

На правах рукописи

ДЕНИСЕНКО Андрей Сергеевич

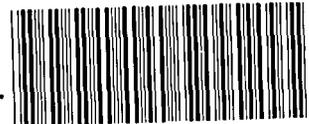
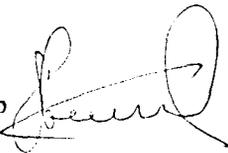
**ИНФОРМАЦИОННАЯ СИСТЕМА ОБРАБОТКИ ДАННЫХ
ФИНАНСОВОГО МОНИТОРИНГА В ОБЛАСТИ БИОРЕСУРСОВ**

Специальность 05.13.01 – системный анализ,
управление и обработка информации
(в информационных системах)

31 МАЙ 2017

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Автор



006657249

Москва
2017

Работа выполнена в Национальном исследовательском ядерном университете «МИФИ»

Научный руководитель:	Крылов Григорий Олегович, доктор физико-математических наук, профессор кафедры финансового мониторинга НИЯУ МИФИ
Официальные оппоненты:	Росс Геннадий Викторович, доктор технических наук, профессор кафедры «Системный анализ в экономике» ФГБОУ ВО «Финансовый университет при Правительстве Российской Федерации» Петровский Михаил Игоревич, кандидат физико-математических наук, доцент кафедры автоматизации систем вычислительных комплексов МГУ им. М.В. Ломоносова
Ведущая организация:	Федеральное государственное автономное научное учреждение «Центр информационных технологий и систем органов исполнительной власти» (ФГАНУ ЦИТИС)

Защита диссертации состоится «28» июня 2017 года в 17 часов 30 минут на заседании диссертационного совета Д 212.130.03 на базе Национального исследовательского ядерного университета «МИФИ» по адресу: 115409, г. Москва, Каширское шоссе, д. 31.

С диссертацией можно ознакомиться в библиотеке Национального исследовательского ядерного университета «МИФИ» и на сайте <http://ods.mephi.ru>.

Отзывы на автореферат в двух экземплярах, заверенные печатью, просьба направлять по адресу: 115409, г. Москва, Каширское шоссе, д. 31, диссертационные советы НИЯУ МИФИ (тел.: +7 (499) 324-84-98).

Автореферат разослан «13» мая 2017 года

Ученый секретарь
диссертационного совета Д 212.130.03
доктор технических наук, доцент

 Леонова Н.М.

Общая характеристика работы

Актуальность диссертационного исследования. Противодействие отмыванию преступных доходов является важной задачей финансового мониторинга¹, решение которой основано на использовании больших (несколько терабайт) информационных массивов гетерогенных данных. Эти массивы поступают в информационную систему (ИС) финансового мониторинга. В целях оценки обстановки (ситуации), являющейся исходным и важнейшим этапом принятия решений, в ИС осуществляется обработка поступающих данных, что позволяет представить субъекты векторами показателей. Суть оценки обстановки заключается в свёртке показателей в единую рейтинговую оценку, характеризующую меру девиантной деятельности субъекта. Существующий в финансовом мониторинге подход к оценке обстановки заключается в том, что эксперты назначают компонентам вектора весовые коэффициенты и складывают полученные результаты. При этом, в соответствии со сложившейся практикой эксперты осуществляют подготовку информации об обстановке на естественном языке.

Однако, получаемые таким способом результаты сопровождаются экспертным субъективизмом и социально-политической мотивацией экспертов. Кроме того, возрастающий объём поступающих данных (приблизительно на 20% ежегодно) приводит к снижению оперативности обработки данных на этапе оценки обстановки. Руководство Росфинмониторинга вынуждено работать с субъективными оценками и сталкиваться с существенно увеличенными сроками получения результата.

В этих условиях возникает противоречие между потребностью руководства Росфинмониторинга и существующей практикой обработки данных, сопровождающейся большими временными и ресурсными затратами, а также экспертным субъективизмом. Данное противоречие связано с недостаточным уровнем системности существующего подхода и может быть устранено путём разработки комплексной методики автоматизации процесса оценки обстановки в ИС финансового мониторинга, основанной на алгоритмах решения частных задач оценки обстановки с учётом их специфики.

Таким образом, исследование и разработка научно-технических решений обработки данных финансового мониторинга для оценки обстановки, исключающих экспертный субъективизм и удовлетворяющих критерию к временным и ресурсным показателям процесса обработки данных, является актуальной задачей.

Разработке и исследованию методов и моделей обработки данных посвящены работы С.А. Айвзяна, М.Г. Завельского, Г.Б. Клейнера, А.Н. Колмогорова, Т.И. Овчинникова, Н.К. Разумовского, В.Я. Райцина, А.Ф. Филиппова, Л.М. Чёрного, Г.О. Крылова и других учёных. Большой вклад в системный анализ процессов в прикладных областях внесли труды

¹ В соответствии с Федеральным законом от 07.08.2001 №115-ФЗ.

зарубежных учёных Р. Кини, М. Месаровича, П. Ромер, Х. Райфа, Т. Саати, Р. Солоу, И. Такахары, Дж. Айтчисона, Дж. Брауна и других.

Между тем, исследование и разработка научно-технических решений обработки данных финансового мониторинга для оценки обстановки, позволяющих исключить экспертный субъективизм и повысить оперативность получения результатов, другими исследователями ранее не осуществлялись.

В соответствии с функцией предпочтения лиц, принимающих решения (ЛПР) в Росфинмониторинге, интерпретация теоретических результатов диссертации осуществляется на примере области биоресурсов.

Цель и задачи исследования. Целью диссертационного исследования является разработка научно-технических решений автоматизированной обработки данных финансового мониторинга для оценки обстановки в области биоресурсов, исключающих экспертный субъективизм при обработке данных и удовлетворяющих критерию к временным и ресурсным показателям процесса обработки данных, для повышения качества принимаемых решений по противодействию легализации преступных доходов.

Достижение поставленной цели предполагает решение следующих задач:

1. Анализ существующих в финансовом мониторинге подходов к оценке обстановки.
2. Постановка и формализация частных задач этапа оценки обстановки, декомпозиция временных и ресурсных показателей.
3. Формирование критерия к временным и ресурсным показателям процесса обработки данных на этапе оценки обстановки.
4. Разработка алгоритма синтеза рейтинговых оценок субъектов обстановки в ИС финансового мониторинга.
5. Разработка алгоритма генерации информации о девиантной деятельности субъектов на естественном языке.
6. Разработка комплексной методик автоматизации процесса оценки обстановки в ИС финансового мониторинга, основанной на алгоритмах синтеза рейтинговых оценок субъектов и идентификации девиантной деятельности.

Объектом диссертационного исследования является информационная система финансового мониторинга.

Предметом диссертационного исследования являются алгоритмы и методы синтеза оценок девиантной деятельности в информационной системе финансового мониторинга.

Методы исследования

В работе использовались методы структурного системного анализа, математической статистики (факторный анализ, кластерный анализ), численные методы линейной алгебры, исследования операций.

Научная новизна состоит в разработке и обосновании научно-технических решений обработки данных финансового мониторинга, отличающихся от существующих подходов тем, что исключают экспертный субъективизм при обработке данных и удовлетворяют критерию к временным и

ресурсным показателям процесса обработки данных. Новыми являются следующие научные результаты:

1. Определены основные этапы обработки данных при оценке обстановки, поставлены и формализованы частные задачи обработки данных в ИС финансового мониторинга для оценки обстановки.

2. Впервые предложена комплексная методика автоматизации процесса оценки обстановки в ИС финансового мониторинга.

3. Разработан алгоритм синтеза рейтинговых оценок субъектов обстановки в ИС финансового мониторинга, учитывающий некоторую априори неизвестную вариацию исходных показателей, характеризующую девиантную деятельность субъектов.

4. Разработан алгоритм генерации информации о девиантной деятельности субъектов на естественном языке путём обработки текущей информации об их деятельности в ИС финансового мониторинга.

5. Впервые проверена гипотеза о применимости логнормального закона для описания зависимости количества субъектов в области биоресурсов от меры их вовлечённости в девиантную деятельность.

6. Разработанные алгоритмы и методика обработки данных финансового мониторинга в области биоресурсов в полной мере применимы к решению класса задач оценки обстановки в других областях деятельности объектов финансового мониторинга.

Практическая значимость работы состоит в следующем:

1. На основе разработанных алгоритмов и методики обработки данных финансового мониторинга модернизирована ИС Росфинмониторинга для обработки данных в области биоресурсов, что позволило:

- синтезировать рейтинговые оценки субъектов обстановки в области биоресурсов (2292 субъекта), лишённые экспертного субъективизма;
- сгенерировать логические правила идентификации девиантных субъектов в области биоресурсов, выраженные на естественном языке, по которым компетентные органы могут своевременно идентифицировать девиантную деятельность в области биоресурсов;
- на порядок сократить длительность процесса обработки данных для оценки обстановки в области биоресурсов (по экспериментальным данным экспертов Росфинмониторинга в 19 раз при формировании рейтинговых оценок и в 3-5 раз при генерации информации о девиантной деятельности в области биоресурсов).

2. С использованием разработанного алгоритма синтеза рейтинговых оценок разработана методика визуализации информации об обстановке, использование которой позволило сформировать карту обстановки в Дальневосточном Федеральном округе в области биоресурсов.

3. Экспериментально подтверждены теоретические положения диссертации: при сопоставлении предоставленной экспертами выборки субъектов (содержащей равное количество девиантных и недевиантных субъектов) с результатами ранжирования субъектов по синтезированным

рейтингам, около 100 субъектов выборки оказались первыми 100 наблюдениями (наиболее девиантными), оставшиеся 100 оказались последними 100 субъектами в списке (недевиантными).

