

Отчёт по теме: «Выбор датасета (dataset)»

Выполнил:

Студент группы БМП-16-2

Новицкий Д. А.

Проверил:

Курочкин И. И.

Оглавление

| | |
|--|---|
| ПРОГНОЗИРОВАНИЕ ПУЛЬСАРСКОЙ ЗВЕЗДЫ | 3 |
| Информация об атрибутах: | 4 |
| Информация о выборке | 5 |

ПРОГНОЗИРОВАНИЕ ПУЛЬСАРСКОЙ ЗВЕЗДЫ

Доктор Роберт Лион

HTRU2 - это набор данных, в котором описывается выборка кандидатов в пульсары, собранных в ходе Вселенной с высоким разрешением.

Пульсары – это редкий тип нейтронных звезд, которые производят радиоизлучение, обнаруживаемое здесь, на Земле. Они представляют значительный научный интерес как зонды пространства-времени, межзвездной среды и состояния материи.

Когда пульсары вращаются, их луч излучения распространяется по небу, и когда он пересекает нашу линию обзора, создает обнаруживаемую картину широкополосного радиоизлучения. Поскольку пульсары вращаются быстро, эта схема периодически повторяется. Таким образом, поиск пульсаров предполагает поиск периодических радиосигналов на больших радиотелескопах.

Каждый пульсар производит немного различную диаграмму эмиссии, которая слегка меняется с каждым вращением. Таким образом, обнаружение потенциального сигнала, известное как «кандидат», усредняется по многим поворотам пульсара, что определяется продолжительностью наблюдения. При отсутствии дополнительной информации каждый кандидат потенциально мог бы описать настоящий пульсар. Однако на практике почти все обнаружения вызваны радиочастотными помехами (RFI) и шумом, что затрудняет поиск законных сигналов.

Инструменты машинного обучения в настоящее время используются для автоматической маркировки кандидатов на пульсар для облегчения быстрого анализа. В частности, широко применяются системы классификации, в которых наборы данных-кандидатов рассматриваются как проблемы двоичной классификации. Здесь законные примеры пульсаров – это положительный класс меньшинства, а ложные примеры - отрицательный класс большинства.

Совокупность данных, представленная здесь, содержит 16 259 ложных примеров, вызванных RFI / шумом, и 1 639 реальных примеров пульсаров. Все эти примеры были проверены аннотаторами-людьми.

Каждая строка сначала перечисляет переменные, а метка класса является последней записью. Используемые метки класса: 0 (отрицательный) и 1 (положительный).

Информация об атрибутах:

Каждый кандидат описывается 8 непрерывными переменными и одной переменной класса. Первые четыре представляют собой простую статистику, полученную из интегрированного профиля импульса (сложенный профиль). Это массив непрерывных переменных, которые описывают версию сигнала с разрешением по долготе, усредненную как по времени, так и по частоте. Остальные четыре переменные аналогично получены из кривой DM-SNR. Они обобщены ниже:

1. Среднее значение интегрированного профиля.
2. Стандартное отклонение интегрированного профиля.
3. Лишний эксцесс интегрированного профиля.
4. Асимметрия интегрированного профиля.
5. Среднее значение кривой DM-SNR.
6. Стандартное отклонение кривой DM-SNR.
7. Избыточный эксцесс кривой DM-SNR.
8. Асимметрия кривой DM-SNR.
9. Класс

Информация о выборке

Всего 17 898 примеров.

1 639 положительных примеров.

16 259 отрицательных примеров.

ИСТОЧНИКИ

<https://archive.ics.uci.edu/ml/datasets/HTRU2>