

doi: 10.3969/j.issn.1000-8349.2018.01.02

# 星团成员辨认方法研究进展

谢安琪<sup>1,2</sup>, 齐朝祥<sup>1,2</sup>

(1. 中国科学院 上海天文台, 上海 200030; 2. 中国科学院大学 天文与空间科学学院, 北京 100049)

**摘要:** 对疏散星团进行成员星辨认是研究疏散星团的性质和确定疏散星团基本属性参数的首要步骤, 对利用疏散星团开展的恒星演化以及银河系结构等研究方向具有重要意义。利用恒星运动学方面的信息(自行和视向速度)来建立可靠的成员星辨认模型, 是成员星辨认领域中一个重要的研究方向。目前利用运动学数据进行成员星辨认较为成熟的模型主要是 Vasilevskis-Sanders 的方法和非参数方法。除此之外, 还有一些新兴的方法: 建立多维模型和聚类算法等, 但这些方法还有待于进一步探索。随着天体测量技术的不断发展, 测量数据精度不断提高, 提高成员星辨认模型可靠性的需求也越来越迫切。综述了用运动学方法对星团进行成员星辨认的研究背景和研究现状, 对各类算法进行了简要介绍, 总结了每种模型的优缺点。最后, 在 Gaia 时代下, 对成员星辨认模型的研究发展进行了展望。

**关键词:** 成员星辨认; 疏散星团; 自行; 视向速度

**中图分类号:** P154.1 **文献标识码:** A

## 1 引 言

疏散星团一直是恒星演化和银河系结构等领域的重要研究对象。在一片天区中把疏散星团的成员星从场星中分辨出来的过程称为成员星辨认。成员星辨认的准确度会直接影响疏散星团基本物理参数的估计和测量, 因此, 建立效率高、准确度高的成员星辨认模型对于确定疏散星团的属性参数以及后续的科学研究具有重要意义。

一般是利用成员星和场星在物理性质上的差异来进行成员星辨认。星团的所有成员星作为一个整体在一些恒星参数上有别于场星, 如相对自行、视向速度、红移、金属丰度等, 通过这些特性可以找到抱团存在的一些恒星, 也就是星团。

要建立一个成员星辨认模型, 首先要明确以哪些物理性质作为判据, 要考虑数据的精确度、场星以及成员星在此物理性质上的差异性和差异程度等, 可以通过一种特性, 也可以综

---

收稿日期: 2017-08-10; 修回日期: 2017-11-24

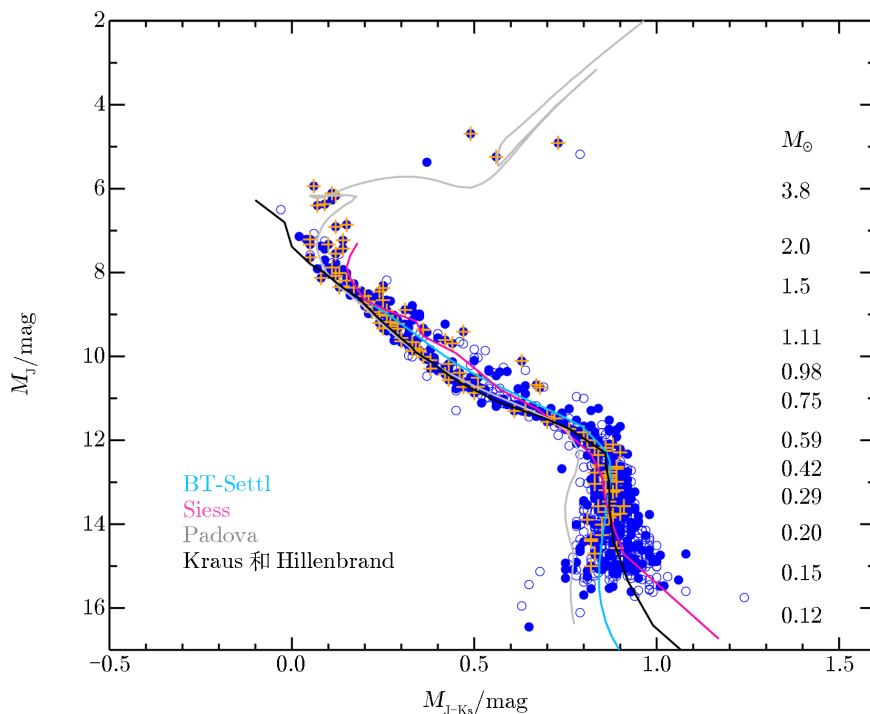
资助项目: 国家自然科学基金 (11573054, 11273003)

通讯作者: 谢安琪, xaq@shao.ac.cn

合采用几种物理特征建立多维的模型,例如综合考虑视向速度、光谱和自行数据等来找到星团;其次是要建立合理的数学模型来刻画场星和成员星在这些性质上的差异,设定判据找出星团成员星。同理,要衡量一个成员星辨认模型辨认结果的可靠程度,主要看此数学模型是否明确地刻画了星团成员星和场星的差异,以及差异的明显程度。

