

С.В. ЕРЕМЕЕВ, Ю.А. КОВАЛЕВ

**Алгоритм классификации
пространственных объектов на
основе модели Random forest**

Муромский институт
(филиал) ФГБОУ ВО
«Владимирский
государственный
университет имени
А.Г. и Н.Г. Столетовых»,
г. Муром

В статье произведен анализ существующих алгоритмов классификации объектов. Разработан алгоритм, который будет классифицировать пространственные объекты на растровых изображениях по топологическим признакам.

Были использованы методы персистентной гомологии, основу которых составляют числа Бетти. Для разработки алгоритма была взята модель классификации Random forest.

В качестве результата реализации алгоритма приводится наименование классификации объекта.

Исследование выполнено при финансовой поддержке РФФИ и администрации Владимирской области в рамках научного проекта № 17-47-330387.

Введение

Любая карта является проекцией объектов реального мира (дорог, рек, озер, линий электропередач или бензоколонок) на некую плоскость.

Одна из основных целей карт - дать возможность сопоставить расположение в пространстве некоторого объекта относительно других объектов.

Главная возможность карт - это возможность с максимальной точностью изображать интересующие нас предметы и их координаты.

Сегодня большинство геоинформационных систем могут производить поиск пространственных объектов на картах, изображениях. Одним из самых точных видов определения объектов является поиск по соответствующим признакам. В статье будут рассмотрены топологические признаки на основе распределения чисел Бетти. Одной из наиболее точных моделей классификации объектов является Random forest.

Постановка задачи

В работе требуется разработать алгоритм, который будет классифицировать пространственные объекты на растровых изображениях по топологическим признакам.

За основу алгоритма будет взято распределение чисел Бетти и модель классификации Random forest [1,2].

Random forest – это алгоритм состоящий из множества решающих деревьев, который был предложен Лео Брейманом и Адель Катлер.

Объекты классифицируются голосованием: любое дерево подмножества относит классифицируемый объект для одного из классов, и приоритетным является тот класс, за который проголосовало наибольшее число деревьев.

Разработка алгоритма

Для классификации признаков пространственных объектов используется алгоритм распределения чисел Бетти.

Числа Бетти — последовательность инвариантов топологического пространства. Каждому пространству X соответствует некая последовательность чисел Бетти B_1 [3].

Нулевое число Бетти совпадает с числом связных компонент;

Первое число Бетти интуитивно представляет собой максимальное число разрезов этого пространства, которые можно сделать без увеличения числа компонент связности.

Алгоритм поиска чисел Бетти реализуется следующим образом:

Шаг 1. На вход подается изображение. Каждый пиксель, который не принадлежит границе, имеет окрестность Мура, т. е. 8 соседей.

Шаг 2. Упорядочим по возрастанию интенсивности все пиксели и будем подниматься вертикально вверх «от подножья гор к вершинам», руководствуясь списком высот.

Шаг 3. При появлении новой вершины в окрестности уже существующей (отмеченной) вершины соединяем их ребром, которое добавляем в фильтрацию.

Значения функции на ребре определим как $f[ab] = \max\{f[a], f[b]\}$, где a, b - соседние вершины.

Шаг 4. Как только в фильтрации появляются три соседние точки, соединяем их треугольником.

Значение функции на треугольнике определим как $f[abc] = \max\{f[a], f[b], f[c]\}$, где a, b, c - соседние вершины.

На рисунке 1 с вершиной 55 появляется первый «треугольник» (11-38-55). На вершине 76 появляется второй «треугольник». Мы запоминаем координаты их появления и даем им индивидуальные номера.

22	123	178	201	154	134
99	38	55	190	111	144
155	11	217	234	76	5
244	166	243	203	22	211
176	156	133	124	200	234
4	35	177	6	205	143

Рис. 1. Компоненты связности и треугольники

Шаг 5. В результате получится фильтрация, состоящая из вершин, ребер и треугольников, упорядоченных по возрастанию функции и по размерности симплекса в случае равных значений функции

Шаг 6. При одинаковых значениях в фильтрации приоритет получает та вершина, которая находится на изображении выше и левее.

Шаг 7. На основе полученной фильтрации можно вычислять числа Бетти.

