

Обработка множеств логических закономерностей с помощью
дисперсионного критерия

Выпускная квалификационная работа

Выполнил: Лисяной А. Е.

Руководитель: д.ф.-м.н., проф. Рязанов В. В.

Факультет Вычислительной Математики и Кибернетики
Кафедра Математических Методов Прогнозирования
МГУ им. М. В. Ломоносова

28 апреля 2015 г.

Задача классификации

Определение задачи классификации

Пусть имеется пространство объектов X и конечное множество имен классов $Y = \{1, \dots, M\}$. Пусть также имеется *обучающая выборка* $X^l = (x_i, y_i)_{i=1}^l$, в которой для каждого объекта x_i известен его класс $y_i \in Y$. Для восстановления целевой зависимости $y^*(x_i) = y_i$ построим алгоритм классификации $a: X \rightarrow Y$, аппроксимирующий y^* на всём пространстве объектов X .

Алгоритмы решения задачи классификации

- Метод логистической регрессии
- Метод опорных векторов
- Решающие деревья
- Нейронные сети
- Логические алгоритмы классификации

Определение логической закономерности

Пусть каждый объект выборки $x \in X^l$ имеет размерность D и пусть $\Omega \subseteq \{1, 2, \dots, D\}$. Предикат

$$\varphi(x) = P^{\Omega, c_1, c_2}(x) = \bigwedge_{j \in \Omega} P^{c_1^j, c_2^j}(f_j(x))$$

называется логической закономерностью класса K , если выполнено:

- ❶ $\exists x \in K: \varphi(x) = 1$
- ❷ $\forall x \notin K: \varphi(x) = 0$
- ❸ $\varphi(x)$ максимизирует некоторый критерий качества предиката Φ .

Обработка множества логических закономерностей

Построенное множество логических закономерностей:

- может содержать большое количество правил
- может содержать похожие правила

Это приводит к тому, что:

- логические закономерности сложно интерпретировать
- по похожим правилам плохо проводить классификацию

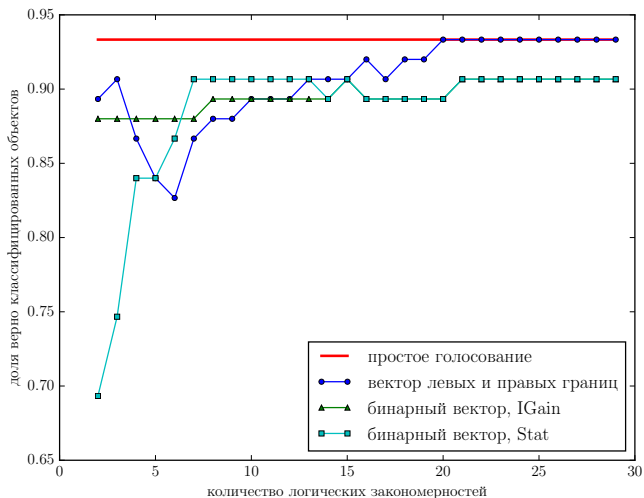
Задача обработки множества логических закономерностей

- По исходному множеству логических закономерностей построить множество меньшей мощности, что должно упростить пользователю задачу интерпретации полученных правил.
- Построенное множество логических закономерностей меньшей мощности должно иметь качество классификации, сравнимое с исходным множеством.

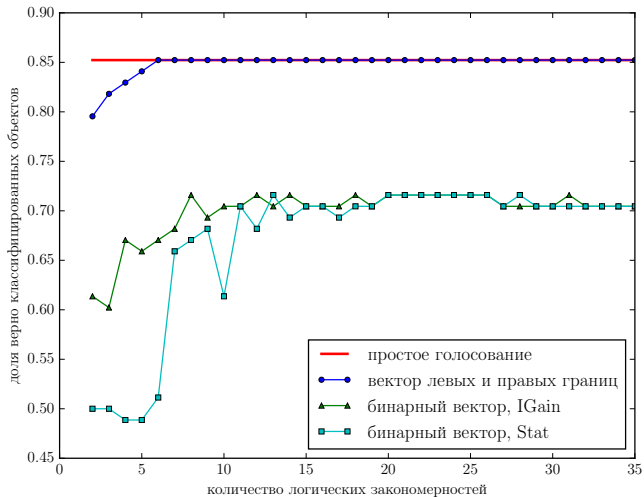
- 1 Для каждого из t правил составить признаковое описание
- 2 Кластеризовать на $k \leq t$ кластеров, найти их центры
- 3 По центрам кластеров построить k новых правил

Выборка	Всего объектов	Объекты по классам	Признаки
Iris	150	50/50/50	4
Wine	178	59/71/48	13
Climate	540	46/494	11
Ionosphere	351	126/255	34

Таблица: Сводная таблица по использованным данным



Выборка Iris. Классификация методом простого голосования



Выборка Wine. Классификация методом простого голосования

- Реализован метод обработки множеств логических закономерностей с помощью кластеризации на основе дисперсионного критерия.
- Проведено сравнение метода обработки, использующего вектор левых и правых границ, и метода обработки, использующего бинаризованное описание логических закономерностей.
- Экспериментально показано, что удастся получить обработанное множество логических закономерностей с меньшим числом элементов и сравнимым с исходным множеством качеством классификации.