

Математические задачи систем оценки и мониторинга сложных процессов. Обзор постановок и результатов

А. П. Рыжов

В работе описывается технология оценки и мониторинга сложных процессов, разрабатываемая на кафедре МАТИС с начала 90-х годов. Приводится содержательная постановка проблемы информационного мониторинга, описываются технологические и математические аспекты разработки систем информационного мониторинга. Приводится обзор основных результатов.

Ключевые слова: системы оценки и мониторинга, нечеткие множества, оценка нечеткости, поиск в нечетких базах данных.

1. Введение

Системы оценки и мониторинга сложных процессов — специализированные человеко-компьютерные интеллектуальные информационные системы, предназначенные для оценки состояния некоторого процесса на базе фрагментарной, ненадежной, возможно недостоверной информации о нем и моделирования возможных вариантов его развития. В отличие от широко применяемых систем мониторинга (например, экологического, медицинского, технического и т. п.), базирующихся на анализе большого количества результатов показаний различных датчиков в рамках некоторой математической модели (например, уравнения теплопроводности), для интересующих нас систем «входом» является не показания приборов, а некоторая информация (в частности, и показания измерительных устройств тоже, но не только они). Говоря совсем популярно, мы разрешаем системам оценки и

мониторинга использовать результаты измерений не только технических устройств, но и человека как измерительного прибора.

2. Содержательная постановка задачи

Многие процессы в бизнесе, экономике, политике и других областях, называемых слабо (или плохо) формализуемыми, невозможно представить в виде набора дифференциальных уравнений, автоматов и других математических средств анализа динамических систем, однако специалисты как-то работают с такими процессами. В общем виде такая работа заключается в оценке текущего состояния процесса на основе всей доступной информации, построении прогнозов его развития и выработке рекомендаций по управлению им, исходя из целей, стоящих перед специалистом. Можно привести следующие примеры процессов из различных областей:

- поведения клиента (маркетинг);
- продвижение кандидата на выборах (политология);
- диагностика (медицина);
- оценка кредитного риска (финансовый анализ);
- оценка страховых рисков (страховое дело);
- развитие технологий/научных исследований (наукометрия, технологии).

Свойство перечисленных проблем — «слабая формализация» (наличие количественных и качественных признаков, отсутствие математических моделей), структурная организация (процесс имеет некоторую структуру) и наличие человека как активного элемента системы.

В качестве примера процессов, не являющихся таковыми, можно привести взаимодействие двух тел или распространение колебаний в однородной среде. Имеются математические модели таких процессов, информация измерима и доступна, результат можно вычислить для любого момента времени.

Будем называть задачу оценки текущего состояния процесса и построении прогнозов его развития **задачей оценки и мониторинга**, а человеко-компьютерные системы, обеспечивающие информационную поддержку подобного рода информационных задач, **системами оценки и мониторинга**.

3. Понятие систем информационного мониторинга

Основными элементами систем оценки и мониторинга являются информационное пространство и аналитик. Существенными являются следующие их свойства.

Информационное пространство представляет собой совокупность различных информационных элементов, которые можно охарактеризовать следующим образом:

- *разнородность носителей* информации, то есть фиксация информации в виде статей, газетных заметок, компьютерном виде, аудио- и видеоинформация и т. п.;
- *фрагментарность*: информация чаще всего относится к какому-либо фрагменту процесса, причем разные фрагменты могут быть по-разному «покрыты» информацией;
- *разноуровневость*: информация может относиться ко всему процессу в целом, к некоторой его части, к конкретному элементу;
- *различная степень надежности*: информация может содержать конкретные данные различной степени надежности, косвенные данные, результаты выводов на основе надежной информации или косвенные выводы;
- *возможная противоречивость*: информация из различных источников может совпадать, слегка различаться или вообще противоречить друг другу;
- *изменяемость во времени*: процесс развивается во времени, поэтому и информация в разные моменты времени об одном и том же его элементе может и должна различаться;
- *возможная тенденциозность*: информация отражает определенные интересы источника информации, поэтому может носить тенденциозный характер. В частном случае она может являться намеренной дезинформацией (например, для политических проблем или для проблем, связанных с конкуренцией).

