

第05章：深入浅出ML之Bayes-Based家族

By ZhouYong

🕒 发表于 2015-11-27

- author: zhouyongsdzh@foxmail.com
- date: 2015-11-15
- weibo: @周永_52ML

内容列表

- 写在前面
- 判别模型与生成模型
- 高斯判别分析
- 朴素贝叶斯算法

文章目录

1. 写在前面
2. 判别模型与生成模型
3. 高斯判别分析
4. 朴素贝叶斯
 - 4.1. 朴素贝叶斯模型
 - 4.2. 拉普拉斯平滑
 - 4.3. NB示例：文本分类
5. Next ...
6. 参考资料

写在前面

本章虽然是谈贝叶斯，但是先不从贝叶斯开始说起。先介绍有监督学习中两大类经典的模型：判别模型与生成模型。

判别模型与生成模型

实际上，机器学习的有监督学习的所有模型大体上可以划分为两大类：一类是判别模型，另一类是生成模型。

- 判别模型

判别模型主要是根据特征值求结果的概率，形式化表示为 $P(y|x; w)$ （比如LR模型）。在参数 w 确定的情况下，求解条件概率 $P(y|x)$ 。

举例：

比如要确定一只羊是山羊还是绵羊，用判别模型的方法是：先从历史数据中学习模型，然后通过提取这只羊的特征来预测出这只羊是山羊的概率和绵羊的概率（即条件概率 $P(y|x)$ ）。

常见的判别模型：

线性回归、对数回归、线性判别分析、SVM、Boosting、CRF、神经网络等

- 生成模型

延续判别模型中的例子，换一种思路考虑：

我们可以根据山羊的特征，首先学习出一个山羊模型；然后根据绵羊的特征学习出一个绵羊模型。然后从某一只羊中提取特征，放到山羊模型中看概率是多少，再放到绵羊模型中看概率是多少，哪个大就是哪个。

上面描述形式化表示为 $P(x|y)P(y)$ （包括 $P(y)$ ）， y 是模型结果， x 是特征（向量）。利用贝叶斯公式可以发现两个模型的关联性：

$$P(y|x) = \frac{P(x, y)}{P(x)} = \frac{P(x|y) \cdot P(y)}{P(x)} \quad (ml.1.5.1)$$

$$P(y|x) = \frac{P(x, y)}{P(x)} = \frac{P(x|y) \cdot P(y)}{P(x)} \quad (ml.1.5.1)$$

在生成模型里面，由于我们关系的是 y 的离散值结果中哪个概率大（比如山羊概率和绵羊概率哪个大），而不是关心具体的概率，因此上式可改写为：

$$\arg \max_y P(y|x) = \arg \max_y \frac{P(x|y)P(y)}{P(x)} \quad (1) \quad (n. ml.1.5.1)$$

$$= \arg \max_y P(x|y)P(y) \quad (2)$$

$$\arg \max_y P(y|x) = \arg \max_y \frac{P(x|y)P(y)}{P(x)} \quad (1) \quad (n. ml.1.5.1)$$

$$= \arg \max_y P(x|y)P(y) \quad (2)$$

其中， $P(y)$ 称为先验概率， $P(y|x)$ 是后验概率， $P(x|y)$ 为似然函数。公式(1)等价于(2)是因为对于离散值 y 来说， $P(x)$ 都是一样的。

常见的生成模型：

隐马尔可夫模型、朴素贝叶斯模型、高斯混合模型、主题模型（LDA等）、受限波尔斯曼机等

- 联系与区别

- 由于 $P(x, y) = P(x|y) \cdot P(y)$ ，判别模型求的是条件概率，生成模型求的是联合概率。

1. 概率论的乘法法则
2. 流派之争：频率学派与贝叶斯学派
3. 先验概率与后验概率

高斯判别分析

这里在介绍朴素贝叶斯算法之前，先介绍一下高斯判别分析模型 (Gaussian Discriminant Analysis)。

- 多变量正态分布

多变量正态分布（又称多值正态分布）描述的是 n 维随机变量的分布情况。单变量正态分布中的期望 μ 在这里变成了向量， σ 也变成了矩阵 Σ ，记作 $N(\mu, \Sigma)$ 。

