多元伪泊松混合分布模型的理论研究

陈奕延[1,2] 李 晔[1,2,3] 张淑芬[3]

[1](英国系统科学学会 伦敦 CR82AD)
[2](北京理工大学自动化学院 北京 100081)
[3](华北理工大学 河北省数据科学与应用重点实验室 唐山 063210)

摘 要 针对混合分布模型中各项权值通常依赖于未知或已知参数而造成的模型不确定问题，提出了一种权值基于 Frobenius 范数的混合分布模型。首先，把多元泊松分布进行截断及均化处理，生成伪多元泊松分布。其次，根据有限可数混合分布的表达式，分别求解伪多元泊松混合分布的集函数矩阵、多线性形式的 Pseudo-Boolean 函数矩阵、多线性 Pseudo-Boolean 函数矩阵的 Frobenius 范数，由此得到新的权值并据此构建多元伪泊松混合分布模型。最后，根据混合分布权值的归一性及非负性证明了模型的正确性并且通过仿真实验来展示构建模型的整个过程，验证了算术平均的合理性。可为今后研究混合分布在机器学习领域的应用及算法设计提供理论基础。

关键词 多元伪泊松混合分布；集函数矩阵；Pseudo-Boolean 函数矩阵；Frobenius 范数
中图分类号 TP 312 O 211.3 **文献标志码** A

A Multivariate Pseudo-Poisson Mixture Distribution Model: a Theoretical Research

CHEN Yiyan[1,2] LI Ye[1,2,3] ZHANG Shufen[3]

[1](*UK Systems Science Association*, *London* CR82AD, *England*)
[2](*School of Automation*, *Beijing Institute of Technology*, *Beijing* 100081, *China*)
[3](*Hebei Key Laboratory of Data Science and Application*, *North China University of Science and Technology*, *Tangshan* 063210, *China*)

Abstract The weight values in mixture distribution models usually depend on unknown or known parameters, which makes the model uncertain. To address this issue, we propose to determine the model weights based on Frobenius. Firstly, the multivariate Poisson distribution was truncated and homogenized to generate a multivariate Pseudo-Poisson distribution. Secondly, set function matrix of multivariate Pseudo-Poisson mixture distribution, multiple linear Pseudo-Boolean function matrix, multiple linear Pseudo-Boolean function matrix's Frobenius norm were solved respectively according to the expression for countable mixture distribution. New weights were calculated and in turn a multivariate Pseudo-Poisson

收稿日期：2017-11-15 **修回日期**：2017-12-26
基金项目：河北省数据科学与应用重点实验室开放课题(20170320002)；英国系统科学学会深度课题(UA1709001F)
作者简介：陈奕延，博士，高级工程师，经济师，研究方向为统计建模、技术经济；李晔(通讯作者)，博士，工程师，研究方向为机器学习、数据分析，E-mail：ly1992bigdata@163.com；张淑芬，教授，研究方向为云计算。

mixture distribution model was constructed. Finally, the correctness of the model was proved according to the normalization and nonnegativeness of the mixture distribution weights and the entire process of building model was demonstrated through simulation experiments. We also verified that arithmetic average is reasonable. The proposed model can provide a theoretical basis for applications and algorithm design of mixture distribution in machine learning.

Keywords multivariate Pseudo-Poisson mixture distribution; set function matrices; Pseudo-Boolean function matrices; Frobenius norm

1 引 言

混合分布模型是一类源于数理统计知识理论的数学模型，在股票、期货、风险管理、健康产业、保费厘定、电力工业、生物医疗、地质勘探等领域中均有广泛的应用[1-11]。在场景识别[12]、TSP 问题[13]、人脸识别[14]、云计算[15]、计算机视觉[16]等计算机相关领域内亦有广泛应用，特别是在机器学习领域。混合分布模型主要为各类算法的设计提供理论基础，如蜂群算法[17]、人工鱼群算法[18]、免疫算法[19]、蚁群算法[20]和遗传算法[21]等。

在机器学习中，常用的连续型随机变量的混合分布及变型有：指数混合分布及其变型[22]、高斯混合分布及其变型[23]、Gamma 混合分布及其变型[24]、*t* 混合分布及其变型[25]等。对于离散型随机变量的混合分布而言，则通常研究泊松混合分布(或称混合泊松分布)及其在相关领域的应用。冯杭和王胜兵[26]利用 EM 算法求解泊松混合分布模型中的参数向量，并结合算例指出该算法对初值敏感的缺陷，最终引入智能优化算法对算例进行改进。该研究使用泰国北部地区学龄前儿童 1982－1985 年发烧、咳嗽或两者皆有的患病频次作为算例，用 EM 算法通过迭代优化估计出该泊松混合分布的权值(该文称之“权重”)。罗修辉等[27]运用 EM 算法对一种新的系统寿命分布——混合指数-泊松分布进行参数估计，并通过随机模拟，验证了 EM 算法在混合模型参数估计的收敛性和有效性。该研究对象是混合指数-泊松分布，通过 100 个来自指数分布分量的数以及 150 个来自泊松分布分量的数，使用 EM 算法随机模拟出分布的权值。虞欢欢[28]为优化金融管理风险，针对资产组合的相关性结构和组合风险计算的复杂性两方面做出实证研究，用混合泊松分布构建资产组合信用风险模型，利用 85 家上市公司的相关数据估算出泊松混合分布的权值(该文称之为“系数”)，并对模型进行了说明。高迎心等[29]引入了基于变异位点的先验概率分布模型，运用基于混合泊松分布的期望最大化(EM)算法对新生突变识别算法进行改进与优化，研究了有亲缘关系的新生突变的识别，并在识别精度与运算速度方面与已有算法进行对比。该研究结果表明，基于混合泊松分布的期望最大化算法在提高运算速度的同时降低了假阳性比率，具有良好的识别效果。

上述研究均存在一个问题，即混合分布的权值往往依赖于已知或未知的参数，而对未知参数的估计又需要使用经验数据或仿真模拟。由于参数估计方法的不同，造成的估计结果也各有差异，数据的获得往往并非一帆风顺，故这种参数依赖性的问题给混合分布的研究造成了障碍。

为解决混合分布中分量参数依赖性的问题，

本文基于多元泊松混合分布的一种变型来进行研究。该混合分布由多元泊松分布经过截断和均化处理后得到伪多元泊松分布，通过集函数矩阵多线性形式的 Pseudo-Boolean 函数矩阵及其相应的 Frobenius 范数得到的权值有限混合而成。该多元混合分布的权值在满足有效市场假说，即算术平均合理性的情况下不依赖于任何参数，仅与混合分布中的分量个数有关。