Шаг 8. Каждая новая вершина добавляет новую компоненту связности, а ребро, в том случае, если оно соединяет разные компоненты, наоборот, убивает компоненту. Появление треугольника означает появление дыры $B_1 = B_1 + 1$.

Шаг 9. Для подсчета количества дыр и их времени жизни нужна та же фильтрация, только проходить ее нужно в обратном порядке, сверху вниз, и добавить один внешний элемент «outside». В процессе обратного прохода при появлении каждого треугольника добавляется дыра.

Шаг 10. При появлении ребра, которое является смежным для двух треугольников, ограничивающих разные дыры, компонента умирает. В случае появления ребра на границе дыра также погибает $B_1 = B_1 - 1$.

Шаг 11. Продолжая движение вниз, мы получим полную фильтрацию для дыр [3].

После находятся максимальные количества дыр и ребер, и на основе данных признаков применяется модель Random forest, которая строится следующим образом:

Шаг 1. Происходит подвыборка обучающей выборки, затем по ней строится дерево.

Шаг 2. Для реализации каждого расщепления в дереве просматривается набор случайных признаков.

Шаг 3. Выбирается наиболее лучший признак и расщепление. В данном алгоритме деревья строятся до того момента, пока выборка не останется пустой [4,5].

Для реализации алгоритма будут использованы признаки максимального количества дыр и ребер эталонного изображения.

В выборку эталонных изображений попадают автомобили и одноэтажные здания. Для обучения алгоритма было взято 10 эталонных изображений автомобилей и 10 изображений одноэтажных зданий. Примеры изображений автомобилей из

данной выборки отображены на рисунке 2(а), на рисунке 2(б) отображены изображения зданий.

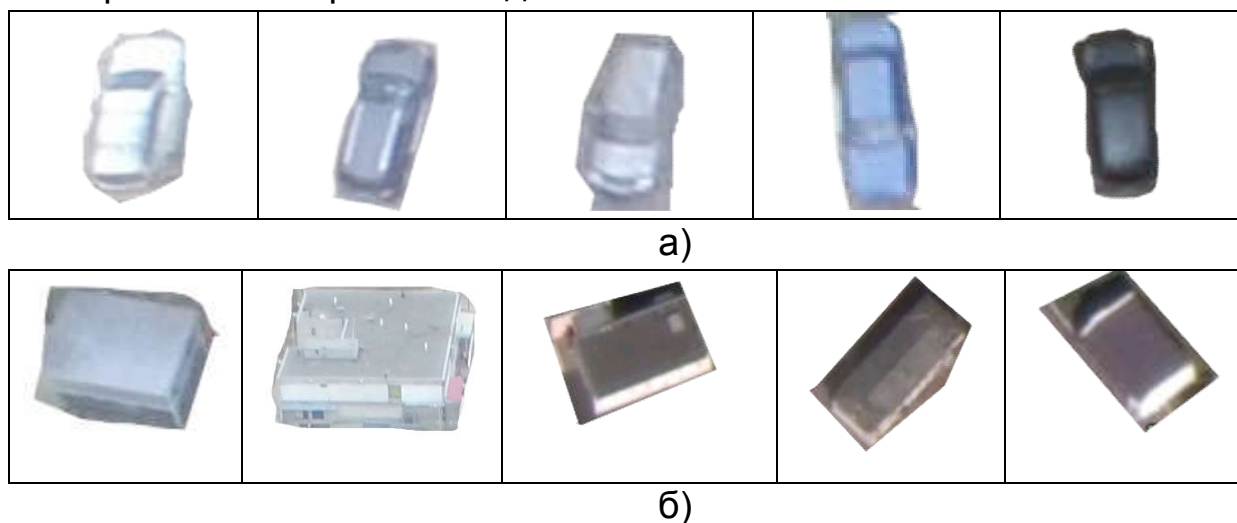


Рис. 2. Изображения, которые использовались для обучения алгоритма:
а) – изображения автомобилей; б) – изображения зданий

Далее система была обучена и построен график отображенный на рисунке 3.