4. Синтезированные рейтинговые оценки и информация о девиантной деятельности субъектов использованы функциональными подразделениями Росфинмониторинга в практической деятельности.

5. Научные результаты диссертации использованы в учебном процессе в рамках курса «Национальная система противодействия легализации преступных доходов и финансированию терроризма» кафедры финансового мониторинга стратегической академической единицы Институт интеллектуальных кибернетических систем НИЯУ МИФИ.

Положения, выносимые на защиту

1. Комплексная методика автоматизации процесса оценки обстановки в ИС финансового мониторинга.

2. Алгоритм синтеза рейтинговых оценок субъектов обстановки в ИС финансового мониторинга.

3. Алгоритм генерации информации о девиантной деятельности субъектов на естественном языке при обработке текущей информации об их деятельности в ИС финансового мониторинга.

4. Результаты внедрения программной реализации разработанных алгоритмов в ИС финансового мониторинга.

Соответствие темы диссертационного исследования паспорту специальности. Диссертационная работа по своему содержанию соответствует пунктам 2 (формализация и постановка задач системного анализа, оптимизации, управления, принятия решений и обработки информации), 4 (разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений и обработки информации), 5 (разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений и обработки информации) паспорта специальности 05.13.01 – системный анализ, управление и обработка информации (в информационных системах).

Апробация работы

Основные результаты диссертационной работы докладывались на следующих Всероссийских и международных конференциях:

1. XXII Всероссийская научная конференция «Информатизация и информационная безопасность правоохранительных органов», Москва, 29-30 мая 2013 г.

2. I Международный конгресс по информационной безопасности национальных экономик в условиях глобализации InfoSecurityFinance, Москва, 15-16 мая 2013 г.

3. VI Международная научно-практическая конференция «Информационные технологии в науке, бизнесе и образовании», InfoSecurityFinance, Москва, 27 ноября 2013 г.

4. II Международный конгресс по информационной безопасности национальных экономик InfoSecurityFinance, Москва, 15 мая 2014 г.

5. XXIV Всероссийская научная конференция «Информатизация и информационная безопасность правоохранительных органов», Москва, 3 июня 2015 г.

6. Международная научно-практическая конференция «Угрозы и риски для экономик стран БРИКС», Москва, 10-12 ноября 2015 г.

7. Всероссийский научный семинар «Актуальные вопросы управления в социально-экономических системах», Москва, 15 марта 2016 г.

Достоверность результатов обеспечивается корректностью применяемого математического аппарата, доказанностью выводов, совпадением теоретических выводов и экспериментальных данных, апробацией на научно-практических конференциях и семинарах, положительным эффектом внедрения результатов исследования.

Публикации. Основные результаты диссертации опубликованы в 7 печатных работах, из них 3 статьи в периодических научных изданиях, рекомендованных ВАК России, 2 статьи в журналах, представленных в базе цитирования Scopus, 2 работы в других изданиях.

Личный вклад автора. Основные научные результаты, заключающиеся в разработке алгоритмов автоматизированной обработки данных финансового мониторинга, получены автором лично.

Объём и структура работы

Диссертация состоит из введения, четырёх глав, заключения, списка литературы, изложена на 136 страницах, в том числе основного текста на 114 страницах. Работа иллюстрирована 28 таблицами и 32 рисунками. Список литературы содержит 163 источника, в том числе 42 на иностранных языках.

Основное содержание работы

Во введении обоснована актуальность выбранной темы, сформулированы цель и задачи исследования, определены научная новизна и практическая значимость работы, представлены основные положения, выносимые на защиту, сведения о практическом внедрении результатов исследования.

В первой главе проведен анализ существующих в финансовом мониторинге подходов к обработке информации на этапе оценки обстановки, предложены показатели, характеризующие процесс обработки информации на данном этапе, проанализирована ИС финансового мониторинга и определены направления её модернизации, поставлена задача исследования.

Анализ существующей в финансовом мониторинге практики показал, что на этапе оценки обстановки (рис. 1) выделяют ряд задач (табл. 1):

$$\langle Z_0 | Z_1, Z_2, Z_3 \rangle, \quad (1)$$

где:

Z_0 – задача выполнения этапа оценки обстановки в контуре принятых решений,

Z_1 – частная задача первичной обработки поступающих данных в целях расчёта специальных признаков (показателей) субъектов, разработанных в финансовом мониторинге, характеризующих финансовые операции субъектов, их руководителей и учредителей,

Z_2 – частная задача формирования рейтинговых оценок субъектов обстановки на основе этих показателей. Суть задачи формирования рейтинговых оценок субъектов обстановки состоит в том, чтобы каждому субъекту поставить в соответствие такой скаляр, что его величина является мерой девиантной деятельности субъекта,

Z_3 – частная задача синтеза информации о девиантной деятельности субъектов. Она заключается в том, что по выборке субъектов, в отношении которых проведены проверки, эксперты формируют информацию об обстановке на естественном языке, содержащую данные о признаках девиантных субъектов и их критических значениях, определяющих принадлежность субъекта к классу девиантных.

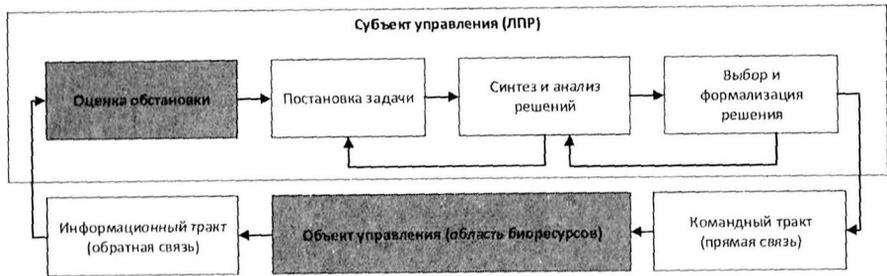


Рис. 1. Этапы принятия решения в финансовом мониторинге

В соответствии с функцией предпочтения лиц, принимающих решения, для описания процесса обработки информации на этапе оценки обстановки используется показатель времени.

Следовательно, общее время на обработку информации на этапе оценки обстановки зависит от оперативности решения частных задач:

$$\langle T_0 | T_1, T_2, T_3 \rangle, \quad (2)$$

где T_0 – общая длительность решения Z_0 , T_1 – длительность решения Z_1 , T_2 – длительность решения Z_2 , T_3 – длительность решения Z_3 .

Этап оценки обстановки можно представить в виде цепочки связанных последовательно выполняемых частных задач (рис. 2). Следовательно, время на обработку информации на этапе оценки обстановки является аддитивным показателем, который можно декомпозировать:

$$T_0 = \sum T_i, \quad i \in \{1, 2, 3\}, \quad (3)$$

$$T_i \in (0; +\infty).$$

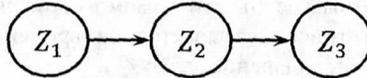


Рис. 2. Частные задачи обработки данных на этапе оценки обстановки

Таблица 1. Частные задачи оценки обстановки

Наименование	Описание данных и алгоритмов
Входные данные для Z_1 :	Большие массивы гетерогенных данных
Результат решения Z_1 :	Вектор показателей $X = (x_1, \dots, x_n)$, $n \in \mathbb{N}$
Алгоритмы обработки данных для решения Z_1 :	Алгоритмы $\alpha = \{\alpha_1, \dots, \alpha_n\}$, $n \in \mathbb{N}$ для расчёта x_1, \dots, x_n
Входные данные для Z_2 :	Вектор показателей x_1, \dots, x_n
Результат решения Z_2 :	Скаляры, сформированные по вектору показателей x_1, \dots, x_n
Алгоритм обработки данных для решения Z_2 :	Алгоритм α_{Z_2} преобразования векторных оценок в скаляры с учётом специфики Z_2
Входные данные для Z_3 :	Выборка векторов $x_1, \dots, x_n, b_i, b_i \in B, B = \{0; 1\}$
Результат решения Z_3 :	Информация об обстановке на естественном языке, сформированная по заданной выборке девиантных субъектов
Алгоритм обработки данных для решения Z_3 :	Алгоритм α_{Z_3} генерации информации о девиантной деятельности, выраженной на естественном языке

Автоматизацию процесса оценки обстановки в ИС финансового мониторинга предложено осуществлять с помощью комплексной методики, изображенной на рис. 3.

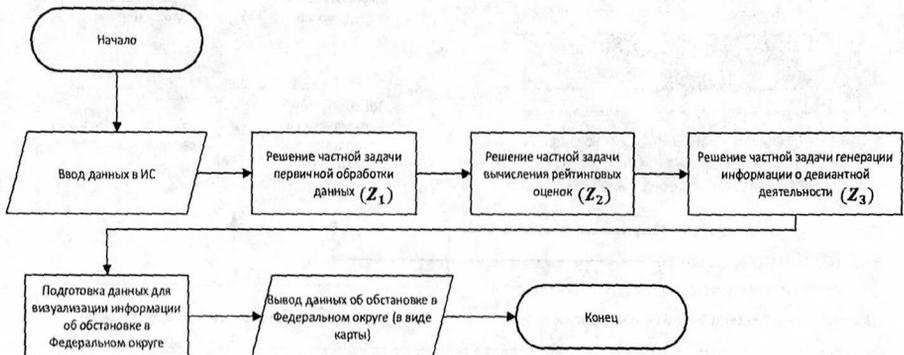


Рис. 3. Комплексная методика автоматизации процесса оценки обстановки в ИС

В настоящее время в ИС финансового мониторинга реализована только функция решения задачи Z_1 (ИС обработки данных, 1 – расчёт признаков, рис. 4). Решение задач Z_2 и Z_3 этапа оценки обстановки осуществляют эксперты,

что сопровождаются экспертным субъективизмом, социально-политической мотивацией экспертов и негативно сказывается на временных показателях. Таким образом, требуется научное решение задач Z_2 и Z_3 .

