

О корректной классификации над произведением частичных порядков

Е.В. Дюкова¹, Г.О. Масляков²

¹Вычислительный Центр ФИЦ ИУ РАН, Вавилова 42, Москва, Россия, 119991

²ООО «Интера-Сервис», Рязанский проспект 8а, Москва, Россия, 109428

Аннотация

Рассматривается одна из ключевых задач машинного обучения – задача классификации на основе прецедентов. Приводится описание схемы синтеза корректных логических алгоритмов классификации при условии, что на множествах значений признаков заданы конечные частичные порядки. Изучается возможность выбора на этапе предварительного анализа обучающей выборки таких частичных порядков, которые обеспечивают качественную классификацию.

Ключевые слова

Машинное обучение, логические алгоритмы классификации, корректный алгоритм классификации по прецедентам, декартово произведение частичных порядков

1. Введение

В задаче классификации под прецедентной (обучающей) информацией понимается совокупность примеров изучаемых объектов, в которой каждый объект представлен в виде числового вектора, полученного на основе измерения или наблюдения ряда его параметров или характеристик, называемых признаками. Каждый пример (обучающий объект или прецедент) приписан к определённому классу объектов. Требуется по признаковому описанию предъявленного объекта, о котором заранее неизвестно, какому классу он принадлежит, определить (распознать) этот класс.

Основное достоинство логического подхода к задаче классификации (распознавания) – возможность получения результата при отсутствии дополнительных предположений вероятностного характера и при небольшом числе прецедентов. Предполагается, что каждый признак принимает ограниченное число допустимых целочисленных значений. Анализ прецедентной информации сводится к поиску в исходных данных определенных закономерностей или элементарных классификаторов, различающих объекты из разных классов. При этом большое внимание уделяется вопросам синтеза алгоритмов, безошибочно классифицирующих материал обучения (корректных алгоритмов). Наиболее известными в данной области являются алгоритмы корректного голосования [1, 3, 4], предложенные впервые в отечественных работах, а также методы Logical Analysis of Data (LAD) и Formal Concept Analysis (FCA) [5]. Последние два направления логической классификации в основном нацелены на обработку бинарных данных. Основополагающие идеи LAD и FCA принадлежат соответственно П. Хаммеру (1986 г.) и Р. Вилле (1981 г.).

Стандартные постановки логической классификации не всегда позволяют решать прикладные задачи со сложными отношениями на множествах допустимых значений признаков. В [3] на базе обобщения классических понятий логической классификации предложены корректные процедуры классификации при условии, что признаковые описания объектов являются элементами декартового произведения конечных частично упорядоченных множеств, и рассмотрены вопросы применения асимптотически оптимальных алгоритмов перечисления элементарных классификаторов общего вида.

В настоящей работе продолжены начатые в [3] исследования. Рассмотрена актуальная и практически не изученная задача выбора «хороших» частичных порядков на множествах значений признаков и для её решения предложены вычислительно эффективные процедуры.

2. Основные результаты

В [2] предложены быстрые методы выбора частичных порядков на множествах допустимых значений признаков, основанные на предварительном анализе обучающей выборки. Показано, что рассмотренные способы упорядочивания значений признаков существенно повышают качество классификации, однако не гарантируют корректность классификации.

Основные результаты настоящей работы касаются исследования возможности построения «хороших» частичных порядков на множествах допустимых значений признаков, обеспечивающих корректную классификацию обучающей выборки. Сформулирован критерий корректности логической классификации над произведением частичных порядков, согласно которому для корректной классификации описание каждого прецедента из каждого класса должно быть независимым от описания любого прецедента из другого класса. Показано, что задача выбора корректных частичных порядков на множествах допустимых значений признаков может быть решена на основе построения неприводимых покрытий булевой матрицы, специальным образом построенной по обучающей выборке. Предложены два основных подхода к рассматриваемой задаче. Первый подход заключается в построении частичных порядков, обеспечивающих корректную классификацию всех обучающих объектов. Второй подход состоит в построении для каждого класса частичных порядков, гарантирующих корректную классификацию только прецедентов данного класса. На прикладных задачах проведено экспериментальное сравнение указанных подходов с методами упорядочивания значений признаков, анонсированными в [2]. Эксперименты проведены для частного случая, когда множество значений признака является цепью или антицепью,

3. Заключение

Авторами продолжены начатые в [2, 3] исследования по разработке методов логического анализа частично упорядоченных данных и применения этих методов в машинном обучении. Изучены вопросы повышения качества логической классификации путём задания на множествах допустимых значений признаков «хороших» частичных порядков. Предложены и исследованы различные подходы к задаче выбора «хороших» частичных порядков.

4. Благодарности

Работа выполнена при частичной финансовой поддержке РФФИ (проект № 19-01-00430-а).

5. Литература

- [1] Баскакова, Л.В. Модель распознающих алгоритмов с представительными наборами и системами опорных множеств / Л.В. Баскакова, Ю.И. Журавлёв // Ж. вычисл. матем. и матем. физ. – 1981. – Т. 21, № 5. С. – 1264-1275.
- [2] Бакланова, А.О. Исследование зависимости качества классификации от выбора частичных порядков на множествах значений признаков / А.О. Бакланова, Е.В. Дюкова, Г.О. Масляков // 13-я международная конференция «Интеллектуализация обработки информации». – М.: Российская академия наук, 2019. – С. 21-23.
- [3] Djukova, E.V. On the Logical Analysis of Partially Ordered Data in the Supervised Classification Problem / E.V. Djukova, G.O. Maslyakov, P.A. Prokofyev // Computational Mathematics and Mathematical Physics. – 2019. – Vol. 59(9). – P. 1542-1552.
- [4] Kovshov, N.V. Algorithms for finding logical regularities in pattern recognition / N.V. Kovshov, V.L. Moiseev, V.V. Ryazanov // Computational Mathematics and Mathematical Physics. – 2008. – Vol. 48(2). – P. 314-328.
- [5] Janostik, R. Interface between Logical Analysis of Data and Formal Concept Analysis / R. Janostik, J. Konecny, P. Krajča // Elsevier European Journal of Operational Research. – 2020. – Vol. 284(2). – P. 792-800.