

Л.А. ФИЛИПКОВСКАЯ, канд. техн. наук, НАУ "ХАИ" (г. Харьков)

ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ МЕЖДУ ОБЪЕМОМ ОБУЧАЮЩЕЙ ВЫБОРКИ И КАЧЕСТВОМ СТРУКТУРНО-АНАЛИТИЧЕСКОГО ПРАВИЛА КЛАССИФИКАЦИИ НА ОСНОВЕ СТАТИСТИЧЕСКОГО АНАЛИЗА

Выделен класс производственных процессов, управление которыми требует классификации (распознавания) рассматриваемых в них производственных ситуаций. Идентификация объекта управления сводится к анализу его параметров. Выработка управляющих воздействий базируется на использовании структурно-аналитического правила классификации. Рассмотрены характеристики терминальных классификаторов в условиях малых выборок.

Ключевые слова: производственная ситуация, распознавание, структурно-аналитическое правило, правило классификации.

Постановка проблемы и анализ литературы. Такие производственные процессы, как техническая подготовка производства [1 – 3], изготовление крупных деталей в машиностроении [4], медицинская диагностика и выработка стратегии лечения [5], характеризуются большим числом и разнотипностью параметров, многоэтапностью функционирования. В [6] выделен класс производственных процессов, управление которыми требует классификации (распознавания) рассматриваемых в них производственных ситуаций. Под производственной ситуацией будем понимать вектор параметров, измеренных в разнотипном пространстве признаков. Идентификация объекта управления сводится к анализу его параметров и построению правила классификации (ПК) в режиме обучения и самообучения по экспериментальным данным, а выработка управляющих воздействий базируется на использовании правила классификации и оценке его эффективности. При этом важную роль играет объем обучающей выборки.

В [7] представлены основные проблемы, для решения которых используется структурно-аналитический метод распознавания образов. Описанию алгоритмов анализа экспериментальных данных и построению правил классификации с целью обоснования простоты технической реализации и интерпретации решающего правила посвящена работа [8].

Цель статьи – исследование терминальных классификаторов в условиях малых выборок, что характерно для реальных задач, описанных выше.

Зависимость между объемом обучающей выборки и качеством структурно-аналитического правила классификации. Методика использования интерактивного программного комплекса классификационной обработки данных [8] состоит из трех уровней: ввод и предобработка

исходных данных по согласованию с пользователем, анализ данных и построение правила классификации с оценкой его качества.

На базе этого программного комплекса был разработан программный стенд, состоящий из генератора таблицы экспериментальных данных, процедур оценки вероятностной ошибки классификации (ВОК) методом "скользящего контроля", а также библиотеки алгоритмов распознавания образов. Внутренние вершины грамматического ПК содержали одномерные решатели, оптимальные по критерию Байеса (ПК Байеса), классификатор для разделения выпуклых оболочек классов (ПК "Минимакс"), а также реализация линейного дискриминанта Фишера (ПК Фишера).

Асимптотическая устойчивость грамматического ПК проверялась на двух нормальных совокупностях с числом реализаций $m = 150$ и числом разнотипных признаков $n = 5$. Теоретическое значение ВОК P^* равнялось 0,286, что следует из решения статистических проблем управления.

ВОК оценивалась по контрольной выборке объемом 1200 объектов, генерированных по тому же закону, что и обучающая выборка.

Результаты моделирования представлены на рис. 1.

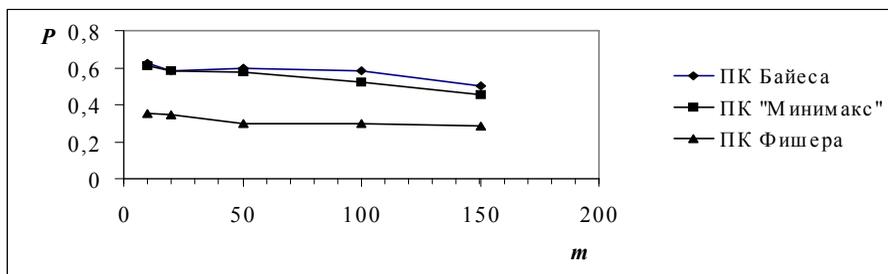


Рис. 1. Диаграмма зависимости ВОК P от объема обучающей выборки m

Эффективность ПК (уровень ВОК) приближается к асимптотическому значению при использовании одномерных классификаторов в условиях малых выборок ($m \sim n$).

