Синтез и анализ гибридных алгоритмов распознавания образов

Лапко А. В., Лапко В. А.

lapko@icm.krasn.ru

Красноярск, Институт вычислительного моделирования СО РАН

Совместное использование в едином решающем правиле разнотипных моделей является перспективным способом наиболее полного учёта априорной информации. В данном направлении получены успешные результаты исследований, к которым можно отнести методы локальной аппроксимации [1], гибридные модели стохастических зависимостей [2], полупараметрические и частично линейные модели [3], непараметрические коллективы решающих правил [4]. При этом особое внимание уделяется алгоритмам восстановления стохастических зависимостей, обеспечивающих учет частичных сведений об их виде и данных обучающих выборок.

Для решения проблемы эффективного использования априорной информации предлагаются гибридные системы распознавания образов, которые обеспечивают сочетание в обобщенном решающем правиле классификации преимущества параметрических и локальных методов аппроксимации, основанных на оценках плотности вероятности типа Розенблатта-Парзена [5].

Синтез гибридных алгоритмов распознавания образов

Пусть исходную информацию при решении двухальтернативной задачи распознавания образов составляют обучающая выборка $V = \big(x^i, \sigma(x^i),\ i = 1,\dots,n\big)$ и априорные сведения $F_{12}(x,\alpha)$ о виде уравнения разделяющей поверхности $f_{12}(x)$ между классами $\Omega_1,\ \Omega_2$ в пространстве $x \in R^k$.

Информация обучающей выборки V формируется на основании данных о значениях признаков x классифицируемых объектов и соответствующим им «указаний учителя» $\sigma(x) = -1$, если $x \in \Omega_1$; 1, если $x \in \Omega_2$.

Для использования в полном объёме априорной информации $(F_{12}(x,\alpha), V)$ воспользуемся принципами гибридного моделирования.

Для этого определим параметры α уравнения разделяющей поверхности $F_{12}(x,\alpha)$ из условия минимума эмпирической ошибки распознавания образов.

По результатам вычислительного эксперимента сформируем выборку расхождений $V_1 = \big(x^i, q(x^i),\ i = 1,\dots,n\big)$ между «решениями» $\bar\sigma(x^i)$ алгоритма классификации, использующего уравнение разделяющей поверхности $F_{12}(x,\bar\alpha)$, и «указаниями учителя» $\sigma(x^i)$ из обучающей выборки V. При этом значения функции расхождений

$$q(x^i) = \begin{cases} 0, & \sigma(x^i) = \bar{\sigma}(x^i); \\ F_{12}(x^i, \bar{\alpha}) + \Delta, & \bar{\sigma}(x^i) = -1 \text{ и } \sigma(x^i) = 1; \\ -(F_{12}(x^i, \bar{\alpha}) + \Delta), & \bar{\sigma}(x^i) = 1 \text{ и } \sigma(x^i) = -1. \end{cases}$$

При наличии ошибки функция расхождения принимает значение, обратное по знаку уравнения разделяющей поверхности $F_{12}(x, \bar{\alpha})$ и превышает его на величину параметра Δ.

Восстановим функцию $q\,(x)$ по выборке V_1 на основе непараметрической регрессии [6]

$$\bar{q}(x) = \frac{\sum\limits_{i=1}^{n} q(x^i)\beta_i(x)}{\sum\limits_{i=1}^{n} \beta_i(x)}, \quad \beta_i(x) = \prod_{\nu=1}^{k} \Phi\left(\frac{x_\nu - x_\nu^i}{c_\nu}\right),$$

где $\Phi(\cdot)$ — ядерная функция, удовлетворяющая условиям положительности, симметричности и нормированности.

Тогда гибридный алгоритм классификации запишется в виде

$$\bar{m}_{12}(x) : \begin{cases} x \in \Omega_1 & \text{если } \bar{f}_{12}(x) \leqslant 0, \\ x \in \Omega_2 & \text{если } \bar{f}_{12}(x) > 0, \end{cases}$$

$$\bar{f}_{12}(x) = F_{12}(x, \bar{\alpha}) + \bar{q}(x).$$

Меняя вид функции $q(x)$, обеспечивающей коррекцию $F_{12}(x, \bar{\alpha})$, получено семейство новых гибридных решающих правил.

Асимптотические свойства гибридных решающих функций

Рассмотрены асимптотические свойства гибридных моделей уравнений разделяющих поверхностей $\bar{f}_{12}(x)$. В частности, при достаточно большом объёме обучающей выборки, смещение

$$\lim_{n\to\infty} \mathsf{E}\left(f_{12}(x) - \bar{f}_{12}(x)\right) = \lim_{n\to\infty} \left[\frac{\left(q(x)p(x)\right)''}{2p(x)} c^2 + \Delta(n) + 0(c^4)\right] = 0,$$

при $\Delta(n) \to 0$ и $c(n) \to 0$. Здесь E — знак математического ожидания; $\left(q(x)p(x)\right)''$ — вторая производная по x произведения $q(x)p(x)$.

Асимптотическое выражение среднеквадратического отклонения имеет вид

$$\mathsf{E}\left(f_{12}(x) - \bar{f}_{12}(x)\right)^2 \sim \left(ncp(x)\right)^{-1} q^2(x) \left\|\Phi\,(u)\right\|^2 + \\ + c^4 \left(\left(q(x)p(x)\right)^{(2)}\right)^2 \left(4p^2(x)\right)^{-1} + \frac{\left(q(x)p(x)\right)''}{p(x)}\; c^2 \Delta\,(n) + \; \Delta^2\,(n)\,,$$

где $\left\|\Phi(u)\right\|^2 = \int \Phi^2(u)\,du$.

Отсюда следует, что $\bar{f}_{12}(x)$ сходится в среднеквадратическом, если $\Delta(n) \to 0$, $c(n) \to 0$ и $nc \to \infty$ при $n \to \infty$.

Асимптотическая несмещённость и сходимость в среднеквадратическом $\bar{f}_{12}(x)$ определяют свойство её состоятельности.

Установлена зависимость скоростей сходимости $\bar{f}_{12}(x)$ от объёма обучающей выборки, вида корректирующей функции и её параметров. На этой основе, исследуя отношение среднеквадратических отклонений модификаций гибридных решающих функций, установлены условия их эффективного использования.

Работа выполнена при поддержке фонда «Научный потенциал».

Литература

[1] *Катковник В. Я.* Линейные и нелинейные методы непараметрического регрессионного анализа // Автоматика, 1979. — № 5. — С. 165–170.

[2] *Лапко А. В.* Имитационные модели неопределённых систем. — Новосибирск: Наука, 1993.

[3] *Хардле В.* Прикладная непараметрическая регрессия. — М.: Мир, 1993.

[4] *Лапко В. А.* Непараметрические коллективы решающих правил. — Новосибирск: Наука, 2002.

[5] *Parzen E.* On the estimation of probability density function and mode // Ann. Math. Stat., 1962. — Vol. 33. — P. 1065–1076.

[6] *Надарая Э. А.* Непараметрические оценки кривой регрессии // Тр. ВЦ АН ГССР, 1965. — Вып. 5. — С. 568.