

Методы машинного обучения для анализа переменных звёзд

Р.Н. Карачурин

НИЯУ МИФИ, Россия
RNKarachurin@mephi.ru

12 февраля 2025 г.

1. Введение

Изучение переменных звёзд остаётся одной из актуальных задач современной астрономии, поскольку изменения их яркости отражают фундаментальные процессы, происходящие в звёздах и двойных системах. Такие объекты, как цефеиды и RR Лиры, играют ключевую роль в определении космических расстояний, а временные ряды наблюдений (кривые блеска) дают возможность более глубоко изучить динамику этих процессов Cox, 1980; Goodfellow и др., 2016.

В последние годы благодаря масштабным астрономическим обзорам, таким как Gaia, OGLE, CRTS и другим, стало возможным собрать огромные объёмы данных, требующих автоматизированных методов анализа. Современные алгоритмы машинного обучения, в особенности нейронные сети, открывают новые перспективы для автоматической классификации переменных звёзд, позволяя объединить статические спектральные признаки и динамические характеристики временных рядов Nun и др., 2020.

Цель данного исследования состоит в разработке комплексной модели, способной шаг за шагом улучшать качество классификации: от базовой бинарной модели до сложного мультиклассового классификатора с интеграцией информации о временной изменчивости объектов.

2. Постановка задачи

Основные задачи, решаемые в рамках проекта, включают:

- Бинарная классификация:** Построить модель, способную отличать переменные звёзды от статичных объектов.
- Многоклассовая классификация:** Разработать систему, позволяющую распределять переменные звёзды по типам (например, цефеиды, RR Лиры, затменные системы) с учётом спектральных и фотометрических характеристик.

- 3. Интеграция временных рядов:** Внедрить анализ кривых блеска для выделения динамических признаков, таких как периодичность и амплитуда изменений яркости, и объединить их с другими данными для повышения точности классификации.

Реализация каждой из указанных задач предполагает последовательное выполнение этапов по сбору и предварительной обработке данных, разработке и обучении моделей, а также оптимизации их гиперпараметров с использованием методов кросс-валидации.

3. Описание исходных данных

Для построения модели используются данные, полученные из множества астрономических обзоров. Эти данные охватывают широкий спектр (от ультрафиолета до рентгеновских диапазонов) и включают как спектральную, так и временную информацию.

3.1. Основные источники данных

Среди используемых источников можно выделить:

- **Спектральные данные:** каталоги ATLAS, Gaia DR3, APASS, Pan-STARRS, предоставляющие информацию о координатах, яркости в различных диапазонах и спектральных характеристиках.
- **Инфракрасные наблюдения:** данные 2MASS и WISE, охватывающие J, H, Ks и другие инфракрасные диапазоны, что позволяет исследовать холодные и пылевые объекты Skrutskie и др., 2006; Wright и др., 2010.
- **Ультрафиолет и рентген:** GALEX, 1RXS, 2RXS, а также данные с eROSITA, применяемые для изучения высокоэнергетических процессов.
- **Кривые блеска:** временные ряды, полученные из обзоров CRTS, ASAS, NSVS, обеспечивают возможность анализа периодичности и амплитудных характеристик переменных звезд Drake и др., 2009.
- **Каталоги переменных звёзд:** AAVSO VSX и данные Gaia DR3 содержат сведения о типах переменных звёзд, их периодах и амплитудах, что является ключевым для обучения моделей.

3.2. Примеры структуры данных

Ниже приведены таблицы с примерами характеристик, извлекаемых из различных источников.

