

ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ МЕЖЗВЕЗДНОГО ПОГЛОЩЕНИЯ СВЕТА ПО ДАННЫМ КАТАЛОГА HIPPARCOS

Амосов Ф.А., Витязев В.В., Цветков А.С.

Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

На основе данных космической миссии Hipparcos для 98827 близких звезд (до 500 пк) получена карта значений градиента покраснения в направлениях, определяемых центрами 3888 равновеликих площадок, построенных методом HealPix. Отсутствующие в каталоге Hipparcos данные о классе светимости 49542 звезд северного экваториального полушария были восстановлены с помощью линейного бинарного классификатора, построенного на обучающем множестве 39807 звезд спектральных классов III и V. Для III класса светимости точность классификатора равна 95%, полнота – 89%. Для V класса светимости соответствующие характеристики равны 91% и 96%.

Введение

Как известно [4], пылевые облака в Галактике приводят к общему ослаблению света звезд (экстинции) и к покраснению звезд, то есть к увеличению их показателей цветов. Распределения покраснения звезд по небесной сфере может служить свидетельством о существовании межзвездной пыли в данном направлении.

Успех выявления пылевых облаков в двумерном и трехмерном варианте определяется имеющимися в нашем распоряжении массовыми каталогами звезд, содержащих информацию о координатах и параллаксах звезд, а также спектральные и фотометрические данные – спектральный класс, класс светимости и звездные величины в различных фотометрических полосах. Одним из таких каталогов является Hipparcos [9]. К сожалению, в каталоге Hipparcos спектральные классы звезд известны в основном только для отрицательных склонений. В настоящей работе поставлена задача устранить этот недостаток путем определения классов светимости звезд северного полушария с помощью метода машинного обучения [6, 10]. Теория обучения машин (machine learning, машинное обучение) – это самостоятельная математическая дисциплина, находящаяся на стыке прикладной статистики, численных методов оптимизации и дискретного анализа. В настоящее время технология машинного обучения нашла широкое применение в различных задачах обработки данных, в том числе и в астрономии [11]. Наша работа основана на построении бинарного классификатора, результатом работы которого является назначение звезде с известной абсолютной звездной величине класса светимости.

Исходные данные

В 2007 году вышла новая редакция астрометрических данных каталога Hipparcos – каталог HIPNEWCAT (HIPparcos NEW astrometric CATalog) [7]. Утверждается, что точность положений, параллаксов и собственных движений всех звезд, ярче $HP = 8$, улучшена в 4 раза, а для всех остальных звезд более чем в 2 раза. Уменьшена взаимная корреляция параметров иногда в 10 раз. Именно эта версия использовалась в работе в качестве источника астрометрических данных.

Каталог Hipparcos для большинства звезд содержит информацию о спектральном типе, полученную из наземных наблюдений. Основной источник – Мичиганский каталог [8] и несколько других каталогов. Однако информация о спектральных классах приведена только для звезд южного экваториального полушария.

В нашей работе использовалась следующая информация:

- положения звезд (каталог Hipnewcat);

- параллаксы звезд (каталог Hipnewcat);
- фотометрия (V_{mag} – каталог Hipparcos);
- класс светимости (каталог Hipparcos).

Дополнительные сведения об используемых данных можно найти в работах [2–3].

Способ получения классов светимости

Как уже было сказано, в каталоге Hipparcos, практически у всех звезд северного экваториального полушария отсутствует класс светимости. Для нашей задачи его наличие чрезвычайно важно, ввиду того, что мы на основе класса светимости и спектрального класса рассчитываем истинное значение $B-V$, то есть $(B-V)_{\text{int}}$. Тем самым, отсутствие класса светимости у половины звезд делает невозможным проведение наших расчетов для всего северного экваториального полушария.

