

Комбинирование классификаторов и потенциальных функций в многомодальном распознавании образов

Татарчук А. И., Елисеев А. П., Моттль В. В.

aitech@yandex.ru, andrevel@gmail.com, mottl@yandex.ru

Москва, МФТИ, ВЦ РАН

Под многомодальным распознаванием образов понимается комплекс задач и методов их решения, связанных с необходимостью принимать суждения о классах объектов, характеризующихся одновременно несколькими видами представлений (модальностей), часто имеющих разную физическую природу. В данной работе проведено теоретическое и экспериментальное исследование принципов многомодального обучения распознавания образов в рамках метода потенциальных функций при разных уровнях комбинирования модальностей.

Уровни комбинирования модальностей

Способ математического выражения информации об объектах в задачах анализа данных принято называть *модальностью* представления объектов. В терминах выбранной модальности каждый объект реального мира $\omega \in \Omega$ характеризуется некоторым значением в пространстве соответствующего обобщенного признака $x(\omega) \in \mathbb{X}$, например, в виде сигнала, изображения, а, в сравнительно простых ситуациях, в виде действительного числа или вектора.

В задачах обучения распознаванию образов по прецедентам такое преобразование приводит к обучающей совокупности, составленной из значений заданного обобщенного признака $x_j = x(\omega_j) \in \mathbb{X}$ и индексов классов $y_j = y(\omega_j) \in Y = \{1, \dots, m\}$ для доступного подмножества объектов $\Omega^* = \{\omega_j, j = 1, \dots, N\} \subset \Omega$. Анализируя предъявленный массив данных, требуется продолжить функцию $y(x(\omega_j))$, заданную в пределах обучающей совокупности $\omega_j \in \Omega^*$, на все множество объектов $\hat{y}(x(\omega)): \mathbb{X} \rightarrow Y$.

Обоснованность предположения, что в результате анализа обучающей совокупности $\{(x_j, y_j), j = 1, \dots, N\}$ удастся построить решающее правило $\hat{y}(x(\omega)): \mathbb{X} \rightarrow Y$, хорошо аппроксимирующее фактическое разбиение генеральной совокупности объектов $\omega \in \Omega$ на множество классов $y(\omega): \Omega \rightarrow Y$, зависит главным образом от того, насколько удачно выбрана модальность представления объектов $x(\omega) \in \mathbb{X}$.

Во многих практических случаях ни одна отдельно взятая модальность не обеспечивает достаточной надежности распознавания, а необходимость повышения обобщающей способности решающего правила приводит к концепции *многомодальных систем*, комбинирующих несколько разных способов представления объектов $(x_i(\omega) \in \mathbb{X}_i, i = 1, \dots, n)$ в единой процедуре распознавания образов $\hat{y}(x_1(\omega), \dots, x_n(\omega))$.

Различают два уровня комбинирования модальностей [1] — *уровень сенсоров* [2], когда до обучения классификатора формируется единое представление объектов, комбинирующее все частные модальности, и *уровень классификаторов* [3], когда комбинируются классификаторы, обученные по каждой модальности в отдельности.

До последнего времени основное внимание в литературе уделялось принципам комбинирования классификаторов, поскольку считалось, что комбинировать модальности разной физической природы затруднительно, либо в принципе невозможно.

Однако развитие метода потенциальных функций, основанного на идее единообразного представления объектов любой природы в виде элементов линейного пространства, и появление методов комбинирования нескольких разных потенциальных функций [2] открывает путь к комбинированию модальностей фактически на уровне сигналов сенсоров.

