DOI: 10.7641/CTA.2017.60134

基于最大熵马尔科夫模型的绩效评价方法

朱 磊[1], 牛绿茵[2], 宋士吉[2†], 张玉利[2,3]

(1. 中国大洋协会办公室, 北京 100860; 2. 清华大学 自动化系, 北京 100084; 3. 清华大学 工业工程系, 北京 100084)

摘要: 本文提出了一种基于最大熵马尔科夫模型的绩效评价方法. 该方法采用马氏模型来定量化建模专家打分过程, 采用特征函数表征打分规则, 通过在训练集上最大化熵来获得符合专家经验的最优的打分模型. 与传统方法相比, 所提出的方法可以融合各种打分规则、专家经验和指标逻辑关系得到综合打分结果. 为了提高模型的训练和打分的效率, 本文提出了基于改进迭代算法的参数估计方法, 并利用Viterbi算法进行快速打分计算. 利用中国大洋协会绩效评价指标体系历史数据进行的仿真实验表明, 与BP神经网络方法和最大熵方法进行对比, 本文所提出的方法具有更高的打分正确率.

关键词: 绩效评价方法; 最大熵马尔科夫模型; 最大熵; 隐马尔科夫模型; Viterbi算法; 改进迭代算法

中图分类号: TP273 **文献标识码**: A

Performance evaluation based on maximum entropy Markov model

ZHU Lei[1], NIU Lü-yin[2], SONG Shi-ji[2†], ZHANG Yu-li[2,3]

(1. China Ocean Mineral Resources R & D Association, Beijing 100860, China;
2. Department of Automation, Tsinghua University, Beijing 100084, China;
3. Department of Industrial Engineering, Tsinghua University, Beijing 100084, China)

Abstract: This paper presents a new performance evaluation method based on the maximum entropy Markov model, which quantifies the process of scoring as a Markov process, represents the scoring rules by characteristic functions and obtains the optimal model parameters by maximizing the maximum entropy over a training sample set. Compared with other traditional methods, this method has the ability to combine complex scoring rules, expert experience with logical connection of the evaluated items to get comprehensive evaluation results. To improve the efficiency of training and scoring, this paper adopts the improved iterative scaling algorithm to obtain near-optimal model parameters and uses the Viterbi algorithm to quickly calculate the final evaluation results. The proposed method has been applied in the history data of China Ocean Mineral Resources R&D Association's evaluation system for simulation. The experimental results show that this method has higher accuracy compared with BP networks and the classical maximum entropy model.

Key words: performance evaluation method; maximum entropy Markov model; maximum entropy methods; hidden Markov models; Viterbi algorithm; improved iterative scaling algorithm

1 引言(Introduction)

项目的绩效考核, 是指在整个生命周期中, 运用一定的评价方法、量化的指标以及恰当的评价标准, 对项目进行整体评价. 该评价过程不仅能对项目进行有效监管, 还能为部门的下一步项目规划提供决策依据, 因此具有十分重要的作用.

大洋协会作为绩效评价项目的试点工作, 依照公平、客观、科学的原则, 建立了一套科学合理的绩效评价体系. 该体系涵盖了项目的整个周期, 包括40个四级指标、19个三级指标、6个二级指标、4个一级指标和1个综合指标, 该体系的树状图如图1所示.

绩效评价方法, 就是在已有的绩效评级体系上, 通过项目的各个末级指标得分, 求得各个高级指标得分, 并最终获得综合得分的过程. 为满足绩效考核的公平、公正、科学、客观的原则, 该评价算法也应满足经验性、客观性和全局性.

目前, 有很多的方法可以用于项目的绩效评价上, 大致可分为主观方法和客观方法. 主观方法多以专家

收稿日期: 2016−03−13; 录用日期: 2016−12−29.
†通信作者. E-mail: shijis@mail.tsinghua.edu.cn; Tel.: +86 10-62782721.
本文责任编委: 陈增强.
国家海洋局大洋专项基金项目(DYXY–125), 国家自然科学基金项目(61273233), 中国博士后科学基金特别项目(61503211), 教育部高等学校博士点基金项目(20130002130010)资助.
Supported by China Ocean Association (DYXY–125), National Natural Science Foundation of China (61273233), China Postdoctoral Science Foundation (61503211) and Specialized Research Fund for Doctoral Program of Higher Education of China (20130002130010).

的经验确定指标重要度, 从而推得指标权重, 来求得综合评分. 主要方法有主成分分析法(principal component analysis, PCA)[1]、Delphi (德尔菲) 方法[2]、层次分析法(analytic hierarchy process, AHP)[3]、决策试验和评价实验(decision making trial and evaluation laboratory, DEMATEL)方法[4]等. 例如, DEMATEL方法利用矩阵和图论的原理, 通过计算指标的中心度和原因度, 以得到指标间相对影响关系, 从而求得权重矩阵, 结合决策矩阵得到综合评分. 层次分析法则将指标分层处理, 在划分好指标层次后, 同级指标之间两两对比, 列出判断矩阵, 利用行平均法、特征值法[5]等来求得权重. 网络分析法(analytic network process, ANP)[6]为AHP的改进算法, 该方法考虑了指标间的相互联系, 更具有适用性. 客观方法则将数据本身的统计信息作为评判依据来确定指标重要度或直接对数据进行分类排序. 主要有数据包络分析(data envelopment analysis, DEA)法[7]、灰色关联度方法[8]、熵信息方法、最大离差法、模糊聚类法[9]等. 有些学者还将多种方法进行组合[11], 以求达到更加稳定的评价效果[12]. 随着机器学习的不断发展, 神经网络方法[10]因为具有非线性自适应的学习能力, 被引入项目智能化评价领域中, 此类方法利用已评价的项目数据作为训练集, 以指标分数为输入, 综合评分为输出, 训练出网络参数, 以达到代替人工打分的智能评价效果.