恒星可被直接测量到的物理性质一般分为两大类:亮度和运动。因此,目前成员星辨认的方法也主要分为测光方法和运动学方法。测光方法是利用恒星不同波段的测光数据画出区域内所有恒星的颜色-星等图,然后找到在主序上的恒星<sup>[1-3]</sup>。但对于很多星团,直接利用多色测光资料和颜色星等图找到清晰的主序是十分困难的,因此也难以给出定量的判定结果。另外,主序的形状和宽度也会受到很多因素的影响,包括恒星自转、星团年龄、双星系统等,颜色星等图主要反映星团演化以及星团成员的年龄信息。

很多利用测光方法进行成员星判定的研究,会综合使用运动学数据对恒星进行筛选。Wang 等人<sup>[4]</sup>先利用 PPMXL 星表的自行数据对所有恒星进行初步筛选,然后使用 2MASS 中 J 和 K 波段的测光数据,最后对 M44 星团进行成员星辨认,结果如图 1 所示。从颜色-星等图中可以发现, M44 的成员星(橙色的点)形成一条主序,可以用等龄线拟合,从而得到恒星的一些物理参数。



注: 橙色点表示成员星概率较高的点, 即视为 M44 的成员星, 蓝色点代表成员概率较低点。蓝色、红色、灰色以及黑色实线分别代表 BT-Settle<sup>[5]</sup>, Siess<sup>[6]</sup>, Padova<sup>[7]</sup>, Kraus 和 Hillenbrand<sup>[8]</sup> 这几种等龄线库。 $M_{\odot}$  表示太阳的质量。

图 1 M44 星团在 J 和 J-Ks 波段的颜色-星等图<sup>[4]</sup>

运动学方法是利用恒星的运动速度(自行和视向速度), 假定星团的所有成员星集体的运动规律会有别于场星, 建立严格的数学模型, 计算区域内每颗恒星是成员星的概率。两种方法各有优劣之处: 测光方法的优势在于数据丰富, 可以综合利用多波段的颜色-星等数据来找到成员星的分布区域, 缺点是难以给出很多星团量化的判定结果; 运动学方法可以弥补这一缺点, 但是当缺乏恒星视向速度数据或者自行数据精度差的时候, 会大大降低运动学方法判断的可靠性。

对于近距离的星团来说, 自行和视向速度方面的数据比较丰富、精度也较高, 采用运动学方法进行成员星辨认的可靠性要高于测光方法。因此对于近距离的星团, 从运动学角度建立效率高、准确度高的成员星辨认模型对于了解星团的基本属性以及进一步的研究具有重要作用。对于远距离的星团, 在无高精度运动学数据的情况下, 只能采用测光的方法。

在实际研究中, 除了模型准确度之外, 还要考虑到数据的测量难易度、完整度和准确度, 目前运动学方面较为常用的星表包括依巴谷星表、第谷星表 2、UCAC 系列和 PPMXL, 以及未来的 Gaia 星表。为了更充分地利用数据, 大多数研究会综合使用测光数据和运动学数据, 用来做成员星辨认的补充判据, 或者结果的交叉验证。如 Wu 等人<sup>[3]</sup>通过建立模型拟合恒星的 SED (spectral energy distribution) 数据来进行成员星辨认, 确定星团的基本物理参数, 然后与用运动学方法得到的成员星判定结果进行比对, 从而综合评判模型结果的可靠性。同理, 当数据精度越来越高的时候, 反过来会要求模型参数解算精度不断提高。因此, 新一代天体测量卫星 Gaia 带来更多更高精度的测量数据的同时, 也会对新一代成员星辨认模型的准确度和可靠性提出更高的要求。

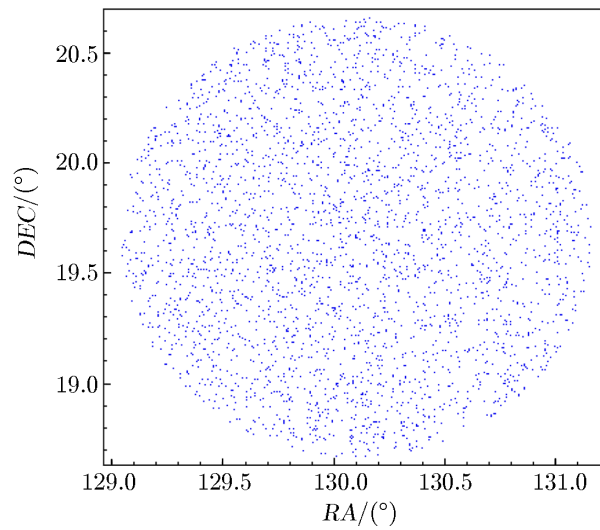
本文将主要介绍运用运动学方法进行成员星辨认的相关研究进展。第 2 章介绍目前最为成熟的 Vasilevskis-Sanders 方法并讨论模型的优缺点和局限性, 第 3 章介绍另一种非参数方法的发展情况, 第 4 章介绍目前建立辨认模型的新尝试和进展以及 Gaia 时代带来的巨大挑战, 最后一章进行总结与展望。

