

УДК 004.42, 004.67, 681.138.4

Н. В. Размочаева, Д. М. Клионский
Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Исследование проблемы представления данных с использованием корреляционного анализа в задаче автоматизации процесса управления розничными продажами

Рассматривается одна из граней проблемы автоматизации процесса управления розничными продажами, а именно проблема представления данных большой размерности. Проблема заключается в том, что для описания данных используется большое количество характеристик (параметров). Исследуется один из популярных подходов к сокращению размерности пространства признаков – корреляционный анализ. Приведены результаты применения корреляционного анализа к данным, полученным при реализации процесса розничной торговли через автоматизированные системы (торговые автоматы). Посредством корреляционного анализа были исключены параметры с сильными зависимостями, в которых наблюдается сильная линейная связь. Вычисленная корреляционная правка показала, что обнаруженная сильная линейная связь не случайна. Результаты корреляционного анализа подтверждены построением классификатора на основе ансамбля лесов деревьев решений. Классификатор показал, что в параметрах со слабой корреляционной связью нет других сильных зависимостей. Результаты анализа были проверены экспертной группой (группа маркетологов из области розничной торговли), которая подтвердила их состоятельность.

Представление данных, анализ данных, интерпретация данных, корреляционный анализ, автоматизация процессов, розничные продажи, управление, машинное обучение

В условиях современного уровня информационных технологий, когда объемы хранимой информации крайне велики, особенно остро стоит проблема не только хранения, но и представления, обработки данных. Большинство исследований по работе с данными нацелены на новый вектор развития, а именно – решают задачи извлечения полезной информации (так называемых знаний). При решении такого рода задач в последнее время все большую популярность набирают методы машинного обучения [1], [2], в которые обычно включают методы интеллектуального анализа данных [3, с. 68], в том числе методы глубокого обучения [3, с. 89], классификации и регрессии [3, с. 102], кластеризации [3, с. 159] и многие др.

Предварительная обработка данных играет важную роль, отвечая на вопрос о применимости

некоторых методов машинного обучения. В настоящей статье представлены результаты исследования применения на этапе предварительной обработки корреляционного анализа как метода снижения размерности пространства в задаче представления данных о продажах товаров.

Задачи, которые необходимо решить для достижения цели, следующие:

1. Провести обзор предметной области, существующих методов.
2. Провести анализ данных (применить корреляционный анализ).
3. Сделать заключение об адекватности получаемых результатов.

Актуальность рассмотрения задачи представления данных обусловлена важностью представления данных для нескольких групп пользовате-

лей, в основном, для сотрудников организации (менеджеры среднего звена и топ-менеджеры, занимающиеся маркетинговым анализом, маркетологи). Сотрудники анализируют продажи товаров и формируют ассортимент товаров на основе результатов анализа.

Проблема заключается в избыточности характеристик (параметров) товаров. Так как данные о продажах поступают из разных источников, то для разных случаев – разные параметры избыточны. Целевая программная система, где необходимо выполнять анализ достаточности (избыточности) характеристик товаров, – это настольное приложение для операционной системы Windows [4]. Цель получения компактного представления данных – повышение удобства работы пользователей с данными и повышение скорости выполнения анализа.

Природа обрабатываемых данных. Исходные данные – данные о продажах товаров, формируемые за определенный период времени [5]. Обычно период времени определяется частотой обслуживания точек реализации (если, например, точка обслуживается раз в неделю, то период будет составлять неделю). Данные, представляющие собой результаты процесса продаж, могут быть сформированы различными способами. На рис. 1 представлена схематичная ER-диаграмма предметной области (подпись n означает характер связи «один ко многим»).

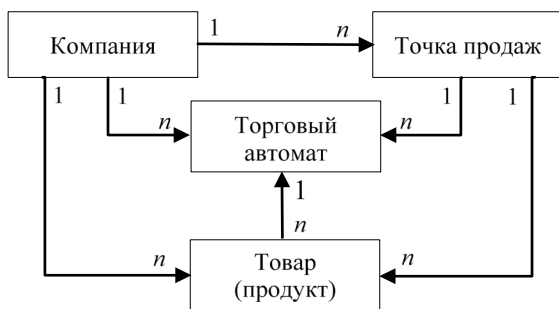


Рис. 1

Как видно из рис. 1, имеется 4 основных сущности: «Компания» (имеется в виду та, что реализует товары), «Точка продаж» (место, где располагаются 1 или несколько торговых автоматов), «Торговый автомат», «Товар». Иллюстрируемые связи нужно читать так: (1) у одной «Компании» может быть множество «Точек продаж» (контактов с разными потребителями), множество «Торговых автоматов» и множество «Товаров»; (2) за каждой «Точкой продаж» закреплена только 1 «Компания», но может располагаться множе-

ство «Торговых автоматов» и, соответственно, «Товаров»; аналогично у «Торгового автомата» только 1 «Точка продаж» и только одна владеющая «Компания», но при этом, естественно, множество «Товаров»; множество «Товаров» принадлежат только одному «Торговому автомату», только одной «Компании» и, соответственно, только одной «Точке продаж».