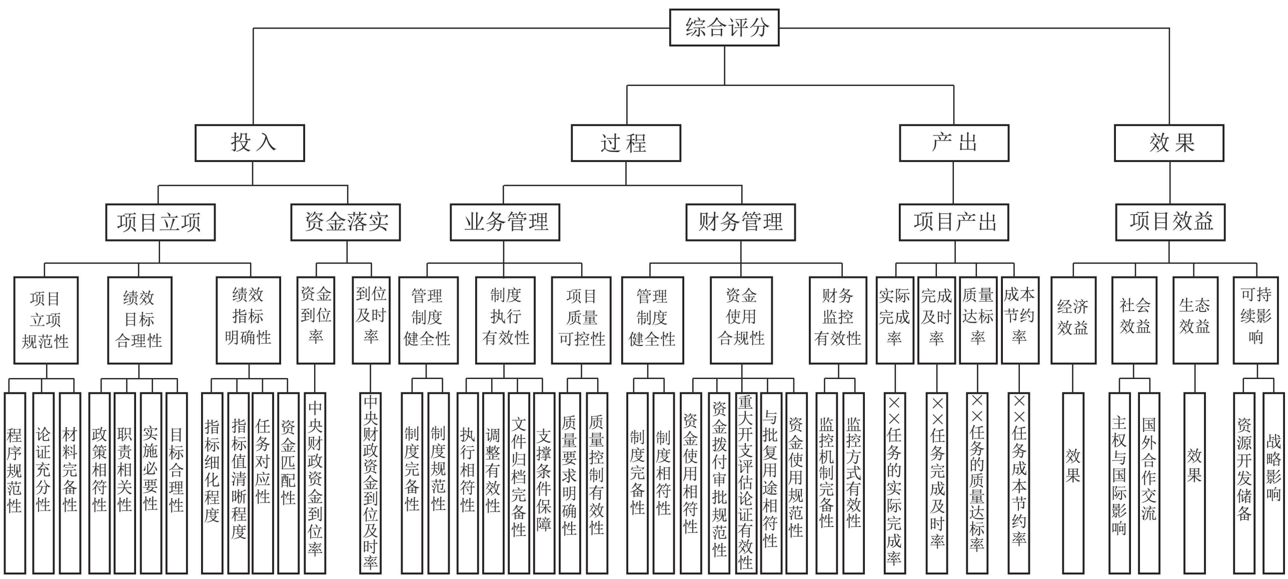

图 1 大洋协会绩效评价体系树状图

Fig. 1 Tree view of COMRA's performance appraisal system

然而上述的方法中, 以层次分析法为代表的主观方法, 是由专家偏好确定指标权重, 一方面利用表示指标重要度的方法存在一定局限性, 另一方面整体结果容易受到人为因素影响, 使结果随机性较大. 而依赖数据的客观方法, 一方面无法体现专家经验, 另一方面其所得结果也具有一定的随机性, 不同的数据或数据量都会对结果产生影响. 例如在DEA中, 数据量过小会严重影响数据有效性的判断. 组合的方法, 其实质也是不同结果的加权平均, 也无法解决极端值情况, 有可能导致波动大的非重要项成为主导项影响整体结果. 而神经网络作为典型的黑箱学习模型, 因其中间神经元指代意义不明确, 得不到评价体系的整体构架. 另外, 上述的方法都只将单一指标的重要程度作为评价规则, 均无法体现例如几个指标的综合影响、指标间对比度影响等等需要增设的复杂打分规则.

本文将评价体系中的各级指标看做随机变量进行建模, 用概率代替权重, 率先利用在基因测序和自然语言处理领域应用广泛的最大熵马尔科夫模型[17]对打分过程进行建模. 该模型一方面可以利用特征函数解释多种多样包含主客观信息的复杂打分规则, 又可以用马尔科夫性来描述同级指标之间原本的逻辑关系, 是一个可以详细描述打分过程的模型. 相比于传统方法, 该方法具有一定的优势: 1) 利用特征函数, 可以挖掘已打分项目中主观的专家打分知识和客观的统计规律, 并能够灵活的增设复杂打分规则; 2) 利用了马尔科夫性, 解释各项得分之间的逻辑关系, 能让打分过程更直观. 因此具有一定的使用环境: 1) 该方法可以习得某个专家的打分风格和直观经验, 在某位专家缺席情况下可以代替该专家进行打分; 2) 该方法打分时间短, 成本小, 可在项目数量较多, 指标繁杂, 工作量较大时, 可以代替专家打分, 减少工作量; 3) 该方法可以作为辅助决策的手段, 通过学习历史数据得到的打分器, 可以修正部分人为打分误差, 或为人工决策提供依据.

本文将通过以下4个部分来对该评价方法进行解释: 第1部分是对模型进行理论分析; 第2部分是实际

建模和方法推导; 第3部分则通过仿真实验的方法验证该模型, 并对结果进行分析; 最后一部分是结果和结论.

2 最大熵马尔科夫模型(Maximun entropy Markov model)

2.1 模型原理(Principle of the model)

为介绍该模型, 首先引入隐马尔科夫模型和最大熵模型.

