

Технологии информационного общества**ПОВЫШЕНИЕ КАЧЕСТВА КАТАЛОГОВ КРАТНЫХ ЗВЕЗД С
ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА****Сажин Михаил Васильевич**

*Доктор физико-математических наук, профессор
МГУ имени М.В.Ломоносова, Государственный астрономический институт имени П.К. Штернберга
(ГАИШ), главный научный сотрудник
Москва, Российская Федерация
sazhinm@gmail.com*

Семенов Валериан Никитич

*Кандидат физико-математических наук
МГУ имени М.В.Ломоносова, Государственный астрономический институт имени П.К. Штернберга
(ГАИШ), старший научный сотрудник
Москва, Российская Федерация
valeras24@gmail.com*

Сорокин Сергей Владимирович

*Кандидат физико-математических наук, доцент
Тверской государственный университет, старший научный сотрудник
sergey@tversu.ru*

Райков Александр Николаевич

*Доктор технических наук, профессор
Институт проблем управления имени В.А. Трапезникова РАН, ведущий научный сотрудник
МГУ имени М.В. Ломоносова, Национальный центр цифровой экономики, руководитель департамента
интеллектуальных технологий
Научно-аналитический журнал «Информационное общество», член Редакционного совета
Москва, Российская Федерация
alexander.n.raikov@gmail.com*

Аннотация

Разработан и предлагается метод выявления оптических двойных звезд на основе использования астрометрических каталогов в сочетании с методами искусственного интеллекта (ИИ). Исследование проведено на примере каталога миссии HIPPARCOS и каталога Pan-STARRS (PS1) на массиве порядка 100 тыс. объектов, имеющих около 80 полей данных. При этом из анализа были исключены такие поля, которые включали ссылки на другие каталоги и источники данных. С применением методов ИИ, а именно, двух видов моделей, ансамбля полносвязных нейронных сетей и ансамбля деревьев решений, на примере указанных каталогов проведен вычислительный эксперимент. При обучении оптимизировалась метрика бинарной кросс-энтропии. Показано, что надежность предсказания двойственности звезд достигает 90-95%, что помогает обнаружить дополнительные двойные звезды по сравнению с классическими методами. Отмечено, что алгоритмы машинного обучения достаточно устойчиво выделяют группу значимых признаков, связанных со статистическими характеристиками наблюдаемых величин. Таким образом, обоснована плодотворность создания соответствующей платформы ИИ для проведения дальнейших исследований.

Ключевые слова

большие данные; деревья решений, искусственный интеллект, каталог двойных звезд, кратные звезды, машинное обучение, качество каталогов звезд, нейронная сеть

© Сажин М.В., Семенов В.Н., Сорокин С.В., Райков А.Н., 2022.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>
https://doi.org/10.52605/16059921_2022_05_106

Введение

Целью работы является повышение точности астрономических измерений за счет применения методов машинного обучения (из сферы методов искусственного интеллекта) на основе деревьев решений и полносвязной нейронной сети. Исследование проведено на примере оценки соотношения количества одиночных и кратных (на примере двойных) звезд.

Двойные звезды образуют динамические системы, которые обращаются под действием силы взаимного притяжения вокруг общего центра масс. Измерительные подходы классифицируют представление таких звезд как визуально-двойные, спектрально-двойные и затменные переменные [1]. В соответствии с этой классификацией методы обнаружения подразделяются на астрометрические, спектральные и фотометрические.

Современное развитие классических методов астрономических измерений сопровождается применением все более сложных моделей, которые могут включать более десяти свободных параметров. Как следствие, доля обнаруженных звезд, демонстрирующих нелинейные движения, постоянно возрастает и, соответственно, растет риск ошибочной оценки соотношения двойных и одиночных звезд. Помимо этого, классические методы обработки астрометрических измерений сталкиваются с проблемой экспоненциально роста объема вычислений от числа параметров.

В данной работе предложено дополнить классические методы оценки упомянутого соотношения методами машинного обучения. Они позволяют работать с множеством плохо определенных параметров и уменьшают объем вычислений. Обращение к методам искусственного интеллекта (ИИ), имеет высокий потенциал, который постоянно наращивается, прежде всего, за счет развития самих методов ИИ, накопления и аналитической обработки больших данных, роста вычислительной мощности компьютеров, развития средств семантической интерпретации, включая построения когнитивных (плохо формализуемых) семантик [2].

