

# Построение интерпретируемых моделей глубокого обучения в задаче социального ранжирования

Гончаров Алексей Владимирович

Московский физико-технический институт  
Кафедра интеллектуальных систем, ФУПМ.  
Научный руководитель: д.ф.-м.н. В. В. Стрижов.

Москва-2018

Предложить решение задачи кредитного скоринга и систему, создающую эксплуатируемую модель без участия аналитика. Модель строится по принципу глубокой сети как интерпретируемая суперпозиция.

## Проблема

- В индустрии модель строится поэтапно, на каждом этапе оптимизируется локальный критерий качества

## Требования

- Интерпретируемость модели оценивается экспертами в предметной области
- Высокая точность согласно эксплуатационному критерию

## Предлагается

- Единая процедура оптимизации структуры и параметров

Наблюдается быстрый рост сложности при медленном росте качества в задаче социального ранжирования:

Модель	Качество ( $\mathcal{L}$ )	Качество ( $\mathcal{T}$ )	Число параметров
Logistic Regression	55%	53%	12
Deep Neural Nets	59%	57%	240
Random Forest	61%	57%	1400
XGBoost	63%	58%	12000

Мотивация: предложить процедуру с ограниченной сложностью, но высоким качеством.

## Дано

- $\mathcal{D} = \{(\mathbf{x}_i, y_i) : i = 1, \dots, m\}$ ,  $\mathbf{x}_i \in \mathbb{R}^l \times \mathbb{C}^c \times \mathbb{B}^b$ ,  $y_i \in \{0, 1\}$
- Индексы объектов  $\{i = 1, \dots, m\} = \mathcal{I} = \mathcal{L} \sqcup \mathcal{T}$
- Индексы признаков  $\mathcal{A} = \mathcal{A}_l \sqcup \mathcal{A}_c \sqcup \mathcal{A}_b$
- Матрица плана  $\mathbf{X} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{X} = (\chi^1, \dots, \chi^n)$ .

## Вид модели

Модель классификации  $F$  — суперпозиция  $F = f \circ f_f \circ f_g \circ f_s$

- 1  $f_s$  — сегментация линейных признаков
- 2  $f_g$  — группировка категориальных признаков
- 3  $f_f$  — порождение новых признаков

$$f_f \circ f_g \circ f_s : \mathbf{X} \mapsto \hat{\mathbf{X}}.$$

- 4  $f$  — логистическая регрессия

$$f(\hat{\mathbf{X}}, \mathbf{w}) = \frac{1}{1 + \exp(-\hat{\mathbf{X}}\mathbf{w})}.$$

В задачах социального ранжирования (кредитного скоринга) также для оценки моделей используют следующие критерии качества:

## Эксплуатационные критерии

- AUC — площадь под ROC-кривой
- Сложность, число признаков
- Устойчивость во времени

## Оптимизационные критерии

- Правдоподобие модели
- Устойчивость — ортогональность признаков
- Точность, полнота

Основным критерием, влияющим на бизнес-процессы банка принят AUC. Поэтому он выбран как целевой критерий при решении задачи оптимизации.

## Критерий качества

Для процедур сегментации, группировки и порождения признаков эксплуатационный критерий качества, согласно банковским требованиям, — AUC или площадь под ROC-кривой:

$$R(H, P, \mathbf{w}) = \frac{\sum_{i \in \mathcal{L}} \sum_{j \in \mathcal{L}} [y_i < y_j] [F(\mathbf{x}_i) < F(\mathbf{x}_j)]}{\sum_{i \in \mathcal{L}} \sum_{j \in \mathcal{L}} [y_i < y_j]}.$$

## Задача оптимизации

$$H^*, P^*, \mathbf{w}^* = \operatorname{argmin}_{H, P, \mathbf{w}} R(H, P, \mathbf{w}),$$

где  $P = p_f \sqcup p_g \sqcup p_s$  — параметры сегментации, группировки и порождения признаков

$H = h_f \sqcup h_g \sqcup h_s$  — гиперпараметры соответственно

Процедуры в суперпозиции при построении модели:

