

Отчет по алгоритмам ленивой классификации для объектов, представленных бинарными признаками.

Тесты проводились на данных из репозитория tic-tac-toe.

Описание алгоритмов:

1. Метод ленивой классификации на основе генераторов (a1.py).

Вычисляется пересечение описания тестируемого объекта с объектами плюс и минус контекста. Далее идет проверка фальсифицируемости гипотезы о значимости такого пересечения – для плюс-гипотез ищем объекты в минус-контексте, для минус-гипотез, соответственно, в плюс-контексте. Голосуют только не фальсифицированные гипотезы. Для каждого объекта подсчитывается число голосов, решение принимается в пользу большего числа голосов.

2. Метод ленивой классификации с цензурированием голосом (a2.py).

Метод аналогичен предыдущему.

Различие состоит в способе взвешивания голосов. В данном методе не учитываются голоса, если пересечение описаний тестового объекта и текущего объекта плюс или минус-контекста слишком «общее» – должно включать в себя не менее 60% признаков. Фальсифицируемые гипотезы не могут участвовать в голосовании.

3. Метод ленивой классификации с цензурированием голосов и нормировкой (a3.py)

Метод аналогичен предыдущему.

Также имеет ограничивающее условие для размера пересечения (60%).

В дополнение к этому разрешает фальсифицированным гипотезам участвовать в голосовании. Условие участия – поддержка этих гипотез должна быть больше, чем замыкание в противоположном контексте.

Дополнительно имеет нормировку числа голосов на размер контекста.

4. Метод ленивой классификации с цензурированием и перевзвешиванием голосов (a4.py)

Метод аналогичен предыдущему.

Имеет единственное существенное дополнение – вес голоса зависит от разности долей пересечения гипотезы с плюс-контекстом и минус-контекстом.

Результаты кросс-валидации:

Алгоритм	True Positive	True Negative	False Positive	False Negative	NA
a1.py	626	272	55	0	5
a2.py	625	254	68	0	11
a3.py	501	332	0	79	46
a4.py	497	306	1	80	74

Метрики точности классификации:

Алгоритм	True Positive Rate	True Negative Rate	Positive Predictive Value	Negative Predictive Value
a1.py	1	0,831804281	0,919236417	1
a2.py	1	0,788819876	0,901875902	1
a3.py	0,863793103	1	1	0,807785888
a4.py	0,86135182	0,996742671	0,997991968	0,792746114