Комбинирование потенциальных функций в многомодальном распознавании образов

Моттль В. В., Татарчук А. И., Красоткина О. В., Сулимова В. В.

mottl@yandex.ru, aitech@yandex.ru, krasotkina@yandex.ru, vsulimova@yandex.ru

Москва, Вычислительный центр РАН, Тула, ТулГУ

Естественно, что физический объект не может быть непосредственно воспринят компьютером, поэтому в качестве посредника между объектами реального мира $\omega \in \Omega$ и процедурой анализа данных, в частности, распознавания образов, всегда выступает тот или иной конструктивный способ выражения доступной информации, который принято называть *модальностью* представления объектов. Например, набор биометрических характеристик, используемых при идентификации личности [1], как правило включает в себя фотопортрет, отпечатки пальцев, подпись, и другие. Специфика в анализе данных социологических опросов населения [2] состоит в том, что интересующие исследователя свойства людей как элементов популяции выражаются специальной формулировкой вопросов в анкете, каждый из которых определяет множество допустимых ответов.

В терминах выбранной модальности каждый объект $\omega \in \Omega$ отображается в пространство значений соответствующего *обобщенного признака* $x(\omega)\colon \Omega \to \mathbb{X}$, например, в виде сигнала, изображения, а в сравнительно простых случаях в виде действительного числа или вектора.

Невозможность обеспечить требуемое качество классификации объектов на основе какой-либо одной модальности привела к огромному разнообразию используемых способов их представления и к появлению концепции *многомодальных систем*, комбинирующих сразу все доступные представления объектов $(x_i(\omega) \in \mathbb{X}_i, i = 1, \ldots, n)$ в единой процедуре распознавания $\hat{y}\big(x_1(\omega), \ldots, x_n(\omega)\big)\colon \mathbb{X}_1 \times \cdots \times \mathbb{X}_n \to Y = \{1, \ldots, m\}$.

В таких ситуациях ключевым моментом является *уровень комбинирования модальностей*, на котором происходит слияние доступной информации перед принятием итоговых суждений о классах объектов. Наиболее естественной представляется идея формирования единого представления объектов непосредственно на *уровне сенсоров*, формирующих исходные представления объектов. Однако до последнего времени считалось, что комбинировать выходные сигналы сенсоров, часто имеющие различную физическую природу, проблематично, либо не представляется возможным в принципе, поэтому основное внимание уделялось комбинированию разнородной информации на *уровне классификаторов*, построенных независимо по каждой модальности.

Развитие беспризнаковой методологии распознавания образов [3], основанной на понятии *потенциальной функции*, позволило унифицировать представление объектов в виде элементов линейного пространства, а появление методов комбинирования нескольких разных потенциальных функций [4, 5, 6] дало возможность комбинировать модальности представления объектов фактически на уровне сигналов сенсоров.

В то же время по-прежнему остается открытым вопрос о соотношении комбинирования модальностей на уровне сенсоров и на уровне классификаторов. В работе [7] проведено исследование условий, при которых один из известных принципов комбинирования классификаторов получается как частный случай комбинирования потенциальных функций.

Специфика потенциальной функции $K(x', x'')\colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$, заданной на множестве значений некоторого обобщенного признака $x(\omega) \in \mathbb{X}$, заключается в том, что ее значения могут интерпретироваться непосредственно как скалярное произведение объектов $x' = x(\omega')$, $x'' = x(\omega'')$ в воображаемом линейном пространстве $\tilde{\mathbb{X}} \supseteq \mathbb{X}$, в которое данная потенциальная функция погружает шкалу значений выходного сигнала соответствующего сенсора, минуя промежуточное понятие вектора числовых признаков.

Задав, как минимум, по одной потенциальной функции для каждой модальности $K_i(x'_i, x''_i)$, $x'_i, x''_i \in \mathbb{X}_i$, $i = 1, \ldots, n$, удобно рассматривать порождаемые этими функциями линейные пространства совместно как декартово произведение $\tilde{\mathbb{X}} = \tilde{\mathbb{X}}_1 \times \cdots \times \tilde{\mathbb{X}}_n$ с комбинированной потенциальной функцией $K(\mathbf{x}', \mathbf{x}'')$, $\mathbf{x}' = (x'_1, \ldots, x'_n)$, $\mathbf{x}'' \in \tilde{\mathbb{X}}$ в роли скалярного произведения. При таком подходе основная нагрузка по интерпретации исходных представлений объектов перекладывается на этап формирования потенциальных функций, но в данной работе этот аспект не рассматривается.

Большинство методов комбинирования потенциальных функций основано на формировании результирующей потенциальной функции как линейной комбинации исходных функций $K(\mathbf{x}', \mathbf{x}'') = \sum_{i=1}^{n} \alpha_i K(x'_i, x''_i)$ с коэффициентами $\alpha_i \geqslant 0$. Очевидно, что если некоторый коэффициент равен нулю, то соответствующая модальность не будет участвовать в принятии решений. Такой подход означает замену дискретной задачи перебора подмножеств модальностей с целью выбора подмножества, наиболее адекватного анализируемому массиву данных $\hat{I} \subseteq I = \{1, \ldots, n\}$, на непрерывную задачу поиска весов $\hat{I} = \{i \in I\colon \alpha_i > 0\}$. Специфика конкретного метода комбинирования заключена в выборе критерия оптимизации.

В работе [4] был предложен *метод опорных потенциальных функций* (Support Kernel Machines или SKM), особенность которого состоит в стремлении процедуры комбинирования выбирать нулевые веса $\alpha_i = 0$ для большинства функций и оставлять ненулевыми $\alpha_i > 0$ только веса для активных (опорных) функций. Однако, такая стратегия обучения приводит к двойственной оптимизационной задаче квадратичного программирования при квадратичных ограничениях, решение которой вычислительно существенно сложнее стандартной задачи квадратичного программирования при линейных ограничениях, лежащей в основе классического метода опорных векторов [3].

