

A. V. Goncharov

Interpreted deep learning models in the social ranking tas

?

Key words: ?

## 1 Key values

Set:  $D = \{(\mathbf{x}_i, y_i) : i = 1, \dots, m\}$

Object:  $\mathbf{x}_i \in \mathbb{R}^l \times \mathbb{C}^c \times \mathbb{B}^b$ , object indexes:  $\{i = 1, \dots, m\} = \mathcal{I}$ ;  $\mathcal{I} = \mathcal{L} \sqcup \mathcal{T}$

Plan matrix:  $\mathbf{X} = (\boldsymbol{\chi}^1, \dots, \boldsymbol{\chi}^n)$ , feature indexes:  $\mathcal{A} = \mathcal{A}_l \sqcup \mathcal{A}_c \sqcup \mathcal{A}_b$

Target:  $y_i \in \{0, 1\}$

Model:  $F = f \circ f_f \circ f_g \circ f_s$

Feature transformation:  $f_f \circ f_g \circ f_s : \mathbf{X} \mapsto \hat{\mathbf{X}}$ , it's optimization criterion:  $Q(\mathbf{w}) = \frac{\sum_{i \in \mathcal{L}} \sum_{j \in \mathcal{L}} [y_i < y_j] [F(\mathbf{x}_i) < F(\mathbf{x}_j)]}{\sum_{i \in \mathcal{L}} \sum_{j \in \mathcal{L}} [y_i < y_j]}$ .

Classification model:  $f(\hat{\mathbf{X}}, \mathbf{w}) = \frac{1}{1 + \exp(-\hat{\mathbf{X}}\mathbf{w})}$ , it's optimization criterion:  $L(\mathbf{w}) = -\ln P(D|\mathbf{w}) = -\sum_{i \in \mathcal{L}} (y_i \ln \mathbf{w}^\top \mathbf{x}_i + (1 - y_i) \ln(1 - \mathbf{w}^\top \mathbf{x}_i))$ .

Parameter set:  $P = p_f \sqcup p_g \sqcup p_s$

Hyperparameter set:  $H = h_f \sqcup h_g \sqcup h_s$

Global optimization problem:  $H^*, P^*, w^* = \underset{H, P, \mathbf{w}}{\operatorname{argmin}} Q(\mathbf{w})$ .

*Проблема 1: лог.регр. оптимизируется по правдоподобию в решении, а параметры по Рос-аис. Как записать единую постановку задачи? Или не говорить, что регрессия так оптимизируется?*

## 2 model

Old one:

$$h_s^*, p_s^* = \underset{h_s, p_s}{\operatorname{argmin}} S_s(h_s, p_s),$$

$$h_g^*, p_g^* = \underset{h_g, p_g}{\operatorname{argmin}} S_g(h_g, p_g),$$

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} L(\mathbf{w}),$$

*Проблема 2: в старой модели нет оптимизации критериев напрямую. Там есть*

просто предложенный алгоритм подбора параметров. Как записать старую постановку задачи тогда?

Проблема 3: в старой модели нет оптимизации генератора признаков. Его можно просто не включать тогда в описание задачи?

New one:

Iterational:

$$\begin{aligned} h_s^*, p_s^* &= \operatorname{argmin}_{h_s, p_s} Q(h_s, p_s), \\ h_g^*, p_g^* &= \operatorname{argmin}_{h_g, p_g} Q(h_g, p_g), \\ h_f^*, p_f^* &= \operatorname{argmin}_{h_f, p_f} Q(h_f, p_f), \\ \mathbf{w}^* &= \operatorname{argmin}_{\mathbf{w}} L(\mathbf{w}), \end{aligned}$$

Проблема 4: как записать в постановке задачи то, что это итерационный метод? Мат постановка одинакова для задач по сути, но алгоритмы разные. Как это разбединить?

Statement 1. For optimization task:  $S(x_1, \dots, x_n) \rightarrow \min$  with initial point  $X^0 = x_1^0, \dots, x_n^0$  the optimization procedure:

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

delivers the quality not better, than optimization procedure:

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

that is:  $S(\hat{x}_1, \dots, \hat{x}_n) \leq S(\tilde{x}_1, \dots, \tilde{x}_n)$ .

