

РАСПОЗНАВАНИЕ ВОЗДУШНЫХ ОБЪЕКТОВ ПО ВЫСОТНО-СКОРОСТНЫМ
ХАРАКТЕРИСТИКАМ С ИСПОЛЬЗОВАНИЕМ ДЕРЕВА ПРИНЯТИЯ РЕШЕНИЙЧЖО ТУ¹, А. А. КОРСУНОВА², М. В. СВИНАРСКИЙ²¹Белорусский государственный университет информатики и радиоэлектроники
(г. Минск, Республика Беларусь)²Военная академия Республики Беларусь (ВАРБ)
(г. Минск, Республика Беларусь)

E-mail: myloveispublic@gmail.com

Аннотация. В статье представлено исследование по разработке модели классификации воздушных объектов по траекторным признакам с использованием алгоритма дерева решений. В качестве объектов классификации были выбраны пять типовых представителей воздушной обстановки: беспилотные летательные аппараты, крылатые ракеты, вертолеты, винтовые самолеты и турбореактивные самолеты. В качестве классификационных признаков использовались значение высоты и скорости полета типовых объектов. Рассмотрены особенности реализации алгоритма дерева принятия решений. Представлены результаты математического моделирования.

Abstract. The article presents a study on the development of a model for classifying air objects by trajectory features using a decision tree algorithm. Five typical representatives of the air situation were selected as classification objects: unmanned aerial vehicles, cruise missiles, helicopters, propeller-driven aircraft and turbojet aircraft. The altitude and flight speed of typical objects were used as classification features. The features of the implementation of the decision tree algorithm are considered. The results of mathematical modeling are presented.

Введение

В современных военных операциях идентификация класса (типа) наблюдаемого объекта является ключевым фактором обеспечения безопасности. Развитие средств нападения, таких как беспилотные летательные аппараты и гиперзвуковые ракеты, значительно усложнило воздушное пространство. В таких условиях своевременное и точное распознавание наиболее опасных воздушных объектов имеет решающее значение для систем противовоздушной обороны, обеспечивая оперативное реагирование. Большинство существующих систем распознавания воздушных объектов основано на извлечении траекторных характеристик из принимаемого сигнала, и последующей классификации объекта на основе этих данных. Наиболее распространенными траекторными признаками, используемыми для распознавания, являются высота и скорость полета объекта, его ускорение и направление движения [1]. В качестве примера на рисунке 1 представлены типовые высотно-скоростные характеристики наблюдаемых объектов.

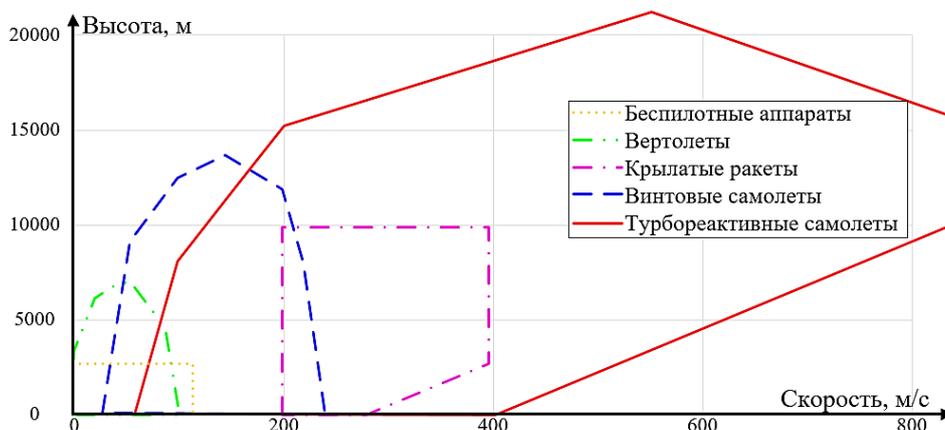


Рис. 1. Типовые высотно-скоростные характеристики наблюдаемых объектов

Классификация объекта заключается в определении принадлежности текущей оценки высоты и скорости к определенной области, на основании которой принимается решение о классе объекта [1]. Однако, поскольку области высотно-скоростных характеристик различных объектов пересекаются (рисунок 1), точность принятых решений о классе наблюдаемого объекта остается низкой. В связи с этим повышение качества принимаемых решений о классе объекта является актуальной задачей.