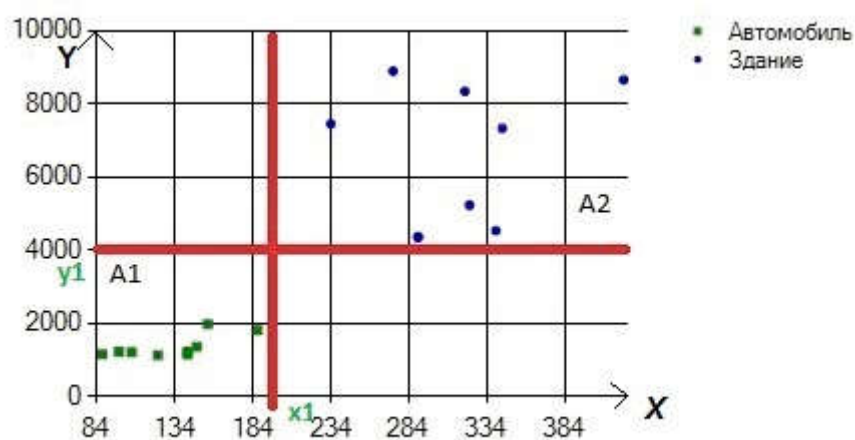


Рис. 3. График обучения алгоритма по выборке

На графике из рисунка 3 видно, что произошла классификация объектов. Где A_1 – это автомобили, A_2 – это здания, X - это максимальное количество ребер, Y - максимальное количество дыр, Y_1 - это координата от которой идут координаты зданий, X_1 - это координата от которой начинаются координаты автомобилей. На основе данной классификации строится дерево, являющееся результатом данного алгоритма. Данное дерево отображено на

рисунке 3, где видно, что левый потомок совпадает с теми элементами множества, у которых предикат верен, а правый с теми у которых предикат возвращает ложь.

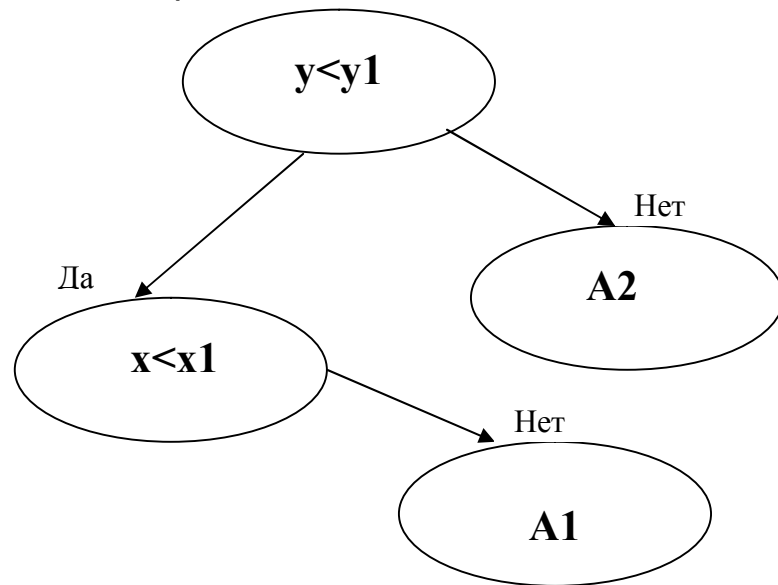


Рис. 4. Дерево решений

Экспериментальные результаты

Для тестирования алгоритма были взяты изображения автомобилей и зданий. На рисунке 5 для проверки алгоритма было взято изображение автомобиля.