Важная особенность в том, что данные имеют иерархическую структуру (рис. 2). Иерархия выражается в том, что, например, выборка данных о продажах товаров по всей компании будет включать в себя особенности каждой (отдельно взятой) точки продаж. Аналогично и с точками продаж: если рассматривать данные о продажах в рамках одной точки продаж, то эти данные будут включать в себя особенности продаж товаров в каждом отдельном автомате. Такое включение чревато тем, что «особенности» включаемых источников будут заглушать друг друга. Здесь же стоит отметить, что рассмотрение данных о продажах товаров по всей компании не целесообразно – данных слишком много (глобальный взгляд), анализ трудоемок. И рассматривать продажи в рамках одного торгового автомата тоже нецелесообразно, но несколько по другой причине: в рамках одной точки продаж могут эксплуатироваться несколько торговых автоматов, и, ограничиваясь одним, можно потерять специфику всей точки продаж (в рамках одного торгового автомата – частный взгляд).

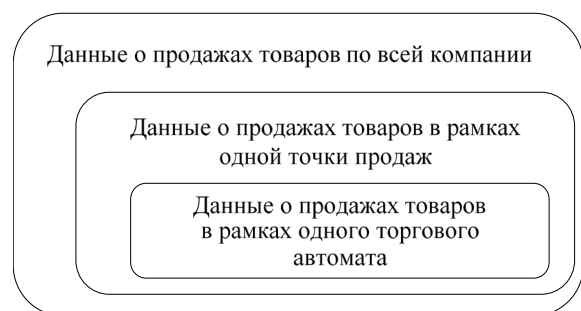


Рис. 2

Получается, что целесообразнее рассматривать данные о продажах товаров в рамках одной точки реализации.

Каждый источник представляется в табличном виде (рис. 3). Особенно ценную информацию несут данные о продажах в рамках одного клиента (точки реализации), так как они отражают особенности предпочтения конкретных покупателей.

	P_1	P_2	...	P_K
T_1				
T_2				
...				
T_N				

Рис. 3

На рис. 3 использованы следующие обозначения: T_i – наименования товаров ($i \in [1, N]$); N – количество товаров; P_j – наименования параметров товаров ($j \in [1, K]$); K – количество параметров.

Существующие методы анализа товаров. Среди существующих методов анализа значимости параметров наибольшее распространение получили: ABC-анализ (рис. 4), XYZ-анализ (табл. 1) и совмещенный (ABC + XYZ)-анализ (табл. 2) [6].

		P_1	P_2	...	P_K
А-группа ($\leq 80\%$)	T_1				
	T_2				
	T_3				
В-группа (80–95 %)	...				
	...				
	...				
С-группа (95–100 %)	T_{N-1}				
	T_N				

Рис. 4

Таблица 1

Иллюстрация соотношения групп А, В, С и X, Y, Z	
Группы	Оценки и диапазоны, %
A/X	80 / (< 15)
B/Y	15 / (15–20)
C/Z	5 / (> 50)

Для применения данных нужна априорная информация – целевой (важный по определенным критериям) параметр. Задача определения критериев для выделения важного столбца требует отдельного рассмотрения.

Принцип работы методов общий: по выделенному параметру вычисляется общая сумма значений, рассчитываются процентные значения каждого параметра, далее вычисляются накопленные проценты, и в зависимости от заданных априори границ (например, 80, 95 %) определя-

ются группы товаров. Например, для ABC-анализа: товары, которые обеспечивают вклад $\leq 80\%$ в значение параметра – А-группа; В-группа – вклад $\leq 15\%$, и товары, вклад в параметр которых незначительный ($\leq 5\%$), – С-группа [7]. С-группа считается критичной, продажи товаров С-группы должны быть отдельно проанализированы и, возможно, исключены из рассматриваемого ассортимента (заменены на более рентабельные товары). Алгоритм подбора товаров на замену подробно рассматривается в [6].