隐马尔科夫模型是一个首创于20世纪70年代的统计模型[13]. 其本质上是两个随机过程的组合: 状态不可直接观测的马尔科夫过程和由不可观测的状态产生观测量的随机过程. 隐含不可观的状态可由观测值推测得到. 简易模型图如图2所示.

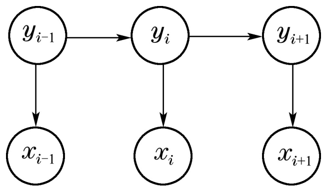

图 2 隐马尔科夫模型简易图

Fig. 2 Simple model of hidden Markov model

从图2可知. y是隐含的状态变量序列, x是观测变量序列, 将两个状态之间转移(即从y_{i-1}到y_i)的概率, 定义为条件分布概率$\mathrm{P}(y_i|y_{i-1})$, 将状态y_i产生观测x_i的概率定义为$\mathrm{P}(x_i|y_i)$. 假设初始的概率矩阵为$\mathrm{P}_0(y)$, 则产生序列x的概率为

$$\mathrm{P}(x)=\sum_{y}\prod_{i=1}^{T}\mathrm{P}(y_i|y_{i-1})\mathrm{P}(x_i|y_i). \quad (1)$$

利用隐马尔科夫模型中的观测概率和条件分布概率, 可以通过学习表征出上、下级的打分规则和同级评分指标之间的客观联系, 能得到已知观测的情况下满足最大观测概率和最大条件分布概率的序列, 即已知低级指标的情况下, 满足打分规则和同级指标间关系的最合理的高级评分. 然而, 其观测量需独立的假设条件, 即不能存在几个指标相互作用的打分规则, 限制了该模型在实际中的应用.

最大熵模型是以最大熵原理为基本原理的统计学模型[14]. 即在无偏假设下, 满足已知部分成立的未知变量的最不确定的一种推断.

为建立最大熵模型, 要做以下几个准备工作:首先, 引入信息熵的定义, 表征信息的不确定性:

$$H(X)=E(I(X))=-\sum_{x\in X}\mathrm{P}(x)\log\mathrm{P}(x), \quad (2)$$

其次: 引入特征函数$f(x,y)$来表征规则, $f(x,y)$是一个二值函数, 表示当x,y满足某一事实时, 该特征函数值为1, 否则为0.

$$f_i(x,y)\in 0,1,\ i=1,2,\cdots,m. \quad (3)$$

针对一个大小为T的训练数据集$D=\{(x,y)\}$, 可通过数据本身的统计信息得到特征函数的经验期望和估计出的条件概率参数计算特征函数的模型期望.

$$\tilde{E}(f_i)=\frac{1}{n}\sum_{x,y}\mathrm{P}(x,y)f_i(x,y), \quad (4)$$

$$E(f_i)=\frac{1}{n}\sum_{x,y}\mathrm{P}(x)\mathrm{P}(y|x)f_i(x,y). \quad (5)$$

通过调节条件概率分布P, 可以得到不同的模型期望. 能使经验期望与模型期望相等的条件概率分布可构成集合C:

$$C=\{\mathrm{P}|E_p(f_i)=\tilde{E}_p(f_i),\ i=1,2,3,\cdots,m\}. \quad (6)$$

在集合C中, 能找到一组条件概率分布, 使整个模型满足最大熵的原理, 即为最优的分布. 由此, 可将最大熵模型转化为一个解凸函数的约束优化问题:

$$\begin{cases}\max\limits_{P\in C}H(\mathrm{P})=-\sum\limits_{x,y}\mathrm{P}(x)\mathrm{P}(y|x)\log\mathrm{P}(y|x),\\ \text{s.t.}\begin{cases}E_p(f_i)=\tilde{E}_p(f_i),\ i=1,2,3,\cdots,m,\\ \sum\limits_{y}\mathrm{P}(y|x)=1.\end{cases}\end{cases} \quad (7)$$

普通的梯度方法不能合理地解出这样的问题, 采用拉格朗日对偶原理来将这个有约束的优化问题变成一个无约束的优化问题, 即

$$\begin{aligned}\varLambda(p,\tilde{\lambda})=&H(y|x)+\sum_{i=1}^{m}\lambda_i(E_p(f_i)-\tilde{E}_p(f_i))+\\&\lambda_{m+1}(\sum_{y\in Y}p(y|x)-1),\end{aligned} \quad (8)$$

$$\begin{aligned}\varLambda(p,\tilde{\lambda})=&-\sum_{x,y}\tilde{p}(x)p(y|x)\log p(y|x)+\\&\sum_{i=1}^{m}\lambda_i(\sum_{x,y}p(x,y)f_i(x,y)-\\&\sum_{x,y}p(x)p(y|x)f_i(x,y))+\\&\lambda_{m+1}(\sum_{y\in Y}\mathrm{p}(y|x)-1),\end{aligned} \quad (9)$$

其中: $\lambda_i, 1\leqslant i\leqslant m$是关于第1个约束的拉格朗日系数, λ_{m+1}是关于第2个约束的拉格朗日系数. 将该式子对条件概率分布$p(y|x)$求偏导, 使其结果为0, 即

$$\begin{aligned}\frac{\delta\varLambda(p,\tilde{\lambda})}{\delta p(y|x)}=&-\tilde{p}(x)(1+\log p(y|x))-\\&\sum_{i=1}^{m}\lambda_i p(x)f_i(x,y)+\lambda_{m+1}=0,\end{aligned} \quad (10)$$

