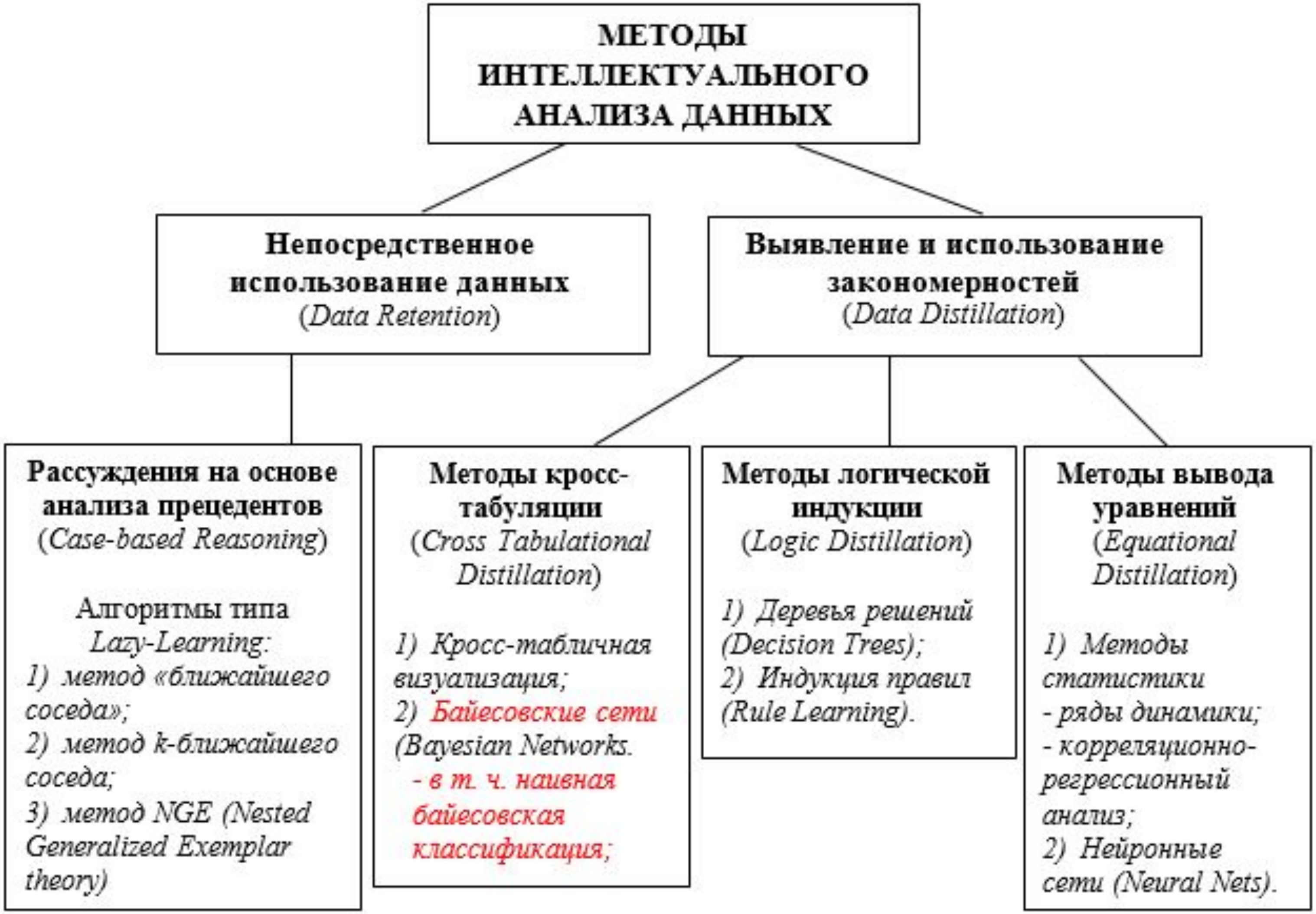
Лекция 6

БАЙЕСОВСКИЕ МЕТОДЫ КЛАССИФИКАЦИИ

Байесовские методы. Байесовская классификация



Классификация технологических методов ИАД [1]

Байесовские методы. Байесовская классификация

Суть методов кросс-табуляции

Кросс-табуляция является простой формой анализа, широко используемой в генерации отчетов средствами систем оперативной аналитической обработки (OLAP).

