

Алгоритм распознавания, основанный на построении метрических закономерностей*Дедовец М. С., Сенько О. В.*

senkoov@ccas.ru

Москва, ВЦ РАН, ВМиК МГУ

Ранее были предложены связанные с решением задач распознавания средства анализа данных, основанные на поиске закономерностей. При этом под закономерностью понимается подобласть признакового пространства, содержащая объекты одного класса (полная закономерность) или преимущественно одного класса (частичная закономерность). При этом геометрическая форма подобласти определяется заранее заданной моделью. Так, наиболее часто используемыми моделями являются модели логических закономерностей [1, 2, 3], в которых искомые области признакового пространства имеют форму гиперпараллелепипедов. Рассматривались также закономерности, формируемые с помощью линейных границ с произвольной ориентацией относительно координатных осей [4]. Несомненно, что априорное задание геометрической формы является существенным ограничением, затрудняющим выявление закономерностей, реально существующих в данных, но не удовлетворяющих сделанным предположениям. В связи с этим был предложен новый тип закономерностей, который далее будет называться метрическим. Под метрической закономерностью понимается подобласть признакового пространства, задаваемая как окрестность некоторого набора точек обучающей выборки, принадлежащих к одному классу K с минимальным включением объектов, не принадлежащих K .

При задании закономерностей используется понятие G -смежности между объектами одного из классов. Предположим, что у нас задана некоторая метрика ρ . Два объекта S_1 и S_2 из пересечения класса K и обучающей выборки назовём G -смежными, если в обучающей выборке не существует такого объекта S , не принадлежащего классу K , что одновременно выполняются два неравенства: $\rho(S, S_1) \leq \rho(S_1, S_2)$ и $\rho(S, S_2) \leq \rho(S_1, S_2)$. Смысл отношения смежности между объектами одного класса состоит в том, что оно обеспечивает отсутствие между ними объектов других классов. Классу K может быть сопоставлен граф, вершинами которого являются объекты класса. Двум вершинам ставится в соответствие ребро, если соответствующие объекты являются G -смежными. Под метрической закономерностью в настоящем исследовании нами понималась окрестность компоненты связности графа G -смежности. Объект S принадлежит метрической закономерности, если существует хотя бы одна пара объектов S_1 и S_2 таких, что: $\rho(S, S_1) \leq \rho(S_1, S_2)$ и $\rho(S, S_2) \leq \rho(S_1, S_2)$.

Был разработан алгоритм распознавания, основанный на голосовании по представительным системам метрических закономерностей. При этом закономерности строились в пространствах, задаваемых парами или тройками признаков. В систему включались только те закономерности, которые содержали не менее $0.25 m_K$ объектов, где m_K — число объектов класса K в обучающей выборке. Для вычисления оценки объекта S за класс K использовалась формула $\gamma(S, K) = \sum_{S_i \in K} V(i)/V_0(i)$, где суммирование ведётся по всевозможным объектам обучающей выборки из класса K , $V_0(i)$ — число метрических закономерностей, в которые вошёл объект S_i , $V(i)$ — число метрических закономерностей, в которые одновременно вошли объекты S и S_i .

Оценка эффективности алгоритма проводилась на совокупности реальных прикладных задач. Проведённые исследования показали достаточно высокую точность распознавания на контрольной информации, сравнимую с точностью, даваемой альтернативными подходами — методом статистически взвешенных синдромов и методом опорных векторов. Можно сделать вывод о перспективности модели и необходимости дальнейших исследований в данном направлении.

Работа выполнена при поддержке РФФИ, проекты № 06-01-00492, № 06-01-08045.

Литература

- [1] Журавлёв Ю. И., Рязанов В. В., Сенько О. В. Распознавание. Математические методы. Программная система. Применения. — Москва: Фазис, 2006.
- [2] Богомолов В. П., Виноградов А. П., Журавлёв Ю. И., Катериночкина Н. Н., Ларин С. Б., Рязанов В. В., Сенько О. В. Программная система распознавания ЛОРЕГ: алгоритмы распознавания, основанные на голосовании по системам логических закономерностей. — М.: ВЦ РАН, 1998. — 63 с.
- [3] Кузнецов В. А., Сенько О. В., Кузнецова А. В. и др. Распознавание нечетких систем по методу статистически взвешенных синдромов и его применение для иммуногематологической нормы и хронической патологии // Химическая физика, 1996. — Т. 15, № 1. — С. 81–100.
- [4] Dokukin A. A., Senko O. V. About new pattern recognition method for the universal program system Recognition. Proc. of the Int. Conf. I.Tech-2004, Varna (Bulgaria), 14–24 June 2004. — Pp. 54–58.
- [5] Сенько О. В., Кузнецова А. В. Алгоритмы распознавания, основанные на голосовании по системам закономерностей различных типов // Докл. всеросс. конф. ММРО-12. — Москва, 2005, — С. 200–203.