第60卷 第6期 天 文 学 报 Vol.60 No.6
2019年11月 ACTA ASTRONOMICA SINICA Nov., 2019

doi: 10.15940/j.cnki.0001-5245.2019.06.008

A 3D CNN molecular clump verification method for MWISP project *

ZHOU Piao[1,3] LUO Xiao-yu[1,3] ZHENG Sheng[1,3†] JIANG Zhi-bo[2] ZENG Shu-guang[1,3]

(1 College of Science, China Three Gorges University, Yichang 443002)
(2 Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210008)
(3 Center of Astronomy and Space Science Research, China Three Gorges University, Yichang 443002)

Abstract In order to investigate the differences between the molecular clouds that are associated with massive star forming regions and those not, we perform single dish simultaneous observations of ^{12}CO J = 2-1 and J=3-2 lines toward a sample of 59 Spitzer Extended Green Objects (EGOs) in the northern sky. Combining our results with the archive data of ^{12}CO J=1-0 observations toward the same sample of objects, we statistically investigate the correlations between the CO line widths and strengths for both those molecular clouds associated with EGO objects (EGO clouds) and those not (non-EGO clouds). We compare the different statistical behaviors between the two groups of molecular clouds, and interpret the differences in terms of density, temperature, and velocity field distributions. Particularly, it is found that both EGO and non-EGO clouds have similar mass ranges.

Key words Deep Learning, Convolutional Neural Networks, Molecular clumps, MWISP project

Classified index: P145; **Document code:** A

1 Introduction

The structures that can be detected by CO in interstellar medium are collectively referred to as molecular clouds, and the dense molecular cloud structure inside which stars may form is called clumps[1–2]. It has become the consensus of astronomers that stars originate in molecular cloud clumps. So far, the processes that convert molecular

Received 2019-12-24, revised version 2020-05-10

*Supported by National Key R & D Program of China no. 2017YFA0402700, the National Natural Science Foundation of China (NSFC) U1731124, No.10873037, No. 11873093 and No. 11803091

†zsh@ctgu.edu.cn

gas into stars is still unclear[3]. The Milky Way Imaging Scroll Painting (MWISP) project is dedicated to the large-scale survey of CO, and its isotopic molecules (i.e. ^{12}CO, ^{13}CO and $C^{18}O$), with (J=1-0) lines, along the northern Galactic Plane. The project is implemented with the Purple Mountain Observatory Delingha (PMODLH-13.7m) telescope[4]. The data product can be used to identify molecular clumps and analyze their properties, etc., thus provides a promising opportunity to study the earliest stages of star formation.

For a large-scale project, such as the MWISP survey, one needs full-automatic methods to analyse the data, especially in detecting clumps in various environments. While a number of methods have been developed for identifying clumps, they are actually not intelligent enough to do the work automatically. The typical algorithms include GaussClumps, ClumpFind, FellWalker, Reinhold, etc.

2 Method

2.1 Detection process

The whole process of the detection is shown in Fig. 1. Firstly we use ClumpFind to identify clump candidates within the original 3D data. The second step is to verify the candidates as positive or negative. Finally the catalog of the clumps is exported for future scientific studies.

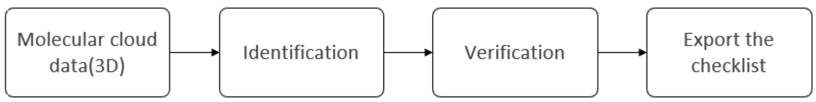

Fig. 1 A schematic view of the molecular clump detection process.

Table 1 Hyper parameters of two CNN models. (We have used the same set of hyper parameters for both the simulated and M16 datasets.

	Voxnet	Multi-view CNN
Activation function	ReLu	ReLu
Output layer	softmax	softmax
Learning rule	Adam	SGDM
Learning rate	0.0001	0.001
Batch size	400	128

2.2 Scoring the performance

According to the input and output, four statistics are obtained. They are: True-Positive(TP), True-Negative(TN), False-Positive(FP) and False-Negative(FN).

- TP: Number of samples belonging to the TRUE class, which are correctly labeled by the classifier as positive.
- TN: Number of samples belonging to the FALSE class, which are correctly labeled by the classifier as negative.
- FP: Number of samples belonging to the FALSE class, which are incorrectly labeled by the classifier as positive.
- FN: Number of samples belonging to the TRUE class, which are incorrectly labeled by the classifier as negative.