В качестве примеров успешного применения ИИ в области астрономии можно привести работу [3], где представлен классификатор астрономических событий, или статью [4], где рассматриваются фотометрические и спектроскопические наблюдения быстро изменяющихся источников, которые образуются после взрыва астрономического объекта. Для классификации сигналов от ядер галактик, сверхновых звезд, астероидов и т.д. используется сверточная нейронная сеть. При этом, к астрономическим изображениям добавляются метаданные, что помогает достичь высокого уровня точности ($\approx 94\%$).

Существует два типа методов классификации – на основании шаблона и кривой блеска. Первый способен различать более богатую таксономию событий, второй использует только первое предупреждение о событии. Современные астрометрические инструменты способны оценить уровень хаоса, вызванного взрывом объектов, оценить размер звезды-компаньона, распознавать, аннотировать и классифицировать большие данные, полученные с обзорных телескопов.

Настоящее исследование производится в условиях крайне высокой эмпирической оценки вероятности двойственности объектов, неполноты каталогов и отсутствии общепризнанной астрофизической теории. Статья сначала делает обзор классических работ по обнаружению двойных звезд и описывает каталог HIPPARCOS. Далее описываются результаты вычислительного эксперимента с применением методов машинного обучения и ИИ. При этом выделяются значимые и контрольные признаки двойственности, производится отбор признаков для проведения классификации одиночных и двойных звезд. Осуществляется построение классификации и проверка ее устойчивости к изменению наблюдательной селекции обучающей выборки. В заключении дается оценка доли двойных звезд по сравнению с одиночными звездами в каталоге.

1 Наземные наблюдения и каталогизация двойных звезд

Идея о существовании во Вселенной физических двойных и кратных звездных систем была впервые высказана Джоном Митчеллом, который применил методы статистики к изучению звезд, и продемонстрировал, что гораздо больше звезд встречается парами или группами, чем можно объяснить случайными совпадениями [5]. Он оценил вероятность появления тесной группы и пришел к выводу, что звезды в таких звездных системах могут притягиваться друг к другу. Его работа над двойными звездами повлияла на исследования Уильяма Гершеля по той же теме, оформившиеся в первый каталог двойных звезд [6].

Предложенный изначально в метод выделения в качестве двойных звезд объектов с малым угловым расстоянием (оптические двойные, очень близкие друг к другу на небе звезды) в дальнейшем было дополнено другими методами, прежде всего, астрометрическим. Многолетние наблюдения оптических и астрометрических двойных звезд позволяют в некоторых случаях определить орбиты их компонент. В настоящее время речь идет о сотнях таких объектов. Это единственный прямой метод определения скоплений звезд.

Астрометрическими методами можно изучать достаточно широкие звездные пары. Для более близких компонент эффективнее оказываются фотометрические (наблюдение взаимных затмений или переменности, связанной с отклонением формы компонент от сферической под влиянием взаимной гравитации) или спектральные методы (наблюдение перемещения спектральных линий в составном спектре звезды под влиянием орбитального движения и, соответственно, проявления эффекта Доплера).

В большинстве случаев двойственность объекта выявляется с использованием критериев типа критерия Рэлея: значимое раздвоение спектральной линии будет выявлено, если линии разойдутся на расстояние, сопоставимое с эффективной шириной спектральной линии, затменный тип переменности будет классифицирован по кривой блеска, если глубина затмения превысит фотометрическую ошибку настолько, что удастся пронаблюдать характерную форму. Аналогично выделялись раньше и оптические двойные звезды: если расстояние между компонентами превысит полуширину функции рассеяния точки (Point Spread Function, PSF). Существенно, что уже более ста лет точность определения астрометрических параметров звезд значительно лучше, чем PSF. С этим связано развитие астрометрических методов обнаружения двойных звезд по особенностям собственных движений. Настоящая работа также направлена на выяснение критериев обнаружения двойственности астрометрических двойных звезд, который был бы на порядки чувствительнее критерия Рэлея.

