RAA 2016 Vol. **16** No. **12**, 184 (5pp) doi: 10.1088/1674–4527/16/12/184

http://www.raa-journal.org *http://iopscience.iop.org/raa*

***R**esearch in*
***A**stronomy and*
***A**strophysics*

An application of the *k*-th nearest neighbor method to open cluster membership determination

Xin-Hua Gao

School of Information Science and Engineering, Changzhou University, Changzhou 213164, China; *xhgcczu@163.com*

Received 2016 August 10; accepted 2016 September 12

Abstract We introduce a non-parametric method for open cluster membership determination in three-dimensional (3D) velocity space (proper motion and radial velocity). Clean 3D cluster members can be obtained by statistically analyzing the Euclidean distance between each star and its k-th nearest neighbor in 3D velocity space. We use 513 sample stars in the direction of open cluster M67 to construct a 3D velocity space and test our method; 291 3D cluster members are obtained. The color-magnitude diagram, proper motions, radial velocities and spatial distribution of these 3D cluster members demonstrate the effectiveness of our method. From the 291 3D cluster members, the mean radial velocity and absolute proper motion of M67 are $V_{\rm r} = +33.46 \pm 0.05$ km s^{-1} and ($PM_{\rm RA}$, $PM_{\rm DEC}$) = $(-7.64 \pm 0.07, -5.98 \pm 0.07)$ mas yr^{-1}, respectively. In addition, we use 640 sample stars with precise proper motions and radial velocities in the direction of open cluster NGC 188 to test our method. The test results also show that our method is effective.

Key words: open clusters and associations: individual (M67)— Hertzsprung-Russell and C-M diagrams— stars: kinematics and dynamics

1 INTRODUCTION

Open clusters (OCs) are key tools for the study of stellar and Galactic disk evolution (Friel 1995; Chen et al. 2003; Bonatto et al. 2006; Wu et al. 2009). Membership determination is the first step in analyzing the fundamental properties of OCs because most of the known OCs are located in the Galactic thin disk where field star contamination is very serious (Bonatto et al. 2006; Dias et al. 2002). Up to now, several different methods have been proposed for membership determination (Vasilevskis et al. 1958; Sanders 1971; Missana & Missana 1990; Cabrera-Cano & Alfaro 1990; Zhao & He 1990; Javakhishvili et al. 2006; Wu et al. 2006; Gao 2014; Gao et al. 2015; Sampedro & Alfaro 2016), which can be divided into two types: parametric and non-parametric methods. Non-parametric methods are attractive in membership determination of OCs because they can work well without depending on any prior assumption about the distributions of cluster or field stars. We have used the non-parametric DBSCAN clustering method to determine memberships of two OCs, NGC 188 and NGC 6819 (Gao 2014; Gao et al. 2015), but it is difficult to determine appropriate input parameters (*Eps*, *Mpts*) in 3D velocity space. We empirically determined the input parameters in our previous work.

In this paper, we introduce a non-parametric method for membership determination of OCs, which can work well when precise proper motion (PM) and radial velocity (RV) data are available. This method can be used to search for cluster structure by analyzing the distribution of local point density in 3D velocity space. The basic principle of this method was first proposed by Ester et al. (1996), which has been used to roughly estimate the input parameters (*Eps*, *Mpts*) of the DBSCAN clustering algorithm in 2D space.

2 METHOD

The basic principle of this method is simple and it can be summarized as follows: (1) compute Euclidean distance d_{ij} between the i-th and j-th sample star ($i \neq j$) in 3D velocity space; (2) sort the distance d_{ij} in ascending order; (3) determine the distance of the k-th nearest neighbor D_k from the i-th star. Briefly speaking, we determine the Euclidean distance between the i-th star and its k-th nearest neighbor in 3D velocity space. Having obtained D_k for each star, the local point density $\rho(i)$ of the i-th star can be estimated using the following formula

$$\rho(i) = \frac{k+1}{\frac{4}{3} \cdot \pi \cdot D_k^3}. \quad (1)$$

We can draw two useful conclusions based on the above formula: (1) Given an appropriate value of k, the mean values of D_k for cluster members are much smaller than those of field stars. This happens because OCs can reasonably be regarded as dense structures in 3D velocity space and the typical velocity dispersion of OCs has been found to be much smaller than that of field stars (Girard

