

# ИНТЕГРАЦИЯ МНОЖЕСТВЕННОЙ РЕГРЕССИИ В АЛГОРИТМЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ, ДЛЯ ПОСТРОЕНИЯ ПОДСИСТЕМ ПРЕДИКТИВНОЙ АНАЛИТИКИ, И ПРИНЯТИЯ НА ИХ ОСНОВЕ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ

INTEGRATION OF MULTIPLE REGRESSION INTO DATA MINING ALGORITHMS TO BUILD PREDICTIVE ANALYTICS SUBSYSTEMS AND MAKE MANAGEMENT DECISIONS BASED ON THEM

V. Zinovev  
A. Nesterov

*Summary:* A study of the applicability of multiple regression, in the framework of building algorithms for data mining, for making management decisions. A study was conducted on the subject of developing mechanisms for integrating multiple regression analysis into adaptive data mining models. The principle of applying multiple regression in the framework of building models of data mining is substantiated. This work was supported by a grant from the Russian University of Economics. G.V. Plekhanov on the topic «Intelligent system for the analysis of satellite data to predict the economic consequences of the dynamics of the global distribution of drinking water reserves and fire hazard».

*Keywords:* multiple regression, regression analysis, data mining, hybrid intelligent system.

**Зиновьев Владимир Иванович**

Аспирант, МГПУ ИЦО, г. Москва  
legrang@live.com

**Нестеров Андрей Владимирович**

Д.ф.-м.н., профессор, МГПУ ИЦО, г. Москва  
andrenesterov@yandex.ru

*Аннотация:* Проведено исследование применимости множественной регрессии, в рамках построения алгоритмов интеллектуального анализа данных, для принятия управленческих решений. Проведено исследование на предмет разработки механизмов интеграции множественного регрессионного анализа в адаптивные модели интеллектуального анализа данных. Обоснован принцип применения множественной регрессии, в рамках построения моделей интеллектуального анализа данных. Работа поддержана грантом РЭУ им. Г.В. Плеханова по теме «Интеллектуальная система анализа спутниковых данных с целью прогнозирования экономических последствий динамики глобального распределения запасов питьевой воды и пожарной опасности».

*Ключевые слова:* множественная регрессия, регрессионный анализ, интеллектуальный анализ данных.

## Введение

С каждым днем обработка данных становится все более актуальной потребностью, так как в современном мире количество информации растет в геометрической прогрессии. Приблизительные расчеты большинства аналитических компаний показали, что каждый год количество информации в планетарном масштабе увеличивается в среднем в 2 раза. При этом 90% всех существующих данных сформированы за последние несколько лет. Так, в 2005 году приблизительный объем данных, созданных человечеством, равнялся 0,15 36 (1 Зеттабайт =  $10^{21}$  байтов). По прогнозам, к 2025 году эта цифра увеличится приблизительно в 1000 раз, и составит  $150 \cdot 10^{21}$  байтов. Это означает, что поиск необходимых релевантных данных с каждым годом становится все более нетривиальной задачей, к тому же усложняющейся возрастающим уровнем информационного зашумления. Обнаружение данных (англ. Data Mining)

становится все более ориентировано на формы интеллектуального анализа данных (далее ИАД), которые позволяют выявить ранее неизвестные закономерности, определяют структуру и правила [1]. В классическом понимании, ИАД это поиск, исследование и обнаружение машинными алгоритмами необходимых наборов данных, из больших массивов информации (зачастую слабо связанных, или не связанных совсем).

## Исследование применимости множественной регрессии, в рамках построения алгоритмов интеллектуального анализа данных

Анализ данных можно разделить на косвенный и прямой. Результатом косвенного анализа выступают классификация данных, и дальнейшая их декомпозиция на кластеры (кластеризация). Результатом прямого анализа является непосредственное прогнозирование результатов.

Существует множество методов ИАД. Одним из них является метод множественного регрессионного анализа (далее МРА). Общая концепция МРА состоит в выявлении связей между искомой (зависимой) переменной, и N количеством регрессоров или предикторов. Ограничение метода МРА заключается в обнаружении числовых зависимостей, без привязки к причинам возникновения явления [2].

Простейший вид МРА - это линейная регрессия, при которой зависимость прогнозируемого фактора от остальных является линейной

$$y = \sum_{i=1}^n a_i x_i + \varepsilon = a_1 x_1 + a_2 x_2 + \dots + a_n x_n + \varepsilon.$$

Для построения статистической модели, входные данные должны удовлетворять ряду требований: выборка должна быть достаточного объема, должна отсутствовать мультиколлинеарность факторов регрессии, остатки должны удовлетворять условию гомоскедастичности (однородности для временных рядов).

Все эти условия, расчетный алгоритмический блок, проводящий МРА, должен проверять автоматически, так же автоматически должны проверяться гипотезы о распределении остатков (нормальное, треугольное, равномерное, дискретное и т.д.) с помощью соответствующих критериев [3].

После получения функций распределения прогнозных величин, их можно использовать при игровых методах принятия решений (определение вероятностей состояния природы); в моделях нечетких выводов; при нечеткой кластеризации, классификации, при построении нечетких деревьев решений, при использовании нечетких нейронных сетей [4].

Кажущаяся простота МРА нивелируется увеличением количества параметров, а также изменением вариативности периода анализа. Т.е. чем объемнее диапазон параметров, и шире период анализируемых метрик, тем сложнее становятся расчеты множественной корреляции в рамках прогнозируемого параметра. На определенном этапе предиктивной аналитики становится малоэффективным использование стандартных методов, а использование более сложных алгоритмов ИАД приводит к пропорциональному усложнению расчетов и повышению уровня компликативных деформаций в оценке функциональной зависимости усредненных значений результативного признака от факторных параметрических признаков. В данном контексте происходит повышение коллинеарности между независимыми переменными, а также повышению уровня гетероскедастичности, что приводит к увеличению погрешности измерений. МРА позволяет построить предиктивную мо-

дель с большим количеством факторов, и помимо этого определить степень влияния каждого из факторов обособленно, а также их совокупного воздействия на моделируемый критерий [5].

Таким образом, МРА в структуре ИАД является одним из фундаментальных алгоритмов, позволяющим строить на своей основе более сложные модели, а также использовать их в полноценных системах принятия решений.

