

Задача обучения с экспертной информацией *

А. В. Грабовой¹, В. В. Стрижов²

Аннотация: Ключевые слова: смесь экспертов; байесовский выбор модели; априорное распределение.

DOI: 00.00000/0000000000000000

1 Введение

2 Постановка задачи

Пусть задано множество объектов Ω , а также подмножество наблюдаемых объектов Ω'

$$\Omega' \subset \Omega, \quad (2.1)$$

где $|\Omega'| = N$. Пусть для Ω задана некоторая экспертная информация $E(\Omega)$.

На основе экспертной информации $E(\Omega)$ введем отображения из множества объектов Ω :

$$K_y^{E(\Omega)} : \Omega \rightarrow \mathbb{R}, \quad K_x^{E(\Omega)} : \Omega \rightarrow \mathbb{R}^n, \quad (2.2)$$

где n количество признаков, причем предполагаем, что $n \ll N$. Применив отображения $K_x^{E(\Omega)}$ и $K_y^{E(\Omega)}$ для множества наблюдаемых объектов Ω' получаем выборку:

$$\mathfrak{D}(\Omega', E(\Omega)) = \{(\mathbf{x}, y) \mid \mathbf{x} = K_x^{E(\Omega)}(\omega), y = K_y^{E(\Omega)}(\omega), \forall \omega \in \Omega'\}, \quad (2.3)$$

Предполагается, что существуют нетривиальные отображения $K_y^{E(\Omega)}, K_x^{E(\Omega)}$, и $\mathbf{w} \in \mathbb{R}^n$, такие, что:

$$y \approx \mathbf{x}^T \mathbf{w}, \quad \forall (\mathbf{x}, y) \in \mathfrak{D}(\Omega', E(\Omega)), \quad (2.4)$$

то есть получаем выборку, которая является задачей линейной регрессии по нахождению неизвестного вектора \mathbf{w} (аналогично можно ввести задачу для логистической регрессии).

*Работа выполнена при поддержке РФФИ и правительства РФ.

¹Московский физико-технический институт, grabovoy.av@phystech.edu

²Московский физико-технический институт, strijov@ccas.ru

В случае, когда экспертная информация представляется в виде объединения нескольких экспертов:

$$\mathbf{E}(\Omega) = \mathbf{E}_0(\Omega) \cup \mathbf{E}_1(\Omega_1) \cup \mathbf{E}_2(\Omega_2) \cup \dots \cup \mathbf{E}_K(\Omega_K), \quad \bigcup_{i=1}^K \Omega_i = \Omega \quad (2.5)$$

в этом случае будем говорить о задаче смеси K экспертов. Каждая информация $\mathbf{E}_k(\Omega_k)$ описывает локальную информацию о каком-то подмножестве объектов Ω_k для всех $k = 1 \dots K$. Информация эксперта $\mathbf{E}_0(\Omega)$ описывает глобальную информацию о всем множестве объектов Ω .

В случае задачи смеси K экспертов вводятся отображения:

$$\begin{aligned} K_y^{\mathbf{E}_1(\Omega_1)} : \Omega \rightarrow \mathbb{R}, \quad K_x^{\mathbf{E}_1(\Omega_1)} : \Omega \rightarrow \mathbb{R}^{n_1}, \\ \dots \\ K_y^{\mathbf{E}_K(\Omega_K)} : \Omega \rightarrow \mathbb{R}, \quad K_x^{\mathbf{E}_K(\Omega_K)} : \Omega \rightarrow \mathbb{R}^{n_K}, \end{aligned} \quad (2.6)$$

где получаем множество отображений во множество локальных моделей, в которых учтены информации от каждого эксперта.

Также, как в и задаче одного эксперта вводятся предположения, что каждая локальная модель является линейной:

$$\forall k \in \{1, \dots, K\}, \quad y \approx \mathbf{x}^T \mathbf{w}_k, \quad \forall (\mathbf{x}, y) \in \mathfrak{D}(\Omega'_k, \mathbf{E}(\Omega_k)). \quad (2.7)$$

Заметим, что истинного разбиения Ω на множества $\{\Omega_k\}_{k=1}^K$ нету. Рассмотрим вектор функцию π :

$$\pi : \Omega \rightarrow \mathbb{R}^K, \quad \sum_{k=1}^K \pi_k(\omega) = 1, \quad \forall \omega \in \Omega, \quad (2.8)$$

где π назовем шлюзовой функцией.

Предположим, что все $\left\{ \left(K_x^{\mathbf{E}_k(\Omega_k)}, K_y^{\mathbf{E}_k(\Omega_k)} \right) \right\}_{k=1}^K$ являются заданными отображениями. Используя локальные модели, построим глобальную мультимодель, которая описывает все множество объектов Ω :

$$\sum_{\omega \in \Omega'} \sum_{k=1}^K \pi_k(\omega, \mathbf{V}) \left(K_y^{\mathbf{E}_k(\Omega_k)}(\omega) - \mathbf{w}_k^T K_x^{\mathbf{E}_k(\Omega_k)}(\omega) \right)^2 + R(\mathbf{V}, \mathbf{W}, \mathbf{E}(\Omega)) \rightarrow \min_{\mathbf{V}, \mathbf{W}} \quad (2.9)$$

где $\mathbf{W} = [\mathbf{w}_1^T, \dots, \mathbf{w}_K^T]$, а $R(\mathbf{V}, \mathbf{W}, \mathbf{E}(\Omega))$ является некоторой регуляризацией параметров, которая также основывается на экспертной информации, \mathbf{V} — параметры шлюзовой функции.

3 Пример

Список литературы

- [1] Yuksel Seniha Esen, Wilson Joseph N., Gader Paul D Twenty Years of Mixture of Experts // IEEE Transactions on Neural Networks and Learning Systems. 2012. Issues. 23, No 8. pp. 1177–1193.

- [2] *Bishop C.* Pattern Recognition and Machine Learning. — Berlin: Springer, 2006. 758 p.