假设有 n 个随机变量 X_1, X_2, \dots, X_n 。 μ 的第 i 个分量是 $E(X_i)$ ，其中 $\Sigma_{ii} = Var(X_i)$ ， $\Sigma_{ij} = Cov(X_i, X_j)$ 。多变量概率分布对应的概率密度函数如下：

$$P(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} \cdot |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) \quad (ml.1.5.2)$$

$$P(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} \cdot |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) \quad (ml.1.5.2)$$

其中 $|\Sigma|$ 是 Σ 的行列式， Σ 是协方差矩阵，并且是对称半正定的。

(二维空间示意图)

注： μ 决定中心位置， Σ 决定投影椭圆的朝向和大小。

- 高斯判别分析模型

如果输入特征 x 是连续性随机变量，那么可以使用高斯判别分析模型来确定 $P(x|y)$ 。 $P(x|y)$ 。模型如下：

$$\begin{aligned} y &\sim \text{Bernoulli}(\phi) \\ x|y=0 &\sim \mathcal{N}(\mu_0, \Sigma) \\ x|y=1 &\sim \mathcal{N}(\mu_1, \Sigma) \end{aligned} \quad (ml.1.5.3)$$

$$\begin{aligned} y &\sim \text{Bernoulli}(\phi) \\ x|y=0 &\sim \mathcal{N}(\mu_0, \Sigma) \\ x|y=1 &\sim \mathcal{N}(\mu_1, \Sigma) \end{aligned} \quad (ml.1.5.3)$$

其中，输出结果 y 服从伯努利分布（ ϕ 为参数）；在给定的模型下，特征（向量）符合多变量高斯分布。

回到上面山羊与绵羊的例子，在山羊模型下，它的体征如胡须长度、角大小、毛长度等连续型变量符合高斯分布，他们组成的特征向量符合多变量高斯分布。

下面给出该模型中涉及到的概率密度函数：

$$\begin{aligned} P(y) &= \phi^y (1 - \phi)^{1-y} \\ P(x|y=0) &= \frac{1}{(2\pi)^{\frac{n}{2}} \cdot |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu_0)^T \Sigma^{-1}(x - \mu_0)\right) \\ P(x|y=1) &= \frac{1}{(2\pi)^{\frac{n}{2}} \cdot |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu_1)^T \Sigma^{-1}(x - \mu_1)\right) \end{aligned} \quad (ml.1.5.4)$$

$$\begin{aligned} P(y) &= \phi^y (1 - \phi)^{1-y} \\ P(x|y=0) &= \frac{1}{(2\pi)^{\frac{n}{2}} \cdot |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu_0)^T \Sigma^{-1}(x - \mu_0)\right) \\ P(x|y=1) &= \frac{1}{(2\pi)^{\frac{n}{2}} \cdot |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu_1)^T \Sigma^{-1}(x - \mu_1)\right) \end{aligned} \quad (ml.1.5.4)$$

最大似然估计（对数似然）公式：

$$\begin{aligned}
\ell(\phi, \mu_0, \mu_1, \Sigma) &= \log \prod_{i=1}^m P(x^{(i)}, y^{(i)}; \phi, \mu_0, \mu_1, \Sigma) \\
&= \log \prod_{i=1}^m P(x^{(i)} | y^{(i)}; \mu_0, \mu_1, \Sigma) \cdot P(y^{(i)}; \phi)
\end{aligned} \tag{ml.1.5.5}$$

m 表示样本数，这里有两个 μ ，表示在不同的结果模型下，特征均值不同。但我们假设协方差相同。反映在图形上表现为：不同模型中心位置不同，但形状相同。这样就可以用直线来进行分隔判别。

将(ml.1.5.4)带入(ml.1.5.5)，求导后，得到参数估计公式：

$$\begin{aligned}
\phi &= \frac{1}{m} \sum_{i=1}^m 1\{y^{(i)} = 1\} & (1) \\
\mu_0 &= \frac{\sum_{i=1}^m 1\{y^{(i)} = 0\} x^{(i)}}{\sum_{i=1}^m 1\{y^{(i)} = 0\}} & (2) \\
\mu_1 &= \frac{\sum_{i=1}^m 1\{y^{(i)} = 1\} x^{(i)}}{\sum_{i=1}^m 1\{y^{(i)} = 1\}} & (3) \\
\Sigma &= \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu_{y^{(i)}})(x^{(i)} - \mu_{y^{(i)}})^T & (4)
\end{aligned} \tag{ml.1.5.6}$$