**Разработка механизмов интеграции множественного регрессионного анализа в адаптивные модели интеллектуального анализа данных**

В качестве ядра интеллектуального анализа данных выступают алгоритмы интеллектуального анализа, решающие спектр основополагающих задач классификации (поиска и противопоставления зависимого параметра, нескольким независимым), развертки ассоциативных правил (поиска и установки основного ассоциативного набора среди множества подобных).

МРА позволит на основе имеющихся наблюдений  $X = \{(x_1, x_2, \dots, x_n)_j\}, j = 1, 2, \dots, N$ , где n - число одновременно измеряемых величин (факторов), N- число измерений, произвести прогнозирование значений анализируемых зависимых факторов [6].

При интеграции модели МРА в ИАД, учитываем особенности представления зависимых факторов (в дальнейшем будем использовать один зависимый фактор), как функции от независимых факторов, на которую накладываются случайные (или неучтенные) факторы  $y = f(x_1, x_2, \dots, x_n, \varepsilon)$ , где  $y$  - зависимая величина,  $(x_1, x_2, \dots, x_n)$  - факторы регрессии,  $\varepsilon$  - случайные (неучтенные) факторы.

Это означает, что основная функциональная роль алгоритмов МРА в системе комплексных алгоритмов ИАД – это восстановление функциональной зависимости и характеристик случайной величины по известным наблюдениям (выборке) X.

В дальнейшем эксперименте будем иметь дело с линейной регрессией, при которой зависимость  $y$  от  $(x_1, x_2, \dots, x_n)$  линейная

$$y = \sum_{i=1}^n a_i x_i + \varepsilon = a_1 x_1 + a_2 x_2 + \dots + a_n x_n + \varepsilon,$$

в которой нам необходимо получение уравнения тренда

$$y = \sum_{i=1}^n a_i x_i = a_1 x_1 + a_2 x_2 + \dots + a_n x_n,$$

доверительного интервала для прогнозируемой величины. Основные методы определения коэффициентов регрессии  $a_i, i = 1, \dots, n$  описаны в литературе [7].

В процессе интеграции, исследуя остатки и изучая статистические характеристики величины  $\mathcal{E}$ , а так же задав доверительную вероятность, необходимо построить доверительный интервал для прогнозируемой величины  $\mathcal{Y}$ .

Зачастую этих данных вполне достаточно для задачи принятия решения. Однако в случае сложных, нетривиальных задачах принятия решений, так же можно использовать результаты МРА для построения комплексных алгоритмов ИАД, что характерно и обосновано в случаях, когда в основе принятия решения лежат игровые методы, либо методы нечеткой логики [8]. Для задач принятия решений с использованием игровых методов, или методов нечеткой логики, необходимо восстанавливать более детальные характеристики прогнозируемой величины, такие, как плотность распределения, после чего можно оценивать вероятность состояния природы, или степень соответствия той или иной функции принадлежности.

Для реализации механизма интеграции МРА в адаптивные модели ИАД, определим основной концепт структуры проводимого анализа, который будет состоять из последовательных алгоритмов, позволяющих изменить парадигму предиктивных моделей на уровне, предшествующем построению моделей классификации (См. Рисунок 1).

Разработанный концепт содержит следующие основные структурные элементы:

1. «Входные параметры» – параметры задачи, которой необходимо экспертное решение.
2. «Выходные параметры» – формализованное экспертное решение, предполагающее: максимальный уровень удовлетворения решению задачи; максимальную степень возможности реализации; минимальное количество рисков (финансовых, технологических, экологических и др.).
3. «Воздействие факторов внешней среды» – микро и макропоказатели (организации, инфраструктуры, климатические, экономические, транспортные и др.).
4. «Обратная связь» – информация различной формы и характера, поступающая от блока «приемника» результата принятия решения (пользователь системы, программный модуль системы), к входным параметрам, для их корректировки или загрузки на повторную обработку системой, в случае недостаточной достоверности или недостаточного соответствия требуемым результатам [9].
5. «Процесс принятия решения» – включает в себя следующие организационные компоненты:
  - 5.1. «Интерпретатор взаимодействия» – представляет собой подсистему взаимодействующих между собой модулей, предназначенных для обработки входящих данных, их преобразования, кодирования, интерпретации, подготовки информационного обмена с интеллектуальной базой данных, комплексной интеграции с моделями принятия решений. Таким образом интерпретаторы взаимодействия являются буфером, позволяющим минимизировать вре-

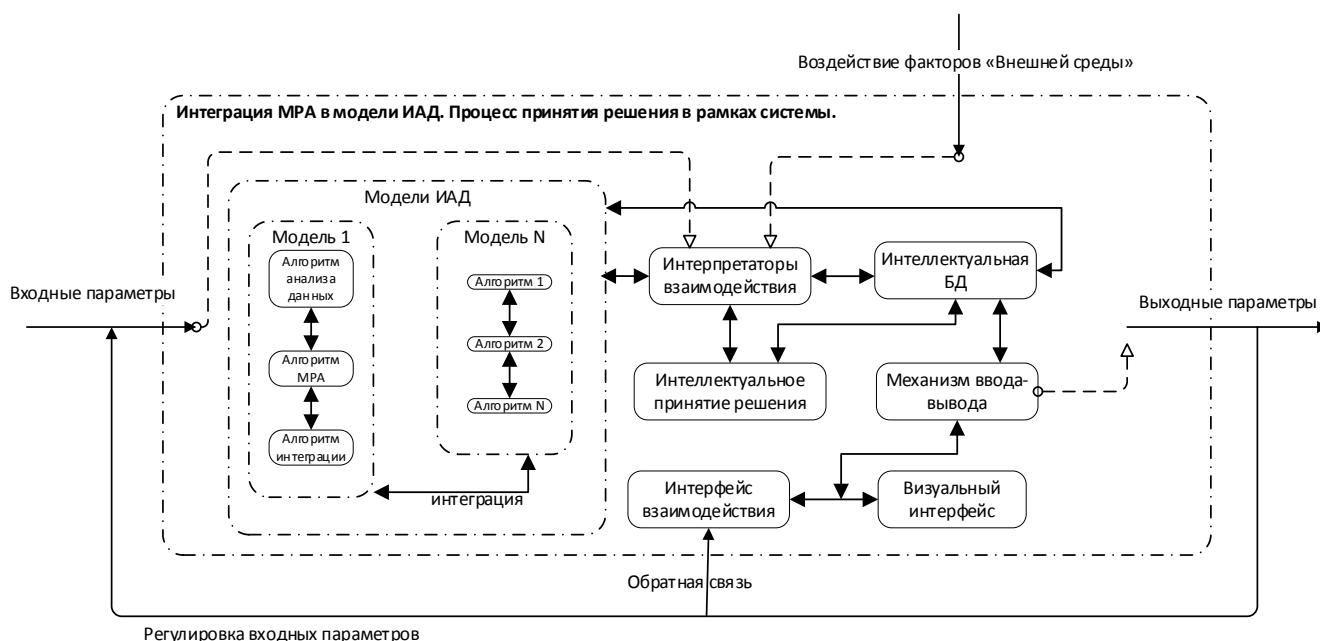


Рис. 1. Концепт интеграции МРА в адаптивную модель ИАД

менные издержки, а также максимизировать уровень валидности блоков обработанных и подготовленных данных [10].