Для повышения достоверности решений о классе наблюдаемого объекта, основанных на траекторных признаках, перспективным направлением является применение алгоритмов машинного обучения с учителем. Особый интерес представляет алгоритм дерева решений, который отличается простотой интерпретации полученных результатов и высоким качеством принятия решений. Алгоритм дерева решений позволяет строить иерархическую структуру правил, основанную на анализе траекторных признаков.

Особенности реализации алгоритма дерева решений

Применение алгоритма машинного обучения для решения практических задач требует последовательного выполнения ряда этапов, от подготовки данных до оценки качества полученной модели [2]. Основные этапы этого процесса представлены на рисунке 2.

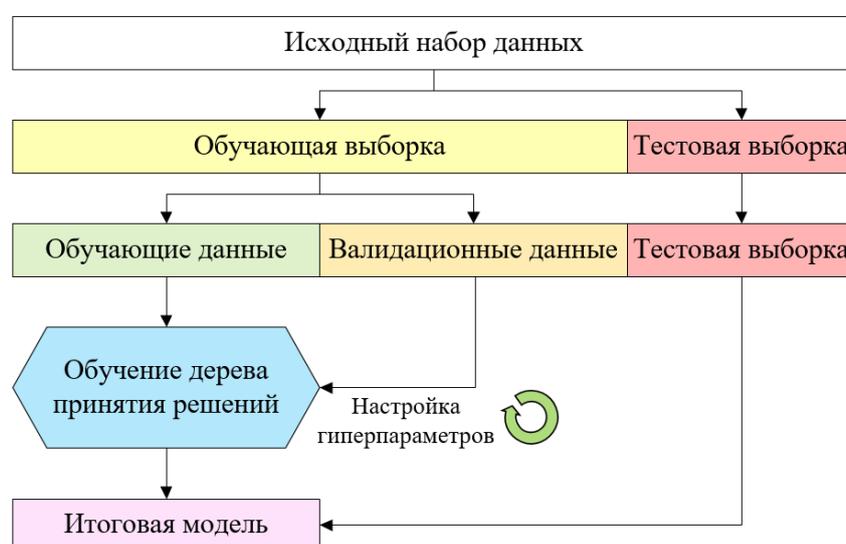


Рис. 2. Последовательность операций при подготовке модели

Процесс обучения включает в себя следующие этапы:

1) Первый шаг заключался в сборе и подготовке данных о высотно-скоростных характеристиках типовых наблюдаемых объектов.

2) Собранные данные разделялись на обучающую и тестовую выборки. Обучающая выборка использовалась для обучения модели, а тестовая – для оценки ее эффективности функционирования.

3) Для подбора оптимальных гиперпараметров модели, обучающая выборка дополнительно разделялась на обучающие и валидационные данные. Обучающие данные использовались для непосредственного обучения алгоритма, а валидационные – для подбора оптимальных значений гиперпараметров. Гиперпараметры – это параметры модели, которые задаются пользователем и влияют на процесс обучения. Их оптимальные значения определяются исходя из максимальной эффективности анализируемой модели. На практике наибольшее распространение получили следующие гиперпараметры [2]:

- *максимальная глубина дерева (max depth)*: Ограничивает максимальное количество уровней в дереве, предотвращая переобучение. Более глубокие деревья могут лучше подгоняться к обучающим данным, но могут быть менее обобщаемыми на новых данных.;

- *минимальное количество наблюдений в узле (min samples split)*: Устанавливает минимальное количество объектов, необходимое для разделения узла. Следует отметить, что min samples split работает как пороговое значение для разделения узла в дереве решений, если количество выборок в узле меньше, чем min samples split, узел не будет разделен и превратится в конечный узел;

- минимальное количество наблюдений в листе (*min samples leaf*): Определяет требуемое минимальное количество объектов, которые должны присутствовать в конечном узле. Это предотвращает бесконечное разделение узлов и гарантирует, что каждый лист имеет достаточное количество объектов для надежного прогнозирования.

Важно отметить, что перечисленные фиксированные гиперпараметры влияют только на количество узлов и листьев в дереве решений.

4) Обучение алгоритма происходит на основе данных обучающей выборки при фиксированных гиперпараметрах. В этом процессе алгоритм учится находить связи между классификационными признаками и классами объектов. В качестве параметров дерева решений, определяемых в процессе обучения, выделяют классификационные признаки, по которым происходит разделение, соответствующие им значения порогов, а также общее количество узлов в дереве принятия решений.

