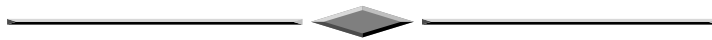


КАРТОГРАФИЯ И ГЕОИНФОРМАТИКА



УДК 519.254

МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В РЕГИОНАЛЬНЫХ СИСТЕМАХ АЭРОКОСМИЧЕСКОГО МОНИТОРИНГА

Александр Владимирович Замятин

Национальный исследовательский Томский государственный университет, 634050, Россия, г. Томск, пр. Ленина, 36, доктор технических наук, заведующий кафедрой теоретических основ информатики, тел./факс: (3822)52–94–96, e-mail: zamyatin@mail.tsu.ru.

Обсуждается возможность построения системы аэрокосмического мониторинга с применением методов интеллектуального анализа данных, таких как ассоциативные методы, а также классификационные методы на основе байесовского решающего правила и методов на основе деревьев принятия решений. Обсуждаются примеры задач регионального аэрокосмического мониторинга (экологический контроль, региональный лесопожарный мониторинг, мониторинг локальных природных и антропогенно-спровоцированных явлений и катастроф и др.) и соответствующих им методов анализа данных (регрессия, обнаружение аномалий, пространственное прогнозирование и др.) при поддержке принятия решений. Рассматриваются проблемы адаптации этих методов к архивам данных дистанционного зондирования Земли, предполагающие учет как особенностей методов анализа данных, так и специфических аспектов построения систем мониторинга. Обсуждается пример структуры транзакционных данных для системы аэрокосмического мониторинга, включающей показания набора измерителей (датчиков) некоторой физической величины, справедливого для локации размещения датчика в некоторой окрестности. Предлагаются подходы к применению методов анализа данных в системах мониторинга с использованием дополнительного архива транзакционных данных такой структуры о состоянии среды, сформулированы концептуальные требования к такой системе, ее обобщенная структура и технология применения с беспилотным летательным аппаратом.

Ключевые слова: интеллектуальный анализ данных, система аэрокосмического мониторинга, данные дистанционного зондирования Земли, беспилотный летательный аппарат, классификация, регрессия, обнаружение аномалий, пространственное прогнозирование.

Введение

Стремительная технологическая эволюция последних лет в сфере информационно–коммуникационных технологий позволила сформировать существенный задел в части развитой программно-аппаратной инфраструктуры, поддерживающей постоянное пополнение архивов данных различной природы и назначения [1]. Отчетливо эти тенденции проявились при развитии наземных

сегментов систем аэрокосмического мониторинга, архивы которых ежедневно пополняются терабайтами данных дистанционного зондирования Земли (ДЗЗ), позволяя увеличивать и без того значительные архивы данных [2]. Наземный сегмент типовой региональной системы мониторинга, как правило, обеспечивает [3]:

- сбор заявок и распространение данных ДЗЗ различных спутниковых систем потребителям;
- подготовку данных для управления съемкой с летательного аппарата на основе соответствующих целеуказаний;
- прием и регистрацию данных ДЗЗ с заданной территории за сеанс связи;
- предварительную обработку с заданным уровнем;
- архивацию и каталогизацию принятых и обработанных данных ДЗЗ.

Видно, что основной задачей этой и других аналогичных систем является прием и накопление (архивирование) данных. Быстро растущие объемы пополняемых архивов данных существенно превышают способности человека в их практической полезной обработке и извлечении из них соответствующих знаний [3, 4]. Это характерно и для других областей человеческой деятельности, крайне ограниченных сегодня в возможностях содержательной обработки данных и извлечении полезных знаний [5, 6]. Именно поэтому в последние годы стремительное развитие получает область Data Mining (в отечественной литературе наиболее используемая аналогия – интеллектуальный анализ данных, ИАД), направленная на поиск и разработку методов извлечения из имеющихся данных знаний, позволяющих обеспечивать поддержку обоснованных и практически полезных управленческих решений.

Наибольший интерес к технологиям интеллектуальной обработки данных проявляют компании, работающие в условиях высокой конкуренции и имеющие четкую группу потребителей (розничная торговля, финансы, связь, маркетинг). Эти компании пытаются найти связь между «внутренними» (цена, востребованность продукта, компетентность персонала) и «внешними» (экономические показатели, конкуренция, демография клиентов и т. п.) факторами, позволяя оценить (прогнозировать) уровни продаж, удовлетворенности клиентов, доходов и др., существенно повышая свою экономическую эффективность.

Вместе с тем, достижения в этой области при построении региональных систем аэрокосмического мониторинга используются лишь отчасти [7–10]. Архивные и получаемые данные ДЗЗ применяют, главным образом, для экспертного визуального анализа (или применяют методы автоматизированной обработки – неконтролируемую и контролируемую классификацию [10]). Решение задач извлечения знаний и поддержки принятия решений с такими исходными данными и традиционно используемыми методами их обработки крайне затруднительно. Поэтому исследования перспективных возможностей области интеллектуального анализа данных для задач аэрокосмического мониторинга являются крайне актуальными. Актуальными будут и детали создания таких систем, способствующие развитию в них потенциала извлечения знаний и поддержки принятия решений.

