

Звездные скопления: развитие знаний на основе интенсивного использования данных

© С.В. Верещагин,

© Н.В. Чупина,

© А.С. Фионов

Институт астрономии РАН

Москва

svvs@ya.ru

chupina@inasan.ru

fionov@mail.ru

Аннотация

Представлены результаты анализа информации о рассеянных звездных скоплениях (РЗС). Изучены динамика роста числа публикаций по РЗС, рост объема каталогов РЗС и звездных обзоров. Показано, что применение фильтров позволит обнаружить более 10 тысяч новых РЗС. Поставлена задача использования методов и инструментов интенсивного использования данных для работы со структурированной и неструктурированной информацией о РЗС.

1 Введение

На примере информационного обеспечения исследований рассеянных звездных скоплений рассмотрены особенности интенсивного использования данных в астрономии (см. также [20]). Цель работы – подготовка требований к разработке методов автоматического поиска и изучения звездных скоплений путем оперативного использования наиболее полной информации о РЗС. Объединение неструктурированной информации, содержащейся в публикациях, со структурированной информацией в обзорах звезд и каталогах РЗС, поиск ранее неизвестных скоплений и индексация звезд, входящих в их состав, позволят получить новые результаты об устройстве и кинематике, как диска Галактики, так и отдельных РЗС. Например, изучение морфологии области в Орионе позволило найти новую группу звезд [3]. Обнаружены и изучаются группировки звезд внутри скопления М67 [2]. Эти работы показали важность индексации звезд для верификации результатов путем подключения дополнительных параметров этих звезд из каталогов и публикаций (фотометрия, параллаксы, химический состав) и для поиска новых звезд, входящих в состав этих группировок.

В [9] (и других работах этих авторов) представлена информационная система, позволяющая осуществлять поиск новых РЗС. Система включает информационный блок,

состоящий из структурированной информации о каждом из скоплений, звездах в составе скопления и неструктурированную информацию, такую, как диаграммы и карты. Таким образом, существует тенденция перехода от простых файлов-каталогов к информационным системам, содержащим многообразную информацию об изучаемых объектах. Новый обзор GAIA (Global Astrometric Interferometer for Astrophysics) [15] позволит продолжить эту работу в направлениях поиска новых и изучения известных РЗС. Ниже сделана оценка ожидаемого числа открытий скоплений в ближайшем будущем и выполнен анализ роста числа публикаций по теме РЗС при помощи Astrophysical Data System [1]. Обсуждение и выводы содержат формулировки требований к возможности применения инструментов интенсивного использования данных к разнородной информации о РЗС.

2 Рассеянные звездные скопления

РЗС представляют собой группировки с численностью звезд от десятков до тысяч в их составе, [10]. На небе скопления выглядят как нерегулярные группы различных размеров, часто с заметным повышением концентрации звезд к центру. Внутри скопления звезды связаны гравитационно, часть из них уже покинула скопление путем диссипации. Звезды в скоплении имеют примерно одинаковые возрасты, хотя существуют скопления с продолжающимся звездообразованием и по этой причине могут содержать молодые звезды. Несмотря на обилие данных наблюдений и множество публикаций, внутреннее устройство РЗС и система скоплений в диске изучены недостаточно. Для РЗС наиболее надежно определены расстояния от Солнца, возрасты, химические составы и т. п. Располагаясь на различных расстояниях от Солнца, они являются ключевыми объектами для получения новых результатов, важных для понимания деталей строения Галактики. Особое место занимают исследования внутреннего устройства и кинематики скоплений. Для этого важно пополнять информацию об отдельных звездах в составе скоплений путем индексации и применения инструментов интенсивного использования данных.

Большинство РЗС образовалось очень давно, но молодых скоплений наблюдается больше, чем

Труды XVIII Международной конференции DAMDID/RCDL'2016 «Аналитика и управление данными в областях с интенсивным использованием данных», Ершово, 11-14 октября 2016

старых. Такое распределение отражает факт распада скоплений со временем: взаимодействуя друг с другом, некоторые звёзды приобретают скорости, большие, чем скорость отрыва, и покидают скопление. Молодые рассеянные скопления расположены вблизи плоскости Галактики на высоте не превосходящей 200 пк, что близко к характеристикам распределения поглощающей материи. В большей степени отходят от плоскости Галактики наиболее старые скопления, некоторые из них достигают высоты до 2.2 кпк. Для старых скоплений не заметна концентрация к плоскости Галактики. Возможно, быстрее разрушаются те скопления, которые движутся вблизи галактической плоскости. Исследования здесь далеки от завершения – альтернативная гипотеза состоит в предположении, что ранее скопления рождались в более толстом слое диска, чем в настоящее время.