Алгоритм дерева решений представляет собой иерархическую структуру, аналогичную блок-схеме, где каждый внутренний узел соответствует признаку (атрибуту), узел отражает правило принятия решения, а лист дерева представляет решение о классе объекта. Процесс разделения данных осуществляется рекурсивно, пока не будет достигнут один из критериев останова:

- превышена максимальная глубина (общее количество узлов) дерева;
- в анализируемой выборке мало наблюдений в узле (листе);
- все объекты в выборке относятся к одному классу;
- в выборке присутствует только один объект.

Построение дерева решений основано на оптимизации разделения данных на подмножества. Для этого необходимо определить целевую функцию, которая оптимизируется алгоритмом обучения дерева. При выборе признака для разделения узла часто используется жадный алгоритм, который на каждом шаге выбирает наилучший из доступных признаков [2]. Для оценки качества разделения узла используется прирост информации. Прирост информации – это мера, которая оценивает, насколько смешаны классы в узле. Этот критерий оптимизируется алгоритмом дерева решений, и он служит в качестве функции стоимости для алгоритма. Математически прирост информации (IG) рассчитывается следующим образом [2]:

$$IG(th) = S_0 - \sum_{i=1}^q \frac{n_i}{N} S_i, \quad (1)$$

где th – условие, по которому производится разбиение выборки; $S_{0(i)}$ – метрика исходной (i -й) выборки; q – число групп после разбиения; n_i – число объектов в группе после разбиения выборки; N – общее количество элементов в выборке.

Таким образом, прирост информации представляет собой просто разность между метрикой родительского узла и суммой метрик дочерних узлов. Чем ниже метрика дочерних узлов, тем выше прирост информации. Для двоичных деревьев принятия решений каждый родительский узел разбивается на два узла:

$$IG(th) = S_0 - \frac{n_{\text{левый}}}{N} S_{\text{левый}} - \frac{n_{\text{правый}}}{N} S_{\text{правый}}. \quad (2)$$

Для расчета прироста информации существует несколько метрик (S): энтропия, неопределенность Джини, ошибка классификации. Эти показатели измеряют примеси или беспорядок в наборе данных и играют ключевую роль в определении оптимального признака для разделения данных.

Энтропия – это степень хаоса (неопределенности) в системе. Математически энтропия ($S = E$) рассчитывается следующим образом [2]:

$$E = - \sum_{j=1}^M p_j \log_2(p_j), \quad (3)$$

где M – общее количество классов в выборке; p_j – вероятность появления j -го класса из выборки.

Уменьшение энтропии увеличивает однородность набора данных, а значит, позволяет лучше ответить на вопрос о принадлежности класса.

Неопределенностью Джини – это мера разнообразия в распределении вероятностей классов. Метрика Джини эффективна с точки зрения вычислений и хорошо работает для бинарных разбиений. Математически неопределенность Джини ($S = G$) определяется следующим выражением [2]:

$$G = 1 - \sum_{j=1}^M (p_j)^2. \quad (4)$$

Более низкие значения примеси Джини указывают на более чистый набор данных. Если все элементы в узле принадлежат к одному классу, то неопределённость Джини равна 0, а в случае равномерного распределения классов в узле неопределённость Джини равна 0.5.

Ошибка классификации – величина, отображающая долю неправильно классифицированных элементов в узле: чем меньше данное значение, тем меньше загрязнённость в узле. Математически ошибка классификации ($S = CE$) определяется следующим выражением [2]:

$$CE = 1 - \max(p_j). \quad (5)$$

5) В процессе подбора оптимальных гиперпараметров модели дерева решений, было сформировано множество моделей с различными комбинациями гиперпараметров. Оптимальная модель выбиралась на основе максимальной эффективности классификации на валидационной выборке. Для объективной оценки способности модели предоставлять правильные предсказания на новых данных, была проведена оценка ее эффективности на отложенной тестовой выборке. Данный подход позволил выбрать оптимальную модель дерева решений, обеспечивающую максимальную точность классификации воздушных объектов по траекторным признакам.