求得最优解, 即

$$\begin{aligned}p_{\tilde{\lambda}}^{*}=&\exp(\sum_{i=1}^{m}\lambda_i f_i(x,y))\exp(-\frac{\lambda_{m+1}}{\tilde{p}(x)})=\\&\frac{1}{Z_{\tilde{\lambda}}(x)}\exp(\sum_{i=1}^{m}\lambda_i f_i(x,y)).\end{aligned} \quad (11)$$

该分布即为满足最大熵的条件概率分布. 其中, $Z_{\bar{\lambda}}(x)$是为了让概率分布归一化的因子, 定义为

$$Z_{\bar{\lambda}}(x) = \sum_{y\in Y} \exp(\sum_{i=1}^{m} \lambda_i f_i(x,y)). \tag{12}$$

称拉格朗日系数λ_i是模型中各个特征函数$f_i(x,y)$的参数向量, 其实际意义为: 该系数表征了每个特征函数对条件分布的影响程度, 即表征了不同打分规则对得分概率的影响程度.

最大熵模型引入了特征函数来控制模型中已知数据对未知数据的适应度和拟合度, 解决了隐马尔科夫模型对已知数据独立性假设的缺陷. 于是将最大熵模型引入到隐马尔科夫模型中, 即得到了最大熵马尔科夫模型[18]. 该模型不仅消除了变量间的独立性假设, 还有利于变量间关系的表征.

将隐马尔科夫模型的条件概率用最大熵模型中的特征函数表征, 并通过最大熵的约束条件求得最优概率, 即得到最大熵马尔科夫模型[15]. 每个未知状态的条件概率为

$$p_{y_{i-1}}(y_i|x_i) = \frac{1}{Z_{\bar{\lambda}}(x_i,y_{i-1})} \exp(\sum_{a=1}^{m} \lambda_a f_a(x_i,y_i)). \tag{13}$$

利用简单的模型图表述最大熵马尔科夫模型, 如图3所示.

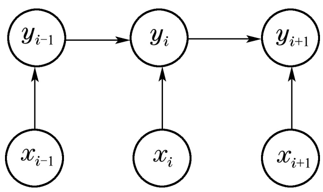

图 3 最大熵马尔科夫模型简易图

Fig. 3 Simple model of maximum entropy Markov model

在该马尔科夫链中, 当前位置的状态, 可以由前一个位置的状态和当前的观测值共同决定.

2.2 模型优势(Advantage of the model)

类比于最大熵模型, 最大熵马尔科夫模型通过引入了特征函数来表征状态与特征之间的关系, 使模型具有规则的适应性. 状态与观测之间不再仅有单一的对应关系, 而是一个最大熵的概率关系, 可以由多个关联的观测决定一个未知状态, 让模型的适用范围更广. 在绩效评价方法中, 体现为对各个规则的适应性更强, 兼顾对主观经验与客观规律的学习.

类比于隐马尔科夫模型, 最大熵马尔科夫模型又具有统计的意义, 可以从统计学的角度表述未知的状态变量之间的关系. 从而辅助了模型的规则性, 使模型更加完善. 在绩效评价中, 体现为涵盖更多客观统计规律, 表征指标间相互关系, 让模型更全面.

3 模型在绩效评价上的应用(Application of the model on performance evaluation)

3.1 建模分析(Modeling analysis)

在绩效评价体系的基础上, 可将评价体系树状图转化成易于建模的概率图. 并按层级结构对有向图进行拆分, 每一级指标与其下级指标构成一个隐马氏链. 对每一层级利用最大熵马尔科夫模型进行建模, 然后逐层计算, 就能得到各级结果和最终结果.

以通过6个二级指标得到4个一级指标这一过程为例, 绩效评价如图4所示.

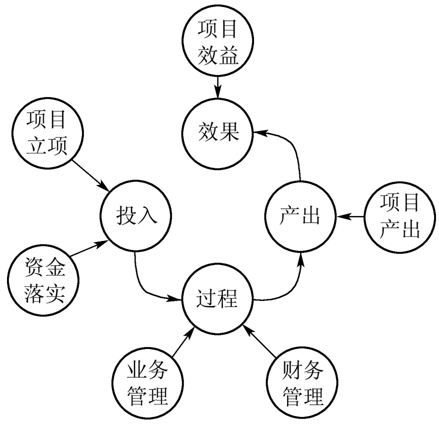

图 4 绩效评价体系图

Fig. 4 Figure of performance evaluation

二级指标得分看作为已知的观测量x, 一级指标得分为未知的隐藏状态量y. 有向图中的边长表示两个变量之间的概率即两个指标的关系. 例如, 过程得分由下级指标"业务管理"和"财务管理"的得分产生, 并受到其前一阶段"投入"得分的影响.

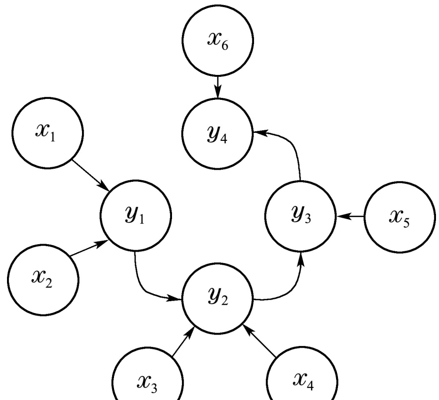

图 5 绩效评价体系模型

Fig. 5 Model of performance evaluation

规定观测量X和状态量Y为随机变量, $x_i \in \{0, 1, 2, 3, 4, 5\}(i = 1, 2, 3, 4, 5, 6)$, $y_i \in \{0, 1, 2, 3, 4, 5\}$ $(i = 1, 2, 3, 4)$表示各个指标的得分, 上下级指标的关系用条件概率$p(y_i|x_i)$表示, 同级指标假设遵循马尔科夫性, 关系用$p(y_i|y_{i-1})$ $(i = 2, 3, 4)$来表示.

