Нелинейные монотонные композиции классификаторов

Гуз И. С.
ivanguz@mail.ru
Москва, МФТИ

Объединение нескольких алгоритмов в композицию во многих случаях позволяет повысить качество классификации. Линейные и выпуклые композиции хорошо исследованы как теоретически, так и эмпирически [1, 2]. Нелинейные монотонные композиции были предложены в [3] и относительно мало исследованы. Цель данной работы — сравнить эффективность монотонных и линейных композиций, а также исследовать влияние структуры композиции и параметров метода настройки на обобщающую способность монотонной композиции.

Стандартная постановка задачи классификации на два класса

Пусть X — множество допустимых описаний объектов, $Y = \{-1, 1\}$ — множество меток классов, $y^*\colon X \to Y$ — неизвестная *целевая зависимость*, $X^\ell = (x_i)_{i=1}^\ell \subset X \times Y$ — обучающая выборка, $y_i = y^*(x_i)$. Требуется построить алгоритм классификации $a\colon X \to Y$, аппроксимирующий целевую зависимость $y^*(x)$ на всём множестве X.

Пусть фиксирован *метод обучения* μ, который по выборке X^ℓ с заданными весами объектов $W^\ell = (w_i)_{i=1}^\ell$ строит *алгоритмический оператор* $b\colon X \to \mathbb{R}$ путём минимизации функционала средней взвешенной ошибки $Q(b) = \sum_{i=1}^\ell w_i \bigl[a(x_i) \neq y_i\bigr]$, где $a(x) = \operatorname{sign} b(x)$ — алгоритм классификации, соответствующий оператору b.

Монотонные композиции классификаторов

Композицией алгоритмических операторов $b_1, \ldots, b_T$ называется алгоритм вида $a(x) = F\bigl(b_1(x), \ldots, b_T(x)\bigr)$, где отображение $F\colon \mathbb{R}^T \to Y$ называется *корректирующей операцией* [1].

Требование монотонности F как отображения из $\mathbb{R}^T$ в Y означает, что $F(b_1, \ldots, b_T)$ не должно уменьшаться при увеличении выходных значений операторов $b_1, \ldots, b_T$, Это вполне естественное требование, если предполагать, что операторы настраиваются на решение одной и той же задачи. Монотонные корректирующие операции образуют более широкое семейство функций по сравнению с выпуклыми (линейными с неотрицательными коэффициентами). Это позволяет точнее настраиваться на данные, но, возможно, повышает риск переобучения.

Пара объектов $x_j, x_k \in X^\ell$ называется *дефектной парой* набора алгоритмических операторов $b_1, \ldots, b_T$, если $y_j < y_k$ и $b_t(x_j) \geqslant b_t(x_k)$ для всех $t = 1, \ldots, T$. Число дефектных пар можно рассматривать как функционал качества композиции [4]. Операторы $b_1, \ldots, b_T$ строятся и добавляются в композицию последовательно. Перед настройкой очередного

Метод \ Задача	ionospere	house-vote	bupa	diabetes
Monotone (SVM)	9.7% (3)	**3.2**% (5)	**31.3**% (2)	23.6% (2)
Monotone (Parzen)	**8.0**% (2)	5.6% (5)	32.7% (3)	30.2% (2)
AdaBoost (SVM)	11.5% (65)	4.1% (40)	**30.7**% (15)	**22.7**% (15)
AdaBoost (Parzen)	12.0% (15)	6.0% (11)	33.0% (34)	29.0% (23)
SVM	13.1%	4.5%	42.2%	23.0%
Parzen	15.0%	6.2%	33.8%	30.7%

Таблица 1. Доля ошибок на контрольных данных, усредненная по 50 разбиениям. В скобках указано среднее число алгоритмов в композиции.

оператора b_t методом μ вес w_i каждого объекта x_i, $i = 1, \ldots, \ell$, устанавливается пропорционально числу дефектных пар набора операторов $b_1, \ldots, b_{t-1}$, в которых участвует данный объект. Кроме того, вводится параметр λ, управляющий стратегией настройки очередного оператора: при $\lambda = 0$ оператор b_t настраивается только на аппроксимацию обучающей выборки без учёта ранее построенных операторов $b_1, \ldots, b_{t-1}$; при $\lambda = 1$ оператор b_t настраивается только на компенсацию ошибок, допущенных композицией $F(b_1, \ldots, b_{t-1})$; при промежуточных значениях $\lambda \in (0, 1)$ реализуется компромиссная стратегия настройки.