Специализированные каталоги двойных звезд публикуются с конца XVIII века, например каталог Herschel датирован 1785 г., а каталог CCDM – 2002 г. Результатам этих работ посвящено множество публикаций. По ним можно заметить, что рост числа открытых систем происходит достаточно умеренно: в начале XX века астрономы пользовались визуальными каталогами, содержащими в сумме около миллиона звезд, в середине века стала доступна фотографическая Карта неба с примерно 4.5 миллионами звезд, развитие космической астрономии потребовало каталогов в десятки миллионов объектов, а в XXI веке в научный оборот были введены электронные версии фотографических каталогов объемом около миллиарда звезд. Объем же каталогов двойных звезд за это время вырос всего примерно в 10 раз.

Приведенная картина характерна для обзорных каталогов звезд, ограниченных самой слабой наблюдаемой звездной величиной. Большинство звезд в таком каталоге [7] незначительно ярче предела обнаружения и выявление двойственности в таком случае проблематично.

2 Модели кратных звезд и каталог HIPPARCOS

В исследованиях режимов коллапса протозвездных облаков общего устоявшегося мнения пока нет. Теоретические модели зарождения одиночных звезд считаются довольно хорошо и получается достаточно убедительная начальная функция масс, которая в популяционных расчетах впоследствии дает разумные результаты. Так, согласно [8] для разных масс звезд построены соответствующие функциональные зависимости.

В теоретических моделях коллапса протозвездных облаков с начальным моментом вращения, когда сжатие газа сначала должно приводить к появлению тороидальной структуры, которая затем уже распадается на отдельных протозвезд, формируя кратную звездную систему (в простейшем случае – двойную) общепризнанной устоявшейся теории пока нет. Учет разных физических механизмов (например, степени влияния магнитного поля) приводит к существенно различающимся результатам, ни один из которых не подтверждается полностью статистикой. Статистика же ближайших к Солнцу звезд говорит о частом проявлении признаков двойственности (около 50%, см. выше).

Высокоточные оптические измерения требуют длиннофокусного инструмента, обладающего, следовательно, небольшим полем зрения. Космический аппарат (КА) HIPPARCOS стал первым специализированным астрометрическим научным спутником. Задача проведения высокоточных измерений по всей небесной сфере определила схему прибора с двумя рабочими

полями диаметром примерно 0.9° , отстоящими одно от другого на $\approx 58^\circ$. КА вращался с периодом 120 минут вокруг оси, перпендикулярной плоскости, в которой лежат входные зрачки [9], а сама ось (по плану наблюдений) медленно прецессировала по конусу 43° вокруг направления на Солнце. Все эти движения в совокупности приводили к достаточно равномерному покрытию наблюдениями небесной сферы и, с другой стороны, не позволяли солнечным батареям спутника сильно отклониться от направления на Солнце и потерять мощность.

Для повышения стабильности измерений производилась регистрация не просто прохождения звезды в поле зрения электровакуумного прибора, а ее проход по решетке из 2660 щелей. Регистрируемые координаты таким образом оказывались одномерными, они привязывались к большому кругу, одному и тому же для нескольких оборотов спутника. В процессе обработки параметры кругов увязывались друг с другом для всей небесной сферы, затем вычислялись сферические координаты отдельных звезд, их параллаксы и собственные движения.

Результаты работы спутника HIPPARCOS показали точность измерения координат лучше наземных наблюдений примерно в 100 раз. Это дало возможность одновременно с координатами за время эксперимента определить параллаксы и собственные движения с сопоставимой точностью – около 1 мсек дуги. То есть эксперимент длительностью в 3.5 года дал результат, сравнимый, а местами (в части высокоточных параллаксов) превосходящий результаты астрометрической деятельности в течение столетия.

Очевидно, что любые мыслимые критерии обнаружения двойственности объекта по измерениям какого-либо одного параметра: эллипсоидальность видимого изображения, отклонение собственного движения от прямой линии, аномальная фотометрия, раздвоение спектральных линий – уже были, как правило, использованы авторами наблюдений или их ближайшими последователями и ничего нового дать не могут. Общий принцип обнаружения двойственности в таком случае примерно один и тот же: критерий Рэлея в последнем случае или значимая по сравнению с функцией PSF эллиптичность изображения в первом случае – аналогичны друг другу и не дают возможности обнаружить двойственность, если разделение компонент меньше погрешностей, ширины спектральной линии или той же PSF.