Интересно, что в составе одного скопления можно наблюдать звезды самых разных типов, кратные системы и группировки. Сосредоточение разнообразных объектов на площадке в несколько квадратных градусов на небе, занимаемой скоплением, – хорошая лаборатория для исследования звезд. Особенно удобны такие наблюдения для поиска звезд с экзопланетами.

3 Рост публикаций об РЗС

В начале 1980-х гг. число известных РЗС, содержащихся в “карточном” каталоге Рупрехта, было около 1200. В настоящее время по данным [14] их насчитывается 3784 и это далеко не предел. Всего по приближенным оценкам в Галактике более 100 тысяч РЗС, и основные открытия еще впереди.

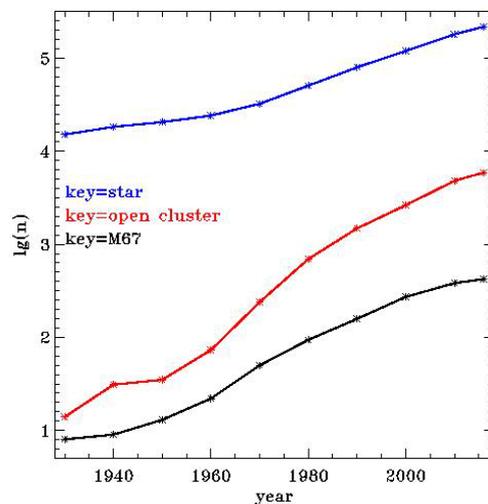
Таблица 1 Рост информации со временем

Дата, до (год)	“M67”	“open (and) cluster”	“star”	k
1930	8	14	15103	1079
1940	9	31	18313	590
1950	13	35	20590	588
1960	22	73	24138	330
1970	50	239	32429	136
1980	94	695	50711	73
1990	158	1483	79757	54
2000	272	2632	119757	46
2010	383	4815	180543	37
2016	422	5914	218392	37

С помощью системы ADS [1] мы оценили рост информации по РЗС, звездам и отдельному скоплению М67. В Таблице 1 приведено количество публикаций (по наличию в заголовке статьи ключей “open (and) cluster”, “star” и “M67”) до указанного года. Коэффициент k – отношение чисел в колонках “star” и “open cluster”. С его помощью показан рост популярности исследований скоплений по сравнению с исследованиями звезд: начиная с 1980-х гг. он заметно уменьшился (с ростом года в колонке

“Дата”), что свидетельствует о возрастании числа запросов по ключу “open (and) cluster” относительно “star”.

На Рис. 1 видно, что информация об РЗС (ключ “open (and) cluster”) растет опережающим темпом по сравнению с более общей темой про звезды (ключи “star” и “M67”). Отметим увеличения скорости роста числа статей по РЗС в 1950-е годы и по звездам в 1970-е. Для РЗС это было связано, скорее всего, с резким увеличением числа наблюдений. Для звезд с 1970-х годов появление крупных ЭВМ позволило резко увеличить исследования по их физике и эволюции. Отметим, что число публикаций по звездам значительно превосходит число публикаций по РЗС. Это вполне естественно, если учесть огромное превосходство в численности и разнообразии звезд в сравнении с РЗС. Сегодня намечился спад публикаций по РЗС, что мало заметно и для звезд. Для М67 рост также замедляется, что иллюстрируют кривые на Рис. 1.



