

Показатели и алгоритмы оценивания качества результатов идентификации состояния прилегающих территорий

Зеленцов Вячеслав Алексеевич¹ — докт. техн. наук, профессор, главный научный сотрудник лаборатории информационных технологий в системном анализе и моделировании Санкт-Петербургского федерального исследовательского центра РАН. Область научных интересов: системный анализ, методы и информационные технологии поддержки принятия решений, космические информационные технологии, методы обработки аэрокосмической информации в интегрированных системах наземно-космического мониторинга. E-mail: v.a.zelentsov@gmail.com

Мочалов Виктор Федорович² — старший научный сотрудник Военного института (научно-исследовательского) Военно-космической академии имени А. Ф. Можайского. Область научных интересов: обработка и анализ материалов мультиспектральной и гиперспектральной аэрокосмической съемки, в том числе высокого пространственного разрешения для определения параметров растительности, устойчивости экосистем с учетом данных наземных измерений

¹ Санкт-Петербургский федеральный исследовательский центр РАН, Россия, 199178, Санкт-Петербург, 14-я линия В.О., 39

² Военно-космическая академия имени А. Ф. Можайского, Россия, 197198, Санкт-Петербург, ул. Ждановская, 13

Для цитирования: Зеленцов В. А., Мочалов В. Ф. Показатели и алгоритмы оценивания качества результатов идентификации состояния прилегающих территорий // Интеллектуальные технологии на транспорте. 2024. № 3 (39). С. 73–82. DOI: 10.20295/2413-2527-2024-339-73-82

Аннотация. Рассматриваются вопросы применения методов машинного обучения для оценивания состояния территорий в непосредственной близости к инфраструктурным объектам железной дороги. Представлены основные источники исходных данных, а также алгоритмы оценивания качества результатов автоматизированной идентификации элементов ландшафта на примере лесной растительности. Идентификация состояния лесной растительности осуществляется на базе спектрально-яркостных признаков, определяемых с помощью материалов мультиспектральной аэрокосмической съемки. Предложены показатели качества результатов автоматизированной обработки материалов съемки и алгоритмы расчета рассмотренных показателей. Приведены примеры оценивания качества результатов применения методов машинного обучения.

Ключевые слова: мультиспектральная аэрокосмическая съемка, спектрально-яркостные признаки, методы машинного обучения, качество идентификации, показатели и алгоритмы оценивания

Благодарности. Исследования выполнены в рамках бюджетной темы FFZF-2022-0004. При проведении разработок использованы ресурсы Центра коллективного пользования «Северо-Западный центр мониторинга и прогнозирования развития территорий». URL: <https://ckp-rf.ru/catalog/ckp/2079622/>

Введение

Идентификация состояния лесных насаждений вблизи объектов железнодорожного транспорта с точки зрения пожарной безопасности является важной практической задачей. Требуется своевременно реагировать на факты нарушений порядка содержания лесных массивов, особенно в пожароопасный период. Для этого в лесных массивах необходимо выявлять, в частности, следующие нарушения, способствующие самопроизвольному возникновению и распространению пожаров: несвоевременный вывоз мусора, порубочных остатков; возникновение заболеваний лесных насаждений, приводящих к усыханию растительности; несвоевременное устранение последствий ветровала. Лесные насаждения являются одним из наиболее распространенных элементов ландшафта на прилегающих территориях. Предложенные показатели качества могут характеризовать также качество идентификации состояния водных объектов, полей, горных массивов, других природных и антропогенных объектов. Представление достоверных данных о выявленных нарушениях можно обеспечить в ходе проведения наземных обследований, но эти работы, как правило, сопровождаются большими материальными и временными затратами. Одним из способов, позволяющих сократить влияние указанных выше недостатков, является проведение мультиспектральной аэрокосмической съемки [1] и ее углубленная автоматизированная обработка с помощью интеллектуально насыщенных алгоритмов.