Результаты математического моделирования

В исследовании была разработана модель классификации воздушных объектов по траекторным признакам. Представителями воздушной обстановки рассматривались пять классов объектов: беспилотные летательные аппараты, крылатые ракеты, вертолеты, винтовые самолеты, турбореактивные самолеты. В качестве классификационных признаков использовались высота и скорость объекта. Исходный набор данных состоял из 200 объектов, по 40 представителей каждого класса. Для обучения модели был применен алгоритм дерева решений. Данные были разделены случайным образом в соотношении 80% на обучающую выборку и 20% на тестовую. Обучающая выборка была дополнительно разделена на обучающие и валидационные данные в соотношении 50% на 50%. Для поиска оптимальных гиперпараметров модели дерева решений был использован алгоритм поиска по сетке с вариациями следующих параметров: максимальная глубина дерева (max depth) от 3 до 20, минимальное количество наблюдений в узле (min samples split) от 2 до 20, минимальное количество наблюдений в листе (min samples leaf) от 1 до 20, критерий разбиения энтропия (3), неопределённость Джини (4), ошибка классификации (5). Оптимальные гиперпараметры были выбраны на основе максимальной эффективности классификации на валидационной выборке. В результате поиска были определены следующие оптимальные параметры: критерий – энтропия, max_depth – 5, min_samples_leaf – 4, min_samples_split – 14. Структура полученного дерева решений была представлена в виде графа (рисунок 3).

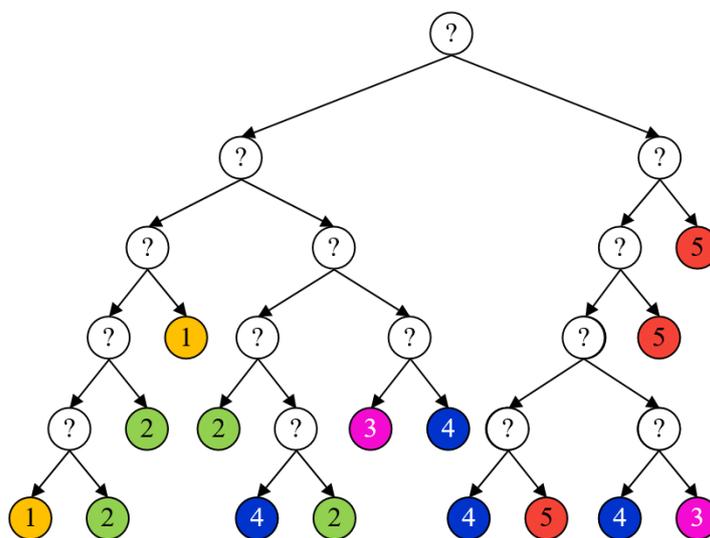


Рис. 3. Структура дерева принятия решений

Качество радиолокационного распознавания класса воздушного объекта оценивалось с помощью следующих метрик: accuracy (точность), precision (точность), recall (полнота), F1-score [2]. Оценка качества принятых решений по тестовой выборке сведена в таблицу 1.

Таблица 1. Показатели качества алгоритма радиолокационного распознавания

Класс объекта	Показатели качества		
	Precision	Recall	F1-score
Беспилотный аппарат	0,85	0,97	0,91
Вертолет	0,71	0,86	0,77
Крылатая ракета	0,73	0,71	0,72
Винтовой самолет	0,72	0,43	0,54
Турбореактивный самолет	0,77	0,82	0,79
Среднее значение	0,76	0,76	0,75

Анализ полученных результатов позволяет утверждать, что обученная модель позволила классифицировать воздушные объекты по траекторным признакам с хорошей точностью (accuracy = 0.76). Не высокая эффективность распознавания классов объектов объясняется пересечением значений классификационных признаков (высоты и скорости). Таким образом, результаты математического моделирования подтверждают возможность использования алгоритма дерева принятия решений для распознавания воздушных объектов по траекторным признакам.

Заключение

Проведенное исследование демонстрирует перспективность применения алгоритма дерева решений для классификации воздушных объектов по траекторным признакам. Полученные результаты свидетельствуют о хорошей эффективности модели, достигающей 76% точности классификации. Однако, анализ отдельных классов объектов выявил необходимость дальнейшего усовершенствования модели, в частности, для повышения точности классификации винтовых самолетов. Перспективным направлением дальнейших исследований является учет ошибок измерения высотно-скоростных характеристик наблюдаемых объектов, а также расширение перечня классификационных признаков.

Таким образом, использование алгоритмов машинного обучения с учителем открывает новые возможности для повышения достоверности распознавания воздушных объектов, что является актуальной задачей для обеспечения безопасности в современных условиях.

Список использованных источников

1. Ширман Я. Д. Радиоэлектронные системы. Основы построения и теория: справочник / Я. Д. Ширман [и др.]; под ред. Я. Д. Ширмана. М. : Радиотехника, 2007. – 512 с.
2. Рашка С., Мирджалили В. Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikitlearn и TensorFlow 2. СПб. : ООО «Диалектика», 2020. – 848 с.