- 5.2. «Интеллектуальная БД» – интеллектуальная база данных, предназначенная для хранения, обработки, и предоставления метазнаний [11], правил, процедур, а также моделей принятия решений.
- 5.3. «Модели ИАД» (модели принятия решения) – содержит набор упорядоченных алгоритмов, процедур, триггеров и представлений, сформированных по целевому назначению в обособленные модели [12], предназначенные для максимально качественной и быстрой обработки поступающих с интерпретатора данных заложенными в них различными методами [13]. В зависимости от комбинации интегрированных алгоритмов, позволяющих как формировать качественные параметрические концепции принятия решений, так и приводить к неполноценным результатам, на данном этапе необходима качественная аналитика, систематизация требований, и помимо этого четко прослеживаемая систематизация исходных параметров [14]. *Именно в этом сегменте концепта, интегрируем алгоритм МРА в общую структуру ИАД.*
- 5.4. «Интеллектуальное принятие решения» – подсистема, предназначенная для непосредственного принятия детерминированных и вероятностных решений, исходя из комплексного или индивидуального подхода.
- 5.5. «Механизм ввода-вывода» – компонент типовой архитектуры системы, предназначенный для возможности как ввода информации в систему, так и ее вывод в необходимой интерпретации [15].
- 5.6. «Интерфейс взаимодействия» – модуль, отвечающий за непосредственное взаимодействие системы принятия решений и конечного пользователя или интеграционного модуля.
- 5.7. «Визуальный интерфейс» – модуль, необходимый для визуального взаимодействия с пользователем системы, и предоставляющий ему конечную информацию в виде информационных сообщений различного вида, в зависимости от первоначальных требований и целевого использования (текст, графики, диаграммы, сигналы взаимодействия, и т.д.).

Таким образом, принцип интеграции модели МРА в структуру адаптивных моделей ИАД был определен. Для дальнейшего практического примера интеграции, смоделируем общую схему, согласно которой будет происходить процесс (См. Рисунок 2).

#### Пример процесса практической интеграции модели МРА, в алгоритмы ИАД

Для наглядной демонстрации интеграции механизмов МРА и алгоритмов ИАД, проведем экспериментальный анализ корреляции курса доллара США, от N количества признаков. Далее на основании результатов корреляции будет разработана предиктивная модель курса доллара на конкретную дату.

На предварительном этапе, согласно ранее описанному алгоритму действий, загружаем подготовленный файл в разработанный прототип системы принятия решения (См. Рисунок 3)

Где: «Date» – календарная дата выборки;  
 «Cost\_dollars» – стоимость доллара на календарную дату;  
 «Cost\_brent» – стоимость нефти марки Brent;  
 «Cost\_gold» – стоимость унции золота;  
 «Inflation\_rate» – курс инфляции;  
 «Key\_rate» – ключевая ставка ЦБ РФ.

В зависимости от настроенной и выбранной категории, формируется интерактивный аналитический график, отражающий корреляцию категориального независимого фактора с исследуемым зависимым (См. Рисунок 4, Рисунок 5).

Заключительным шагом предварительного этапа является интеграция существующих признаков и модели МРА.

На первом этапе построения многофакторной регрессионной модели необходима разработка спецификации, которая включает в себя выборку факторных признаков и выбор вида уравнения регрессии. Факторы, включаемые во множественную регрессию должны иметь количественное измерение, и не должны находиться в функциональной связи, так как это повлечет за собой снижение результативности модели. В качестве интервала для сбора статистических данных анализируемых показателей выступает временной период с 13.01.2015 по 16.10.2018 год (3160 (632x5) значений). На основе проведенного теоретико-экономического анализа, в качестве используемых переменных для построения модели МРА были взяты следующие показатели:

Зависимая переменная (моделируемый показатель):  
 — «Стоимость доллара США (в рублях)»;

Независимые переменные (факторы):  
 — «Средняя Стоимость барреля нефти»;  
 — «Средняя Стоимость унции золота»;  
 — «Уровень инфляции (%)»;  
 — «Ключевая ставка ЦБ России (%)»;

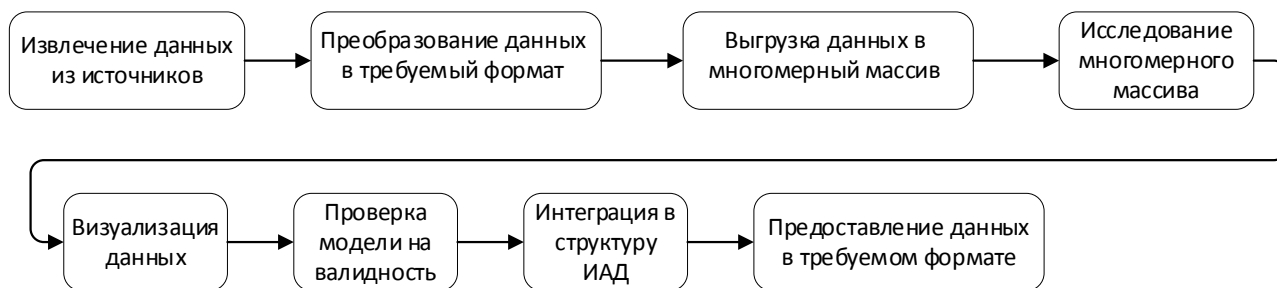


Рис. 2. Процесс практической интеграции МРА в ИАД

	Date	Cost_dollars	Cost_brent	Cost_gold	Inflation_rate	Key_rate
0	16.10.2018	65.7508	80.895	1232.15	0.30	7.5
1	12.10.2018	66.7727	80.355	1223.40	0.30	7.5
2	11.10.2018	66.1832	81.215	1212.60	0.30	7.5
3	10.10.2018	66.4032	84.030	1192.40	0.30	7.5
4	09.10.2018	66.9685	84.590	1190.85	0.30	7.5
...	...	...	...	...	...	...
627	20.01.2015	64.9732	48.630	1289.40	3.85	17.0
628	16.01.2015	64.8337	49.240	1268.75	3.85	17.0
629	15.01.2015	66.0983	48.810	1246.65	3.85	17.0
630	14.01.2015	64.8425	47.400	1235.00	3.85	17.0
631	13.01.2015	62.7363	46.240	1234.70	3.85	17.0