Современные алгоритмы обработки материалов съемки часто основаны на методах машинного обучения [2–5]. При идентификации состояния прилегающих территорий обработка материалов съемки предполагает решение задачи так называемой семантической сегментации (semantic segmentation) элементов ландшафта в пределах рассматриваемой сцены. При решении такой задачи необходимо заранее иметь углубленное семантическое описание известных элементов ландшафта. Для них на основе материалов съемки выполняется построение обучающих и контрольных выборок. Для железнодорожных объ-

ектов должен быть составлен перечень типовых нарушений, оказывающих влияние на безопасность железной дороги. Географически привязанные участки растительного покрова должны быть зарегистрированы и снабжены семантическим описанием. На каждом участке выполняются измерения спектральных отражательных характеристик элементов ландшафта с учетом сезонной изменчивости. То есть должна быть создана специализированная система исходных данных.

В дальнейшем можно будет применять методы обработки материалов съемки, основанные на учете спектрально-яркостных признаков идентификации элементов ландшафта [1]. Спектрально-яркостные признаки опираются на результаты измерений отражательных характеристик растительности в нескольких спектральных каналах видимого и ближнего инфракрасного диапазонов спектра. По результатам проведения атмосферной коррекции осуществляется определение абсолютных физических величин — коэффициентов спектральной яркости. Знание абсолютных физических величин обеспечивает возможность обработки материалов съемки от мультиспектральной аппаратуры, размещенной на борту различных космических аппаратов и авиационных носителей, включая беспилотные летательные аппараты. При этом также могут рассчитываться и учитываться в ходе обработки различные специализированные индексы, характеризующие состояние растительности в условиях известных антропогенных воздействий [1]. Сложности формирования исходных данных и дальнейшее применение методов автоматизированной обработки подтверждают необходимость оценивания качества результатов обработки.

Одной из типовых задач обработки результатов дистанционных измерений является отнесение каждого пикселя изображения анализируемого участка территории к типу поверхности, обладающему заранее известными свойствами. Глубина семантического описания определяется содержанием практической задачи и может представлять собой, например, описание видового состава растительности, возраст и другие количественные

характеристики, имеющие значение для безопасности инфраструктурных объектов железной дороги. Кроме того, должны учитываться и отражаться в семантическом описании различные виды антропогенного воздействия на элементы ландшафта, в частности, последствия пожаров, ветровалов, загрязнения бытовыми и промышленными отходами, нефтепродуктами и др.

Оценивание качества обработки материалов съемки может выполняться на основе различных показателей, и от их выбора зачастую зависят выводы о достоверности полученных результатов. Для демонстрации роли показателей оценивания качества результатов обработки материалов съемки рассмотрим пример семантической сегментации трех типов поверхностей после возникновения нарушения в состоянии лесного массива вблизи одного из условных объектов железной дороги [6]. В качестве исходных данных приняты результаты измерения спектральных отражательных характеристик известных элементов ландшафта. Измерения выполнены с помощью мультиспектральной аппаратуры, размещенной на борту космического аппарата Sentinel-2.

Проведен анализ изменения состояния растительности и спектральных отражательных характе-

ристик элементов ландшафта до неблагоприятного воздействия на основе материалов съемки от 17 августа 2019 года и после воздействия на основе материалов съемки от 2 июня 2020 года. Семантическое описание территории выполнено в ходе проведения наземных обследований. Изображения фрагмента сцены в псевдоцветах приведены на рис. 1.

Семантическое описание содержало данные о трех основных элементах ландшафта с условными метками класса 1, 2 и 3. С помощью семи методов машинного обучения осуществлена семантическая сегментация указанных элементов ландшафта. Верификация результатов идентификации выполнялась в ходе наземных обследований в выбранных заранее контрольных точках. Отдельные точки контроля отмечены на рис. 1 в виде значков пикселей.

При идентификации нарушения в состоянии лесной растительности получены результаты, приведенные в табл. 1. Размер контрольной выборки составлял 102 пикселя. Было идентифицировано три типа поверхностей, обозначенных соответствующими метками класса. Номера меток отражают состояние территорий с высокой, низкой степенью нарушения и ненарушенные территории соответственно.