Аналитики являются активным элементом систем оценки и мониторинга и, наблюдая и изучая элементы информационного пространства, делают выводы о состоянии проблемы и перспективах ее раз-

вития с учетом перечисленных выше свойств информационного пространства. Обычно аналитики образуют некоторую структуру (министерство, агентство, консультационную службу, и т. п.) В этом случае каждый аналитик «нижнего уровня» имеет дело с некоторой частью процесса и работает с элементами информационного пространства, аналитики «более высокого уровня» имеют дело с более крупными фрагментами процесса или процессом в целом, и работают уже с выводами предыдущих аналитиков. При этом они могут ознакомиться с выводами более низкого уровня вплоть до элементов информационного пространства.

Взаимодействие основных элементов систем оценки и мониторинга представлено на рис. 1.

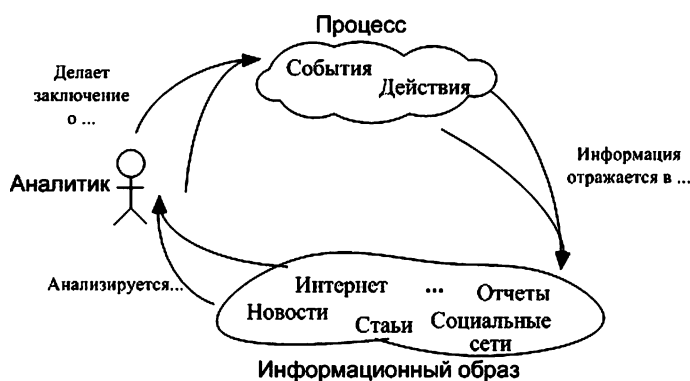


Рис. 1. Взаимодействие основных элементов систем оценки и мониторинга.

4. Технологические основы разработки систем оценки и мониторинга

Учитывая особенности информации и конкретных способов ее обработки, основные черты технологии оценки и мониторинга можно изложить следующим образом.

Эта технология базируется на использовании ряда приемов, позволяющих обрабатывать подобного рода информацию. В частности:

- для реализации возможности обработки информации из разнородных источников, в базе данных системы хранятся как сами документы, так и ссылки на них с оценкой содержащейся в них информации, данной аналитиком;
- для возможности обработки фрагментарной информации используется структурная модель процесса в виде графа;
- обработка разноуровневой информации достигается за счет предоставления пользователю возможности отнести оценку конкретного информационного материала к разным вершинам модели;
- обработка информации различной степени надежности и обладающей возможной противоречивостью или тенденциозностью достигается за счет использования лингвистических оценок данной информации;
- изменчивость во времени учитывается фиксацией даты поступления информации при оценке конкретного материала, то есть время является одним из элементов описания объектов системы.

Таким образом, системы, построенные на базе этой технологии, позволяют иметь развивающуюся во времени модель процесса на основе оценок аналитиков, подкрепленную ссылками на все информационные материалы, выбранные ими, с общими и частными оценками состояния процесса или его аспектов. Использование времени как параметра системы позволяет проводить ретроспективные анализ и строить прогнозы развития процесса (отвечать на вопросы типа «Что будет, если...?»). В последнем случае возникает возможность выделения «критических путей», то есть таких элементов модели, малое изменение которых может вызвать значительные изменения в состоянии всего процесса. Знание таких элементов имеет большое практическое значение и позволяет выявить «слабые места» в процессе на текущий момент времени, разработать мероприятия по блокированию нежелательных ситуаций или провоцированию желательных, то есть в некоторой степени управлять развитием процесса в интересах организации, его отслеживающей.

Концептуальная схема систем оценки и мониторинга представлена на рис. 2.