其中， ϕ 是训练样本中结果 $y = 1$ 的样本占比； μ_0 和 μ_1 分别表示 $y = 0$ 和 $y = 1$ 的样本中特征均值； Σ 是样本特征方差均值。

(高斯判别分析模型 示意图)

• 高斯判别分析（GDA）与LR的关系

如果将GDA用条件概率方式表示的话，可表示为：

$$P(y = 1 | x; \phi, \mu_0, \mu_1, \Sigma) \tag{n.ml.1.5.2}$$

y 是 x 的函数，其中 $\phi, \mu_0, \mu_1, \Sigma$ 都是参数。我们进一步推导，可得：

$$P(y = 1 | x; \phi, \mu_0, \mu_1, \Sigma) = \frac{1}{1 + \exp(-w^T x)} \tag{n.ml.1.5.3}$$

这里 w 是 $\phi, \mu_0, \mu_1, \Sigma$ 的函数。恰好公式(n.ml.1.5.3)是LR模型的表达式。

总结如下：如果 $P(x|y)$ 符合多变量高斯分布，那么 $P(y|x)$ 符合Logisti回归模型。反之不成立，因为GDA有着更强的假设条件和约束。

如果我们认定训练数据满足多变量高斯分布，那么GDA能够在训练集上是最好的模型。然而，我们往往事先并不知道训练数据满足什么样的分布，不能做很强的假设。**Logistic回归的条件假设要弱于GDA，因此更多时候采用LR方法。**

例如，训练数据满足泊松分布，

$x|y=0 \sim \text{Poisson}(\lambda_0)$; $x|y=1 \sim \text{Poisson}(\lambda_1)$ 。那么可以得到 $P(y|x)$ 也是Logistic回归模型。（此时如果采用GDA，效果就会比较差，因此训练样本的特征分布是泊松分布，非多变量高斯分布。）

朴素贝叶斯

朴素贝叶斯模型

在高斯判别分析中，我们要求特征向量 x 是连续型的实数向量。如果 x 是离散型向量的话，可以考虑采用朴素贝叶斯的分类方法。

- 多项式分布（Multinomial Distribution）

这里以经典的邮件分类（文本分类的一种应用）为例，假如邮件要分为两类：垃圾邮件和正常邮件。采用最简单的特征描述方法，首先要找一个词库，然后将每封邮件中的内容表示成一个向量（根据词库进行过滤后），向量中每一维都是词库中的一个词的0/1值（1表示该词在词库中出现，0表示未出现）。

比如一封邮件中出现了“的”和“打折”，没有出现“通知”，“提交”，“预约”等，那么形式化表示为：

\$\$

$\phi(x) =$

$\begin{bmatrix}$

$\left[$

```
\begin{array}{cc}
1 \\
0 \\
\vdots \\
1 \\
\vdots \\
0
\end{array}
```

\right] \; \leftarrow \;

\left[

```
\begin{array}{cc}
  的\\
  提交 \\
  \vdots \\
  打折 \\
  \vdots \\
  通知 \\
\end{array}
```

\right]

\end{equation} \quad (\text{exp.ml.1.5.1})

\$\$

假设词库中有10000个词，那么 x 是10000维的。

此时如果要建立多项式分布模型，对该问题来说，就是把每封邮件看作一次随机试验，那么结果的可能性就有 x^{10000} 种。(意味着参数 p_i 有 x^{10000} 个，参数太多，不可能直接用该思路建模。)

多项式分布：

假设某随机试验有 k 个可能结果 X_1, X_2, \dots, X_k ，它们的概率分布分别是 p_1, p_2, \dots, p_k 。在 N 次采样的总结果中， X_1, X_2, \dots, X_k 出现的次数分别为 n_1, n_2, \dots, n_k ($N = \sum_{i=1}^k n_i$)。那么这次事件出现的概率 P 可表示为（公式中 x_i 代表出现 n_i 次）：

$$P(X_1 = n_1, \dots, X_k = n_k) = \begin{cases} \frac{N!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} & \text{当 } N = \sum_{i=1}^k n_i \text{ 时,} \\ 0 & \text{其它} \end{cases}$$