2　研究步骤

多元伪泊松混合分布模型的设计按以下 8 个步骤进行：

(1) 设定范式及约束条件；

(2) 对多元泊松分布进行截断处理；

(3) 在截断处理的基础上，再对其进行均化处理，得到伪多元泊松分布；

(4) 定义伪多元泊松混合分布的集函数矩阵；

(5) 根据集函数矩阵，写出相应的多线性 Pseudo-Boolean 函数矩阵；

(6) 根据 Lovász 延拓计算出 Frobenius 范数，构造新的权值；

(7) 根据新的权值，写出多元伪泊松混合分布模型的数学表达式；

(8) 根据混合分布权值的归一性及非负性，证明其数学表达式正确。

3　分布模型的设计

3.1　范式设定及约束条件

设随机变量 $X_{ij}\sim PN_{ij}\left(\lambda_{ij}\right)$，$PN$ 为泊松分布 (Poisson) 的英文缩写；λ_{ij} 为参数。令 X_{ij} 之间相互独立，且 $Z_i=\sum_j X_{ij}$，可知 Z_i 是一个由有限可数的 X_{ij} 线性加和组成的多元泊松随机变量，$Z_i\subseteq\mathbb{N}$。令 $Z_i\sim MPN_i\left(\lambda_{i1},\lambda_{i2},\cdots,\lambda_{im(i)}\right)$，$MPN$ 为多元泊松 (Multivariate Poisson) 的英文缩写；$m(i)$ 为第 i 个多元泊松分布中 X_{ij} 的个数，设：

$$\exists\tilde{\zeta}_{MPN}=\{Z_1,Z_2,\cdots,Z_l\} \tag{1}$$

对 $\forall e,f,g,h\in\mathbb{Z}^+$，有 $\lambda_{eg}\neq\lambda_{fh}$，$\tilde{\zeta}_{MPN}$ 是一个由独立不同参数的多元泊松分布组成的有限可数分布簇集，已知 Z_i 对 $\tilde{\zeta}_{MPN}$ 中的任一元素 $MPN_i\left(\lambda_{i1},\lambda_{i2},\cdots,\lambda_{im(i)}\right)$ 皆有不等的权值 ρ_i，使得 l 个多元泊松随机变量的混合分布模型的范式为：

$$\left\langle\overset{MD}{MPN}\right\rangle=\sum_{i=1}^{l}\rho_i MPN_i\left(\lambda_{i1},\cdots,\lambda_{im(i)}\right) \tag{2}$$

其中，MD 为英文 Multivariate Mixture Distribution 的缩写；$\sum_{i=1}^{l}\rho_i=1$，$0\leqslant\rho_i\leqslant1$，则称 $\left\langle\overset{MD}{MPN}\right\rangle$ 为有限可列的多元泊松混合分布。

泊松分布的取值是无限可列的。但为解决实际问题，令 $X_{ij}=\{x_{ij1},\cdots,x_{ijn(j)}\}$，其中 $n(j)$ 表示第 i 个多元泊松随机变量中第 j 个随机分量的第 $n(j)$ 个观察值，令 $k\leqslant n(j)$，$x_{ijk}\in\mathbb{N}$。假设从第 $n(j)$ 项开始施行截断和均化处理，原始的多元泊松分布将变成新的有限分布，即伪多元泊松分布。则根据以上条件，模型的约束条件为：

(1) $X_{ij}\sim PN_{ij}\left(\lambda_{ij}\right)$；

(2) X_{ij} 是相互独立的；

(3) $Z_i=\sum_j X_{ij}$；

(4) $X_{ij}=\{x_{ij1},\cdots,x_{ijn(j)}\}$；

(5) $Z_i\sim MPN_i\left(\lambda_{i1},\lambda_{i2},\cdots,\lambda_{im(i)}\right)$；

(6) 随机变量满足有效市场假说。

3.2　多元泊松分布的截断处理

多元随机变量 $Z_i\subseteq\mathbb{N}$，Z_i 在服从多元泊松分布的条件下，其取值的个数是无穷多个。由于现实生活中的局限，很多问题的取值是有限个，这里需对原始的多元泊松分布 $MPN_i\left(\lambda_{i1},\lambda_{i2},\cdots,\lambda_{im(i)}\right)$ 进行截断处理。

截断即设定阈值来限制调查对象的范围，超过阈值的对象不做调查，但需将其录入总体分布

中，以保证随机分布的完整性。例如，对收入进行调查，小于 5 000 元的用实际收入表示，大于 5 000 元不作调查，但需将其作为全部收入分布的一部分，这就是截断。

由于 $Z_i=\sum_j X_{ij}$，所以先对 X_{ij} 进行截断。设服从泊松分布的随机变量 X_{ij} 的阈值为 μ_{ij}，X_{ij} 取值构成集合 $V_{ij}=\{x_{ij1},x_{ij2},\cdots,x_{ijn(j)}\}$，$x_{ijn(j)}\leqslant\mu_{ij}<x_{ijn(j)+1}$，故将 X_{ij} 的取值 x_{ijk} 自第 $n(j)$ 项以后截断，$n(j)\in\mathbb{Z}^+$。将第 $n(j)+1$ 项以后的各点概率值 $p_{i[n(j)+a_{ij}]}$，$a_{ij}\in\mathbb{Z}^+$ 累积到第 $n(j)$ 项上，此时 $X_{ij}=\{x_{ij1},x_{ij2},\cdots,x_{ijn(j)}\}$，$x_{ijk}\in\mathbb{N}$，存在一个排序，有 $x_{ij1}<x_{ij2}<\cdots<x_{ijn(j)}$。设 $\langle X_{ij}\rangle$ 为随机变量 X_{ij} 中取值 x_{ijk} 的个数，$\langle X_{ij}\rangle=\langle\{x_{ij1},x_{ij2},\cdots,x_{ijn(j)}\}\rangle=\langle V_{ij}\rangle=n(j)$。假设 $\sum P_{[-n(j)]}=\sum_{a_{ij}=1}^{\infty}p_{ij[n(j)+a_{ij}]}$，则 $x_{ijn(j)}$ 的概率值为：

$$p^*_{ijn(j)}=p_{ijn(j)}+\sum P_{[-n(j)]} \tag{3}$$

$p_{ijk}=\dfrac{(\lambda_{ij})^{x_{ijk}}}{(x_{ijk})!}e^{-\lambda_{ij}}$，此时泊松随机变量 X_{ij} 的分布函数为：

$$P_{ij}(X_{ij}=x_{ijk})=\begin{cases}\dfrac{(\lambda_{ij})^{x_{ijk}}}{(x_{ijk})!}e^{-\lambda_{ij}},x_{ijk}<x_{ijn(j)}\\ \sum\limits_{k=n(j)}^{\infty}\dfrac{(\lambda_{ij})^{x_{ijk}}}{(x_{ijk})!}e^{-\lambda_{ij}},x_{ijk}=x_{ijn(j)}\end{cases} \tag{4}$$

由于 $Z_i=\sum_j X_{ij}$，设 $\langle Z_i\rangle=\prod_{j=1}^{m(i)}n(j)=q$，$Z_i=\{Z_{i1},Z_{i2},\cdots,Z_{iq}\}$，$z_{iv}\in\mathbb{N}$，可知：