С другой стороны, очевидно, что оптическое разделение компонент предполагаемой двойной звезды, большие погрешностей координатных измерений, должно как-то проявиться в результатах наблюдений. Настоящая работа состоит в проверке с применением средств ИИ факта, что для неразрешенных, как и для разрешенных, двойных звезд будут появляться аномалии в погрешностях измеряемых величин.

3 Применение ИИ для повышения качества выявления двойственных звезд

В качестве систем ИИ применялись инструменты нейронных сетей и деревьев решений. Каталог HIPPARCOS содержит множество характеристик звезд (всего в каталоге 77 полей данных). Из них в настоящей работе были исключены из анализа поля такие, которые включали ссылки на другие каталоги и источники данных. Из рассмотрения также были исключены значения небесных координат, с одной стороны, потому что спутник исследовал малый околосолнечный объем в Галактике, где ее структура еще не очень заметна. С другой стороны, выбранный при составлении каталога способ сканирования неба космическим аппаратом делает выделенными областями полюса эклиптики, где число измерений меньше, а статистические погрешности выше, чем в плоскости эклиптики. Следовательно, сами погрешности измерений в каталоге неминуемо будут зависеть от координат, и это только усложняет реализацию алгоритма ИИ.

Поле спектрального класса звезды представлено в каталоге в текстовом виде, что делает невозможным его прямой анализ алгоритмами ИИ. Для обработки этого поля текстовая строка спектра была разделена на построенный вручную вектор признаков, описывающих спектр, которые уже могут быть использованы как входы системы ИИ. Таким образом, данные каталога были преобразованы в пространство 73 признаков, включающее 34 числовых значения взятых непосредственно из каталога, а также 39 признаков, выделенных из спектрального класса полей.

Для обработки данных использовались два вида моделей: ансамбль нейронных сетей и ансамбль деревьев решений, обученных по алгоритму XGBoost (XGB, eXtreme Gradient Boosting, оптимизированная распределенная библиотека ускорения градиентного поиска с использованием алгоритмов машинного обучения [10]).

В качестве нейронных сетей использовались полносвязные нейронные сети, включающие два скрытых слоя с функцией активации ReLU (Rectified Linear Unit, функция рампы) из 200 и 100 нейронов, соответственно, и выходной слой из одного нейрона с сигмоидной функцией активации. Перед подачей данных на входной слой нейронной сети они пропускались через слой нормализации. Для обучения использовался алгоритм Adam (сокращение от Adaptive Moment Estimation, – метод адаптивной оценки моментов), оптимизировалась метрика бинарной кросс-энтропии. Нейронная сеть была реализована с помощью библиотеки Keras, входящей в пакет Tensorflow.

Для создания ансамбля нейронных сетей все записи каталога HIPPARCOS были разделены на 25 выборок случайным образом с сохранением в каждой из них одинаковой пропорции целевого признака двойственности. Далее, по очереди, каждая из выборок применялась для обучения одной нейронной сети с использованием всех оставшихся 24 в качестве контрольной выборки для обучения. Полученные 25 нейронных сетей были объединены в ансамбль, выходное значение которого было рассчитано для всех данных каталога. Аналогичным образом был обучен ансамбль из деревьев решений.

Так как есть основания полагать, что в каталоге HIPPARCOS отмечены не все кратные системы, которые таковыми на самом деле являются, в настоящей работе предложено использовать для выявления кандидатов в двойные системы подход, который часто используется в анализе данных для выявления ошибок разметки: обучить классификатор на исходном наборе данных и затем рассматривать как кандидаты на неправильную разметку объекты с максимальной ошибкой [11].

Библиотека, реализующая алгоритм XGB, позволила получить информацию о важности признаков для обученного классификатора. Рис.1 показывает, что алгоритм предложил в качестве значимых параметров разнообразные статистические характеристики в каталоге, а малозначимыми посчитал, например, характеристики спектральные. То есть, алгоритмы ИИ подтвердили известный по каталогам двойных звезд результат, что идентификация двойственности слабо зависит от спектральных характеристик пары.