利用最大熵马尔科夫模型对评价体系进行解释, 已知观测量x_i ($i=1,2,3,4,5,6$), 依据最大熵原理和马尔科夫性, 可求出每个未知状态量(上级指标)在该组观测下的满足最大熵的最优条件概率$p(y_i|x_i, y_{i-1})$, 再利用此条件概率, 可推理出状态序列Y.

为求出这一概率, 该模型引入了特征函数. 每个特征函数描述一个事件, 例如

$$f_1(x,y)=\begin{cases}1,\ x_1=5,\ y_1=5,\\ 0,\ 其他.\end{cases} \tag{14}$$

描述x_1指标得分是5, y_1指标得分是5这一事件, 或

$$f_2(x,y)=\begin{cases}1,\ x_{\max}=4,\ y_1=3,\\ 0,\ 其他.\end{cases} \tag{15}$$

描述x_1, x_2最大得分为4, 最终y_1得分为3这一事件. 每个特征函数都有一个权重系数, 来描述事件出现的概率.

依据特征函数, 针对指标y_2可列得该指标在最大熵马尔科夫模型下的条件概率为

$$p(y_2|x_3,x_4,y_1)=\\ \frac{1}{Z_{\tilde{\lambda}}(x_3,x_4,y_1)}\exp(\sum_{a=1}^{m}\lambda_a f_a(\{x_3,x_4,y_1\},y_2)), \tag{16}$$

其中m表示与y_2有关的特征函数有m个.

利用上述公式, 即可通过已知的低级指标得到熵最大的上级指标的取值概率, 从而得到概率最大的上级指标取值序列.

3.2 特征选择(Feature selection)

特征选择, 即设定合适的特征函数. 特征函数, 是最大熵马尔科夫模型对于规则的认知, 合理而易于区分的特征对于提高评价系统的性能至关重要.

本文通过对各个指标值间关系进行分析, 发现低级评分将直接影响高级评分. 与此同时, 为提高整个系统的鲁棒性、客观性, 以及运行效率, 更多的特征被考虑进来: 通过方均差来表征数据的有效性, 从而对已有的上下级得分特征进行筛选; 利用各观测值的最低分、方差等信息, 修正结果, 防止系统因为个别数据偏差产生大偏差, 提高系统的鲁棒性; 通过增加一些统计规律, 增加评分系统的客观性. 一系列分析后, 最终特征函数见表1.

表 1 特征函数设定

Table 1 Feature function selection

各项得分	最大分	最小分	平均分	方差最大项分
scole	max	min	mean	maxvar

3.3 参数估计(Parameter estimation)

参数估计, 即利用训练集估计出条件概率分布公式中各个特征函数的权值, 下面介绍两种迭代算法进行参数估计.

广义迭代算法(generalized iterative scaling, GIS)[15], 本质是利用期望逐步逼近来习得权重的一种迭代算法. 具体步骤如下:

1) 设定初始时的权值$\lambda_a^{(0)}=1,\ a\in 1,2,3,\cdots,m$.

2) 对于所有的$a\in 1,2,3,\cdots,m$:

a) 计算每个特征函数的经验期望

$$\tilde{E}(f_a)=\sum_{x\in X,y\in Y}f_a(x,y). \tag{17}$$

b) 计算每个未知变量的条件概率分布

$$p_{y_{i-1}}(y_i|x_i)=\frac{1}{Z(x_i,y_{i-1})}\exp(\sum_a\lambda_a f_a(x,y)), \tag{18}$$

其中: $p_{y_{i-1}}(y_i|x_i)$表示在前一个状态和观测给定的情况下状态的条件概率, y_{i-1}是y_i的前一个状态, $Z(x_i, y_{i-1})$是归一化因子, 且满足式子

$$Z(x_i,y_{i-1})=\sum_{y_i\in Y}\exp(\sum_{a=1}^{k}\lambda_a f_a(x,y)). \tag{19}$$

该归一化因子保证了状态y_i的概率总和为1, 确保$p_{y_{i-1}}(y_i|x_i)$是一个条件概率.

c) 计算每个特征函数f_a的期望

$$\mathrm{E}(f_a)^{(n)}=\sum_{x_i\in X,y_i\in Y}p_{y_{i-1}}(y_i|x_i)f_a(x,y). \tag{20}$$

3) 利用GIS算法更新, 利用迭代公式:

$$\lambda_a^{(n+1)}=\lambda_a^{(n)}+\frac{1}{C}\log(\frac{\tilde{E}(f_a)}{E(f_a)^{(n)}}). \tag{21}$$

4) 重复步骤2)–3), 直到$|\lambda_a^{(n+1)}-\lambda_a^{(n)}|<\varepsilon$, 得到的$\lambda_a$即为最终的参数. 可证, 此算法是收敛的.

GIS算法最终可以收敛到一个数值, 是最大熵模型中估计参数λ_a的有效方法, 但是迭代次数较多, 并不是很理想, 为了进一步降低迭代次数, 引入了改进的迭代算法, 改进的迭代尺度(improved iterative scaling, IIS)算法[16].

IIS算法的具体步骤与GIS算法相类似, 主要的区别在于对迭代公式中$\Delta\lambda_a$的计算上, 具体步骤为

1) 初始化权值: $\lambda_a^{(0)}=1, a\in 1,2,3,\cdots,m$.