Для исправления этого недостатка используем метод машинного обучения. Натренируем классификатор, который будет определять класс светимости для звезды по двум факторам – ее показателю цвета $B-V$ и ее абсолютной звездной величине M_V . Этим факторам должно быть достаточно, т.к. классы светимости теоретически разделимы на диаграмме Герцшпрунга-Рессела.

В каталоге Hipparcos доля звезд, которые не относятся ни к III, ни к V классам светимости мала (16.3%, 8058 из 49285). Поэтому мы упростим задачу – обучим линейный бинарный классификатор, который будет предсказывать III или V класс. Сделаем это с помощью метода опорных векторов [6]. В качестве обучающего множества возьмем все звезды, у которых присутствует класс светимости III и V. Таких звезд 39807. Распределение этих звезд по классам светимости показано в табл. 1,

Таблица 1. Структура обучающего множества

Класс светимости	$B-V < 0.6$	$B-V > 0.6$	всего
III	1947	16681	18628
V	15549	5630	21179
III & V	17496	22311	39807

Диаграмма Герцшпрунга-Рессела, соответствующая нашему обучающему множеству, показана на рис. 1.

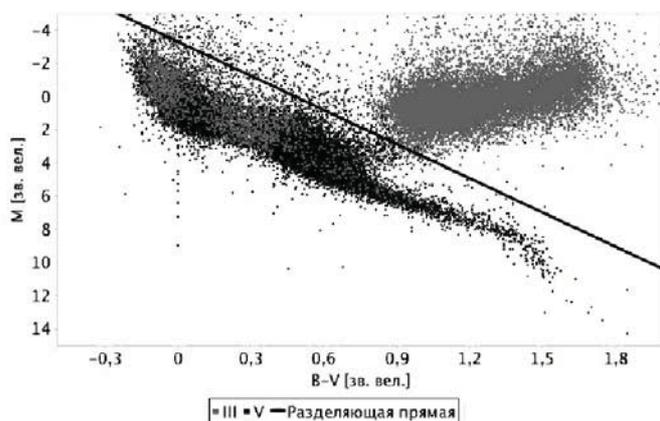


Рис. 1. Диаграмма Герцшпрунга-Рессела обучающего множества звезд (табл. 1). Серым цветом отмечены звезды III класса, темным – V. На рисунке также изображена разделяющая прямая.

Результатом обучения классификатора является разделяющая классы прямая, уравнение которой имеет следующий вид:

$$F(B-V, M) = 3.0752 \cdot (B-V) + 0.4485 \cdot M + 1.4793 = 0, \quad (1)$$

где $B-V$ – показатель цвета, M – абсолютная звездная величина.

Мы видим, что при показателе цвета $B-V > 0.6$ классификатор работает практически идеально, но при $B-V < 0.6$ и $F(B-V, M) > 0$ он всем звездам предсказывает V класс при большой доле звезд III класса в обучающем множестве (10.3%, 1758 из 17029). При $B-V < 0.6$ звезды III и V класса неразделимы, поэтому для этой половины можно принять другое решение – результат работы классификатора будет взвешенным средним III и V классов, то есть некоторым средним классом, $(B-V)_{int}$, у которого будет равен

$$(B-V)_{int} = w_1 \cdot (B-V)_{int}(III) + w_2 \cdot (B-V)_{int}(V). \quad (2)$$

В этой формуле веса w_1 и w_2 , ($w_1 + w_2 = 1$) логично взять в соответствии с априорной вероятностью классов в этой области (0.1 и 0.9). В таблице двумерной спектральной классификации [1] у спектрального типа G2 показатель цвета при III классе 0.733, при V – 0.630. Это максимальная разница между III и V классом в интересующем нас диапазоне. Класс, соответствующий средневзвешенному решению, будет иметь показатель цвета

$$(B-V)_{int} = 0.1 \cdot 0.733 + 0.9 \cdot 0.630 = 0.64, \quad (3)$$

то есть, даже в наихудшем случае отличие от V класса минимальное – 0.01^m , гораздо ниже уровня ошибок показателя цвета в каталоге Hipparcos. Поэтому в дальнейшем мы всегда будем использовать решение классификатора в этой области по V классу.