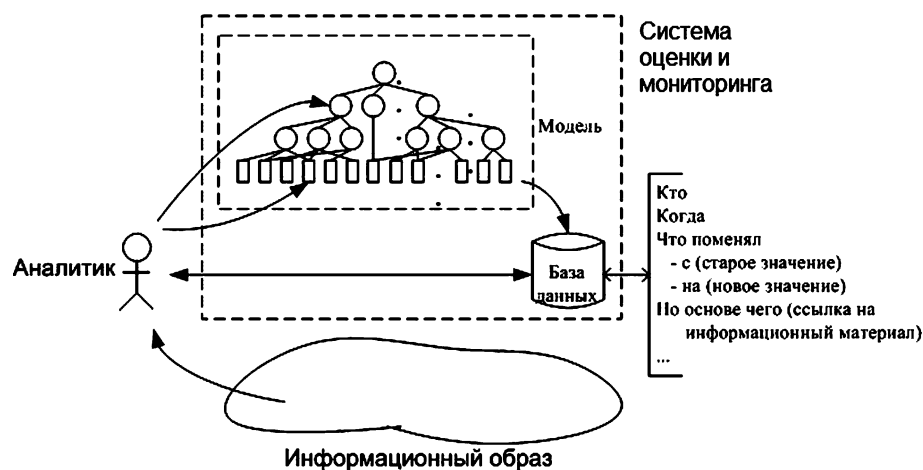


Рис. 2. Концептуальная схема систем информационного мониторинга.

5. Математические основы систем информационного мониторинга

Системы оценки и мониторинга можно отнести к классу иерархических нечетких дискретных динамических систем. Теоретическую основу такого класса систем составляет методы анализа иерархий, теория нечетких множеств, теория измерений, теория автоматов, дискретная математика, которые были разработаны в работах Саати (Saaty) [13], Месаровича (Messarovich) [14], Заде (Zadeh) [1], Пфанцгаля (Pfanngal) [5], Кудрявцева В. Б. [3] и других авторов.

Для эффективного практического применения предложенных технологических решений (раздел 4) необходима проработка ряда теоретических проблем, результаты которой приводятся ниже.

Предполагается, что аналитик описывает степень противоречивости полученной информации и имеющейся или возможной реализации некоторых процессов в предметной области в виде лингвистических значений. Субъективная степень удобства такого описания зависит от набора и состава таких лингвистических значений. Поясним сказанное на модельном примере.

Пример 1. Пусть нас просят описать рост человека. Рассмотрим две крайние ситуации.

Ситуация 1. Разрешено использовать только два значения: «низкий» и «высокий».

Ситуация 2. Разрешено использовать много значений: «очень низкий», «не очень высокий», ..., «не высокий и не низкий», ..., «очень — очень высокий».

Ситуация 1 является неудобной. Действительно, для многих людей оба разрешенных значения могут не подходить, и, описывая их, мы выбираем между двумя «плохими» значениями.

Ситуация 2 также является неудобной. Действительно, при описании конкретного человека могут подходить несколько из разрешенных значений. Мы опять испытываем дискомфорт, но теперь от того, что мы вынуждены выбирать между двумя или более «хорошими» значениями. Может ли быть оптимальное в этом смысле множество лингвистических значений?

Предполагается, что система отслеживает развитие проблемы, то есть изменение ее во времени. Предполагается также, что она интегрирует оценки различных аналитиков. Это значит, что один объект может описываться разными аналитиками. Поэтому желательно иметь гарантии того, что разные аналитики описывают один и тот же объект наиболее «единообразно».

Исходя из сделанных замечаний мы можем сформулировать первую проблему следующим образом.

Проблема 1. Можно ли, учитывая некоторые особенности восприятия человеком объектов реального мира и их описания, сформулировать правило выбора оптимального множества значений признаков, по которым описываются эти объекты? Возможны два критерия оптимальности:

Критерий 1. Под оптимальными понимаются такие множества значений, используя которые человек испытывает минимальную неопределенность при описании объектов.

Критерий 2. Если объект описывается некоторым количеством экспертов, то под оптимальными понимаются такие множества значений, которые обеспечивают минимальную степень рассогласования описаний.

Показано, что мы можем сформулировать методику выбора оптимального множества значений качественных признаков. Эта методика базируется на введенном в [9] понятии степени нечеткости

лингвистической шкалы, которая является обобщением понятия степени нечеткости множества и обладает рядом естественных свойств, изученных в [10]. Более того, показано, что такая методика является устойчивой, то есть возможные при построении функций принадлежности естественные маленькие ошибки не оказывают существенного влияния на выбор оптимального множества значений [11]. Множества, оптимальные по критериям 1 и 2 совпадают.

Технология оценки и мониторинга предполагает хранение информационных материалов (или ссылки на них) и их лингвистических оценок в базе данных системы. В связи с этим возникает следующая проблема.