- 朴素贝叶斯假设

我们先换一种思路，我们要求的是 $P(y|x)$ ，根据生成模型定义，可以求 $P(x|y)$ 和 $P(y)$ 。假设 x 中的特征都是独立的，这个作为朴素贝叶斯假设。

特征之间相互条件独立是朴素贝叶斯模型的前提条件。如果一封邮件是垃圾邮件 $y = 1$ ，且这封邮件出现词“buy”与这封邮件是否出现“price”无关，那么“buy”和“price”之间是**条件独立**的。

在给定 Z 的情况下， X 和 Y 条件独立，形式化表示为：

$$P(X|Z) = P(X|Y, Z) \quad (n.ml.1.5.5)$$

也可以表示为：

$$P(X, Y|Z) = P(X|Z) \cdot P(Y|Z) \quad (n.ml.1.5.6)$$

那么对于一条样本来说，有下式成立：

$$\begin{aligned} P(x_1, \dots, x_{10000}) &= P(x_1|y)P(x_2|y, x_1)P(x_3|y, x_1, x_2) \cdots P(x_{10000}|y, x_1, x_2, \dots, \\ &= P(x_1|y)P(x_2|y) \cdots P(x_{10000}|y) \\ &= \sum_{i=1}^n P(x_i|y) \end{aligned} \quad (ml.1.5.7)$$

公式(ml.1.5.7)与NLP中的n-gram语言模型非常类似，这里相当于1-gram。同时也可以看出，朴素贝叶斯假设是约束性很强的假设。通常来说，“buy”与“price”是有关系的，而在NB这里，假设条件独立。

注：条件独立与独立是两个不同的概念。

• 朴素贝叶斯模型 (Naive Bayes Model)

建立朴素贝叶斯的形式化模型：

$$\begin{aligned} \phi_{j|y=1} &= P(x_j = 1|y = 1) \quad (1) \\ \phi_{j|y=0} &= P(x_j = 1|y = 0) \quad (2) \\ \phi_y &= P(y = 1) \quad (3) \end{aligned} \quad (ml.1.5.8)$$

而我们想要的是模型在所有训练样本上的概率累积最大，即最大似然估计：

$$\mathcal{L}(\phi_y, \phi_{i|y=0}, \phi_{i|y=1}) = \prod_{i=1}^m P(x^{(i)}, y^{(i)}) \quad (ml.1.5.9)$$

注意：这里是联合概率分布积最大（非LR中的条件概率积），说明朴素贝叶斯是省城模型。

公式(ml.1.5.9)求解，可得：

$$\phi_{j|y=1} = \frac{\sum_{i=1}^m 1\{x_j^{(i)} = 1 \wedge y^{(i)} = 1\}}{\sum_{i=1}^m 1\{y^{(i)} = 1\}} \quad (1)$$

$$\phi_{j|y=0} = \frac{\sum_{i=1}^m 1\{x_j^{(i)} = 1 \wedge y^{(i)} = 0\}}{\sum_{i=1}^m 1\{y^{(i)} = 0\}} \quad (2) \quad (ml.1.5.10)$$

$$\phi_y = \frac{\sum_{i=1}^m 1\{y^{(i)} = 0\}}{m} \quad (3)$$

(3)式表示 $y=1$ 的样本数在全部样本中的占比。前两个分别表示在 $y = 1$ 和 $y = 0$ 的样本中，特征 $x_j = 1$ 的比例。

然而我们最终要求的是：

$$\begin{aligned} P(y = 1|x) &= \frac{P(x|y = 1)P(y = 1)}{P(x)} \\ &= \frac{\left(\prod_{j=1}^n P(x_j|y = 1)\right) P(y = 1)}{\left(\prod_{j=1}^n P(x_j|y = 1)\right) P(y = 1) + \left(\prod_{j=1}^n P(x_j|y = 0)\right) P(y = 0)} \end{aligned}$$

> 实际上，只要求分子即可，分母对于 $y=1$ 和 $y=0$ 都一样。

>

> 朴素贝叶斯模型可以扩展到 x 和 y 都有多个值的情况。对于特征是连续的情况，可采用分段的方法将连续值转化为离散值。这涉及到了特征离散化的一些技术，比如等分、等差、信息增益、基尼指数等。（该部分在第11章：特征工程方法中会详细介绍。）