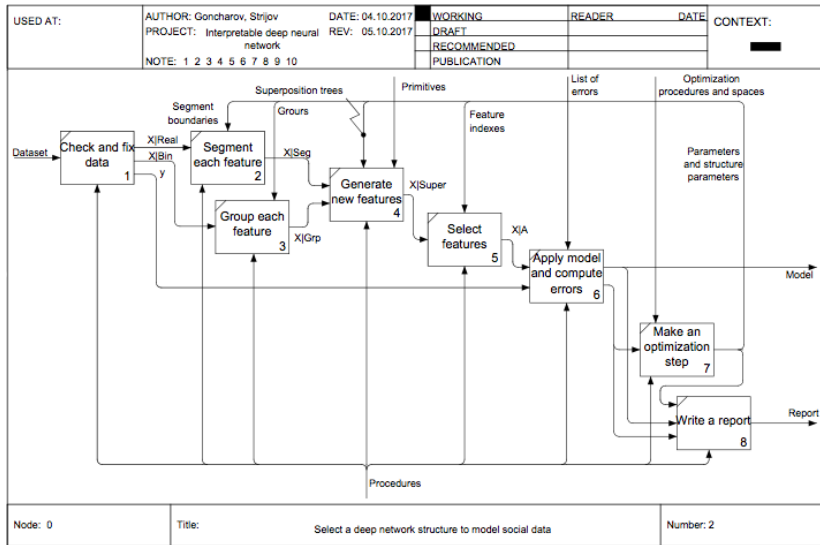
- Сегментация
- Группировка
- Порождение признаков
- Построение линейной модели

## Базовая модель

Последовательная оптимизации в каждой задачи локальных критериев качества, не связанных с целевым критерием.

## Предлагается

Единая процедура оптимизации: пошаговая итерационная оптимизация заданного целевого критерия качества.





*Утверждение 1. Для задачи оптимизации:  $S(x_1, \dots, x_n) \rightarrow \min$  при начальном приближении  $X^0 = x_1^0, \dots, x_n^0$  процедура оптимизации:*

$$\hat{x}_i = \operatorname{argmin}_{x_i} S_i(\hat{x}_1, \dots, \hat{x}_{i-1}, x_i, x_{i+1}^0, \dots, x_n^0), i = 1, \dots, n,$$

*доставляет качество не лучше, чем процедура оптимизации:*

$$\tilde{x}_1, \dots, \tilde{x}_n = \operatorname{argmin}_{x_1, \dots, x_n} S(x_1, \dots, x_n),$$

*то есть:  $S(\hat{x}_1, \dots, \hat{x}_n) \leq S(\tilde{x}_1, \dots, \tilde{x}_n)$ .*

Пусть линейный признак  $\chi \in [a, b]$ .

Параметрами  $p_s$  процедуры сегментации признака  $\chi$  является набор координат узлов разбиения, причем

$$p_s^j \in [a, b], j = 0, \dots, h_s; \quad a = p_s^0 < \dots < p_s^j < \dots < p_s^{h_s} = b.$$

Исходная матрица плана пополняется множеством признаков — индикаторами принадлежности сегменту:

$$\{\chi^j\}_{j=1}^{h_s} \in \mathbb{B}^{h_s} : \chi^j = [\chi \in [p_s^{q-1}; p_s^q]], q = 1, \dots, h_s.$$

Для каждого линейного признака решается следующая оптимизационная задача:

$$h_s^*, p_s^*, w^* = \operatorname{argmin}_{h_s, p_s, w} R(h_s, p_s, w | D, \mathcal{A}_l, \mathcal{L}),$$

# Процедура группировки

Пусть категориальный признак  $\chi$  имеет  $|C|$  категорий.  
Структурный параметр  $h_g$  — множество новых категорий,  
 $|h_g| \leq |C|$ . Параметр группировки — сюръекция  $p_g : C \rightarrow h_g$ .

Исходное признаковое пространство изменяется:  
категориальный признак  $\chi$  заменяется категориальным  
признаком  $\chi_h$ .