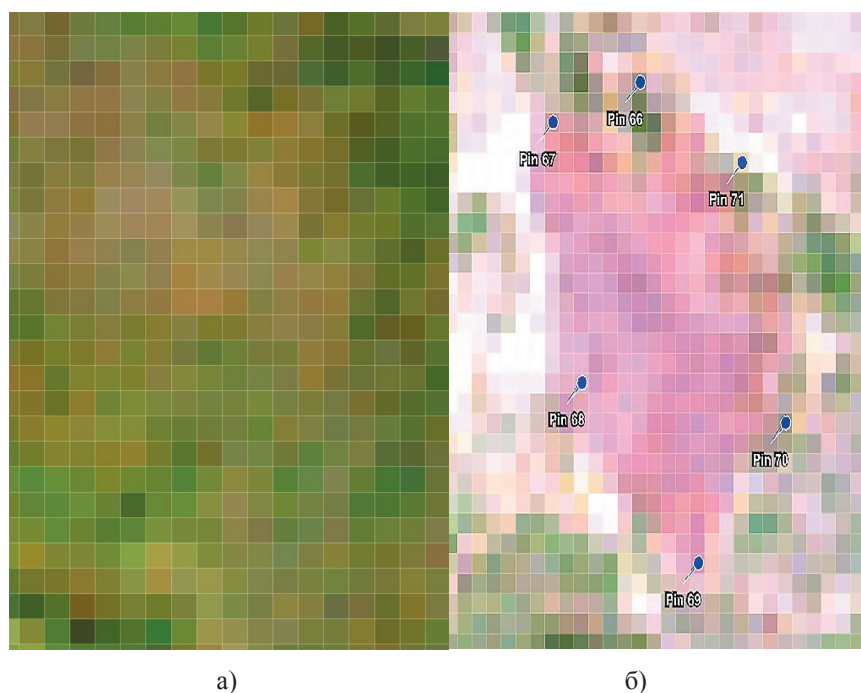


Рис. 1. Фрагменты сцены: а) до возникновения нарушения; б) после возникновения нарушения

Таблица 1

Результаты идентификации элементов ландшафта и данные о качестве идентификации

№ пикселя	Элемент ландшафта	Алгоритмы машинного обучения						
		<i>PPN</i>	<i>SVM</i>	<i>Tree</i>	<i>LR</i>	<i>Forest</i>	<i>KNN</i>	<i>Ensemble</i>
1	2	2	2	2	2	2	2	2
2	1	1	1	1	1	1	1	1
3	2	2	2	2	2	2	2	2
...								
100	1	1	1	1	1	1	1	1
101	1	1	1	1	1	1	1	1
102	3	3	3	3	3	3	3	3
Accuracy		1,0	1,0	0,97	1,0	0,98	0,98	1,0

Методы обработки материалов съемки или методы машинного обучения в табл. 1 имеют следующие обозначения:

- *PPN* — метод коррекции ошибки;
- *SVM* (support vector machine) — метод опорных векторов;
- *Tree* — дерево принятия решений;
- *LR* — метод логистической регрессии;
- *Forest* — метод случайного леса (random forest);
- *KNN* — метрический алгоритм k -ближайших соседей (k -nearest neighbors).

Комплексное сочетание алгоритмов (*Ensemble*) реализовано путем анализа результатов всех приведенных ранее алгоритмов машинного обучения. Для оценивания качества результатов обработки использовался традиционный обобщенный показатель — *Accuracy*. Значение показателя определяется путем анализа соотношений количества пикселей для верно идентифицированных элементов ландшафта к общему числу пикселей в пределах контрольной выборки для анализируемого элемента ландшафта.

Показатель *Accuracy* в таблице принимает значение равное единице или близкое к ней. Причиной такой ситуации может быть недостаточный размер контрольной выборки или низкая информативность привлекаемого показателя оценивания качества обработки.

Приведенный пример позволяет сделать вывод о том, что применяемый показатель (*Accuracy*) для

оценивания качества семантической сегментации элементов ландшафта носит неинформативный характер. Требуется обеспечение обоснованного выбора показателей качества тематической обработки, учитывающих семантическое описание идентифицируемых поверхностей, и разработка соответствующих алгоритмов их расчета.

Целью описываемых исследований является разработка предложений по возможному составу показателей и алгоритмов для оценивания качества обработки материалов съемки, основанных на анализе особенностей конкретных задач тематической обработки.

Основная часть

Источники исходных данных. В качестве источников исходных данных выступают материалы мультиспектральной космической съемки. Основные технические характеристики бортового съемочного оборудования приведены в табл. 2. Спектральная чувствительность оборудования и условное пространственное разрешение в наиболее информативных каналах сопоставимы.

