第5卷第2期
1994年2月

软件学报
JOURNAL OF SOFTWARE

Vol. 5, No. 2
Feb. 1994

线性规划在知识获取中的应用*

陈建华　李定坤

（福州大学计算机科学系，福州 350002）

摘要　本文提出一个利用线性规划来实现从实例归纳学习的实用方法，简称为PKA方法. 该方法适用于许多分类专家系统的知识获取.

关键词　专家系统，知识获取，线性规则.

知识获取一直是发展专家系统的一个主要瓶颈. 因此机器学习的研究日益引起人们的兴趣[1-3]. 对许多分类问题领域专家往往只能告诉知识工程师少量关于分类的不完全且模糊的知识，单凭这些知识建造的专家系统很难向用户提供良好的建议. 然而领域专家们实践的大量成功实例和极少数失败的例子隐含着他们的集体智慧和经验教训. 于是如何从大量实例中归纳抽取出规律性的知识，即进行有效的机器学习，是成功地建造这类专家系统的关键.

对于符号机器学习，目前已提出一些从实例学习的方法和系统. 其中最著名的是Michalski的AQ11，Quinlan的ID3和洪家荣的AQ15、GS、AE1和AE5. 它们已用于一定范围的问题并已收到良好效果. 不过由于实际问题的复杂性，现有方法远不能满足发展专家系统的需要. 一些分类问题专家系统和统计模式识别系统常用判别函数作为主要分类器. 确定判别函数有以下两种主要方法[4]：

1. Bayes方法. 对输入的特征向量X，此法以各类的后验概率$P(\omega_i|X)$作为判别函数，取函数值最大的那个判别函数所对应的类ω_i作为X应属于的类，即$X\sim\omega_i$. 据Bayes公式，为确定$P(\omega_i|X)$，须知各类的先验概率$P(\omega_i)$和各ω_i关于X的似然函数$P(X|\omega_i)$. 通常估计$P(\omega_i)$的困难不大. 对于$P(X|\omega_i)$如能预测它的函数形式，如假定是正态分布形式，那么只要通过对实例集（样本集）的统计计算便可估计出均值向量μ_i和协方差矩阵Σ_i，从而确定了$P(X|\omega_i)\sim N(\mu_i,\Sigma_i)$[4]. 这些公式的意义见第2节. Bayes法的优点是它有较坚实的概率统计理论基础. 但对特征向量X维数较大的复杂情况用单一的正态分布函数来表示判别函数并不一定符合实际情况.

2. 一般函数法. 该法不用概率统计概念，而是人为地假设判别函数的形式. 常用广义线性判别函数$G_i(X)=\sum_{j=1}^{r}C_{ij}Y_{ij}(X)$. 其中特征函数$Y_{ij}(X)$是特征向量的任意函数，是预先假定

* 本文1991年4月29日收到，1991年10月8日定稿

本文得到福建省自然科学基金资助. 作者陈建华，35岁，讲师，1990年硕士毕业于福州大学，主要研究领域为智能CAD. 李定坤，56岁，教授，主要研究领域为智能CAD.

本文通讯联系人：陈建华，福州350002，福州大学计算机科学系

的;C_{ij}是系数,由实例进行调整确定.若用最小二乘法总可以通过实例集导出伪逆矩阵[4],从而定出权系数C_{ij}.该法优点是可由具体领域知识灵活地假定特征函数的形式和个数.但在考虑特征函数的形式时随意性大,缺少较坚实的理论基础.

对这类从实例归纳学习以确定判别函数的分类问题,在其实例集中成功的实例即正例往往占绝大多数;失败的实例即反例一般是极少数;同时在正例集中往往有一部分特别成功的实例即典型正例.通常要求通过归纳学习所得到的判别函数能完全描述典型正例并完全排除反例,否则该系统就很难被用户接受.然而上述两个主要方法都未能满足这一要求.数学规划已成功地应用于许多领域[5-7],但在知识获取中的应用还不多见.本文提出一个利用线性规划实现从实例归纳学习的实用方法.该法力图吸取Bayes法和一般函数法的优点而克服它们无法完全描述典型正例和完全排除反例的弱点.为便于说明本法的思想和实现步骤,我们用一经简化的高层建筑结构体系选型的例子(以下简称"结构选型")贯穿以下各节.