An ideal classifier requires both high recall rate (r) and detection precision (p), which are defined as:

$$r = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \, , p = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \, , \tag{1}$$

In order to combine the two parameters into a single score, we define F_1, which is the harmonic mean of p and r, i.e.

$$F_1 = 2 \cdot \frac{p \cdot r}{p + r} \, , \tag{2}$$

where the definition of F_1 indicates that r and p are of the same weight. F_1 ranges between 0 and 1, which can signify the performance of different methods. The value 1 indicates a perfect performance while 0 indicates poor.

3 Conclusion

We propose here a 3D CNN method, to identify and verify molecular clumps automatically for the MWISP data. The whole process of the method has two steps, identification and verification. The candidates are firstly obtained by classical algorithm ClumpFind with very low threshold and then verified with 3D CNN models. The method has the advantage over the existing molecular clump detection algorithms in that it does not need frequent interaction between machine and human in optimizing parameters. The experimental results on simulated data demonstrate that the performance of the 3D CNN method is better than those of the traditional methods in different kinds of environments. Applying this method to the real MWISP data (M16) suggests that it works satisfactorily as well. The Voxnet is a promising tool in automatically detecting molecular clumps with little human interactions for the large amount of 3D data in the big data era.

Reference

[1] Williams J P, Blitz L, Mckee C F. Physi, 2000, : 97
[2] Kauffmann J, Pillai T, Goldsmith P F. ApJ, 2013, 779: 185
[3] Francesco J D, Sadavoy S, Motte F, et al. A&A, 2010, 518: L91
[4] Zuo Y X, Li Y, Sun J X, et al. ChA&A, 2011, 35: 439
[5] Stutzki J, Guesten R. ApJ, 1990, 356: 513
[6] Williams J P, Geus E J D, Blitz L. ApJ, 1994, 428: 693
[7] Berry D. A&C, 2014, 10: 22
[8] Berry D S, Reinhold K, Jenness T, et al. ASCL, 2013, 376: 425
[9] Krizhevsky A, Sutskever I, Hinton G E. In International Conference on Neural Information Processing Systems. 2012: 1097–1105
[10] Russakovsky O, Deng J, Su H, et al. International Journal of Computer Vision, 2015, 115: 211
[11] Simonyan K, Zisserman A. Computer Science, 2014, : 1409
[12] Szegedy C, Liu W, Jia Y, et al. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: arXiv:1409.4842
[13] He K, Zhang X, Ren S, et al. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: arXiv:1512.03385
[14] Girshick R, Donahue J, Darrell T, et al. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2014, 580-587
[15] He K, Zhang X, Ren S, et al. ITPAM, 2014, 37: arXiv:1406.4729
[16] Ren S, He K, Girshick R, et al. ITPAM, 2015, 39: arXiv:1506.01497
[17] Redmon J, Divvala S, Girshick R, et al. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: arXiv:1506.02640
[18] Petrillo C E, Tortora C, Chatterjee S, et al. MNRAS, 2017, 472: 1129
[19] George D, Huerta E A. PhRvD, 2016, 97: 044039
[20] Maturana D, Scherer S. In IEEE/RSJ International Conference on Intelligent Robots and Systems. 2017: 922-928
[21] Hang S, Maji S, Kalogerakis E, et al. In Proceedings of the IEEE international conference on computer vision. 2015: 945-953
[22] Li C, Wang H C, Wu Y W, et al. RAA, 2020, 20: 031
[23] Song C. Master Degree Thesis, Shanghai Normal University, 2018

请补充中文标题和中文摘要

李志广[1,2,3] HEUVEL Edward[4]
(1 中国科学院云南天文台 昆明 650011)
(2 中国科学院天体结构与演化重点实验室 昆明 650011)
(3 中国科学院大学 北京 100049)
(4 Anton Pannekoek Institute for Astronomy, University of Amsterdam, Amsterdam NL-1098)

摘要 文中公式符号、图表、参考文献格式及引用规范等同中文模板. 为了极大地方便用户进行文章排版和录入, 我们制作了该LaTeX模板, 模板给出了基本的编写文章的模板样例. 注意摘要采用第3人称表述, 避免使用本人、本工作、我们等字眼. 请在http://www.twxb.org/下载最新模板, 如无必要不建议增加额外的扩展包、自定义命令, 以免造成正文字体、参考文献等排版格式不统一的情况! 如无特殊情况, 文章中的图片和表格环境需与模板一致. 摘要不宜引用正文中的文献、图、表、公式. 摘要中首次出现的英文缩写需注明全称, 如VLBI应写为VLBI (Very Long Baseline Interferometry)或VLBI (Very Long Baseline Interferometer), 中文名称也应写全称, 如中科院紫台应写为中国科学院紫金山天文台.

关键词 恒星: 形成, 星际介质: 云, 星际介质: 谱线与谱带, 星际介质: 分子, 亚毫米: 星际介质