Другая стратегия комбинирования [5] приводит к итерационной процедуре вычисления весов $(\alpha_1, \ldots, \alpha_n)$, но в этом случае веса неинформативных потенциальных функций лишь стремятся к нулю, но никогда его не достигают. Будем называть такую стратегию комбинирования *методом релевантных потенциальных функций* (Relevance Kernel Machines или RKM) по аналогии с методом релевантных векторов, предложенным в работе [8] для поиска линейных дискриминантных функций как альтернатива классическому методу опорных векторов.

Вообще говоря, методология обучения распознаванию объектов двух классов на основе концепции оптимальной линейной дискриминантной функции не предусматривает принятия каких-либо предположений о вероятностной модели генеральной совокупности. Такой концепции адекватна качественная модель в виде в линейного пространства признаков, в котором объективно существует некоторая гиперплоскость, такая, что объекты двух разных классов отображаются, в основном, по разные стороны от нее, без количественного уточнения, в какой степени это предположение может нарушаться. При этом выбор значений направляющего элемента $\hat{\vartheta} \in \tilde{\mathbb{X}}$ и порога $\hat{b} \in \mathbb{R}$ линейной дискриминантной функции

$$f(\mathbf{x}(\omega) \,|\, \vartheta, b) = K(\vartheta, \mathbf{x}(\omega)) + b \begin{cases} < 0 \to y(\mathbf{x}(\omega)) = -1 \\ > 0 \to y(\mathbf{x}(\omega)) = 1 \end{cases}$$

полностью задает некоторую классификацию множества объектов $\omega \in \Omega$ в комбинированном линейном пространстве обобщенных признаков $\tilde{\mathbb{X}} = \tilde{\mathbb{X}}_1 \times \cdots \times \tilde{\mathbb{X}}_n$ со скалярным произведением $K(\mathbf{x}', \mathbf{x}'')$, $\mathbf{x}', \mathbf{x}'' \in \tilde{\mathbb{X}}$.

В данной работе предлагается *квази-статистический подход* к решению задачи комбинирования потенциальных функций, который полностью охватывает методы опорных и релевантных потенциальных функций. В качестве модели генеральной совокупности объектов предлагается рассматривать два параметрических семейства плотностей распределения $\varphi_{-1}(\mathbf{x}(\omega) \,|\, \vartheta, b)$ и $\varphi_1(\mathbf{x}(\omega) \,|\, \vartheta, b)$, связанных с двумя классами объектов и сконцентрированных преимущественно по разные стороны гиперплоскости $f(\mathbf{x}(\omega) \,|\, \vartheta, b) = 0$. Однако невозможно выбрать эти плотности распределения так, чтобы они количественно отражали возможность попадания объектов за пределы своего класса и не вносили бы при этом иной информации о значениях признака, поскольку такое требование связано с предположением о равномерном распределении в бесконечных областях пространства, что приводит к равенству нулю соответствующей плотности. В книге [9] рекомендуется задавать «безразличное» априорное распределение в виде ненулевой функции-константы, несмотря на то, что ее интеграл по бесконечной области не существует. Функции такого вида были названы *несобственными плотностями распределениями*.

В основу предлагаемого подхода положен принцип максимизации плотности апостериорного распределения в пространстве параметров модели генеральной совокупности, который приводит к стандартному байесовскому правилу обучения.

Предположение о нормальном распределении направляющего вектора $\vartheta = (\vartheta_1, \ldots, \vartheta_n) \in \tilde{\mathbb{X}}$ с независимыми компонентами немедленно приводит к критерию обучения, полностью идентичному критерию по методу опорных потенциальных функций. В то же время, принятие в качестве априорного распределения компонент направляющего вектора распределения Лапласа приводит к методу релевантных потенциальных функций.

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

Литература

[1] *Ross A., Jain A.* Multimodal biometrics: An overview // 12th European Signal Processing Conference, Vienna, Austria, 2004. — С. 1221.

[2] *Галицкий Е. Б., Моттль В. В., Татарчук А. И.* Обучение распознаванию образов в анализе данных опросов населения. // ММРО-12, Москва, 2005.

[3] *Vapnik V.* Statistical Learning Theory. John-Wiley & Sons, Inc. 1998.

[4] *Bach F. R., Lankriet G. R. G., Jordan M. I.* Multiple kernel learning, conic duality, and the SMO algorithm // 21th International Conference on Machine Learning, Banff, Canada, 2004.

[5] *Sonnenburg S., Rätsch G., Schäfer C.* A general and efficient multiple kernel learning algorithm // 19th Annual Conference on Neural Information Processing Systems, Vancouver, Canada, 2005.

[6] *Mottl V., Krasotkina O., Seredin O., Muchnik I.* Kernel fusion and feature selection in machine learning // 8th IASTED International Conference on Intelligent Systems and Control, Cambridge, USA, 2005.

[7] *Татарчук А. И., Елисеев А. П., Моттль В. В.* Комбинирование классификаторов и потенциальных функций в многомодальном распознавании образов // ММРО-13 (в настоящем сборнике). — 2007. — С. **??**–**??**.

[8] *Bishop C. M., Tipping M. E.* Variational relevance vector machines // 16th Conf. on Uncertainty in Artificial Intelligence, Morgan Kaufmann, 2000. — Pp. 46–53.

[9] *Де Гроот М.* Оптимальные статистические решения. — Москва: Мир, 1974.