- Логарифм числа статей (n) с ключами в заголовках статей “star”, “open (and) cluster”, “M67”, вышедших до указанной даты (year)

4 Звездные обзоры

Поиски и открытия новых РЗС происходят в основном путем применения различных фильтров к распределениям звезд из современных обзоров, краткая информация о некоторых наиболее крупных из них представлена в Таблице 2. В Таблице 2 представлены поверхностные звездные плотности (P), численности звезд, процент охвата неба и предельные звездные величины рассмотренных обзоров. Обзор GAIA [15] включает приблизительно 1 млрд. звезд, что составляет приблизительно 1% звездного населения Млечного Пути. Предельная звездная величина равна 20 в системе G (на интервале длин волн от 400 до 1000 нм). Микросекундная точность измерений позволяет получить новую информацию о кинематике и строении РЗС. Точность достигнута во многом благодаря сверхдальному (более 1 млн. км от Земли) расположению ИСЗ в точке Лагранжа (L2),

исключающей влияние на положение аппарата Земли–Луны и Солнца. Исключаются засветка от Земли, Луны (отраженного света от Земли), а также влияние переходов из света в тень. Последние мешают работе близких аппаратов, таких, как телескоп Хаббл. Кроме того, поддержание аппарата в точке L2 наиболее энергетически выгодно [7].

Таблица 2 Поверхностные звездные плотности в звездных обзорах

Обзор	Число звезд	Охват неба, %	P, зв./ кв. град.	Пред. зв. вел.
GAIA	10**9	100	24240	G=20
UCAC4	0.113*10**9	100	2739	R=16
2MASS	0.300*10**9	70	4848	J=15.8

UCAC4 (United States Naval Observatory CCD Astrograph Catalog) [19] представляет собой компилятивный каталог, включающий 113,780,093 звезд главным образом на интервале от 8 до 16 зв. величины в полосе пропускания между V и R. Ошибки положений составляют от 15 до 20 мкс для звезд в диапазоне от 10 до 14 зв. величин. Получены собственные движения для большинства звезд. Для этого использовались около 140 звездных каталогов с существенным различием эпох наблюдений. Эти данные дополнены фотометрией из 2MASS приблизительно для 110 миллионов звезд. В частности, 5- полосная (B, V, г, г, I) фотометрия подключена более чем для 50 миллионов звезд [6]. Обзор UCAC4 полон для ярких звезд (приблизительно до R = 16 зв. величины). По номерам звезд обеспечена связь с каталогом Hipparcos [18], где есть дополнительные данные, например, параллаксы.

2MASS (Two Micron All-Sky Survey) [17] – это обзор неба в ИК диапазоне в полосах пропускания J (1.25 μm), H (1.65 μm) и Ks (2.17 μm). Осуществлен в 1997–2001 годах с помощью телескопов в северном (Обсерватория имени Уиппла, Аризона) и южном (Чили) полушариях. Каталог точечных источников содержит положения и потоки для приблизительно 300 млн. звезд до J=15.8.

5 Фильтры для поиска РЗС

Применение фильтров к звездным обзорам позволяет многократно увеличить число открытых РЗС. Большие перспективы открываются в ближайшем будущем, как уже говорилось, с поэтапной реализацией проекта GAIA. Первые результаты появятся в общем доступе в сентябре 2016 года. Доступные данные будут постепенно увеличиваться в ближайшие годы [15]. Объем обзора достигает 200 Тб.

О методах выделения пространственных флуктуаций звездной плотности (фильтрах) см. [13]. Для выявления мест повышенной концентрации звезд применяются методы “ближайшего соседа”, детализации полигональной сетки Вороного,

нахождения минимального остовного дерева. Чтобы понять, являются ли выделенные уплотнения звезд реальными РЗС, применяются кинематический, фотометрический, статистический и другие критерии.

Кинематический критерий [4] основан на методе диаграмм апексов (AD-диаграмм), разработанном авторами. Он применялся нами для исследования кинематики Гиад, Яслей, скоплений и групп в Орионе, потока Большой Медведицы. AD-диаграмма представляет собой распределение апексов звезд в экваториальной системе координат. Их координаты получаются из решения геометрической задачи, в которой находятся пересечения векторов пространственных скоростей звезд с небесной сферой, при этом начала векторов перемещены в точку наблюдений. По аналогии с обычным апексом координаты этих точек можно назвать индивидуальными апексами звезд. Как отмечается в [22], РЗС обнаруживались даже с помощью простого визуального просмотра распределений звезд – по данным обзоров DSS и 2MASS в [12] были обнаружены 66 кандидатов в скопления. В работе [6] путем автоматического поиска флуктуаций плотности по картам распределения звездной плотности, построенных по данным 2MASS, обнаружено 681 скопление. В [11] разработан новый эффективный метод поиска пиков звездной плотности на звездных картах, построенных по многим (79) обзорам. Этот метод основан на свертке карт плотности со специальным двумерным фильтром.