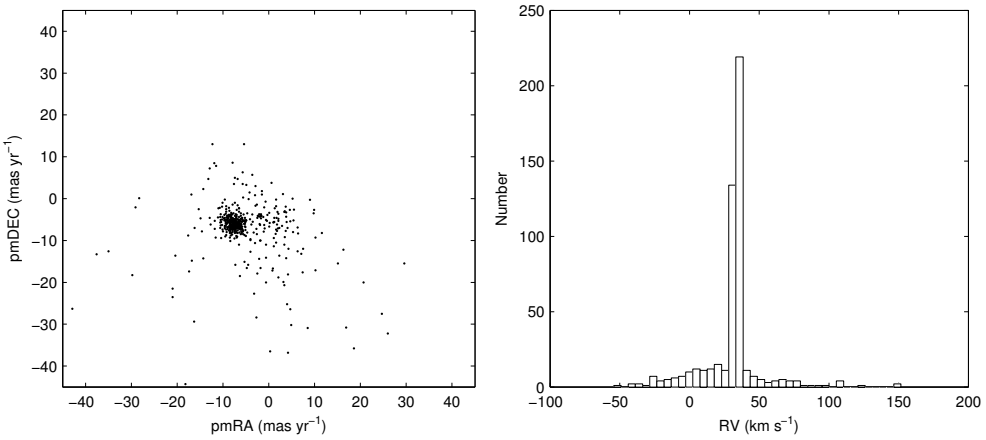

Fig. 1 PM VPD (*left*) and RV histogram (*right*) of the 513 sample stars.

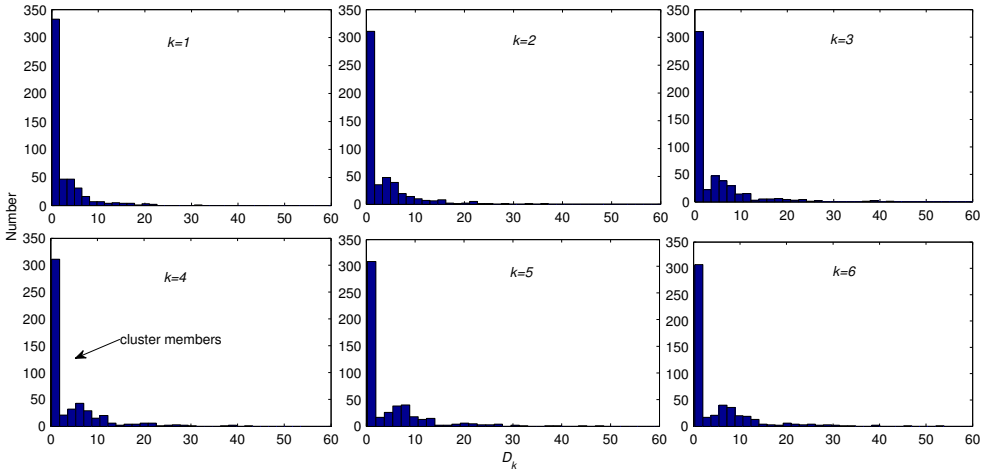

Fig. 2 Illustrations of the D_k for our sample stars with six different values of k. Only the stars with D_k less than 60 are displayed here for clarity.

et al. 1989). (2) Assuming a constant local point density for cluster members, a change in k does not lead to large changes in D_k. We will show that these two assumptions can be used for membership determination of OCs in the next section. To our knowledge, the k-th nearest neighbor method has never been used for OC membership determination.

3 CLUSTER MEMBERS OF M67 AND NGC 188

M67 is one of the best-studied OCs, and the fundamental parameters of this cluster have been well determined. We use 513 sample stars in the direction of the OC M67 to test this method. These sample stars have PM and RV data. The PMs and photometric data (2MASS) of these sample stars are selected from the PPMXL catalog (Roeser et al. 2010), and the observational errors of the PMs in both right ascension and declination directions are less than 3 mas yr^{-1}. The RVs of the sample stars are selected from the catalog compiled by Geller et al. (2015). Their measured precisions range from about 0.1 to 0.8 km s^{-1}. Our sample stars only include single stars. Binary candidates which show significant RV variability have been removed. In other words, only 513 single stars with precise PPMXL PMs are used to test our method. Although the total number of sample stars is not large, their kinematic data are precise enough for our purpose.