Двумерная кросс-таблица представляет собой матрицу значений, каждая ячейка которой лежит на пересечении значений атрибутов.

Расширение идеи кросс-табличного представления на случай гиперкубической информационной модели является основой многомерного анализа данных, поэтому эта группа методов рассматривается как симбиоз многомерного оперативного анализа и интеллектуального анализа данных.

К методам ИАД группы кросс-табуляции относится также использование байесовских сетей (***Bayesian Networks***).

Байесовские методы. Байесовские сети

Байесовская сеть (или байесова сеть, байесовская сеть доверия) - ***графическая вероятностная модель, представляющая собой множество переменных и их вероятностных зависимостей.***

Математический аппарат байесовых сетей создан американским ученым *Джудой Перлом,* лауреатом Премии Тьюринга (2011 г.).

Формально: ***байесовская сеть*** - это направленный ациклический граф, каждой вершине которого соответствует случайная переменная, а дуги графа кодируют отношения условной независимости между этими переменными.

Вершины могут представлять переменные любых типов, быть взвешенными параметрами, скрытыми переменными или гипотезами.

Байесовские методы. Байесовские сети

Байесовская сеть - это полная модель для переменных и их отношений

Следовательно, она может быть использована для того, чтобы давать ответы на вероятностные вопросы.

Этот процесс вычисления апостериорного распределения переменных по переменным-свидетельствам называют ***вероятностным выводом***, что дает универсальную оценку для приложений, где нужно выбрать значения подмножества переменных, которое минимизирует функцию потерь, например, вероятность ошибочного решения.

Байесовская сеть позволяет получить ответы на следующие типы вероятностных запросов:

1) нахождение вероятности свидетельства,

2) определение априорных маргинальных вероятностей,

3) определение апостериорных маргинальных вероятностей, включая:

- *прогнозирование*, или прямой вывод - определение вероятности события при наблюдаемых причинах),
- *диагностирование*, или обратный вывод (абдукция), - определение вероятности причины при наблюдаемых следствиях,
- *межпричинный* (смешанный) *вывод* или трансдукция, - определение вероятности одной из причин наступившего события при условии наступления одной или нескольких других причин этого события.

4) вычисление наиболее вероятного объяснения наблюдаемого события,

5) вычисление апостериорного максимума.

Байесовские методы. Байесовская классификация

- Ранее байесовская классификация использовалась для формализации знаний экспертов в экспертных системах.
- В настоящее время байесовская классификация применяется и в качестве одного из методов Data Mining.
- Байесовские методы получили достаточно широкое распространение и активно используются в самых различных областях знаний.

История вопроса: *формула Байеса* была опубликована в 1763 году спустя 2 года после смерти *Томаса Байеса*.
Методы, использующие ее, получили широкое распространение только к концу 20 века - расчеты требуют определенных вычислительных затрат, и они стали возможны лишь с развитием информационных технологий.

Байесовские методы. Байесовская классификация

Байесовский метод опирается на теорему о том, что если плотности распределения классов известны, то алгоритм классификации, имеющий минимальную вероятность ошибок, можно выписать в явном виде.

Для оценивания плотностей классов по выборке применяются различные подходы (в частности, параметрический, непараметрический и оценивание смесей распределений).

В ***байесовских классификаторах*** используется *критерий, минимизирующий вероятность принятия ошибочного решения,* поэтому байесовские алгоритмы являются статистически оптимальными.

Но при этом алгоритмы требуют в идеале полного знания многомерных функций распределения наблюдаемых признаков для каждого класса. Необходимость такого знания обусловлена использованием ***формулы Байеса***, которая лежит в основе ***байесовских методов*** принятия решения.

Байесовский подход основан на предположении, что задача выбора решения сформулирована в терминах теории вероятностей и известны все представляющие интерес вероятностные величины. В основе байесовской классификации лежит ***правило Байеса.***

Байесовские методы. Байесовская классификация

Теорема Байеса позволяет рассчитать апостериорную вероятность $P(c|x)$ на основе $P(c)$, $P(x)$ и $P(x|c)$:

$$P(c|x) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

$P(c|x)$ – апостериорная вероятность данного класса c

(т.е. данного значения целевой переменной) при данном значении признака x.