Практический интерес представляет задача нахождения связей между ВОК и объемом обучающей выборки m при различном числе ветвей N_v грамматического ПК. Была смоделирована аналогичная описанной выше статистическая задача с обучением распознаванию $s = 2$ и $s = 3$ классов с $n = 3$ признаками.

Результаты моделирования представлены на рис. 2 и рис. 3 и указывают на стремление ВОК к 1 у более сложных грамматических ПК, имеющих большое число ветвей N_v при заданных s и m .

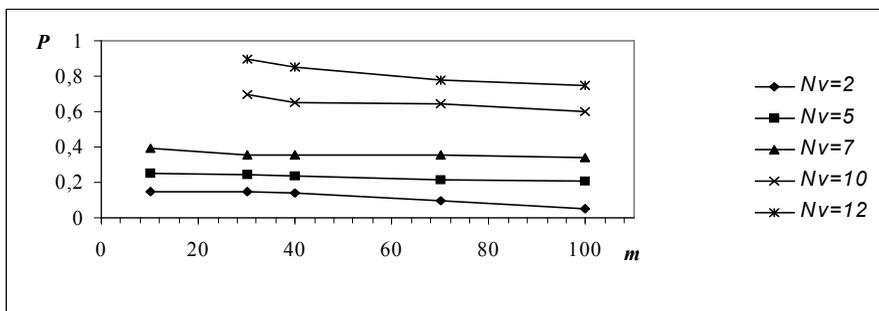


Рис. 2. Диаграмма зависимости ВОК P от объема обучающей выборки m и $s = 2$

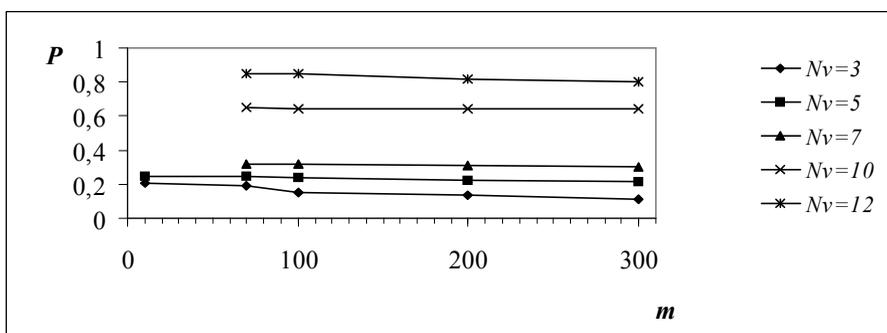


Рис. 3. Диаграмма зависимости ВОК P от объема обучающей выборки m и $s = 3$

Приведенные факты позволяют сделать следующие выводы, полезные для построения ПК после предварительного анализа данных:

- для малых выборок реальных задач обосновано применение простых в реализации и интерпретации решающих правил с достаточным уровнем ВОК;
- так как число терминальных классификаторов прямо пропорционально числу ветвей грамматического ПК N_v , то из всех структурно-полных систем бинарной таблицы для восстановления ПК предпочтительнее выбирать ту, которая имеет наименьшее число терминальных классификаторов.

Выводы. В работе были предложены рекомендации для построения структурно-аналитического ПК после предварительного анализа данных [9]. Таким образом, полученные результаты необходимы для выработки практического критерия предпочтения при выборе правила классификации с помощью интерактивного программного комплекса классификационной обработки данных. Опираясь на эти исследования, а также на теоретические и алгоритмические результаты работы [8], обеспечивается повышение эффективности управления производственных процессов за счет

использования оптимальных в смысле минимума эмпирического риска структурно-аналитических моделей анализа и распознавания производственных ситуаций, получения надежных правил их классификации, сокращения стоимости продукции, а также материальных и временных затрат.