В соответствии с общепринятыми частными показателями оценки эффективности автоматизированных систем управления² для описания процессов обработки данных в ИС финансового мониторинга наряду с показателем времени T предложено использовать ресурсный показатель r количества экспертов (операторов), принимающих участие в решении задач:

$$\langle T, r \rangle, \quad T \in (0; +\infty), \quad r \in [1; R_{кр}], \quad (4)$$

где $R_{кр}$ – доступное количество операторов в финансовом мониторинге для участия в решении задач на этапе оценки обстановки.

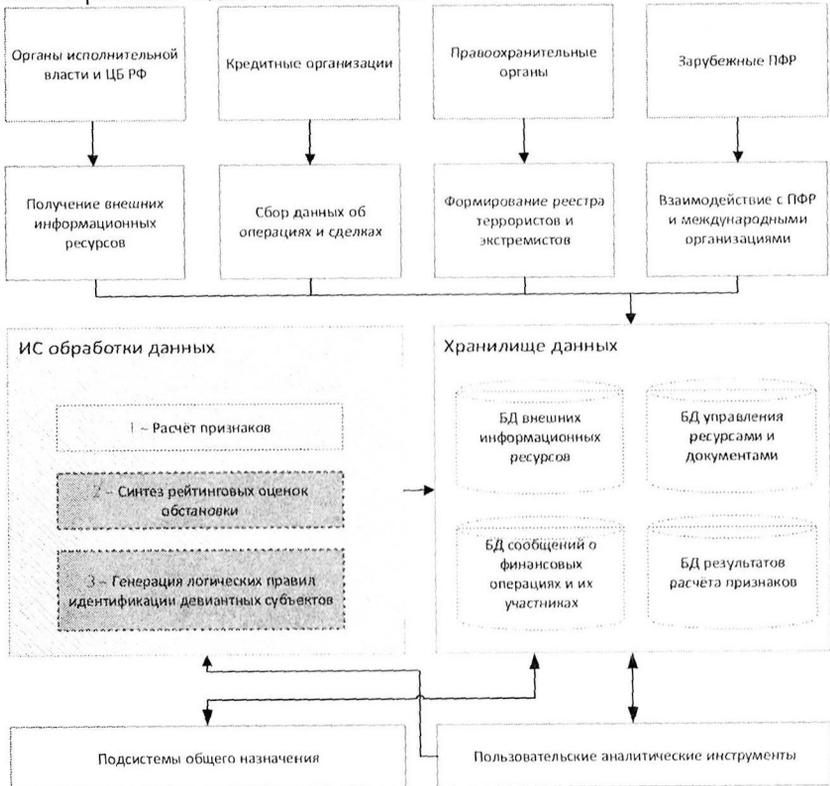


Рис. 4. Общая схема ИС Росфинмониторинга

² ГОСТ 24.702-85 «Единая система стандартов автоматизированных систем управления. Эффективность автоматизированных систем управления. Основные положения». — М.: Государственный комитет СССР по стандартам, 1985.

На основе указанных показателей с учётом функции предпочтения ЛПР в работе предложен критерий для обоснования модернизации ИС финансового мониторинга. Требуется найти такие $\langle T'_i, r'_i \rangle$, что:

$$T'_i < T_i; r'_i < r_i, \quad (5)$$

$$i \in \{2; 3\},$$

где T'_i – длительность решения задачи Z_i в ИС финансового мониторинга, T_i – длительность решения Z_i экспертами, r'_i – количество операторов ИС, принимающих участие в решении задачи Z_i , r_i – количество экспертов, принимающих участие в решении задачи Z_i .

Исследование задач Z_2 и Z_3 и разработка алгоритмов их решения с учётом специфики задач повысят качество результата (поскольку позволят исключить субъективный фактор), а программная реализация алгоритмов в ИС позволит сократить временные и ресурсные затраты за счёт автоматизации решения данных задач.

Формальная постановка задачи Z_2 состоит в нахождении скалярной функции векторного аргумента:

$$f_2(X) = z, \quad (6)$$

$$X = (x_1, \dots, x_n),$$

где X – вектор признаков x_1, \dots, x_n , z – скаляр, полученный в результате свёртки вектора признаков.

Примером скалярной функции является величина штрафа от местоположения (каждому вектору ставится в соответствие скаляр).

Задача Z_2 обладает спецификой, которая заключается в том, что при вычислении рейтинговых оценок субъектов требуется определить некоторую заранее неизвестную вариацию исходных показателей субъектов, которая характеризует девиантную деятельность, и синтезировать рейтинговые оценки именно по этой вариации. Специфика обусловлена особенностью исходных показателей. Только некоторая часть вариации исходных показателей характеризует девиантную деятельность субъекта. Следовательно, решение задачи состоит в разработке алгоритма, основанного на системном анализе исходных показателей, позволяющего определить и учесть указанную вариацию при синтезе рейтинговых оценок.

При постановке задачи приняты следующие ограничения:

- исследуется только один временной срез данных, итеративный характер оценки обстановки не учитывается;
- единовременные затраты на разработку и внедрение предложенного решения, затраты на подготовку (переподготовку) кадров не входят в рамки исследования;
- расчёт коэффициента эффективности от эксплуатации модернизированной ИС в работе не осуществляется;

- интерпретация теоретических результатов осуществляется только на примере области биоресурсов, содержащей 2292 субъекта (заданной ЛПР в соответствии с высокой практической значимостью «вскрытия рыбной мафии» в Дальневосточном Федеральном округе);

- с учётом структуры и логистики добычи биоресурсов в качестве базиса признакового пространства априорно заданы 13 показателей (признаков), характеризующих трансграничные и внутрироссийские финансовые операции субъектов, их руководителей и физических лиц-учредителей, а также показатель численности населения в регионе регистрации субъекта; алгоритмы вычисления признаков реализованы в ИС на языке PL/SQL, их значения рассчитаны для всех субъектов заданной области, в результате чего сформирована матрица исходных данных. В табл. 2 представлен её фрагмент. Расшифровка признаков дана в диссертации.

Использование принятых ограничений обоснованно, поскольку они соответствуют большинству ситуаций, возникающих при оценке обстановки в финансовом мониторинге.

Таблица 2. Фрагмент сформированной матрицы исходных данных

Наименование	Пр 1	Пр 2	Пр 3	Пр 4	Пр 5	Пр 6	Пр 7	Пр 8	Пр 9	Пр 10	Пр 11	Пр 12	Пр 13
ООО "ИКОМА"	18441609.80	4389200	0	3885306	26855000	0	0	0	0	0	0	0	0
ООО "ВПК"	13722935.9	17144130.26	12442143.6	0	3077942.63	2767232	9720679	0	0	0	3077942.63	2767232	0
ООО "ПЕЛЛЕВИК"	50301832.1	312038020	0	1950336	19789320	2208889	0	0	0	0	2130638.09	2370638.66	0
ООО "СРЧ"	166979466	121019480	8221735	0	13121249	0	0	0	0	0	0	0	41323007
ООО "САХАЛИНРЫБПРОД"	39851838	0	3367204	0	0	0	0	0	0	0	0	0	809619
ООО "УФ"	265262080	246702460.9	0	0	2141561036	620986	15229176	0	0	0	0	0	0
ООО "РЫБПРОД ДП"	33143984	661820	0	0	0	0	0	0	0	0	0	0	13152279
ООО "РАЙ-ФЛО"	3301308	0	0	0	0	0	0	0	0	0	1562026.27	4558146.57	334202036
ООО "РК ТАШИНИ"	4622020	802932	802262	0	0	0	0	0	0	0	0	0	1481378
ООО "ПРЭ ПАЗСВКИ"	2438030	1702800	0	0	0	0	646506	0	922000	0	0	0	610236.68
ООО "КАРА"	0	0	0	0	0	0	0	0	0	0	0	0	2667030
ООО "СВЕРХРЫБ-ПАЗИД"	0	0	0	0	0	1663036	0	0	0	0	0	0	15481031
ООО "МАРИНАРИ-М"	33348930	4791030	0	0	2856204.03	0	8014622.9	0	0	0	0	0	1489221.1
ООО "МОРСКИЕ ВЕЛ"	0	0	0	0	0	0	0	0	0	0	0	0	4467030
ООО "КОСМОС-РИ"	0	0	0	0	0	1034000	1044000	0	0	0	0	0	0
РЫБКО-ОБСЛУЖИВАТЕЛЬСКИЕ	122279930	8242660.8	0	0	7176034	8581403	0	0	0	0	0	0	6191000
ООО "СЭТИРИ"	0	0	0	0	0	0	0	0	0	0	0	0	0
ООО "ЧЕРМУХИНО"	13964092	0	2294790.7	0	0	0	333212.83	0	0	2036762.45	10730100	0	36502000
ООО "СКОРОС"	27482029	0	0	0	0	0	3885644.8	0	0	0	0	0	0
ООО "КОСМОСРИ ШИП"	70345120.9	1028115	4053100.15	1181788	30689716.74	0	378921.54	0	0	0	0	0	0
ООО "КОСМОС"	26100667	108343689.1	0	0	1030300	7292000	0	0	0	0	0	0	283000
ООО "КОСМОС-РИ"	5217740	3403330	0	0	0	0	0	0	0	0	0	0	2491000
ООО "СКОС"	0	0	0	0	0	0	0	0	0	0	0	0	0
ООО "РК ТЕХСАР"	181177816	0	0	0	0	0	0	0	0	0	0	0	0
ООО "КОТЛАС"	15639080	6943600	0	0	0	0	0	0	0	0	0	0	0
ООО "ПОРТОВОЙ РЫБАК"	749008	749008	0	0	1180308	595200	0	0	0	0	0	0	0
ООО "ПРИКОСТРЕШ"	6717439.4	7699030	0	0	668493.91	304180	2332036	8096229	0	0	0	0	9778840
ООО "СВЕТЛОСНОВА ЗОО"	26871894	11028186.18	0	0	2032020	7032020	0	0	0	0	0	0	54692240
ООО "ЮТА"	18198176	6614480	0	0	2191200	2191200	0	0	0	0	0	0	0
ООО "СН"	39139000	148793433.8	4693981.80	10620	4103000	0	0	0	0	0	0	0	13668030
ООО "МАРИНАРИ"	6688131.5	13428536.19	7302763	2160000	1210000	0	0	0	0	0	0	0	2603000
ЗАО "МЭГЕЛМА"	130608736	216832000	16680390	0	0	0	0	0	0	0	0	0	0
ООО "ВЕНТЮРИНГ"	1391866.8	0	0	0	803481038	4957618	0	0	0	0	0	0	0
ООО "МФ"	3772241.0	11184200	0	0	244191020	2441431	0	0	0	0	0	0	0
ООО "ВНЗ"	22762627.9	1813020	0	0	2090530	12600973.84	1876811	0	0	0	0	0	0
ООО "МОРСКОЕ БИРО"	0	0	0	0	0	0	0	0	0	0	0	0	0
ООО "СКОС-РИ-ПАЗИД"	95981127.2	13917020	0	0	1768000	0	0	0	0	0	0	0	0
ООО "СВЕРХРЫБ-ПАЗИД"	18100231	14811139	15289504.2	0	0	0	880927.85	0	0	0	0	0	0
ООО "РЫБОВОДНЫЙ КОМПЛЕКС"	4581807.9	40581070	1451060	0	3191397.48	957490	0	0	0	0	0	0	0
ООО "РЫБОВОДНЫЙ КОМПЛЕКС"	28710020	833000	0	0	1669000	0	0	0	0	0	0	0	0
РЫБ-ПРОД-СЕРВИС	178181790	4791000	0	0	4627630	2445430	827280	0	0	0	0	0	0
ООО "БЕСТЕР"	8186020.8	10846000	30060	2181000	4495000	0	0	0	0	0	0	0	0
ООО "МПК"	27953038	28742308	387410268	0	4812500	0	0	0	0	0	0	0	0