拉普拉斯平滑

朴素贝叶斯方法有一个致命的缺点就是对数据的稀疏问题过于敏感。

比如，5.2.1节中介绍的邮件分类，加入现在又来了一封邮件，邮件中出现“NIPS call for papers”。其中NIPS在现有的词典库中不存在（10000个词）。现在我们使用更大的词典库来分类（词的数目由10000变为50000），假设NIPS这个词在词典中的位置是50000。而这个词在训练数据中未曾出现过，那么此时算条件概率的结果是：

$$\begin{aligned} \phi_{50000|y=1} &= \frac{\sum_{i=1}^m 1\{x_{50000}^{(i)} = 1 \wedge y^{(i)} = 1\}}{\sum_{i=1}^m 1\{y^{(i)} = 1\}} = 0 \\ \phi_{50000|y=0} &= \frac{\sum_{i=1}^m 1\{x_{50000}^{(i)} = 1 \wedge y^{(i)} = 0\}}{\sum_{i=1}^m 1\{y^{(i)} = 0\}} = 0 \end{aligned} \quad (exp. ml.1.5.2)$$

由于NIPS在以前的所有邮件中，故而其对应的条件概率为0。带入公式(*ml.1.5.11*)，得到最终的条件概率也为0。

上述导致条件概率为0的主要原因是朴素贝叶斯的假设－特征之间条件独立，而模型又实用相乘的方式得到结果。为解决此问题，也为训练样本中未出现的特征值，赋予一个较小的值（非0）。具体方法如下：

假设离散随机变量 x 有 $\{1, 2, \dots, k\}$ 个值，这里用 $\phi_i = P(x = i)$ 来表示每个值的概率。在 m 个训练样本中， x 的观察值表示为 $\{x^{(1)}, x^{(2)}, \dots, x^{(k)}\}$ 。其中每个观察值对应 k 个值中的一个，那么根据之前的方法可以得到：

$$\phi_j = \frac{\sum_{i=1}^m 1\{x^{(i)} = j\}}{m} \quad (exp. ml.1.5.3)$$

即 $x = j$ 出现的比例。

拉普拉斯平滑法(Laplace Smooth)将每个 k 值出现次数事先都加1，通俗地讲就是假设他们都出现过一次，那么修改后的表达式为：

$$\phi_j = \frac{\sum_{i=1}^m 1\{x^{(i)} = j\} + 1}{m + k} \quad (ml.1.5.12)$$

每个 $x = j$ 的分子都加1，分母加 k 。有 $\sum_{j=1}^k = 1$ 成立。那么在邮件分类问题中，修改后的公式为：

$$\phi_{j|y=1} = \frac{\sum_{i=1}^m 1\{x_j^{(i)} = 1 \wedge y^{(i)} = 1\} + 1}{\sum_{i=1}^m 1\{y^{(i)} = 1\} + 2} \quad (1) \quad (n. ml.1.5.7)$$

$$\phi_{j|y=0} = \frac{\sum_{i=1}^m 1\{x_j^{(i)} = 1 \wedge y^{(i)} = 0\} + 1}{\sum_{i=1}^m 1\{y^{(i)} = 0\} + 2} \quad (2)$$

NB示例：文本分类

上面介绍的邮件分类其实就是一个典型的文本分类的例子，用于该文本分类的朴素贝叶斯模型又称作多变量伯努利事件模型。

邮件生成过程描述：

（第1步）首先随机选定了邮件的类型（垃圾邮件与正常邮件，即 $P(y)$ ），（第2步）然后观察词库中的每一个词，随机决定是否要出现在邮件中，出现则标记为1，否则标记为0。然后将出现的词组成一封邮件。决定一个词是否出现时根据概率 $P(x_i|y)$ ，那

么一封邮件的概率可表示为：

$$\text{一封邮件的生成概率} = P(x, y) = P(y) \cdot \prod_{i=1}^n P(x_i | y) \quad (exp.ml.1.5.3)$$