1 方法的主要思想

设高层建筑的结构体系[8]有m类,如框架,框——剪,剪力墙,…,等等.依次记它们为$\omega_1,\omega_2,\cdots,\omega_m$.用$\Omega$表示它们的集合,即$\Omega=\{\omega_1,\omega_2,\cdots,\omega_m\}$.设影响结构体系选择的主要因素有n个,如,建筑物总高,地震烈度,场地土类别、层数、底层有否大空间要求,…,等等,依次记它们为$x_1,x_2,\cdots,x_n$,并组成如下特征向量:

$$X=\{x_1,x_2,\cdots,x_n\}^T$$

对一输入的特征向量X,系统的任务就是建议一类合理的结构体系.

知识工程师从领域专家和有关资料中得到的知识,虽然还不足以使系统向用户提供一个良好的建议,但这些知识对于建造该系统毕竟是十分宝贵的.对一输入的特征向量X,通常可根据上述知识先从类型集Ω中排除一些类,即对具体这一工程这些类肯定不会采用.如以规则表示这些知识则有如下形式:

IF 层数大于25 THEN 不采用框架类

IF 底层有大空间要求 THEN 不采用剪力墙类

………

设经过这样排除后剩下的类型全体记为Ω_X,接着用判别函数法从Ω_X中选出合理的类ω_i,即$X\sim\omega_i$,记号"~"表示对应.这一分类过程可用图1表示.

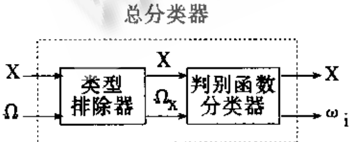

图1

给每一类型定义一个判别函数,其形式为广义线性函数:

$$G_i(X)=\sum_{j=1}^{r}C_{ij}Y_{ij}(X)\qquad i=1,2,\cdots,m \tag{1.1}$$

这样选取可吸取一般函数法具有灵活性的优点.在构造$Y_{ij}(X)$时,将结合领域知识尽可能吸取Bayes法的优点;然后用线性规划确定权系数C_{ij},以保证归纳所得到的类型函数能完全满足描述典型正例和排除反例的要求.

2　$Y_{ij}(X)$的构造和定义

在构造和定义 $Y_{ij}(X)$时，首先应根据具体领域的特点和已有的、显式的领域知识细心确定特征函数的个数及它们的自变量.形式上，$Y_{ij}(X)$与各特征分量有关，而实质上应使各 Y_{ij}只与少数几个特征分量有关；某个特征分量可能在多个 Y_{ij}的自变量表中出现.

在"结构选型"例子中，分别给每类定义 r 个特征函数，即 $Y_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,r)$.为说明构造这些函数的基本思想，我们较具体地讨论其中的一个函数.如，从领域专家处得知每类结构体系关于建筑物总高的适用范围，一般都与地震烈度和场地土类别有密切关系.于是对每一类都可构造这样一个特征函数 $Y_{i1}(x_1,x_2,x_3)$.其中 x_1 是总高，x_2 是地震烈度，x_3 是场地土类别.定义：

$$Y_{i1}(x_1,x_2,x_3)=P(x_1,x_2,x_3|\omega_i) \qquad i=1,2\cdots,m$$

式中 $P(x_1,x_2,x_3|\omega_i)$是各类关于(x_1,x_2,x_3)的似然函数.一般可假定它们是正态分布形式，通过对实例库的统计计算可得出均值向量 μ_i 和协方差矩阵 Σ_i，从而确定 $P(x_1,x_2,x_3|\omega_i)\sim N(\mu_i,\Sigma_i)$[4].

设 X_o 是由 V 个特征分量构成的向量，则其似然函数的正态分布形式为：

$$P(X_o|\omega)=\frac{1}{(2\pi)^{v/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}d^2\right]$$

式中
$$d^2=(X_o-\mu)^T\Sigma^{-1}(X_o-\mu)$$

其中 μ 是 V 维均值向量，Σ 是协方差矩阵.设相应的实例有 s 个，各实例的 X_o 值分别为 $X_o^{(1)},X_o^{(2)}\cdots,X_o^{(s)}$，那么均值向量 μ 和协方差矩阵 Σ 的估计值为：

$$\mu=\frac{1}{S}\sum_{k=1}^{s}X_o^{(k)},\Sigma=\frac{1}{S}\sum_{k=1}^{s}(X_o^{(k)}-\mu)(X_o^{(k)}-\mu)^T$$