Задача Z_3 синтеза информации об обстановке на естественном языке, содержащей данные о признаках девиантных субъектов и их критических значениях, сводится к задаче генерации логических правил импликации с помощью которых девиантные субъекты можно идентифицировать по значениям их признаков. В общем виде задачу можно представить как задачу идентификации девиантных субъектов путём отображения векторного множества субъектов на множество чисел, таких, что они позволяют определить класс субъекта:

$$f_3: A \rightarrow B, \quad (7)$$

$$f_3(x_1, \dots, x_n) = b_i, \quad b_i \in B, \quad B = \{0; 1\},$$

где A — множество заданных субъектов, каждый из которых представлен вектором признаков (x_1, \dots, x_n) , B — множество чисел, $B = \{0; 1\}$, где 1 обозначает принадлежность субъекта к классу девиантных, 0 — недевиантных. Требуется разработать алгоритм генерации логических правил импликации, в соответствии с которыми каждому элементу множества A сопоставляется единственный элемент множества B . Сгенерированные логические правила можно выразить на естественном языке и, таким образом, решить задачу Z_3 .

Решение задачи определения класса объекта по набору его исходных характеристик позволяют осуществить методы классификации. Таким образом, для подготовки информации об обстановке на естественном языке требуется осуществить классификацию субъектов из заданной выборки на девиантных и недевиантных. При постановке задачи ЛПР сформулированы следующие требования, которые требуется учесть при выборе метода:

- результатом извлечения правил классификации должны быть логические правила импликации, использующие исходные признаки субъектов;
- метод должен обладать интуитивно понятной классификационной моделью;
- допустимая доля ошибок классификации должна быть не более 10%.

Во второй главе разработан алгоритм синтеза рейтинговых оценок субъектов обстановки на примере области биоресурсов для модернизации ИС финансового мониторинга.

В результате проведения сравнительного анализа методов, позволяющих осуществлять скаляризацию многокритериальных оценок, к числу перспективных отнесены метод k -средних кластерного анализа и метод главных компонент факторного анализа.

Автором протестирован метод кластерного анализа для решения указанной задачи, поскольку он предполагает выделение групп (кластеров, классов) анализируемых объектов. Первичный анализ информационного массива показал, что наблюдения могут быть разделены на два кластера (рис. 5). За функционал качества разбиения взята сумма внутриклассовых дисперсий³:

$$Q_1(C) = \sum_{m=1}^p \sum_{O_i \in C_m} d^2(O_i, \bar{X}(m)), \quad (8)$$

где p — число классов, C_m — m класс в классификации, $\bar{X}(l)$ — центр класса C_m .

Интерпретация результатов, полученных методом k -средних показала, что сформированные кластеры неудовлетворительны для решения поставленной задачи. В диссертации проанализированы свойства обоих кластеров, представлены рассчитанные расстояния объектов до центроидов кластеров.

³ Айвазян, С.А. Прикладная статистика. Теория вероятностей и прикладная статистика. Том №1. / С.А. Айвазян, В.С. Мхитарян. — М.: Юнити, 2001. — 656 с.

Анализ показал, что отнесение объектов к кластерам имеет формальный характер, преимущественно из-за больших значений дисперсии некоторых признаков, что негативно сказывается на возможности определить кластер с девиантными субъектами и использовать меры близости точек в качестве рейтинговых оценок субъектов обстановки.

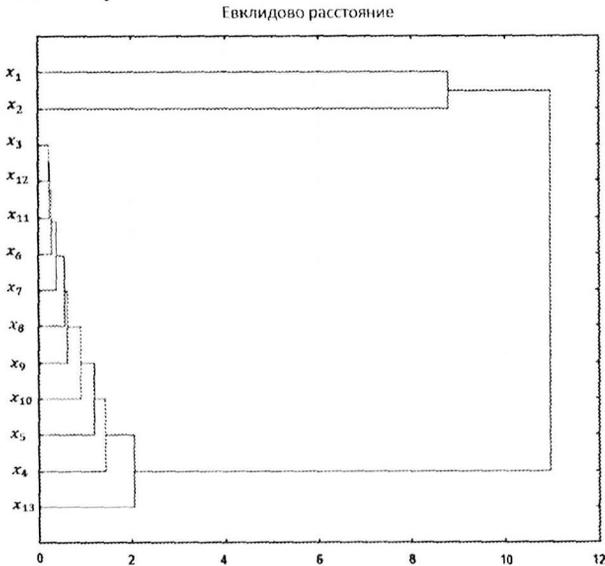


Рис. 5. Дендрограмма для оценки числа кластеров по исходным данным

В связи с этим в работе протестирован метод главных компонент факторного анализа. В терминах метода главных компонент формализована и поставлена задача синтеза рейтинговых оценок субъектов обстановки.

В основе метода главных компонент лежат всевозможные линейные преобразования исходных признаков, позволяющие перейти к главным компонентам. Формальная постановка задачи снижения размерности выглядит следующим образом³:

$$z^{(w)} = c_{j1}(x^{(j)} - \mu^{(1)}) + \dots + c_{jp}(x^{(j)} - \mu^{(p)}), \quad (9)$$

где $\mu^{(v)}$ – математическое ожидание $x^{(v)}$, а мерой информативности является

$$\text{выражение } I_{p'}(Z(X)) = \frac{\sum_{j=1}^{p'} \mathbf{D}z_j}{\sum_{j=1}^p \mathbf{D}x_j}, \quad \sum_{j=1}^p \mathbf{D}x_j \neq 0,$$

где $\sum_{j=1}^p \mathbf{D}x_j$ – дисперсия признаков. Как видно из исходных данных дисперсия признаков исходного пространства не равна нулю. Таким образом, переходим к p' главным компонентам.

Модель метода главных компонент³:

$$y'_{ij} = \sum_{r=1}^n a_{jr} f_{ir}, \quad (10)$$

где $r = 1, 2, \dots, n$, $j = 1, 2, \dots, n$, f_{ir} – значение r -й главной компоненты для i -го наблюдения; a_{jr} – вес r -й компоненты в j -й переменной (факторная нагрузка r -й главной компоненты на j -ю переменную); y'_{ij} – нормированное значение j -го признака для i -го наблюдения.

Главные компоненты характеризуют разную вариацию исходных показателей. Полный вклад r -го фактора в дисперсию всех n признаков определяет ту долю общей дисперсии, которую данная главная компонента обуславливает. Этот вклад вычисляется по формуле³:

$$V_r = \sum_{j=1}^n a_{jr}^2.$$

Проецирование исходных признаков на новое подпространство, образованное главными компонентами, позволит определить меру проявления исследуемого латентного свойства в каждом объекте. Статистические оценки значений r -го фактора по признакам i -го наблюдения можно найти по формуле³:

$$f_{ir} = \frac{1}{\lambda_r} (a_{1r} y_{i1} + a_{2r} y_{i2} + \dots + a_{nr} y_{in}), \quad (11)$$

где λ_r – собственные числа, расчёт которых предполагает метод главных компонент.

Таким образом, анализ данных с помощью главных компонент позволяет использовать главные компоненты для исследования структуры вариации исходных показателей и определения априорно неизвестной вариации, характеризующей исследуемое свойство.

Алгоритм синтеза рейтинговых оценок субъектов обстановки с использованием метода главных компонент представлен на рис. 6.