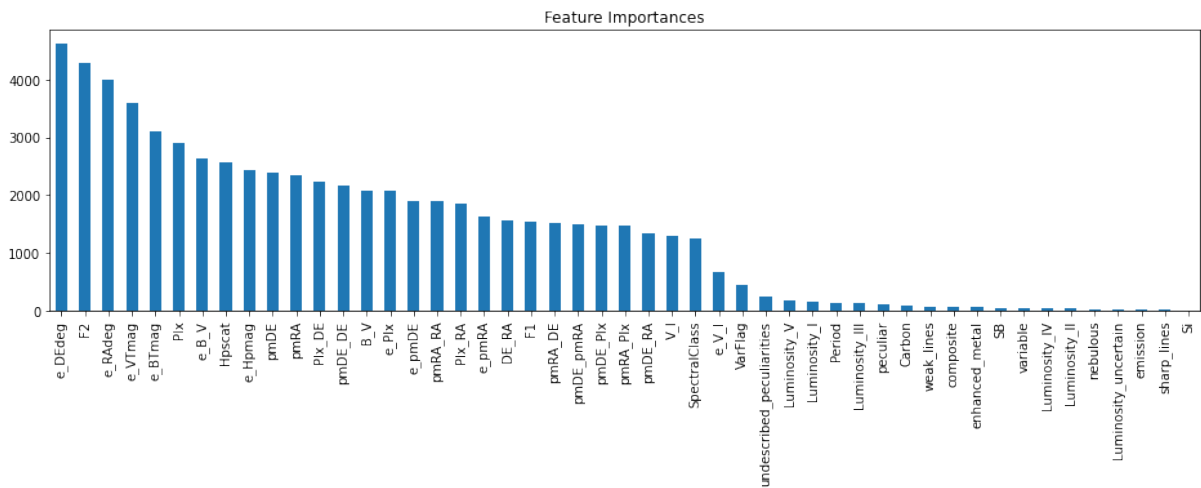


Рисунок 1 – Поля каталога HIPPARCOS, ранжированные по степени значимости

На Рис. 2 в качестве критерия важности по оси абсцисс приведено значение числа использований переменной в вершинах деревьев, усреднённого по всем членам ансамбля. Проведённые вычислительные эксперименты показали, что значения этого показателя согласованы между членами ансамбля и устойчивы к изменениям на случайных этапах обучения.

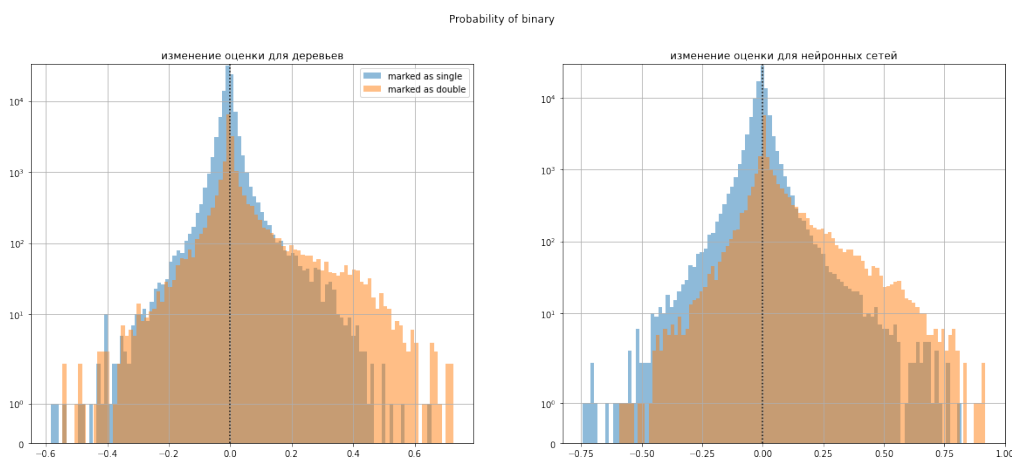


Рисунок 2 – Изменение вероятностей классификации звезд двойными и одиночными для разных алгоритмов ИИ

Обучающей выборкой для алгоритмов ИИ служит подмножество звезд каталога [12], обладающих свойством признанных двойными. Сама процедура установления двойственности по признаку аномальной близости соседних звезд подвержена сильной наблюдательной селекции: двойственность по каталогам для ярких звезд (которых немного) оказывается значительно вероятнее, чем для более многочисленных слабых звезд. С целью проверки влияния данного эффекта на работу алгоритмов ИИ, была проведена независимая тренировка, с исключением фотометрических величин.