$P_i\left(Z_i=\sum_j x_{ijk}\right)=P_i(X_{i1}+\cdots+X_{im(i)}=Z_i)$，同时有 $X_{im(i)}=x_{im(i)n[m(i)]}=r_{im(i)}$，$0\leqslant r_{im(i)}\leqslant\mu_{ij}$。因为 X_{ij} 相互独立，则：

$$\begin{aligned}&P_i(X_{i1}+\cdots+X_{im(i)}=z_i)\\&=P_i\left(X_{i1}=r_{i1},\cdots,X_{im(i)}=\left(z_i-\sum_{j=1}^{m(i)-1}r_{ij}\right)\right)\\&=P_{i1}(X_{i1}=r_{i1})\cdots P_{im(i)}\left(X_{im(i)}=\left(z_i-\sum_{1}^{m(i)-1}r_{ij}\right)\right)\\&=P_{i1}(X_{i1}=x_{i1k})\cdots P_{im(i)}\left(X_{im(i)}=z_i-\sum_{j=1}^{m(i)-1}x_{ijk}\right)\end{aligned} \tag{5}$$

将公式(4)代入公式(5)，可得：

$$P_i(Z_i=z_{iv})=\begin{cases}\prod\limits_{j=1}^{m(i)}\left[\dfrac{(\lambda_{ij})^{x_{ijk}}}{(x_{ijk})!}e^{-\lambda_{ij}}\right],x_{ijk}<x_{ijn(j)}\\ \prod\limits_{j=1}^{m(i)}\left[\sum\limits_{k=n(j)}^{\infty}\dfrac{(\lambda_{ij})^{x_{ijk}}}{(x_{ijk})!}e^{-\lambda_{ij}}\right],x_{ijk}=x_{ijn(j)}\end{cases} \tag{6}$$

其中，$1\leqslant v\leqslant n(j)^{m(i)}=q$；$z_{iv}\in\mathbb{N}$ 且 $0\leqslant z_{iv}\leqslant\sum_{j=1}^{m(i)}\mu_{ij}$。

通过截断处理，可将不可数取值的多元离散随机变量 z_{iv} 的值限定在一个有限域 $\left[0,\sum_{j=1}^{m(i)}\mu_{ij}\right]$ 内。但这样的处理会产生一些问题，故对此还需进行均化处理。

3.3 多元泊松分布的均化处理

泊松分布的取值遍布自然数域，即取值有无穷多项，故完成截断处理后的分布在点 $x_{ijn(j)}$ 处有一个相对前 $n(j)-1$ 个点的概率而言非常巨大的概率值 $p^*_{ijn(j)}$，这会造成概率在某一点处的过度堆积，以致于 $p^*_{ijn(j)}\approx1$ 且 $p^*_{ijn(j)}\xrightarrow{\infty}1$，同时导致 $p^*_{ijn(j)}\gg P_{ij}[X_{ij}=x_{ijk}\,|\,1\leqslant k\leqslant n(j)-1]$。为保证随机变量波动的均匀性，须在截断后对其进行均化处理。

均化原本是化学工业上的一个概念，狭义的均化特指通过采用一定的工艺措施，使物料化学成分的波动振幅降低，达到物料化学成分均匀一

致的过程。这种方法多用于化工作业的操作中。这里的均化是指广义的均化，特指为了防止离散随机波动过于极端而将极端的概率值进行处理后均匀(算术平均、几何平均等)分散到其余各点的行为。在离散条件下，若有 l 个极端概率，则舍弃这 l 个概率值对应的点，将这 l 个点的概率累积后再进行均化处理。

特别注意，对泊松分布而言，其随机变量取值有无穷多项，从某一项 $n(j)$ 处进行截断处理，显然剩余的项依然是无穷多项，且 $\sum P_{[-n(j)]}$ 是单调非减的，则显然可知 $\sum P_{[-n(j)]} \gg p_{ijn(j)}$。

设 $\left\langle \left\{ x_{ij1}, x_{ij2}, \cdots, x_{ijn(j)} \right\} \right\rangle$ 为截断处理后随机变量 X_{ij} 中取值 x_{ijk} 的个数，由于之前采取了截断措施，故 $\left\langle \left\{ x_{ij1}, x_{ij2}, \cdots, x_{ijn(j)} \right\} \right\rangle = n(j)$。将第 $n(j)$ 项以后(不包括第 $n(j)$ 项)的概率累积值 $\sum P_{[-n(j)]}$ 按算术平均数平均分割成 $n(j)$ 份，每一份为 $p_{ij}^{h} = \left\{ \sum P_{[-n(j)]} \right\} \Big/ n(j)$，$p_{ij}^{h}$ 称为算术均化增量，将 p_{ij}^{h} 分别加到前 $n(j)$ 项的概率值 $P_{ij}\left(X_{ij} = x_{ijk}\right)$ 上，可得：

$$P_{ij}\left(X_{ij} = x_{ijk}\right) = \frac{\left(\lambda_{ij}\right)^{x_{ijk}}}{\left(x_{ijk}\right)!} e^{-\lambda_{ij}} + \left(\sum P_{[-n(j)]}\right) \Big/ n(j) \tag{7}$$

设上述分布为伪泊松分布，$X_{ij} \sim \Omega_{ij}^{\tau}\left(\lambda_{ij}\right)$，由 $Z_i = \sum_j X_{ij}$，设 $\langle Z_i \rangle = \prod_{j=1}^{m(i)} n(j) = q$，$Z_i = \left\{ z_{i1}, z_{i2}, \cdots, z_{iq} \right\}$，$z_{iv} \in \mathbb{N}$，$X_{ij}$ 相互独立，根据约束条件可知：

$$P_i\left(Z_i = z_{iv}\right) = \prod_{j=1}^{m(i)} \left[\frac{\left(\lambda_{ij}\right)^{x_{ijk}}}{\left(x_{ijk}\right)!} e^{-\lambda_{ij}} + \left(\sum P_{[-n(j)]}\right) \Big/ n(j) \right] \tag{8}$$

设均化处理后的离散随机分布为 Ω_i^{τ}，可知 $Z_i \sim \Omega_i^{\tau}$，为便于行文，本文称 Ω_i^{τ} 为多元伪泊松分布。

截断处理和均化处理的目的是在原有泊松分布的基础上得到新的分布，即得到新的多元伪泊松分布。为使图中文字公式清晰便于查阅，本文更改一下表示符，设 $p_{ijk} = p(i,j,k)$，生成多元伪泊松分布的具体操作流程如图 1 所示。

3.4　有限可列的多元伪泊松混合分布

由上述章节可知，原始泊松分布经过截断和均化处理后的分布 Ω_i^{τ} 称为多元伪泊松分布。假设 $\exists \tilde{\psi}_{\tau} = \left\{ \Omega_1^{\tau}, \Omega_2^{\tau}, \cdots, \Omega_l^{\tau} \right\}$ 为分布 Ω_i^{τ} 的簇，针对簇中任一 Ω_i^{τ}，$\exists \Omega^{\mathbb{M}}$ 使得：

$$\Omega^{\mathbb{M}} = \rho_1^{\mathbb{M}} \Omega_1^{\tau} + \cdots + \rho_l^{\mathbb{M}} \Omega_l^{\tau} = \sum_{i=1}^{l} \rho_i^{\mathbb{M}} \Omega_i^{\tau} \tag{9}$$