## 2 经典的 Vasilevskis-Sanders 方法

### 2.1 模型的提出与发展

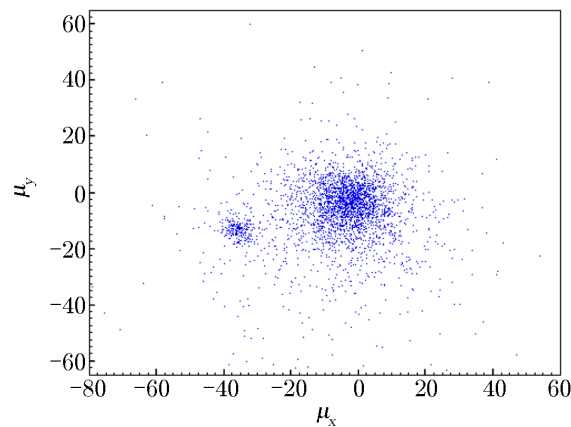
在位置空间中, 星团的成员星会成团存在, 不易与周围场星区别开来。图 2 为 M44 星团所在天区的恒星位置分布图, 可以看到通过位置空间的聚团效应寻找星团的方法并不可靠。而在速度空间, 由于星团的成员星相对于场星会有一个整体的速度, 一般会形成两个明显集合, 分别是星团成员星和场星集合。M44 星团所在天区的恒星自行矢点图如图 3 所示, 数据来源于 UCAC5<sup>[9]</sup>。图中可以看到所有的点聚集成两个集合。UCAC5 星表的自行精度在  $0.0012'' \cdot a^{-1}$  左右, 两个集合的平均自行之间的差异远大于测量误差, 因此从运动学上将星团成员星分离出来是可行的。

Vasilevskis-Sanders 方法利用运动学特性对疏散星团进行成员星辨认, 是目前使用最为广泛的方法。Vasilevskis 等人<sup>[10]</sup>提出了描述星团运动学数据的分布函数, 认为星团成员自行



注：利用 UCAC5<sup>[9]</sup> 的数据绘制， $RA$  和  $DEC$  分别表示赤经和赤纬，数据点为距离 M44 中心  $1^\circ$  天区内的恒星。

图 2 M44 所在天区的恒星位置分布



注：利用 UCAC5<sup>[9]</sup> 的数据绘制， $\mu_x$  和  $\mu_y$  表示恒星在赤经和赤纬方向的自行分量，数据点为距离 M44 中心  $1^\circ$  天区内的恒星。

图 3 M44 星团所在天区的恒星自行矢点图

分布可以用标准二维圆形正态分布函数来描述，分布函数中的弥散度为星团成员自行的内禀弥散度。而场星的自行内禀弥散度还受到太阳视差动和银河系转动等方面的影响，因此自行分布无法用二维圆形正态分布来描述。但是，在初始阶段可以用二维椭圆正态分布函数来拟合，其中，二维椭圆正态分布函数的主轴平行于银河系平面。Sanders<sup>[11]</sup> 在继承 Vasilevskis 等人的假设基础上，提出了“成员概率”的概念，用以表征这颗恒星属于星团成员的可能性大小；并发展出含 8 个参数的数学模型，计算出每颗恒星的“成员概率”，再利用最大似然

法对模型进行参数估计。

Vasilevskis-Sanders 方法提出后, 又经过了多人的完善和发展。Slovak<sup>[12]</sup>利用模拟的自行数据对 Vasilevskis-Sanders 模型进行测试, 证明了模型结果的唯一性和收敛性, 并指出此模型只能应用于没有明显内部运动的星团, 因此最适合用来研究疏散星团。另外, 在某些情况下, 该模型也适用于球状星团, 取决于球状星团周围场星和团星的速度差异与星团本身速度弥散度的比值。之后 Zhao 和 He<sup>[13]</sup>在 Vasilevskis-Sanders 模型基础上, 引进了每颗恒星自行数据的测量误差。除此之外, 星等对估计模型参数的影响也考虑进来。Shao 和 Zhao<sup>[14]</sup>将此方法进一步改进, 建立了一套严格的包含多种判据的数学模型, 应用于具有多重结构的复杂星团的成员星辨认工作中。Tian 等人<sup>[15]</sup>应用这种改进的模型, 成功计算出同一片天区中两个星团的参数和成员概率<sup>[16]</sup>。在疏散星团的成员星辨认研究领域, Vasilevskis-Sanders 方法是一种比较经典, 而且发展较为完善的研究方法。

## 2.2 模型介绍

假设  $\phi_c$  和  $\phi_f$  分别表示星团和场星自行分布的概率密度函数, 每颗星属于星团的概率可以表示为:

$$p_c = \frac{\phi_c}{\phi_c + \phi_f} \quad (1)$$

Vasilevskis-Sanders 模型认为星团和场星的分布与位置和自行有关系, 因此  $\phi_c$  和  $\phi_f$  可以用位置和自行的概率密度函数  $\Phi^r$  和  $\Phi^v$  的乘积表示出来, 即

$$p_c = \frac{\Phi_c^r \Phi_c^v}{\Phi_c^r \Phi_c^v + \Phi_f^r \Phi_f^v} \quad (2)$$