С учетом этого соглашения проведем оценку качества работы классификатора с помощью процедуры 10-fold кросс-валидации. Полученные оценки параметров классификатора (точность, полнота и F1-мера) приведены в следующей таблице:

Решение классификатора →	III	V	Класс	Точность	Полнота	F1-мера
III	16636	1992	III	95%	89%	92%
V	783	20396	V	91%	96%	93%

Как видим, классификатор имеет приемлемое качество, и с его помощью мы определили классы светимости 49542 звезд северного экваториального полушария. Этот результат позволил составить звездный каталог со следующей информацией:

- номер звезды в каталоге Hipparcos;
- галактические долготы и широты звезд, град;
- звездная величина в полосе V;
- спектральный класс;
- класс светимости.

Карты градиентов покраснения звезд

Межзвездное поглощение может быть описано избытком цвета, которое мы будем называть «покраснением»:

$$E = E_{B-V} = (B-V)_{obs} - (B-V)_{int} \quad (4)$$

где $(B-V)_{obs}$ – ее видимый показатель цвета звезды (с учетом межзвездного поглощения), а $(B-V)_{int}$ – теоретический показатель цвета звезды (без учета межзвездного поглощения). Значение $(B-V)_{obs}$ мы можем получить на основе данных фотометрии звезды из каталога. Значение $(B-V)_{int}$ берется из приведенной в [4] двумерной таблицы «спектральный класс, класс светимости – показатель цвета». То есть, для получения $(B-V)_{int}$ звезды нам нужно знать ее спектральный класс и класс светимости.

В дальнейшем мы будем обозначать конусы, высекаемые соответствующими пикселями разбиения Healpix [5], через C_i , где $i = 0, \dots, N_{pix}-1$. У нас $N_{pix} = 3888$.

Такое разбиение позволит нам:

- 1) рассмотреть ход покраснения в каждом конусе как одномерную функцию $E(r)$. Это корректно, ввиду того, что конусы достаточно узкие;
- 2) сделать наши результаты «независимыми», т.к. конусы не пересекаются;
- 3) поместить в каждый конус примерно одинаковое число звезд, чтобы избежать недостатка звезд в некоторых конусах.

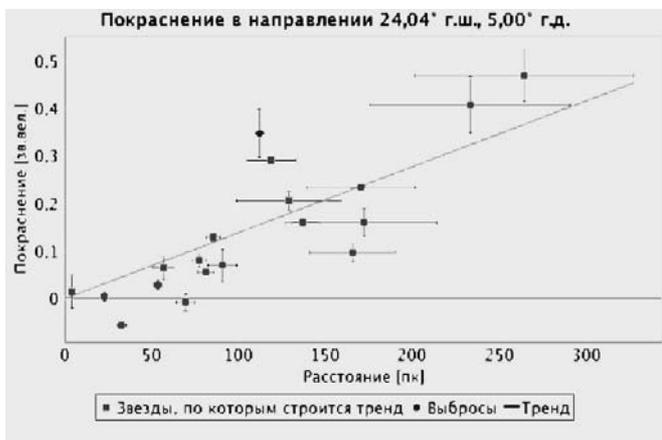


Рис. 2. Ход покраснения звезд по расстояниям в направлении $l = 5.00$, $b = 24.04$. По горизонтали расстояния в пк, по вертикали покраснение (зв. вел.). При вычислении градиента k отброшена одна точка. Данные из каталога Hipparcos.