Figure 1 shows the PM vector-point diagram (VPD) and RV histogram of our sample stars. Given different values of k ($k = 1, 2, 3, 4, 5, 6$), we calculate D_k for each star. Figure 2 shows that values of D_k for about 300 stars exhibit only a slight increase with increasing k. As mentioned in Section 2, these stars can be regarded as cluster member candidates. As can be seen in Figure 2, we can use a threshold value of D_k to segregate cluster members. A threshold value of $D_4 = 1.5$ ($k = 4$) is adopted and 291 cluster members are obtained. The mean values of D_4 for the cluster members and field stars are about 0.66 and 9.71,

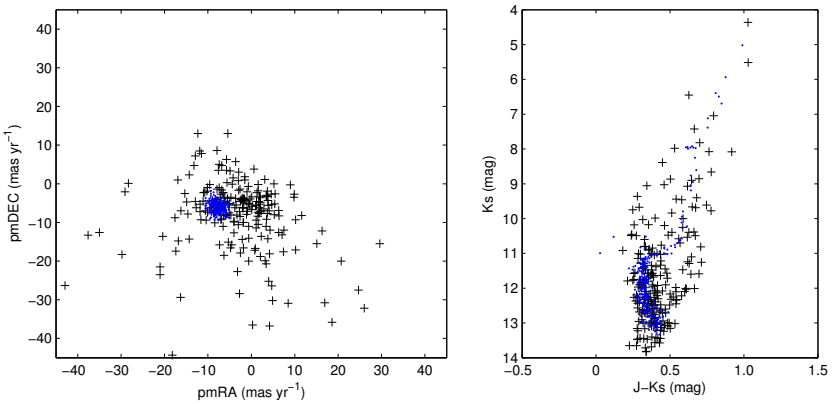

Fig. 3 (*Left*) The PM VPD of the 291 cluster members and field stars. (*Right*) The CMD of the 291 cluster members and field stars. In each panel, the *dots* and *pluses* represent the cluster members and field stars, respectively.

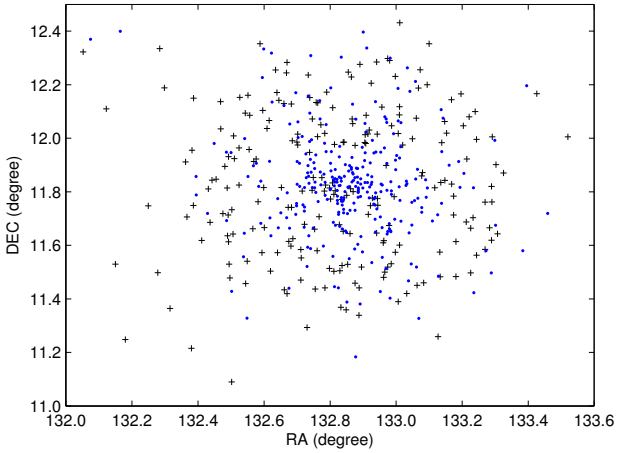

Fig. 4 Spatial distribution of the 291 cluster members (*dots*) and field stars (*pluses*).

respectively. As can be seen in Figure 3, the PM VPD and color-magnitude diagram (CMD) of the 291 cluster members demonstrate the effectiveness of our method.

Figure 4 shows the spatial distribution of the 291 cluster members and field stars. The cluster members are significantly more concentrated toward the cluster center compared with field stars. Among our 291 cluster members, 286 stars ($\sim 98\%$) are determined to be high-probability RV cluster members ($P > 0.7$) in the catalog of Geller et al. (2015), which indicates once again that our method is very effective.