632 rows × 6 columns

Рис. 3. Загруженная выборка

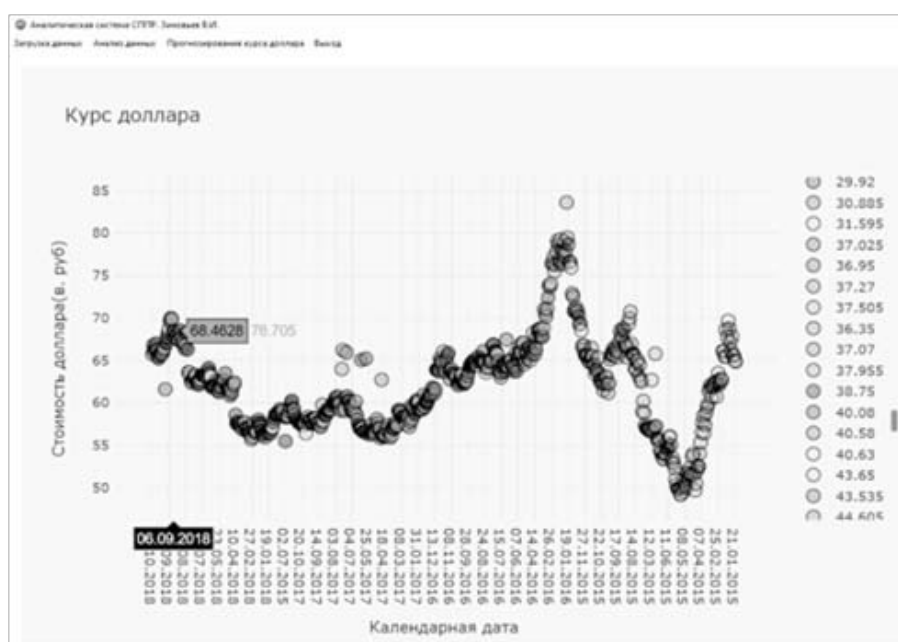


Рис. 4. Интерактивный график корреляции на примере параметров: Cost\_Dollars – Cost\_brent



Рис. 5. Интерактивный график корреляции на примере параметров: Cost\_Dollars – Key\_rate

Дата	Regr				
	Стоимость Доллара США (в рублях)	Средняя Стоимость барреля нефти	Средняя Стоимость унции золота	Уровень инфляции (%)	Ключевая ставка ЦБ России (%)
27.03.2015	56,4271	57,495	1198,65	1,21	14
26.03.2015	57,3879	58,04	1208,7	1,21	14
25.03.2015	58,771	55,935	1192,95	1,21	14
24.03.2015	59,4452	55,82	1189,9	1,21	14
20.03.2015	59,8308	54,61	1177,65	1,21	14
19.03.2015	61,3483	55,055	1165,85	1,21	14
18.03.2015	61,751	54,76	1160,95	1,21	14
17.03.2015	62,1497	53,475	1150,45	1,21	14
13.03.2015	60,9595	56,175	1155,6	1,21	15
12.03.2015	62,6797	57,95	1156,6	1,21	15
11.03.2015	60,6649	57,3	1155,15	1,21	15
06.03.2015	61,8457	60,71	1180,8	1,21	15
05.03.2015	61,8745	61,34	1202,35	1,21	15
04.03.2015	62,3649	60,965	1202,9	1,21	15
03.03.2015	62,2248	61,495	1207,75	1,21	15
27.02.2015	60,7109	61,67	1211,65	2,22	15
26.02.2015	62,5906	61,145	1213,2	2,22	15
25.02.2015	63,5083	60,305	1206,3	2,22	15
20.02.2015	62,1307	60,485	1206,5	2,22	15
19.02.2015	62,4001	59,395	1214,05	2,22	15
18.02.2015	62,8353	60,855	1207,3	2,22	15
17.02.2015	62,6632	61,635	1215,85	2,22	15
13.02.2015	66,0994	60,425	1228,95	2,22	15
12.02.2015	66,0585	57,99	1226	2,22	15
11.02.2015	65,4469	56,655	1227,6	2,22	15
10.02.2015	65,7817	58,405	1238,25	2,22	15
06.02.2015	68,6113	58,785	1248,55	2,22	15
05.02.2015	65,447	56,42	1265	2,22	15
04.02.2015	67,7727	56,825	1267,05	2,22	15
03.02.2015	69,664	57,46	1271,15	2,22	15
30.01.2015	68,7303	50,92	1271,75	3,85	17
29.01.2015	67,1506	48,81	1266,8	3,85	17
28.01.2015	67,8153	49,005	1285,15	3,85	17
27.01.2015	65,5937	48,835	1287,35	3,85	17
23.01.2015	65,4	49,175	1292,05	3,85	17
22.01.2015	65,5558	49,235	1293,45	3,85	17
21.01.2015	64,9862	48,9	1295,8	3,85	17
20.01.2015	64,9732	48,63	1289,4	3,85	17
16.01.2015	64,8337	49,24	1268,75	3,85	17
15.01.2015	66,0983	48,81	1246,65	3,85	17
14.01.2015	64,8425	47,4	1235	3,85	17
13.01.2015	62,7363	46,24	1234,7	3,85	17

Рис. 6. Прямоугольная матрица данных

В результате была определена прямоугольная матрица размерностью 632x5 (См. Рисунок 6).

Результативный признак представлен как  $Y$ .

Независимые признаки представлены как  $X_1, X_2, X_3, X_4$  (См. Рисунок 7).

