

Литература

1. Baker J.W., M. Schubert, Faber M. H. On the Assessment of Robustness, Journal of Structural Safety. - 2008, т.30, с. 253-267.
2. Starossek, U., Haberland M. Disproportionate Collapse: Terminology and Procedures. Journal of Performance of Constructed Facilities. - 2010, 24(6), с.519-528.
3. Махутов Н.А., Резников Д.О. Использование сценарного анализа для оценки прочностной надежности сложных технических систем//Проблемы машиностроения и автоматизации. - 2015. № 1. С. 4-13.
4. Доронин С.В., Филиппова Ю.Ф. Оценка живучести структурно-сложных силовых конструкций на основе сценариев накопления повреждений//Вестник инженерной школы ДВФУ. - 2019. № 4(41) стр. 39-54.
5. Махутов Н.А., Резников Д.О. Оценка прочностной живучести технических систем//Проблемы безопасности и чрезвычайных ситуаций. - 2019. №4 с. 47-57.
6. Махутов Н.А., Резников Д.О., Зацаринный В.В. Два типа сценариев аварий в сложных технических системах//Проблемы безопасности и чрезвычайных ситуаций. - 2014. № 2. С. 28-41.
7. Махутов Н.А., Резников Д.О. Многоуровневая оценка живучести сложных технических систем с учетом масштабно-структурной иерархии процессов накопления повреждений и разрушения//Безопасность в техносфере. - 2016. Т. 5. №4.С. 3-17
8. Reznikov D.O. Assessment of constructional robustness of a complex technical system: A multi-level approach. AIP Conference Proceedings 2053, 040080 (2018); <https://doi.org/10.1063/1.5084518>. Published Online: 19 December 2018

Сведения об авторе

Резников Дмитрий Олегович, ведущий научный сотрудник, Федерального государственного бюджетного учреждения науки Институт машиноведения им. А.А. Благонравова РАН (ИМАШ РАН). e-mail: imashreznikoff@yandex.ru, тел. +7 (495) 623 58 35

УДК 004.853

DOI: 10.36535/0869-4176-2020-05-4

АНАЛИЗ ДАННЫХ ПО ЛЕСНЫМ ПОЖАРАМ

Кандидат физ.-мат. наук Л.Р. Борисова
Финансовый университет при Правительстве РФ, Московский физико-технический институт (государственный университет)

Доктор сельхоз. наук, кандидат техн. наук Ю.В. Подрезов
ФБГУ ВНИИ ГОЧС (ФЦ)
Московский физико-технический институт

Кандидат биолог. наук А.В. Кузнецова
Институт Биохимической физики им. Н.М. Эмануэля РАН

З.В. Тимошенко
ВИНИТИ РАН

Цель работы – прогноз лесных пожаров на основе известных литературных данных о пожарах в парке Монтезиньо. Построена модель множественной регрессии зависимости площади лесного пожара от погодных условий. Проведен анализ данных методами машинного обучения. Получено высокое значение правильности распознавания, если группирующий показатель – это числовой индекс ожидаемых темпов распространения огня, индекс легкости воспламенения топлива FFMC или площадь лесного пожара. Выполнен дисперсионный анализ данных, подтверждающий вывод о значимости фактора температуры в распространении лесных пожаров.

Ключевые слова: дисперсионный анализ, регрессия, машинное обучение, статистические факторы, метеорологические данные, лесные пожары, анализ данных.

ANALYSIS OF FOREST FIRE DATA

Ph.D. (Phys.-Mat.) L.R. Borisova

Financial University under the Government of RF, Moscow Institute of physics and technology (state University)

Dr. of agricultural sciences, Ph.D (Tech) J.V. Podrezov

FC VNII GOCHS EMERCOM of Russia

Moscow Institute of physics and technology (state University)

Ph.D. (Bio.) A.V. Kuznetsova

Institute Emanuel Biochemical Physics RAS

Z.V. Timoshenko

VINITI RAN

The purpose of this work is to forecast forest fires based on the well-known literature data on fires in The Montezinho Park. A model of multiple regression of the dependence of the forest fire area on weather conditions is constructed. The data was analyzed using machine learning methods. A high recognition accuracy value is obtained if the grouping indicator is a numerical index of the expected fire propagation rate, the index of ease of ignition of fuel or the area of a forest fire. A dispersion analysis of the data confirming the conclusion about the significance of the temperature factor in the spread of forest fires was performed.

