

Построение классификатора на имеющемся наборе данных.

Евдокимов Данил, Стельмах Татьяна, Назаров Максим, Владимирова Элина

30 декабря 2020 г.

1 Резюме

Была решена задача распределения данных по известным классам в зависимости от определенных признаков каждого конкретного объекта на примере классификации набора организмов по царствам живой природы в согласовании с некоторыми их особенностями и характеристиками.

Входные данные - список различных живых организмов с уникальным набором признаков для каждого. Результатом работы программы является распределение объектов - организмов - по массивам - царствам.

2 Постановка задачи

Необходимо построить классификатор по известному набору данных, т.е. в зависимости от имеющихся признаков поместить каждый объект набора в определенный класс.

3 Допущения

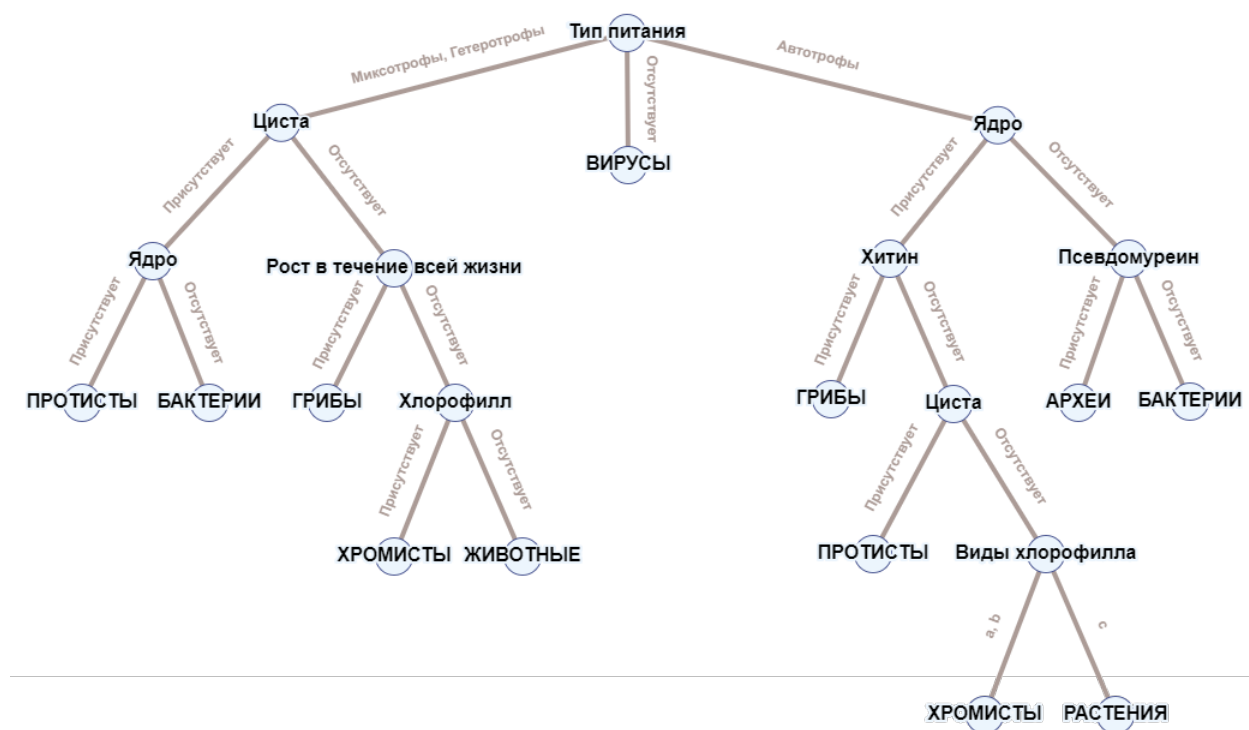
- Классы не пересекаются, т.е. 1 объект может принадлежать только 1 классу.
- Любой объект гарантированно принадлежит какому-либо из классов.
- Набор признаков объекта однозначно определяет его принадлежность к тому или иному классу.
- Все объекты считываются из файла с определенным названием и расширением - *animals* - *Lucm1.csv*. Предполагается корректное задание каждого объекта.

4 Описание решения

В нашем случае проводилось распределение живых организмов по биологическим царствам в зависимости от значений 10 признаков:

1. Тип питания	Автотрофы, гетеротрофы и миксотрофы
2. Наличие ядра	Присутствует/Отсутствует
3. Способность образовывать цисту	Присутствует/Отсутствует
4. Рост в течение всей жизни	Присутствует/Отсутствует
5. Наличие хлорофилла	Присутствует/Отсутствует
6. Тип хлорофилла	a, b, c
7. Наличие хитина	Присутствует/Отсутствует
8. Наличие псевдомуреина	Присутствует/Отсутствует

Прежде всего задаются списки, в которых будет храниться информация об индексах объектов, после чего производится считывание информации из данного файла. Для классификации используется структура "дерева": при проходе по всем объектам на каждом шаге для каждого элемента происходит обращение к конкретному признаку, что постепенно сужает выбор между потенциальными классами и в итоге приводит к однозначному определению класса объекта.



5 Результаты

Для подборки организмов был сконструирован классификатор, корректно и быстро определяющий для каждого организма царство, к которому он принадлежит. Правильность работы программы проверена вручную.

Идея

Тактическая концепция

Подбор данных

Код и проверка корректности его работы

Отчет

Назаров Максим

Назаров Максим, Владимирова Элина, Стельмах Татьяна

Евдокимов Данил, Стельмах Татьяна

Евдокимов Данил, Стельмах Татьяна

Евдокимов Данил, Владимирова Элина