Перспективные подходы к развитию потенциала систем мониторинга

Выделяют несколько стандартных типов закономерностей, которые позволяют относить используемые методы к методам извлечения знаний ИАД – ассоциация (включая последовательную ассоциацию), классификация (контролируемая и неконтролируемая), регрессионный анализ [11, 12]. Также в этой области рассматривают различные аспекты предварительной обработки данных (очистка данных, снижение размерности признакового пространства), обнаружение аномалий, визуализации и др. [13–15].

Ассоциация (англ. association) имеет место в случае, если несколько событий связаны друг с другом. В случае, если события связаны во времени, то имеет место последовательность (англ. sequential patterns). Метод применяют в торговой отрасли при нахождении типичных шаблонов покупок при анализе рыночной корзины (англ. market basket analysis). Примером практической реализации поиска ассоциативных правил является алгоритм «Априори» (англ. «Apriori», «A Priori») [16]. Данные в этом случае должны быть транзакционными (англ. transaction data). Их структура представляет собой таблицу, столбцы которой соответствуют некоторым параметрам, а каждая запись – набор значений определенного момента времени.

При анализе ассоциаций с учетом времени используют методы анализа последовательности (англ. sequence) и поиск последовательных ассоциаций (англ. sequential association – сиквенциальный анализ). Сиквенциальный анализ широко используется, например, в телекоммуникационных компаниях, для анализа данных об авариях на различных узлах сети [17]. Алгоритмы решения задач сиквенциального анализа построены на базе алгоритма «Априори», но отличаются учетом дополнительного параметра – времени совершения транзакции (например, «AprioriAll», «AprioriSome», «DynamicSome») [17]. Эти алгоритмы используют подход к генерации и отбору кандидатов, часто встречающихся последовательностей, а их применение предполагает наличие транзакционных данных, причем каждый параметр содержит время получения значения.

Методы классификации реализуются в случае, если имеется несколько заранее сформированных классов (групп, типов) объектов (контролируемая классификация) или если классы (группы, типы) заранее не заданы, а их количество и состав определяются автоматически (неконтролируемая классификация). Математический аппарат решения задачи классификации может быть самым разнообразным [10, 11, 18, 19]. Часто применяют методы байесовской классификации с нахождением апостериорной вероятности $p(\omega_j | x)$ события $x \in \omega_j$, вычисляемого исходя из априорной вероятности $p(\omega_i)$ и условной плотности распределения $p(x | \omega_i)$ [18]:

$$p(\omega_j | x) = \frac{p(\omega_j)p(x | \omega_j)}{\sum_{k=1}^M p(\omega_k)p(x | \omega_k)},$$

где ω_i – класс (тип), $i=1, \dots, M$; M – количество predetermined типов; вероятность $p(\omega_i)$ называется априорной вероятностью класса ω_i ; функция $p(x | \omega_i)$ описывает условную плотность распределения вектора признаков x в классе ω_i . Ключевым параметром классификации является условная плотность вероятности $p(x | \omega_i)$, которая может быть определена различными параметрическими и непараметрическими методами (к ближайших соседей, Розенблатта – Парзена и др.) при условии достаточного набора репрезентативных данных обучающих выборок [10].

Сегодня такие методы широко применяют при классификации исходных данных ДЗЗ. В этом случае мультиспектральное аэрокосмическое изображение (АИ) представляет собой совокупность значений яркостей, полученных в различных спектральных диапазонах, которая может быть обозначена как матрица $I[m, n, k]$, где m – индекс строки, n – индекс столбца, а k – индекс номера канала АИ. Исходная структура данных АИ хорошо подходит для применения методов классификации.

Для данных ДЗЗ широко применяют методы классификации, не требующие оценки плотности распределения – опорные векторы, искусственные нейронные сети и др. [19]. Один из наиболее широко используемых в ИАД методов контролируемой классификации, позволяющий обеспечивать поддержку принятия решений, основан на использовании деревьев принятия решений (англ. decision trees). В основе этого метода лежит использование ориентированного дерева как связного ациклического графа [20]. В общем случае, для решения задачи классификации необходимо спуститься по дереву от вершины до листа, выполняя соответствующие действия в узлах и выбирая при этом соответствующую дугу. Такая иерархическая структура позволяет реализовать способ представления правил классификации, основанный в каждом узле дерева на логической конструкции «что – если».

Построение дерева реализуется с помощью обучающей выборки примеров $T = \{x_i, f(x_i) = \omega_i\}$, где x_i – переменные, каждой из которых соответствует некоторый набор атрибутов (атрибут – условие перемещения по дуге) $Q_i = \{Q_j, j=1 \dots q\}$, а ω_i – классы, которым принадлежат переменные. Для классификации методом дерева решений разбивают множество T на некоторые подмножества. Для этого выбирается один из признаков x , имеющий два и более отличных друг от друга значений x_1, x_2, \dots, x_n . T разбивается на подмножества T_1, T_2, \dots, T_n , где каждое подмножество T_i содержит все примеры, имеющие значение $f(x = x_i)$ для выбранного признака. Эта процедура рекурсивно продолжается до тех пор, пока конечное множество не будет состоять из примеров, относящихся к одному и тому же классу. Фиксируя эти преобразования в виде элементов дерева решений, выполняется его построение сверху вниз. Параметром, определяющим сложность формируемого дерева, является крите-