2) 对于任意的$a\in 1,2,3,\cdots,m$, 求解关于$\Delta\lambda_a$的方程:

$$\tilde{E}_p(f_a)=\\ \sum_{x,y}\tilde{p}(x)p(y|x)f_a(x,y)\exp(\Delta\lambda_a\sum_{a=1}^{n}f_a(x,y)), \tag{22}$$

其中$\tilde{E}(f_a)$是每个特征函数的经验期望.

要解此方程, 又分为两种情况:

a) 当$\sum_{a=1}^{n}f_a(x,y)=M$, M为常数时,

$$\Delta\lambda_a=\frac{1}{M}\times\log(\frac{\tilde{E}(f_a)}{E(f_a)^{(n)}}). \tag{23}$$

b) 当$\sum\limits_{a=1}^{n} f_a(x,y)$非常数时, 则不能得到求$\Delta\lambda_a$的直接公式, 此时, 利用拟牛顿方法(L–BFGS)进行数值分析, 得到数值解.

3) 利用公式$\lambda_a^{(n=1)} = \lambda_a^{(n)} + \Delta\lambda_a$, 计算$\lambda_a^{(n+1)}$.

4) 重复步骤2)和3), 直到$|\lambda_a^{(n+1)} - \lambda_a^{(n)}| < \varepsilon$, 得到的$\lambda_a$即为最终的参数.

在IIS算法中, $\sum\limits_{a=1}^{n} f_a(x,y)$要小于GIS算法中的$C$, 于是, 可以明显较少迭代次数.

3.4 解码算法(Decoding algorithm)

利用GIS或者IIS, 可得模型参数, 由此可推导出在已知观测值得条件下每个隐状态的条件概率$p_{y_{i-1}}(y_i|x_i)$. 利用Viterbi算法[13]来处理本文中的解码问题, 即通过观测序列(即低级得分)和条件概率, 求得状态序列(即高级得分).

Viterbi算法是隐马尔科夫模型中常用的解码算法. 是代替枚举法的简单有效的方法. 其思路是:在任意的位置, 通过记录该位置下某一状态产生之前观测的最大概率和产生此概率时前一个位置的状态值, 就可以利用递推的关系依次求得所有位置上为某状态时的最大概率. 找到最后一个位置上的概率最大的状态值, 就可以利用之前记录的前一个位置的最优状态值推导出所有位置的最优状态值.

具体的算法步骤为: 首先, 对于一个隐马尔科夫模型, 参数为$\lambda = (A, B, \pi_0)$引入一些辅助变量: $\alpha_t(i) = p(x_1x_2\cdots x_t, y_t = s_i|\lambda)$表示第$t$位置, 观测值为$x_t$时, 状态值$y_t$处于状态$s_i$; $b_i(x_t) = p(x_t|y_t = s_i, \lambda)$表示在状态为$s_i$的条件下产生观测值为$x_t$的概率.

根据$\alpha_t(i)$和$b_i(x_t)$表示的意义, 可以求得递推关系为

$$\begin{cases} \alpha_1(i) = p(x_1, y_1 = s_i) = \pi_i b_i(x_1), \\ \alpha_t(i) = [\sum\limits_{j} \alpha_{t-1}(j)a_{ji}] \times b_i(x_t). \end{cases} \tag{24}$$

由此, 可以得到产生某一观测序列的概率为

$$p(X|\lambda) = \sum_{i=1}^{N} \alpha_T(i). \tag{25}$$

引入变量

$$\delta_t(i) = \max p(x_1x_2\cdots x_t, y_t = s_i|\lambda)$$

表示t位置正处于状态s_i, 且沿着状态序列$y_1y_2\cdots y_t$产生出的观测序列为$x_1x_2\cdots x_t$的最大概率.

$$\varphi_t(i) = \arg\max p(x_1x_2\cdots x_t, y_t = s_i|\lambda)$$

表示产生最大概率时前一个状态的值.

于是, Viterbi算法的大致过程如下:

1) 初始化: $\delta_1(i) = \max p(x_1, y_1 = s_i|\lambda) = b_i(x_1)$.

2) 通过迭代公式:

$$\begin{cases} \delta_t(i) = \max\limits_{1\leqslant j\leqslant N} [\delta_{t-1}(j)a_{ji}] \times b_i(x_t), \\ \varphi_t(i) = \arg\max\limits_{1\leqslant j\leqslant N} \delta_{t-1}(j)a_{ji} \end{cases} \tag{26}$$

来计算$t = 1, 2, \cdots, T$时的$\delta_t(i)$与$\varphi_t(i)$.

3) 计算出$p(X|\lambda) = \max \delta_T(i)$, 记录下此位置(即最后一个位置)的状态, 即$y_T = s_i$, 利用此时的$\delta_T(i)$, 得到$\varphi_T(i) = s_j$, 即倒数第2个位置的状态, 再利用$\delta_{T-1}(j)$依次倒推. 直到求得$\varphi_2(i)$即为第1个位置时的状态.

由此, 利用Viterbi算法求得各个层级的指标得分.

4 实验与分析(Experiments and analysis)

为验证该模型方法的有效性, 本文以win7下的MATLAB为编程工具, 编写了仿真程序.

4.1 样本选取(Sample selection)

利用编写的人工打分系统, 让8位专家对项目绩效进行打分. 通过项目的40个四级指标, 逐级评判, 得到19个三级指标、6个二级指标、4个一级指标和最终的综合指标. 并将数据集随机分为训练集1、训练集2、测试集1、测试集2进行模型形成测试. 为了进一步验证模型是否能应用项目绩效评价中, 本文又用中国大洋协会2014年和2015年的部分绩效考核数据集进行测试. 总结为表2.

