Порядковая классификация с использованием частично упорядоченных наборов признаков

Папай Иван Дмитриевич

Московский физико-технический институт

Кафедра интеллектуального анализа данных ФПМИ МФТИ
Научный руководитель: д-р физ.-мат. наук В. В. Стрижов

2025

Порядковая классификация

Объект исследования

Определим задачу ранговой классификации. По имеющемуся набору объектов, каждый из которых задаётся множеством частично упорядоченных признаков, требуется воссоздать ранг каждого из них.

Цель исследования

Разработать метод решения данной задачи и сравнить его с уже существующими альтернативными подходами.

Пример датасета

В Football Player List под объектами имеются в виду футболисты, популярность каждого из которых требуется оценить по шкале от 1-го до 5-го порядка(неизвестный и невероятно популярный соответственно).

Постановка задачи ранговой классификации

Дана выборка

$$\{(X_i, y_i)\}_{i=1}^{m},$$

где $X_i = [x_{i1}, ..., x_{ij}, ..., x_{in}]$ это объект, который требуется классифицировать, y_i это метка класса. Объект X_i является n-мерным вектором, каждый компонент которого принадлежит ч.у.м.-у X_j.

Задача - найти монотонную функцию

$$f : X \to \hat{y},$$

такую, что отклонение её предсказаний от реальных меток классов будет минимальным.

$$S(f) = \sum_{i=1}^{n} |y_i - f(X_i)| \to min,$$

Частично упорядоченные множества

Другими словами, элемент j вектора X_i принадлежит множеству X_j, в свою очередь подчинённое некоторому частичному порядку $\succeq$ со следующими свойствами:

1. рефлексивность, $\forall a \in X \ (a \succeq a)$,
2. антисимметричность, $\forall a, b \in X$, $(a \succeq b) \wedge (b \succeq a) \Rightarrow (a = b)$,
3. транзитивность, $\forall a, b, c \in X \ (a \succeq b) \wedge (b \succeq c) \Rightarrow (a \succeq c)$.

Под объектом X имеется в виду декартово произведение ч.у.м.-ов $X_1, ..., X_n$,

$$X = X_1 \times X_2 \times \cdots \times X_n,$$

множества $X_1, ..., X_n$ являются множествами значений признаков.

Матрицы частичных порядков

Каждый частичный порядок $\succeq$, определённый на мн-ве X_j, описывается бинарной функцией $z_j(x_i, x_k)$

$$z_j(x_i, x_k) = \begin{cases} 1, & \text{при} \quad x_{ij} \succeq x_{kj}, \\ 0, & \text{при} \quad x_{ij} \not\succeq x_{kj}, \end{cases}$$

$$\begin{tikzcd} x_1 \arrow[r] & x_2 \arrow[r] \arrow[dr] & x_3 \\ & & x_4 \end{tikzcd}$$

Этот граф соответствует матрице Z_j:

$$Z_j = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Мы строим матрицы Z_j для каждого из частичных порядков X_j.

Конусы и генераторы

Выпуклый конус в $\mathbb{R}^m$ это такое множество $\mathcal{X}$, что

$$\mathcal{X} = \{\chi \mid A\chi \leq \mathbf{0}, \in \mathbb{R}^m\},$$

Теорема

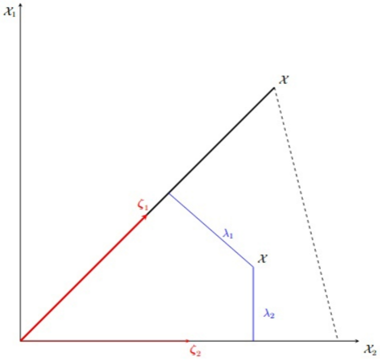
Вектор χ, принадлежащий конусу $\mathcal{X}$ может быть единственным образом разложен в ЛК с неотрицательными коэффициентами,

$$\chi = \sum_{k=1}^{m} \lambda_k \zeta_k, \quad \lambda_k \geq 0,$$

где ζ_k это генератор конуса $\mathcal{X}$,

$$\zeta_k(i) = \begin{cases} 1, & \text{if} \quad x_i \succeq x_k, \\ 0, & \text{if} \quad x_i \not\succeq x_k, \end{cases}$$