从以上公式可知，为定义 Y_{ij}，必须有一个足够丰富的实例库.若一个实例用一条记录表示，那么每条记录至少应包含 n+2 个字段的值.这些字段是："结构体系类别"、"成功程度"，"x_1"，"x_2"，…，"x_n".其中"成功程度"字段的取值为'特别成功'，'成功'或'失败'.在估计均值向量和协方差矩阵时，只对'特别成功'和'成功'的实例进行统计计算.至于如何从极少数'失败'的实例中吸取教训的问题将在第 3 节中讨论.

3　C_{ij}的确定

一旦 Y_{ij}被定义后剩下的任务就是确定权系数 C_{ij}.可在领域专家的帮助下从实例库中精选出一批典型正例.在"结构选型"例子中，典型正例就是'特别成功'的工程.这些工程或是由著名的、经验丰富的工程师所设计，或是经同行专家评议认为特别成功的，等等.我们总希望所归纳学习得到的判别函数能完全描述这些典型正例.设第 i 类有 α_i 个典型正例，其相应

的特征矢量为$X_i^{(1)}, X_i^{(2)}, \cdots, X_i^{(\alpha_i)}$. 为了使判别函数能完全描述它们，应使得：

$$\sum_{j=1}^{r} C_{ij} Y_{ij}(X_i^{(k)}) \geqslant \sum_{j=1}^{r} C_{lj} Y_{lj}(X_i^{(k)}) \tag{3.1}$$

$$i=1,2,\cdots,m;\quad l=1,2,\cdots,m,\quad l\neq i;\quad k=1,2,\cdots,\alpha_i$$

因Y_{ij}均由正例统计来定义所以要求权系数为正数，即

$$C_{ij}\geqslant 0 \qquad i=1,2,\cdots,m;\quad j=1,2,\cdots,r \tag{3.2}$$

若用一正数乘(3.1)式，则不等式关系不变. 因此为了唯一确定C_{ij}，不妨要求

$$\sum_{i=1}^{m}\sum_{j=1}^{r} C_{ij} = m\times r \tag{3.3}$$

在满足上述所有约束的情况下，我们总希望在典型正例中能将相应最佳类型的函数值尽可能地与其他类型的函数值拉开来. 为此引入一个新的未知变量$\varepsilon\geqslant 0$，并用下式代替(3.1)：

$$\sum_{j=1}^{r} C_{ij} Y_{ij}(X_i^{(k)}) - \sum_{j=1}^{r} C_{lj} Y_{lj}(X_i^{(k)}) - \varepsilon \geqslant 0 \tag{3.4}$$

$$i=1,2,\cdots,m;\quad l=1,2,\cdots,m,\quad l\neq i;\quad k=1,2,\cdots,\alpha_i$$

于是得到如下线性规划：

$$\left.\begin{array}{l} \max\quad \varepsilon \\ \text{S.t.}\begin{cases} \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{r} C_{ij} = m\times r \\ \displaystyle\sum_{j=1}^{r} C_{ij} Y_{ij}(X_i^{(k)}) - \sum_{j=1}^{r} C_{lj} Y_{lj}(X_i^{(k)}) - \varepsilon \geqslant 0 \\ \qquad i=1,2,\cdots,m; l=1,2,\cdots,m, l\neq i; k=1,2,\cdots,\alpha_i \\ C_{ij}\geqslant 0, \varepsilon\geqslant 0 \qquad i=1,2,\cdots,m; j=1,2,\cdots,r \end{cases}\end{array}\right\} \tag{3.5}$$

对上述线性规划，现已有不少成熟算法，如，单纯形法[5]. 一旦从(3.5)解出C_{ij}和ε，判别函数便被完全确定，从而建立了初始系统.

值得指出的是，对每一典型正例，其相应的特征向量$X_o^{(k)}$应经过“类型排除器”检查，那些被排除的类型就不必放在(3.1)中进行比较. 这样可大幅度地缩小线性规划的规模，从而节省内存和计算时间.