其中，$\sum_{i=1}^{l} \rho_i^{\mathbb{M}} = 1$，$0 \leqslant \rho_i^{\mathbb{M}} \leqslant 1$。

3.5　多元伪泊松分布的集函数矩阵

集函数是计算机技术领域内常用的一类函数，国内外对其也有一定研究[30-35]。离散型随机变量 X_{ij}，设 $X_{ij} \sim \Omega_{ij}^{\tau}\left(\lambda_{ij}\right)$，$X_{ij}$ 的全体取值为 V_{ij}，则 $V_{ij} = \left\{ x_{ij1}, x_{ij2}, \cdots, x_{ijn(j)} \right\}$。若 A_{ij} 为 V_{ij} 的任意子集，则对 $\forall A_{ij}$，$\exists A_{ij} \subseteq V_{ij}$ 使得 f_{ij}^{S} 为集函数，$f : 2^{V_{ij}} \mapsto \mathbb{R}$，设 $i \in \mathbb{Z}^{+}$ 且 $1 \leqslant i \leqslant l$，首先求 X_{ij} 的集函数。假设空集的集函数 $f_{ij}^{S}(\phi) = 0$，继续定义其余子集的集函数，设 $\forall A_{ij} = \left\{ r_{ij1}, r_{ij2}, \cdots, r_{ijk(j)} \right\}$，$r_{ijk(j)} \in A_{ij} \subseteq V_{ij}$，$k(j) \leqslant n(j)$，令：

$$f_{ij}^{S}\left(\left\{ r_{ij1}, \cdots, r_{ijk(j)} \right\}\right) = \sum_{k=1}^{k(j)} \frac{\left(\lambda_{ij}\right)^{r_{ijk}}}{\left(r_{ijk}\right)!} e^{-\lambda_{ij}} + \frac{k(j)\left(\sum P_{[-n(j)]}\right)}{n(j)} \tag{10}$$

这里 $\frac{\left(\sum P_{[-n(j)]}\right)}{n(j)}$ 是对原始泊松分布在点 $x_{ijn(j)}$ 处进行截断处理后 $n(j)+1$ 项及其以后的概率累积后的算术平均值，若 $A_{ij} = V_{ij}$，$k(j) = n(j)$，此时有：

$$\begin{aligned} & f_{ij}^{S}\left(\left\{ x_{ij1}, x_{ij2}, \cdots, x_{ijn(j)} \right\}\right) \\ & = \sum_{k=1}^{n(j)} \frac{\left(\lambda_{ij}\right)^{x_{ijk}}}{\left(x_{ijk}\right)!} e^{-\lambda_{ij}} + \left(\sum P_{[-n(j)]}\right) = 1 \end{aligned} \tag{11}$$

由此可知，随机变量 X_{ij} 的集函数为：

$$f_{ij}^{S}\left(A_{ij}\right)=\begin{cases}0, A_{ij}=\phi\\ \sum_{k=1}^{k(j)}\frac{\left(\lambda_{ij}\right)^{r_{ijk}}}{\left(r_{ijk}\right)!}e^{-\lambda_{ij}}+\frac{k(j)\left(\sum P_{\left[-n(j)\right]}\right)}{n(j)}, A_{ij}=\phi\end{cases}\tag{12}$$

其中，$k(j)\leqslant n(j)$。对于 $Z_i\sim\Omega_i^{\tau}$，$Z_i=\sum_j X_{ij}$，因为其是多元的，故可用矩阵表示，设 $\boldsymbol{S}_i$ 为一个 $1\times m(i)$ 阶的矩阵，矩阵中元素为 $f_{ij}^{S}\left(A_{ij}\right)$，则：

$$\boldsymbol{S}_i=\left[f_{i1}^{S}\left(A_{i1}\right)f_{i2}^{S}\left(A_{i2}\right)\cdots f_{im(i)}^{S}\left(A_{im(i)}\right)\right]\tag{13}$$