Таблица I: Примеры параметров из спектральных и фотометрических каталогов

Каталог	Параметр	Описание	Пример значения
ATLAS	Идентификатор	Уникальный номер объекта	ATLAS-123456
	RA (часы)	Прямое восхождение	12.3456
	Dec (градусы)	Склонение	-45.6789
	Магнитуда (G)	Яркость в полосе G	15.2
2MASS	Идентификатор	Уникальный номер звезды	2MASS J12345678+1234567
	Магнитуды (J, H, Ks)	Измерения в ИК диапазоне	12.3, 11.8, 11.5
WISE	Идентификатор	Уникальный номер объекта	WISE J123456.78+123456.7
	Магнитуды (W1–W4)	Измерения в инфракрасном диапазоне	10.5, 10.3, 9.8, 9.5
APASS	Идентификатор	Номер объекта	APASS 123456
	Магнитуды (B, V, g', r', i')	Фотометрия в оптических диапазонах	14.2, 13.8, 13.5, 13.3, 13.1

Таблица II: Примеры параметров временных рядов и классификации переменных звёзд

Каталог	Параметр	Описание	Пример значения
CRTS	Идентификатор	Уникальный номер объекта	CRTS J123456.7+123456
	Дата наблюдения	Момент измерения яркости	2455000.1234
	Магнитуда	Измеренная яркость	15.0
AAVSO VSX	Тип переменной	Классификация переменной звезды	RR Lyr
	Период (дней)	Период изменения яркости	0.567

4. Методы и подходы

Разработка классификатора переменных звёзд осуществляется в несколько этапов, каждый из которых направлен на постепенное усложнение модели и повышение её точности.

4.1. Этап 1: Бинарная классификация

На данном этапе разрабатывается базовая модель для разделения переменных и непеременных объектов. Основное внимание уделяется:

- **Предварительной обработке данных:** нормализация, фильтрация шумов, удаление аномальных значений и заполнение пропусков.

- **Использованию классических алгоритмов:** метод опорных векторов (SVM), случайный лес или градиентный бустинг (например, LightGBM) для первичной классификации В. е. al., 2019.

4.2. Этап 2: Многоклассовая классификация по типам переменных звёзд

После успешного создания бинарного классификатора модель расширяется для детального разделения переменных звёзд по типам. Здесь учитываются:

- **Спектральные и фотометрические признаки:** анализ цветовых индексов, температур, металличности и других параметров.
- **Учёт ошибок измерений:** интеграция информации об ошибках для повышения надёжности классификации.
- **Применение ансамблевых методов и нейронных сетей:** для моделирования нелинейных зависимостей между признаками К. е. al., 2020.

4.3. Этап 3: Интеграция временных рядов (анализ кривых блеска)

Финальный этап включает обработку временных рядов, полученных из обзоров CRTS, ASAS, NSVS и других источников:

- **Анализ кривых блеска:** с использованием методов, таких как периодограмма Ломба-Скаргла, для выявления периодичности и амплитуды изменений.
- **Извлечение динамических признаков:** построение фазированных кривых блеска, которые затем объединяются с остальными спектральными и фотометрическими данными.
- **Применение нейронных сетей:** сверточных (CNN) и рекуррентных (LSTM) для обработки одномерных временных рядов и интеграции динамической информации Nun и др., 2018; Balázs и др., 2020.

4.4. Оптимизация и кросс-валидация

Для повышения устойчивости и точности модели используются методы оптимизации гиперпараметров, а также кросс-валидация (например, 5-кратная) для предотвращения переобучения и выбора оптимальных параметров моделей М. е. al., 2018.

5. План проведения исследования

Проект реализуется в несколько этапов, каждый из которых включает выполнение следующих задач:

5.1. Этап 1: Сбор и предварительная обработка данных

1. Сбор информации из различных астрономических каталогов (ATLAS, 2MASS, WISE, APASS, Gaia DR3, GALEX, CRTS, AAVSO VSX и др.).
2. Очистка данных от шумов, удаление выбросов и заполнение пропущенных значений.
3. Нормализация и стандартизация фотометрических и спектральных параметров.

5.2. Этап 2: Разработка базового бинарного классификатора

1. Выбор и тестирование классических алгоритмов (SVM, случайный лес, градиентный бустинг).
2. Обучение модели на предварительно обработанных данных.
3. Оценка точности модели с использованием стандартных метрик (accuracy, precision, recall, F1-score).