$P(c)$ – априорная вероятность данного класса.

$P(x|c)$ – правдоподобие, т.е. вероятность данного значения признака при данном классе.

$P(x)$ – априорная вероятность данного значения признака.

Байесовская классификация. Постановка задачи

Рассмотрим обучающую выборку из n объектов, каждый из которых принадлежит одному из K классов и характеризуется набором m числовых признаков a_1, a_2,... a_m.

Пусть имеется n_k объектов k-ого класса, так что

$$N=\Sigma n_k,\ k=1,\ldots,\text{K}.$$

Значение j-ого признака i-ого объекта из k-ого класса обозначим x_{ijk}.

Тогда этот объект можно охарактеризовать вектором-строкой $x_{ik}=(x_{i1k},\ldots,x_{ijk},\ldots,x_{imk})$. Эту строку будем рассматривать как i-ю реализацию векторной случайной величины ξ_k, подчиняющейся распределению вероятностей с плотностью $p(x_1,\ldots,x_m|k)$, своей для каждого класса k.

Пусть теперь наблюдается объект, для которого необходимо определить, к какому классу он относится. Объект характеризуется только набором m числовых признаков $x_1,\ldots,x_m$.

Байесовская классификация. Общая структура байесовского классификатора

В основе классификатора лежит следующее ***правило***. Классификатор вычисляет ***апостериорную вероятность*** $P(k|x)$ каждого класса k, которому может принадлежать испытуемый объект, и относит этот объект к апостериорно наиболее вероятному классу $\hat{k}$:

$$\hat{k} = \arg\max_k \ln P(k|x_1, \ldots, x_m)$$

Апостериорная вероятность вычисляется по ***формуле Байеса***:

$$P(k|x_1,\ldots,x_m)=P(k)p(x_1,\ldots,x_m|k)/p(k),$$

где $P(k)$ – априорная вероятность того, что объект относится к k-ому классу, $p(k)$ и $p(x_1,\ldots,x_m|k)$ - безусловная и условная многомерные плотности распределения вектора признаков, компоненты которого обычно статистически зависимы.

Таким образом, байесовский классификатор предполагает, что многомерная совместная плотность распределения признаков известна для всех классов.

Байесовская классификация. Общая структура байесовского классификатора

Аналитическое представление многомерной плотности вероятности известно только для нормального распределения.

При этом многомерная нормальная плотность распределения дает подходящую модель для одного важного случая, а именно когда значения векторов признаков x для данного класса k представляются непрерывнозначными, слегка искаженными версиями единственного типичного вектора, или вектора-прототипа, μ_k.

Именно этого ожидают, когда классификатор выбирается так, чтобы выделять те признаки, которые, будучи различными для образов, принадлежащих различным классам, были бы, возможно, более схожи для образов из одного и того же класса.

Байесовская классификация. Общая структура байесовского классификатора

Многомерная нормальная плотность распределения в общем виде представляется выражением

$$p(x) = \frac{1}{(2\pi)^{\frac{m}{2}} \det R^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T R^{-1}(x-\mu)}$$

где μ – m-компонентный вектор среднего значения,
R – ковариационная матрица размера $m \times m$,
T – знак транспонирования.
Если все недиагональные элементы равны нулю, то $p(x)$ сводится к произведению одномерных нормальных плотностей компонент вектора x).

Байесовская классификация. Общая структура байесовского классификатора

Для многомерного нормального распределения удаётся выразить в аналитически замкнутой форме (с точностью до несущественных слагаемых) алгоритм байесовской классификации:

$$\hat{k} = \arg\max_{k}\left(\ln P(k) - \frac{1}{2}\ln\det R_k - \frac{1}{2}(x_k - \mu_k)R_k^{-1}(x_k - \mu_k)^T\right)$$

где μ_k – m-вектор-строка математических ожиданий значений признаков объектов класса k,
R_k – $m\times m$-матрица ковариаций векторов признаков класса k. Диагональные элементы матрицы образуют m-вектор D_k дисперсий признаков объектов класса k.