$\boldsymbol{S}_i$ 为多元随机变量 Z_i 的集函数矩阵。

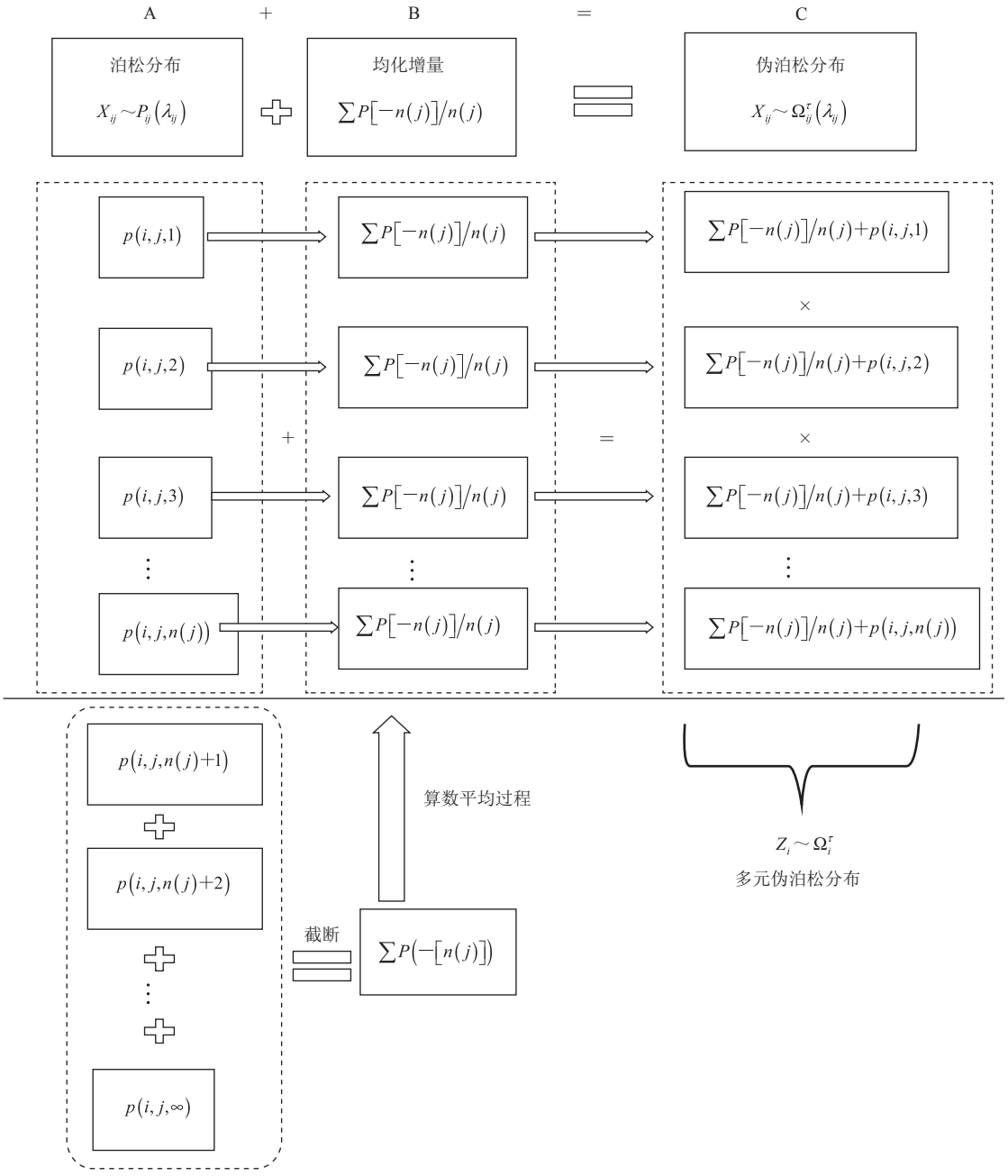

图 1　生成多元伪泊松分布的流程图

Fig. 1　Flow chart of multivariate pseudo-poisson distribution generating

3.6　多线性 Pseudo-Boolean 函数矩阵

Pseudo-Boolean 函数也是一种计算机技术领域常用的函数[36-38]，其定义域为 0～1 变量且对应到实数域上的函数 $f:\mathbb{B}^n\mapsto\mathbb{R}$，其中，$n=n(j)$。设 $X_{ij}\sim\Omega_{ij}^{\tau}(\lambda_{ij})$，则相应的多线性 Pseudo-Boolean 函数的表达式为：

$$f_{ij}^{B}\left(x_{ij1},\cdots,x_{ijn(j)}\right)=\sum_{A_{ij}\subseteq V_{ij}}\hat{f}_{ij}^{S}\left(A_{ij}\right)\prod_{k\in A_{ij}}x_{ijk}\tag{14}$$

其中，$\hat{f}_{ij}^{S}\left(A_{ij}\right)$ 是 $f_{ij}^{S}\left(A_{ij}\right)$ 的估计值。根据莫比乌斯变换可知：

$$\hat{f}_{ij}^{S}\left(A_{ij}\right)=\sum_{U_{ij}\subseteq A_{ij}}(-1)^{\langle A_{ij}\rangle-\langle U_{ij}\rangle}f_{ij}^{S}\left(\mathbf{1}_{U_{ij}}\right)\tag{15}$$

其中，$\langle A_{ij}\rangle$ 与 $\langle U_{ij}\rangle$ 分别代表集合 A_{ij} 与其子集 U_{ij} 中元素的个数；$\mathbf{1}_{U_{ij}}$ 为指示向量，任意集函数都可以用其指示向量标记出来。设 $A_{ij}\subseteq V_{ij}$，$\mathbf{1}_{A_{ij}}\in\mathbb{B}^{n(j)}=\{0,1\}^{n(j)}$，令位置函数为：

$$\mathbf{1}_{A_{ij}}\left(x_{ijk}\right)=\begin{cases}1,x_{ijk}\in A_i\\0,x_{ijk}\notin A_i\end{cases}\tag{16}$$

例如，对 $\{x_{ijk}\}\subseteq V_{ij}$ 其指示向量的表达式为：

$$\mathbf{1}_{\{x_{ijk}\}}=\begin{pmatrix}\mathbf{1}_{A_{ij}}\left(x_{ij1}\right)\\\mathbf{1}_{A_{ij}}\left(x_{ij2}\right)\\\vdots\\x_{ijk}\\\vdots\\\mathbf{1}_{A_{ij}}\left(x_{ijn(j)}\right)\end{pmatrix}=\begin{pmatrix}0\\0\\\vdots\\x_{ijk}\\\vdots\\0\end{pmatrix}\tag{17}$$

由此可求出 $\mathbf{1}_{A_{ij}}$ 的形式。根据定义，任意一个集函数 $f:2^{V_{ij}}\mapsto\mathbb{R}$ 都可等价于一个 Pseudo-Boolean 函数 $f:\mathbb{B}^{n(j)}\mapsto\mathbb{R}$，即 $f_{ij}^{S}\left(A_{ij}\right)=f_{ij}^{S}\left(\mathbf{1}_{A_{ij}}\right)$，将该等式代入公式(15)，可得：

$$\hat{f}_{ij}^{S}\left(A_{ij}\right)=\sum_{U_{ij}\subseteq A_{ij}}(-1)^{\langle A_{ij}\rangle-\langle U_{ij}\rangle}f_{ij}^{S}\left(U_{ij}\right)\tag{18}$$

然后将公式(18)代入公式(14)中，可得：

$$f_{ij}^{B}\left(x_{ij1},\cdots,x_{ijn(j)}\right)=\sum_{A_{ij}\subseteq V_{ij}}\left[\sum_{U_{ij}\subseteq A_{ij}}(-1)^{\langle A_{ij}\rangle-\langle U_{ij}\rangle}f_{ij}^{S}\left(U_{ij}\right)\right]\prod_{k\in A_{ij}}x_{ijk}\tag{19}$$

设 $\langle A_{ij}\rangle=k(j)$，易证，当 $k(j)=1$ 时，$\langle U_{ij}\rangle=1$ 或 0，显然：

$$f_{ij}^{S}\left(U_{ij}\middle|k(j)=1\right)=\frac{\left(\lambda_{ij}\right)^{x_{ijk}}}{\left(x_{ijk}\right)!}e^{-\lambda_{ij}}+\left(\sum P_{[-n(j)]}\right)\Big/n(j)\tag{20}$$

若 $k(j)=0$，则 $\langle U_{ij}\rangle=0$，$f_{ij}^{S}(\phi)=0$，若 $k(j)\geqslant 2$：

$$f_{ij}^{S}\left(U_{ij}\middle|2\leqslant k(j)\leqslant n(j)\right)=\sum_{k=1}^{k(j)}\frac{\left(\lambda_{ij}\right)^{x_{ijk}}}{\left(x_{ijk}\right)!}e^{-\lambda_{ij}}+\frac{k(j)\left(\sum P_{[-n(j)]}\right)}{n(j)}\tag{21}$$

由此可知：

$$f_{ij}^{S}\left(U_{ij}\right)=\begin{cases}0,U_{ij}=\phi\\\sum_{k=1}^{k(j)}\frac{\left(\lambda_{ij}\right)^{x_{ijk}}}{\left(x_{ijk}\right)!}e^{-\lambda_{ij}}+\frac{k(j)\left(\sum P_{[-n(j)]}\right)}{n(j)},U_{ij}\neq\phi\end{cases}\tag{22}$$