Проблема 2. Можно ли определить показатели качества поиска информации в нечетких (лингвистических) базах данных и сформулировать правило выбора такого множества лингвистических значений, использование которого обеспечивало бы максимальные показатели качества поиска информации?

Показано, что можно ввести показатели качества поиска информации в нечетких (лингвистических) базах данных и формализовать их. Показано, что возможно сформулировать методику выбора оптимального множества значений качественных признаков, которое обеспечивает максимальные показатели качества поиска информации. Более того, показано, что такая методика является устойчивой, то есть возможные при построении функций принадлежности естественные маленькие ошибки не оказывают существенного влияния на выбор оптимального множества значений. Данные результаты изложены в [7].

Важной проблемой при разработке и систем оценки и мониторинга является задача выбора адекватных операторов агрегирования информации в модели процесса. Эта задача возникает в силу того, что системы оценки и мониторинга ориентированы на обработку разноразмерной фрагментарной информации. Это означает, что при вводе информации в систему пользователь может осуществлять привязку информационных материалов к узлам различных уровней иерархии в модели проблемы (и, соответственно подтверждать/изменять их оценки). Это коренным образом отличает системы оценки и мониторинга от аналогичных иерархических систем (например, поддержки принятия решений), в которых разрешается менять только оценки

нижнего уровня иерархии. Указанное допущение позволяет использовать системы оценки и мониторинга при решении значительно более широкого класса практических задач, однако платой за это является необходимость разработки соответствующей теории и создания инструментария выбора адекватных операторов агрегирования информации. Поясним суть проблемы на следующем примере.

Пример 2. Пусть текущее состояние модели имеет вид, изображенный на рис. 3.

Будем для простоты считать, что оценки пользователя представляются в виде числа из отрезка $[0,1]$. Пусть, далее, в качестве операторов агрегирования информации используется операция взятия максимума. Как не трудно видеть, в этом случае модель является согласованной. Рассмотрим две ситуации.

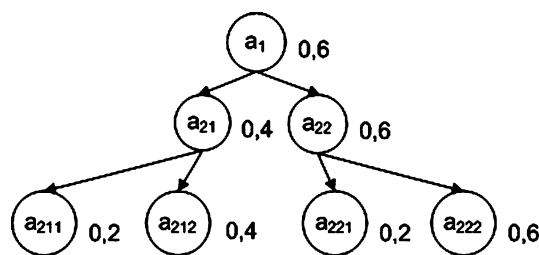


Рис. 3. Упрощённый пример состояния модели.

Ситуация 1. Пользователь изменил значение в узле a_{211} с текущего на 0.3. В этом случае согласованность остается.

Ситуация 2. Пользователь изменил значение в узле a_{22} с текущего на 0.5. В этом случае введенная оценка входит в противоречие как с оценками нижнего уровня, так и с оценкой верхнего уровня. Как быть в этом случае? Если оценки узлов a_1 и a_{212} являются не результатом вычислений в системе, а введены пользователем на основе предыдущей информации, мы должны для сохранения согласованности модели либо «заставить» пользователя изменить их, либо выбрать другие операторы агрегирования информации. Первый путь, несмотря на насильственный характер, иногда оказывается применимым при условии развитой подсистемы работы с архивом (например, в ситуации, когда оценки, которые необходимо менять, являются достаточно «старыми» и/или поставлены пользователем на основе недостаточно

надежной информации). Однако, если пользователь одинаково уверен в правильности всех оценок, необходимо менять оператор агрегирования информации.

Проблема 3. Можно ли предложить процедуры выбора операторов агрегирования информации в нечетких иерархических динамических системах, минимизирующих противоречивость модели процесса в системах оценки и мониторинга?

Можно выделить следующие подходы к решению этой проблемы, базирующиеся на различных интерпретациях операторов агрегирования информации: геометрический, логический и подход на основе обучения, включающий в себя обучение на основе генетических алгоритмов и обучение на основе нейронных сетей. Описание этих подходов, формулировка и решение возникающих в их рамках задач представлены в [6, 8, 13].

6. Аналитические возможности систем оценки и мониторинга

В рамках систем оценки и мониторинга возможно решение прямой и обратной задачи.