6 Сравнение поверхностных плотностей

Каково же число потенциальных открытий РЗС в обзоре GAIA? Мы сделали такую оценку путем сравнения средних плотностей звезд в скоплениях (ρ) и обзорах (P). Плотности P приведены в Таблице 2. В РЗС ρ зависит, например, от возрастов (Рис. 2 и 3) и, главным образом, от расстояния скопления от Солнца (Рис. 2). На Рис. 2 и Рис. 3 показаны эти зависимости, которые мы построили по данным каталога MWSC II [10].

Конкретный пример - скопление M67 рассмотрено особо. Мы взяли его средние параметры – площадь 3 кв. град. и численность 1000 звезд в его составе. В этом случае плотность ρ равна 318 звезд на кв. град. Как видим из Таблицы 2, все звезды M67 с большим запасом входят в рассмотренные обзоры, поскольку P любого из обзоров, как минимум на порядок, превосходит ρ для M67.

Плотность в GAIA (Таблица 2) примерно в три раза превосходит плотность максимально плотного РЗС в MWSC II (до 8000 звезд на кв. град., Рис. 2). Это означает, что в GAIA содержится по крайней мере в три раза больше РЗС, чем в [10] - 3006. Следовательно, в GAIA содержится приблизительно 10 тыс. ранее неизвестных РЗС, которые ждут первооткрывателя. Отметим, что в 2MASS и UCAC4

входят далеко не все скопления [10], что можно определить по их значениям P относительно GAIA в Таблице 2.

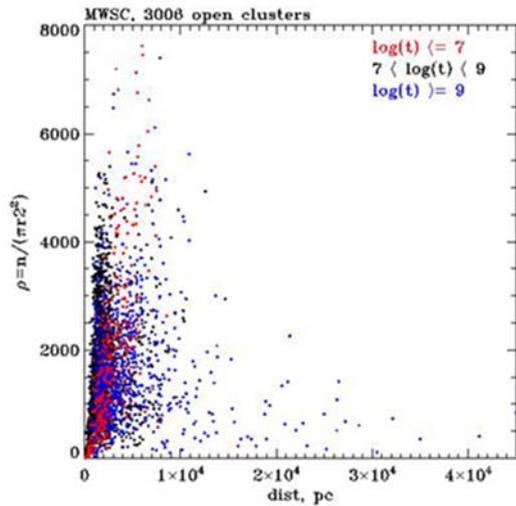


Рисунок 2 Распределение РЗС из MWSC II. По оси ординат – поверхностная плотность ρ звезд в скоплениях, по оси абсцисс – расстояние скоплений от Солнца (dist). Цветом обозначены РС разных возрастных диапазонов, подписанных в справа сверху

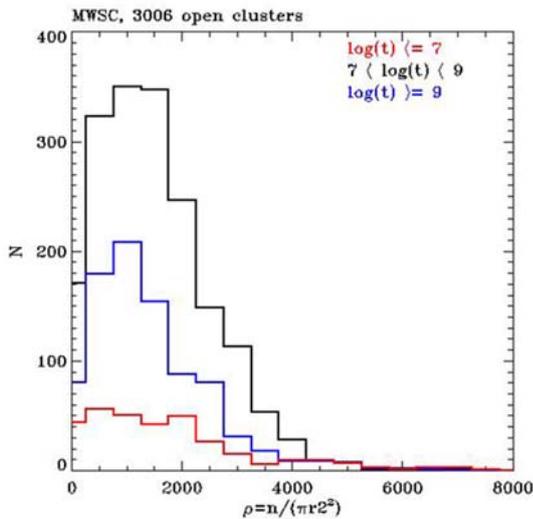


Рисунок 3 Число скоплений (N) в зависимости от ρ . Гистограммы построены отдельно для РЗС разных возрастов, обозначенных цветом и подписанных справа сверху