Обучение по единственной модальности

В рамках метода потенциальных функций [4] для построения двух-классового классификатора по единственной модальности $x_i(\omega)$ необходимо выбрать потенциальную функцию $K(x_i(\omega'), x_i(\omega''))$, $\omega', \omega'' \in \Omega_i^*$, адекватную искомой классификации, определить направляющий элемент $\vartheta \in \mathbb{X}_i$ и сдвиг $b \in \mathbb{R}$ линейной дискриминантной функции:

$$f_i(x_i(\omega)) = K(\vartheta, x_i(\omega)) + b \begin{cases} < 0 & \rightarrow \hat{y}(x_i(\omega)) = -1; \\ > 0 & \rightarrow \hat{y}(x_i(\omega)) = 1. \end{cases}$$

Если $f_i(x_i(\omega)) = 0$, то, строго говоря, решение о принадлежности объекта $\omega \in \Omega$ к одному из классов не может быть принято. Такие точки пространства $\mathbb{X}_{\emptyset, i} = \{x_i(\omega) \in \mathbb{X}_i \mid f_i(x_i(\omega)) = 0\}$ будем называть *нейтральными точками* [5] обобщенного признака \mathbb{X}_i .

Обучение по непересекающимся обучающим совокупностям: Метод нейтральной точки

При построении мультимодальных систем типична ситуация, когда разные модальности разрабатываются разными группами специалистов независимо друг от друга. В этом случае общая обучающая совокупность оказывается состоящей из непересекающихся подмножеств объектов $\Omega^* = \bigcup_{i=1}^m \Omega_i^*$, $\Omega_i^* \cap \Omega_j^* = \emptyset$, $i \neq j$, в пределах каждого из которых известны значения только одного обобщенного признака $x_i(\omega_j)$, $\omega_j \in \Omega_i^*$, следовательно, не представляется возможным напрямую строить решающее правило распознавания по методу комбинирования потенциальных функций [2].

В работе [5] предлагается рассматривать задачу обучения по непересекающимся совокупностям как задачу обучения по неполным данным.

Для объектов каждой совокупности $\omega_j \in \Omega_i^*$ неизвестные фактические значения их признаков по каждой из других модальностей $x_l(\omega_j)$, $l \neq i$ заменяются одним общим значением $\hat{x}_{\phi,l} \in \mathbb{X}_{\phi,l}$, которое представляет собой нейтральную точку соответствующего обобщенного признака.

При таком способе восполнения недостающих значений обобщенных признаков метод комбинирования потенциальных функций [2] сводится к методу комбинирования классификаторов, известный в англоязычной литературе как Sum Rule of Classifier Fusion.

Экспериментальное сравнение комбинирования классификаторов и потенциальных функций

В работе [6] показано, что при независимых модальностях, имеющих примерно равную информативность, комбинирование классификаторов по методу Sum Rule предпочтительнее по ошибке на генеральной совокупности в сравнении с комбинированием потенциальных функций.

В случае независимых модальностей с существенно разной информативностью, а так же зависимых модальностей комбинирование потенциальных функций дает лучшие результаты на генеральной совокупности.

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

Литература

- [1] *Ross A., Jain A.* Multimodal biometrics: An overview // 12th European Signal Processing Conference (EUSIPCO), Vienna, Austria, 2004. — С. 1221–1224.
- [2] *Mottl V., Tatarchuk A., Seredin O., Krasotkina O., Sulimova V.* Combining pattern recognition modalities at the sensor level via kernel fusion // 7th International Workshop on Multiple Classifier Systems, Prague, Czech Republic, 2007. — С. 1–12.
- [3] *Kittler J., Hatef M., Duin R., Matas J.* On combining classifiers. // IEEE Trans. on Patt. Anal. and Mach. Intelligence. — 1998. — Т. 20, No 3. — С. 226–239.
- [4] *Vapnik V.* Statistical Learning Theory. John-Wiley & Sons, Inc. 1998.
- [5] *Windridge D., Mottl V., Tatarchuk A., Eliseyev A.* The neutral point method for kernel-based combining disjoint training data in multi-modal pattern recognition // 7th International Workshop on Multiple Classifier Systems, Prague, Czech Republic, 2007. — С. 13–21.
- [6] *Windridge D., Mottl V., Tatarchuk A., Eliseyev A.* The relationship between kernel and classifier fusion in kernel-based multi-modal pattern recognition: An experimental study // Int. Conf. on Machine Learning and Cybernetics, August 19-22, 2007, Hong Kong, China.