表 2 样本选取

Table 2 Sample selection

训练集1	400个项目
训练集2	400个项目
测试集1	100个项目
测试集2	100个项目
历史数据集	150个项目

4.2 仿真过程(Simulation)

实验 1 利用8位专家打分得到的数据集进行训练测试.

步骤 1 将训练集1依照特征函数规则, 将数据存入矩阵中.

步骤 2 利用IIS算法, 得到各级的特征函数的权重参数. 已知将指标得分划分为0到5, 共6个等级, IIS具体算法为:

第1步 初始化$\lambda_a^{(0)} = -100, \epsilon = 0.001, C =$特征个数, $a = 1, \cdots, 6 \times C$;

第2步 计算经验期望$\tilde{E}_p(f_i)$;

第3步 利用$\lambda_a^{(n)}$计算$P_{y_{i-1}}(y_i|x_i)$, $E_p(f_a)$;

第4步 $\lambda_a^{(n+1)} = \lambda_a^{(n)} + \dfrac{1}{C} \times \log \dfrac{\tilde{E}_p(f_a)}{E_p(f_a)}$;

第5步 重复第3–4步直到$|\lambda_a^{(n+1)} - \lambda_a^{(n)}| < \epsilon$.

依照上述算法可以学习出关于第i个指标的特征函数权重, 对该评价体系中的每项高级指标进行IIS算法学习, 即可得到整体的参数矩阵.

步骤 3　利用学习出的特征值权重参数λ_a, 得到条件分布$P_{y_{i-1}}(y_i|x_i)$, 利用Viterbi算法, 得到最终各级得分. 设某层级评价过程有T个评价指标, Viterbi算法具体步骤为:

第1步　初始化

$$\delta_1(i) = P(y_1 = i|x_1)p(x_1),\ i = 0, \cdots, 5;$$

第2步　利用$P_{y_{t-1}}(y_t|x_t)$计算$\delta_t(i)$和$\varphi_t(i)$, 至$t = T$;

第3步　计算$s_i = \arg\max_i \delta_T(i)$, 记录$y_T = s_i$;

第4步　$y_{t-1} = \varphi_t(y_t)$, 直到$t = 2$.

依次可以得到某一层的所有高级得分, 逐层计算, 可以得到体系中的所有得分和综合得分.

步骤 4　由此, 得到模型1, 利用模型计算出的综合得分与真实的综合得分进行比较, 计算出正确率为

$$e = \frac{\#\{\mathrm{scole}_{\mathrm{test}} = \mathrm{scole}_{\mathrm{ture}}\}}{100} \times 100\%.$$

步骤 5　对训练集2进行处理, 完成步骤1到步骤4, 得到模型2.

步骤 6　利用模型1对测试集1进行测试, 并与专家组2的打分结果进行对比分析, 计算正确率, 并将结果与BP神经网络和最大熵模型做对比.

步骤 7　利用模型2对测试集2进行测试, 并与专家组1的打分结果进行对比分析, 计算正确率, 并将结果与BP神经网络和最大熵模型做对比.

实验 2　利用大洋协会的历史数据对该模型进行验证.

首先, 数据预处理:在随机抽取的150个项目中, 项目的各级指标得分均是按百分制, 而且实际得分区间在85–95 分之间, 需要将得分去均值化、离散化.

其次, 训练模型: 利用交叉验证的方法找到最好的训练模型. 具体为随机抽取50个样本作为训练集, 对模型进行训练, 对余下的模型进行测试, 得到第1次实验正确率, 再重复5次这样的实验, 选择正确率最高的训练模型.

最后, 测试结果: 利用该模型对剩余100个样本进行打分, 得到的结果与BP神经网络和最大熵模型做对比.

4.3　结果分析(Result analysis)

本文利用仿真步骤4中计算的正确率, 来作为模型的性能指标, 两个实验得到的结果如表3所示.

对于大洋协会的历史数据测试集的100个项目, 可比较其真实分数分布与本文模型评得分数分布的比较, 见图6.

从表3中可知, 最大熵马尔科夫模型的正确率明显高于BP神经网络和最大熵模型: BP神经网络的正确率在70%左右, 而且该网络并不是很稳定, 对于同样的训练集和测试集, 在设定的精度要求范围内, 每次实验的正确率都不稳定, 本次实验结果大致描述了该算法的一个平均情况. 而最大熵模型要求未知量是独立的, 在统计指标之间有相互联系的绩效评价体系中, 并不能发挥有效的作用. 因此正确率在80%左右. 而最大熵马尔科夫模型即满足了最大熵的条件又考虑了各级间的关系, 因此取得了很好的效果.

表 3 实验结果

Table 3 Experimental result

	实 验	正确率/%
BP神经网络	测试集1	70
	测试集2	69
	历史数据集	73
最大熵模型	测试集1	76
	测试集2	80
	历史数据集	80
最大熵马尔科夫模型	测试集1	88
	测试集2	84
	历史数据集	86

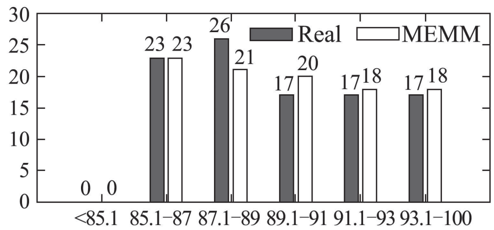

图 6 得分结果分布图

Fig. 6 Distribution of performance evaluation

从图6中, 也能显示出最大熵马尔科夫模型较好的分类效果, 但由于项目在89分附近太多, 而该模型的评分趋于均匀化, 导致3, 4两类分类误差较大.