5.3. Этап 3: Создание многоклассового классификатора

1. Расширение модели для классификации переменных звёзд по типам (цефеиды, RR Лиры, затменные системы и пр.).
2. Интеграция спектральных, фотометрических признаков и информации об ошибках измерений.
3. Оценка производительности модели для каждого класса с использованием соответствующих метрик.

5.4. Этап 4: Анализ временных рядов и интеграция динамических данных

1. Извлечение и обработка временных рядов (кривых блеска) из астрономических наблюдений.
2. Фазировка данных и построение периодических кривых блеска.
3. Обучение сверточных и рекуррентных нейронных сетей с целью интеграции динамических признаков в итоговую модель.

5.5. Этап 5: Финальный анализ, визуализация результатов и составление отчёта

1. Проведение детального анализа результатов классификации, построение графиков, таблиц и диаграмм.
2. Сравнение эффективности различных моделей и выбор оптимальной архитектуры.
3. Подготовка научной публикации с описанием методологии, результатов и выводов.

6. Приложение

6.1. Теоретическая справка по астрофизике

Переменные звёзды представляют собой объекты, яркость которых изменяется со временем вследствие внутренних и внешних процессов. Такие изменения могут быть обусловлены пульсационными процессами, затмениями в двойных системах или аккреционными процессами в системах с компактными объектами.

6.1.1. Спектральный анализ

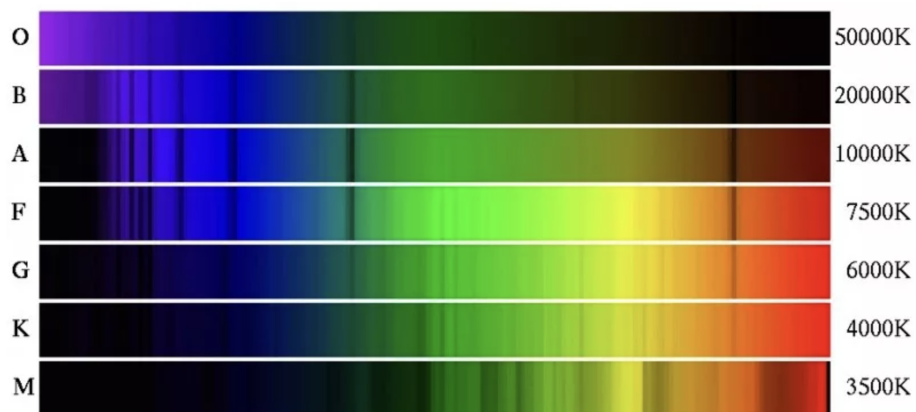


Рис. 1: Пример спектра звезды различной температуры

Спектр звезды отражает распределение её излучения по длинам волн и позволяет определить физические параметры, такие как температура, химический состав, давление и скорость движения. Классификация по спектральным типам (O, B, A, F, G, K, M) основана на особенностях спектральных линий.

6.1.2. Кривые блеска

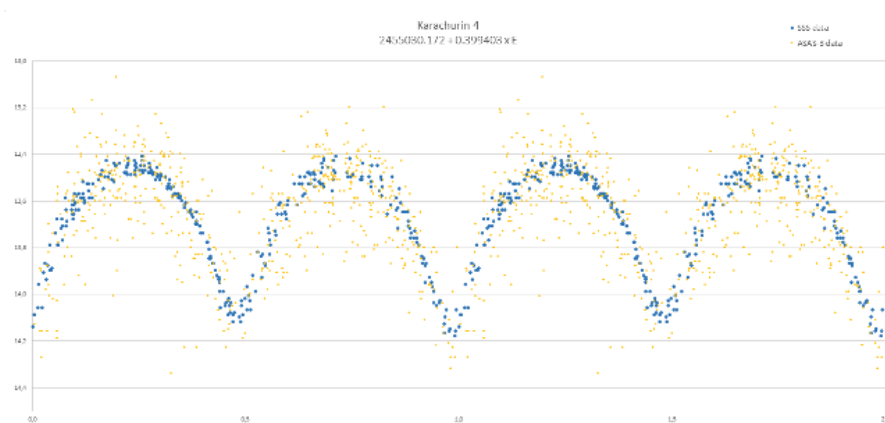


Рис. 2: Пример кривой блеска звезды (вариант 1)