Будем моделировать покраснение звезд вдоль луча зрения линейной функцией $E(r) = k \cdot r$. Свободный член можно считать нулевым, т.к. у ближайших звезд покраснения не должно быть ввиду очень малого объема пыли на луче зрения до них. Параметр k модели, соответствующей каждому конусу C_i , находится с помощью метода наименьших квадратов. При этом мы проводили отбор данных, так как метод наименьших квадратов очень чувствителен к выбросам, т.е. большие ошибки значений покраснения и/или параллакса могут очень сильно испортить результат. Для устранения этого недостатка после построения тренда по всем звездам в конусе, мы выбрасывали те звезды, которые давали самые большие отклонение от найденного тренда. Затем, новое значение тренда определялось уже только по оставшимся звездам. После выброса 10% самых плохих звезд процесс выбросов заканчивался (см. рис. 2).

Очевидно, что сильные выбросы дают далекие звезды, для которых параллакс определен с большой ошибкой. В нашей работе для построения градиентов покраснения использовались звезды на гелиоцентрических расстояниях до 500 пк. Кроме того, выбросы могут давать и звезды с большими ошибками в покраснении. Специальное исследование показало, что уровень ошибок покраснения, полученного по данным каталога Hipparcos, составляет 0.03^m , что является приемлемым значением. Тем не менее, некоторые звезды имеют большие ошибки в покраснении и могут быть выброшены в нашем методе вычисления тренда.

Распределение коэффициента k по небесной сфере

Тем самым, ход покраснения в конусе C_i мы описываем одним числом k_i — скоростью роста покраснения в этом конусе. Она, как мы ранее выясняли, должна коррелировать с наличием пыли. Поэтому, составив карту распределения коэффициента k , мы составим двумерную панораму распределения пыли в окрестности Солнца с радиусом до 500 пк.

На рис. 3 показана карта распределения градиентов покраснения для звезд обучающего множества. Черным цветом закрашены те площадки, для которых звезды обучающего множества не позволяют вычислить покраснение, так как для них не известны классы светимости. Число таких площадок равно 1678.

Как было сказано выше, определение спектрального класса для остальных звезд, не входящих в обучающее множество, нами производилось с помощью двумерного линейного классификатора. Карта градиентов покраснения, построенная с помощью классификатора, показана на рис. 3 справа. Сравнение левой и правой карт на рис. 3 показывает, что примененный нами метод машинного обучения для сортировки звезд по III и V классам светимости позволил получить информацию о межзвездном покраснении практически для всей небесной сферы.

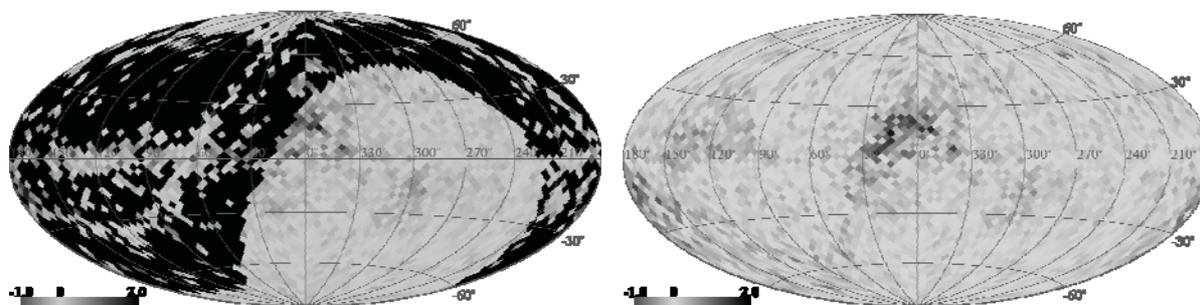


Рис. 3. Слева: распределение градиента покраснения, построенного по звездам, для которых можно определить покраснение только по данным каталога Hipparcos. Черным цветом отмечены 1678 пикселей, в которых нет даже 5-ти звезд. Справа: распределение градиента покраснения, построенного по звездам, для которых можно определить значение покраснения (94199 звезд). Границы цветовой шкалы указаны в зв. вел. на кпк.

Заключение

Перечислим основные результаты, полученные в работе.