Показатели и способы оценивания качества результатов обработки материалов съемки. При выборе показателей качества идентификации состояния элементов ландшафта предлагается предварительно провести анализ смыслового

Основные характеристики бортового съемочного оборудования

Принадлежность	Европейское космическое агентство, ESA	США, NASA	Россия	Китай
Носитель	Sentinel-2	Landsat 9	Канопус-В	БПЛА
Наименование или обозначение оборудования	MSI	OLI	MCC	Parrot Sequoia
Количество спектральных каналов, шт.	12	8	4	4
Условное пространственное разрешение, м	10–30	20	12	–
Границы спектрального диапазона, нм	432,2–2289,9	411,3–2389,6	460–840	530–810
Ширина полосы захвата, км	185	290	19	–
Периодичность съемки, сутки	5	5	3	–

содержания решаемой практической задачи управления прилегающей территории.

Подобным вопросам посвящена, в частности, работа [7], в которой приведен перечень часто применяемых показателей и порядок их расчета. Традиционные технологии обработки предусматривают формирование обучающей и контрольной выборок. Если обучающая выборка применяется для настройки алгоритмов обработки, то контрольная — представляет собой совокупность достоверных данных из независимых источников и служит для оценивания результатов обработки.

Недостатками большинства работ являются: попытка специалистов, занимающихся тематической обработкой материалов съемки, найти универсальные, обобщенные показатели для решаемых задач и всех идентифицируемых элементов ландшафта; низкая концентрация внимания на основных видах элементов ландшафта, требующих повышенной точности их идентификации; не всегда учитывается сбалансированность исходных данных, характеристики относительных размеров идентифицируемых поверхностей.

Приведем примеры различных практических задач, от содержания которых зависит выбор того или иного показателя качества обработки.

Обязательное выделение на снимках элементов ландшафта, требующих повышенного приоритетного внимания. Например, при выявлении пожароопасных участков леса такими элементами яв-

ляются места высыхания растительности, участки с ветровалом, участки, загрязненные мазутом или другими нефтепродуктами.

В других часто решаемых задачах в ходе семантической сегментации различных видов поверхностей в пределах анализируемого контура требуется провести идентификацию двух или более видов лесной растительности, имеющих сопоставимое значение.

Очевидно, что качество обработки материалов съемки в этих двух задачах должно оцениваться по-разному. Однако и в том, и в другом случае основу оценивания составляют значения первичных показателей, отражающих количественные значения правильно и ошибочно идентифицированных пикселей изображения. На основе значений первичных показателей в дальнейшем вычисляются комплексные показатели.

Предлагаемый общий подход к выбору показателей для оценивания качества обработки состоит в следующем. Идентифицируемые элементы ландшафта разделяются на два класса элементов, обладающих условно положительными и отрицательными свойствами с точки зрения решаемой задачи. Например, в задаче выявления пожароопасных участков леса к элементам ландшафта с отрицательными свойствами относятся места высыхания растительности и участки с ветровалом. Тогда результаты идентификации можно представить в виде значений следующих первичных показателей качества обработки [3]:

TP (True Positive) — количество пикселей, представляющих территории с истинно неповрежденными участками растительности;

TN (True Negative) — количество пикселей, представляющих территории с истинно поврежденными участками растительности;

FP (False Positive) — количество пикселей, представляющих территории с ошибочно неповрежденными участками растительности;

FN (False Negative) — количество пикселей, представляющих территории с ошибочно поврежденными участками растительности.

Первичные показатели определяются на основе анализа данных контрольной выборки, которая, как и обучающая выборка, формируется на основе исходных данных – выбранных фрагментов материалов съемок с известным семантическим описанием. Но в отличие от обучающей выборки, данные контрольной выборки не используются в обучении модели автоматизированной обработки материалов съемки. Значения данных контрольной и обучающей выборок считаются достоверными.

Иллюстрация порядка определения значений первичных показателей качества идентификации двух поверхностей приведена на рис. 2.

На рис. 2 по оси абсцисс отложены условные пороговые численные значения, отражающие положение гиперпараметра, или разделяющей поверхности (Threshold), на идентифицируемые элементы ландшафта в рассматриваемом примере.