$\Phi^r$  和  $\Phi^v$  可以用高斯分布函数表示出来, 对星团的第  $i$  颗成员星, 相应的频率函数表示为:

$$\phi_c^v = \frac{1}{2\pi(\sigma_c^2 + \epsilon_{x_i}^2)^{1/2}(\sigma_c^2 + \epsilon_{y_i}^2)^{1/2}} \exp \left\{ -\frac{1}{2} \left[ \frac{(\mu_{x_i} - \mu_{x_c})^2}{\sigma_c^2 + \epsilon_{x_i}^2} + \frac{(\mu_{y_i} - \mu_{y_c})^2}{\sigma_c^2 + \epsilon_{y_i}^2} \right] \right\} \quad (3)$$

其中,  $(\mu_{x_i}, \mu_{y_i})$  是星团中第  $i$  颗星的在量度平面(照相底片、CCD 探测器平面)上的自行分量,  $(\mu_{x_c}, \mu_{y_c})$  是星团中心的自行分量,  $\sigma_c$  是星团成员自行分量的内禀弥散度,  $(\epsilon_{x_i}, \epsilon_{y_i})$  为第  $i$  颗星的观测误差。对场星来说, 相应的频率函数为:

$$\phi_f^v = \frac{1}{2\pi(1 - \gamma^2)^{1/2}(\sigma_{x_f}^2 + \epsilon_{x_i}^2)^{1/2}(\sigma_{y_f}^2 + \epsilon_{y_i}^2)^{1/2}} \exp \left\{ -\frac{1}{2(1 - \gamma^2)^{1/2}} \left[ \frac{(\mu_{x_i} - \mu_{x_f})^2}{\sigma_{x_f}^2 + \epsilon_{x_i}^2} - \frac{2\gamma(\mu_{x_i} - \mu_{x_f})(\mu_{y_i} - \mu_{y_c})}{(\sigma_{x_f}^2 + \epsilon_{x_i}^2)^{1/2}(\sigma_{y_f}^2 + \epsilon_{y_i}^2)^{1/2}} + \frac{(\mu_{y_i} - \mu_{y_c})^2}{\sigma_{y_f}^2 + \epsilon_{y_i}^2} \right] \right\} \quad (4)$$

其中,  $(\mu_{x_i}, \mu_{y_i})$  是场星中第  $i$  颗星在量度平面上的自行分量,  $(\mu_{x_f}, \mu_{y_f})$  是场星的中心自行分量,  $(\sigma_{x_f}, \sigma_{y_f})$  是自行分布的内禀弥散度,  $(\epsilon_{x_i}, \epsilon_{y_i})$  为第  $i$  颗星的观测误差,  $\gamma$  为相关系数。

对于星团成员星和场星的位置分布频率函数的表示, Balaguer-Nunez 等人<sup>[16]</sup>用正态分布

表示星团的位置分布。场星的位置分布被认为是均匀分布，即：

$$\Phi_c^r = \frac{n_c}{2\pi r_c^2} \exp \left\{ \frac{1}{2} \left[ \left( \frac{x_i - x_c}{r_c} \right)^2 + \left( \frac{y_i - y_c}{r_c} \right)^2 \right] \right\}, \quad (5)$$

$$\Phi_f^r = \frac{1 - n_c}{\pi r_{\max}^2}. \quad (6)$$

Krone-Martins 等人<sup>[17]</sup>提出可以直接使用恒星的坐标来表示位置分布：

$$\Phi_c^r = \begin{cases} n_c, & 2D \\ \frac{1}{1 + g^{-1} \exp \left\{ -\frac{1}{2(1-\rho_{\text{pos}}^2)} \left[ \frac{(\alpha_i - \alpha)^2}{\sigma_\alpha^2} + \frac{(\delta_i - \delta)^2}{\sigma_\delta^2} - \frac{2\rho_{\text{pos}}(\alpha_i - \alpha)(\delta_i - \delta)}{\sigma_\alpha \sigma_\delta} \right] \right\}}, & 4D \end{cases}, \quad (7)$$

$$\Phi_c^f = \begin{cases} 1 - n_c, & 2D \\ \frac{1}{1 + g \exp^{-1} \left\{ -\frac{1}{2(1-\rho_{\text{pos}}^2)} \left[ \frac{(\alpha_i - \alpha)^2}{\sigma_\alpha^2} + \frac{(\delta_i - \delta)^2}{\sigma_\delta^2} - \frac{2\rho_{\text{pos}}(\alpha_i - \alpha)(\delta_i - \delta)}{\sigma_\alpha \sigma_\delta} \right] \right\}}, & 4D \end{cases}, \quad (8)$$

其中， $n_c$  和  $1 - n_c$  是星团和场星位置分布函数的权重系数， $g$  是成员星与场星的数量比， $\alpha_i$  和  $\delta_i$  是第  $i$  颗星的赤道坐标， $\alpha$  和  $\delta$  为星团中心坐标， $\sigma_\alpha$  和  $\sigma_\delta$  是位置分布的中误差， $\rho_{\text{pos}}$  为赤经与赤纬之间的相关系数。二维 (2D) 情况下不考虑位置分布，四维 (4D) 情况下需要将位置信息加入到模型之中。模型参数估计采用最大似然法，对于  $N$  颗星有如下似然函数：

$$L(\theta_j; j = 1, \dots, n) = \ln \prod_1^N \Phi(\mu_{x_i}, \mu_{y_i}), \quad (9)$$

$$\frac{\partial L}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \ln \prod_1^N \Phi(\mu_{x_i}, \mu_{y_i}). \quad (10)$$