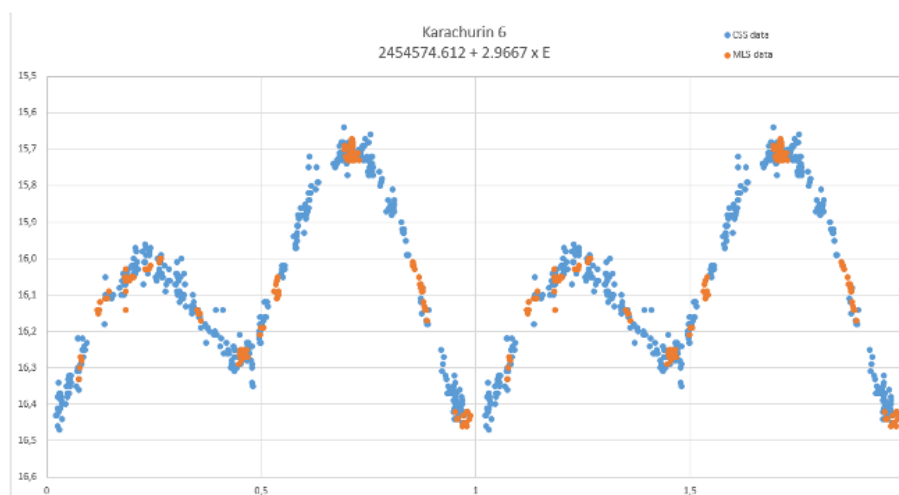


Рис. 3: Пример кривой блеска звезды (вариант 2)

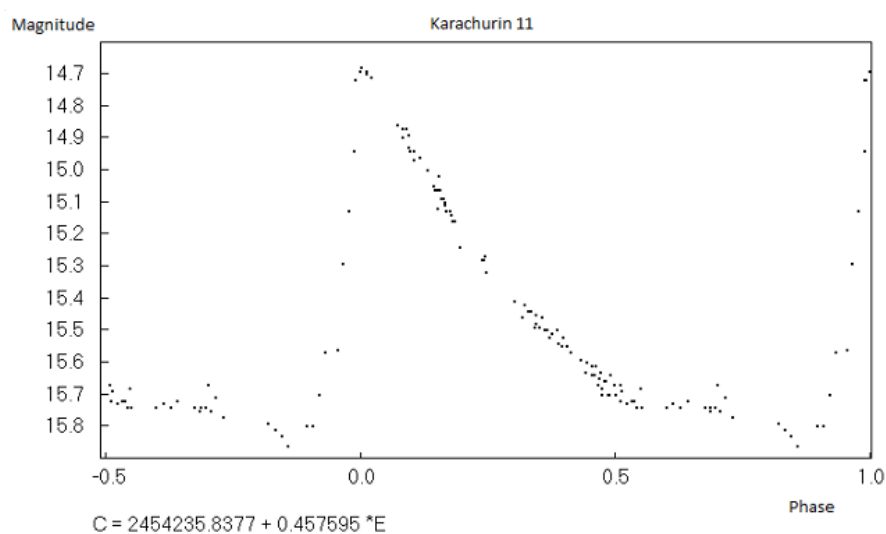


Рис. 4: Пример кривой блеска звезды (вариант 3)

Кривые блеска демонстрируют изменения яркости звезды во времени. Основными характеристиками являются период, амплитуда и форма кривой. Эти

параметры позволяют классифицировать переменные звёзды и выявлять особенности их поведения.