По оси ординат в виде гистограммы представлены значения количества пикселей, отражающих соответствующие элементы ландшафта. Условно положительными свойствами (P) обладает ряд 2. Условно отрицательными свойствами обладает ряд 1. Красной линией обозначено текущее расположение разделяющей поверхности. Слева от красной линии голубая поверхность, ограниченная голубым контуром, соответствует значению *TP*, а ограниченная зеленым контуром — *FP*. Справа от разделяющей поверхности зеленая фигура, ограниченная зеленым контуром, соответствует значению *TN*, а ограниченная голубым контуром — *FN*.

В зависимости от выбора порогового значения виртуальной плоскости определяется пространственное расположение красной линии и рассчитываются значения промежуточных показателей *TP*, *TN*, *FP*, *FN* путем суммирования числа пикселей в пределах выделенных контуров.

На основе приведенных первичных показателей определяются значения комплексных показателей, с помощью которых оценивается качество обработки [3]:

Полнота, чувствительность (Recall):

$$REC = \frac{TP}{TP + FN} \quad (1)$$

отражает долю элементов ландшафта, которые характеризуются условно положительными свойствами. С помощью показателя оценивается способность привлекаемого алгоритма

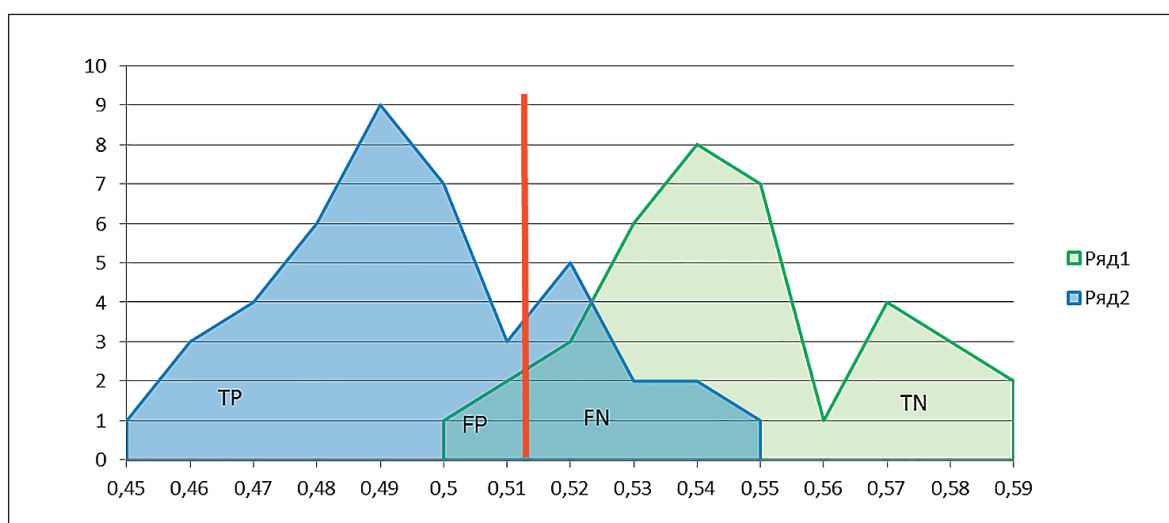


Рис. 2. Представление значений первичных показателей качества обработки

идентифицировать искомый, важный для потребителя вид ландшафта. Значение REC, рассчитанное по формуле (1), приближается к единице, когда количество пикселей с ошибочно отрицательными свойствами стремится к нулю. Таким образом показатель может свидетельствовать о вероятности успешной идентификации элементов ландшафта с положительными свойствами. На основе анализа значения REC можно принять решение о необходимости проведения дополнительных обследований или привлечения дополнительных алгоритмов обработки.

Точность (Precision):

$$PRE = \frac{TP}{TP + FP} \tag{2}$$

Точность — это отношение между истинно положительными и всеми положительными результатами идентификации. Значения REC и PRE характеризуют различные, конкурирующие свойства алгоритмов автоматизированной обработки. Поэтому целесообразно проводить их совместный анализ.

На рис. 3 приведен график изменения значений показателя PRE, рассчитанного по формуле (2), в зависимости от значений показателя REC, рассчитанного по формуле (1). Изменения отражают движение разделяющей поверхности, или красной линии, приведенной на рис. 2 от значения 0,5 до 0,55.

По оси абсцисс представлены значения показателя REC, по оси ординат — PRE, которые соответствуют числовым значениям, приведенным в табл. 3.