Список литературы: 1. *Евгеньев Г.Б.* Систематология инженерных знаний: учебн. пособие для вузов. – М.: МГТУ им. Н.Э. Баумана, 2001. – 246 с. 2. *Игнатъева А.В., Максимов М.М.* Исследование систем управления. – М.: Юнити-Дана, 2000. – 157 с. 3. *Филипковская Л.А., Шостак И.В.* Интеллектуальная интегрированная система автоматизированной обработки производственных ситуаций структурно-аналитическим методом распознавания образов // Автоматизированные системы управления и приборы автоматики. – Х., 2001. – Вып. 115. – С. 65–69. 4. Автоматизации управления предприятием / *В.В. Баронов, Г.Н. Калянов, Ю.И. Попов* и др. – М.: ИНФРА-М, 2000. – 239 с. 5. *Аспарухов О.К., Георджева С.А., Кисъов В.Т.* Методы за вземане на решения в медецината, основани на теорията на размытите множества // Биоавтоматика. – 1986. – № 2. – С. 68–77 (болг.) 6. *Филипковская Л.А.* Структурно-аналитическая модель распознавания образов в управлении производственными процессами // Вестник НТУ "ХПИ". – Х.: НТУ "ХПИ". – 2001. – № 6. – С. 279–283. 7. Структурно-аналитический метод машинного распознавания образов с разнотипными признаками. / *Стоян Б.Г., Проценко В.С., Сироджа И.Б.* и др. // Теория R-функций и актуальные проблемы прикладной математики. – К., 1986. – С. 212–244. 8. *Филипковская Л.А.* Информационная технология классификационной обработки данных производственных ситуаций // Вісник НТУ "ХП". – Х.: НТУ "ХП". – 2003. – № 7. – Т. 2. – С. 93–98. 9. Перспективы использования объектного подхода в ресурсноберегающем заготовительно-штамповочном производстве / *А.И. Долматов, В.В. Третьяк, В.Ю. Гранин, Л.А. Филипковская* // Авиационно-космическая техника и технология. – Х.: "ХАИ". – 2007. – 11/47 – С. 245–254.

УДК 658.012.011.56: 681.327.12

Дослідження залежності між об'ємом навчальної вибірки і якістю структурно-аналітичного правила класифікації на основі статистичного аналізу / Філіпковська Л.О. // Вісник НТУ "ХП". Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХП", 2008. – № 24. – С. 185 – 188.

Виділено клас виробничих процесів, управління якими вимагає класифікацію (розпізнавання) розглянутих у них виробничих ситуацій. Ідентифікація об'єкта управління зводиться до аналізу його параметрів. Вироблення управлінських впливів базується на використанні структурно-аналітичного правила класифікації. Розглянуто характеристики термінальних класифікаторів в умовах малих вибірок. Іл.: 3. Бібліогр.: 9 назв.

Ключові слова: виробнича ситуація, розпізнавання, структурно-аналітична правило, правило класифікації.

UDC 658.012.011.56: 681.327.12

Research of dependence between volume of training sample and quality of a structural-analytical classification rule on the basis of the statistical analysis / Filipkovskaja L.A. // Herald of the National State University "KhPI". Subject issue: Information science and modelling. – Kharkov: NSU "KhPI", 2008. – № 24. – P. 185 – 188.

The class of productions, management with which demands classification (recognition) of industrial situations considered in them, is allocated. Identification of management object is reduced to the analysis of its parameters. Manufacture of managing influences is based on use of a structural - analytical classification rule. Characteristics of terminal qualifiers in conditions small sample of the data are considered. Figs: 3. Refs: 9 titles.

Key words: an industrial situation, recognition, structural-analytical rule, a classification rule.

Поступила в редакцію 05.05.2008