При проверке на корректность данных прямоугольной матрицы было выявлено, что все 632 значения могут участвовать в построении корреляционной матрицы (См. Рисунок 8).

Далее было произведено построение корреляционной матрицы (См. Рисунок 9), на которой видно, что сильной корреляции ( $>0,7$ ) между независимыми признаками  $X_1, X_2, X_3, X_4$  не наблюдается.

Для проверки отсутствия мультиколлинеарности был произведен расчет стандартизированных коэффициентов (См. Рисунок 10), и построена матрица стандартизированных коэффициентов  $\beta_j$ , на основании данных которой сделан вывод о том, что параметры в рамках нормы, и мультиколлинеарность отсутствует (См. Рисунок 11).

Для проверки использования критериев Фишера и Стьюдента, делаем предположения относительно поведения остатков  $\varepsilon_j$ , представляющих собой независимые случайные величины со средним значением равным 0, постоянной дисперсией, и подчиняющихся нормальному распределению [16], после чего производим предварительные расчеты (См. Рисунок 12).

Далее была построена частотная гистограмма остатков, на которой несмотря на небольшие отклонения четко прослеживается симметричность значений (См. Рисунок 13)

Для более детальной оценки параметров регрессии, проверяем их на отсутствие смещения, и эффективность, для чего построен нормально-вероятностный график (См. Рисунок 14). На графике видно, что систематических отклонений фактических данных от теоретической нормальной прямой не наблюдается, и дисперсия укладывается в концепцию гомоскедастичности (для каждого значения фактора  $x_j$ , остатки  $\varepsilon_j$  имеют одинаковую дисперсию), что означает факт нормализации распределения остатков [17].

Для проверки наличия или отсутствия остатков от предсказанных значений строим диаграмму рассеивания (соотношение остатков и предсказанных значений), на которой по степени хаотичности рассеивания видно, что остатки не зависят от предиктивных значений (См. Рисунок 15).

Следующим этапом оценки приемлемости модели является определение уровня значимости  $p$ -value, которое равняется нулю (См. Рисунок 16), что говорит о том, что модель корректна, и будет работать намного лучше, чем наивный прогноз по усредненным значениям.

Факторы, включенные в множественную регрессию, должны объяснять вариативность разбросов значений независимых переменных. Таким образом, при включенных в модель факторов  $x_i = (X_1, X_2, X_3, X_4)$  становится выше коэффициент детерминации  $R^2$  ( $R^2 = 0,37573384$ ), и при этом закономерно уменьшается остаточная дисперсия [18]. При коэффициенте детерминации  $R^2 > 0,3$  можно с уверенностью утверждать, что уровень точности модели будет на достаточно высоком уровне [19].

Для проведения финального эксперимента по прогнозированию курса доллара возьмем дату 09.03.2020, более известную на финансовых биржах как «черный понедельник», отмеченный падением моделируемого показателя (нефти марки Brent) на рекордно низкие отметки. В качестве предиктивных значений будут выступать зафиксированные на 09.03.2020 показатели по заложенным в модели показателям  $x_i = (X_1, X_2, X_3, X_4)$ , где:  $X_1 = 36,950$  (См. Рисунок 17),  $X_2 = 1667,700$  (См. Рисунок 18),  $X_3 = 2,500$  (реальный средний показатель),  $X_4 = 6$  (См. Рисунок 19).

Результатом проведения эксперимента стало фактически полное совпадение прогнозируемого значения курса доллара США (Predicted = 74,3333), по заданным параметрам на указанную дату (См. Рисунок 19, Рисунок 20).

Приведем пример определения степени уверенности принятия решения, при двух возможных вариантах «покупать» или «продавать», с использованием элементов нечеткой логики [20],

Пусть каждое решение имеет функцию принадлежности (См. Рисунок 22).

По горизонтали отложен диапазон изменения курса рубля: 45 – 90. По вертикали отложены значения функции принадлежности - для решения «продать» -  $y = f_1(x)$  – график №1, для решения «купить» -  $y = f_2(x)$  – график №2.

Проанализировав остатки ряда, можно сделать вывод о том, что случайная величина – курс доллара – имеет нормальное распределение параметров [21] – случайная нормально распределенная величина  $N(74,333;5)$ . На этом же рисунке приведен график нормально распределенной величины  $y = f_3(x) = kN(74,333;5)$  – линия серого цвета (масштаб по вертикали изменен для наглядности).

1 - Стоимость Доллара США (в рублях)	1 - Стоимость Доллара США (в рублях)
2 - Средняя Стоимость барреля нефти	2 - Средняя Стоимость барреля нефти
3 - Средняя Стоимость унции золота	3 - Средняя Стоимость унции золота
4 - Уровень инфляции (%)	4 - Уровень инфляции (%)
5 - Ключевая ставка ЦБ России (%)	5 - Ключевая ставка ЦБ России (%)
6 - Ключевая ставка ФРС США	

Рис. 7. Определение зависимых и независимых параметров

```
missing data were casewise deleted

632 cases were processed
632 valid cases accepted
```

Рис. 8. Проверка корректности данных прямоугольной матрицы

Variable	Correlations (Regr in regr.stw)				
	Средняя Стоимость барреля нефти	Средняя Стоимость унции золота	Уровень инфляции (%)	Ключевая ставка ЦБ России (%)	Стоимость Доллара США (в рублях)
Средняя Стоимость барреля нефти	1,000000	0,256563	-0,187930	-0,462298	-0,419321
Средняя Стоимость унции золота	0,256563	1,000000	-0,186749	-0,358490	-0,263447
Уровень инфляции (%)	-0,187930	-0,186749	1,000000	0,678505	0,213538
Ключевая ставка ЦБ России (%)	-0,462298	-0,358490	0,678505	1,000000	0,046968
Стоимость Доллара США (в рублях)	-0,419321	-0,263447	0,213538	0,046968	1,000000

Рис. 9. Корреляционная матрица

```
Multiple Regression Results

Dependent: Стоимость Долл Multiple R = ,61297132 F = 94,34483
R?= ,37573384 df = 4,627
No. of cases: 632 adjusted R?= ,37175128 p = 0,000000
Standard error of estimate: 4,394666381
Intercept: 114,76472204 Std.Error: 3,811374 t( 627) = 30,111 p = 0,0000

Средняя Стоим b*=-,55 Средняя Стоим b*=-,26 Уровень инфля b*=-,493
Ключевая став b*=-,64
```

