Исследование стратегий обучения ранговой классификации по методу опорных векторов

Куракин А. В., Татарчук А. И., Моттль В. В.
alekseyvk@yandex.ru, aitech@yandex.ru, mottl@yandex.ru
Москва, МФТИ, ВЦ РАН

При решении практических задач распознавания образов типичной является ситуация, когда на конечном множестве значений $Y = \{0, \ldots, m\}$ некоторой целевой характеристики объектов $y(\omega) \in Y$, $\omega \in \Omega$, подлежащей восстановлению, задано отношение порядка. Задачу обучения с ранжированными классами (рангами) принято называть *задачей ранговой классификации* или *задачей восстановления ранговой регрессии*. Для двух рангов задача представляет собой классической задачу распознавания образов с двумя классами.

В основе большинства подходов к решению такой задачи лежит естественное предположение о модели ранговой зависимости, состоящее в том, что объективно существует скрытая характеристика объектов, выраженная некоторой действительнозначной функцией $f\colon \Omega \to \mathbb{R}$. Однако, в отличие от задачи восстановления регрессии, значение такой скрытой характеристики известно только через сравнение его с некоторым набором порогов $h_1 < \cdots < h_m$, разделяющих действительную ось на $m+1$ интервалов. Таким образом, для каждого объекта конечной обучающей совокупности $\omega \in \Omega^* = \{\omega_1, \ldots, \omega_N\} \subset \Omega$ известно лишь, в какой из промежутков $y(\omega) \in Y = \{0, \ldots, m\}$ на оси он попал. При этом все признаковое пространство $\mathbb{R}^n$, в котором объекты представлены векторами своих признаков $\mathbf{x}(\omega) \in \mathbb{R}^n, \omega \in \Omega$, разделяется на упорядоченные области набором параллельных гиперплоскостей с общим направляющим вектором $\mathbf{a} \in \mathbb{R}^n$:

$$\begin{cases} \qquad f(\mathbf{x}(\omega)) < h^{(1)} & \to \hat{y}(\omega) = 0; \\ h^{(1)} \leqslant f(\mathbf{x}(\omega)) < h^{(2)} & \to \hat{y}(\omega) = 1; \\ \ldots & \ldots \\ h^{(m)} \leqslant f(\mathbf{x}(\omega)) & \to \hat{y}(\omega) = m; \end{cases}$$

где $f(\mathbf{x}(\omega)) = \mathbf{a}^{\mathbf{T}}\mathbf{x}(\omega)$.

При такой интерпретации природы ранговости целевой характеристики, задача обучения сводится к поиску направляющего вектора $\mathbf{a} \in \mathbb{R}^n$ и набора числовых порогов $h_1 < \cdots < h_m \in \mathbb{R}$.

Основой метода опорных векторов [1], получившего широкое распространение при решении двухклассовой задачи распознавания, является концепция оптимальной разделяющей гиперплоскости, максимизирующей зазор между объектами двух классов. Однако в случае нескольких гиперплоскостей зазоров тоже несколько, и по этой причине возможны различные стратегии обучения [2, 3, 4].

В настоящее время существует два основных обобщения метода опорных векторов для решения задачи ранговой классификации.

Идея так называемой стратегии обучения с фиксированным зазором (Fixed Margin Strategy) состоит в максимизации зазора между объектами ближайших соседних рангов в обучающей совокупности. Альтернативная стратегия обучения с суммированием зазоров (Sum of Margins Strategy) основана на максимизации суммы зазоров между всеми рангами [2]. В данной работе рассматривается только первая стратегия обучения с фиксированным зазором, представляющая собой прямое обобщение метода опорных векторов.

