

Использование машинного обучения для управления навигацией робота по показаниям датчиков расстояний

Н.А. Ломовцева
Ульяновский государственный
технический университет
Ульяновск, Россия
natalya.lomovtseva@gmail.com

Ю.Е. Кувайскова
Ульяновский государственный
технический университет
Ульяновск, Россия
u.kuvaiskova@mail.ru

И.Н. Карпунина
Ульяновский институт
гражданской авиации
Ульяновск, Россия
karpunina53@yandex.ru

В.Н. Клячкин
Ульяновский государственный
технический университет
Ульяновск, Россия
v_kl@mail.ru

Д.М. Ястребов
Научно-производственное
объединение «Марс»
Ульяновск, Россия
mars@mv.ru

Abstract—По показаниям ультразвуковых датчиков для измерения расстояний, расположенных на роботе, требуется спрогнозировать направление перемещения робота. Это задача мультиклассовой классификации, решаемая методами машинного обучения. Решение проводилось в системе Statistica с применением метода Random Forest.

Keywords— мультиклассовая классификация, перемещение робота, система Statistica, Random forest.

1. ВВЕДЕНИЕ

По показаниям ультразвуковых датчиков для измерения расстояний, расположенных на роботе, требуется спрогнозировать направление перемещения робота. Предполагается наличие неизвестной зависимости между показаниями датчиков и перемещениями робота. По результатам предварительных испытаний известна совокупность пар: набору показателей X соответствует состояние Y – это исходная выборка результатов наблюдений. По этим данным надо построить алгоритм для восстановления зависимости, с помощью которой по новым данным о показаниях датчиков можно выдать достаточно точный ответ о направлении его перемещения. Это задача мультиклассовой классификации, решаемая методами машинного обучения, или обучения по прецедентам (с учителем) [1-3].

Существует множество методов решения этой задачи. Это и базовые статистические методы (наивный байесовский классификатор, дискриминантный анализ, метод опорных векторов и др.), и нейронные сети, и композиционные методы (бэггинг и бустинг). Для бинарной классификации более эффективны агрегированные методы. В настоящей работе решение проводилось в системе Statistica [4] с применением метода Random Forest (случайный лес), показавшего лучшие результаты именно для мультиклассовой классификации.

2. ПРОВЕДЕНИЕ МАШИННОГО ОБУЧЕНИЯ

Случайный лес – один из наиболее эффективных методов машинного обучения, используемых при решении задач классификации и регрессии [2-3]. Это модель, состоящая из множества независимых деревьев решений и использующая две ключевые концепции при построении каждого дерева, которые и делают этот лес случайным:

1. Случайная выборка наблюдений из обучающего набора данных.
2. Случайный набор предикторов из всех известных предикторов при принятии решений о разбиении узлов.

Модуль случайного леса включен в систему Statistica, позволяет варьировать различные факторы, влияющие на качество обучения, и может быть использован для решения поставленной задачи прогнозирования перемещения робота по результатам показаний датчиков расстояний, расположенных на роботе.

Точность прогнозирования случайного леса зависит от ряда факторов – количества предикторов в случайном наборе, объема подвыборки (случайной выборки из обучающего набора), количества деревьев, максимальной глубины деревьев (максимального количества уровней), максимального количества узлов в деревьях, минимального числа объектов в листьях, минимального количества объектов в дочернем узле [6-7].

Для численного исследования использовались данные, полученные на кафедре инженерной телеинформатики Федерального университета штата Сеара в Форталезе (Бразилия) [5]: 24 ультразвуковых датчика для измерения расстояния были расположены в различных частях робота. Фиксировалось направление перемещения робота: отклик Y , характеризующий перемещение робота, представлен номером соответствующего класса: 1 – движение вперед, 2 – плавный поворот направо, 3 – резкий поворот направо, 4 – плавный поворот налево. Всего получено 5456 наблюдений.

Обозначения показаний датчиков: X_1 – датчик в передней части робота, опорный угол -180° ; X_2 – опорный угол -165° ; X_3 – опорный угол -150° ; X_4 – опорный угол -135° ; ... X_{12} – опорный угол -15° ; X_{13} – датчик в задней части робота, опорный угол 0° ; X_{14} – опорный угол 15° ; ... X_{23} – опорный угол 150° ; X_{24} – опорный угол 165° .

Исследовалось влияние количества случайных предикторов, используемых при принятии решений о разбиении узлов деревьев, максимальной глубины деревьев, а также количества деревьев на точность прогнозирования случайным лесом направлений перемещения робота.

Выбор этих факторов в каждой конкретной задаче может существенно улучшить качество классификации.

3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

При значениях этих факторов, установленных в системе Statistica по умолчанию, построенная матрица неточностей (рис. 1) показывает, что случайный лес присвоил метку 1-го класса 237 наблюдениям на тестовой выборке, из которых верно было распознано 235, что составляет 99,16%, такой же процент правильно распознанных наблюдений от общего числа наблюдений 1-го класса (движение вперед) на тестовой выборке и 41,89% от всех наблюдений на тестовой выборке.

Случайный лес обнаружил 2-й класс (плавный поворот направо) в 77 случаях, все из которых верные (100%) и составляют 98,72% от общего количества наблюдений 2-го класса на тестовой выборке.

Все наблюдения 3-го класса (резкий поворот направо) на тестовой выборке были распознаны верно, что составляет 100%.

Процент правильно распознанных наблюдений 4-го класса (плавный поворот налево) от общего количества класса на тестовой выборке – 92,86% (рис. 1).