Рис. 10. Расчет стандартизованных коэффициентов

		Regression Summary for Dependent Variable: Стоимость Доллара США (в рублях) (Regr in regr.stw)					
		R= ,61297132 R?= ,37573384 Adjusted R?= ,37175128					
		F(4,627)=94,345 p<0,0000 Std Error of estimate: 4,3947					
N=632		b*	Std Err. of b*	b	Std Err. of b	t(627)	p-value
	Intercept			114,7647	3,811374	30,1111	0,000000
	Средняя Стоимость барреля нефти	-0,554796	0,036435	-0,2643	0,017356	-15,2271	0,000000
	Средняя Стоимость унции золота	-0,257096	0,034073	-0,0192	0,002545	-7,5455	0,000000
	Уровень инфляции (%)	0,492843	0,043864	4,4351	0,394728	11,2358	0,000000
	Ключевая ставка ЦБ России (%)	-0,636075	0,050067	-1,6218	0,127658	-12,7045	0,000000

Рис. 11. Матрица стандартизованных коэффициентов



```

Dependent: Стоимость Д Multiple R : ,61297192 F = 94,34483
R2: ,37573384 df = 4,627
No. of cases: 632 adjusted R2: ,37175129 p = 0,000000
Standard error of estimate: 4,394666381
Intercept: 114,76472204 Std.Error: 3,811374 t( 627) = 30,111 p < 0,0000
    
```