Краткая характеристика совмещенных групп в табл. 2:

1) товарам в группах AX (высокая стоимость) и BX (средняя стоимость) присуща хорошая выкупаемость (не требуется обеспечивать большой запас) и стабильный оборот – эти две характеристики, действующие интегрально, называют высокой прогнозируемостью;

2) товары, попавшие в группы AY и BY, обладают хорошим оборотом, но не обладают должной стабильностью, вследствие чего прогнозируемость невысокая;

3) товары в группах AZ и BZ все также характеризуются хорошим оборотом, но при этом о стабильности не идет и речи, в связи с чем прогнозируемость оказывается на низком уровне;

4) для товаров группы CX характерны низкая стоимость, стабильность и высокая прогнозируемость;

5) товары в группах CY и CZ обладают низкой стоимостью и слабой прогнозируемостью, обычно их стремятся исключить из ассортимента.

Существуют подходы, при которых XYZ-анализ используется для решения задач управления запасами на складах, но такая задача достойна отдельного рассмотрения. Подробнее с совмещенными группами, представленными в табл. 2, можно ознакомиться в [7].

Помимо упомянутых ранее недостатков (наличие априорной информации) применимость указанных методов на больших данных весьма затруднительна, так как в процессе используется вся информация, что требует больших временных затрат и дальнейшего анализа (интерпретации)

Таблица 2

Вид анализа	Горизонтальный анализ заданного параметра по всем товарам			
	А-группа	В-группа	С-группа	
Вертикальный анализ временного ряда заданного параметра по каждому товару	Х-группа	АХ	ВХ	СХ
	Y-группа	AY	BY	CY
	Z-группа	AZ	BZ	CZ

Сравнение существующих методов		
Название	Достоинства	Недостатки
ABC-анализ	Группировка по заданному параметру	Требуются дополнительные ресурсы
XYZ-анализ	Выявление временных особенностей в продажах	1. Требуются дополнительные ресурсы. 2. Неточности для «невременных данных»
Совмещенный анализ	Группировка товаров по заданному параметру с учетом временных особенностей	1. Требуются дополнительные ресурсы. 2. Неточности для «невременных» данных

результатов [8]. Практика показывает, что указанные методы прекрасно работают на временных данных (и чаще всего используются для анализа временных данных). Но, к сожалению, данные для рассматриваемой задачи не являются временными рядами, и на настоящем этапе исследования временные ряды по рассматриваемым товарам невозможно получить. Поэтому в описываемом исследовании использовались другие методы, рассмотренные далее.

В табл. 3 приведено резюме сравнительного анализа ABC- и XYZ-методов и их совмещенного варианта.

Для исследования характеристик товаров были выбраны методы регрессионного анализа в силу их простоты и доступности в применении. Применение корреляционного анализа призвано показать, а возможно ли в принципе сокращение размерности пространства признаков (анализируется первичная предпосылка – наличие линейных связей). Положительные результаты корреляционного анализа послужат отправной точкой для применения других методов сокращения размерности пространства признаков.

Если понимать регрессионный анализ в общем смысле, то стоит уточнить, что на сегодняшний день науке известны следующие алгоритмы регрессии в области машинного обучения: линейная, гребневая, изотоническая, логистическая, лассо-регрессия, байесовская, логическая, квантильная, LAD-регрессия, Джекнайф-регрессия (метод складного ножа), экологическая и некоторые др. Далее будет рассмотрено применение линейной регрессии для исследуемых данных. Другие виды регрессии будут исследованы в дальнейшем.

Анализ параметров данных. Данные представляют собой числовые наборы значений параметров по каждому товару (табличное или матричное представление будет обсуждаться далее). Описание отдельно взятого товара можно понимать как вектор значений параметров. Данные представляют собой сведения о продажах в рамках одной точки продаж (в силу ранее рассмот-

ренных причин). Размерности данных таковы: количество рассматриваемых товаров – 93 шт., количество параметров для анализа – 12. Во избежание разглашения коммерческой тайны конкретные наименования товаров и названия параметров не будут упоминаться в настоящей статье. Однако уточним, что исследуемые параметры можно условно разделить на 2 группы: денежные параметры, выраженные в денежных единицах (например, рубли и тысячи рублей), и количественные параметры, выраженные в штуках (на соответствующих графиках будут выделены полужирным, см. дальнейшие уточнения).