рий выбора переменной при построении очередного узла дерева. Именно этот критерий отличает распространенные алгоритмы построения деревьев решений (ID3, C4.5, CART), реализующие выбор атрибута на основании значения прироста информации. Например, если множество T со свойством $S = f(x_i)$ классифицировано признаком x_i с атрибутом Q , имеющим q возможных значений, то прирост информации определяется как

$$\text{Gain}(T, S) = H(T, S) - \sum_{k=1}^q \frac{|T_k|}{|T|} H(T_k, S).$$

Этот метод непосредственно к данным ДЗЗ не используется в связи с тем, что крайне вероятно построение дерева решений чрезвычайной глубины и сложности, неприменимого на практике. Вместе с тем, особых препятствий для применения этого метода классификации на транзакционных данных нет.

Хранимая ретроспективная информация позволяет использовать методы регрессионного анализа, заключающиеся в поиске зависимости между признаками и прогнозировании динамики значений в них на будущие периоды времени. Если пространство объектов обозначить как X и множество возможных ответов Y , то существует неизвестная целевая зависимость $y^*: X \rightarrow Y$, значения которой известны только на объектах обучающей выборки $X^\ell = (x_i, y_i)_{i=1}^\ell$, $y_i = y^*(x_i)$. Требуется построить алгоритм, который называют функцией регрессии $f: X \rightarrow Y$, аппроксимирующий целевую зависимость y^* . Задачу обучения по прецедентам (x_i, y_i) , позволяющую найти регрессионную зависимость y^* , называют восстановлением регрессии [21]. Практических примеров применения этого метода для исходных данных ДЗЗ крайне мало. Существенно более распространенным является его применение при использовании транзакционных данных.

Приведенные семейства методов ИАД при их применении в задачах аэрокосмического мониторинга позволяют реализовать поддержку принятия решений по спектру ранее недоступных направлений. Примеры задач, методов ИАД и решений, поддержку принятия которых возможно обеспечить, предложены в табл. 1. Основной перечень методов ИАД применим исключительно в случае, если есть возможность оперирования не только исходными данными ДЗЗ, но и соответствующими транзакционными данными в исследуемой области. Пример обобщенной структуры таких данных предложен в табл. 2. Каждая запись такой таблицы представляет собой показания набора измерителей (датчиков) некоторой физической величины (температуры и влажности окружающей среды, атмосферного давления, содержания в атмосфере некоторого вещества и т. п.), справедливого для локации размещения датчика (координаты пространства $\{x_i; y_i\}$) в некоторой окрестности (окружности диаметром r_i). Такими датчиками также могут быть датчики состояния сооружений и агрегатов [8].

Примеры задач регионального аэрокосмического мониторинга
и методов анализа данных при поддержке принятия решений

Примеры направлений мониторинга	Примеры процессов мониторинга	Примеры принимаемых решений, поддержку которых следует обеспечить	Примеры методов ИАД	Тип необходимых данных
Экологический контроль	<ul style="list-style-type: none"> • определение уровня ПДК веществ и соединений; • прогноз изменения уровня предельно-допустимой концентрации (ПДК) веществ и соединений в различных областях наблюдаемой территории 	фактическое (прогнозное, высокая вероятность возможности) достижение в одной или нескольких контролируемых областях уровня ПДК, превышающего допустимые уровни	<ul style="list-style-type: none"> • регрессия; • обнаружение аномалий 	транзакционные
Региональный лесопожарный мониторинг	<ul style="list-style-type: none"> • вероятность пожарной опасности; • наличие очага возгорания; • фактические границы пожара и прогнозная динамика их изменений 	<ul style="list-style-type: none"> • вероятность пожарной опасности выше заданного порогового значения; • границы распространения пожара критически близки к заданным буферным зонам 	<ul style="list-style-type: none"> • регрессия; • ассоциация (в том числе последовательная); • обнаружение аномалий; • пространственное прогнозирование 	<ul style="list-style-type: none"> • ДЗЗ; • транзакционные
Мониторинг локальных природных и антропогенно-спровоцированных явлений и катастроф	<ul style="list-style-type: none"> • вероятность возникновения катастрофической ситуации; • фактические границы разрушений и прогнозная динамика их изменений; • фактические границы наводнения и прогнозная динамика их изменений 	<ul style="list-style-type: none"> • вероятность возникновения катастрофической ситуации в конкретной области выше заданного порогового значения; • границы распространения разрушений критически близки к заданным буферным зонам 	<ul style="list-style-type: none"> • регрессия; • ассоциация; • пространственное прогнозирование 	<ul style="list-style-type: none"> • ДЗЗ; • транзакционные