$\chi$	=	1	2	3	...	C	C — число категорий
		↓	↓	↓		↓	
$\chi_h$	=	$\gamma_1$	$\gamma_2$	$\gamma_3$	...	$\gamma_C$	$\gamma_i \in h_g$

Для каждого категориального признака решается следующая  
оптимизационная задача:

$$h_g^*, p_g^*, w^* = \operatorname{argmin}_{h_g, p_g, w} R(h_g, p_g, w | D, \mathcal{A}_C, \mathcal{L}),$$

Множество интерпретируемых порождающих функций  $G$

Описание	Формула	in	N in	out
Negate binary	$\bar{x}$	bin	1	bin
Logarithm	$\log(x)$	lin	1	lin
Logistic sigmoid	$\frac{1}{1+\exp(x)}$	lin	1	lin
Square root	$\sqrt{x}$	lin	1	lin
Inverse	$\frac{1}{x}$	lin	1	lin
Multiplication	$x * y$	any	2	lin
Sum	$x + y$	any	2	lin

*Определение.* Суперпозиция — формула, представимая в виде дерева, в каждом узле которого находится функция из  $\{f_i\}$ , а в каждом листе — признак из  $\{\chi_i\}$ .

Сложность суперпозиции — число использованных в ней элементов из  $\{f_i\}$  и  $\{\chi_i\}$  с повторениями.

- Измеряемые признаки  $\mathbf{X} = \{\chi\}$
- Заданные экспертами порождающие функции  $G = \{g(b, \chi)\}$
- Правила порождения:  $\mathcal{G} \ni G$ , суперпозиция  $g_k \circ g_l \in \mathcal{G}$  построена с учетом ограничений на сложность, типы входных и выходных переменных.
- Правила упрощения:  $g_u$  не принадлежит  $\mathcal{G}$ , если существует правило  $g_u \rightarrow g_v \in \mathcal{G}$

Параметрами порождения признаков является набор супераозиций  $p_f = \{\mathbb{F}_i\}$ :

$$h_f, p_f, w = \operatorname{argmin}_{h_f, p_f, w} R(h_f, p_f, w | D, \mathcal{A}_l \sqcup \mathcal{A}_b, \mathcal{L}).$$

## Сегментация

$p_s^i = p_s^{i-1} - \lambda \nabla R(p_s)$ , — задача решается градиентным спуском.

## Группировка

$p_g : C \rightarrow h_g$  — сюръективное отображение между двумя конечными множествами. Задача решается генетическим алгоритмом.

## Порождение признаков

$F = \{f_i\}$  — множество суперпозиций элементарных функций. Задача поиска оптимального набора решается генетическим алгоритмом

## Построение модели

$w$  — непрерывные веса модели. Задача решается методом градиентного спуска.

Последовательно выполняются следующие шаги:

- 1 Инициализация начальных точек для градиентных методов и начальных поколений для генетических алгоритмов.
- 2 Шаги градиентного метода оптимизации для процедуры сегментации для каждого признака
- 3 Итерация генетической оптимизации для процедуры группировки для каждого признака
- 4 Итерация генетической оптимизации для процедуры порождения
- 5 Возврат ко 2 шагу до сходимости функционала качества

Вычислительный эксперимент проводился на выборке "Statlog (German Credit Data)", UCI.

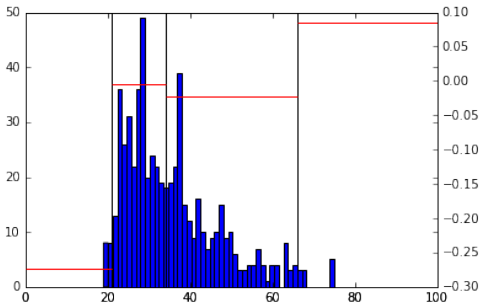
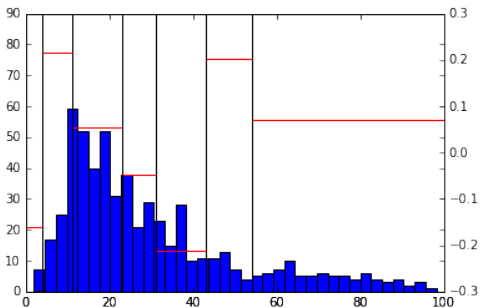


Рис.: Пример сегментации линейного признака и соответствующий вес в модели.



Вычислительный эксперимент проводился на выборке "Statlog (German Credit Data)", UCI.