Программная реализация предложенного алгоритма внедрена в ИС финансового мониторинга. Заштрихованные элементы требуют участия оператора ИС для выбора главной компоненты. На данном этапе требуется выбрать главную компоненту, по которой будет осуществляться синтез рейтинговых оценок. На первом этапе требуется определить количество информативных главных компонент. Существуют следующие подходы:

- по критерию Кайзера, в соответствии с которым отбираются факторы с собственными значениями равными или большими 1 (отбираются только факторы, выделяющие дисперсию, эквивалентную дисперсии хотя бы одной исходной переменной);
- по величине кумулятивной характеристики дисперсии (главные компоненты должны описывать не менее 70% дисперсии исходных данных).

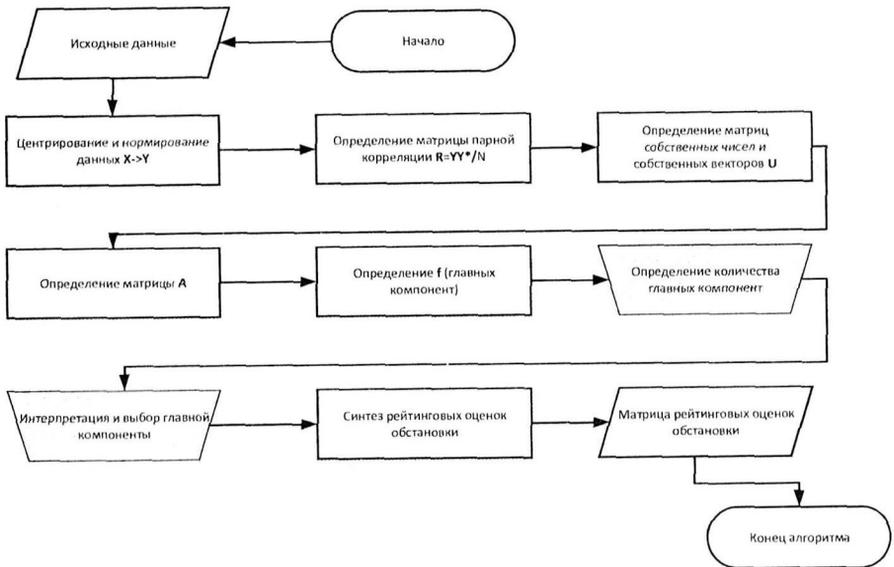


Рис. 6. Алгоритм синтеза рейтинговых оценок субъектов обстановки в ИС финансового мониторинга

Применение обоих правил показало, что информативными являются первые 5 главных компонент (из исходных 13). Расчёты с проекциями исходных признаков на каждую из главных компонент представлены в табл. 3.

На втором этапе осуществляется выбор целевой главной компоненты, по которой осуществляется синтез рейтинговых оценок. В диссертации по правилу «логические дроби» проведена интерпретация главных компонент с использованием исходных данных и привлечением ведущих экспертов Росфинмониторинга. Синтез рейтинговых оценок субъектов обстановки предложено осуществлять по первой главной компоненте.

Наибольшая положительная корреляция первой главной компоненты (ГК₁) наблюдается с признаками, характеризующими финансовые операции руководителя (табл. 3, x_5, \dots, x_8), и малая отрицательная – с численностью населения в регионе регистрации субъекта (табл. 3, x_{14}).

Таблица 3. Матрица факторных нагрузок

Признак		ГК ₁	ГК ₂	ГК ₃	ГК ₄	ГК ₅
X ₁	Признак подозрительности финансовых операций субъекта по данным кредитных организаций	0,38	0,56	0,63	0,30	-0,02
X ₂	Признак, характеризующий трансграничные финансовые операции субъекта	0,27	0,51	0,58	0,39	0,17
X ₃	Признак, характеризующий финансовые операции субъекта в наличной форме	0,48	0,18	0,21	-0,04	-0,41
X ₄	Признак, характеризующий финансовые операции учредителя субъекта в наличной форме	0,21	0,34	0,32	-0,34	-0,25
X ₅	Признак, характеризующий финансовые операции руководителя субъекта в наличной форме	0,84	-0,38	-0,05	0,03	0,03
X ₆	Признак, характеризующий валютнообменные финансовые операции руководителя субъекта	0,83	-0,45	-0,09	0,09	0,02
X ₇	Признак, характеризующий трансграничные финансовые операции руководителя субъекта	0,81	-0,44	-0,09	0,07	0,02
X ₈	Признак подозрительности финансовых операций руководителя субъекта по данным кредитных организаций	0,50	0,08	0,14	-0,38	0,25
X ₉	Признак, характеризующий трансграничные финансовые операции учредителя субъекта	0,23	0,39	0,04	-0,66	0,17
X ₁₀	Признак, характеризующий валютнообменные финансовые операции субъекта	0,26	0,57	-0,56	0,14	-0,07
X ₁₁	Признак, характеризующий особенность проведения финансовых операций руководителя субъекта в наличной форме	0,20	0,41	-0,60	0,34	-0,04
X ₁₂	Признак, характеризующий особенность проведения финансовых операций учредителя субъекта в наличной форме	0,28	0,52	-0,58	0,02	0,09
X ₁₃	Признак подозрительности финансовых операций учредителя субъекта по данным кредитных организаций	0,11	0,26	-0,21	-0,37	0,08
X ₁₄	Численность населения региона регистрации субъекта	-0,02	0,01	0,06	0,11	0,82

Синтезированные по первой главной компоненте рейтинговые оценки субъектов обстановки представлены в табл. 4.

Таблица 4. Синтезированные рейтинговые оценки субъектов обстановки

Наименование субъекта	Рейтинг	Регион
ООО КОМПАНИЯ "ТУНАЙЧА"	69,35	Сахалинская обл.
ООО АРТЕЛЬ "НАРОДЫ СЕВЕРА"	22,07	Камчатский край
ОАО ТУРНИФ	15,82	Приморский край
ООО "САЛМО"	11,82	Сахалинская обл.
ООО "МОРСКОЙ ВЕТЕР"	10,43	Хабаровский край
ООО "БЛАФ-ЮЖНЫЙ"	9,53	Камчатский край
ООО "РХ "ВЛАДИМИРОВО"	9,28	Сахалинская обл.
ООО "ТРАЛ-МАСТЕР"	8,11	Хабаровский край
ООО "ПО САХАЛИНРЫБАКСОЮЗ"	7,84	Сахалинская обл.
ООО "ДАЛЬФИШ"	7,01	Хабаровский край

Для верификации полученных рейтинговых оценок субъектов найдены частоты $p_k = \frac{N_{1k}}{N_{2k}}, N_{2k} \neq 0$, где N_{1k} – количество девиантных субъектов (априорные данные в ИС Росфинмониторинга), N_{2k} – общее количество субъектов, принадлежащие интервалу рейтинговых оценок в соответствии с их рейтингами (рис. 7).

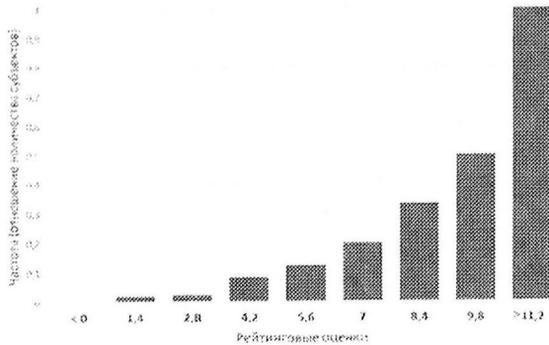


Рис. 7. Частоты p_k девиантных субъектов, рассчитанные с учётом априорных данных

Из графика (рис. 8) видно, что большинство рейтинговых оценок имеют близкие к нулю либо отрицательные значения. Отрицательные рейтинги означают отсутствие исследуемого латентного свойства, которое характеризует первая главная компонента.

В целях экспериментальной проверки полученных результатов экспертами предоставлена выборка, содержащая около 200 субъектов, в отношении которых экспертами проведены проверки в соответствии с синтезированными рейтингами. Выборка содержит равное число девиантных и недевиантных субъектов. При сопоставлении субъектов выборки с результатами ранжирования по синтезированным рейтингам около 100 субъектов выборки оказались первыми 100 наблюдениями (наиболее девиантными), оставшиеся 100 оказались последними 100 наблюдениями в списке ранжирования (недевиантными).

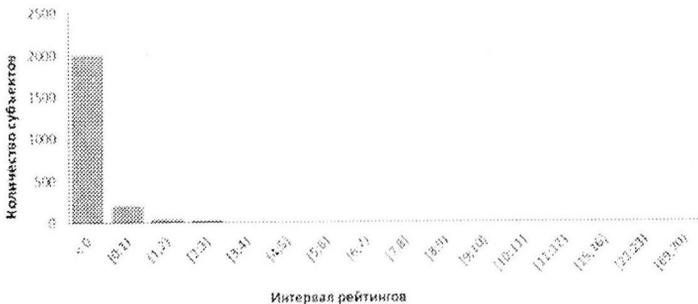


Рис. 8. Распределение рейтинговых оценок по интервалам

Для интерпретации распределения рейтинговых оценок автором проведён анализ закона их распределения, в результате чего сформулирован вывод о логнормальной зависимости (рис.9) количества субъектов от величины их рейтингов. Вывод основан на проверке гипотезы по критерию χ^2 при уровне значимости $p \leq 0,05$.

Таким образом, логнормальный закон применим для описания зависимости количества субъектов в области биоресурсов от меры их вовлечённости в девиантную деятельность. Известно, что логнормальный закон имеет вид⁴:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0, \sigma > 0, \mu \in \mathbb{R}. \quad (12)$$

В исходной выборке 101 субъект (около 5%) обладает рейтинговыми оценками выше среднего ожидаемого (математического ожидания) $M = 0,95$. Субъекты, у которых рейтинговые оценки выше установленного порогового значения, обладают существенно выраженной девиантной деятельностью (свойство, которое характеризует первая главная компонента).