We use the 291 3D cluster members to calculate the weighted mean RV and its corresponding uncertainty using the following formulas

$$V_{\rm r} = \frac{\sum(rv_i \times p_i)}{\sum p_i}, \quad (2)$$

$$\sigma V_{\rm r} = \left[\frac{\sum((rv_i - V_{\rm r})^2 \times p_i)}{(N-1) \times \sum p_i}\right]^{1/2}, \quad (3)$$

$$p_i = 1/(erv_i)^2, \quad (4)$$

where $V_{\rm r}$ and $\sigma V_{\rm r}$ are the weighted average RV and the corresponding uncertainty respectively, rv_i is RV of the i-th cluster member, p_i is weight of the i-th cluster member, erv_i is observational error of the i-th cluster member and N is the total number of cluster members. We find M67 to have a mean RV of $V_{\rm r} = +33.46 \pm 0.05$ km s^{-1}, which is quite consistent with the value derived by Geller et al. (2015). A mean PM of ($PM_{\rm RA}$, $PM_{\rm DEC}$) = (-7.46 ± 0.07, -5.98 ± 0.07) mas yr^{-1} is also derived using the same formulas, which is consistent with the value determined by Dias et al. (2014).

Additionally, we use sample stars with precise RVs and PMs in the direction of OC NGC 188 to further test our method. The PMs are selected from the catalog of Platais et al. (2003) and the RVs are selected from the catalog of Geller et al. (2008). A total of 640 single stars with RVs and PMs can be used for our test. A threshold value of $D_4 = 1.5$ ($k = 4$) is adopted and 353 cluster members are obtained (see Figs. 5–7), 341 of which ($\sim 97\%$) are determined to be high-probability RV cluster members ($P > 0.7$) in the RV catalog of Geller et al. (2008). The

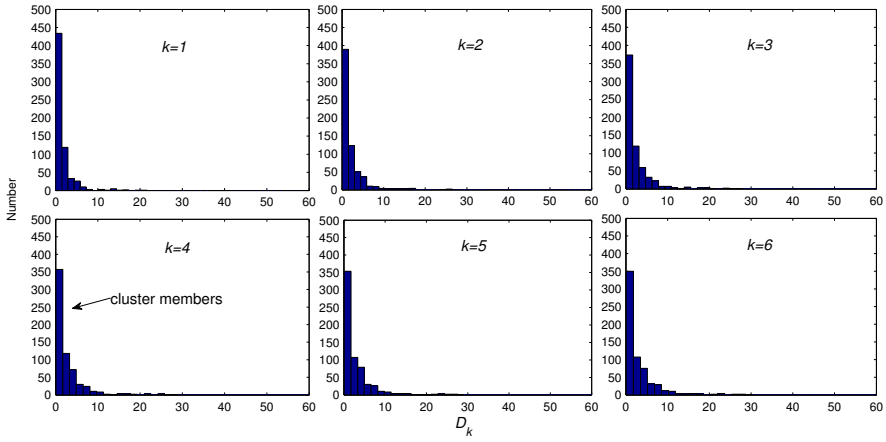

Fig. 5 The same as Fig. 2 but for NGC 188.

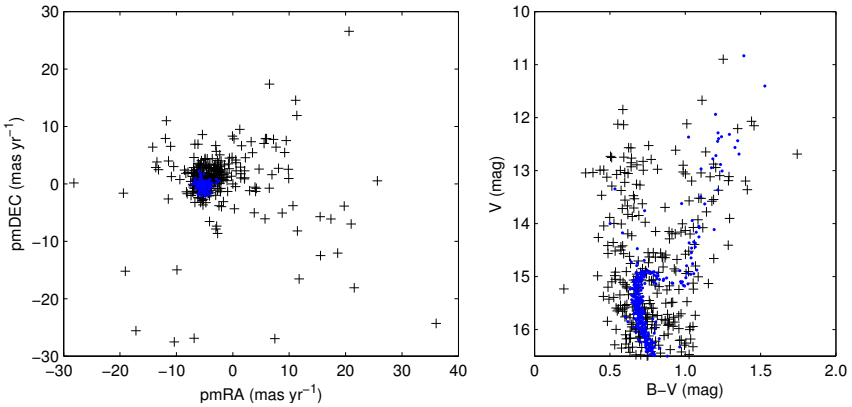

Fig. 6 The same as Fig. 3 but for NGC 188.