Keywords: variance analysis, regression, machine learning, statistical factors, meteorological data, forest fires, data analysis.

Лесным пожаром называют стихийное и неуправляемое распространение огня по лесным площадям. Крупные лесные пожары представляют опасность для многих стран мира, и, прежде всего, для Российской Федерации с ее наибольшим среди других государств объемом лесных зон. Данные по распространению пожаров в Российской Федерации постоянно обновляются, и представлены в открытом доступе на сайте Федерального агентства лесного хозяйства. Наличие исторических баз данных по таким показателям, как влажность, температура воздуха, количество осадков, сила и направление ветра, дают возможность выработать гипотезы по зависимости частоты и размеров пожаров от предложенных величин. Ранее нами были представлены результаты исследований с использованием методов машинного обучения, таких как метод дерева решений, метод

случайного леса и метод к-ближайших соседей. Были построены ROC-кривые для каждого из этих методов. Получены оценки точности. Среди всех имеющихся моделей большую точность показало простое дерево решений, так же давшее наибольшую скорость расчета результата, однако точность всех трех моделей небольшая. Для проведенного ранее анализа были использованы открытые данные по погодным условиям и данные о лесных пожарах в Российской Федерации [1-3].

В настоящей работе анализируются более структурированные данные по пожарам в Португалии, чем рассмотренные ранее данные в Российской Федерации. В России для оценки пожарной опасности применяют индекс Нестерова. Комплексный показатель пожарной опасности В.Г. Нестерова, используемый в РФ, содержит меньше параметров, чем Канадская система индекса погоды лесных пожаров (FWI), состоящая из шести компонентов, которые учитывают влияние влажности топлива и ветра на поведение пожара. Показатель лесной пожарной опасности (Forest Fire Danger Meter, FFDMM), используемый в Канадской системе, рассчитывается на основе температуры, относительной влажности, скорости ветра, запаса ЛГМ, наклона поверхности и фактора сухости. Он позволяет предсказывать вероятность возникновения пожара, скорость его распространения, его интенсивность, трудность подавления. Данные содержат большой массив из 517 строк, что позволяет проводить исследования с использованием разных статистических методов [1,4,5].

Сначала по данным Р. Cortez and А. Morais, мы построили уравнение множественной регрессии, взяв за результирующий признак площадь распространения пожара. Однако получили статистически незначимую модель с низким коэффициентом корреляции. Модель множественной регрессии – следующая [4]:

$$y=2,49-0,023x_1+0,076x_2-0,006x_3-0,7x_4+0,856x_5-0,26x_6+1,52x_7-2,56x_8 \quad (1)$$

В этом уравнении переменные $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$ обозначают *FFMC* — индекс легкости воспламенения топлива из системы *FWI* в интервале значений 18,7–96,2; *DMC* — индекс нормы влажности угля из системы *FWI* в интервале значений 1,1–291,3; *DC* — индекс нормы засухи от системы *FWI* в интервале значений 7,9–860,6; *ISI* — индекс первоначального распространения из системы *FWI* в интервале значений от нуля до 56,1; температуру в интервале значений 2,2–33,3°C; относительную влажность от 15,0 до 100 %; скорость ветра от 0,4 до 9,4 км/ч; дождь снаружи от 0,0 до 6,4 мм/м² (соответственно). Модель была построена для данных о площади пожара леса от 0,00 до 1090,84 га, но для большей информативности на рис.1 представили данные только до 200 га.

Множественный коэффициент корреляции равен 0,126, модель статистически незначима (значимости F-критерия составила 0,41), значения *p-value* – большие для всех коэффициентов уравнения регрессии.

Дискриминантный анализ показал, что наиболее значимым параметром является температура, что не является удивительным. Такой же вывод мы получили при анализе данных по лесным пожарам в Российской Федерации [1].

Данные дискриминантного анализа представлены на рис. 2.

Из дискриминантного анализа данных по Р. Cortez and А. Morais следует, что на факт лесного пожара на 87% влияет высокая температура окружающего воздуха. Из представленных в Канадской системе индексов интересен индекс *ISI* - числовой индекс ожидаемых темпов распространения огня. Для данных Р. Cortez and А. Morais этот индекс изменяется в диапазоне от 0 до 56,1, причем последнее значение принимается всего один раз [4].