Вспомогательная иллюстрация



Задача оптимизации

Поскольку целевой вектор $\hat{y} \in \sum_{i=1}^{n} \mathcal{X}_i$, применим теорему выше для декомпозиции вектора $\hat{y}$ в линейную комбинацию конусов $\mathcal{X}_1, ..., \mathcal{X}_n$,

$$\hat{y} = \sum_{j=1}^{n} \sum_{k=1}^{m} \lambda_{jk} \zeta_{jk}, \quad \lambda_{jk} \in \mathbb{R}_+,$$

Вектор ζ_{jk} является k столбцом матрицы Z_j. Отсюда определим u для каждого объекта,

$$u(x) = \sum_{j=1}^{n} w_j \sum_{k=1}^{m} \lambda_{jk} z_j(x, x_k), \tag{1}$$

Таким образом мы свели задачу к минимизации лосса следующей функции по параметрам w и λ

$$f_{w,\lambda}(x_i) = \phi \left(\sum_{k=1}^{m} \lambda_k \Psi(x_i, x_k) \right) \tag{2}$$

Альтернативный подход: Изотоническая регрессия

Задача изотонической регрессии ставится следующим образом:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \sum_{i=1}^{n} w_i (x_i - y_i)^2 \tag{3}$$
$$\text{s.t.} \quad Ax \leq 0$$

Где матрица определена следующим образом:

$$A = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{pmatrix} = \begin{pmatrix} a_1^\top \\ \vdots \\ a_{n-1}^\top \end{pmatrix} \in \mathbb{R}^{n-1 \times n} \quad ,$$

Альтернативный подход: Косые решающие деревья

Решающие деревья задаются последовательностью условий на ЛК на каждом из узлов. Более общно говоря, возьмем в качестве примера $X = x_1; x_2; \ldots x_d;\ C_j$ где C_j метка класса и x_i вещественные признаки. Проверка условий на узлах имеет вид:

$$\sum_{i=1}^{d} a_i x_i + a_{d+1} > 0. \tag{4}$$

Здесь $a_1; \ldots; a_{d+1}$ вещественные. Так как такие условия будут аналогичными тем же в обычных решающих деревьях при криволинейной замене координат, мы называем класс таких деревьев косыми.

Стохастическая модификация прежнего алгоритма

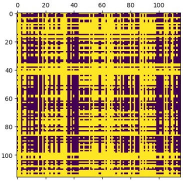
Заменим матрицы Z_j, элементы которых принимали бинарные значения, на P, векторы которой лежат в вероятностных симплексах.

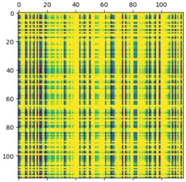
$$P_0 = \begin{pmatrix} P(x_1 \succeq x_1 | \mu_1 \succeq \mu_1) & P(x_1 \succeq x_2 | \mu_1 \succeq \mu_2) & \\ \ddots & P(x_i \succeq x_j | \mu_i \succeq \mu_j) & \cdots \\ & & \ddots \\ & & P(x_n \succeq x_n | \mu_n \succeq \mu_n) \end{pmatrix}$$

Оцениваем изначальное приближение матрицы по выборке(сэмплированием), а затем, меняя распределения бернуллиевских распределений компонент матриц, оптимизируем старый функционал, только уже по трём параметрам.

Сравнение методов решения задачи

Визуализируем матрицу попарных сравнений между оригинальными метками классов и матрицу, соответствующую полученным предсказаниям.





Сравнение методов на датасете футбольных игроков.

Algorithm	Learn error(MAE)	Test error(MAE)
Partial Orders	1.14 $\pm$0.05	1.69 $\pm$0.2
Isotonic Regression	0.98 $\pm$0.2	1.28 $\pm$0.4
Oblique Decision Trees	0.47 $\pm$0.5	1.06 $\pm$0.71
Stochastic Partial Orders	1.48 $\pm$0.14	1.32 $\pm$0.05

На защиту выносится

- Был разработан новый метод для решения задачи ранговой классификации.
- Проведён сравнительный анализ с другими методами, решающими данную задачу.
- Стохастическая модификация прежнего алгоритма показала большую устойчивость к шуму.