Proof:  $\forall x_i \neq \tilde{x}_i : S(\tilde{x}_1, \dots, x_i, \dots, \tilde{x}_n) \leq S(\tilde{x}_1, \dots, \tilde{x}_i, \dots, \tilde{x}_n)$ .  $\square$

A(N):

for  $i = 1, \dots, N$ :

$$\hat{x}_1 = \operatorname{argmin}_{x_1} \text{step } Q(\hat{x}_1 \sqcup \dots \sqcup \hat{x}_n)$$

...

for  $i = 1, \dots, N$ :

$$\hat{x}_n = \underset{x_n}{\operatorname{argmin}} \text{step } Q(\hat{x}_1 \sqcup \dots \sqcup \hat{x}_n).$$

The numerical optimization procedure  $A(N)$  for solving the task  $Q(x_1, \dots, x_n) \rightarrow \min$ , is not better, than: for  $j = 1, \dots, M$ :  $A(\frac{N}{M})$ .

### 3 segmentation

$$\chi \in [a, b]$$

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

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

$$h_s^*, p_s^*, w^* = \underset{h_s, p_s, w}{\operatorname{argmin}} Q(h_s, p_s, w | D, \mathcal{A}_t, \mathcal{L}),$$

Проблема 5: Как я понимаю, такая оптимизационная задача не стоит глобально. Это локальная задача в этом куске про сегментацию. Нужно ли ее здесь ставить?

old one:

$$\begin{aligned} \text{count}_q &= \frac{1}{m} \sum_{j \in \mathcal{L}} [\chi_j \in [p_s^{q-1}; p_s^q]], \\ \text{woe}_q &= \log \frac{g_j}{b_j} = \log \frac{\sum_{j \in \mathcal{L}} [\chi_j \in [p_s^{q-1}; p_s^q]] [y_j=1]}{\sum_{j \in \mathcal{L}} [\chi_j \in [p_s^{q-1}; p_s^q]] [y_j=0]}, \\ S_s &= [\text{count}_q \geq 0.04] \quad [|\text{woe}_q - \text{woe}_{q-1}| \geq 0.1]. \end{aligned}$$

Проблема 6: Нет напрямую оптимизации этого функционала. Просто идет алгоритм выбора этих значений так, чтобы они удовлетворяли требованиям. Как это писать?

new one:

for iteration = 1, \dots, N :

for  $\chi \in X$  :

$$p_s^i = p_s^{i-1} - \lambda \nabla Q(p_s)$$

$$p_s^i = \text{Correction}(p_s^i)$$

$$/ \text{Correction}(p_s) : \text{удаляется } p_s^q, \text{ если } p_s^q - p_s^{q-1} < 2 /$$

Проблема 7: Это - алгоритм подбора параметров. Нужно ли здесь его приводить?

### 4 grouping

$$p_g : C \rightarrow h_g$$

$$\begin{array}{ccccccc}
\chi & = & 1 & 2 & 3 & \dots & C \\
& & \downarrow & \downarrow & \downarrow & & \downarrow \\
\chi_h & = & \gamma_1 & \gamma_2 & \gamma_3 & \dots & \gamma_C
\end{array}
\quad C - \text{categories number}$$

$|h_g|$  — new categories number,  $\gamma_i \in h_g$   
 $h_g^*, p_g^*, w^* = \underset{h_g, p_g, w}{\operatorname{argmin}} Q(h_g, p_g, w | D, \mathcal{A}_c, \mathcal{L}),$

Проблема 8: Как я понимаю, такая оптимизационная задача не стоит глобально. Это локальная задача в этом куске про группировку. Нужно ли ее здесь ставить?

old one:

$$\begin{aligned}
count_c &= \frac{1}{m} \sum_{j \in \mathcal{L}} [\chi_j = c], \\
woe_c &= \log \frac{g_j}{b_j} = \log \frac{\sum_{j \in \mathcal{L}} [\chi_j = c][y_j = 1]}{\sum_{j \in \mathcal{L}} [\chi_j = c][y_j = 0]}, \\
S_g &= [count_c \geq 0.04] \quad \prod_{i,j} [|woe_i - woe_j| \geq 0.1].
\end{aligned}$$

new one:

```

for iteration = 1, ..., N :
  for χ ∈ X :
    for j = 1, ..., N :
      A append crossing(aχi, aχk), where i, k are random
      A = select M best of A
      pg = select one best of A

```

## 5 feature construction

$$\begin{aligned}
p_f &= \{\mathbb{F}_i\} \\
h_f^*, p_f^*, w^* &= \underset{h_f, p_f, w}{\operatorname{argmin}} Q(h_f, p_f, w | D, \mathcal{A}_l \sqcup \mathcal{A}_b, \mathcal{L}),
\end{aligned}$$

new one:

```

for iteration = 1, ..., N :
  F append crossing(ai, ak), где i, k — случайные
  F = choose M best of F
  pf = choose one best of F

```