Наряду с дефектными парами при расчёте весов w_i можно учитывать специальным образом определяемые дефектный тройки [4], что значительно увеличивает вычислительную сложность алгоритма. Поэтому в данной работе исследуется вопрос о целесообразности учета троек.

Эксперименты и результаты

Исследование монотонной коррекции проводились на четырёх реальных задачах из репозитория UCI. В качестве метода обучения μ использовались два метода: байесовский классификатор с локальным восстановление плотности по Парзену-Розенблатту и переменной шириной окна; и метод опорных векторов (SVM). Выпуклая композиция, с которой проводилось сравнение, строилась алгоритмом AdaBoost [5].

Для оценивания обобщающей способности использовался скользящий контроль: выборка разбивалась 50 раз случайным образом на обучающую (80%) и контрольную (20%). Оптимальное число T операторов в композиции определялось по минимуму средней доли ошибок на контрольной выборке. Результаты представлены в Таблице 1.

Оптимальное число операторов в монотонных композициях в большинстве случаев равно 2 или 3, что гораздо меньше, чем в линейных композициях. При дальнейшем увеличении T число дефектных пар быстро исчерпывается до нуля и возникает переобучение.

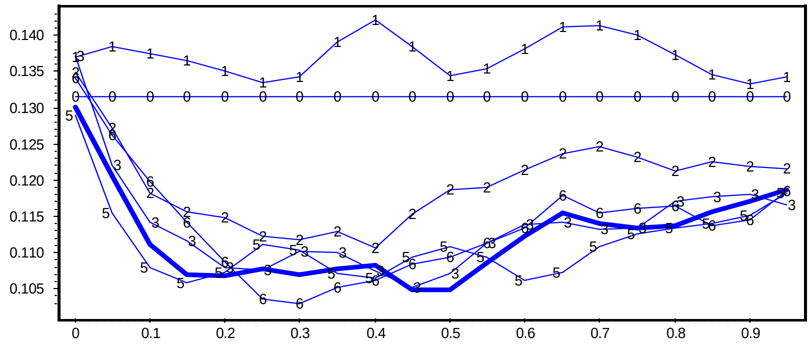

Рис. 1. Зависимость доли ошибок на контроле от параметра λ на примере монотонной коррекции SVM в задаче `ionosphere`. Цифрами 1–6 отмечено значение T. Выделена кривая $T = 4$, для которой положение минимума наиболее устойчиво относительно λ. Линия, помеченная цифрой 0, показывает уровень ошибок на контроле для отдельного SVM.

Параметр λ имеет смысл брать в пределах 0.1–0.4, Рис 1. Настройка базовых алгоритмов только лишь на устранение совокупного дефекта предыдущих алгоритмов ($\lambda = 1$) ведёт к переобучению.

При $T = 2$ учет дефектных троек немного улучшает обобщающую способность; при $T \geqslant 3$ только ухудшает. Поскольку подсчёт дефектных троек — трудоемкая операция, то от него вообще можно отказаться.

Таким образом, монотонные композиции, так же как и линейные, позволяют улучшать качество классификации, но при этом состоят из гораздо меньшего числа алгоритмов.

Работа выполнена при поддержке РФФИ, проект №05-01-00877, и программы ОМН РАН «Алгебраические и комбинаторные методы математической кибернетики».

Литература

[1] *Журавлёв Ю. И.* Об алгебраическом подходе к решению задач распознавания или классификации // Пробл. кибернетики. — 1978. — Т. 33. — С. 5–68.

[2] *Kuncheva L.* Combining pattern classifiers. — John Wiley & Sons, Inc., 2004.

[3] *Рудаков К. В., Воронцов К. В.* О методах оптимизации и монотонной коррекции в алгебраическом подходе к проблеме распознавания // Докл. РАН. — 1999. — Т. 367, № 3. — С. 314–317.

[4] *Воронцов К. В.* Оптимизационные методы линейной и монотонной коррекции в алгебраическом подходе к проблеме распознавания // ЖВМ и МФ. — 2000. — Т. 40, № 1. — С. 166–176.

[5] *Freund Y., Schapire R. E.* Experiments with a new boosting algorithm // International Conference on Machine Learning. — 1996. — Pp. 148–156.