Результаты работы консорциума HIPPARCOS, опубликованные в [12], включали решение системы нелинейных уравнений редукции для определения кинематических параметров звезд. Система уравнений решалась методом итераций.

Проведенный эксперимент с машинным обучением показал устойчивость значений основных статистических характеристик в каталоге для процесса классификации объектов как двойных звезд (рис.2).

Так как результатом работы предложенных моделей является вероятность двойственности для каждой из звезд каталога HIPPARCOS, в качестве кандидатов в двойные звезды были рассмотрены такие звезды, для которых эта вероятность превысит некоторый порог. Показано, какое число новых кандидатов в двойные системы может быть выделено с использованием предложенных моделей для разных значений порога.

4 Сопоставление и верификация результатов работы ИИ

Сравнение было проведено по самой объемной работе по выявлению дополнительных двойных звезд в каталоге HIPPARCOS [12, 13], где применялся астрометрический метод, с использованием дополнительной информации из каталога Tycho-2. Показано, что большинство кратных звезд, выделенных в этой статье, получили при оценке их с помощью системы ИИ низкие вероятности двойственности. С другой стороны, показано, что большинство из звезд, предлагаемых в кандидаты на двойственность моделями ИИ, не были выявлены. В связи с этим интерес может представлять включение выявленных в этой статье двойных звезд в обучающую выборку, обучение и исследование полученных моделей.

Для верификации полученных результатов применения ИИ к объектам каталога HIPPARCOS полученные списки наиболее вероятных двойных звезд были сопоставлены с двойными звездами из каталога объектов Pan-STARRS (телескоп $D = 1.8$ м установлен на Гавайском острове Мауи, на вершине вулкана Халеакала, фотоприемник – мозаика CCD, 1.4 млрд пикселей) [14]. Поиск соседей производился в окрестности звезды радиусом $5''$ (Это размер рабочей области фотоприемника HIPPARCOS, производившего измерения). Показано, что точность предсказания двойственности предложенным в настоящей статье методом – 90-95%.

Заключение

Представленный в статье методологический подход и проведенные вычислительные эксперименты показывают, что использование методов ИИ позволяет извлечь дополнительную информацию и выявить кратные (на примере двойных) звёздные системы, которые не удавалось ранее выявлять классическими методами. Это является результатом вскрытия методами ИИ сложных зависимостей между астрометрическими и фотометрическими характеристиками с оценкой погрешностей этих характеристик. Классические методы основаны на анализе изолированных характеристик или небольших групп характеристик, ограничены точностью их измерения.

Ограничением предложенного метода является то, что он не позволяет ввести строгого критерия двойственности звёзд. В классических методах критерии такого рода формулируются на основе известных физических закономерностей перед началом анализа. Однако, теоретически сформулировать подобного рода критерий, который бы описывал взаимоотношения между несколькими десятками характеристик, не представляется возможным.

Непосредственным следствием такой ситуации является то, что выходные значения предложенных в настоящей работе моделей анализа данных с применением машинного обучения нельзя рассматривать как вероятности двойственности звезд. Так как доля двойных звезд в обучающей выборке занижена, то следует ожидать, что выходные значения также будут занижены по сравнению с истинными вероятностями.

Важное значение имеет выявление средствами ИИ группы значимых признаков двойственности звезд. Показано, что идентификация двойственности слабо зависит от спектральных характеристик пары звезд.

Методы ИИ оказались устойчивыми к наблюдательной селекции в самой обучающей выборке. Кроме того, параметры объектов выходного каталога HIPPARCOS, которые к кратности/двойственности звезд, согласно исследованиям других авторов, отношения не имеют, показали низкую значимость, что явилось дополнительным доказательством результативности подхода с ИИ.

Предложенный в настоящей работе подход с использованием методов ИИ позволяет повысить качество идентификации двойных звезд, а также помогает получать новые знания и стимулировать генерацию новых идей, по сравнению с классическим подходом. Предложенный для проведения экспериментов подход с машинным обучением, как комплексный инструмент, может быть применен в дальнейшем также для исследования иных космических объектов и явлений.