Примеры направлений мониторинга	Примеры процессов мониторинга	Примеры принимаемых решений, поддержку которых следует обеспечить	Примеры методов ИАД	Тип необходимых данных
Региональный мониторинг динамики ледников, процессов заболачивания и опустынивания, засоления, половодий, паводков	<ul style="list-style-type: none"> фактические границы ледников, болот, пустыни и т. п.; прогнозирование динамики изменения границ ледников, болот, пустыни и т. п. 	фактические или прогнозируемые границы (площадь) территории ледника, пустыни, болота и т. п. превысили заданное пороговое значение (близки к заданным буферным зонам)	<ul style="list-style-type: none"> регрессия; ассоциация; пространственное прогнозирование 	<ul style="list-style-type: none"> ДЗЗ; транзакционные
Инвентаризация, контроль строительства и мониторинг состояния транспортной, энергетической, информационной инфраструктуры и прилегающих к ним территорий, разведка месторождений нефти и газа	<ul style="list-style-type: none"> оценка соответствия дорожного покрытия существующим стандартам в зимний и летний периоды; контроль отсутствия высоких кустарников под линиями электропередачи; контроль состояния магистрального газо- и нефтепроводов и окружающей природной среды 	<ul style="list-style-type: none"> неудовлетворительное состояние дорожного покрытия; предельная высота кустарников под конкретными сегментами линий электропередачи; наличие фрагментов нефтепродуктов, демонстрирующих вероятное повреждение фрагмента газо- или нефтепровода 	классификация	ДЗЗ

Таблица 3

Пример структуры транзакционных данных для системы аэрокосмического мониторинга

«Датчик/время» № 1		«Датчик/время» № 2		...
Датчик 1 (x_1, y_1, r_1)	Время фиксации параметра	Датчик 2 (x_2, y_2, r_2)	Время фиксации параметра	...
Значение параметра	Дата, время	Значение параметра	Дата, время	...
...
Значение параметра	Дата, время	Значение параметра	Дата, время	...

Следует отметить, что записи таблицы транзакционных данных могут содержать пропуски, погрешности в измерениях и ошибки, существенно затрудняющие без дополнительной предварительной обработки данных практическую интерпретацию результатов. Именно решению таких задач посвящен отдельный раздел ИАД, получивший название очистка данных (англ. Data Cleaning, Data Cleansing, Data Scrubbing). Раздел имеет самостоятельное значение, а его исследованию и деталям практического применения посвящены отдельные емкие исследования, результаты которых следует учитывать при предварительной обработке данных в системах аэрокосмического мониторинга [15, 16].

Концептуальные требования к системе мониторинга с извлечением знаний

Анализ методов ИАД и условий их применения в системе аэрокосмического мониторинга позволяет сформулировать принципиальные требования к такой системе, наделенной функциями извлечения знаний:

- наличие подсистемы сбора и хранения не только данных ДЗЗ, но и транзакционных данных, имеющих географическую привязку (пример структуры данных – в табл. 2);
- наличие подсистемы интеллектуального анализа данных, использующей как данные ДЗЗ, так и данные архива транзакционных данных, а также подсистемы поддержки принятия решений по результатам применения методов ИАД.

Задача извлечения знаний в системах аэрокосмического мониторинга (как некоторой новой совокупности информации о свойствах объектов, закономерностях процессов и явлений ландшафтного покрова для принятия решений), предполагает оперирование такими данными ДЗЗ архива системы мониторинга, которые в совокупности фиксируют определенную динамику изменений, про-

исходящую на ландшафтном покрове. Собственно, лишь факт наличия зафиксированных изменений на разновременных архивных и актуальных данных ДЗЗ позволяет предположить с высокой достоверностью наличие в исследуемом ландшафте интересных для более глубокого анализа изменений, например, фактов возникновения обезлесения, нерегламентированной урбанизации, схода селевых потоков, движения ледовых и снежных покровов, границ водных объектов, возникновения лесного пожара. Очевидно, что для анализа и поддержки принятия решений в задачах мониторинга (см. табл. 1) порой достаточно лишь некоторого подмножества данных ДЗЗ, фиксирующих во времени такие значимые происходящие изменения. Аналогичные требования логично предъявить и к подмножеству архива транзакционных данных, количество записей и объем которого может быть также значительным.

Поэтому следует сформулировать еще одно принципиальное требование к системе мониторинга – наличие подсистемы эффективной фильтрации архивных и актуальных данных (транзакционных и ДЗЗ), оставляющей лишь подмножество данных, фиксирующих во времени значимые происходящие изменения. Это позволит существенно снизить объем данных для трудоемкого экспертного анализа при решении конкретных задач аэрокосмического мониторинга. Для решения этой задачи могут быть применены известные и оригинальные методы идентификации изменений ландшафта (англ. change detection) [4, 22]. Однако в существующем множестве методов до сих пор ведется поиск подходов, позволяющих применять их обоснованно и с высокой адекватностью в каждом конкретном случае [22–26]. При этом задел в исследованиях эффективного применения методов идентификации изменений в больших архивах разновременных данных ДЗЗ присутствует в некоторой степени лишь в работах под руководством автора [22, 27]. В методах поиска в значительных по объему архивах данных ДЗЗ сравнительно небольшого поднабора разновременных АИ предложено использовать оригинальную метрику – масштабность изменений ландшафта. Масштабность определяется как $M^d = \sum_{i=1}^N d_i \cdot S^w / S$, где S – общая площадь исследуемой области ландшафта; $S^w = (2r + 1)^2$ – площадь скользящего окна; r – порядок окна; $i = 1 \dots N$, где N – число областей с высокой концентрацией изменений $d_i = n_i / S^w$; n_i – число измененных пикселей в скользящем окне.