Рис. 12. Количественный расчет параметров остатков

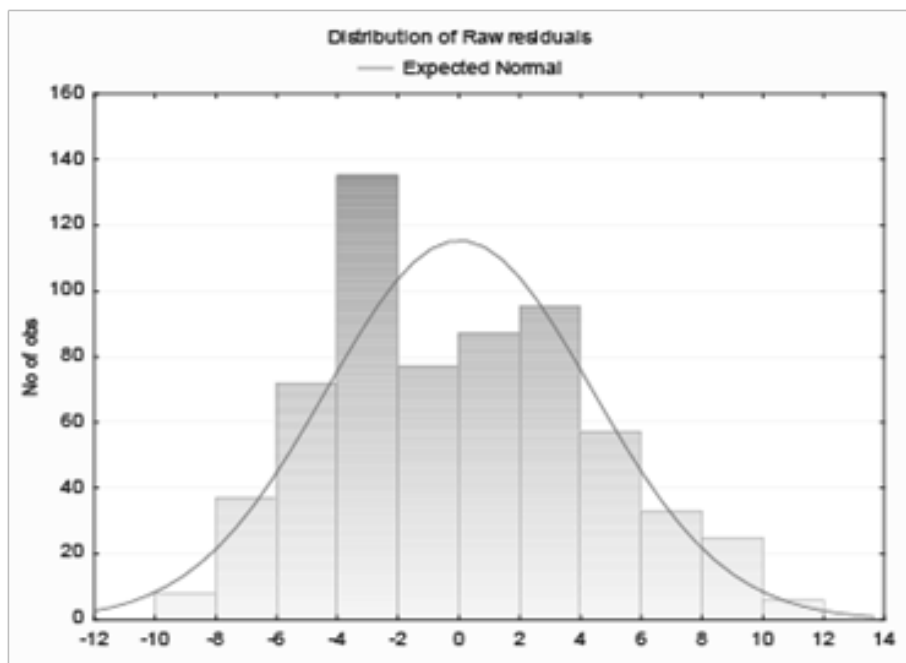


Рис. 13. Частотная гистограмма остатков

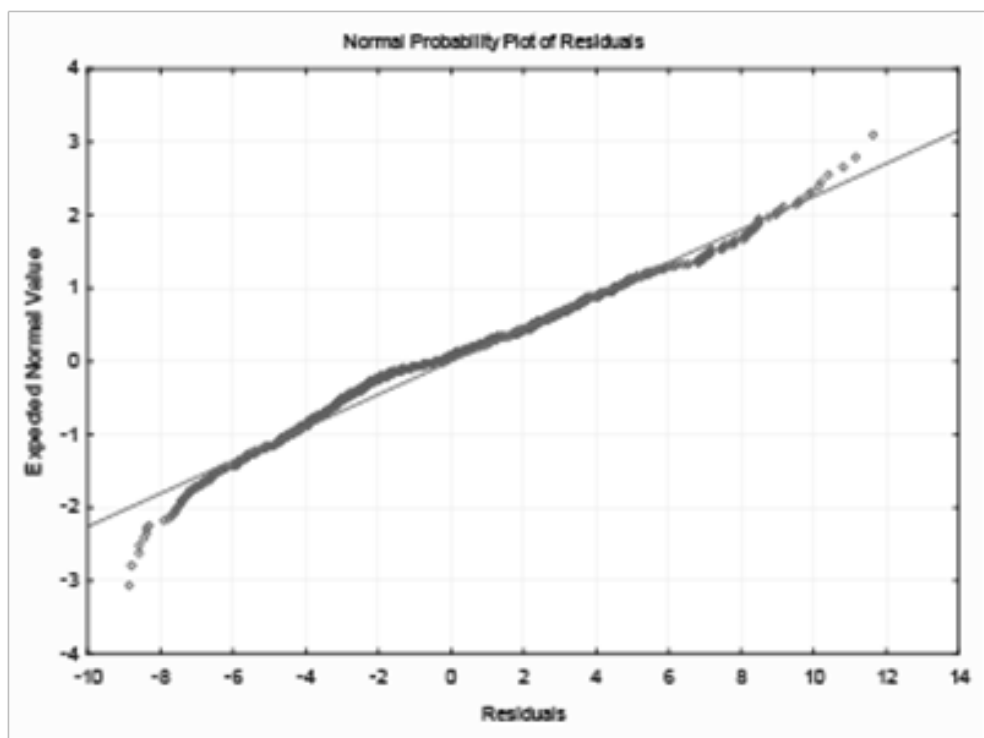


Рис. 14. Нормально-вероятностный график

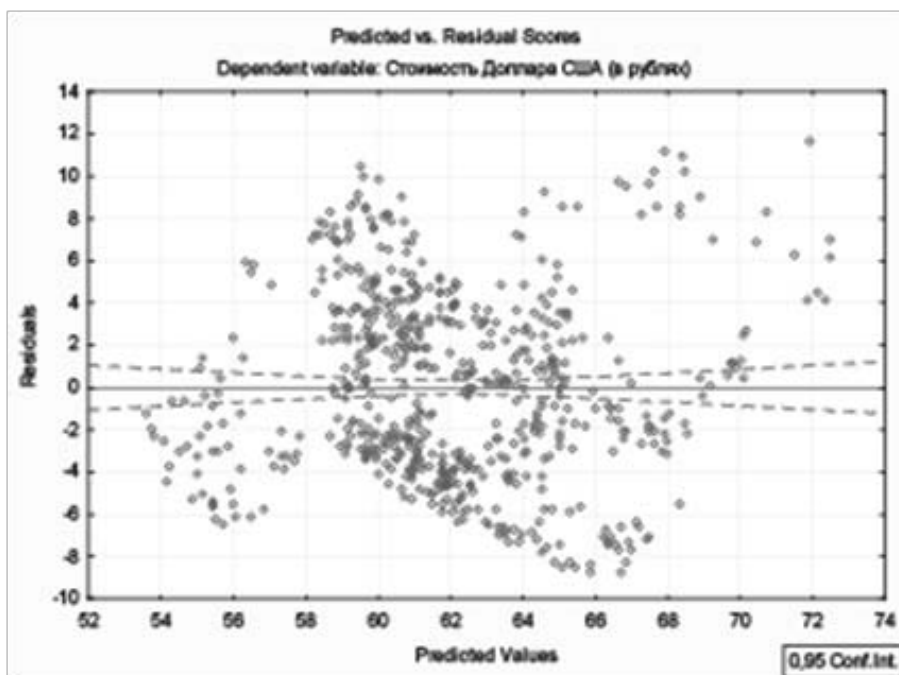


Рис. 15. Диаграмма рассеивания

Analysis of Variance; DV: Стоимость Доллара США (в рублях) (Regr in regr.stw)					
Effect	Sums of Squares	df	Mean Squares	F	p-value
Regress.	7288,36	4	1822,090	94,34483	0,00
Residual	12109,31	627	19,313		
Total	19397,67				

Рис. 16. Уровень значимости p-value



Рис. 17. Параметр  $X_1$ , (Цена за баррель нефти марки Brent по состоянию на 09.03.2020)



Рис. 18. Параметр  $X_2$  (Цена за тройскую унцию золота по состоянию на 09.03.2020)

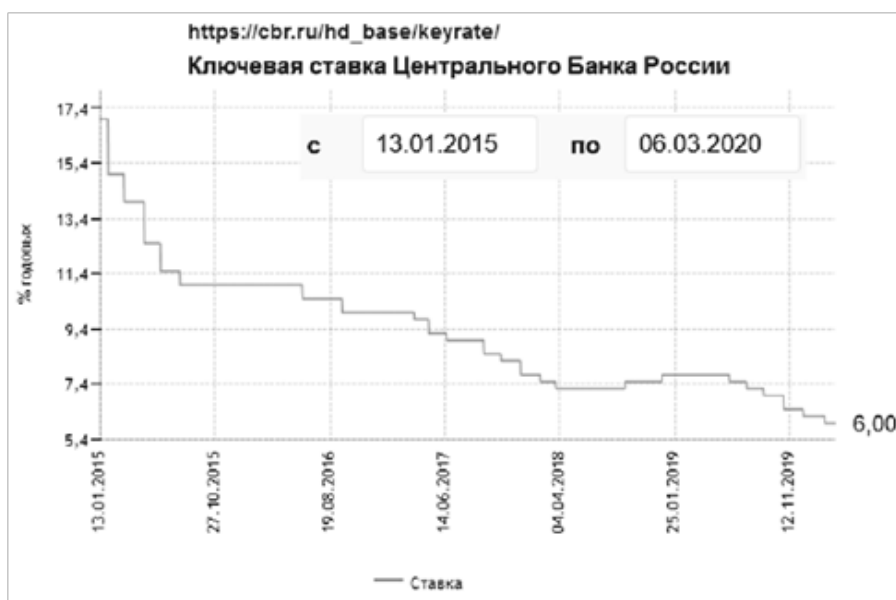


Рис. 19. Параметр  $X_4$  (Ключевая ставка ЦБ России по состоянию на 09.03.2020)

Variable	Predicting Values for (Regr in regr.stw) variable: Стоимость Доллара США (в рублях)		
	b-Weight	Value	b-Weight * Value
Средняя Стоимость барреля нефти	-0,26428	36,950	-9,7650
Средняя Стоимость унции золота	-0,01920	1667,700	-32,0232
Уровень инфляции (%)	4,43508	2,500	11,0877
Ключевая ставка ЦБ России (%)	-1,62184	6,000	-9,7310
Intercept			114,7647
Predicted			74,3333
-95,0%CL			71,0739
+95,0%CL			77,5926

Рис. 20. Прогнозируемое значение по заданным параметрам



Рис. 21. Реальное значение котировок доллара США по состоянию на 09.03.2020

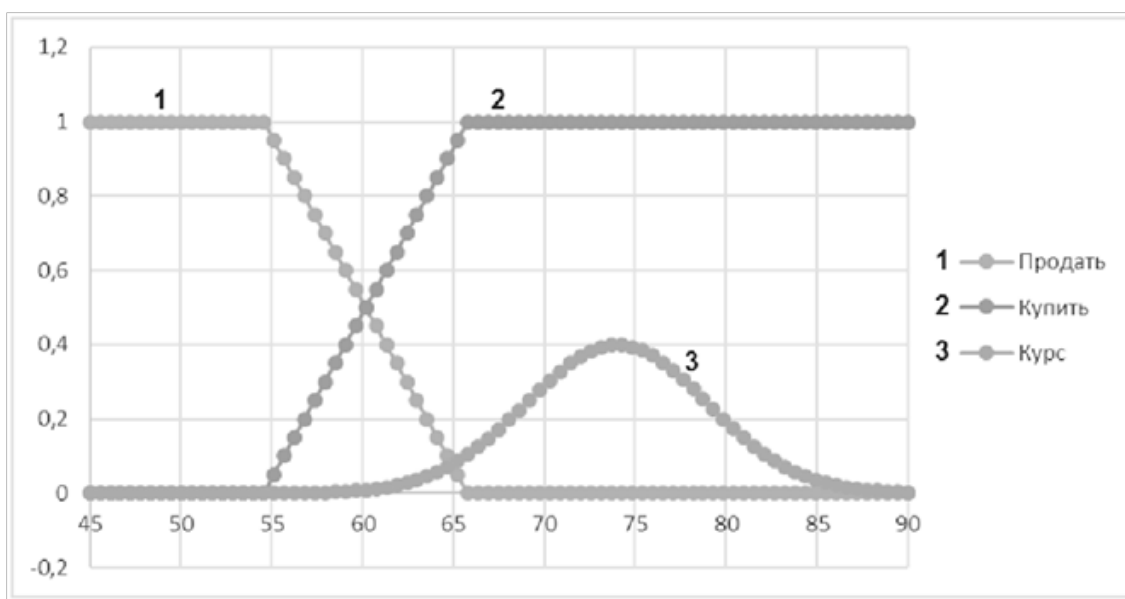


Рис. 22. Варианты принятия решения

Нормированные степени уверенности легко вычислить как [22]