Согласно всем статистическим методам обработки данных это значение можно принять за выброс [7].

Изучим зависимость между температурой – основным фактором, влияющим за распространение лесного пожара и индексом *ISI*, выбросим из массива данных единственную точку, при которой значение индекса *ISI* было 56,1 (рис. 3).

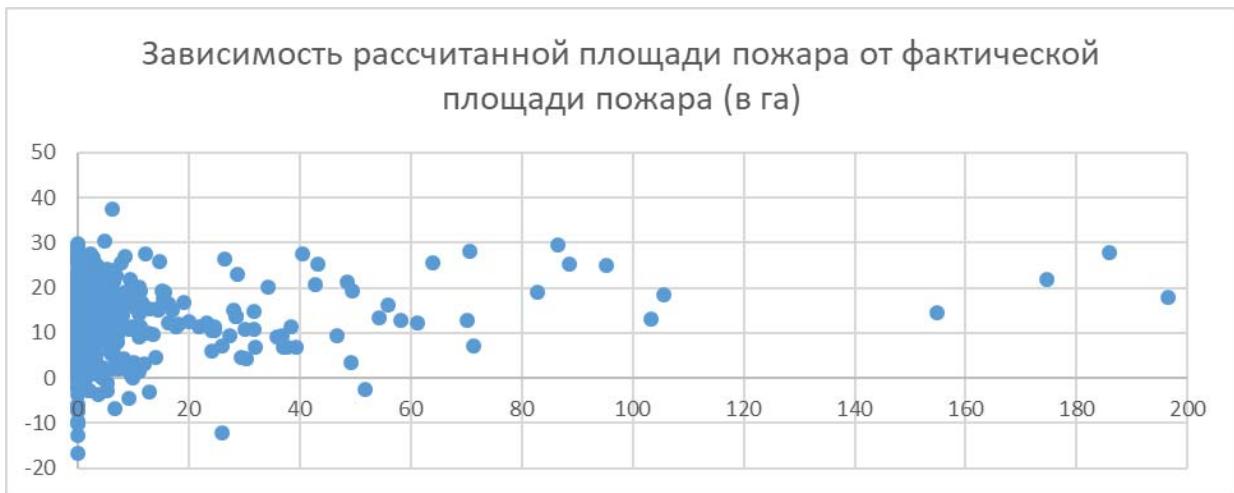


Рис. 1. Зависимость рассчитанной площади пожара от фактической (в га) по данным уравнения (1)