Таблица 3

Значения комплексных показателей качества обработки в зависимости от порогового значения

№	Порог	REC	PRE	OA
1	0,50	0,65	0,97	0,80
2	0,51	0,78	0,92	0,85
3	0,52	0,89	0,87	0,87
4	0,53	0,93	0,78	0,83
5	0,54	0,99	0,69	0,76
6	0,55	1,00	0,63	0,69

В табл. 3 приведен пример расчета значений комплексных показателей с учетом значений первичных показателей, приведенных на рис 2. При этом в столбце «Порог» отражаются значения пересечения красной линии с осью абсцисс.

В четвертом столбце табл. 3 представлены значения обобщенного показателя OA (Overall Accuracy), характеризующего долю правильно идентифицированных пикселей:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Показатель OA описывает общую точность привлекаемого метода обработки. Этот показатель особенно полезно использовать, когда каждый класс идентифицируемых поверхностей одинаково важен. Он рассчитывается как отношение количества пикселей с правильно определенными свойствами элементов ландшафта к их общему количеству. На рис. 4 приведен график изменения показателя OA и числовых значений порога разделяющей поверхности в зависимости от номера точек отсчета.

Анализ данных, приведенных на рис. 4, позволяет сделать вывод о возможности оптимизации выбора порогового значения разделяющей поверхности для достижения максимального значения показателя OA, рассчитываемого по формуле (3).

На базе рассмотренных показателей может быть сформирован комплексный показатель как объединение показателей REC и PRE в виде гармонического среднего:

$$F1 = 2 \cdot \frac{PRE \cdot REC}{PRE + REC} \tag{4}$$

Мера F1, рассчитанная в формуле (4), учитывает показатели PRE и REC как равноценные, а если более важным является один из этих показателей, применяется мера [7]:

$$F_{\beta} = \frac{(1 + \beta^2) \cdot PRE \cdot REC}{\beta^2 \cdot PRE + REC} \tag{5}$$

Порядок выбора значения β в формуле (5) с учетом важности одного из показателей PRE или REC определяется неравенством, которое отражает тот факт, что для оценивания качества обработки в задаче, для которой приоритетным является

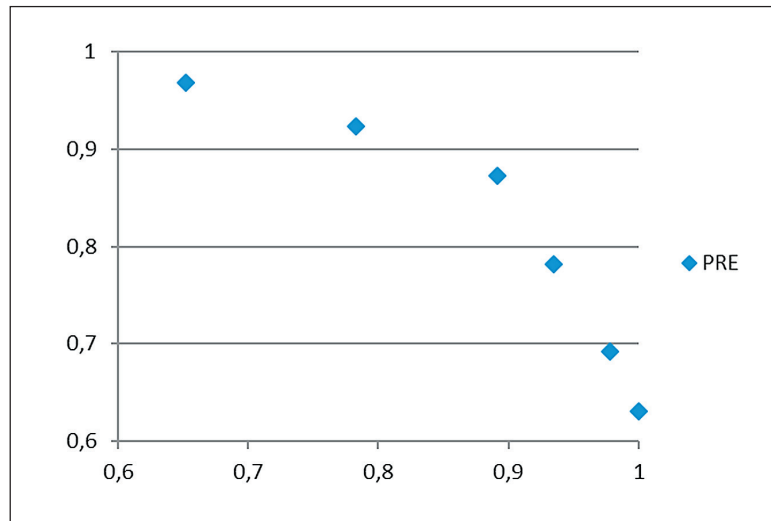


Рис. 3. Изменения значений показателя PRE в зависимости от REC

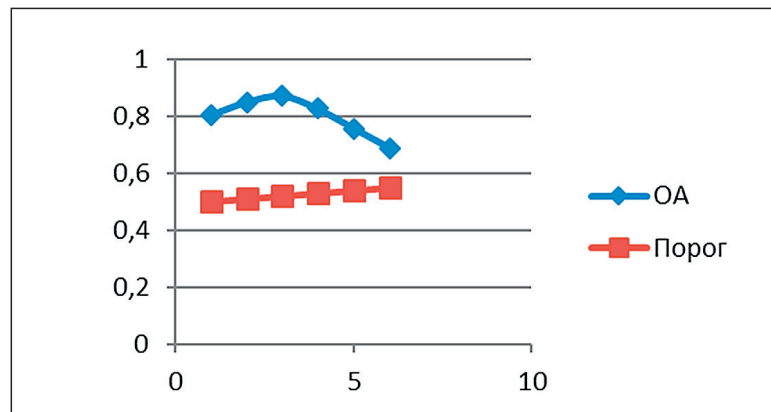


Рис. 4. Изменения значений показателя OA и порога разделяющей поверхности

достижение повышенного показателя PRE с заданным уровнем REC, выбирается значение $\beta < 1$.