Прямая задача заключается в нахождении критических путей — таких элементов модели, малое изменение которых приводит к изменению всех выше лежащих узлов, включая верхний. Практическая значимость критических путей обсуждена в разделе 4. Для большого класса операторов агрегирования информации возможно вычисление степени критичности каждого элемента модели. Задача также может решаться перебором для достаточно компактной модели.

Обратная задача позволяет оптимизировать бюджет на достижение определённого уровня целевого показателя. Если задан бюджет, и мы знаем стоимость изменения состояния узла модели, то возможно нахождение тех узлов, изменение которых находится в рамках заданного бюджета и даёт максимальной эффект повышения корневого узла модели. Формализовать такую постановку можно следующим образом. Пусть модель содержит N узлов, к i -му узлу приписано m_i значений: $a_i = \{a_i^1, a_i^2, \dots, a_i^{m_i}\}$ ($i = 0, 1, \dots, N - 1$) (рис. 4). Пусть задано: X — бюджет, c_i — стоимость изменения состояния i -го узла

(перевода его из состояния a_i^k в состояние a_i^{k+1} , $k = 1, \dots, m_i - 1$; будем для простоты записи считать такую стоимость одинаковой для всех k). Обозначим через Δa_i силу изменения узла a_i . Эта величина равна q для ситуации, когда состояние узла поменялось с a_i^k на a_i^{k+q} . Для множества номеров узлов $I = \{1, 2, \dots, N - 1\}$ обозначим $S(I)$ множество всех его непустых подмножеств.

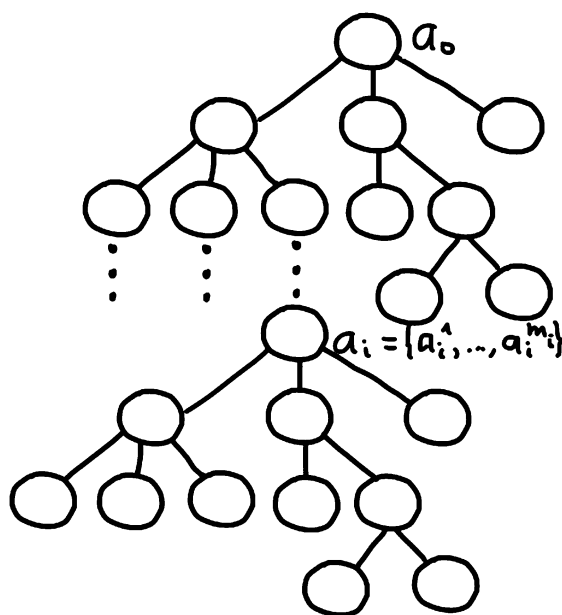


Рис. 4. Пример модели процесса.

Задача 1. Найти множество номеров узлов $s \in S(I)$: $\Delta a_0 \rightarrow \max$, $\sum_{i \in s} c_i \leq X$.

Это задача нахождения максимального эффекта в рамках заданного бюджета.

При заданных условиях возможно и решение сопряжённой задачи:

Задача 2. Найти множество номеров узлов $s \in S(I)$: $\sum_{i \in s} c_i \rightarrow \min$, $\Delta a_0 \geq q$.

Это задача нахождения минимального бюджета, позволяющего достичь необходимого эффекта.

Примерами таких задач, при наличии соответствующих моделей, могут быть оценка и повышение капитализации технологических стартапов [7]; повышение рейтинга банков, университетов, спортивных команд (обычно структура рейтинга известна — например, [20]); повышение инвестиционной привлекательности компаний, регионов, стран (обычно рейтинговые агентства публикуют структуру оценки); снижение уровня технологических рисков опасных производств (при наличии карты технологических рисков); повышение устойчивости развития бизнеса [11] и многие другие задачи.

Перечисленные задачи возможно решать как в рамках фрагментов модели (например, оптимизация работы соответствующих подразделений компании), так и модели в целом (например, оптимизация работы компании в целом).

Аналитические возможности систем оценки и мониторинга суммированы на рис. 5.