**Рис.:** Пример сегментации линейного признака и соответствующий вес в модели.

Вычислительный эксперимент проводился на выборке "Statlog (German Credit Data)", UCI.

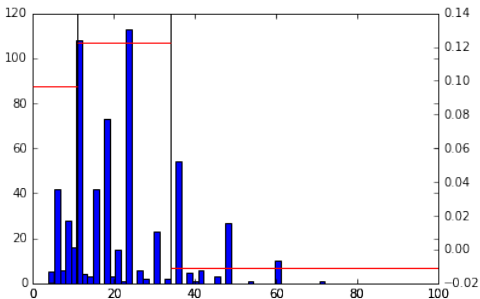


Рис.: Пример сегментации линейного признака и соответствующий вес в модели.

Вычислительный эксперимент проводился на выборке "Statlog (German Credit Data)", UCI.

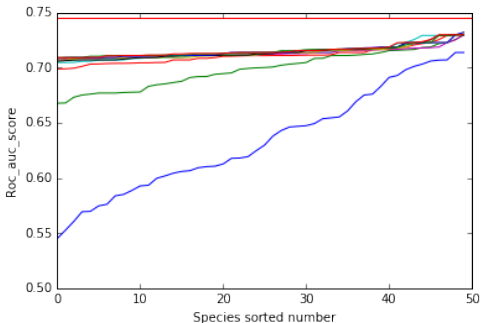


Рис.: Сходимость процедуры оптимизации группировки.

# Вычислительный эксперимент

Вычислительный эксперимент проводился на выборке "Statlog (German Credit Data)", UCI.

$$\chi = 1 \quad 2 \quad 3 \quad 4 \quad 5$$

5 — число исходных категорий

$$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow$$

$$\chi_h = 1 \quad 2 \quad 2 \quad 3 \quad 3$$

3 — число новых категорий,  $\gamma_i \in h_{gi}$

$$\chi = 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5$$

6 — число исходных категорий

$$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow$$

$$\chi_h = 3 \quad 1 \quad 3 \quad 2 \quad 4 \quad 1$$

4 — число новых категорий,  $\gamma_i \in h_g$

Например, признак кредитной истории, содержащий 5 категорий разбивается на три категории:

"Просрочил платеж в прошлом" и "Критический аккаунт",

"Не брал",

"Выплачивал" и "Выплачивал до настоящего момента".

# Вычислительный эксперимент

Вычислительный эксперимент проводился на выборке "Statlog (German Credit Data)", UCI.

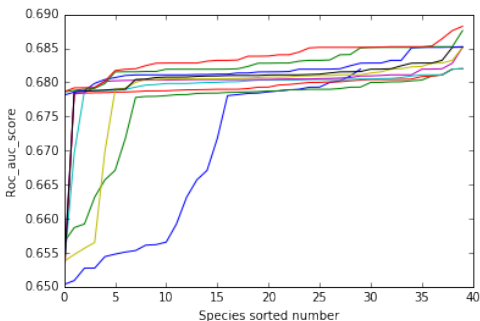


Рис.: Сходимость процедуры оптимизации порождения признаков.

Пример порожденных признаков:

$$\chi'_1 = \frac{1}{\chi_1 + \chi_2 + \chi_3}$$

$$\chi'_2 = \sqrt{2\chi_2} + \sqrt{\chi_1 + \chi_2}$$

Вычислительный эксперимент проводился на выборке "Statlog (German Credit Data)", UCI.

Модель	Качество ( $\mathcal{L}$ ) Roc-Auc	Качество ( $\mathcal{T}$ ) Roc-Auc	Число параметров
Logistic Regression	0.751	0.704	6
<b>End-to-end LR</b>	0.782	0.730	25
XGBoost	0.912	0.729	5000
NN	0.834	0.720	3000

- Предложена система создания эксплуатируемой модели без участия аналитика (end-to-end)
- Модель построена по принципу глубокого обучения как интерпретируемая суперпозиция
- Получено интерпретируемое решение
- Решена проблема поэтапной оптимизации локальных критериев качества

## Планы

- Усовершенствовать имеющуюся процедуру оптимизации модели
- Изменить процедуру порождения признаков
- Создание адаптивной системы онлайн обучения мультимодели на полученном пространстве признаков