Благодарности

Результаты работы доложены на XLIX Общественно-научных чтениях, посвящённых памяти Ю.А.Гагарина, 9-12 марта 2022 года.

Работа выполнена при частичной поддержке Российского научного фонда, грант № 21-18-00184 «Социогуманитарные основания критериев оценки инноваций, использующих цифровые технологии и искусственный интеллект».

Литература

1. Hilditch R.W. An Introduction to Close Binary Stars. Cambridge, Cambridge University Press, 2001.
2. Raikov A. Cognitive Semantics of Artificial Intelligence: A New Perspective. Springer Singapore, Springer Briefs in Applied Sciences and Technology edition, 2021. <https://doi.org/10.1007/978-981-33-6750-0>
3. Becker I., Pichara K., Catelan M., et al. Scalable end-to-end recurrent neural network for variable star classification. Monthly Notices of the Royal Astronomical Society, vol. 493 (2), 2020, pp. 2981-2995. <https://doi.org/10.1093/mnras/staa350>
4. Carrasco-Davis R., Reyes E., et al. Alert Classification for the ALERCE Broker System: The Real-time Stamp Classifier. The Astronomical Journal, 2021, 162:231 (27 p) <https://doi.org/10.3847/1538-3881/ac0ef1>

5. Michell J. An Inquiry into the Probable Parallax, and Magnitude of the Fixed Stars, from the Quantity of Light Which They Afford us, and the Particular Circumstances of Their Situation, by the Rev. John Michell, B. D. F. R. S. Philosophical Transactions of the Royal Society of London, Series I, 57:234–264, January 1767.
6. Herschel W. Catalogue of double stars. by William Herschel, esq. f. r. s. Philosophical Transactions of the Royal Society of London, 75:40–126, 1785.
7. Kharchenko N.V., Roeser S. All-sky compiled catalogue of 2.5 million stars. https://www.researchgate.net/publication/234499169_All-sky_Compiled_Catalogue_of_25_million_stars_Kharchenko_2009 (Обращение 29.05.2022)
8. Duquennoy A. and Mayor M. Multiplicity among solar-type stars in the solar neighborhood. II - Distribution of the orbital elements in an unbiased sample. *A&A*, 500:337–376, 1991.
9. The Hipparcos mission. Pre-launch status. Volume I: The Hipparcos satellite., volume 1, June 1989
10. Chen T. and Guestrin C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'16, pp. 785–794, New York, NY, USA, 2016. Association for Computing Machinery.
11. Zhu X., Wu X., and Chen Q. Eliminating class noise in large datasets. In Tom Fawcett and Nina Mishra, editors, Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), 2003, Washington, DC, USA, pp. 920–927. AAAI Press, 2003.
12. ESA Special Publication. The HIPPARCOS and TYCHO catalogues. Astrometric and photometric star catalogues derived from the ESA HIPPARCOS Space Astrometry Mission, vol. 1200 of ESA Special Publication, 1997.
13. Makarov V. V. and Kaplan G. H. Statistical Constraints for Astrometric Binaries with Nonlinear Motion. *The Astronomical Journal*, vol 129, No 5, pp. 2420–2427, 2005.
14. Chambers K. C., et al. VizieR Online Data Catalog: The Pan-STARRS release 1 (PS1) Survey - DR1 (Chambers+, 2016). VizieR Online Data Catalog, p. II/349, 2017.

IMPROVING QUALITY OF MULTIPLE STAR CATALOGS WITH USING ARTIFICIAL INTELLIGENCES

Sazhin Mikhail Vasilievich

*Doctor of phys.-math. sciences, professor
Lomonosov Moscow State University, Sternberg Astronomical Institute, chief researcher
Moscow, Russia Federation
sazhinm@gmail.com*

Sementsov Valerian Nikitich

*Candidate of phys.-math. sciences
Lomonosov Moscow State University, Sternberg Astronomical Institute, senior researcher
Moscow, Russia Federation
valeras24@gmail.com*

Sorokin Sergey Vladimirovich

*Candidate of phys.-math. sciences, associate professor
Tver State University, senior researcher
Tver, Russian Federation
sergey@tversu.ru*