根据归纳法，易证明当 $k(j)\geqslant 2$ 时，$\hat{f}_{ij}^{S}\left(A_{ij}\right)=0$，则：

$$f_{ij}^{B}\left(x_{ij1},\cdots,x_{ijn(j)}\right)=\sum_{n(j)}\left[f_{ij}^{S}\left(U_{ij}\middle|k(j)=1\right)\left(x_{ijk}\middle|k\in A_{ij}\right)\right]\tag{23}$$

因 $\left(x_{ijk}\middle|k\in A_{ij}\right)=x_{ijk}$，$k\in A_{ij}$，$x_{ijk}\in\{0,1\}$，显然可得：

$$f_{ij}^{B}\left(x_{ij1},\cdots,x_{ijn(j)}\right)=\begin{cases}1,x_{ijk}=1\\0,x_{ijk}=0\end{cases}\tag{24}$$

这里使用全集形式，令 $A_{ij}=\{x_{ij1},\cdots,x_{ijn(j)}\}$，$x_{ijk}=1$，则 $f_{ij}^{B}\left(x_{ij1},\cdots,x_{ijn(j)}\right)\equiv 1$。其目的是为了使多线性 Pseudo-Boolean 函数具有现实意义，即所有随机变量的取值均在样本空间中出现，同时保证后续计算 Frobenius 范数的权值有意义。对于 $Z_i\sim\Omega_i^{\tau}$，$Z_i=\sum_j X_{ij}$，因其是多元的，故可用矩阵表示，设 $\boldsymbol{B}_i$ 为一个 $1\times m(i)$ 阶的矩阵，矩阵中元素为 $f_{ij}^{B}\left(x_{ij1},x_{ij2},\cdots,x_{ijn(j)}\right)$，简记为 f_{ij}^{B}，则：

$$\boldsymbol{B}_i=\begin{bmatrix}f_{i1}^{B} & f_{i2}^{B} & \cdots & f_{im(i)}^{B}\end{bmatrix}\tag{25}$$

$\boldsymbol{B}_i$ 称为多元随机变量 Z_i 的 Pseudo-Boolean 函数

矩阵。

3.7 Frobenius 范数

Frobenius 范数作为范数的一种，在计算机领域中有很好的运用[39-43]。运用 Frobenius 范数可得到一个由矩阵中的元素组成的代数式，其定义为：

$$\|\boldsymbol{A}\|_F=\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}\right|^2\right)^{1/2} \tag{26}$$

其中，$\boldsymbol{A}$ 为 $m\times n$ 阶矩阵；a_{ij} 为矩阵中的元素；$\|\boldsymbol{A}\|_F$ 表示矩阵 $\boldsymbol{A}$ 的 Frobenius 范数。对于矩阵 $\boldsymbol{B}_i^L$，其 Frobenius 范数的表达式为：

$$\left\|\boldsymbol{B}_i^L\right\|_F=\left(\sum_{j=1}^{m(i)}\left|f_{ij}^L\right|^2\right)^{1/2} \tag{27}$$

由此可根据 Frobenius 范数来构造多元伪泊松混合分布的权值。

3.8 根据 Frobenius 范数构建的权值

根据公式(27)，可构建由 Frobenius 范数构成的权值，设一共有 l 个多元伪泊松分布进行混合，则：

$$\rho_i^{\mathbb{M}}=\frac{\left\|\boldsymbol{B}_i^L\right\|_F}{\sum_{i=1}^{l}\left\|\boldsymbol{B}_i^L\right\|_F} \tag{28}$$

显然，$0\leqslant\frac{\left\|\boldsymbol{B}_i^L\right\|_F}{\sum_{i=1}^{l}\left\|\boldsymbol{B}_i^L\right\|_F}\leqslant\sum_{i=1}^{l}\left(\frac{\left\|\boldsymbol{B}_i^L\right\|_F}{\sum_{i=1}^{l}\left\|\boldsymbol{B}_i^L\right\|_F}\right)=1$，由此证明了权值 $\rho_i^{\mathbb{M}}$ 具有非负性即归一性，即 $0\leqslant\rho_i^{\mathbb{M}}\leqslant1$ 且 $\sum_{i=1}^{l}\rho_i^{\mathbb{M}}=1$。若令 $f_{ij}^B\left(x_{ij1},\cdots,x_{ijn(j)}\right)\equiv1$，则显然有：

$$\rho_i^{\mathbb{M}}=\frac{\left\|\boldsymbol{B}_i^L\right\|_F}{\sum_{i=1}^{l}\left\|\boldsymbol{B}_i^L\right\|_F}=\frac{1}{l} \tag{29}$$

其中，$l\neq0$。根据曼昆提出的有效市场假说[44]可知，当该理论成立时，混合分布的混合方式是随机游走的，故没有一种分布在混合分布中所占的权值是高于其余分布的，所以在有效市场假说的前提下，权值的大小是相等的，且它是一个与混合分布中分布的个数成反比的量，即一个算术平均数。本文将这种满足有效市场假说，权值为算术平均数的情况称之为算术平均的合理性，此时权值仅依赖于混合分布中分布的个数，而不依赖于其他参数。

3.9 基于 Frobenius 范数权值的多元伪泊松混合分布模型

对于 $Z_i\sim\Omega_i^{\tau}$，一共有 l 个多元伪泊松分布进行混合，设混合分布为 $\Omega^{\mathbb{M}}$，权值为 $\rho_i^{\mathbb{M}}$，则基于 Frobenius 范数权值的多元伪泊松混合分布模型的表达式为：

$$\Omega^{\mathbb{M}}=\sum_{i=1}^{l}\left(\frac{\left\|\boldsymbol{B}_i^L\right\|_F}{\sum_{i=1}^{l}\left\|\boldsymbol{B}_i^L\right\|_F}\right)\Omega_i^{\tau} \tag{30}$$

当 $f_{ij}^B\left(x_{ij1},\cdots,x_{ijn(j)}\right)\equiv1$ 时，有：

$$\Omega^{\mathbb{M}}=\sum_{i=1}^{l}(1/l)\Omega_i^{\tau} \tag{31}$$

4 仿真实验

模型可用来解决软件工程领域的问题，为突出模型的计算层次以及逻辑推理步骤，本文不会直接令 $\rho_i^{\mathbb{M}}=1/l$，而是根据计算层次来推导得出这一结论。另外由于获取现实数据存在难度，本文通过仿真实验的方式来进行模拟，所有程序均在 R 中完成编辑和运行。

某工业园有 A、B、C 三家软件研发企业，三家企业均在一年内开发了多款软件，每款软件的开发过程均满足瀑布模型。在开发过程中，每个阶段(可行性研究、需求分析、需要设计、详细设计、编码调试、单元测试、集成测试、确认测试、运行维护、退役)的风险损失次数满足 $X_{ij}\sim\Omega_{ij}^{\tau}\left(\lambda_{ij}\right)$，$X_{ij}$ 相互独立，且风险损失次数满