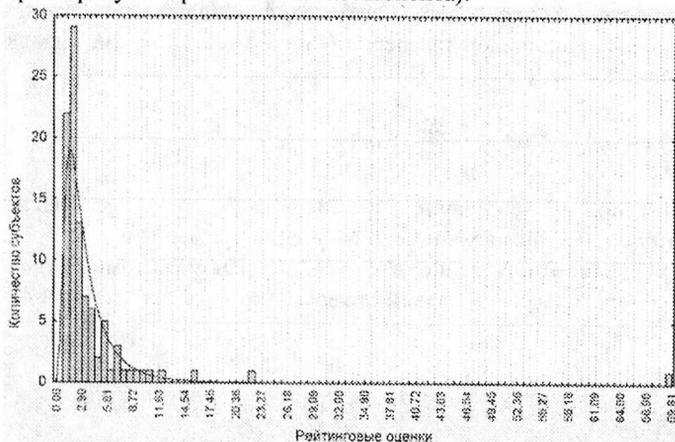


Рис. 9. Асимптотическая аппроксимация частот логнормальной функцией

В третьей главе разработан алгоритм генерации информации о девиантной деятельности субъектов.

В работе проведён сравнительный анализ методов, позволяющих классифицировать объекты с помощью интуитивно понятных правил, которые можно интерпретировать на естественном языке с использованием исходных признаков субъектов. К числу перспективных отнесён метод деревьев решений.

Метод требует использования обучающей и контрольной выборки для разработки модели классификации субъектов. Выборки заданы ЛПР по результатам проведённых проверок, основанных на рейтинговых оценках

⁴ Aitchison, J., Brown, J.A.C. The lognormal distribution. Cambridge University Press, Cambridge, 1957. — 176 p.

обстановки, синтезированных в ходе диссертационного исследования. Исходная выборка состоит из 192 наблюдений, содержит равное количество девиантных субъектов и обычных. 70% наблюдений (134) отобраны для формирования обучающей выборки, 30% наблюдений (58) приняты за контрольную выборку, на которой проверены построенные модели.

В исследовании проведён сравнительный анализ алгоритмов CART и QUEST деревьев решений для генерации логических правил. Алгоритм CART на каждом шаге построения дерева последовательно сравнивает все возможные разбиения для всех атрибутов и выбирает наилучший атрибут и наилучшее разбиение для него. В качестве функции оценки качества разбиения выборки алгоритм CART использует индекс *Gini*. Индекс *Gini* в общем виде можно выразить следующим образом (для выборки A с k классами)⁵:

$$Gini(A) = 1 - \sum_{i=1}^k p_i^2, \quad (13)$$

где p_i – вероятность i -го класса в выборке A . Индекс *Gini* можно выразить разделив выборку A с числом примеров N на A_1 и A_2 с K_1 и K_2 наблюдениями соответственно⁵:

$$Gini_{split}(A) = \frac{K_1}{N} \cdot Gini(A_1) + \frac{K_2}{N} \cdot Gini(A_2), \quad N \neq 0. \quad (14)$$

Как известно, наилучшим считается то разбиение, для которого $Gini_{split}(A)$ минимально. Если K – число примеров в узле-предке, L, H – число примеров соответственно в левом и правом потомке, l_i и h_i – число экземпляров i -го класса в левом/правом потомке, то оценим качество разбиения⁵:

$$Gini_{split} = \frac{L}{K} \cdot \left(1 - \sum_{i=1}^n \left(\frac{l_i}{L} \right)^2 \right) + \frac{H}{K} \cdot \left(1 - \sum_{i=1}^n \left(\frac{h_i}{H} \right)^2 \right) \rightarrow \min, \quad H \neq 0, L \neq 0, K \neq 0. \quad (15)$$

Применив тождественные преобразования (15) можно переписать⁵:

$$G_{split} = \frac{1}{L} \cdot \sum_{i=1}^n l_i^2 + \frac{1}{H} \cdot \sum_{i=1}^n h_i^2 \rightarrow \max, \quad H \neq 0, L \neq 0. \quad (16)$$

Данные о разделяющих признаках и их значениях при формировании дерева представлены в табл. 5.

Применение алгоритма CART позволило сгенерировать правила, в соответствии с которыми идентификация субъектов классов производится с высокой точностью (по экспериментальным данным – без ошибок).

⁵ Han, J., Kamber, M., Pei, J. Data mining. Concepts and techniques. — Elsevier, 2012. — 744 p.

Таблица 5. Данные о разделяющих признаках и их значениях, полученные по алгоритму CART деревьев решений

№ узла дерева (левые ветви)	№ узла дерева (правые ветви)	Количество девиантных/недевиантных субъектов в узле левой ветви	Количество девиантных/недевиантных субъектов в узлах правой ветви	Класс девиантных/недевиантных	Значение разделяющего признака k_i	Разделяющий признак x_i
2	3	67	67	0	21 442 266	x_6 (характеризует финансовые операции руководителя субъекта)
4	5	67	31	0	54 539 762	x_{12} (характеризует финансовые операции учредителя субъекта)
		0	36	1		
6	7	67	17	0	2 930 000	x_3 (характеризует финансовые операции субъекта)
		0	14	1		
8	9	67	8	0	15 630 637	x_8 (характеризует финансовые операции руководителя субъекта)
		0	9	1		

Таким образом, использование алгоритма CART позволило сгенерировать правила идентификации девиантных субъектов по их исходным признакам:

- $x_6 \geq 21\,442\,266$,
на естественном языке: девиантными являются субъекты, у которых сумма операций вида, который характеризует признак x_6 , совершённых руководителем субъекта, больше или равна 21 442 266 руб., узел №3;
- $x_6 < 21\,442\,266 \wedge x_{12} \geq 54\,539\,762$,
на естественном языке: девиантными являются субъекты, у которых сумма операций вида, который характеризует признак x_6 , совершённых руководителем субъекта, менее 21 442 266 руб., но сумма операций вида, который характеризует признак x_{12} , совершённых учредителем субъекта больше или равна 54 539 762 руб., узел №5.

Высокая точность результатов подтверждается оценками вероятности ошибочной классификации, полученной при тестировании построенной модели на контрольной выборке.

Вероятность ошибочной классификации по алгоритму CART удовлетворяет неравенству $0,00 < \tilde{p} < 0,06$ при коэффициенте доверия⁶ $P = 0,975$.

Проверка результатов классификации по алгоритму QUEST по таблицам показала, что вероятность ошибочной классификации $0,08 < \tilde{p} < 0,28$ при $P = 0,975$.

⁶ Расчёты произведены в соответствии с табличными значениями в справочнике Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. — М.: Наука, 1983. — 416 с.

При определении оценок по таблицам взяты две цифры после запятой, при этом табличные значения не округлялись.

Рассчитаем указанные оценки другим способом⁷ и сравним результаты. Общее количество субъектов контрольной выборки $n = 60$, количество ошибок $m_1 = 0$, относительная частота $\omega = \frac{m_1}{n} = 0$, t определено по таблице значений аргументов функции Лапласа^{6,7} $\Phi(t)$. Отсюда находим интервальные \tilde{p}_1 и \tilde{p}_2 вероятности ошибочной классификации по алгоритму CART:

$$\tilde{p}_1 = \frac{n}{t^2+n} \left[\omega + \frac{t^2}{2n} - t \sqrt{\frac{\omega(1-\omega)}{n} + \left(\frac{t}{2n}\right)^2} \right] \Rightarrow \tilde{p}_1 \approx 0,00,$$

$$\tilde{p}_2 = \frac{n}{t^2+n} \left[\omega + \frac{t^2}{2n} + t \sqrt{\frac{\omega(1-\omega)}{n} + \left(\frac{t}{2n}\right)^2} \right] \Rightarrow \tilde{p}_2 \approx 0,06.$$

Получим аналогичные оценки для результатов классификации по алгоритму QUEST: $n = 60$, $m_2 = 10$, $\omega = \frac{m_2}{n} \approx 0,17$, t определено по таблице значений аргументов функции Лапласа^{6,7} $\Phi(t)$. Отсюда находим $\tilde{\tilde{p}}_1$ и $\tilde{\tilde{p}}_2$:

$$\tilde{\tilde{p}}_1 = \frac{n}{t^2+n} \left[\omega + \frac{t^2}{2n} - t \sqrt{\frac{\omega(1-\omega)}{n} + \left(\frac{t}{2n}\right)^2} \right] \Rightarrow \tilde{\tilde{p}}_1 \approx 0,13.$$

$$\tilde{\tilde{p}}_2 = \frac{n}{t^2+n} \left[\omega + \frac{t^2}{2n} + t \sqrt{\frac{\omega(1-\omega)}{n} + \left(\frac{t}{2n}\right)^2} \right] \Rightarrow \tilde{\tilde{p}}_2 \approx 0,24.$$

Таким образом, сформулированы следующие выводы:

1. Экспериментально подтверждена возможность автоматизированной идентификации девнантной деятельности субъектов по признакам \tilde{k} , $\tilde{k} \subseteq \tilde{N}$, где \tilde{N} – исходное множество признаков для заданной выборки A .

2. Правила, сгенерированные с помощью алгоритма CART, обладают более высокой точностью по сравнению с правилами, полученными по алгоритму QUEST, что подтверждается оценками вероятности ошибочной классификации, полученными с помощью двух методов. Результаты оценок, полученных обоими способами, не противоречат друг другу, что подтверждает корректность выбора алгоритма CART.