注意：这里的 n 是词典中词的数目； x_i 表示一个词是否出现，只有两个值0和1，概率之和为1； x 向量是长度为 n 的0/1值。

现在我么换一种思路，不先从词典入手，而是选择从邮件入手。让 (i) 表示邮件中的第 i 个词， x_i 表示这个词在词典中的位置，那么 x_i 的取值范围是 $\{1, 2, \dots, |V|\}$ （ $|V|$ 是字典中词的数目）。这样，一封邮件可以表成 n 维向量 (x_1, x_2, \dots, x_n) （ n 为邮件中词的个数，可以变化，因为每封邮件长度可不同）。

这就相当于重复投掷 $|V|$ 面的骰子，将观察值纪录下来就形成了一封邮件。当然每个面的概率服从 $P(x_i | y)$ ，而且每次实验条件独立，这样我们得到的邮件概率为 $P(y) \prod_{i=1}^n P(x_i | y)$ 。

注意：这里的公式虽然与共公式 $(exp.ml.1.5.3)$ 相同，但符号意义完全不同。本式中的 n 表示邮件中词的数目； x_i 表示 $|V|$ 中的一个值； x 向量中元素是词典中的位置。

形式化表示如下：

m 个训练样本： $\{(x^{(i)}, y^{(i)}); i = 1, \dots, m\}$ ；第 i 条样本表示为：
 $x^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_{n_i}^{(i)})$ ，第 i 个样本有 n_i 个词。

我们仍然按照朴素贝叶斯的方法求最大似然估计，似然函数表示为：

$$\begin{aligned} \mathcal{L}(\phi, \phi_{i|y=0}, \phi_{i|y=1}) &= \prod_{i=1}^m P(x^{(i)}, y^{(i)}) \\ &= \prod_{i=1}^m \left(\prod_{j=1}^{n_i} P(x_j^{(i)} | y^{(i)}; \phi_{i|y=0}, \phi_{i|y=1}) \right) \cdot P(y^{(i)}; \phi_y) \end{aligned} \quad (n.m)$$

求解结果为：

$$\begin{aligned} \phi_{k|y=1} &= \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} 1\{x_j^{(i)} = k \wedge y^{(i)} = 1\}}{\sum_{i=1}^m 1\{y^{(i)} = 1\} \cdot n_i} \\ \phi_{k|y=0} &= \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} 1\{x_j^{(i)} = k \wedge y^{(i)} = 0\}}{\sum_{i=1}^m 1\{y^{(i)} = 0\} \cdot n_i} \\ \phi_y &= \frac{\sum_{i=1}^m 1\{y^{(i)} = 1\}}{m} \end{aligned}$$

与之前的分子相比，分母多了 n_i ，分子由0/1变成了 k .

Next ...

- 贝叶斯网络
- 概率图模型

参考资料

- [判别模型、生成模型与朴素贝叶斯方法](#)
- 微博: [@JerryLead](#)

更多信息请关注: [计算广告与机器学习 – CAML 技术共享平台](#)

📁

深入浅出机器学习

🏷️

Gaussian Discriminant Analysis

Naive Bayes



上一篇:

◀ [第06章：深入浅出ML之Boosting家族](#)

下一篇:

▶ [第04章：深入浅出ML之Kernel-Based家族](#)

分类

OpenMIT ³

分布式机器学习 ¹

强化学习与智能决策 ⁴

概率与统计 ²

深入浅出机器学习 ⁹

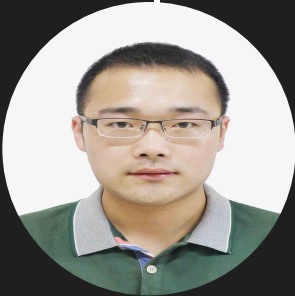
标签

- 计算广告学²
- Agent²
- 参数服务器²
- Dynamic Programming¹
- DP¹
- Policy Evalation¹
- Policy Improvement¹
- Policy Iteration¹
- Value Iteration¹
- Greedy Policy¹
- 连续随机变量¹
- MDP¹
- Markov Decision Process¹
- RL¹
- Environments¹
- GD¹
- FTRL¹
- AdaGrad¹
- AdaDelta¹
- Adam¹

友情链接

码农圈
Jark's Blog

 RSS 订阅



Hello, Welcome to CAML technology sharing platform.
I'm Zhou Yong, engaged in algorithms work on computational advertising and machine learning.

