Задача обучения с экспертной информацией *

А. В. Грабовой[1], **В. В. Стрижов**[2]

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

DOI: 00.00000/00000000000000

1 Введение

2 Постановка задачи

Пусть задано множество объектов $\mathbf{\Omega}$, а также подмножество наблюдаемых объектов $\mathbf{\Omega}'$

$$\mathbf{\Omega}' \subset \mathbf{\Omega}, \tag{2.1}$$

где $|\mathbf{\Omega}'| = N$. Пусть для $\mathbf{\Omega}$ задана некоторая экспертная информация $\boldsymbol{E}(\mathbf{\Omega})$.

На основе экспертной информации $\boldsymbol{E}(\mathbf{\Omega})$ введем отображения из множества объектов $\mathbf{\Omega}$:

$$K_y^{\boldsymbol{E}(\mathbf{\Omega})} : \mathbf{\Omega} \to \mathbb{R}, \quad K_x^{\boldsymbol{E}(\mathbf{\Omega})} : \mathbf{\Omega} \to \mathbb{R}^n, \tag{2.2}$$

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

$$\mathfrak{D}\left(\mathbf{\Omega}', \boldsymbol{E}(\mathbf{\Omega})\right) = \{(\mathbf{x}, y)\, |\mathbf{x} = K_x^{\boldsymbol{E}(\mathbf{\Omega})}(\omega),\ y = K_y^{\boldsymbol{E}(\mathbf{\Omega})}(\omega),\ \forall \omega \in \mathbf{\Omega}'\}, \tag{2.3}$$

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

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

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.

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

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

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

$$\boldsymbol{E}(\boldsymbol{\Omega}) = \boldsymbol{E}_0(\boldsymbol{\Omega}) \cup \boldsymbol{E}_1(\boldsymbol{\Omega}_1) \cup \boldsymbol{E}_2(\boldsymbol{\Omega}_2) \cup \cdots \cup \boldsymbol{E}_K(\boldsymbol{\Omega}_K), \quad \cup_{i=k}^{K}\boldsymbol{\Omega}_k = \boldsymbol{\Omega} \tag{2.5}$$

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

В случае задачи смеси K экспертов вводятся отображения:

$$\begin{aligned} K_y^{\boldsymbol{E}_1(\boldsymbol{\Omega}_1)} : \boldsymbol{\Omega} \to \mathbb{R}, \quad K_x^{\boldsymbol{E}_1(\boldsymbol{\Omega}_1)} : \boldsymbol{\Omega} \to \mathbb{R}^{n_1}, \\ \cdots \\ K_y^{\boldsymbol{E}_K(\boldsymbol{\Omega}_K)} : \boldsymbol{\Omega} \to \mathbb{R}, \quad K_x^{\boldsymbol{E}_K(\boldsymbol{\Omega}_K)} : \boldsymbol{\Omega} \to \mathbb{R}^{n_K}, \end{aligned} \tag{2.6}$$

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

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

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

Заметим, что истинного разбиения $\boldsymbol{\Omega}$ на множества $\{\boldsymbol{\Omega}_k\}_{k=1}^{K}$ нету. Рассмотрим вектор функцию $\boldsymbol{\pi}$:

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

где $\boldsymbol{\pi}$ назовем шлюзовой функцией.

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

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

где $\mathbf{W} = [\mathbf{w}_1^{\mathsf{T}}, \cdots, \mathbf{w}_K^{\mathsf{T}}]$, а $R(\mathbf{V}, \mathbf{W}, \boldsymbol{E}(\boldsymbol{\Omega}))$ 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.