Эффективный отбор разновременных данных ДЗЗ с высокой масштабностью изменений региональной системы мониторинга делает в этом случае целесообразным применение беспилотных летательных аппаратов (БПЛА), позволяющих проводить экспертный анализ по данным высокого разрешения лишь для выбранных участков ландшафта с зафиксированными значимыми изменениями. При этом следует применять БПЛА типа конвертоплан, способный переходить из режима вертикального взлета и посадки в режим горизонтального полета путем поворота винтов. Такой БПЛА может без использования взлет-

ной полосы быть направлен в область, признанную по результатам поиска наиболее интересной с точки зрения оперативного анализа и подтверждения характера зафиксированных изменений. Запуск и пилотирование по маршруту технически могут быть реализованы в автоматизированном режиме с привлечением оператора лишь в момент подтверждения старта и достижения интересующей области с высокой масштабностью изменений ландшафта.

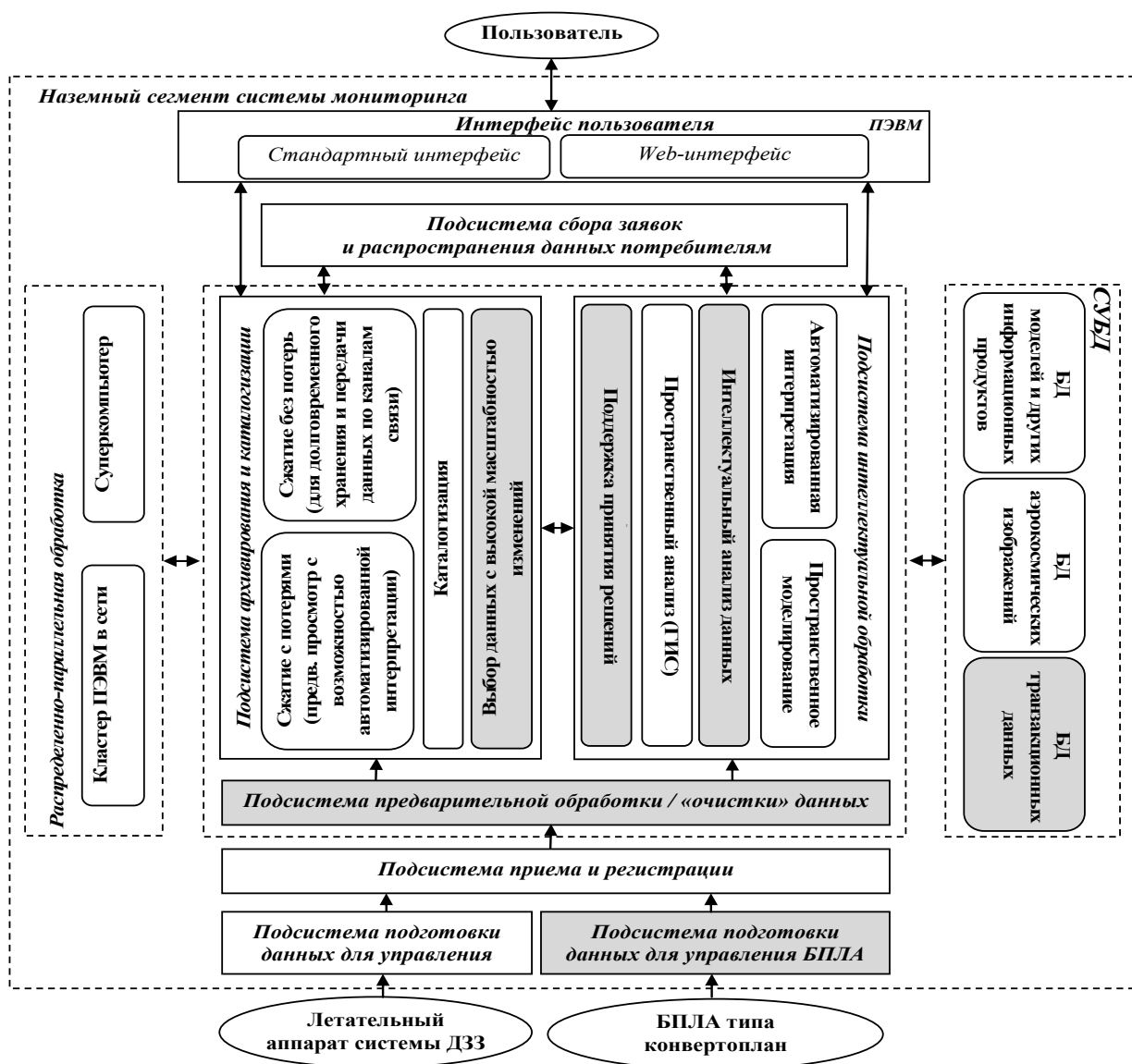
Такая комбинация применения данных периодического спутникового мониторинга сравнительно невысокого разрешения и точечного использования БПЛА, передающего данные высокого разрешения, позволяет не только добиться более широкой функциональности системы мониторинга, но и минимизировать экономические затраты получения знаний о происходящих изменениях на территории большой площади. В связи с вышеизложенным, следует сформулировать еще одно требование к системе мониторинга – наличие БПЛА типа конвертоплан и подсистемы управления им.