Classification matrix (data.sta)						
Response: Y						
Test set sample; Number of trees: 100						
	Observed	Class Predicted 1	Class Predicted 2	Class Predicted 3	Class Predicted 4	Row Total
Number	1	235		2		237
Column Percentage		99.16%	0.00%	0.90%	0.00%	
Row Percentage		99.16%	0.00%	0.84%	0.00%	
Total Percentage		41.89%	0.00%	0.36%	0.00%	42.25%
Number	2		77	1		78
Column Percentage		0.00%	100.00%	0.45%	0.00%	
Row Percentage		0.00%	98.72%	1.28%	0.00%	
Total Percentage		0.00%	13.73%	0.18%	0.00%	13.90%
Number	3			218		218
Column Percentage		0.00%	0.00%	98.64%	0.00%	
Row Percentage		0.00%	0.00%	100.00%	0.00%	
Total Percentage		0.00%	0.00%	38.86%	0.00%	38.86%
Number	4	2			26	28
Column Percentage		0.84%	0.00%	0.00%	100.00%	
Row Percentage		7.14%	0.00%	0.00%	92.86%	
Total Percentage		0.36%	0.00%	0.00%	4.63%	4.99%
Count	All Groups	237	77	221	26	561
Total Percent		42.25%	13.73%	39.39%	4.63%	

Рис. 1. Матрица неточностей случайного леса (количество случайных предикторов – 5, максимальная глубина – 10, количество деревьев – 100)

На рис. 2 представлен процесс построения случайного леса: по горизонтальной оси отложено количество деревьев, по вертикальной – доля ошибок классификации на обучающей (красная линия) и тестовой (синяя линия) выборках. Видно, что процесс практически стабилизировался при использовании 50 деревьев.

Наилучший результат прогнозирования случайным лесом направлений перемещения робота обеспечивается при построении 50 деревьев с максимальной глубиной – 10, которые принимают решения о разбиении узлов с использованием семи случайных предикторов.

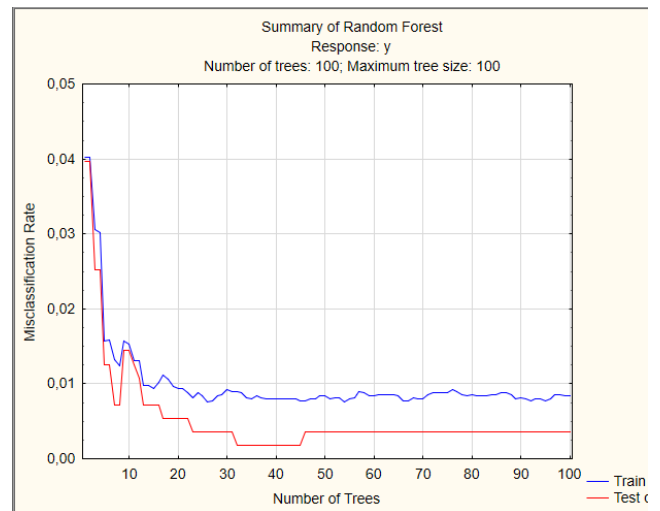


Рис. 2. Построение случайного леса

4. ЗАКЛЮЧЕНИЕ

Проведенное исследование показало, что мультиклассовая классификация в системе Statistica с применением случайного леса обеспечивает достаточно качественное прогнозирование направления перемещений робота по результатам показаний датчиков расстояний.

Рассмотренный подход может быть использован и при диагностике сложных технических систем, когда необходимо оценить по результатам наблюдений за показателями функционирования, в каком из множества состояний находится исследуемый объект.

ЛИТЕРАТУРА

- [1] Witten, I.H. Data mining practical machine learning tools and techniques / I.H. Witten, E. Frank. – SF: Morgan Kaufmann Publ, 2005.
- [2] Клячкин, В.Н. Прогнозирование и диагностика стабильности функционирования технических объектов: монография / В.Н. Клячкин, В.Р. Крашенинников, Ю.Е. Кувайскова // Москва: Русайнс, 2020.
- [3] Wang, B. A Multicategory Kernel Distance Weighted Discrimination Method for Multiclass Classification / B. Wang, H. Zou // Technometrics. – 2019. – Vol. 61(3). – P. 396-408. DOI: 10.1080/00401706.2018.1529629.
- [4] Боровиков, В.П. STATISTICA. Искусство анализа данных на компьютере / В.П. Боровиков. – СПб: Питер, 2003.
- [5] Freire, A.L. Short-Term Memory Mechanisms in Neural Network Learning of Robot Navigation Tasks: A Case Study / A.L. Freire, G.A. Barreto, M. Veloso, A.T. Varela // Proceedings of the 6th Latin American Robotics Symposium (LARS). – Valparaíso-Chile, 2009. – P. 1-6.
- [6] Wyner, A.J. Explaining the success of adaboost and random forests as interpolating classifiers / A.J. Wyner // The Journal of Machine Learning Research. – 2017. – Vol. 18(1). – P. 1558-1590.
- [7] Kuvayskova, Yu. Recognition and Forecasting of a Technical Object State based on its Operation Indicators Monitoring Result / Yu. Kuvayskova, V. Klyachkin, V. Kracheninnikov // International Multi-Conference on Industrial Engineering and Modern Technologies. – 2020. – P. 9271119. DOI: 10.1109/FarEastCon50210.2020.9271119.