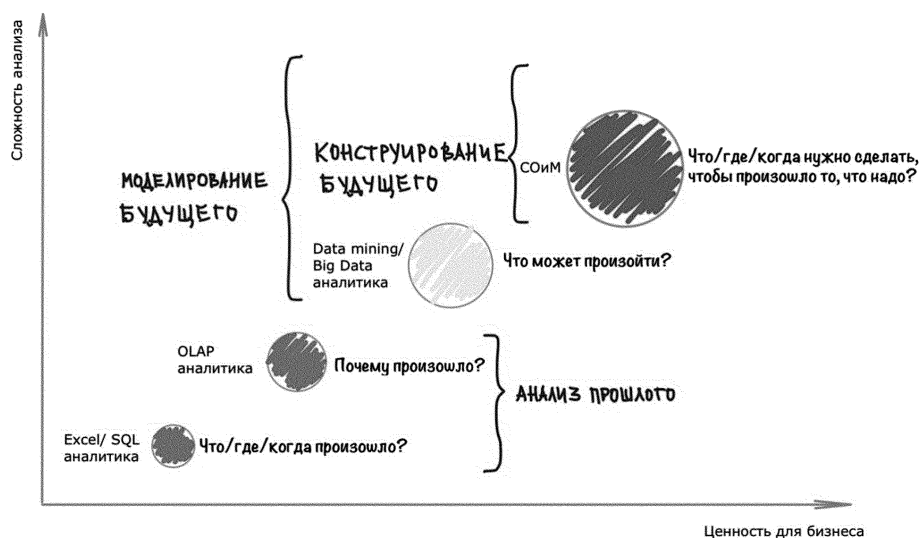


Рис. 5. Эволюция аналитических инструментов.

Текущий уровень почти всех компаний — OLAP аналитика; лишь немногие компании используют Big Data аналитику; аналитические инструменты систем оценки и мониторинга используются в единичных международных и правительственных организациях, а также в крупном международном высокотехнологическом бизнесе. Использо-

вание аналитических возможностей систем оценки и мониторинга — это естественный эволюционный этап развития аналитических инструментов.

7. Заключение

Итак, системы информационного мониторинга:

- решают задачу мониторинга;
- позволяют аналитику вводить информацию из всех возможных (доступных) источников «естественным» образом (вводятся только оценки);
- хранят историю развития процесса, оценивают его текущее состояние, прогнозируют и моделируют будущее.

Системы информационного мониторинга эффективны, когда нет (нельзя построить) математическую модель процесса в виде уравнений, автоматов, и т. п. Разработка таких систем возможна, когда можно построить «семантическую модель» процесса в виде набора понятий и их взаимосвязей, и когда поступает и анализируется реальная информация (возможно обучение или настройка). Возможна разработка оптимальных систем с точки зрения удобства ввода информации аналитиком, согласованности мнений аналитиков, информационного обеспечения ввода информации, и моделирования.

На основе описанных методов разработаны системы оценки и мониторинга процессов создания странами специальных видов технологий и материалов в интересах управления гарантией Международного Агентства по Атомной Энергии [18], риска развития сердечнососудистых заболеваний в РФ [1], создания изделий микроэлектроники (оценка технологических стартапов) в интересах компании Cadence Design Systems [4]. Возможности применения подобных систем в оценке и мониторинге кризисных ситуаций и в стратегическом управлении компаниями описаны в [17] и [19] соответственно.

Автор хотел бы поблагодарить профессора В. Б. Кудрявцева и коллектив кафедры МАТИС и лаборатории ПТК за многолетнюю поддержку этого направления, профессора Л. Заде (США), профессора Е. Керре (Бельгия), профессора Я. Нишиваки (Япония), А. Фатту (Австрия) и В. Мацке (Германия) за плодотворную работу над проектами на базе технологии информационного мониторинга.