6.1.3. Классификация переменных звёзд

Переменные звёзды условно делятся на две большие группы:

1. **Внутренне переменные звёзды**, где изменения яркости обусловлены внутренними пульсациями. В эту группу входят:
 - **Цефеиды** – обладают законом период-светимость и используются для определения космических расстояний.
 - **Звёзды типа RR Лиры** – характеризуются короткими периодами и асимметричными кривыми блеска.
 - **Другие пульсирующие переменные**, включая δ Щита и HADS, с различными подтипами, отражающими особенности пульсаций.
2. **Внешне переменные звёзды**, изменения яркости которых связаны с внешними факторами, например, затмениями в двойных системах:
 - **Затменно-переменные звёзды (EA, EB, EW)** – системы, в которых один компонент периодически закрывает другой.
 - **Релятивистские и аккрецирующие системы**, где изменения яркости вызваны процессами аккреции или взаимодействиями в двойных системах.

6.2. Физические процессы переменности

Основными механизмами, приводящими к изменению яркости звёзд, являются:

- **Пульсации**, вызванные неустойчивостью внутренних слоёв звезды.
- **Аккреция**, когда вещество переходит с одной звезды на другую в двойной системе.
- **Эруптивные процессы**, характерные для молодых звёзд или объектов с активной магнитной активностью.
- **Затмения**, возникающие при прохождении одного компонента двойной системы перед другим.

Такая комплексная классификация и анализ позволяют использовать переменные звёзды для калибровки космических расстояний и исследования эволюционных процессов в звёздах Freedman и Madore, 2010; Hilditch, 2001; Riess и др., 2016.

Список литературы

al., Bazin et (2019). «Photometric light curves classification with machine learning». В: *arXiv preprint*. URL: <https://ar5iv.labs.arxiv.org/html/1909.05032>.

- al., Kim et (2020). «Variable star classification with machine learning methods». B: *Monthly Notices of the Royal Astronomical Society* 491, с. 3805—3816. URL: <https://academic.oup.com/mnras/article/491/3/3805/5625784>.
- al., Mackenzie et (2018). «Deep multi-survey classification of variable stars». B: *arXiv preprint*. URL: <https://arxiv.labs.arxiv.org/html/1810.09440>.
- Balázs, Csörgő и др. (2020). «Image-based classification of variable stars - First results on OGLE data». B: *arXiv preprint arXiv:2006.07614*. URL: <https://arxiv.org/abs/2006.07614>.
- Cox, J. P. (1980). «Stellar Pulsation Theory». B: *Annual Review of Astronomy and Astrophysics* 18, с. 15—42.
- Drake, A. J. и др. (2009). *The Catalina Real-time Transient Survey (CRTS)*. URL: <http://crts.caltech.edu/>.
- Freedman, W. L. и B. F. Madore (2010). «The Hubble Constant». B: *Annual Review of Astronomy and Astrophysics* 48, с. 673—710.
- Goodfellow, Ian, Yoshua Bengio и Aaron Courville (2016). *Deep Learning*. MIT Press.
- Hilditch, R. W. (2001). *An Introduction to Close Binary Stars*. Cambridge University Press.
- Nun, Igor и др. (2018). «Deep multi-survey classification of variable stars». B: *arXiv preprint arXiv:1810.09440*. URL: <https://arxiv.org/abs/1810.09440>.
- Nun, Igor, Carl Mackenzie, Josh Long и др. (2020). «Deep multi-survey classification of variable stars». B: *Monthly Notices of the Royal Astronomical Society* 491.3, с. 3805—3816. URL: <https://academic.oup.com/mnras/article/491/3/3805/5625784>.
- Riess, A. G. и др. (2016). «A 2.4% Determination of the Local Value of the Hubble Constant». B: *The Astrophysical Journal* 826.1, с. 56.
- Skrutskie, M. F. и др. (2006). «The Two Micron All Sky Survey (2MASS)». B: *Astronomical Journal* 131, с. 1163. URL: <https://irsa.ipac.caltech.edu/Missions/2mass.html>.
- Wright, E. L. и др. (2010). «The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance». B: *Astronomical Journal* 140, с. 1868—1881. URL: <https://irsa.ipac.caltech.edu/Missions/wise.html>.