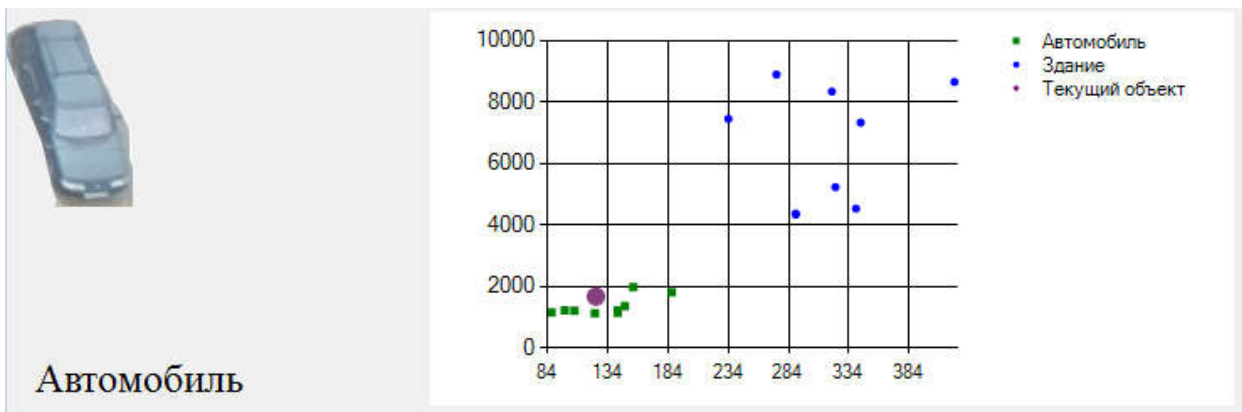


Рис. 5. Определение класса выбранного изображения автомобиля.

На рисунке 6 выбран объект одноэтажного здания. Результатом является вывод его классификации в соответствии с обученной моделью на основе алгоритма Random forest.

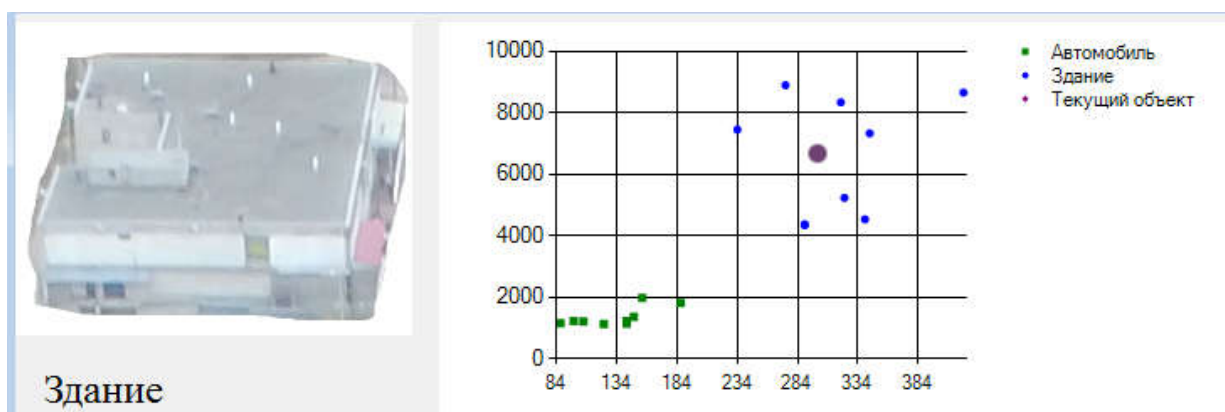


Рис. 6. Определение класса выбранного изображения одноэтажного здания.

Заключение

В статье разработан алгоритм классификации объектов на основе модели Random forest. Подробно описан данный алгоритм, в основе которого лежит модель Random forest, приведены результаты его применения к растровым изображениям.

Данный алгоритм полезен в строительстве при необходимости поиска места для постройки новых зданий и объектов, а также для автолюбителей, например, для поиска своего автомобиля на карте в черте города.

Литература

1. Еремеев С.В., Андрианов Д.Е., Баринов А.Е., Титов Д.В. Алгоритмы поиска объектов по пространственным характеристикам в задачах муниципальных ГИС // Известия юго-западного государственного университета. 2012. №2. С. 37-41. (Часть 3. Серия управление, вычислительная техника, информатика, медицинское приборостроение)
2. Еремеев С.В. Пространственные структуры в геоинформационных системах // Алгоритмы, методы и системы обработки данных . 2007. №12. С. 71-74.
3. Н.Г. Макаренко, Ф.А. Уртьев, И.С. Князева, Д.Б. Малкова, И.Т. Пак, Л.М. Каримова Распознавание текстур на цифровых изображениях методами вычислительной топологии // Журнал «Современные проблемы дистанционного зондирования Земли из космоса». 2015 г.
4. Abdulsalam H., Skillicorn D. B., Martin P. Classification Using Streaming Random Forests // IEEE Transactions on Knowledge and Data Engineering. 2011. Vol. 23, N. 1. P. 22–36.
5. Biau G., Devroye L., Lugosi G. Consistency of random forests and other averaging classifiers // Journal of Machine Learning Research. 2008. Vol. 9. P. 2015–2033.