Raikov Alexander Nikolaevich

*Doctor of engineering sciences, professor
Lomonosov Moscow State University, National Center for Digital Economy, head of department of intellectual technologies
V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, leading researcher
Research and analytical journal "Information Society", member of the Editorial board
Moscow, Russia Federation
Alexander.N.Raikov@gmail.com*

Abstract

The article has developed and proposed a method for detecting optical binary stars based with using of astrometric catalogs in combination with artificial intelligence (AI) methods. The study was carried out on the example of the HIPPARCOS mission catalog and the Pan-STARRS (PS1) catalog on an array of about 100 thousand objects with about 80 data fields. At the same time, fields that included links to other catalogs and data sources were excluded from the analysis. With the use of AI methods, namely, two types of models, an ensemble of fully connected neural networks and an ensemble of decision trees, a computational experiment was carried out using the example of these catalogs. During training, the binary cross-entropy metric was optimized. It is shown that the reliability of stellar binary prediction reaches 90-95%, which helps to detect additional binary stars compared to classical methods. It is noted that machine learning algorithms quite steadily identify a group of significant features associated with the statistical characteristics of the observed values. Thus, the fruitfulness of creating an appropriate AI platform for further research is justified.

Keywords

big data, decision trees, artificial intelligence, double star catalog, multiple stars, machine learning, quality of star catalogs, neural network

References

1. Hilditch R.W. An Introduction to Close Binary Stars. Cambridge, Cambridge University Press, 2001.
2. Raikov A. Cognitive Semantics of Artificial Intelligence: A New Perspective. Springer Singapore, Springer Briefs in Applied Sciences and Technology edition, 2021. <https://doi.org/10.1007/978-981-33-6750-0>
3. Becker I., Pichara K., Catelan M., et al. Scalable end-to-end recurrent neural network for variable star classification. Monthly Notices of the Royal Astronomical Society, vol. 493 (2), 2020, pp. 2981-2995. <https://doi.org/10.1093/mnras/staa350>

4. Carrasco-Davis R., Reyers E., et al. Alert Classification for the ALeRCE Broker System: The Real-time Stamp Classifier. *The Astronomical Journal*, 2021, 162:231 (27 p)
<https://doi.org/10.3847/1538-3881/ac0ef1>
5. Michell J. An Inquiry into the Probable Parallax, and Magnitude of the Fixed Stars, from the Quantity of Light Which They Afford us, and the Particular Circumstances of Their Situation, by the Rev. John Michell, B. D. F. R. S. *Philosophical Transactions of the Royal Society of London, Series I*, 57:234–264, January 1767.
6. Herschel W. Catalogue of double stars. by William Herschel, esq. f. r. s. *Philosophical Transactions of the Royal Society of London*, 75:40–126, 1785.
7. Kharchenko N.V., Roeser S. All-sky compiled catalogue of 2.5 million stars.
https://www.researchgate.net/publication/234499169_All-sky_Compiled_Catalogue_of_25_million_stars_Kharchenko_2009 (Обращение 29.05.2022)
8. Duquennoy A. and Mayor M. Multiplicity among solar-type stars in the solar neighborhood. II - Distribution of the orbital elements in an unbiased sample. *A&A*, 500:337–376, 1991.
9. The Hipparcos mission. Pre-launch status. Volume I: The Hipparcos satellite., volume 1, June 1989
10. Chen T. and Guestrin C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'16*, pp. 785–794, New York, NY, USA, 2016. Association for Computing Machinery.
11. Zhu X., Wu X., and Chen Q. Eliminating class noise in large datasets. In Tom Fawcett and Nina Mishra, editors, *Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003)*, 2003, Washington, DC, USA, pp. 920–927. AAAI Press, 2003.
12. ESA Special Publication. The HIPPARCOS and TYCHO catalogues. Astrometric and photometric star catalogues derived from the ESA HIPPARCOS Space Astrometry Mission, vol. 1200 of ESA Special Publication, 1997.
13. Makarov V. V. and Kaplan G. H. Statistical Constraints for Astrometric Binaries with Nonlinear Motion. *The Astronomical Journal*, vol 129, No 5, pp. 2420–2427, 2005.
14. Chambers K. C., et al. VizieR Online Data Catalog: The Pan-STARRS release 1 (PS1) Survey - DR1 (Chambers+, 2016). *VizieR Online Data Catalog*, p. II/349, 2017.