На рис. 5 представлены графики зависимостей некоторых параметров товаров, где по осям x и y отложены значения первого и второго параметров соответственно, масштабированные к интервалу $[0,1]$.

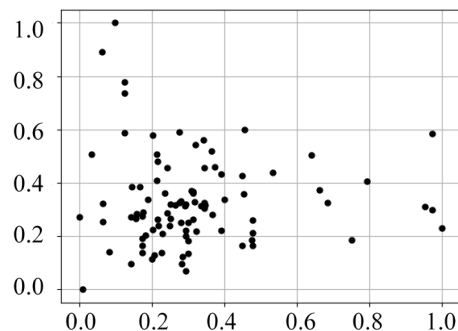


Рис. 5

По рис. 5 нельзя точно судить, каким образом параметры товаров взаимодействуют друг с другом и как можно использовать их взаимосвязи для оптимизации продаж (и есть ли эти взаимосвязи вообще).

Для проверки достаточности имеющегося набора параметров (12 штук) использовался корреляционный анализ, предусматривающий вычисление коэффициента корреляции r :

$$r_{p_i, p_j} = \frac{\sum_{k=1}^N (p_{i,k} - \bar{p}_i)(p_{j,k} - \bar{p}_j)}{\sqrt{\sum_{k=1}^N (p_{i,k} - \bar{p}_i)^2 \sum_{k=1}^N (p_{j,k} - \bar{p}_j)^2}},$$

$$i \in [1, K], j \in [1, K],$$

Таблица 4

Сила связи	Характеристика связи	
	Прямая (положительный коэффициент)	Обратная (отрицательный коэффициент)
Полная	1	-1
Сильная	[0.7; 1)	[-0.7; -1)
Средняя	[0.3; 0.7)	[-0.3; -0.7)
Слабая	(0.3; 0)	(-0.3; 0)
Связь отсутствует	0	0

Таблица 5

Номер параметра	0	1	2	3	4	5	6	7	8	9	10	11
0	1	0.89	0.74	0.87	0.8	0.82	0.05	0.04	0.02	0.05	0.27	0.35
1	0.89	1	0.96	0.98	0.71	0.96	0.09	0.16	0.12	0.21	0.28	0.29
2	0.74	0.96	1	0.95	0.59	0.95	0.11	0.21	0.19	0.34	0.26	0.23
3	0.87	0.98	0.95	1	0.72	0.96	0.21	0.24	0.01	0.25	0.23	0.28
4	0.8	0.71	0.59	0.72	1	0.77	0.27	0.32	0.05	0.18	0.09	0.36
5	0.82	0.96	0.95	0.96	0.77	1	0.19	0.28	0.13	0.31	0.16	0.29
6	0.05	0.09	0.11	0.21	0.27	0.19	1	0.9	0.45	0.44	0.38	0.02
7	0.04	0.16	0.21	0.24	0.32	0.28	0.9	1	0.01	0.69	0.34	0.04
8	0.02	0.12	0.19	0.01	0.05	0.13	0.45	0.01	1	0.4	0.18	0.05
9	0.05	0.21	0.34	0.25	0.18	0.31	0.44	0.69	0.4	1	0.01	0.02
10	0.27	0.28	0.26	0.23	0.09	0.16	0.38	0.34	0.18	0.01	1	0.21
11	0.35	0.29	0.23	0.28	0.36	0.29	0.02	0.04	0.05	0.02	0.21	1

где p_i – i -й параметр; $p_{i,k}$ – k -е значение i -го параметра; \bar{p}_i – среднее значение i -го параметра; K – количество параметров.

При вычислениях использовалась корреляционная поправка:

$$\left| \frac{r}{S_r} \right| = \left| r \left(\frac{1-r^2}{\sqrt{n-1}} \right)^{-1} \right|,$$

где n – размер выборки; r – коэффициент корреляции (сокращенная запись для r_{p_i, p_j}); S_r =

$$= \frac{1-r^2}{\sqrt{n-1}} \text{ – поправочный (корректирующий)}$$

множитель.

Значения корреляционной поправки помогут определить, была ли обнаружена зависимость случайной или нет. Пороговое значение для корреляционной поправки равно 3.

В результате вычислений было обнаружено, что некоторые параметры обладают сильной линейной зависимостью, что означает их избыточность.

Исключаем сильнокоррелирующие (порог коэффициента корреляции установим равным 0.7) параметры из рассмотрения. Значение порога было выбрано на основании классификации силы корреляции в зависимости от значения коэффициента корреляции (табл. 4) [9].

