Глоссарий к Теме 11
Анализ с помощью CHAID

CHAID (Chi Squared Automatic Interaction Detection) анализ – используется для построения прогностической модели, основанной на системе классификации. Это метод эффективного поиска взаимосвязи между предикторными переменными и категориальным откликом.

Exhaustive CHAID (Исчерпывающий CHAID) – является модификацией CHAID и позволяет обойти некоторые недостатки CHAID:

- В алгоритме CHAID слияние категорий останавливается как только обнаруживается, что все оставшиеся категории статистически различны. ExCHAID продолжает сливать категории переменной-предиктора, пока их не останется только
- ExCHAID сливает категории в разных комбинациях, для того, чтобы найти оптимальное разбиение на 2 категории для данной переменной-предиктора (такое, что p-значение будет минимальным).

CRT (Classification and Regression Trees) – программа деревьев классификации, которая при построении дерева осуществляет полный перебор всех возможных вариантов одномерного ветвления (бинарный).

QUEST (Quick, Unbiased, Efficient Statistical Tree) – это программа деревьев классификации, в которой используются улучшенные варианты метода рекурсивного квадратичного дискриминантного анализа и которая содержит ряд новых средств для повышения надежности и эффективности деревьев классификации, которые она строит. (бинарный).

Дерево решений (decision tree) – это способ представления данных в иерархической, последовательной структуре, где каждому объекту соответствует единственный узел, дающий решение.

Корневой узел – верхний разбиваемый узел, представляющий всю выборку.

Узел-сын– новые узлы, получившиеся в результате разбиения.

Узел-отец – узел, который был расщеплен.

Терминальные узлы – окончательные узлы, которые в дальнейшем не разбиваются. Их еще называют листьями, потому что в них рост дерева решений останавливается. Лист представляет собой наилучшее окончательное решение.

Глубина дерева – количество уровней, образующихся от узлов, не считая родительский узел (можно установить вручную в SPSS).

Правило остановки –критерий определения "подходящего размера" дерева классификации; состоит из глубины дерева, минимальное количество наблюдений в узле-отце и в узле-сыне.

Номинальный тип – характерен для категориальных переменных с дискретными значениями, когда значениям не приписывается конкретный порядок. Можно объединять любые категории, если они не различаются значимо.

Порядковый тип – характерен для переменных с дискретными значениями, когда задан порядок значений. Две категории могут быть объединены, только если есть возможность присоединения к ним промежуточных категорий.

Непрерывный тип – по умолчанию CHAID преобразует непрерывную числовую предикторную переменную в порядковую, имеющую 10 категорий с приблизительно равным числом наблюдений. Эти категории формируются путем объединения соседних значений с исходной переменной.

© Академия НАФИ
Актуальная аналитика на nafi.ru Онлайн-опросы за вознаграждение: tetotvet.ru