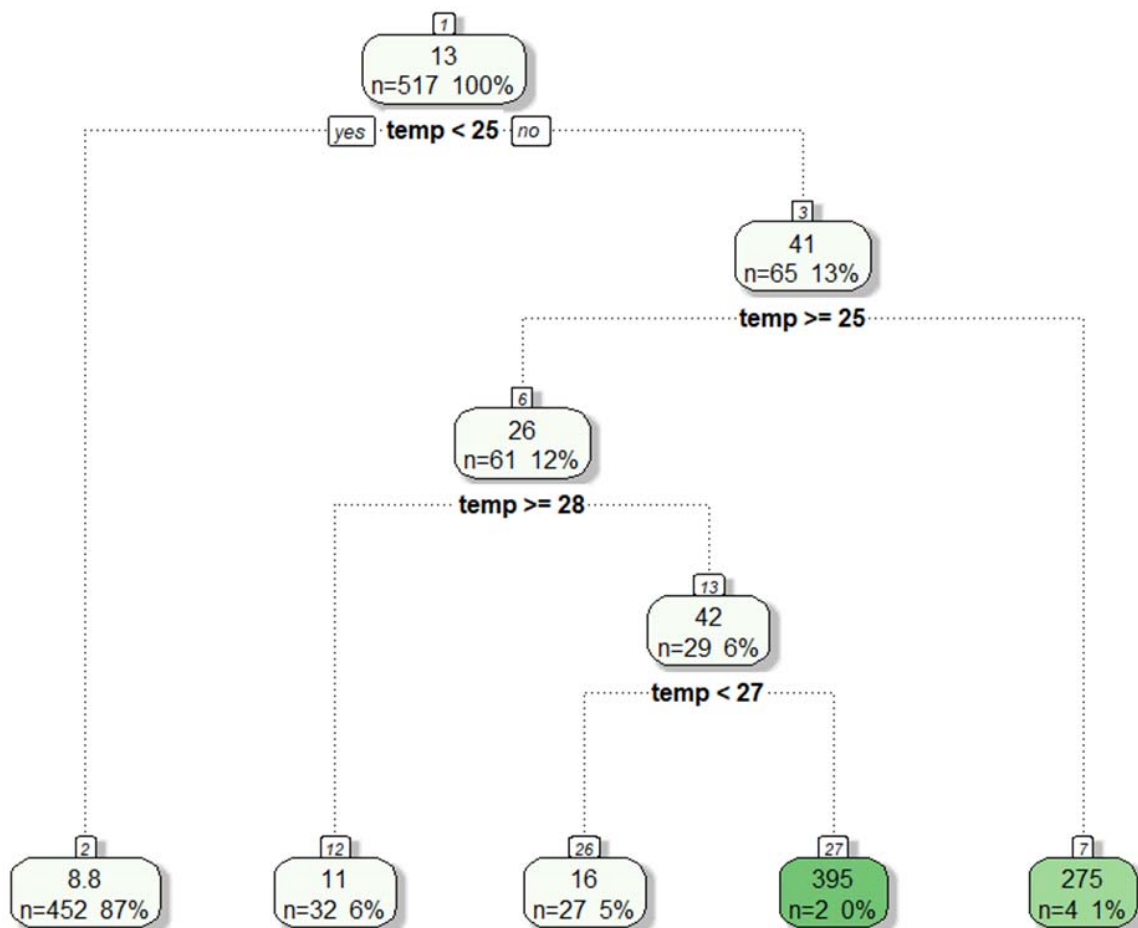


Рис. 2. Результаты дискриминантного анализа данных, представленных в работе [4]

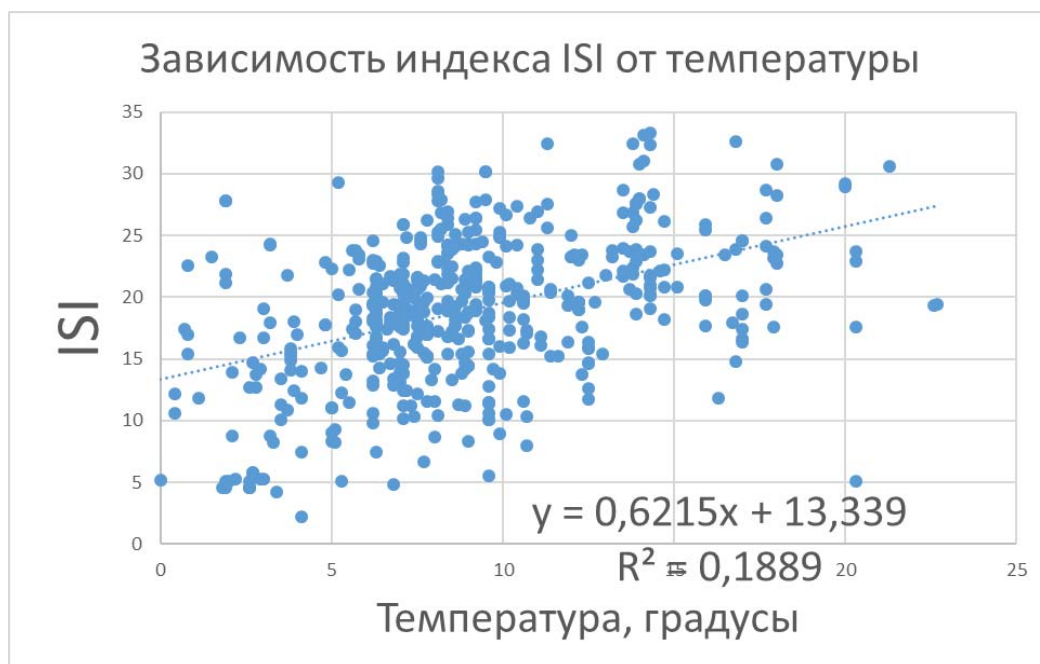


Рис. 3. Результаты аппроксимации данных по индексу ISI от температуры окружающего воздуха (в градусах), представленных в работе, линейной зависимостью [4]

Уравнение линейной модели: $y = 3,19 + 0,3x$. В этой модели коэффициент корреляции равен 0,19, что говорит о наличии пусть слабой, но устойчивой положительной связи между температурой и индексом *ISI*. Модель – статистически устойчива. *P*-значение для обоих коэффициентов в уравнении практически равно 0 (10^{-7} и 10^{-25} соответственно), значимость *F*-критерия равна 10^{-25} .