- Создан бинарный классификатор, позволяющий звезде с известной абсолютной звездной величиной назначить класс светимости III или V. Для III класса светимости точность классификатора равна 95%, полнота – 89%. Для V класса светимости соответствующие характеристики равны 91% и 96%.
- Для 49542 звезд каталога HIPPARCOS с помощью бинарного классификатора определены классы светимости. Электронная версия этого каталога доступна по адресу <http://github.com/amosov-f/dust-detector/tree/R-1.0/dust/src/test/resources/predict-lumin.txt>
- Для 94199 звезд каталога HIPPARCOS определены покраснения звезд по показателю цвета B-V.
- Для звезд с расстояниями до 500 пк получена карта значений градиента покраснения в направлениях, определяемых центрами 3888 равновеликих площадок, построенных методом HealPix. Электронная версия этого каталога доступна по адресу <http://github.com/amosov-f/dust-detector/tree/R-1.0/dust/src/test/resources/table-k.txt>
- Определена статистическая надежность результатов для каждой площадки.
- Произведено сравнение карт покраснения, полученных нами в оптическом диапазоне, с аналогичными результатами, полученными в ближней инфракрасной области по данным каталога 2MASS. Практическое совпадение этих карт свидетельствует о надежной работе использованного нами бинарного классификатора.

Работа выполнена при поддержке гранта СПбГУ 6.37.343.2015.

Литература

1. В. Страйжис. Многоцветная фотометрия звезд. – Вильнюс, 1977.
2. Цветков А.С. Руководство по практической работе с каталогом Hipparcos. – СПб: СПбГУ, (2005).
3. Цветков А.С. Руководство по работе с каталогом ТУСНО-2. – СПб: СПбГУ, (2006).

4. *J. Binney and M. Merrifield. Galactic Astronomy. – Princeton: Princeton Univ. Press, 1998.*
5. *Gorski et al. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere // Astrophys. J. 622, 759 (2005).*
6. *Hastie, T.; Tibshirani, R. & Friedman, J. The Elements of Statistical Learning. – Springer New York Inc., New York, NY, USA (2001).*
7. *Hipparcos, the New Reduction of the Raw Data van Leeuwen F. // Astron. Astrophys. 474, 653 (2007).*
8. *Houk N. et al. Michigan Catalog of Two-dimensional Spectral Types for HD Stars. Vol. 1–5. Ann. Arbor: Univ. Michigan Dept. Astron. 1978, 1982, 1975, 1988, 1999.*
9. *Perryman M.A.C., Lindegren L., Kovalevsky J., Hog E., Bastian U., Bernacca P.L., Creze M., Donati F., Grenon M., Grewing M., van Leeuwen F., van der Marel H., Mignard F., Murray C.A., Le Poole R.S., Schrijver H., Turon C., Arenou F., Froeschle M., Petersen C.S. The Hipparcos Catalogue // A&A. 1997. 323L. 49 P.*
10. *V.N. Vapnik. The Nature of Statistical Learning Theory. – Berlin: Springer Verlag, 1995.*
11. *Zeljko Ivezić, Andrew J. Connolly, Statistics, Data Mining, and Machine Learning in Astronomy. – Princeton University Press. 533 p., 2014.*

INTERSTELLAR EXTINCTION PARAMETERS DERIVED FROM THE HIPPARCOS CATALOGUE

Amosov F., Vityazev V., Tsvetkov A.
Saint-Petersburg State University, St. Petersburg, Russia

A map of extinction gradients based on data from the Hipparcos catalogue in the directions, defined by the centers of 3888 equal size HealPix areas is presented. Missing data on luminosity classes of 49542 stars in northern equatorial hemisphere have been restored by linear binary classifier built on the training set of 39807 stars of the III and V spectral classes. The classifier precisions and recalls are 95% and 89% for the luminosity class III and 91% and 96% for the V class luminosity, respectively.