На основе CART разработан алгоритм автоматизированной генерации правил, программная реализация которого внедрена в ИС финансового мониторинга. Схема алгоритма представлена на рис. 10. Заштрихованный блок требует участия оператора ИС для выполнения алгоритма.

⁷ Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике / В.Е. Гмурман. — М.: Высш. Школа, 2003. — 400 с.

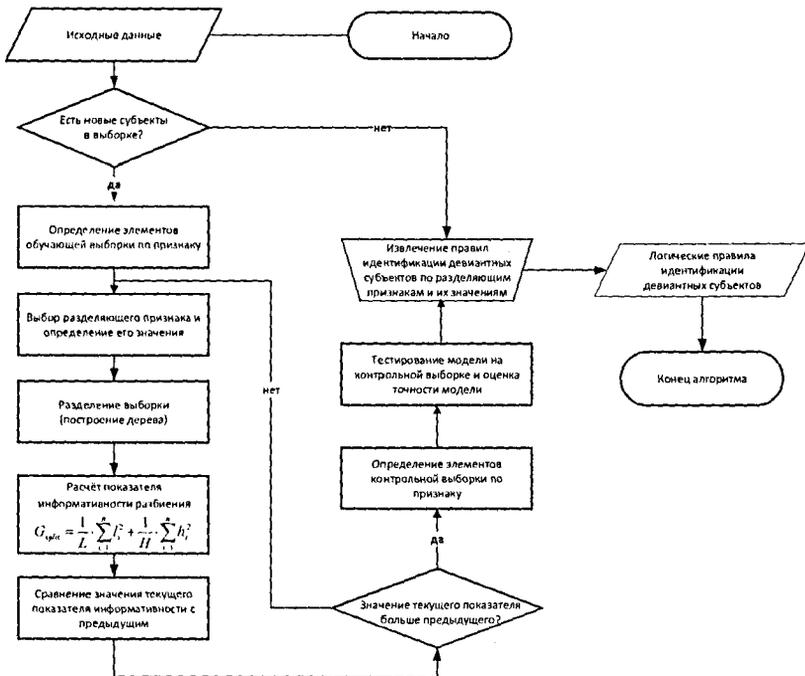


Рис. 10. Алгоритм генерации информации о девиантной деятельности субъектов на основе текущей информации о субъектах в ИС финансового мониторинга

В четвёртой главе представлены результаты реализации разработанных алгоритмов в ИС, констатирован положительный эффект модернизации ИС.

На основе разработанных в диссертации алгоритмов модернизирована ИС Росфинмониторинга. Использование ИС для обработки данных особенно эффективно при больших выборках, с которыми, как правило, сталкивается финансовый мониторинг (например, несколько десятков гигабайт данных по 2292 субъектам в области биоресурсов). Это подтверждается экспериментальными данными эксплуатации модернизированной ИС: временные затраты на синтез рейтинговых оценок сокращены в 19 раз, временные затраты на синтез информации о девиантной деятельности субъектов сокращены в 10 раз, ресурсные затраты при решении каждой из задач сокращены в 3-5 раз.

Рейтинговые оценки субъектов обстановки, полученные в результате работы модернизированной ИС финансового мониторинга отображены в

интерфейсах аналитических инструментов ИС финансового мониторинга (рис.11, рис. 12).

Результаты

Описание подзаголовка колонки для сортировки по этой колонке

Имя колонки	ИФП	СФП	Рейтинг	(A) Прогноз	(B) Оценка по БЭП	(C) Поддержка по-прежнему	(D) Внешние показатели	(E) IT-Данные	(F) Сведения
ООО "РАЙПРОМНЕВСИ"	0201060574	109620004095	5,92	10 350 599,84	2 277 735,00	0	0	0	0
ООО "БЕРИТ"	1107000251	1094100942996	5,70	14 684 860,00	800 000,00	0	0	0	0
ООО "ДЭВИТ"	0207000265	109060093795	5,60	16 794 000,00	5 584 000,00	0	0	0	0
ООО "САМУРАИ"	1001001188	1096501009554	5,78	156 305 214,76	5 281 680,00	0	0	0	0
ООО "ШИКАС"	4101142141	1114101000235	5,75	0	0	0	0	0	0
ООО "МАРТ"	1107000300	10924010229944	5,72	143 158 080,10	5 493 850,00	0	2 250 000,00	0	0
ИСТОК ФИВ ЗАО	2507000500	1092001790031	5,70	10 366 880,06	5 244 256,06	0	5 244 256,06	0	307 043,59
ООО "ВОССТРАНРАЙВОТ"	4101152432	1114101000166	5,69	965 379 322,63	39 684 950,93	0	0	0	107 677 616,50
ООО "КИМРАЧ-ФИВ"	1102000400	1094141000525	5,56	0	0	0	0	0	0
ООО "МРД"	2705010246	1092709015046	5,65	31 932 928,85	30 650 000,00	0	8 450 000,00	0	0
ООО "К ПЕРМОНЦ"	2705010253	1112503001339	5,59	376 622 700,23	11 593 354,43	0	10 825 354,43	0	687 047 667,30
ООО "ВЕРИАН"	1102000419	1094100944114	5,56	0	0	0	0	0	0
ООО "ПОЛИНА КИ ЛЕД"	0505000059	1090500970479	5,53	1 320 000,00	0	0	0	0	0
ООО "ИВИННЕ СТАР КАМ"	0201001264	1092401015603	5,40	110 652 690,36	74 360 712,36	0	65 469 206,36	0	109 951 670,14
ООО "СТИЛЬ"	4109000240	1094101219997	5,42	95 867 436,79	94 200 374,00	0	0	0	295 633 370,50
ООО "ПЯМТА"	4101000432	1119417000073	5,41	39 850 000,00	500 000,00	0	0	0	0
ООО "СОВЕР"	7101196229	1114101000334	5,36	2 400 000,00	2 400 000,00	0	0	0	0
ООО "ВОССТРАНЦИ ПРОМЦИС"	0201156287	1114101000336	5,36	12 600 000,00	10 000 000,00	0	0	0	0
ООО "СМАУМ"	4101146243	1114101000335	5,31	0	0	0	0	0	0
ООО "САМАУМ"	0201156280	109500049812	5,30	1 090 024,60	0	0	1 090 024,60	0	0
ООО "СКОРБИНА"	0501200030	1114501000317	5,29	29 379 286,00	4 000 000,00	0	0	0	0
ООО "К ПЕРМОНЦ"	2508000002	1102500000339	5,25	4 000 000,00	800 000,00	0	800 000,00	0	0
ООО "ИВАНОВ"	1107000052	1097101119204	5,24	298 160 130,53	142 091 249,00	0	1 166 905,00	0	0
ЗАО "ЭКОПРОММАШИНАКОМПАНИ"	0101000000	1096510000008	5,20	179 167 745,52	69 650 000,00	0	0	0	0
ООО "БАР"	0503000272	10920000754150	5,14	0	0	0	0	0	0
ООО "ВЕГЕР УДАЧА"	1107001480	1094191020922	5,09	10 299 518,00	1 642 618,00	0	2 456 900,00	0	0
ООО "СОЛНЦЕ УДАЧА"	0203010961	1124177000003	5,07	99 453 755,38	6 600 000,00	0	0	0	0
ООО "ПАР"	2705010244	1092709015005	5,00	190 631 446,40	31 671 365,00	0	82 215 736,00	0	3 816 999,60

Рис. 11. Скриншот интерфейса ИС с результатами синтеза рейтинговых оценок субъектов обстановки на основе разработанного алгоритма

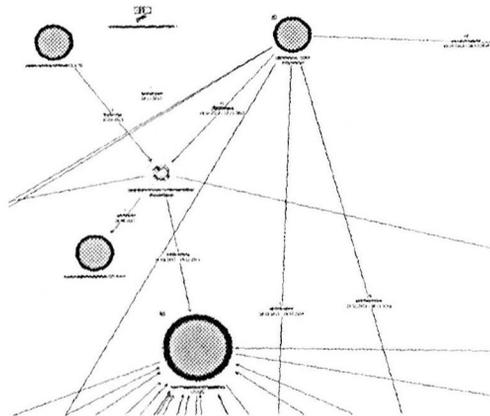


Рис. 12. Использование рейтинговых оценок в пользовательских интерфейсах ИС

В табл. 6 представлены результаты сравнительного анализа ИС Росфинмониторинга до и после исследования в части алгоритмов обработки данных для оценки обстановки и выходных данных.

Таблица 6. Результаты сравнительного анализа ИС Росфинмониторинга

Группа критерия сравнения	Критерий сравнения	До проведения исследования	После внедрения результатов исследования
Алгоритмы обработки информации	<ul style="list-style-type: none"> Алгоритм синтеза рейтинговых оценок субъектов обстановки Алгоритм генерации информации о девиантной деятельности на естественном языке 	-	+
Результат обработки информации	<ul style="list-style-type: none"> Рейтинговые оценки субъектов обстановки Логические правила идентификации девиантных субъектов 	-	+

В диссертации предложена методика визуализации информации об обстановке (схема на рис. 13). В основе методики лежит предложенный алгоритм синтеза рейтинговых оценок субъектов обстановки, адаптированный для обработки агрегированных данных по субъектам Федерации ДФО. Методика позволяет определить вариацию исходных показателей, характеризующую девиантную деятельность в области биоресурсов в субъектах Федерации ДФО.