$$K_1 = \int_{-\infty}^{+\infty} f_1(x)f_3(x)dx, K_2 = \int_{-\infty}^{+\infty} f_2(x)f_3(x)dx,$$

$$k_1 = \frac{K_1}{K_1 + K_2}, k_2 = \frac{K_2}{K_1 + K_2}.$$

Это дает степень уверенности принимаемого решения: продать = 0,009402; купить = 0,009402.

Таким образом, даже простая модель множественной регрессии показала свою эффективность при правильно подобранных параметрах, а также возможность интеграции в алгоритмы ИРА, для принятия решения.

### Заключение

Ранее алгоритмы МРА использовались обособленно, и не интегрировались в

сложные алгоритмы. Т.е. при необходимости проведения неких статистических исследований, определялись структурные компоненты предикторы, и формировался результат по моделируемому показателю.

По результатам предварительных исследований, а также проведенного эксперимента:

- был выявлен значительный потенциал применения МРА в более сложных комплексных алгоритмах ИАД;
- описан принцип применения МРА в ИАД;
- обоснованы методы применения МРА в ИАД;
- проведен эксперимент интеграции МРА в ИАД.

МРА является удобным инструментом в различных задачах принятия решения, и может быть использован как для простого интервального прогнозирования, так и для принятия решений на основе игровых методов, нечеткой логики, нечеткой классификации.

### ЛИТЕРАТУРА

1. Рафалович, Владимир Data mining, или Интеллектуальный анализ данных для занятых. Практический курс / Владимир Рафалович. - М.: И-Трейд, 2014. - 102 с.
2. Марманис, Х. Алгоритмы интеллектуального Интернета. Передовые методики сбора, анализа и обработки данных / Х. Марманис, Д. Бабенко. — М.: Символ, 2018. — 480 с.
3. Дрейпер, Н. Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. - М.: Вильямс, 2016. - 912 с.
4. Рутковская, Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский. - М.: РиС, 2013. - 384 с.
5. Маккинни, Уэс Python и анализ данных / Уэс Маккинни. - М.: ДМК Пресс, 2013. - 663 с.
6. Син, Такахаси Занимательная статистика. Регрессионный анализ. Манга / Такахаси Син. - М.: ДМК Пресс, 2017. — 239
7. Смирнов, В. Д. Методы корреляционно-регрессионного анализа в эконометрических исследованиях: учебное пособие / В.Д. Смирнов. - М.: Белокопытов Алексей Вячеславович, 2018. - 951 с.
8. Пегат А. Нечеткое моделирование и управление / пер. с англ. А.Г. Подвесовского и Ю.В. Тюменцева; под ред. Ю.В. Тюменцева. — М.: БИНОМ. Лаборатория знаний, 2009. — 798 с.
9. Бритков В. Б. Методическое пособие по курсу «Системы поддержки принятия решений»; Ленанд - Москва, 2006. - 972 с.
10. Трофимова Л. А., Трофимов В. В. Методы принятия управленческих решений. Учебник; Юрайт - Москва, 2014. - 336 с.
11. Ромашкова О.Н., Пономарева Л.А. Модель эффективного управления объединенной образовательной системой (структурой) // В книге: Новые информационные технологии в научных исследованиях материалы XXII Всероссийской научно-технической конференции студентов, молодых ученых и специалистов. Рязанский государственный радиотехнический университет. 2017. С. 16-18.
12. Ромашкова О.Н., Яковлев Р.И. Анализ моделей и методов для оценки живучести инфокоммуникационных сетей в условиях чрезвычайных ситуаций // Т-Сотт: Телекоммуникации и транспорт. 2012. Т. 6. № 7. С. 165-170.
13. Трофимова Л. А., Трофимов В. В. Методы принятия управленческих решений. Учебник; Юрайт - Москва, 2014. - 336 с.
14. Юдин Д. Б. Вычислительные методы теории принятия решений; Либроком - Москва, 2013. - 320 с.
15. Мендель А. В. Модели принятия решений; Юнити-Дана - Москва, 2010. - 464 с.
16. Соколов, Г.А. Введение в регрессионный анализ и планирование регрессионных экспериментов в экономике / Г.А. Соколов. - М.: ИНФРА-М, 2017. - 109 с.
17. Кабаков, Р. Р в действии. Анализ и визуализация данных в программе R / Р. Кабаков. — М.: ДМК, 2016. — 588 с.
18. Кулаичев, А.П. Методы и средства комплексного анализа данных: Учебное пособие / А.П. Кулаичев. — М.: Форум, НИЦ ИНФРА-М, 2016. — 512 с.
19. Лесковец, Ю. Анализ больших наборов данных / Ю. Лесковец, А. Раджараман. — М.: ДМК, 2016. — 498 с.
20. Круглов В.В., Длин М.И., Голунов Р.Ю. Нечеткая логика и искусственные нейронные сети — М.: Физматлит, 2001. — 224 с.
21. Горяинова, Е.Р. Прикладные методы анализа статистических данных: Учебное пособие / Е.Р. Горяинова, А.Р. Панков, Е.Н. Платонов. — М.: ИД ГУ ВШЭ, 2018. — 310 с.
22. Боровиков, В.П. Популярное введение в современный анализ данных в системе STATISTICA: Учебное пособие для вузов / В.П. Боровиков. — М.: Гор. линия-Телеком, 2018. — 288 с.

© Зиновьев Владимир Иванович (legrang@live.com), Нестеров Андрей Владимирович (andrenesterov@yandex.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»