Применение методов машинного обучения к данным [4] позволит выявить скрытые закономерности в этих данных.

Проанализируем данные Р. Cortez and А. Morais с помощью метода Статистически Взвешенных Синдромов (СВС), хорошо зарекомендовавшего себя при анализе больших массивов в различных сферах. В качестве целевого показателя (группирующей) был выбран числовой индекс ожидаемых темпов распространения огня (*ISI*). Граница разбиения на классы – среднее значение по показателю - 9. Общее количество объектов 517. Количество объектов первого класса (со значениями по *ISI* ниже границы 9) – 310, второго класса - 207 [4,6,7].

На рис. 4 показан принцип действия метода СВС. С помощью метода Оптимально Достоверных Разбиений, основанного на принципе Хи-квадрат, ставится граница таким образом, чтобы с одной стороны преобладали объекта одного класса, а с другой стороны – объекты второго класса. Затем на одномерных и двумерных разбиениях проводится взвешенное голосование по базовым множествам. Проверка точности распознавания осуществляется с помощью скользящего контроля (cross validation, CV).

В результате работы метода СВС получили долю правильных ответов алгоритма (Accuracy) - 0,824. Это намного больше ранее проведенного распознавания по данным о связи факта пожара в Российской Федерации с погодными условиями. Эти результаты свидетельствуют о том, что использование таких параметров, как влагосодержание лесной подстилки и почвы, характеристик пламени, антропогенного фактора и грозовой активности, что не учитывалось при распознавании в работе Борисовой Л.Р., Подрезова Ю.В. и Анисимова А.А. "Анализ изменения статистических факторов возникнове-

ния лесных пожаров в Российской Федерации., существенно повышает как правильность прогноза, так и точность прогноза" [1].

На рис. 5 представлены результаты распознавания данных по распространению лесных пожаров, если группирующая – это числовой индекс ожидаемых темпов распространения огня (*ISI*), который не рассчитывается в Российской Федерации.

