- Проведен вычислительный эксперимент на двух наборах данных. Результаты сравнивались с работами зарубежных авторов.

## Публикации

- Попова М. С., Стрижов В. В. Выбор оптимальной модели классификации физической активности по измерениями акселерометра // Информатика и ее применения, 2015. Т. 9. Вып 1. С. 79–89.
- Попова М. С., Стрижов В. В. Построение суперпозиции нейронных сетей глубокого обучения для классификации временных рядов // Системы и средства информатики (принято в журнал).