足 $0\leqslant X_{ij}\leqslant 4$，即损失超过 4 次以后就需要进行截断和均化处理，$\Omega_{ij}^{\tau}\left(\lambda_{ij}\right)$ 为参数 λ_{ij} 的从第 4 项后开始进行截断和算术均化处理的伪泊松分布。每个企业一年内软件开发过程全部软件的总风险损失次数为 $Z_i\sim\Omega_i^{\tau}$，$Z_i=\sum_j X_{ij}$，Ω_i^{τ} 为多元伪泊松分布。设三家软件研发企业一年内的混合总风险损失次数满足伪泊松混合分布 $\Omega^{\mathbb{M}}=\sum_{i=1}^{3}\rho_i^{\mathbb{M}}\Omega_i^{\tau}$，$0\leqslant\rho_i^{\mathbb{M}}\leqslant 1$ 且 $\sum_{i=1}^{l}\rho_i^{\mathbb{M}}=1$，假设损失次数满足有效市场假说，求三家软件研发企业的平均总混合风险损失次数，相实验关数据如表 1 所示。

使用 R 语言进行程序汇编，按照建模的步骤对基于 Frobenius 范数权值的多元伪泊松混合分布模型编程，然后代入表 1 中的数据进行仿真运算，得到结果后将其输出，结果如表 2 所示。

由实验结果可知，三家企业一年内总共开发的 377 款软件中，任意一款软件在满足瀑布模型的开发过程中发生的平均总混合风险损失次数为 6.562 619 次，显然可知 $Z_{\max}=40$，$Z_{\mathrm{mix}}=0$，较小 4 分位数 $Q_1=(n+1)/4=10.500\ 000>6.562\ 619$。表明工业园内三家企业一年内开发的 377 款软件产品的平均质量较好，平均可靠性较高，且混合分布的权值均为 0.333 333 3，满足了有效市场假说，即间接证明了算术平均的合理性。

5　模型比较与优缺点

虞欢欢[28]与本文均以泊松混合分布为基础，且均应用于风险分析领域，但虞欢欢[28]使用混合泊松模型来代替广泛使用的 Copula 模型对资产组合的相关性结构和组合风险计算的复杂性两方面做出实证研究。该研究与本文研究有相近之处，抛开研究目的不同这一点，两者主要有以下几点不同：

(1)权值的构造不同。虞欢欢[28]提出的模型通过结构模型求出债务人历史违约概率来得到混合泊松分布模型的权值。该权值通过历史经验数

表 1　仿真实验的相关数据图

Table 1　Data graph of simulation experiment

企业名称	研发阶段	参数	损失次数	软件总数
A	可行性设计	0.670	0,1,2,3,4	100
	需求分析	2.130	0,1,2,3,4	
	需求设计	0.800	0,1,2,3,4	
	详细设计	0.460	0,1,2,3,4	
	编码调试	0.310	0,1,2,3,4	
	单元测试	0.280	0,1,2,3,4	
	集成测试	0.920	0,1,2,3,4	
	确认测试	1.010	0,1,2,3,4	
	运行维护	0.230	0,1,2,3,4	
	退役	0.150	0,1,2,3,4	
B	可行性设计	0.270	0,1,2,3,4	120
	需求分析	1.320	0,1,2,3,4	
	需求设计	0.690	0,1,2,3,4	
	详细设计	0.380	0,1,2,3,4	
	编码调试	0.180	0,1,2,3,4	
	单元测试	0.760	0,1,2,3,4	
	集成测试	0.282	0,1,2,3,4	
	确认测试	1.210	0,1,2,3,4	
	运行维护	0.430	0,1,2,3,4	
	退役	0.210	0,1,2,3,4	
C	可行性设计	1.560	0,1,2,3,4	157
	需求分析	0.350	0,1,2,3,4	
	需求设计	0.260	0,1,2,3,4	
	详细设计	0.780	0,1,2,3,4	
	编码调试	1.040	0,1,2,3,4	
	单元测试	0.690	0,1,2,3,4	
	集成测试	0.370	0,1,2,3,4	
	确认测试	2.030	0,1,2,3,4	
	运行维护	0.140	0,1,2,3,4	
	退役	0.380	0,1,2,3,4	

表 2　仿真实验运算输出结果

Table 2　Output of simulation experiment

企业名称	Frobenius 范数	权值
A	3.162 278	0.333 333 3
B	3.162 278	0.333 333 3
C	3.162 278	0.333 333 3
期望	6.562 619	

据求得，而本文是通过多线性 Pseudo-Boolean 函数矩阵的 Frobenius 范数构造混合分布的权值，权值的构造可以使用往期的历史数据，也可以用当期观察数据，构造的方法并不完全依赖于历史经验数据。

(2) 模型结构不同。虞欢欢[28]的研究实际是一种风险分析方法，该方法主要包含信用风险结构模型、混合泊松模型两个模型。其中，信用风险结构模型用来构建混合分布模型的权数，因此虞欢欢[28]的研究方法实际上是两个模型，或者可认为是一个组合式模型。而本文自始至终只有一个模型，每一步研究得出的量均是模型的一部分。

(3) 研究对象不同。虞欢欢[28]主要针对的是信用资产的风险组合，而本文研究的是一般风险，并没有特殊的约定，仅在仿真实验中使用软件工程中软件开发瀑布模型的风险及风险值对模型进行了实现。

(4) 适用度不同。虞欢欢[28]提出的模型更适用于研究信用违约造成的风险，针对性强但普适性稍弱，尤其在缺少违约回收率数据时，使用该模型进行风险分析的精度会受到影响。而本文研究的模型可用于所有风险，只要风险满足或近似满足于多元伪泊松混合分布便可使用模型进行风险分析，但针对性并不强。

(5) 模型分量不同。虞欢欢[28]模型的分量均为泊松分布，而本文研究的混合模型中的分量是伪泊松分布，即经过了截断与均化处理的泊松分布。

但是，本文所提模型同样存在不足之处：

①模型的计算复杂度较大，映射关系层层叠加，操作过程中容易产生非系统风险；

②约束条件过多，模型的适用范围由此缩小；

③仅考虑各分布线性有限混合的情况，未考虑无限混合以及非线性情况。

6 总结与未来展望

多元伪泊松分布模型是一种特殊的混合分布模型，它能有效解决复杂的统计计算问题，同时可通过调整混合分布中分布的个数、截断的位置、随机变量的元数等数值型属性，满足不同人群解决不同问题的需求。

未来可继续深入研究这一问题，如将多线性 Pseudo-Boolean 函数进行 Lovász 延拓，将定义域从 $\{0,1\}$ 拓展到 $[0,1]$；或者从减少约束条件、模型复杂程度的角度上着手，同时可引入更多专业情景，在符合研究对象所属学科知识的客观条件下对模型进行改良。

参 考 文 献

[1] 曾裕峰, 向修海. 随机波动模型的扩展: 理论与中国股市的实证 [J]. 经济学, 2016, 16(1): 205-228.

[2] 肖佳文, 杨政. 混合分布的 VaR 非参数估计: 对期货市场的实证分析 [J]. 系统工程学报, 2016, 31(4): 471-480.