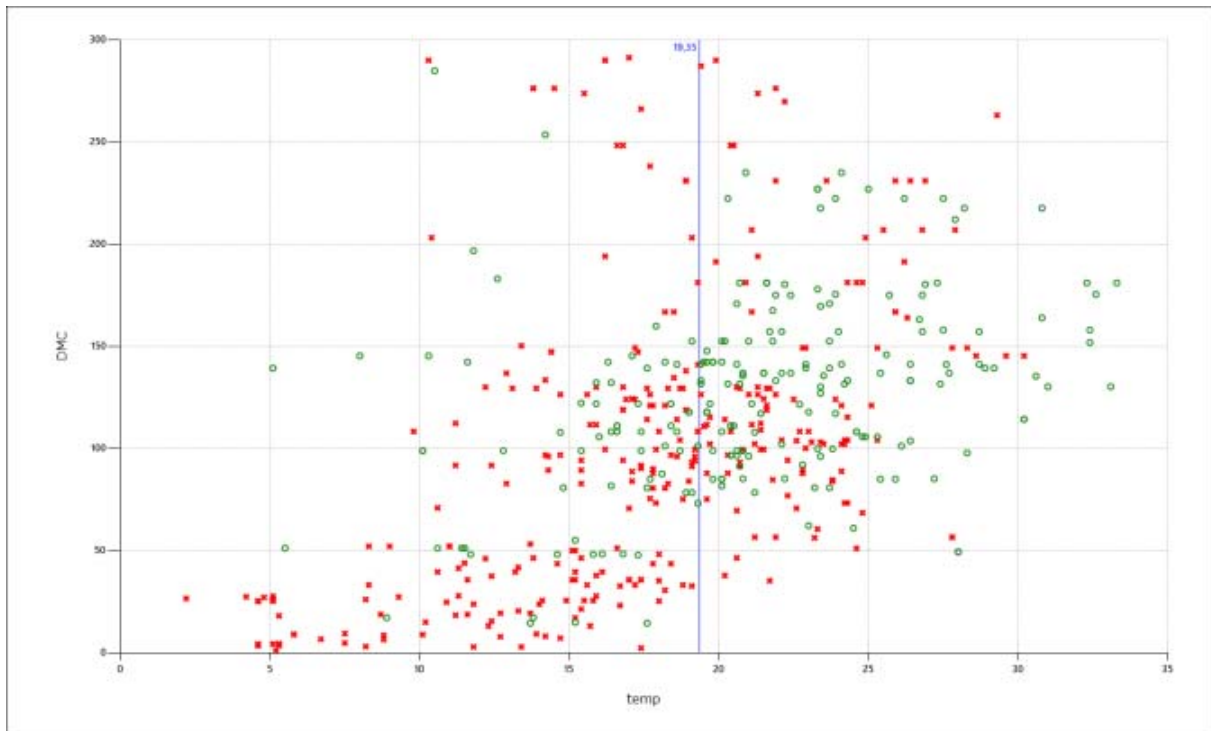


Рис. 4. Одномерная диаграмма. В первом квадранте преобладает первый класс – \times (объекты со значениями $ISI < 9$). Во втором квадранте преобладает второй класс – \circ ($ISI > 9$)

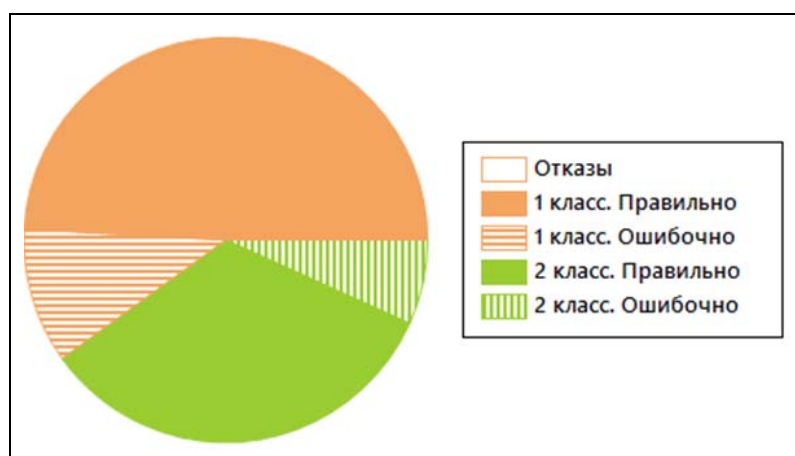


Рис. 5. Диаграмма качества классификации данных по лесным пожарам, рассчитанная по методу статистически взвешенных синдромов [4]

Таблица 1

Качество классификации для целевой FFMC =91,2

Accuracy (Правильность)	0,6914
Precision (Точность)	0,6359
Sensitivity (Чувствительность)	0,6121
Specificity (Специфичность)	0,7483

Таблица 2

Одномерные закономерности, выявленные при целевой функции DC

n	Показатель	Граница разбиения	Ниже границы		Выше границы		X ²	p <
			1 класс	2 класс	1 класс	2 класс		
1	sep_month	0,5	263 (100 %)	82 (32,3 %)	0 (0 %)	172 (67,7 %)	266,3	0,0005
2	DMC	52,75	105 (39,9 %)	14 (5,5 %)	158 (60,1 %)	240 (94,5 %)	86,18	0,0005
3	mar_month	0,5	209 (79,5 %)	254 (100 %)	54 (20,5 %)	0 (0 %)	58,12	0,0005
4	FFMC	90,25	98 (37,3 %)	34 (13,4 %)	165 (62,7 %)	220 (86,6 %)	38,67	0,0005
5	jul_month	0,5	232 (88,2 %)	253 (99,6 %)	31 (11,8 %)	1 (0,4 %)	28,83	0,0005
6	feb_month	0,5	243 (92,4 %)	254 (100 %)	20 (7,6 %)	0 (0 %)	20,05	0,0005
7	aug_month	0,5	147 (55,9 %)	186 (73,2 %)	116 (44,1 %)	68 (26,8 %)	16,9	0,0005
8	jun_month	0,5	246 (93,5 %)	254 (100 %)	17 (6,5 %)	0 (0 %)	16,94	0,0005
9	area	1,705	176 (66,9 %)	123 (48,4 %)	87 (33,1 %)	131 (51,6 %)	18,09	0,001
10	temp	25,45	247 (93,9 %)	211 (83,1 %)	16 (6,1 %)	43 (16,9 %)	15	0,001
11	apr_month	0,5	254 (96,6 %)	254 (100 %)	9 (3,4 %)	0 (0 %)	8,83	0,002
12	dec_month	0,5	254 (96,6 %)	254 (100 %)	9 (3,4 %)	0 (0 %)	8,83	0,002
13	oct_month	0,5	261 (99,2 %)	241 (94,9 %)	2 (0,8 %)	13 (5,1 %)	8,69	0,007
14	RH	47,5	194 (73,8 %)	150 (59,1 %)	69 (26,2 %)	104 (40,9 %)	12,53	0,012
15	wind	7,8	260 (98,9 %)	239 (94,1 %)	3 (1,1 %)	15 (5,9 %)	8,71	0,032