Вычислим коэффициент корреляции по всем параметрам товара (попарная корреляция). Сгруппируем полученные значения в корреляционную матрицу (табл. 5). Проанализировав коррелирующие параметры, представленные в таблице, видим, где наблюдается сильная корреляция (выделено цветом).

Рассмотрим отдельно сильнокоррелирующие параметры. Для каждой коррелирующей пары вычислим корреляционную поправку. Результаты вычислений представлены в табл. 6. Полужирным выделены количественные параметры, все остальные – денежные параметры.

Сильная корреляция наблюдается как у количественных параметров, так и у параметров, выраженных в денежном эквиваленте (табл. 5), а также и между этими параметрами. Вычисленные значения корреляционной поправки (табл. 6, последний столбец) показывают, что обнаруженные зависимости между параметрами не являются случайными.

Это подтверждается соответствующими графиками (рис. 6), из которых видно, что между обозначенными параметрами действительно есть сильная линейная зависимость.

На рис. 6 введены следующие обозначения: — – линейная регрессия; • – исходные данные. Необходимо уточнить, что корреляционный анализ проводился с данными, претерпевшими масштабирование к интервалу [0;1] следующим образом:

Таблица 6

Коррелирующие параметры		Коэффициент корреляции	Корреляционная правка
Первый параметр	Второй параметр		
0	2	0.74	31.90
0	4	0.80	42.45
0	5	0.82	47.01
0	3	0.87	67.02
0	1	0.89	85.98
1	4	0.71	27.88
1	2	0.96	255.04
1	5	0.96	220.50
1	3	0.98	475.02
2	3	0.95	186.10
2	5	0.95	174.74
3	3	0.72	28.58
3	5	0.96	210.02
4	5	0.77	36.17
6	7	0.90	90.21

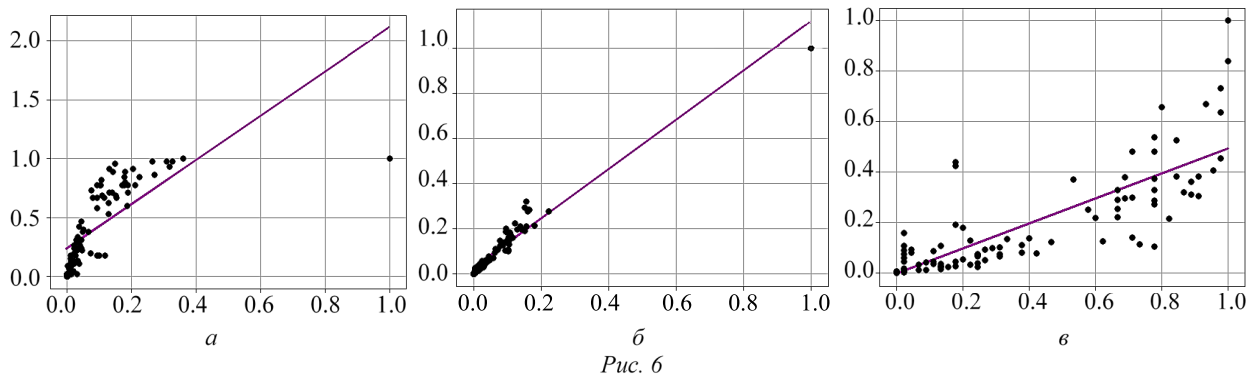


Рис. 6

$$R = X_{\max} - X_{\min},$$

$$X'_i = \frac{X_i - X_{\min}}{R},$$

где R – размах выборки; X_{\max} и X_{\min} – максимальное и минимальное значения выборки соответственно; X'_i – преобразованное значение (масштабированное); X_i – i -е значение исходной выборки. Масштабирование представляет собой своего рода отображение $f: \{X_i\} \rightarrow \{X'_i\}$.

На рис. 6 представлены 3 графика обнаруженных сильных зависимостей: a – зависимость количественного параметра от количественного параметра; $б$ – зависимость денежного параметра от денежного параметра; $в$ – зависимость количественного параметра от денежного параметра.

Интерпретация полученных результатов.

Опыт показывает, что нельзя оставлять без внимания характеристики, которые были исключены по причине наличия сильной линейной связи. Чтобы убедиться в правомочности исключения, можно выполнить графическое представление изменения исключенных параметров по товарам. Поведение

сильнокоррелирующих параметров проиллюстрировано на рис. 7 (на оси x – номера товара).