$$\begin{cases} \beta < 1 \rightarrow PRE \\ \beta > 1 \rightarrow REC. \end{cases}$$

В табл. 4 приведены примеры выбора показателей для оценивания качества идентификации элементов ландшафта в зависимости от вида решаемой практической задачи.

Заключение

Представленные показатели для оценивания качества результатов обработки материалов съемки и порядок их расчета базируются на учете особенностей каждой конкретной задачи обработки данных. Показатели могут применяться как в случае применения традиционных алгоритмов, предусма-

тривающих расчет индексов, отражающих состояние элементов ландшафта, так и в случае применения алгоритмов машинного обучения, включая алгоритмы искусственных нейронных сетей.

Важным моментом при идентификации состояния прилегающих территорий является обоснованное формирование обучающих и контрольных выборок, обеспечение независимости этих выборок в алгоритмах машинного обучения и при оценивании качества выходных результатов.

Необходимо отметить, что, опираясь на предварительный анализ значений приведенных показателей качества, можно обосновывать требования к алгоритмам обработки в части оптимизации выбора значения гиперпараметра или порога разделяющей поверхности.

Примеры практических задач и показатели для оценивания их качества

Задача	Наименование задачи	Характеристика задачи, требования к качеству ее решения	Формула расчета качества идентификации
A	Идентификация двух видов леса: соснового, свойство (P) и лиственного, свойство (N)	Однородные исходные данные, требования к точности идентификации сопоставимы	$OA = \frac{TP + TN}{TP + TN + FP + FN}$
B	Идентификация участков леса, подверженных опасному заболеванию, свойство (N), на фоне здорового леса, свойство (P)	Преимущественное обязательное выявление элементов ландшафта, обладающих свойством (N)	$REC = \frac{TP}{TP + FN}$
C	Идентификация участков леса, характеризующихся повышенным возможным воздействием на безопасность эксплуатации объекта, свойство (P)	Выявление с максимальной достоверностью только тех элементов ландшафта, которые отличаются заданными свойствами (P)	$PRE = \frac{TP}{TP + FP}$
D	Идентификация двух видов леса: соснового, свойство (P) и лиственного, свойство (N)	Однородные исходные данные, требования по приоритетности идентификации поверхностей со свойствами (P) и (N) могут уточняться в зависимости от складывающейся ситуации	$F_{\beta} = \frac{(1 + \beta^2) \cdot PRE \cdot REC}{\beta^2 \cdot PRE + REC}$

Следует также подчеркнуть, что управление состоянием территорий вблизи железнодорожных объектов — это непрерывный плановый процесс, в котором в качестве регулярных операций должны быть предусмотрены систематическое выполнение съемочных работ, обработка материалов съемки, а также синхронные

наземные обследования на выбранных тестовых участках. Обязательным является выполнение этих операций при возникновении неблагоприятных воздействий на прилегающую территорию, а также по итогам реализации тех или иных мероприятий по управлению развитием территорий.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- Шовенгердт Р. А. Дистанционное зондирование. Методы и модели обработки изображений / пер. с англ. А. В. Кирюшина, А. И. Демьяникова, 3-е изд. М.: Техносфера, 2010. 560 с.
- Корабашев О. З. Анализ и перспективы применения методов машинного обучения для чрезвычайных ситуаций // Интеллектуальные технологии на транспорте. 2024. № 1 (37). С. 12–17. DOI: 10.20295/2413-2527-2024-137-12-17
- Рашка С. Python и машинное обучение / пер. с англ. А. В. Логунова. М.: ДМК Пресс, 2017. 418 с.
- Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Second Edition. Sebastopol (CA): O'Reilly Media, 2019. 848 p.
- Теребиж В. Ю. Введение в статистическую теорию обратных задач. М.: Физматлит, 2005. 376 с.
- Мочалов В. Ф., Хабаров Р. С. Обработка материалов мультиспектральной съемки на основе методов машинного обучения при управлении состоянием лесного массива // Сборник материалов IV Международной научной конференции по проблемам управления в технических системах (ПУТС-2021) (Санкт-Петербург, 21–23 сентября 2021). СПб.: ЛЭТИ, 2021. С. 269–272.
- Maxwell A. E., Warner T. A., Guillén L. A. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studie. Part 1: Literature Review // Remote Sensing. 2021. Vol. 13, iss. 13. Art. 2450. 27 p. DOI: 10.3390/rs13132450