下面讨论如何处理反例. 设有一反例其特征向量为$X^{(*)}$，而其失败的类型为ω_*. 把$X^{(*)}$输入到刚建立起来的初始系统，将有两种可能的结果：一是初始系统给出一个区别于ω_*的好建议ω_g，即系统能够排除这一反例，这时我们构造一个“设想的典型正例”$X^{(*)}\sim\omega_g$；另一是系统给出的建议恰好也是ω_*，即系统不能排除这一反例. 此时应取出判别函数值次大者所对应的类型ω_f，并构造一个“设想的典型正例”$X^{(*)}\sim\omega_f$. 如，有多个反例，则对每个都

进行上述的检查处理.若初始系统能排除所有反例，那么该系统就已是所期望的系统.若有些反例不能被初始系统排除，那么应进一步完善初始系统.办法是将由反例所产生的“设想典型正例”集与真正的典型正例集一起均看作典型正例，重新按前述方法建立新的线性规划(3.5)，并求解出 C_{ij} 和 ε 的新值.将新值代入判别函数表达式(1.1)，便得到所期望的系统.从而可开始向用户提供咨询.

4　实　例

本文方法已应用到“高层建筑结构体系选型电脑咨询系统”中.对某个将要兴建的高层建筑，只要用户输入影响结构体系选型的主要因素(即特征向量)，系统便可给用户建议一个合适的结构体系类型.类型集 Ω 中含有：框架(ω_1)，框——剪(ω_2)，剪力墙(ω_3)，框肢——剪力墙(ω_4)，框筒(ω_5)，筒体框架(ω_6)，筒中筒(ω_7)，多筒(ω_8)，成束筒(ω_9)，多重筒(ω_{10})十种结构体系；特征向量为：建筑物总高(x_1)，地震烈度(x_2)，场地土类别(x_3)，层数(x_4)，用途(x_5)，高宽比(x_6)，底层是否大空间(x_7)，平面方形系数(x_8)，基本风压(x_9)，材料(x_{10})十个主要因素.根据领域知识对每类定义了 6 个特征函数.设 ω_g 为相应的最佳类型，则类型排除器中一部分规则如下：

IF　x_7=“Y”　THEN　$\omega_g \neq \omega_3$

IF　x_7=“N“　THEN　$\omega_g \neq \omega_4$

IF　x_4>25　　THEN　$\omega_g \neq \omega_1$

……

系统在 IBM－PC/XT 上开发，采用 Turbo－Prolog 和 Fortran 语言编程，试运行情况良好.

结束语：本文方法主要适用于以判别函数为主要分类器的分类专家系统，它有如下特点：1.留有足够的空间存放领域专家提供的显式知识.这些知识可用适当的方法(如规则)表示，加上适当的推理机从而构成第 1 节中的“类型排除器”；另在构造特征函数 Y_{ij} 时可充分利用这些知识.2.在定义判别函数时，既吸取了一般函数法的灵活性，又吸取了 Bayes 法具有概率统计理论基础的优点.3.利用线性规划方法确定权系数可保证系统能完全描述典型正例和排除反例.通过定义特征函数 Y_{ij} 为相应的似然函数从而可吸取数量很大的一般正例的经验.体现了“抓两头，计及中间”的思想.4.必须有足够数量的实例并从中精选出一批典型正例.

参考文献

1　徐家福，费宋铭.机器学习.计算机科学，1991(1).

2　Walker A. Knowledge System and Prolog. Addison－Wesley Publishing Co.，1987.

3　洪家荣，刘宁.示例式学习的两个方法及其应用.计算机科学，1990(6).

4　王飞龙.模式识别基础.武汉：湖北科技出版社，1986.

5　魏权龄等.数学规划与优化设计.北京：国防工业出版社，1984.

6　李定坤.弹性薄板小挠度弯曲的 Stoker 问题.中国科学 A 辑，1986(6).

7　钱令希.工程结构优化设计.水利电力出版社，1983.

8 包世华,方鄂华.高层建筑结构设计.北京:清华大学出版社,1985.

APPLICATION OF LINEAR PROGRAMMING TO KNOWLEDGE ACQUISITION

Chen Jianhua and Li Dingkun

(*Department of Computer Science*, *Fuzhou University*, *Fuzhou* 350002)

Abstract In this paper, a practical method by means of linear programming to implement learning by induction from examples has been developed. The method can be applied to the knowledge acquisition of many expert systems of classification.

Key words Expert system, knowledge acquisition, linear programming.