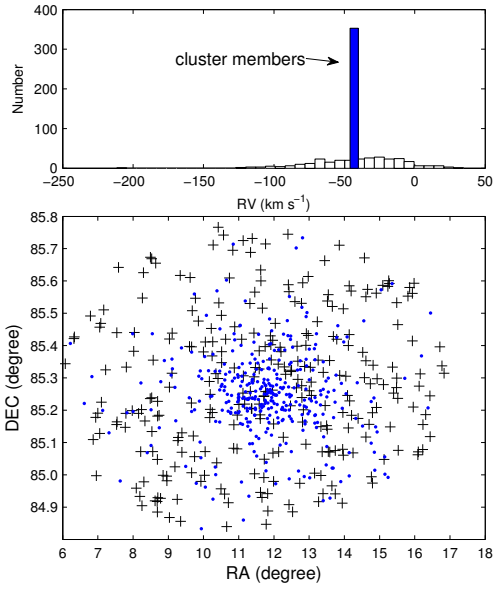

Fig. 7 (*Upper*) RV histogram of the 640 sample stars. (*Lower*) Spatial distribution of the 353 cluster members (*dots*) and field stars (*pluses*).

mean values of D_4 for the cluster members and field stars are about 0.36 and 5.53, respectively. So, we can safely say that our method is very effective for membership determination.

4 CONCLUSIONS AND DISCUSSION

Membership determination is the first step for the study of OCs. We develop a non-parametric method for membership determination of OCs. Our method is simple and easy to implement, and its effectiveness has been confirmed by identifying the 3D cluster members of OC M67 and NGC 188. If a sufficiently large sample of stars with precise PMs and RVs is available for OCs, then good results can be obtained based on our method.

Acknowledgements We are grateful to the anonymous referee for his/her insightful comments. This research was supported by the National Natural Science Foundation of China (NSFC, Grant No. 11403004). This research has made use of the VizieR catalog access tool, CDS, Strasbourg, France.

References

Bonatto, C., Kerber, L. O., Bica, E., & Santiago, B. X. 2006, A&A, 446, 121

Cabrera-Cano, J., & Alfaro, E. J. 1990, A&A, 235, 94

Chen, L., Hou, J. L., & Wang, J. J. 2003, AJ, 125, 1397

Dias, W. S., Alessi, B. S., Moitinho, A., & Lépine, J. R. D. 2002, A&A, 389, 871

Dias, W. S., Monteiro, H., Caetano, T. C., et al. 2014, A&A, 564, A79

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. 1996, in KDD-proceedings, 96, 226

Friel, E. D. 1995, ARA&A, 33, 381

Gao, X.-H. 2014, RAA (Research in Astronomy and Astrophysics), 14, 159

Gao, X.-H., Xu, S.-K., & Chen, L. 2015, RAA (Research in Astronomy and Astrophysics), 15, 2193

Geller, A. M., Mathieu, R. D., Harris, H. C., & McClure, R. D. 2008, AJ, 135, 2264

Geller, A. M., Latham, D. W., & Mathieu, R. D. 2015, AJ, 150, 97

Girard, T. M., Grundy, W. M., Lopez, C. E., & van Altena, W. F. 1989, AJ, 98, 227

Javakhishvili, G., Kukhianidze, V., Todua, M., & Inasaridze, R. 2006, A&A, 447, 915

Missana, M., & Missana, N. 1990, AJ, 100, 1850

Platais, I., Kozhurina-Platais, V., Mathieu, R. D., Girard, T. M., & van Altena, W. F. 2003, AJ, 126, 2922

Roeser, S., Demleitner, M., & Schilbach, E. 2010, AJ, 139, 2440

Sampedro, L., & Alfaro, E. J. 2016, MNRAS, 457, 3949

Sanders, W. L. 1971, A&A, 14, 226

Vasilevskis, S., Klemola, A., & Preston, G. 1958, AJ, 63, 387

Wu, Z.-Y., Zhou, X., Ma, J., & Du, C.-H. 2009, MNRAS, 399, 2146

Wu, Z.-Y., Zhou, X., Ma, J., Jiang, Z.-J., & Chen, J.-S. 2006, PASP, 118, 1104

Zhao, J. L., & He, Y. P. 1990, A&A, 237, 54