Отбор признаков в задаче обучения распознаванию образов при смешении решающего правила

Турков П.А. Красоткина О.В.

Тульский государственный университет

10-я Международная конференция «Интеллектуализация обработки информации»

Крит 2014

Обучающее множество

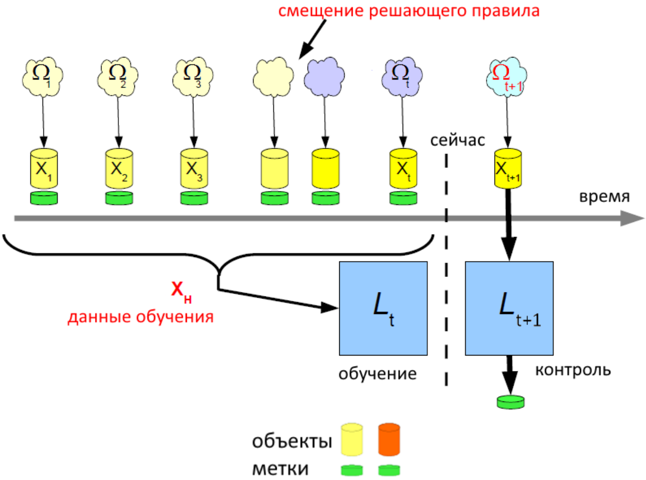
- Пусть каждый объект генеральной совокупности $\omega \in \Omega$ имеет скрытую характеристику $y \in \{1, -1\}$ и представлен точкой в линейном признаковом пространстве $\mathbf{x}(\omega) = \big(x^1(\omega), \dots, x^n(\omega)\big) \in \mathbb{R}^n$.

- Потоковый характер данных требует рассмотрения каждого объекта ω вместе с моментом времени t, когда он поступает. В результате обучающее множество приобретает вид

$$\{(\mathbf{X}_t,\ \mathbf{Y}_t,\ t)\}_{t=1}^{T},$$

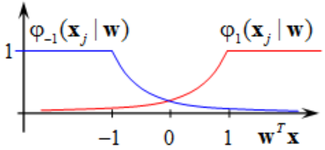
$(\mathbf{X}_t, \mathbf{Y}_t) = \{(\mathbf{x}_{k,t}, y_{k,t})\}_{k=1}^{N_t}$ - подмножество объектов, поступивших в момент времени t.

Обучение при смещении решающего правила
Постановка задачи [1]



Модель генеральной совокупности будем понимать как дискриминантную функцию следующего вида: $f\left(\mathbf{x}(\omega)\right) = \mathbf{w}^T\mathbf{x} > 0$ если $y(\omega) = 1$, и < 0 если $y(\omega) = -1$.

Апостериорные вероятности принадлежности объекта к одному из двух классов описываются как:



$$\varphi_1(\mathbf{x}|\mathbf{w}) = \begin{cases} 1, \mathbf{w}^T\mathbf{x} > 1, \\ \exp\left[-c\left(1 - \mathbf{w}^T\mathbf{x}\right)\right], \mathbf{w}^T\mathbf{x} < 1, \end{cases}$$

$$\varphi_{-1}(\mathbf{x}|\mathbf{w}) = \begin{cases} 1, \mathbf{w}^T\mathbf{x} < -1, \\ \exp\left[-c\left(1 + \mathbf{w}^T\mathbf{x}\right)\right], \mathbf{w}^T\mathbf{x} > -1. \end{cases}$$

Выберем совместное распределение параметров разделяющей гиперплоскости и их дисперсии в виде

$$\psi(w_i|r_i) = \frac{1}{\sqrt{2\pi r_i}} \exp(-w_i^2/r_i)$$

В качестве совместного распределения для дисперсий используем гамма-распределение:

$$\begin{aligned}\gamma\big(1/r_i|\alpha, 1/\beta\big) =& \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \Big(\frac{1}{r_i}\Big)^{\alpha-1} \cdot \exp\Big(-\frac{\beta}{r_i}\Big) \\ \propto& \Big(\frac{1}{r_i}\Big)^{\alpha-1} \cdot \exp\Big(-\frac{\beta}{r_i}\Big)\end{aligned}$$