7 Метаданные

Для одновременной работы с каталогами РЗС, публикациями и звездными обзорами нужны метаданные. Ими являются номера РЗС в различных каталогах, их экваториальные координаты и параметры - радиус, расстояние от Солнца, химический состав и т. п.. В состав метаданных также входит номер звезды внутри скопления. В [9] предложен вариант нумерации WOCS звезд в скоплениях ID (IDW) и индексация ID (IDX). В свою очередь ID основаны на звездной величине в фильтре V и расстоянии звезды от центра скопления. В

качестве метаданных целесообразно использовать и неструктурированную информацию, которая основополагающа при исследовании РЗС - диаграммы Герцшпрунга-Рессела (H-R), двуцветные диаграммы (CMD) и диаграммы собственных движений (PMD).

8 Заключение

Показано присутствие в информации о РЗС характеристик Big Data (так называемых «трех V», [8, 23]): гигантского объема (volume) постоянно растущей структурированной (обзоры, каталоги) и неструктурированной (публикации) многообразной (variety) информации. Степень изученности звездного населения Млечного Пути составляет 1% (как уже говорилось, GAIA включает данные об 1 из 100 млрд. звезд Галактики). В GAIA процент охвата РЗС составляет также около 1% всех РЗС в Галактике, предоставляя возможность открыть еще 10 тыс. скоплений.

Важным в развитии метаданных является проведение индексации звезд, входящих в состав РЗС. Это ускорит наполнение данными, поиск новых членов скоплений и другие исследования. Повышение скорости необходимо также для организации совместного поиска в нескольких обзорах (на сегодня – GAIA, 2MASS), каталогах РЗС и системе публикаций ADS (с возможностью поиска по параметрам).

Скорость прироста (velocity) составляет несколько тысяч публикаций в год, наблюдается многократное увеличение объема обзоров (Таблица 2) и числа открытий РЗС. Применение инструментов интенсивного использования данных к РЗС будет эффективным для получения всесторонней информации о выбранном скоплении. Физические параметры, результаты наблюдений, публикации и другая информация – это обычная подготовка материалов для исследований. Автоматизация с помощью инструментов интенсивного использования данных ускорит процесс сбора материалов.

Подход к объединению структурированной и неструктурированной информации об РЗС существует в CDS [16]. Здесь разработан аппарат поиска каталогов РЗС и данных о звездах без объединения поисковых возможностей. Разработана удобная система для объединения и обработки данных TOPCAT. Система работы с данными GAIA также будет удаленной, [15]. Авторами [10] создана система Каталогов рассеянных скоплений Млечного пути. Система содержит результаты поиска новых РЗС на базе каталога 2MASS, а также множество параметров скоплений, отдельно оформлены точки входа для карт и диаграмм.

Благодарности

Авторы благодарны QI и рецензентам за помощь. Работа поддержана грантом РФФИ 16-52-12027 ННИО_а.