Обобщенная структура системы мониторинга с интеллектуальной обработкой данных и технология ее применения с БПЛА

На основе сформулированных концептуальных требований к системе аэрокосмического мониторинга предложена обобщенная структура системы, реализующая новые функции (на рисунке отмечены цветом) интеллектуального анализа и очистки данных, поддержки принятия решений, а также использования БПЛА.

Предложенная структура системы предполагает следующую технологию ее применения.

1. Регулярное пополнение архива снимками ДЗЗ и банком транзакционных данных.
2. Оригинальная технология поиска в большом массиве данных ДЗЗ областей с высокой масштабностью изменений.
3. Выбор области с масштабными зафиксированными изменениями и отправка к ней БПЛА по автоматически сформированному полетному заданию.
 - 3.1. Обработка данных модулем поддержки принятия решений и получение результатов советующей системы.
 - 3.2. Визуальный анализ данных БПЛА, сопоставление результатов советующей системы, снимков ДЗЗ и БПЛА, транзакционных данных.
4. Принятие окончательного (обоснованного) решения по найденной области.



Обобщенная структура региональной системы аэрокосмического мониторинга с интеллектуальным анализом данных

Заключение

Объемы данных в системах аэрокосмического мониторинга динамично возрастают, превышая возможности человека по их содержательному оперативному анализу. Поэтому возрастает потребность в наделении таких систем высокоавтоматизированными функциями поддержки принятия решений основных задач аэрокосмического мониторинга. Для этого целесообразно использовать имеющийся задел области интеллектуального анализа данных. Однако, до сих пор непосредственное применение этих методов к архивам данных ДЗЗ имеет ограниченный характер и требуется их адаптация, предполагающая как глубокое понимание особенностей методов анализа данных, так и учет специ-

фических аспектов построения систем аэрокосмического мониторинга. Именно решению в определенной степени этой задачи посвящена данная статья и получены следующие основные результаты:

- проведен анализ, который показал перспективность применения методов и технологий ИАД для задач аэрокосмического мониторинга;
- анализ математического аппарата основных методов ИАД показал, что данные ДЗЗ традиционной системы аэрокосмического мониторинга должны быть дополнены архивом транзакционных данных, а также предложен пример структуры таких данных;
- сопоставлены основные задачи и процессы мониторинга, а также примеры методов ИАД и примеры решений, поддержку которых можно обеспечить в системах мониторинга, используя архив транзакционных данных и данных ДЗЗ;
- сформулированы концептуальные требования к системе мониторинга с извлечением знаний, предложены принципы и технология применения БПЛА, а также предложена обобщенная структура системы аэрокосмического мониторинга с возможностями интеллектуального анализа данных и поддержкой принятия решений.

Работа выполнялась при поддержке РФФИ (грант № 14–07–00127а).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Консалтинговая компания IDC [Электронный ресурс]. – Режим доступа : <http://idc-group.ru> (03.05.2015 г.).
2. Построение информационной системы удаленной работы с каталогами данных НЦ ОМЗ / М. А. Бурцев, К. С. Емельянов, В. Ю. Ефремов, Е. А. Лупян, А. А. Мазуров, Л. А. Пахомов, А. А. Прошин, В. П. Саворский // Современные проблемы дистанционного зондирования Земли из космоса. – 2010. – Т. 7. – № 4. – С. 64–71.
3. Копылов В. Н. Основы создания центра космического мониторинга окружающей среды. – Екатеринбург : ПП «Контур», 2006. – 144 с.
4. Кудашев Е. Б. Электронная библиотека спутниковых данных: доступ к коллекциям экологического мониторинга // Космическая наука и технология. – 2003. – № 5/6. – С. 207–210.
5. Big Data Analytics Methodological Training in Statistical Data Science [Электронный ресурс]. – Режим доступа : <http://www.statoo.com/dm/> (03.05.2015 г.).
6. Knowledge Discovery Through Data Mining: What Is Knowledge Discovery? – Tandem Computers Inc., 1996.
7. Замятин А. В. Концепция региональной информационной системы аэрокосмического мониторинга с интеллектуальной распределено-параллельной обработкой данных // Информационные технологии. – 2011. – № 7. – С. 38–43.
8. Интеллектуальные информационные технологии наземно-космического мониторинга сложных объектов: состояние и перспективы развития / О. В. Майданович, М. Ю. Охтилев, В. А. Зеленцов, Б. В. Соколов, Р. М. Юсупов // Сб. трудов конференции «Управление в технических, эргодических, организационных и сетевых системах» (УТЭОСС – 2012). – 2012. – С. 38–52.