Рис. 7, a демонстрирует поведение количественных параметров с сильной корреляцией (кривые 1, 2 и 3 соответствуют значениям количественных параметров под номерами 3, 4, 5 соответственно); рис. 7, $б$ – сильнокоррелирующие денежные параметры (кривые 4, 5 и 6 соответствуют значениям параметров под номерами 1, 2 и 6 соответственно). Из рис. 7 виден линейный характер зависимости параметров друг от друга по каждому номеру товара (в каждой точке кривых – для каждого товара). Распространенный пример, иллюстрирующий такого рода зависимости, – анализ таких параметров, как, например, выручка и себестоимость (понятно, что они отличаются на константу, называемую прибылью).

Отобранные слабокоррелирующие параметры (табл. 7) были переданы для оценки экспертной группе. В ее состав входят маркетологи с большим опытом работы, которые как раз при выполнении анализа товаров столкнулись с проблемой избыточности характеристик. Результаты оценки подразумевают ответ на вопрос: «Достаточно ли

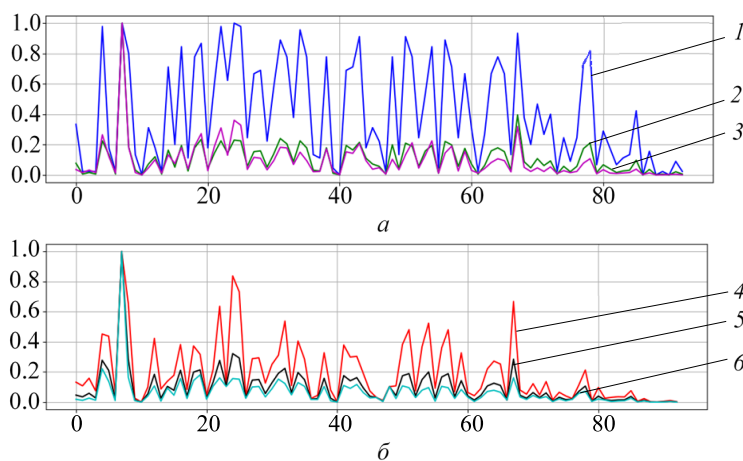


Рис. 7

Таблица 7

Номера параметров	5	7	8	9	10	11
5	1	0.28	0.13	0.31	0.16	0.29
7	0.28	1	0.01	0.69	0.34	0.04
8	0.13	0.01	1	0.40	0.18	0.05
9	0.31	0.69	0.40	1	0.01	0.02
10	0.16	0.34	0.18	0.01	1	0.21
11	0.29	0.04	0.05	0.02	0.21	1

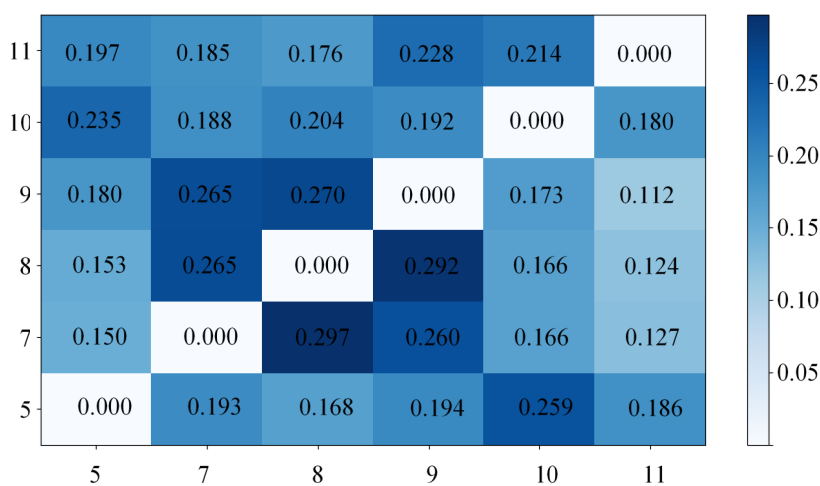


Рис. 8

проводить анализ товаров, описываемых сокращенным набором характеристик, чтобы эффективность от продаж не снижалась?» Экспертная группа подтвердила состоятельность полученного набора параметров. В табл. 7 выделены цветом параметры, для которых, по мнению авторов, значения коэффициента корреляции сильно выделяются на фоне других.