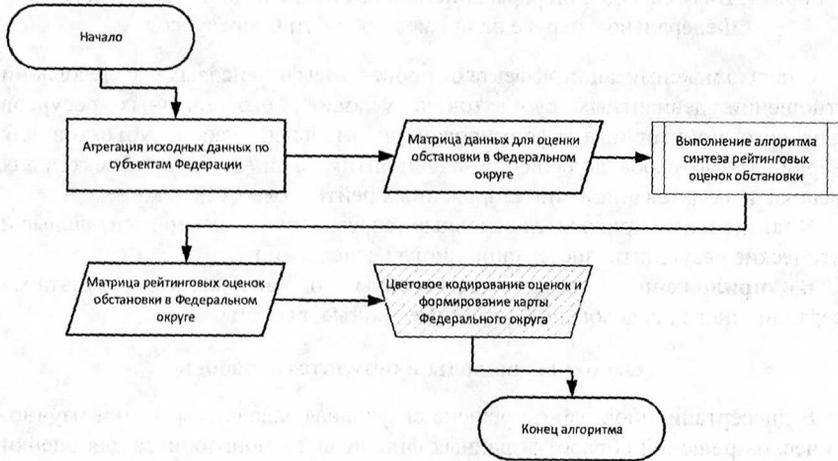


Рис. 13. Схема методики визуализации информации об обстановке в Федеральном округе

Заштрихованный элемент требует участия оператора ИС. Результат применения методики на примере области биоресурсов в Дальневосточном Федеральном округе представлен на рис. 14.

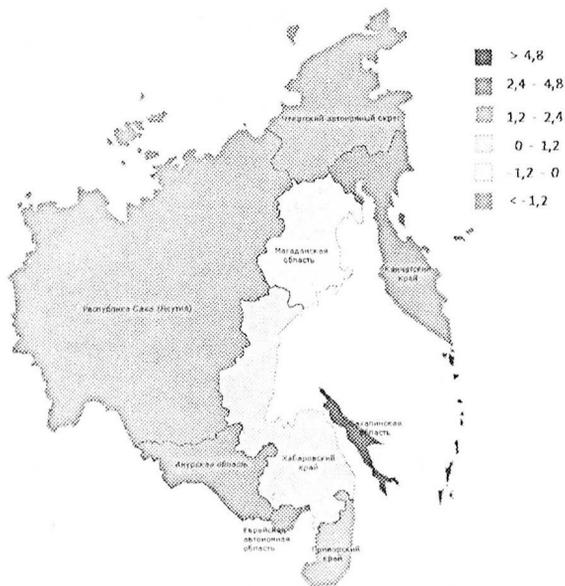


Рис. 14. Визуализация информации об обстановке в Дальневосточном Федеральном округе на примере области биоресурсов

С целью максимизации эффекта от проведения финансовых расследований в отношении девиантных субъектов в условиях ограниченных ресурсов финансового мониторинга, сформирован оптимальный набор субъектов для проверки. Это потребовало решения задачи оптимизации выбора субъектов для проверок с использованием синтезированных рейтинговых оценок.

В заключении приводятся основные теоретические, экспериментальные и практические результаты диссертационного исследования.

В приложении представлены акты о внедрении результатов диссертационного исследования, исходные данные, расчёты.

Основные выводы и результаты работы

В диссертационной работе решена актуальная задача разработки научно-технических решений обработки данных финансового мониторинга для оценки обстановки, исключая экспертный субъективизм и удовлетворяющих критерию к временным и ресурсным показателям процесса обработки данных. Программная реализация предложенных решений позволила модернизировать ИС Росфинмониторинга для автоматизированной обработки данных в области биоресурсов.

Основные результаты диссертационной работы заключаются в следующем:

1. Проведен анализ существующих в финансовом мониторинге подходов к обработке информации на этапе оценки обстановки. Впервые определены основные этапы обработки данных при оценке обстановки.

2. В соответствии с общепринятыми показателями эффективности автоматизированных систем управления автором предложено наряду с принятым в финансовом мониторинге временным показателем использовать ресурсный показатель для описания процесса обработки информации на этапе оценки обстановки. Декомпозированы временные и ресурсные показатели процессов обработки данных при решении частных задач.

3. Сформирован критерий к временным и ресурсным показателям процесса обработки данных при решении частных задач.

4. Впервые формализованы и поставлены частные задачи Z_1, Z_2, Z_3 обработки данных в ИС финансового мониторинга на этапе оценки обстановки.

5. Впервые предложена комплексная методика автоматизации процесса оценки обстановки в ИС финансового мониторинга, потребовавшая разработки алгоритмов синтеза рейтинговых оценок субъектов и идентификации девиантной деятельности.

6. Разработан новый алгоритм синтеза рейтинговых оценок субъектов обстановки в ИС финансового мониторинга, учитывающий некоторую априори неизвестную вариацию исходных показателей, характеризующую девиантную деятельность субъектов. Алгоритм позволяет решить частную задачу Z_2 оценки обстановки. Синтезируемые таким образом рейтинговые оценки лишены экспертного субъективизма. Внедрение программной реализации алгоритма в ИС позволило сформировать рейтинговые оценки субъектов обстановки в области биоресурсов (состоящей из 2292 субъектов).

7. Впервые проверена гипотеза о применимости логнормального закона для описания зависимости количества субъектов в области биоресурсов от меры их вовлечённости в девиантную деятельность. Гипотеза проверена по критерию хи-квадрат при уровне значимости $p \leq 0,05$. Проверка показала справедливость этой гипотезы.

8. Разработан новый алгоритм генерации информации о девиантной деятельности субъектов в целях решения частной задачи Z_3 оценки обстановки. Внедрение программной реализации алгоритма в ИС позволило сгенерировать логические правила на естественном языке, позволяющие компетентным органам своевременно идентифицировать девиантную деятельность в области биоресурсов. Предложенный алгоритм обладает высокой точностью идентификации девиантных субъектов, что подтверждается низкими значениями вероятности ошибки классификации, полученной при тестировании построенной модели на контрольной выборке. Вероятность ошибочной классификации удовлетворяет неравенству $0,00 < \bar{p} < 0,06$ при коэффициенте доверия $P = 0,975$.

9. Разработана новая методика визуализации обстановки по данным финансового мониторинга, что позволило сформировать карту обстановки в Дальневосточном Федеральном округе на примере области биоресурсов.

10. Предложена комплексная методика автоматизации процесса оценки обстановки в ИС финансового мониторинга, в соответствии с которой модернизирована ИС Росфинмониторинга для обработки данных в области биоресурсов, что позволило значительно повысить качество аналитических процессов Росфинмониторинга. Это подтверждается исключением субъективного фактора экспертов при синтезе рейтинговых оценок обстановки и информации о девиантной деятельности субъектов, сокращением временных затрат на синтез рейтинговых оценок в 19 раз, сокращением временных затрат на синтез информации о девиантной деятельности субъектов в 10 раз, сокращением ресурсных затрат при решении каждой из задач в 3-5 раз.

11. Результаты диссертационного исследования внедрены и используются в Центральном аппарате Росфинмониторинга, кафедре финансового мониторинга НИЯУ МИФИ, что подтверждается актами о внедрении результатов диссертационного исследования.

12. Экспериментальные исследования и результаты внедрения подтвердили теоретические положения диссертации о возможности решения актуальной научно-технической задачи обработки данных финансового мониторинга в области биоресурсов.

Основные публикации по теме диссертации

Публикации, представленные в международной базе цитирования Scopus:

1. Denisenko A.S. Application of Principal Components Analysis Results in Visual Network Analysis / A.S. Denisenko, G.O. Krylov // *Biosciences Biotechnology Research Asia*. — 2015. — Vol. 12 (1). — p.609-617. (0,593 п.л.).

2. Денисенко А.С. Анализ данных о деятельности кредитных организаций с использованием программы интерактивного визуального анализа многомерных данных / И.Е. Мильман, А.П. Пахомов, В.В. Пилюгин, Е.Е. Писарчик, А.А. Степанов, Ю.М. Бекетнова, А.С. Денисенко // *Научная визуализация (только электронный вид)*. — 2015. — Т.7. — №1. — С.45-64. (0,681 п.л.).

Публикации, представленные в рецензируемых научных изданиях ВАК России:

3. Денисенко А.С. Факторный анализ и интегральные оценки причастности промышленных предприятий к отмыванию доходов / А.С. Денисенко // *Глобальный научный потенциал*. — 2014. — №8. — С.96-100. (0,275 п.л.).

4. Денисенко А.С. Обработка и интерпретация результатов, полученных путём анализа данных о финансовых потоках компаний методом главных компонент / А.С. Денисенко // *Аудит и финансовый анализ*. — 2015. — №1. — С.458-460. (0,245 п.л.).

5. Денисенко А.С. О применении метода главных компонент в задачах финансового мониторинга / А.С. Денисенко, Г.О. Крылов, И.А. Корнев //

Известия Самарского научного центра РАН. — 2015. — Т. 17. — №2 (4). — С. 1131-1140. (0,51 п.л.).

Публикации в других изданиях:

6. Денисенко А.С. Подходы к классификации финансовых операций и их участников в целях повышения эффективности исполнения возложенных на Росфинмониторинг функций по противодействию отмыванию доходов, полученных преступным путём, и финансированию терроризма / А.С. Денисенко, К.И. Гобрусенко // Информатизация и информационная безопасность правоохранительных органов: сб. матер. XXII Всерос. научн. конф. — Москва, 2013. — С.102-105. (0,378 п.л.).

7. Денисенко А.С. О применении методов факторного анализа в задачах финансового мониторинга / А.С. Денисенко, К.И. Гобрусенко, Г.О. Крылов // Информатизация и информационная безопасность правоохранительных органов: сб. трудов XXIV Всерос. научн. конф. — Москва, 2015. — С.135-139. (0,41 п.л.).

Подписано в печать 19.05.2017. Формат 60x84 1/16

Печ.л. 1,75. Тираж 100 экз.

Отпечатано в типографии ШИЯУ МНФИ.

115409, Москва, Каширское ш., 31