Нестационарные параметры гиперплоскости

- Смещение решающего правила приводит к необходимости понимания параметра $\mathbf{w}$ как функции времени: $\mathbf{w}_t : T \to \mathbb{R}^n$
- Апостериорная плотность распределения классов $y_{j,t} = \pm 1$:

$$\varphi(y_{t,j}|\mathbf{x}_{t,j}, \mathbf{w}_t) = \begin{cases} 1, y_{t,j}\mathbf{w}_t^T\mathbf{x}_{t,j} > 1, \\ \exp\left[-c\left(1 - y_{t,j}\mathbf{w}_t^T\mathbf{x}_{t,j}\right)\right], y_{t,j}\mathbf{w}_t^T\mathbf{x}_{t,j} < 1, \end{cases}$$

- Для обучающего подмножества $\mathbf{X}_t, \mathbf{Y}_t$, полученного в момент времени t совместная плотность распределения:

$$\Phi(\mathbf{Y}_t|\mathbf{X}_t, \mathbf{w}_t) = \prod_{j=1}^{N_t} \varphi(y_{t,j}|\mathbf{x}_{t,j}, \mathbf{w}_t).$$

Нестационарные параметры гиперплоскости

- Будем понимать зависящие от времени параметры решающего правила, как случайные стационарные марковские процессы

$$\mathbf{w}_t = q\mathbf{w}_{t-1} + \boldsymbol{\xi}_t, M\left(\boldsymbol{\xi}_t\right) = \mathbf{0}, M(\boldsymbol{\xi}_t\boldsymbol{\xi}_t^T) = d\mathbf{I}$$

  где дисперсия d определяет динамику изменений в генеральной совокупности.

- Априорная плотность распределения скрытой последовательности параметров решающего правила:

$$\Upsilon(\mathbf{w}_t, t = 2, \ldots, T) = \prod_{t=2}^{T} \left[\mathcal{N}\left(\mathbf{w}_t | \sqrt{1-d}\,\mathbf{w}_{t-1}, d\mathbf{I}\right)\right]$$

Апостериорная плотность совместного распределения последовательности параметров решающего правила и обучающей совокупности:

$$P\big((\mathbf{w}_t, \hat{\mathbf{r}})_{t=1}^{T} | (\mathbf{X}_t, \mathbf{Y}_t)_{t=1}^{T}\big) \propto$$

$$\propto \Upsilon(\mathbf{w}_t, t = \overline{2, T}) \prod_{t=1}^{T} \Phi(\mathbf{Y}_t | \mathbf{X}_t, \mathbf{w}_t)$$

$$\prod_{i=1}^{n} \gamma\big(1/r_i | \alpha, 1/\beta\big) \cdot \prod_{t=1}^{T} \prod_{i=1}^{n} \psi(w_{t,i} | r_i)$$

Оценка оптимальных значений параметров решающего правила:

$$(\hat{\mathbf{w}}_t, \hat{\mathbf{r}})_{t=1}^{T} = \arg \max_{\mathbf{w}_t, t=1,\cdot,T} P\big((\mathbf{w}_t, \hat{\mathbf{r}})_{t=1}^{T} | (\mathbf{X}_t, \mathbf{Y}_t)_{t=1}^{T}\big)$$

Критерий обучения

$$\left(\hat{\mathbf{w}}_t, \hat{\mathbf{r}}, \hat{\boldsymbol{\delta}}_t | (\mathbf{x}_j, y_j), \alpha, \beta, j = 1, ..., N_t, t = 1, ..., T\right) =$$

$$\arg\max_{\mathbf{w},\mathbf{r}} \Bigg[-\sum_{t=1}^{T}\sum_{i=1}^{n} w_{t,i}^2/r_i - C\sum_{t=1}^{T}\sum_{i=1}^{N_t}\delta_{t,j} -$$

$$-\frac{1}{2}\sum_{i=1}^{n}\log r_i - (\alpha-1)\sum_{i=1}^{n}\log r_i - \sum_{i=1}^{n}\frac{\beta}{r_i} -$$

$$-\frac{1}{2d}\sum_{t=2}^{T}(\mathbf{w}_t - \sqrt{1-d}\,\mathbf{w}_{t-1})^T(\mathbf{w}_t - \sqrt{1-d}\,\mathbf{w}_{t-1}) \Bigg] \quad (1)$$