9. Системный анализ актуальных прикладных задач наземно-аэрокосмического мониторинга эколого-технологических объектов, исследуемых в проекте ELRI-184 / Д. В. Жуков, В. А. Матьяш, В. Ф. Мочалов, А. В. Труфанов // Труды СПИИРАН. – 2013. – 5(28). – С. 107121.
10. Лебедев В. В., Гансвинд И. Н. Проектирование систем космического мониторинга. – М. : Наука, 2010 г. – 388 с.
11. Richards J. A., Xiuping Jia. Remote Sensing Digital Image Analysis: An Introduction. Berlin: Springer, 1999. – 363 p.
12. Ian H. Witten, Eibe Frank, Mark A. Hall, Morgan Kaufmann. Data Mining: Practical Machine Learning Tools and Techniques (Third Edition). – Elsevier, 2011. – 629 p.
13. Chandola, V., Kumar, V. Summarization – compressing data into an informative representation // Knowledge and Information Systems. – 2007. – Vol. 12 (3). – P. 355–378.
14. Jain A., Zongker D. Feature Selection: Evaluation, Application, and Small Sample Performance // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1997. – Vol. 19. – No. 2. – P. 153–158.
15. Khandelwal P., Singh K. K., Singh B. K., Mehrotra A. Unsupervised Change Detection of Multispectral Images using Wavelet Fusion and Kohonen Clustering Network // International Journal of Engineering and Technology. – 2013. – Vol. 5. – No. 2. – P. 1401 – 1406.
16. Rahm E., Do H. H. Data Cleaning: Problems and Current Approaches // IEEE Bulletin on Data Engineering. – 2000. – 23(4).
17. Agrawal R., Srikant R. Mining Sequential Patterns // Proc. of the 11th Int'l Conference on Data Engineering, 1995.
18. Дуда Р., Харт П. Распознавание образов; Пер. с англ. – М. : Наука, 1981. – 450 с.
19. Bishop C. M. Neural Networks for Pattern Recognition. Oxford Univ. Press, 1995. – 508 p.
20. Breiman L., Friedman J. H., Olshen R. A., Stone C. T. Classification and Regression Trees. Wadsworth, Belmont, California, 1984. – p. 358.
21. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Множественная регрессия. – 3-е изд. – М. : Диалектика, 2007. – С. 912.
22. Afanasyev A. A., Zamyatin A. V., Cabral P. Land Cover Change Analysis using Change Detection Methods // Information Technologies and Mathematical Modelling, Communications in Computer and Information Science / Dudin et al. (Eds.). – Switzerland: Springer International Publishing, 2014. – Vol. 487. – P. 11–17.
23. Lu D., Mausel P. Change detection techniques // Remote sensing. – 2004. – Vol. 25. – No. 20. – P. 2365–2407.
24. Hussain, M., Chen D., Cheng A., Wei H., Stanley D. Change detection from remotely sensed images: from pixel-based to object-based approaches // ISPRS Journal of Photogrammetry and Remote Sensing. – 2013. – Vol. 80. – P. 91–106.
25. Lu D., Li G., Moran E. Current situation and needs of change detection techniques // International Journal of Image and Data Fusion. – 2014. – Vol. 5. – No. 1. – P. 13–38.
26. Collins J. B., Woodcock C. E. An Assessment of Several Linear Change Detection Techniques for Mapping Forest Mortality Using Multitemporal Landsat TM Data // Remote sensing of Environment. – 1996. – Vol. 56. – No. 1. – P. 66–77.
27. Afanasyev A., Zamyatin A. Hybrid Landscape Change Detection Methods in a Noisy Data Environment // Lecture Notes in Electrical Engineering. – Springer Verlag, 2015 (in press).
28. Замятин А. В., Афанасьев А. А., Кабрал П. Подход к анализу динамики ландшафтного покрова с использованием идентификации изменений и пространственного стохастического моделирования. Автометрия. – 2015. – № 4. – С. 40–52.

Получено 23.09.2016

© А. В. Замятин, 2016

DATA MINING METHODS IN THE REGIONAL SYSTEM OF SPACE MONITORING

Alexander V. Zamyatin

National Research Tomsk State University, 634050, Russia, Tomsk, 36 Lenin Avenue, D. Sc., Professor, Head of Department Theoretical Foundations of Information Science, tel./fax: (3822)52-94-96, e-mail: zamyatin@mail.tsu.ru

The possibility of constructing a system of space monitoring using data mining techniques such as association and classification methods based on Bayesian decision rule using methods on the basis of the adoption of decision trees are discussed. Examples of regional aerospace monitoring problems (environmental monitoring, regional forest fire monitoring, monitoring of local natural and anthropogenically provoked by events and disasters, and others.) and the corresponding data analysis techniques (regression, anomaly detection, spatial prediction, etc.) with the support of decision-making are discussed. The problems of adaptation of these methods to the archives of remote sensing data, involving as accounting features of data analysis methods, and specific aspects of building monitoring systems are presented. We discuss the example of a structure of transactional data for aerospace monitoring system comprising a set of reading meters (sensors) of a physical quantity, which is valid for sensor placement location in a spatial neighborhood. The approaches to the use of data analysis in systems monitoring using the optional archive transactional data such structure on the state of the environment, formulated conceptual requirements for such a system, its structure and the generalized use of technology to unmanned aerial vehicles.

Key words: data mining, system of aerospace monitoring, remote sensing data, unmanned aerial vehicle, classification, regression, anomaly detection, spatial forecast.