Список литературы

- [1] Ахмеджанов Н. М., Жукоцкий А. В., Кудрявцев В. Б., Оганов Р. Г., Расторгуев В. В., Рыжов А. П., Строгалов А. С. Информационный мониторинг в задаче прогнозирования риска развития сердечно-сосудистых заболеваний // Интеллектуальные системы. — 2003. Т. 7, вып. 1–4. — С. 5–38.
- [2] Заде Л. А. Понятие лингвистической переменной и его применение к принятию приближенных решений. — М.: Мир, 1976.
- [3] Кудрявцев В. Б., Алешин С. В., Подколзин А. С. Введение в теорию автоматов. — М.: Наука, 1985.
- [4] Лебедев А. А., Рыжов А. П. Оценка и мониторинг проектов разработки высокотехнологических изделий микроэлектроники // Известия ТРТУ. Тематический выпуск «Интеллектуальные САПР». — 2006. № 8. — ISBN 5–8327–0249–2. — С. 93–99.
- [5] Пфанцгаль И. Теория измерений / Пер. с англ. — М.: Мир, 1976.
- [6] Рогожин С. В., Рыжов А. П. О нечетко заданных классах функций k -значной логики // V Всероссийская конференция «Нейрокомпьютеры и их применение». Сборник докладов. — Москва, 17–19 февраля 1999 г. — С. 460–463.
- [7] Рыжов А. П., Анохина Е. С. Об одном подходе к оценке технологических стартапов средствами технологии оценки и мониторинга сложных процессов // Материалы VII ежегодной международной научно-практической конференции «Инвестиции. Инновации. Информационные технологии». — Москва, 27–28 февраля 2015 г. Российская академия народного хозяйства и государственной службы при Президенте РФ. — С. 25–32.
- [8] Рыжов А. П. Модели поиска информации в нечеткой среде. — М.: Изд-во Центра прикладных исследований при мех.-мат. ф-те МГУ, 2004.
- [9] Рыжов А. П. Об агрегировании информации в нечетких иерархических системах // Интеллектуальные системы. — 2001. Т. 6, вып. 1–4. — С. 314–337.
- [10] Рыжов А. П. О степени нечеткости размытых характеристик // Математическая кибернетика и ее приложения в биологии. Под

ред. Л. В. Крушинского, С. В. Яблонского, О. Б. Лупанова. — М.: Изд-во МГУ, 1987. — С. 60–77.

- [11] Рыжов А. П. Оценка и мониторинг бизнеса на основе сбалансированной системы показателей // Материалы V ежегодной международной научно-практической конференции «Инвестиции. Инновации. Информационные технологии». — Москва, 1–2 марта 2013 г. Российская академия народного хозяйства и государственной службы при Президенте РФ. — ISBN 978–5–7749–0847–9. — С. 47–49.
- [12] Рыжов А. П. Оценка степени нечеткости и ее применение в системах искусственного интеллекта // Интеллектуальные системы. — 1996. Т. 1, вып. 1–4. — С. 95–102.
- [13] Рыжов А. П. Элементы теории нечетких множеств и измерения нечеткости. — М.: Диалог-МГУ, 1998.
- [14] Рыжов А. П., Федорова М. С. Генетические алгоритмы в задаче выбора операторов агрегирования информации в системах информационного мониторинга // V Всероссийская конференция «Нейрокомпьютеры и их применение». Сборник докладов. — Москва, 17–19 февраля 1999 г. — С. 267–270.
- [15] Саати Т. Анализ иерархических процессов. — М.: Радио и связь, 1993.
- [16] Messarovich M. D., Macko D., Takahara Y. Theory of hierarchical multilevel systems. — N.Y.–London: Academic Press, 1970.
- [17] Ryjov A. Basic principles and foundations of information monitoring systems // Monitoring, Security, and Rescue Techniques in Multi-agent Systems. — Springer, 2005. — ISBN 3–540–23245–1, ISSN 16–15–3871. — P. 147–160.
- [18] Ryjov A., Belenki A., Hooper R., Pouchkarev V., Fattah A., Zadeh L. A. Development of an Intelligent System for Monitoring and Evaluation of Peaceful Nuclear Activities (DISNA). — IAEA, STR-310. — Vienna, 1998.
- [19] Ryjov A. Information Monitoring Systems as a Tool for Strategic Analysis and Simulation in Business // International Conference on Fuzzy Sets and Soft Computing in Economics and Finance FSSCEF

2004 Proceedings. — Saint-Petersburg, Russia, June 17–20, 2004. — ISBN 968–489–028–1. — Vol. II. — P. 511–520.

- [20] WASEDA — IAC 10th International E-Government Ranking 2014.
http://www.e-gov.waseda.ac.jp/pdf/2014_e-gov_press_release.pdf.