解上述方程组即可得到  $L(\theta_j; j = 1, 2, 3, \dots, n)$  的最佳估计值。

### 2.3 模型评价

Vasilevskis-Sanders 方法包含了一些前提假设<sup>[18]</sup>：

(1) 天区中的星仅被分成无交集的两大类——星团成员星和场星；

(2) 星团和场星在自行空间都满足二元正态分布；

(3) 星团成员星在自行空间的分布可以用二维圆形正态分布函数来拟合，其中弥散度为星团自行的内禀弥散度。

上述条件只是理想化的情况，实际上会受到很多因素的影响，如在真实的情况下，场星的自行分布有时候会有明显的重尾现象，不是标准的二维椭圆正态分布；实际的观测数据可能有较大的误差；一些星团的成员星分布并不会完全满足二维圆形正态分布等。除了这些影响因素之外，Galadí-Enríquez 等人<sup>[19]</sup>还详细讨论了模型假设可能造成的其他问题，应用二维圆形正态分布函数来描述星团的赤经自行和赤纬自行时，不能将星团的运动沿赤经和赤纬

分解开来, 只能用来描述二维平面中星团的整体运动规律。另外, 在赤经和赤纬两个方向上的观测数据也可能存在系统性的干扰和误差。而且, 用二维椭圆正态分布函数不能完全描述场星的自行分布, 因为场星的自行分布会受到太阳视差动和银河系运动等影响, 比较复杂。当场星的数量与成员星数量相当或者场星数量占优时, 用一个不精确的分布函数来描述场星分布导致的误差可能会淹没掉星团的信号, 极大地影响成员星辨认的准确度。这种情况下要用 Vasilevskis-Sanders 模型, 就需要一个更加精确刻画场星分布的函数, 例如叠加三个高斯分布来拟合场星分布情况。当星团与场星的两个分布函数的中心位置相聚过近从而无法有效地进行区分时, 对于自行较小的星团成员的判定也会带来较大的误差, 比如远距离的星团。应用 Vasilevskis-Sanders 方法时, 场星和星团成员的数量比也需要在一个合理的范围内, 成员过多或者过少, 对判定结果都不利。另外, 当星团距离太阳系过近时, 此方法也不适用<sup>[20]</sup>。

但 Vasilevskis-Sanders 模型仍然是目前为止最为成熟且应用最为广泛的经典模型, 不少星团成员辨认研究工作均在此基础上开展。由于基于同一个模型, 这也有利于不同结果间的比较。如 Dias 等人<sup>[21]</sup> 采用了 Vasilevskis-Sanders 的方法, 使用 UCAC4 星表的数据对所有有记录的星团进行了成员星辨认, 并将辨认的结果与其他人的工作进行了比对, 通过详细讨论结果之间的区别, 评估了不同星表数据的差异性, 以及一些星团特有的性质对于成员辨认结果的影响。

### 3 非参数方法

#### 3.1 模型介绍

为了避开 Vasilevskis-Sanders 方法的一些不足, Cabrera-Cano 和 Alfaro<sup>[18]</sup> 提出了一种非参数方法, 不需要用任何先验函数去拟合星团和场星, 而是根据运动学数据或其他物理特性经验性地找到概率密度函数。

假设在一个区域内有  $N$  颗星, 用  $n_a \times n_b$  的网格线来分割这片区域, 对于每一个网格节点  $(a_i, b_j)$ , 会相应地有一个密度值描述周围星的分布情况。如果区域内成员够多, 网格足够密集, 那么经验频率函数  $\Psi(a_i, b_j)$  会收敛于真实的频率分布函数  $\Phi(a, b)$ 。

问题在于如何计算节点  $(a_i, b_j)$  所在区域的密度值。一个可行的方法是利用核函数  $K(a, b)$ 。假设以  $(a_i, b_j)$  为中心,  $r$  为半径, 统计该圆形区域内所有点的数量, 除以圆的面积  $\pi r^2$  作为密度值, 则这里的核函数可定义为<sup>[19]</sup>:

$$K(a, b) = \begin{cases} \frac{1}{\pi r^2}, & [(a - a_i)^2 + (b - b_i)^2] < r^2 \\ 0, & \text{其他情况} \end{cases} \quad (11)$$

确定了核函数之后, 相应的经验频率函数可以定义为:

$$\Psi(a_i, b_j) = \sum_{k=1}^N K(a_k, b_k) \quad (i = 1, \dots, n_a; j = 1, \dots, n_b) \quad , \quad (12)$$