Кроме того, нами было выполнено распознавание в случае, когда группирующим показателем был индекс легкости воспламенения топлива (FFMC). Очевидно, что данный показатель также является критическим для распространения лесного пожара. В этом случае доля правильных ответов алгоритма (Accuracy) составила 0,69. На скользящем контроле правильно распознан 131 объект в 1 классе (61,21 %), 223 объекта во втором классе (74,83 %). Всего 354 объектов из 517.

При назначении целевой функцией показателя DC - индекса нормы засухи (деление по среднему, равному 664) – доля правильных ответов алгоритма (Accuracy) достигла 91,88%. Наиболее информативными показателями стали в порядке убывания информативности: sep_month, DMC, mar_month, FFMC, jul_month, feb_month, aug_month, jun_month, area, temp, apr_month, dec_month, oct_month, RH, wind. (таблица 2).

В табл. 1 видим, что из 30 показателей, вошедших в базу данных, информативными стали 15. (Число показателей выросло за счет того, что каждый месяц года и каждый день недели были преобразованы в отдельный показатель-столбец). Из них 8 показателей при вычислении значимости с помощью перестановочного теста были на высоком уровне - $p < 0,0005$.

И, наконец, если в качестве целевой переменной взять площадь лесного пожара, то получается практически такой же результат (Accuracy=0,64). Этот результат был получен при разделении на 2 класса на уровне Area = 0,5 (га). Если сдвигать эту границу в сторону увеличения площади пожара, то результат распознавания будет меняться в соответствии с табл. 3.

Качество классификации для целевой «Площадь лесного пожара» при смещении границы разбиения на классы

Граница класса по Area (га)	Число объектов в классах: N1/N2	Число правильных распознаваний в классах	Accuracy (Доля правильных ответов - %)	Precision (Точность)	Sensitivity (Чувствительность)	Specificity (Специфичность)
0,5	257 / 260	145 / 188	64,41	0,6713	0,5642	0,7259
1,0	274 / 243	164 / 167	64,02	0,6862	0,6052	0,6901
3,0	335 / 182	231 / 103	64,6	0,7452	0,6896	0,5659
4,0	355 / 162	310 / 53	70,21	0,7399	0,8732	0,3272
5,0	366 / 151	317 / 52	71,37	0,7620	0,8661	0,3444
6,0	378 / 139	325 / 47	71,95	0,7794	0,8598	0,3381
8,0	405 / 112	368 / 33	77,56	0,8233	0,9086	0,29466
10,0	422 / 95	397 / 19	80,46	0,8393	0,9408	0,2000
15,0	448 / 69	149 / 61	40,62	0,9490	0,3326	0,8841
20,0	458 / 59	190 / 50	46,42	0,9548	0,4148	0,8475
30,0	474 / 43	274 / 31	58,99	0,9580	0,5781	0,7209
40,0	486 / 31	241 / 27	51,84	0,9837	0,4959	0,8710
50,0	493 / 24	297 / 21	61,51	0,9900	0,6024	0,8750
60,0	497 / 20	260 / 19	53,97	0,9962	0,5231	0,9500
70,0	499 / 18	292 / 17	59,77	0,9966	0,5852	0,9444

В табл. 3 в столбце Sensitivity (Чувствительность) отражена точность распознавания для 1 класса. Она максимальна при разделении на классы при границе площади пожара равной 10 га. В столбце Specificity (Специфичность) отражена точность для 2 класса, и она максимальна для границы площади в 60 га. Но при этом делении второй класс уже очень маленький. Всего 20 объектов.

