

Генетический выбор частичных порядков на множестве значений признаков в задаче классификации

Сорокин Олег, 317

ММП ВМК МГУ

Спецсеминар
9 марта 2023 г.

- 1 Обзор статей про частичные порядки
 - Общая постановка задачи
 - Алгоритм классификации и процедуры упорядочения

Постановка задачи классификации для произведений частичных порядков

Рассматривается более общая постановка задачи, приведённая ниже.

$M = \cup_{n=1}^l K_n$, где $K_i \cap K_j = \emptyset$ при $i \neq j$.

Пусть M представимо в виде $N_1 \times \dots \times N_n$, где N_i ($i \in \{1, 2, \dots, n\}$) — конечное множество допустимых значений признака x_i . Не ограничивая общности, можно считать, что N_i имеет наибольший элемент k_i .

Пусть также задан набор прецедентов

$S_1 = (a_{11}, \dots, a_{1n})$, $S_2 = (a_{21}, \dots, a_{2n})$, ..., $S_m = (a_{m1}, \dots, a_{mn})$.

Требуется по предъявленному набору значений признаков (a_1, \dots, a_n) объекта $S \in M$ (класс которого, вообще говоря, неизвестен) определить этот класс.

О рассматриваемом классе алгоритмов

- 1 Обучение: для каждого класса K строится некоторое множество представительных эл. кл. $C^A(K)$.
- 2 Процедура голосования: вычисление оценок вида

$$\Gamma(S, K) = \frac{1}{|C^A(K)|} \sum_{(\sigma, H) \in C^A(K)} P_{(\sigma, H)} * \hat{B}(\sigma, S, H)$$

Быстрая процедура независимого линейного упорядочения значений признаков

Частичные порядки в этой процедуре строятся после анализа частот встречаемости значений признаков.

Определение

Частичный порядок на M называется (A, K) -корректным, если алгоритм A правильно классифицирует каждый объект из $R(K)$.

Замечания

- 1 Порядок на множестве значений каждого признака выбирается независимо от выбора порядков для других признаков.
- 2 Описанная процедура не является корректной в смысле определения, приведённого выше.

Процедура корректного упорядочения значений признаков

В ходе процедуры строится булева матрица B_K особого вида.

Рассматривается некоторый алгоритм A из описанного ранее класса. Если множество $C^A(K)$ содержит все тупиковые представительные эл. кл. класса K , то для алгоритма справедлива

Теорема.

Частичный порядок, заданный на множестве M , является (A, K) -корректным тогда и только тогда, когда существует неприводимое покрытие H матрицы B_K такое, что $\forall j \in \{1, 2, \dots, n\}$ и $\forall a, b \in N_j$ ($a < b$) столбец (j, b, a) не входит в H .

Использованная литература

- О выборе частичных порядков на множествах значений признаков в задаче классификации. Дюкова Е. В., Масляков Г. О.
- О логическом анализе данных с частичными порядками в задаче классификации по прецедентам. Дюкова Е. В., Масляков Г. О., Прокофьев П. А.