其中,  $(a_k, b_k)$  为网格中的第  $k$  个格点。核函数中的参数控制着频率函数的拟合程度, 被称为平滑参数, 过大则会过度拟合, 丢掉很多有用的信息; 过小则结果中的噪声过大, 可能会掩盖掉星团的信号。核函数  $h$  的选择可以参考如下公式:

$$h = \left( \frac{4}{d+2} \right)^{1/(d+4)} \sigma N^{-1/(d+4)}, \quad (13)$$

其中,  $d$  为空间维度, 这里讨论的是二维, 则  $d = 2$ ;  $\sigma$  是样本的标准方差, 表示为:

$$\sigma = \sum \sigma_i^2 / d, i = 1, \dots, d.$$

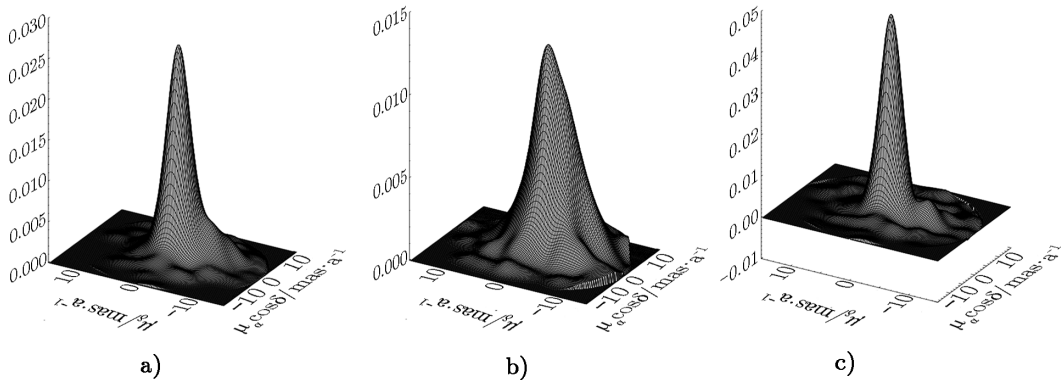
得到频率函数之后, 将其标准化就可以得到概率密度函数:

$$\Phi(a_i, b_j) = \Psi(a_i, b_j) \left[ \sum_{k=1, m=1}^{n_a, n_b} \Psi(a_i, b_j) \Delta a \Delta b \right], \quad (14)$$

其中,  $\Delta a$  和  $\Delta b$  是网格大小。

### 3.2 模型评价

相比于 Vasilevskis-Sanders 模型, 非参数方法的优点在于不需要先对星团和成员星的分布作出假设。利用非参数方法, 从天区中所有恒星在自行空间的整体分布函数中, 分离出场星和星团成员星的分布函数 (如图 4 所示)。除了运动学数据之外, 非参数方法还可以用于测光数据。Zhang 等人<sup>[23]</sup> 采用非参数方法处理了 LAMOST 光谱测光得到的视向速度数据, 成功地完成了成员星辨认工作。天文领域的数据经常不够全面, 质量参差不齐。对于运动学数据不够丰富而测光数据较全面的星团, 可以采用非参数方法, 并针对实际的数据特征进行相应的分析研究。



注: a) 距 NGC 1817 中心点周围  $20'$  天区内恒星的自行分布经验频率函数; b) 场星的经验频率函数; c) 星团成员的经验频率函数。

图 4 NGC 1817 的星团成员及周围场星在自行分布上的经验频率函数<sup>[22]</sup>

非参数方法的缺点也很明显, 如果数据的观测误差过大, 或者星团和场星的数据中心相接近时, 则星团和场星的信号不易区分, 这时候选取适当的平滑参数就显得十分重要。



Balaguer-Nunez 等人<sup>[16]</sup>分别采用了 Vasilevskis-Sanders 方法和非参数方法对 NGC 1817 进行成员星辨认工作, 并对这两个模型的结果做了详细的比对分析, 虽然辨认结果相差不大, 分析发现 Vasilevskis-Sanders 方法对初始参数的选取较为敏感, 因此需要选择合乎物理意义的参数; 非参数方法没有这个问题, 但模型没有包含每个数据的自行误差, 因此计算过程中亮星和暗星, 以及测量误差大的和误差小的都拥有同等权重, 对于结果的可靠性也有所影响。

## 4 Gaia 时代的新挑战

随着空间观测技术的发展, 天体测量的精度也越来越高。Gaia 巡天计划的顺利进行, 预示着天体测量领域将会进入一个新纪元。精确度更高的数据同时也会要求模型参数解算精度的提升, 因此有必要探索新的成员星辨认模型, 使得测量数据得到更好的应用。未来 Gaia 将会发布更多高精度的恒星运动参数数据, GES (Gaia-ESO Public Spectroscopic Survey) 还会提供视向速度、化学成分以及其他测光方面的信息, 因此将这些高质量的数据综合起来, 考虑建立多维度的成员星辨认模型是现在的发展方向, 如将这些数据综合起来建立十二维 (12-Dimension) 的模型<sup>[24]</sup>。Sampedro 和 Alfaro<sup>[25]</sup>提出了一种成员星辨认的方法, 在  $N$  维空间中计算每颗星距离星团中心的距离, 迭代估计出星团中心。 $N$  维数据包括位置、自行、视向速度等, 只要星团成员星的分布比场星密集, 即可进行成员星辨认。