Однако в подтверждение отсутствия избыточности обнаруженного набора параметров провели чисто техническую проверку. Был проведен дополнительный анализ отобранных параметров при помощи метода машинного обучения на языке Python 3.5 (данный язык программирования был выбран по причине наличия библиотеки scikit-learn [10], где все необходимые методы машинного обучения уже реализованы, готовы к использованию).

Для проверки набора параметров воспользовались специальным классификатором ExtraTrees Classifier, построенным на ансамбле лесов случайных деревьев [10, с. 454], (рис. 8). На рис. 8 справа представлена шкала значений вектора весов. По осям x и y отложены номера параметров, на пересечении – значение веса, отражающего силу взаимосвязи параметров.

Классификатор был построен для вычисления оценки влияния параметров друг для друга. Прямая задача классификации не рассматривалась. Классификатор был построен на основе обучающей выборки, разработанной специалистами экспертной группы из тех данных, что были предварительно взяты для анализа.

Достоинством данного классификатора является способность формировать для переданного в

качестве целевого какого-то одного параметра некоторый вектор весов для других параметров, где каждый вес символизирует значимость отдельно взятого параметра по отношению к тому, что был передан на вход как целевой. Из рис. 8 видно, что отобранные параметры слабо влияют друг на друга. Такого рода оценку влияния параметров (признаков) друг на друга называют еще взаимной информативностью признаков.

Построенный классификатор относится к семейству методов машинного обучения, называемых Feature Selection – отбор признаков. К сведению, выделяют другое семейство методов машинного обучения – Feature Engineering – более трудоемкий процесс экстракции признаков, зависящий в подавляющем большинстве от квалификации эксперта.

В настоящей статье рассмотрена одна из граний проблемы автоматизации процесса управления розничными продажами – проблема представления данных большой размерности. Показано, что данные, описываемые большим числом характеристик, можно представлять в виде, пригодном для применения методов машинного обучения, и при этом без потери эффективности решения прикладных задач.

В качестве метода предобработки данных был исследован один из самых популярных – корреляционный анализ. Корреляционный анализ поз-

волил сократить размерность пространства признаков с 12 шт. до 6.

На основе результатов корреляционного анализа были исключены сильнокоррелирующие параметры. Полученные сильные корреляционные зависимости подтверждены вычислением корреляционной правки, значения которой показали, что полученные сильные линейные зависимости не случайны.

Для анализа слабокоррелирующих параметров был построен классификатор на основе ансамбля лесов деревьев решений. Классификатор показал, что в параметрах со слабой корреляционной связью нет других сильных зависимостей. Результаты анализа были проверены экспертной группой, которая подтвердила их состоятельность. Экспертная группа состояла из сотрудников организации, которая предоставляла данные для анализа.

Будущие задачи для развития темы – более глубокое исследование отобранных параметров с целью использования их в автоматизации процесса формирования ассортимента. Стоит отметить, что те параметры, которые были отброшены в ходе анализа, следует рассмотреть отдельно, чтобы убедиться в том, что они действительно не несут в себе полезной для автоматизации процесса формирования ассортимента информации.

СПИСОК ЛИТЕРАТУРЫ

1. Klionskiy D. M., Chernokulsky V. V., Razmochaeva N. V. The Investigation of Machine Learning Methods in the Problem of Automation of the Sales Management Business-process // Proc. of the Intern. Conf. «Quality Management, Transport and Information Security, Information Technologies» (IT&QM&IS-2018), St. Petersburg, Sept. 24–30 2018. P. 376–381. URL: <https://ieeexplore.ieee.org/document/8525008> (дата обращения 19.04.2018).

2. Mikhailov Y. I., Razmochaeva N. V. The Problems of Quality Management Automation in Retail Sales // Proc. of the Intern. Conf. «Quality Management, Transport and Information Security, Information Technologies» (IT&QM&IS-2018), St. Petersburg, Sept. 24–30 2018. P. 372–375. URL: <https://ieeexplore.ieee.org/document/8525056> (дата обращения 19.04.2018).

3. Анализ данных и процессов: учеб. пособие / А. А. Баргесян, М. С. Куприянов, И. И. Холод и др. 3-е изд., перераб. и доп. СПб.: БХВ-Петербург, 2018. 512 с.

4. Семенов В. П., Чернокульский В. В., Размочаева Н. В. Программное приложение для оптимизации розничных продаж // XXI Междунар. конф. по мягким вычислениям и измерениям (SCM'2018), Санкт-Петербург, 23–25 мая 2018 г. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2018. С. 468–471.