Все результаты распознавания, полученные с применением метода взвешенных синдромов, дали правильность распознавания данных по лесным пожарам в Португалии, превышающую аналогичный результат по распознаванию данных по лесным пожарам в Российской Федерации, более чем в 20 раз, что говорит о необходимости учета большего числа параметров, получаемых из статистических данных о лесных пожарах.

Таким образом, в данной статье проанализированы известные литературные данные по зависимости площади лесного пожара от различных погодных условий и погодных индексов Канадской системы.

Сделан вывод о недостаточности использования таких общепринятых статистических методов анализа данных, как регрессионные методы и дисперсионный анализ. Именно методы машинного обучения позволяют создать модель распознавания, которую можно обучать в дальнейшем и получить результаты распознавания данных с хорошей точностью.

Литература

1. Борисова Л.Р., Подрезов Ю.В., Анисимов А.А. Анализ изменения статистических факторов возникновения лесных пожаров в Российской Федерации. //Проблемы безопасности и чрезвычайных ситуаций. –М. - 2019. – № 6. – 33-38.
2. <http://aisori-m.meteo.ru/waisori/select.xhtml>.
3. http://rosleshoz.gov.ru/information_systems.
4. P. Cortez and A. Morais. A Data Mining Approach to Predict Forest Fires using Meteorological Data. In J. Neves, M. F. Santos and J. Machado Eds., New Trends in Artificial Intelligence, Proceedings

of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, December, Guimaraes, Portugal, pp. 512-523. - 2007. APPIA, ISBN-13 978-989-95618-0-9.

5. Van Wagner, C.E. and T.L. Pickett. 1985. Equations and FORTRAN program for the Canadian Forest Fire Weather Index System. Can. For. Serv., Ottawa, Ont. For. Tech. Rep. 33. 18 p.

6. Борисова Л.Р., Кузнецова А.В., Сергеева Н.В., Сенько О.В. Применение методов машинного обучения для сравнения компаний Арктической Зоны РФ по экономическим критериям в соответствии с рейтингом Полярного Индекса. // Компьютерные исследования и моделирование. - 2010. Т.12, № 2. С.201-216.

7. Кузнецова А.В., Сенько О.В., Кузнецова Ю.О. Преодоление проблемы черного ящика при использовании методов машинного обучения в медицине. Журнал «Врач и информационные технологии. Искусственный интеллект в здравоохранении». - 2018, № 7, (Октябрь), с. 74-80. http://www.idmz.ru/media/vit_ru/2018/4/vititm_2018_07pdf.pdf

Сведения об авторах

Борисова Людмила Робертовна, доцент департамента анализа данных, принятия решений и финансовых технологий Финансового университета при Правительстве РФ; доцент кафедры «Высоких технологий в обеспечении безопасности жизнедеятельности» МФТИ (ГУ). Тел. 8-916-245-71-27 E-mail: borisovalr@mail.ru.

Подрезов Юрий Викторович, доцент, главный научный сотрудник научно-исследовательского центра ФГБУ ВНИИ ГЧС (ФЦ); заместитель заведующего кафедрой Московского физико-технического института (государственного университета). Тел.: 8-903-573-44-84; E-mail: uvp4@mail.ru

Кузнецова Анна Викторовна, старший научный сотрудник института Биохимической физики им.Н.М. Эмануэля РАН. Тел.: 8-903-253-84-23; e-mail: azforus@yandex.ru;

Тимошенко Зинаида Владимировна, научный сотрудник ОНИ по транспорту ВИНТИ РАН. тел.: 8-916-959-68-99; e-mail: tzv48@mail.ru.

УДК 001.891-001.892

DOI: 10.36535/0869-4176-2020-05-5

ОЦЕНКА ВКЛАДА ИСПОЛНИТЕЛЕЙ В ВЫПОЛНЕНИЕ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЫ

Доктор техн. наук Р.А. Дурнев
ФГБУ РАРАН

И.В. Жданенко
ФГБУ ВНИИ ГОЧС

Представлен подход к оценке вклада исполнителей в результаты, полученные при выполнении НИР. Данные оценки могут использоваться для определения квалификации исполнителей, целесообразности их привлечения к предполагаемым НИР в НИУ, а также