另一种思路是应用机器学习领域的聚类算法, 与非参数方法相同, 聚类算法也不需要事先对成员星和场星分布做假设, 直接利用原始数据找到分布较为密集的区域, 以此确定星团的中心和成员。聚类算法的核心在于计算出每一个点与星团中心的距离 (大多数采用欧几里得度规), 然后最小化目标函数 (一般是距离函数), 迭代优化找到最优的星团中心位置。具体的聚类算法也分很多种, 如 Gao 等人<sup>[26, 27]</sup>利用 DBSCAN (density-based spatial clustering of applications with noise) 聚类算法建立三维模型处理绝对自行和视向速度数据, 对 NGC 6819 进行了研究。应用最广泛的是 K-means 算法, El Aziz 等人<sup>[28]</sup>详细介绍了 K-means 应用于成员星辨认工作的流程, 并应用于 NGC 188 和 NGC 2266 的辨认工作, 结果证明 K-means 是一种有效的成员星辨认方法。但作为一种新的尝试, 聚类算法的可靠性以及判定结果的准确程度还有待于进一步的研究论证。

## 5 总结与展望

疏散星团是恒星演化和银河系结构等研究方向的重要研究对象, 而成员星辨认是确定疏散星团的属性参数以及后续科学研究的关键的第一步, 因此研究成员星辨认具有重要的意义。成员星辨认的方法主要分为测光方法和运动学方法。两种方法各有优劣之处: 测光方法的优点在于数据丰富, 但无法建立严格的数学模型; 运动学方法可以弥补这一缺点, 但是对于测量数据的精度有一定要求。对于近距离的疏散星团, 由于自行和视向速度方面的数据较为丰富, 精度较高, 因此多采用运动学方法来找出星团成员。

进行成员星辨认需要考虑辨认所用到的物理性质以及数据的测量误差等因素,建立合适的数学模型并刻画出星团成员星和场星之间的差异。目前较为成熟的辨认模型包括经典的 Vasilevskis-Sanders 方法和非参数方法等。Vasilevskis-Sanders 方法假设场星和星团成员星均满足二维圆形正态分布,然后用分布函数去拟合整个天区内的所有恒星,找到星团的中心点和平均移动速度,并计算出每颗恒星的成员概率。事实上,场星的自行分布并不是标准的二维圆形正态分布,星团成员星的分布也会受到观测误差等误差项的影响,另外应用 Vasilevskis-Sanders 方法时还需要考虑场星数量与成员星数量比的大小。但 Vasilevskis-Sanders 方法经过多人的改进,目前仍然是应用最为广泛的经典模型,不少星团成员星辨认工作在此基础上开展。为了解决传统方法的问题,另一种非参数方法直接根据运动学数据或其他物理特性,经验性地找到场星和成员星分布的概率密度函数,这样就避开了先验函数的假设。当星团的运动学数据不够丰富而测光数据较全面时,适合采用非参数方法,针对实际的数据特征进行相应的分析研究。对于非参数模型,如果数据的观测误差过大,或者星团和场星的数据中心相接近时,场星和星团的差异性会不够明显。两种方法各有优劣之处,在具体使用的时候要根据具体的情况选择合适的方法。

随着天体测量的精度不断提高,尤其是 Gaia 巡天计划的顺利进行,未来会有更多更精确的运动学数据。这些高精度数据在提高成员星辨认结果准确度的同时,也对模型本身提出了更高的要求。目前,综合恒星位置、运动学数据、视向速度以及光谱信息等参数,建立多维度的成员星辨认模型是一个新的发展方向,除此之外,还有研究人员尝试利用聚类算法来建立成员星辨认模型。这些新的研究方向还处于探索阶段,未来也将会有更多新的模型和算法被投入应用,这些新的研究结果的可靠性和准确性还有待于更深入的研究。

## 致谢

感谢两位审稿人提出的宝贵意见以及编辑部的帮助。

## 参考文献:

- [1] Becker W. *Quarterly Journal of the Royal Astronomical Society*, 1972, 13: 226
- [2] Baade D. *Astronomy and Astrophysics Supplement Series*, 1983, 51: 235
- [3] Wu Z Y, Zhou X, Ma J, et al. *Publications of the Astronomical Society of the Pacific*, 2006, 118(846): 1104
- [4] Wang P F, Chen W P, Lin C C, et al. *The Astrophysical Journal*, 2014, 784(1): 57
- [5] Allard F. *Proceedings IAU Symposium*, 2013, 299: 271
- [6] Siess L, Dufour E, Forestini M. *arXiv preprint astro-ph/0003477*, 2000
- [7] Bressan A, Fagotto F, Bertelli G, et al. *Astronomy and Astrophysics Supplement Series*, 1993, 100: 647
- [8] Kraus A L, Hillenbrand L A. *The Astronomical Journal*, 2007, 134(6): 2340
- [9] Zacharias N, Finch C, Frouard J. *The Astronomical Journal*, 2017, 153(4): 166
- [10] Vasilevskis S, Klemola A, Preston G. *The Astronomical Journal*, 1958, 63: 387
- [11] Sanders W L. *Astronomy and Astrophysics*, 1971, 14: 226
- [12] Slovak M H. *The Astronomical Journal*, 1977, 82: 818
- [13] Zhao J L, He Y P. *Astronomy and Astrophysics*, 1990, 237: 54