Алгоритм байесовской классификации с обучением состоит из двух этапов:
1) этап обучения;
2) этап классификации

Байесовские методы. Байесовская классификация

Достоинства байесовских сетей как метода Data Mining

- в модели определяются зависимости между всеми переменными, это позволяет легко обрабатывать ситуации, в которых значения некоторых переменных неизвестны;
- байесовские сети достаточно просто интерпретируются и позволяют на этапе прогностического моделирования легко проводить анализ по сценарию «что, если»;
- байесовский метод позволяет естественным образом совмещать закономерности, выведенные из данных, и, например, экспертные знания, полученные в явном виде;
- использование байесовских сетей позволяет избежать проблемы переучивания (overfitting), то есть избыточного усложнения модели, что является слабой стороной многих методов (например, деревьев решений и нейронных сетей).

Байесовские методы. Наивно-байесовский подход (Naive-Bayes Approach)

Суть метода наивно-байесовской классификации

В наивном байесовском классификаторе делается предположение о независимости признаков объекта. Если пренебречь статистическими связями между компонентами вектора признаков, тогда матрица R_k будет диагональной с вектором D_k на главной диагонали и классификатор станет ***наивным байесовским классификатором***.

Также предполагается, что ***маргинальная плотность распределения*** $p(xj|k)$ любого признака является нормальной для любого класса.

Но на практике так бывает далеко не всегда, то есть наблюдаемые данные не подчиняются нормальному закону распределения (в общем случае закон вообще неизвестен) и имеет место статистическая зависимость, поэтому область применения классификатора сужается.

Байесовские методы. Наивно-байесовский подход (Naive-Bayes Approach)

Наивно-байесовский подход имеет следующие недостатки:

- перемножать условные вероятности корректно только тогда, когда все входные переменные действительно статистически независимы;
- хотя часто данный метод показывает достаточно хорошие результаты при несоблюдении условия статистической независимости, но теоретически такая ситуация должна обрабатываться более сложными методами, основанными на обучении байесовских сетей;
- невозможна непосредственная обработка непрерывных переменных - требуется их преобразование к интервальной шкале, чтобы атрибуты были дискретными; однако такие преобразования иногда могут приводить к потере значимых закономерностей;
- на результат классификации в наивно-байесовском подходе влияют только индивидуальные значения входных переменных, комбинированное влияние пар или троек значений разных атрибутов здесь не учитывается.

Оптимальный байесовский классификатор

Так как ***оптимальный байесовский классификатор*** является модификацией наивного байесовского классификатора, то в качестве решающего правила также берётся рассмотренная ранее формула.

Одна из идей оптимизации наивно-байесовского классификатора состоит в том, чтобы, максимально используя обучающую выборку и гауссову копула-функцию, обойти два «наивных» предположения.

Модификация позволяет:

1) учесть статистические связи между наблюдаемыми признаками;

2) адаптировать классификатор к неизвестному действительному распределению путем приведения сглаженных маргинальных функций распределения признаков к нормальному виду.

Другими словами, с помощью нелинейных гауссовых копула-функций негауссовы данные преобразуются в гауссовы, которые можно подавать на вход классификатору.

Литература

1. Л. В. Щавелёв Способы аналитической обработки данных для поддержки принятия решений. (СУБД. - 1998. - № 4-5)
2. Воронцов К.В. Математические методы обучения по прецедентам (теория обучения машин) URL: http://www.ccas.ru/voron
3. Chickering D, Geiger D., Heckerman D. Learning Bayesian networks: The combination of knowledge and statistical data / Machine Learning. 1995. 20. P. 197-243
4. Heckerman D Bayesian Networks for Data Mining / Data Mining and Knowledge Discovery. 1997. № 1. P. 79-119
5. etc, Friedman N., Geiger D., Goldszmidt M. Bayesian Network Classifiers / Machine Learning. 1997. 29. P. 131-165
6. Brand E., Gerritsen R / Naive-Bayes and Nearest Neighbor DBMS. 1998. № 7