5. Размочаева Н. В., Клионский Д. М., Чернокульский В. В. Автоматизация бизнес-процессов в торговле с использованием методов интеллектуального анализа данных // Качество. Инновации. Образование». 2018. № 5. С. 77–85.

6. Чернокульский В. В., Размочаева Н. В. Разработка подхода к решению задачи формирования ассортимента товаров точки розничной торговли // Изв. СПбГЭТУ ЛЭТИ. 2018. № 2. С. 5–10.

7. Семенов В. П., Чернокульский В. В., Размочаева Н. В. Исследование искусственного интеллекта в задачах управления розничной торговлей // II Междунар. науч. конф. по проблемам управления в технических системах (CTS'2017), Санкт-Петербург, 25–27 окт. 2017 г. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2017. С. 346–349.

8. Semenov V. P., Chernokulsky V. V., Razmochaeva N. V. Research of artificial intelligence in the retail management problems // IEEE II Intern. Conf. on Control in Technical Systems (CTS'2017), St. Petersburg, 2017. P. 333–336. URL: <https://ieeexplore.ieee.org/document/8109560/> (дата обращения 19.04.2018).

9. Елисева И. И., Юзбашев М. М. Общая теория статистики: учеб. / под ред. И. И. Елисеевой. 4-е изд., перераб. и доп. М.: Финансы и статистика, 2002. 480 с.

10. Инструкция по использованию библиотеки. learn.org/0.19/_downloads/scikit-learn-docs.pdf (дата
scikit-learn user guide. Release 0.19.2. URL: https://scikit- обращения 19.04.2018).

N. V. Razmochaeva, D. M. Klionskiy
Saint Petersburg Electrotechnical University «LETI»

DATA PRESENTATION PROBLEM INVESTIGATION USING CORRELATION ANALYSIS IN RETAIL SALES MANAGEMENT PROCESS AUTOMATION TASK

One side of the problem of retail sales management process automating aspects, namely, the problem of presenting data of a large dimension is considered. The problem is that a large number of characteristics (parameters) are used to describe the data. One of the most popular approaches to reducing the dimension of feature space is correlation analysis are researched. The results of the correlation analysis application for data obtained as a result of the retail process through automated systems (vending machines) are presented. Parameters with strong dependencies were excluded by correlation analysis (in such parameters has a strong linear relationship). According to the calculated correlation correction, it was found that the detected strong linear relationship is not accidental. The results of the correlation analysis are confirmed by constructing a classifier based on an ensemble of decision trees forests. The classifier showed that there are no other strong dependencies in the parameters with a weak correlation connection. The results of the analysis were checked by an expert group (a group of marketing specialists from the retail industry), which confirmed their consistency.

Data presentation, data analysis, data interpretation, correlation analysis, process automation, retail sales, management, machine learning

УДК 37.014

А. И. Алтухов, В. И. Билан, М. А. Чебурков
Военно-космическая академия им. А. Ф. Можайского

Методика индивидуального оценивания остаточного уровня обученности по дисциплинам профессионального цикла при подготовке инженерных кадров

Предложена методика индивидуального оценивания остаточного уровня обученности по дисциплинам профессионального цикла при подготовке инженерных кадров. Подготовка инженерных кадров направлена на формирование у обучающихся определенного набора знаний, умений и навыков при изучении ими отдельных дисциплин или цикла дисциплин. Завершающим этапом изучения дисциплины является промежуточная аттестация, в ходе которой проверяется уровень сформированности компетенций дисциплины. Одним из проблемных вопросов, возникающих в ходе образовательного процесса, является снижение уровня обученности в связи с процессом естественного забывания обучающимся полученной информации. В основу предлагаемой методики положен математический аппарат, позволяющий производить расчеты значений уровня обученности в заданный момент времени на основе результатов промежуточной аттестации в зависимости от длительности интервала времени, прошедшего от момента ее проведения. Приведены результаты апробации методики на основе полученных экспериментальных данных. Результаты индивидуального оценивания остаточных знаний коррелируют с теоретическими выкладками. Предлагаемая методика позволяет определять остаточный уровень знаний, умений и навыков обучающихся в любой момент времени на протяжении всего срока обучения.

Подготовка специалистов, уровень обученности, компетенция, задачи по предназначению, дисциплины профессионального цикла, реальные условия выполнения задач

Среди задач, решаемых посредством автоматизированных обучающих систем, особое внимание уделяется контролю уровня обученности обучающихся с целью его оценивания и при необходимости его своевременного восполнения. При завершении изучения дисциплины или цикла