- [14] Shao Z, Zhao J. *Acta Astronomica Sinica*, 1996, 37: 377
- [15] Tian K P, Zhao J L, Shao Z Y, et al. *Astronomy and Astrophysics Supplement Series*, 1998, 131(1): 89
- [16] Balaguer-Nunez L, Tian K P, Zhao J L. *Astronomy and Astrophysics Supplement Series*, 1998, 133(3): 387
- [17] Krone-Martins A, Soubiran C, Ducourant C, et al. *Astronomy and Astrophysics*, 2010, 516: A3
- [18] Cabrera-Cano J, Alfaro E J. *Astronomy and Astrophysics*, 1990, 235: 94
- [19] Galadí-Enríquez D, Jordi C, Trullols E. *Astronomy and Astrophysics*, 1998, 337: 125
- [20] van Leeuwen F. *Astronomy and Astrophysics*, 2009, 497(1): 209
- [21] Dias W S, Monteiro H, Caetano T C, et al. *Astronomy and Astrophysics*, 2014, 564: A79 GB/T 7714
- [22] Balaguer-Nunez L, Jordi C, Galadí-Enríquez D, et al. *Astronomy and Astrophysics*, 2004, 426(3): 819
- [23] Zhang B, Chen X Y, Liu C, et al. *arXiv preprint arXiv:1506.04222*, 2015
- [24] Gilmore G, Randich S, Asplund M, et al. *The Messenger*, 2012, 147: 25
- [25] Sampedro L, Alfaro E J. *Monthly Notices of the Royal Astronomical Society*, 2016, 457(4): 3949
- [26] Gao X H, Xu S K, Chen L. *Research in Astronomy and Astrophysics*, 2015, 15(12): 2193
- [27] 高新华, 陈力, 侯振杰. *天文学报*, 2013, 54(5): 439
- [28] El Aziz M A, Selim I M, Essam A. *Experimental Astronomy*, 2016, 42(1): 49

## The Researches of Open Cluster Membership Determination

XIE An-qi<sup>1,2</sup>, QI Zhao-xiang<sup>1,2</sup>

(1. *Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China*; 2. *School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China*)

**Abstract:** The identification of the membership of the open cluster is the first step to study the nature and to determine the basic physical parameters of the open cluster. It is also important for the research of stellar evolution and the structure of the Milky Way. It is an important research direction in the field of membership identification by using the kinematic information (proper motion, radial velocity) of the star to establish the precise membership identification model. At present, the most widely used models of using kinematic information to identify membership are Sanders' method and nonparametric method. Sanders' method assumes both field stars and cluster members distributed according to normal bivariate function, and uses position and proper motion data to calculate every star's membership probability. While nonparametric method makes less restrictive astronomical hypotheses, and allows the multi-use of different variables, such as radial velocities, proper motions, multi-color photometry data, etc. In addition, there are some new methods, including multi-dimensional model and clustering algorithm etc, but these new methods still need to be further explored. With the development of astrometry technology, the increasing accuracy of measurement data, and the demand for improving the model accuracy is becoming very urgent. This paper reviews the research background and the status of the membership identification of the star cluster

by kinematic methods, and introduces the various algorithms in detail. The advantages and disadvantages of each model are also summarized. Finally, in the coming Gaia era, the future research and development of membership identification models are presented.

**Key words:** membership determination; open clusters; proper motion; radial velocity

.....

## 《天文学进展》2018 年征稿启事

一、本刊刊登反映国内外天文学研究的最新进展和作者见解的述评、研究论文、前沿介绍、专题讲座、研究快报。反映新思路、新手段、新成果的短评优先发表。

二、来稿须包括：题名、作者署名、作者单位及所在地与邮政编码、摘要、关键词、正文、参考文献、英文摘要、英文关键词等。

三、来稿应论点明确、叙述精练、条理清晰、深入浅出，以利非本分支学科的读者阅读。

四、摘要应简明扼要地概括正文中的主要信息。采用第三人称的写法，不用“本文”、“作者”等作主语。稿件如果以中文发表，英文摘要应有近一页的篇幅。

五、关键词请尽可能参照 ApJ., MNRAS 和 A&A 共同采用的关键词表选用，可适当扩充。

六、稿件中引用论点、公式、图、表均需注明参考文献(按文中出现顺序编码，在右上角用小方括号标出)。正文后则按编码次序依本刊要求的编排格式列出相应的参考文献。

七、审定稿后需提供清稿样的电子文本。本刊采用中西文 Latex 系统软件排版，请从本刊网站 (<http://202.127.29.4/twxjz/index.htm>) 下载排版模版。

八、图、表务求精、简、规范，以配合文字阐明观点。公式务求精约，杜绝公式推导过程。

九、请勿一稿两投。来稿请通过 E-mail 发送至 [twxjz@shao.ac.cn](mailto:twxjz@shao.ac.cn)。

《天文学进展》编辑部

2018 年 2 月