при ограничениях

$$1 - y_{t,j}\mathbf{w}_t^T\mathbf{x}_{t,j} < \delta_{t,j},$$

$$\delta_{t,j} > 0, j = 1, \ldots, N_t, t = 1, ..., T$$

Оценка параметров решающего правила

Для оценки параметров воспользуемся методом покоординатного спуска, тогда при зафиксированном $\mathbf{r}$ получаем следующую задачу оптимизации

$$\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{w_{t,i}^2}{r_i} + C\sum_{t=1}^{T}\sum_{j=1}^{N_t}\delta_{t,j} +$$

$$+\frac{1}{2d}\sum_{t=2}^{T}(\mathbf{w}_t - \sqrt{1-d}\mathbf{w}_{t-1})^T(\mathbf{w}_t - \sqrt{1-d}\mathbf{w}_{t-1}) \to \min_{\mathbf{w}_t,\boldsymbol{\delta}_t,t=1,\ldots T}$$

с ограничениями

$$1 - y_{t,j}\mathbf{w}_t^T\mathbf{x}_{t,j} < \delta_{t,j},$$
$$\delta_{t,j} > 0, j = 1, \ldots, N_t, t = 1, \ldots, T$$

Оценка параметра селективности

Для оценивания значения гиперпараметра $\mathbf{r}$ запишем критерий (1) в двойственной форме:

$$L\left(\mathbf{w}_t, \mathbf{r}, \boldsymbol{\delta}_t, \lambda_t, \mu_t\right) =$$

$$\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{w_{t,i}^2}{r_i} + C\sum_{t=1}^{T}\sum_{j=1}^{N_t}\delta_{t,j} + (\alpha - \frac{1}{2})\sum_{i=1}^{n}\log r_i + \sum_{i=1}^{n}\frac{\beta}{r_i} +$$

$$+ \frac{1}{2d}\sum_{t=2}^{T}(\mathbf{w}_t - \sqrt{1-d}\mathbf{w}_{t-1})^T(\mathbf{w}_t - \sqrt{1-d}\mathbf{w}_{t-1}) +$$

$$+ \sum_{t=1}^{T}\sum_{j=1}^{N_t}\lambda_{t,j}(1 - y_{t,j}\mathbf{w}_t^T\mathbf{x}_{t,j} - \delta_{t,j}) - \sum_{t=1}^{T}\sum_{j=1}^{N_t}\mu_{t,j}\delta_{t,j}$$

Частные производные по целевым переменным

Откуда следует

$$r_i = \frac{\sum_{t=1}^{T} w_{t,i}^2 + \beta}{\alpha - 1/2}$$

Искусственные данные
Описание данных

- Каждый момент времени двумя нормальными двумерными распределениями с математическими ожиданиями 1 и -1, соответственно, генерируется по 25 объектов каждого класса
- К каждому объекту добавляется еще 98 нормально распределенных значений с математическим ожиданием 0 и дисперсией 1;
- С течением времени центры генерирующих распределений поворачиваются на 0.0314 радиана.

Вектор дисперсий

Номер признака	1	2	3	...	100
Дисперсия	47,6407	48,1681	0,7355	...	0,0052

Список литературы I

Zliobaite I., Learning under Concept Drift: an Overview, http://zliobaite.googlepages.com/Zliobaite_CDoverview.pdf

Частные производные двойственной задачи

$$\frac{\partial L\left(\mathbf{w}, \mathbf{r}, \delta, \lambda, \mu\right)}{\partial w_{t,i}} = 2\sum_{t=1}^{T} w_{t,i}/r_i + \frac{1}{d}\sum_{t=1}^{T} w_{t,i} - \frac{\sqrt{1-d}}{d}\sum_{t=1}^{T} w_{t-1,i} - $$

$$- \sum_{t=1}^{T}\sum_{j=1}^{N_t} \lambda_{t,j} y_{t,j} \mathbf{x}_{t,j,i}$$

$$\frac{\partial L\left(\mathbf{w}, \mathbf{r}, \boldsymbol{\delta}, \boldsymbol{\lambda}, \boldsymbol{\mu}\right)}{\partial r_i} = -\sum_{t=1}^{T} w_{t,i}^2/r_i^2 + (\alpha - \frac{1}{2})\frac{1}{r_i} - \frac{\beta}{r_i^2}$$

$$\frac{\partial L\left(\mathbf{w}, \mathbf{r}, \boldsymbol{\delta}, \boldsymbol{\lambda}, \boldsymbol{\mu}\right)}{\partial \delta_{t,j}} = \lambda_{t,j} + \mu_{t,j} - C$$

Вернуться