Дата поступления: 28.06.2024

Решение о публикации: 16.09.2024

Indicators and Algorithms for Assessing the Identification Results Quality of the Adjacent Territories State

Vyacheslav A. Zelentsov¹ — Dr. Sci. in Engineering, Professor, Chief Researcher at the Laboratory of Information Technologies in System Analysis and Modeling of the St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research interests: system analysis, methods and information technologies of decision support, space information technologies, methods of processing aerospace information in integrated ground-space monitoring systems. E-mail: v.a.zelentsov@gmail.com

Viktor F. Mochalov² — Senior Researcher at the Military Research Institute. Research interests: processing and analysis of multispectral and hyperspectral aerospace survey materials, including high spatial resolution for determining vegetation parameters, ecosystem stability, taking into account ground-based measurement data.

¹ St. Petersburg Federal Research Center of the Russian Academy of Sciences, 14th line V.O., 39, St. Petersburg, 199178, Russia

² Mozhaisky Military Aerospace Academy, 13, Zhdanovskaya str., St. Petersburg, 197198, Russia

For citation: Zelentsov V. A., Mochalov V. F. Indicators and Algorithms for Assessing the Identification Results Quality of the Adjacent Territories State // Intellectual Technologies on Transport. 2024. № 3 (39). Pp. 73–82. DOI: 10.20295/2413-2527-2024-339-73-82 (In Russian)

Abstract. Machine learning methods for identification of territories near railway infrastructure facilities are considered. The main sources of initial data are presented, as well as algorithms for assessing the quality of the results for landscape elements automated identification using the forest vegetation example. Forest vegetation state identification is carried out based on spectral-brightness characteristics determined using multispectral aerospace imaging materials. Quality indicators for the results of survey materials automated processing and algorithms for calculating the considered indicators are proposed. Examples are given to assess the results quality of using machine learning methods.

Keywords: multispectral remote sensing, spectral characteristics, machine learning methods, identification quality

REFERENCES

1. Shovengerdt R. A. Distancionnoe zondirovanie. Metody i modeli obrabotki izobrazhenij / per. s angl. A. V. Kiryushina, A. I. Dem'yanikova, 3-e izd. M.: Tekhnosfera, 2010. 560 s. (In Russian)
2. Koraboshev O. Z. Analiz i perspektivy primeneniya metodov mashinnogo obucheniya dlya chrezvychajnyh situacij // Intellektual'nye tekhnologii na transporte. 2024. № 1 (37). S. 12–17. DOI: 10.20295/2413-2527-2024-137-12-17 (In Russian)
3. Rashka S. Python i mashinnoe obuchenie / per. s angl. A. V. Logunova. M.: DMK Press, 2017. 418 s. (In Russian)
4. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Second Edition. Sebastopol (CA): O'Reilly Media, 2019. 848 p.
5. Terebizh V. Yu. Vvedenie v statisticheskuyu teoriyu obratnyh zadach. M.: Fizmatlit, 2005. 376 s. (In Russian)
6. Mochalov V. F., Habarov R. S. Obrabotka materialov mul'tispektral'noj s'emki na osnove metodov mashinnogo obucheniya pri upravlenii sostoyaniem lesnogo massiva // Sbornik materialov IV Mezhdunarodnoj nauchnoj konferencii po problemam upravleniya v tekhnicheskikh sistemah (PUTS-2021) (Sankt-Peterburg, 21–23 sentyabrya 2021). SPb.: LETI, 2021. S. 269–272. (In Russian)
7. Maxwell A. E., Warner T. A., Guillén L. A. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies. Part 1: Literature Review // Remote Sensing. 2021. Vol. 13, Is. 13. Art. № 2450. 27 p. DOI: 10.3390/rs13132450

Received: 28.06.2024

Accepted: 16.09.2024