

Решение задач распознавания с невыполненной гипотезой компактности

Гуров С. И., Потепалов Д. Н., Фатхутдинов И. Н.

sgur@cs.msu.ru, mmp@cs.msu.ru

Москва, МГУ им. М. В. Ломоносова, факультет ВМиК

Реализованные на сегодняшний день алгоритмы логического синтеза имеют разную эффективность на различных типах схем. Поэтому возникает задача определения наилучшего алгоритма синтеза исходя из тех или иных характеристик входного описания схемы.

В первой решённой задаче зафиксирован набор синтезирующих алгоритмов и выбран критерий качества синтеза. Для совокупности описаний схем (прецедентов) выясняется распределение их по классам: принадлежность данному классу означает, что синтезированная соответствующим алгоритмом схема имеет наилучшие характеристики по указанному критерию. Целью исследования являлась разработка и реализация метода определения областей компетентности алгоритмов, позволяющего для каждого нового описания схемы указать наилучший (в данном смысле) алгоритм её синтеза.

Поставленная задача решалась методами распознавания образов. Суть предложенного решения заключается в построении признакового пространства образов (описаний схем) с последующим решением в нём задачи классификации. Первый, почти не разработанный в теоретическом плане и трудоёмкий на практике этап состоит в: (1) фиксации набора первичных числовых признаков прецедентов; (2) построении достаточно представительной совокупности вторичных признаков как функций от первичных; (3) отборе наиболее информативных вторичных признаков. На втором этапе строится решающее правило, относящее описание схемы в построенном признаковом пространстве к тому или иному классу.

Во второй задаче требовалось определить, сумеет ли имеющийся алгоритм оптимизации схемы обработать её в заданное время. Эта задача решалась аналогично предыдущей.

Разработанные алгоритмы указанных задач оказались эффективными.

Анализ показывал, что схемы из одного и того же класса образовывали некомпактные области в пространствах их исходного описания, т.е. для них не выполнялась т.н. гипотеза компактности (ГК). Это характеризует задачи как чрезвычайно сложные в практическом и теоретическом аспектах, для которых обычные методы распознавания оказываются непригодными.

Гипотеза компактности в задачах распознавания

Наиболее общая неформальная (и, как представляется, наиболее адекватная сегодняшнему уровню понимания проблемы) формулировка ГК предложена ещё в классической монографии [1]: «*образам соответствуют компактные множества в пространстве выбранных свойств*»¹. Предположение о выполнении ГК лежит в основе подавляющего большинства подходов к решению различных типов задач распознавания (а при решении задач кластеризации является определяющим). Действительно, при выполнении ГК полученные в рамках любого из традиционных методов алгоритмы классификации имеют, очевидно, невысокую сложность и, как следствие, легко реализуемы, не требуют больших вычислительных ресурсов по памяти, времени счёта, и т. д. Также важной характеристикой таких алгоритмов являются достаточно высокие оценки их надёжности — прямое следствие выполнения ГК.

Задачи, где ГК не имеет места, указанные «хорошие» свойства алгоритмов распознавания — при использовании обычных подходов — отнюдь не гарантируются. В силу этого на протяжении последних лет они находились вне круга интересов разработчиков: теория, способная дать направления решения таких задач отсутствовала. Здесь надо сказать, что попытки решения задач с невыполненной ГК делались ещё на заре развития теории и практики распознавания образов. В известной монографии М. М. Бонгарда [3] приведён демонстрационный пример решения задачи указанного типа. Однако этот подход на многие годы оставался невостребованным...

Для решения задач данного типа необходимо было понять, что же понимается под неформальным понятием «компактные образы», откуда берётся и как формируется пространство свойств, и т. д. В связи с этим в последнее время были предприняты некоторые попытки формализации ГК [6, 5]. Однако они либо носили частный характер, либо уводили проблему в общефилософскую плоскость [2]. Так что на сегодняшний день этот вопрос является открытым.

В данной работе приведены примеры практического решения двух задач с невыполненной ГК. Решения основывались на подходе М. М. Бонгарда. При этом мы не предлагаем своей формулировки ГК, считая имеющиеся у нас попытки её определения предварительными и оставив их для дальнейших публикаций по данной теме исследования.

¹ Данная формулировка, по сути, лишь несколько уточняет первоначальную, высказанную М. А. Айзерманом в первых работах по распознаванию образов конца 50-х годов XX в.

Решения задач. Результаты

Для решения задач использовался вышеупомянутый подход М. М. Бонгарда. На основе первичных признаков (до 10–12) объектов генерировались вторичные признаки (до десятков тысяч) как функции от первичных, из которых отбирались наиболее информативные. Окончательно задачи классификации (определения областей компетентности имеющихся алгоритмов) решались в сформированном признаковом пространстве. Ошибка классификации, определяемая методом скользящего контроля с одним исключаемым прецедентом, не превосходила нескольких процентов.

В задаче № 1 исследовались 19 практических алгоритмов синтеза схем и 120 описаний реальных схем (первые результаты см. в [4]). В задаче № 2 исследовались 185 схем (разбиение по классам: 175 + 10).

Программные модули, реализующие разработанные алгоритмы, интегрированы в открытую среду SIS [7], информационно совместимую со многими промышленными системами синтеза БИС, в частности используемой в фирме Intel Inc.

В развитие подхода видится целесообразным разработать способ уменьшения числа рассматриваемых вторичных признаков на основе исключения из рассмотрения тождественно равных, но структурно отличных признаков и установления нахождения причинно-следственной связи между информативностью двух различных вторичных признаков.

Работа выполнена при поддержке РФФИ, проект № 07-01-00211.

Литература

- [1] Айзерман М. А., Браверман Э. М., Розоноэр Л. И. Метод потенциальных функций в теории обучения машин. — М.: Наука, 1970. — 384 с.
- [2] Белозерский Л. А. Современный взгляд на гипотезу компактности // Штучный интеллект (Донецк). — 2005. — № 3. — С. 6–12.
- [3] Бонгард М. М. Проблема узнавания. — М.: Наука, 1967.
- [4] Гранкин М. В., Гуров С. И., Фатхутдинов И. Н. Определение областей компетентности алгоритмов при невыполненной гипотезе компактности // Штучный интеллект (Донецк). — 2006. — № 2. — С. 88–98.
- [5] Донской В. И. О метрических свойствах кратчайших эмпирических закономерностей // Уч. записки ун-та им. В. И. Вернадского. — 2003. — № 2. — С. 143–147.
- [6] Загоруйко Н. Г., Елжина В. Н., Киприянова Т. П. Пакет Прикладных Программ ОТЭКС для Обработки Таблиц Экспериментальных данных. Версия 4.0 — www.math.nsc.ru/AP/oteks/Russian/.
- [7] SIS: A System for Sequential Circuit Synthesis // Dep. of Electrical Engineering and Computer Science Univ. of California, Berkeley. — 1992.