[3] 肖佳文. 基于混合分布的 VaR 估计及其应用 [D]. 成都: 电子科技大学, 2015.

[4] 张中文. 基于缺失数据的两正态混合分布的参数估计 [D]. 大连: 大连理工大学, 2008.

[5] 杨彬. 沪市 A 股交易量与收益率及波动性的关系—基于混合分布模型的实证研究 [J]. 统计与决策, 2005, (8): 91-93.

[6] 凌士勤, 杨波, 袁开洪, 等. 基于高频数据的分类信息混合分布 GARCH 模型研究 [J]. 数量经济技术经济研究, 2005, 22(3): 72-77.

[7] 杨维维. 贝叶斯模型在老年人健康管理效果评价中的应用 [D]. 南京: 东南大学, 2016.

[8] 孙维伟, 陈伟珂. 有限混合分布在车险费率厘定中的应用 [J]. 系统工程, 2016 (5): 144-153.

[9] 李程昊. 混合直流输电系统拓扑结构及关键技术研究 [D]. 武汉: 华中科技大学, 2015.

[10] 栾途. 透明质酸粘多糖的分子表征、流变学性质及其物理凝胶的研究 [D]. 上海: 上海交通大学, 2011.

[11] 刘向冲, 侯翠霞, 申维, 等. MML-EM 方法及其在

化探数据混合分布中的应用 [J]. 地球科学(中国地质大学学报), 2011, 36(2): 355-359.
[12] 胡昭华, 姜啸远, 王珏, 等. 混合深度网络在场景识别技术中的应用 [J]. 小型微型计算机系统, 2017, 38(6): 1387-1393.
[13] 何广才. 基于云平台的遗传—蚁群混合型算法的研究及在 TSP 中的应用 [D]. 呼和浩特: 内蒙古农业大学, 2016.
[14] 周颖. 标记分布在人脸图像属性识别上的应用 [D]. 南京: 东南大学, 2016.
[15] 降国栋. 基于云计算的计算机辅助诊断及远程诊断平台研究 [D]. 杭州: 杭州电子科技大学, 2016.
[16] 张晓庆. 基于计算机视觉的颜色恒常性算法研究 [D]. 天津: 天津科技大学, 2016.
[17] 赵挺. 蜂群算法及其仿生策略研究 [D]. 杭州: 浙江大学, 2016.
[18] 陈新. 基于人工鱼群算法的柔性作业车间调度研究 [D]. 大连: 大连理工大学, 2015.
[19] 吴建辉. 混合免疫优化理论与算法及其应用研究 [D]. 长沙: 湖南大学, 2013.
[20] 黄情操, 余达祥. 单变量边缘分布算法与蚁群算法的混合算法收敛性分析 [J]. 现代电子技术, 2012, 35(6): 74-77.
[21] 冯巍巍, 魏庆农, 汪世美, 等. 基于混合遗传算法的偏振双向反射分布函数优化建模 [J]. 红外与激光工程, 2008, 37(4): 743-747.
[22] 全星澄, 李巍. 基于 EM 算法的有限维混合分布参数估计研究 [J]. 统计与决策, 2017 (12): 25-29.
[23] 成英超, 王瑞胡, 胡章平. 一种基于高斯混合模型的协同过滤算法 [J]. 计算机科学, 2017, 44(S1): 451-454.
[24] 唐绩, 朱峰, 路彬彬, 等. 一种基于混合 Gamma 分布的自动目标识别混合 EM 算法 [J]. 现代雷达, 2017, 39(4): 45-49.
[25] 李保珠, 董云龙, 李秀友, 等. 基于 *t* 分布混合模型的抗差关联算法 [J]. 电子与信息学报, 2017, 39(7): 1774-1778.
[26] 冯杭, 王胜兵. 有限混合泊松分布参数优化的改进 EM 算法 [J]. 兵工自动化, 2017, 36(1): 80-82.
[27] 罗修辉, 韦程东, 王一茸. EM 算法在混合分布模型参数估计中的应用研究 [J]. 广西师范学院学报(自然科学版), 2016, 33(3): 35-39.
[28] 虞欢欢. 基于违约强度为混合泊松分布情形的信用资产组合集成风险度量研究 [D]. 杭州: 浙江财经大学, 2015.
[29] 高迎心, 温佳威, 徐尔, 等. 基于混合泊松分布的新生突变识别算法 [J]. 中国生物化学与分子生物学报, 2017 (11): 1168-1174.
[30] Wang DW. Fast hybrid level set model for non-homogenous image segmentation solving by algebraic multigrid [J]. Engineering and Technology Research, 2016, doi: 10.12783/dtetr/iceea2016/6727.
[31] Jiang D, Han DF, Hu XL. The shape optimization of the arterial graft design by level set methods [J]. Applied Mathematics: A Journal of Chinese Universities, 2016, 31(2): 205-218.
[32] 师肖静, 张兴芳. 概率逻辑、不确定逻辑和模糊逻辑之比较 [J]. 计算机工程与应用, 2017, 53(12): 50-52, 69.
[33] 钱玲, 张慧, 张化朋. 非可加测度空间上 Egoroff 定理的伪条件 [J]. 模糊系统与数学, 2016 (1): 77-83.
[34] 杨勇, 潘伟民, 徐春. 水平集函数规则化的 C-V 主动轮廓模型 [J]. 计算机工程与应用, 2008, 44(34): 166-168.
[35] 欧阳耀, 李军. Choquet 积分定义的单调集函数的几个遗传性质(英文) [J]. 东南大学学报(英文版), 2003, 19(4): 423-426.
[36] 李劲, 岳昆, 张德海, 等. 社会网络中影响力传播的鲁棒抑制方法 [J]. 计算机研究与发展, 2016, 53(3): 601-610.
[37] Zhao Y, Cheng DZ. Calculus of Pseudo-Boolean functions [C] // Technical Committee on Control Theory of Chinese Association of Automation, Systems Engineering Society of China, 2012.
[38] Marques Silva J. On applying lower bound estimates in Pseudo-Boolean optimization [C] // Proceedings of Guangzhou Symposium on Satisfiability and Its Applications, 2004.
[39] 傅文进, 吴小俊, 董文华, 等. 基于协同表示的子空间聚类 [J]. 模式识别与人工智能, 2017, 30(3): 251-259.
[40] 吴萱. 基于变分正则化的彩色滤波阵列(CFA)图像去马赛克方法研究 [D]. 南京: 南京理工大学, 2016.
[41] 张语涵. 鲁棒人脸正面化方法研究 [D]. 南京: 南京理工大学, 2016.
[42] 石聪聪. 矩阵 Frobenius 范数不等式及次可加性研究 [D]. 重庆: 重庆大学, 2016.
[43] 翁小清, 沈钧毅. 基于滑动窗口的多变量时间序列异常数据的挖掘 [J]. 计算机工程, 2007, 33(12): 102-104.
[44] Mankiw GN. 经济学原理: 宏观经济学分册第 7 版 [M]. 北京: 北京大学出版社, 2012.