В работе [2] был предложен оптимизационный критерий выбора направляющего вектора $\mathbf{a} \in \mathbb{R}^n$ и набора порогов $h_1, \ldots, h_m \in \mathbb{R}$, которые в соответствующем признаковом пространстве дают наилучшее разделение объектов $\omega \in \Omega$ только двух смежных рангов $y(\omega) = i - 1$ и $y(\omega) = i$, относительно каждой границы $h^{(i)}$. Такая постановка приводит к задаче квадратичного программирования с числом переменных, немногим меньшим удвоенного количества объектов обучающей совокупности $2|\Omega^*|$.

Предложенный подход является привлекательным с вычислительной точки зрения, однако, хотя и в относительно редких случаях, может приводить к некорректной упорядоченности оптимальных величин порогов $\hat{h}^{(i-1)} \geqslant \hat{h}^{(i)}$. Такую постановку будем называть сокращенной стратегией обучения с фиксированным зазором (Truncated Fixed Margin Strategy или TFM).

В работе [3] был предложен другой подход, основанный на идее оптимального разделения каждой границей $h^{(i)}$ объектов из обучающей совокупности, принадлежащих всем низшим рангам $y(\omega) \leqslant i - 1$, от объектов всех высших рангов $y(\omega) \geqslant i$. Такой подход автоматически гарантирует корректность упорядоченности оптимальных величин порогов $\hat{h}^{(0)} < \cdots < \hat{h}^m$. Будем называть такую постановку полной стратегией обучения с фиксированным зазором (Full Fixed Margin Strategy или FFM).

Ценой гарантированной корректности получаемого решения в полной постановке задачи обучения является необходимость решать задачу квадратичного программирования с числом переменных, равным произведению числа объектов и количества искомых границ $m|\Omega^*|$, что в $(m/2)^3$ раз превышает вычислительную сложность сокращенной постановки.

Следует заметить, что в простейшем случае двух рангов $m = 1$ обе рассматриваемые стратегии обучения вырождаются в классическую постановку задачи обучения распознавания с двумя классами по методу опорных векторов.

В работе [5] определены условия, при которых решение оптимизационной задачи квадратичного программирования сокращенной стратегии обучения в точности совпадает с решением соответствующей оптимизационной задачи полной стратеги обучения.

Предложен новый подход к обучению ранговой классификации, основанный на последовательном применении вначале сокращенной стратегии обучения, затем проверке корректности полученного результата и, в заключении, применении полной стратегии обучения, но только в случае, если условия корректности решения для сокращенной стратегии не выполняются.

Другим отличием рассмотренных обобщений метода опорных векторов от его классической постановки состоит в том, что по решениям соответствующих оптимизационных задач квадратичного программирования удается найти оптимальные величины только для одного активного порога, соответствующего минимальному зазору между рангами. Для нахождения оптимальных значений остальных порогов необходимо решать дополнительную задачу линейного программирования. В данной работе указаны способы явного вычисления оптимальных значений порогов разделяющих гиперплоскостей по решению соответствующих двойственных оптимизационных задач полной и сокращенной стратегий обучения ранговой классификации.

Работа выполнена при поддержке РФФИ, проекты № 05-01-00679, № 06-01-08042, № 06-07-89249, а также INTAS, проект № 04-77-7347.

Литература

[1] *Vapnik V.* Statistical Learning Theory. — New York: John-Wiley & Sons, Inc., 1998. — 732 c.

[2] *Shashua A., Levin A.* Ranking with large margin principle: two approaches // Advances in Neural Information Processing Systems, 2003. — С. 937–944.

[3] *Chu W., Keerthi S. S.* New approaches to support vector ordinal regression // 22nd Int. Conf. on Machine Learning, Bonn, Germany, 2005. — Pp. 145–152.

[4] *Waegeman W., Boullart L.* An ensemble of weighted support vector machines for ordinal regression // Transactions on Engineering, Computing and Technology, 2006. — Vol. 12. — Pp. 71–75.

[5] *Mottl V., Tatarchuk A., Kurakin A.* Support vector machines for ranking learning: the Full and the Truncated fixed margin strategies // Int. Conf. on Machine Learning and Cybernetics, Hong Kong, China, August 19-22, 2007.