5　结语(Conclusions)

本文针对大洋协会的绩效考核项目, 利用最大熵马尔科夫模型, 有针对性的研究了一套评价方法. 该方法结合了最大熵模型和隐马尔科夫模型的优势, 不仅能体现出既定的复杂的打分规则, 还能利用统计规律得到客观的同级间指标关系, 使整个打分系统更加全面和稳定, 能够科学地代替人工进行打分. 通过对该模型算法进行仿真实验, 对所得结果进行对比分析, 进一步验证了该模型方法的科学有效.

参考文献(References):

[1] SONG X, ZHANG Y. Financial performance evaluation in an industry — A principal component analysis based on financial ratios and entropy weight [J]. *On Economic Problems*, 2011, 33(6): 024.

[2] TSAI H Y, CHANG C W, LIN H L. Fuzzy hierarchy sensitive with Delphi method to evaluate hospital organization performance [J]. *Expert Systems with Applications*, 2010, 37(8): 5533 – 5541.

[3] ISLAM R, BIN MOHD RASAD S. Employee performance evaluation by the AHP: A case study [J]. *Asia Pacific Management Review*, 2006, 11(3): 163 – 176.

[4] WU H H, CHEN H K, SHIEH J I. Evaluating performance criteria of employment service outreach program personnel by DEMATEL method [J]. *Expert Systems with Applications*, 2010, 37(7): 5219 – 5223.

[5] SAATY T L. How to make a decision: the analytic hierarchy process [J]. *European Journal of Operational Research*, 1990, 48(1): 9 – 26.

[6] BHATTACHARYA A, MOHAPATRA P, KUMAR V, et al. Green supply chain performance measurement using fuzzy ANP-based balanced scorecard: a collaborative decision-making approach [J]. *Production Planning & Control*, 2014, 25(8): 698 – 714.

[7] HSIEH L F, LIN L H. A performance evaluation model for international tourist hotels in Taiwan — An application of the relational network DEA [J]. *International Journal of Hospitality Management*, 2010, 29(1): 14 – 24.

[8] ZHENG G, JING Y, HUANG H, et al. Application of improved grey relational projection method to evaluate sustainable building envelope performance [J]. *Applied Energy*, 2010, 87(2): 710 – 720.

[9] KOU G, PENG Y, WANG G. Evaluation of clustering algorithms for financial risk analysis using MCDM methods [J]. *Information Sciences*, 2014, 275(1): 1 – 12.

[10] AKSOY A, OZTÜRK N. Supplier selection and performance evaluation in just-in-time production environments [J]. *Expert Systems with Applications*, 2011, 38(5): 6351 – 6359.

[11] CHEN J K, CHEN I S. Using a novel conjunctive MCDM approach based on DEMATEL, fuzzy ANP, and TOPSIS as an innovation support system for Taiwanese higher education [J]. *Expert Systems with Applications*, 2010, 37(3): 1981 – 1990.

[12] CHEN F H, HSU T S, TZENG G H. A balanced scorecard approach to establish a performance evaluation and relationship model for hot spring hotels based on a hybrid MCDM model combining DEMATEL and ANP [J]. *International Journal of Hospitality Management*, 2011, 30(4): 908 – 932.

[13] RABINER L R, JUANG B H. An introduction to hidden Markov models [J]. *IEEE ASSP Magazine*, 1986, 3(1): 4 – 16.

[14] BERGER A L, PIETRA V J D, PIETRA S A D. A maximum entropy approach to natural language processing [J]. *Computational Linguistics*, 1996, 22(1): 39 – 71.

[15] DARROCH J N, RATCLIFF D. Generalized iterative scaling for log-linear models [J]. *The Annals of Mathematical Statistics*, 1972, 43(5): 1470 – 1480.

[16] MALOUF R. A comparison of algorithms for maximum entropy parameter estimation [C] //*Proceedings of the 6th Conference on Natural Language Learning*. New Jersey: Association for Computational Linguistics, 2002, 20: 1 – 7.

[17] MCCALLUM A, FREITAG D, PEREIRA F C N. Maximum entropy Markov models for information extraction and segmentation [C] //International Conference on Machine Learning. San Francisco: Morgan Kaufmann, 2000, 17: 591 – 598.

[18] ZHAO X, ZHANG J, CHEN Y, et al. Promoter recognition based on the maximum entropy hidden Markov model [J]. *Computers in Biology and Medicine*, 2014, 51(1): 73 – 81.

作者简介:

朱 磊 (1963–), 女, 高级工程师, 主要从事国际海底区域资源研究系统开发、项目管理和绩效评价等方向的研究, E-mail: zLdan@vip.sina.com;

牛绿茵 (1993–), 女, 硕士研究生, 目前研究方向为复杂系统性能分析与评价, E-mail: niulyhappy@126.com;

宋士吉 (1965–), 男, 教授, 博士生导师, 目前主要从事深海地质科学数学建模与资源评价、复杂制造系统建模、优化与控制技术、大洋信息系统、科考船信息化等方向的研究, E-mail: shijis@mail.tsinghua.edu.cn;

张玉利 (1987–), 男, 博士, 在做博士后研究工作, 目前主要从事运作优化、智能电网、供应链管理等领域的研究, E-mail: zhangyuli@mail.tisnghua.edu.cn.