REFERENCES

1. *Konsaltingovaya kompaniya IDC [Consulting company IDC]*. Retrieved from at <http://idc-group.ru> [in Russian].
2. Burtsev, M. A., Emel'yanov, K. S., Efremov, V. Yu., Lupyan, E. A., Mazurov, A. A., Pakhomov, L. A., Proshin, A. A., & Savorskiy, V. P. (2010). Building an information system remote data directory NTS OMZ. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa [Modern Problems of Remote Sensing of the Earth from Space]*, 7(4), 64–71 [in Russian].
3. Kopylov, V. N. (2006). *Osnovy sozdaniya tsentra kosmicheskogo monitoringa okruzhayushchey sredy [Basis for the creation of the space monitoring center of the environment]*. Ekaterinburg: PP Kontur [in Russian].
4. Kudashev, E. B. (2003). Digital library of satellite data: access to collections of environmental monitoring. *Kosmicheskaya nauka i tekhnologiya [Aerospace Science and Technology]*, 5/6, 207–210 [in Russian].
5. *Big Data Analytics Methodological Training in Statistical Data Science*. Retrieved from at <http://www.statoo.com/dm/> [in Russian].
6. *Knowledge Discovery Through Data Mining: What Is Knowledge Discovery?* (1996). Tandem Computers Inc.
7. Zamyatin, A. V. (2011). The Concept of the regional information system of space monitoring with intelligent distributed–parallel processing. *Informatsionnye tekhnologii [Information Technologies]*, 7, 38–43 [in Russian].
8. Maydanovich, O. V., Okhtilev, M. Yu., Zelentsov, V. A., Sokolov, B. V., & Yusupov, R. M. (2012). Intellectual information technologies of ground–space monitoring of complex objects: status and prospects. In *Trudy konferentsii "Upravlenie v tekhnicheskikh, ergodicheskikh, organizatsionnykh i setevykh sistemakh" [Proceedings of the Conference "Control in Technical, Ergodic, Organizational and Network Systems"]* (pp. 38–52). UTEOSS – 2012 [in Russian].

9. Zhukov, D. V., Mat'yash, V. A., Mochalov, V. F., & Trufanov, A. V. (2013). A systematic analysis of the actual applied problems of ground–space monitoring of ecological and technological objects studied in the project ELRI–184. In *Trudy SPIIRAN [Proceedings of SPIIRAS]*, 5(28). 107–121 [in Russian].
10. Lebedev, V. V., & Gansvind I. N. (2010). *Proektirovanie sistem kosmicheskogo monitoring [Design of systems for space monitoring]*. Moscow: Nauka [in Russian].
11. Richards J. A., & Xiuping, Jia. (1999). *Remote Sensing Digital Image Analysis: An Introduction*. Berlin: Springer.
12. Witten, I. H., Frank E., Hall, M. A., & Kaufmann, M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3d Ed.). Elsevier.
13. Chandola, V., & Kumar, V. (2007). Summarization – compressing data into an informative representation. *Knowledge and Information Systems*, 12(3), 355–378.
14. Jain, A., & Zongker, D. (1997). Feature Selection: Evaluation, Application, and Small Sample Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2).
15. Khandelwal, P., Singh, K. K., Singh, B. K., & Mehrotra A. (2013). Unsupervised Change Detection of Multispectral Images using Wavelet Fusion and Kohonen Clustering Network. *International Journal of Engineering and Technology*, 5(2), 1401–1406.
16. Rahm, E., & Do, H. H. (2000). Data Cleaning: Problems and Current Approaches. *IEEE Bulletin on Data Engineering*, 23(4).
17. Agrawal, R., & Srikant, R. (1995). Mining Sequential Patterns. In *Proc. of the 11th Int'l Conference on Data Engineering*.
18. Duda, R., & Hart, P. (1981). *Raspoznavanie obrazov [Pattern recognition]*. Moscow: Nauka [in Russian].
19. Bishop C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford Univ. Press.
20. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. T. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, California.
21. Draper, N., & Smith, H. (2007). *Prikladnoy regressionnyy analiz. Mnozhestvennaya regressiya [Applied regression analysis. Multiple regression]* (3d ed.). Moscow: Dialektika [in Russian].
22. Afanasyev, A. A., Zamyatin, A. V., & Cabral, P. Land Cover Change Analysis using Change Detection Methods. (2014). In Dudin et al. (Eds.), *Information Technologies and Mathematical Modelling, Communications in Computer and Information Science: Vol. 487* (pp. 11–17). Switzerland: Springer International Publishing.
23. Lu, D., & Mausel, P. (2004). Change detection techniques. *Remote sensing*, 25(20), 2365 – 2407.
24. Hussain, M., Chen, D., Cheng, A., Wei, H., & Stanley, D. (2013). Change detection from remotely sensed images: from pixel–based to object–based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80, 91–106.
25. Lu, D., Li, G., & Moran, E. (2014). Current situation and needs of change detection techniques. *International Journal of Image and Data Fusion*, 5(1), 13–38.
26. Collins, J. B., & Woodcock, C. E. (1996). An Assessment of Several Linear Change Detection Techniques for Mapping Forest Mortality Using Multitemporal Landsat TM Data. *Remote Sensing of Environment*, 56(1), 66–77.
27. Afanasyev, A., & Zamyatin, A. (2015). Hybrid Landscape Change Detection Methods in a Noisy Data Environment. Lecture notes in Electrical Engineering. Springer Verlag.
28. Zamyatin, A. V., Afanas'ev, A. A., & Kabral, P. (2015). Approach to the analysis of the dynamics of landscape cover using the identification of changes and spatial stochastic simulation. *Avtometriya [Autometry]*, 4, 40–52 [in Russian].

Received 23.09.2016

© A. V. Zamyatin, 2016