Литература

- [1] Astrophysical Data System, SAO/NASA ADS Astronomy Abstract Service <http://adsabs.harvard.edu/abs/>
- [2] N.V. Chupina, S.V. Vereshchagin. Stellar clumps within the corona in the open cluster M 67 // *Astronomy and Astrophysics*, V. 334, p. 552–557, 1998.
- [3] N.V. Chupina, S.V. Vereshchagin. Star clusters in the Sword region in Orion // In Proceedings of the 33rd ESLAB Symposium on Star formation from the small to the large scale, p. 347–349, edited by F. Favata, A. Kaas, and A. Wilson, ESA SP 445, Noordwijk, The Netherlands, ESA, 2000.
- [4] N.V. Chupina, V.G. Reva, S.V. Vereshchagin. The geometry of stellar motions in the nucleus region of the Ursa Major kinematic group // *Astronomy and Astrophysics*, V. 371, p. 115–122, 2001.
- [5] R.M. Cutri и др. 2MASS (The Two Micron All-Sky Survey) The 2MASS All-Sky Catalog of Point Sources University of Massachusetts and Infrared Processing and Analysis Center (IPAC/California Institute of Technology), 2003. Cat. II/246.
- [6] D. Froebrich, A. Scholz, C.L. Raftery. A systematic survey for infrared star clusters with $|b| < 20^\circ$ using 2MASS // *Monthly Notices of the Royal Astronomical Society*, V. 374, p. 399, 2007.
- [7] Gaia enters its operational orbit, The European Space Agency (ESA) 2014-01-08.
- [8] Gartner, Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data STAMFORD, Conn., June 27, 2011. <http://www.gartner.com/newsroom/id/1731916>
- [9] K. Tabetha Hole, Aaron M Geller, Robert D. Mathieu, Imants Platais, Søren Meibom, David W. Latham. WIYN Open Cluster Study. XXIV. Stellar Radial-Velocity Measurements in NGC 6819 // *The Astronomical Journal*, V. 138, Issue 1, p. 159–168, 2009.
- [10] N.V. Kharchenko, A.E. Piskunov, E. Schilbach, S. Roeser, R. Scholz, MWSC Milky Way global survey of star clusters. II // *Astronomy and Astrophysics*, V. 558, p. 53, 2013.
- [11] S.E. Koposov, E.V. Glushkova, I.Yu. Zolotukhin. Automated search for Galactic star clusters in large multiband surveys. I. Discovery of 15 new open clusters in the Galactic anticenter region // *Astronomy and Astrophysics*, V. 486. p. 771, 2008.
- [12] M. Kronberger, P. Teutsch, B. Alessi, M. Steine, L. Ferrero, K. Gracjewski, M. Juchert, D. Patchick, D. Riddle, J. Saloranta and 2 coauthors. New galactic open cluster candidates from DSS and 2MASS imagery // *Astronomy and Astrophysics*, V. 384, p. 403, 2006.
- [13] S. Schmeja. Identifying star clusters in a field: a comparison of different algorithms // *Astronomische Nachrichten*, V. 332, No 2, p. 172–184, 2011.
- [14] S. Schmeja, N.V. Kharchenko, A.E. Piskunov, S. Röser, E. Schilbach, D. Froebrich, and R. Scholz. Global survey of star clusters in the Milky Way III. 139 new open clusters at high Galactic latitudes // *Astronomy and Astrophysics*, V. 568, A51, 2014.
- [15] Science with 1 billion objects in three dimensions – Gaia. <http://www.cosmos.esa.int/web/gaia/>
- [16] Strasbourg astronomical Data Center. <http://cds.u-strasbg.fr/>
- [17] The Two Micron All Sky Survey at IPAC. <http://www.ipac.caltech.edu/2mass/overview/access.html>
- [18] The Hipparcos and Tycho Catalogues, ESA SP-1200 – Cat. I / 239.
- [19] USNO CCD Astrograph Catalog (UCAC). <http://www.usno.navy.mil/USNO/astrometry/optical-IR-prod/ucac>
- [20] N. Zacharias, C.T. Finch, T.M. Girard, A. Henden, J.L. Bartlett, D.G. Monet, M.I. Zacharias. The fourth U.S. Naval Observatory CCD Astrograph Catalog (UCAC4) // *The Astronomical Journal*, V. 145, p. 44, 2013. <http://cdsarc.u-strasbg.fr/viz-bin/Cat?I/322A>
- [21] Y. Zhang, Y. Zhao. Astronomy in the Big Data Era // *Data Science Journal*, V. 14, p. 11, 2015. <http://datascience.codata.org/articles/10.5334/dsj-2015-011/>
- [22] Е.В. Глушкова. Комплексное исследование рассеянных звездных скоплений Галактики // Дис. ... доктора физ.-мат. наук. Москва, ГАИШ МГУ, 2014.
- [23] Крис Канаракус. Машина Больших Данных // Журнал «Сети» № 04, 11.10.2011. <http://www.osp.ru/nets/2011/04/13010802/>

Star clusters: the growth of knowledge based on data intensive research

S.V. Vereshchagin, N.V. Chupina, A.S. Fionov

The results of information analysis on the open star clusters (OC) are presented. The growth in number of publications on OC, stellar reviews and OC catalogs has been demonstrated. It is shown that the use of filters allows to detect more than 10 thousand of unknown OC. The main intention of the paper is to show prospects of the use of methods and tools of data intensive research to analyze the structured and unstructured information on OC.