цифровых технологий, инвестировать в людей и обеспечить глобальный доступ и участие в этом процессе.

1. Балынская Н.Р., Коптякова С.В., Зиновьева Е.Г. Анализ использования информационно-коммуникационных технологий в органах муниципального управления // Муниципалитет: экономика и управление. 2019. №1 (26). URL: <https://cyberleninka.ru/article/n/analiz-ispolzovaniya-informatsionno-kommunikatsionnyh-tehnologiy-v-organah-munitsipalnogo-upravleniya> (дата обращения: 31.05.2020).
2. Гавриков Ф. А. Зарубежный опыт внедрения информационных технологий в сферу государственного и муниципального управления // Актуальные проблемы социально-гуманитарного и научно-технического знания. 2017. № 4 (13). С. 11-13.
3. Леньшина А. А. Особенности коммуникативных технологий в деятельности органов государственной власти / А. А. Леньшина. — Текст: непосредственный // Молодой ученый. — 2019. — № 9 (247). — С. 23-25. — URL: <https://moluch.ru/archive/247/56814/> (дата обращения: 31.05.2020).
4. Цифровое будущее государственного управления по результатам / Е. И. Добролюбова, В. Н. Южаков, А. А. Ефремов, Е. Н. Клочкова, Э. В. Талапина, Я. Ю. Старцев. — М. : Издательский дом «Дело» РАНХиГС, 2019. — 114с. — (Научные доклады: государственное управление).

Счастливцев Е.Л., Юкина Н.И.

Методы искусственного интеллекта для оценки качества вод

*Институт вычислительных технологий СО РАН
(Россия, Кемерово)*

doi: 10.18411/sr-10-06-2020-11

idsp: sciencerrussia-10-06-2020-11

Аннотация

Рассматривается задача оценки качества природных вод с применением методов искусственного интеллекта: градиентный бустинг (eXtreme Gradient Boosting), случайный лес (Random Forest), логистическая регрессия (Logistic Regression), метод *k*-ближайших соседей (k Nearest Neighbors) и нейронная сеть. В зависимости от точности предсказаний моделей искусственного интеллекта подбирается наиболее эффективный метод оценки качества вод. Точность предсказаний моделей вычисляется по метрикам качества (accuracy, precision, recall, f1).

Ключевые слова: модели машинного обучения, качество вод, нейронные сети, прогноз, точность предсказаний, метрики качества

Abstract

We consider the problem of assessing the quality of natural waters using artificial intelligence methods: gradient boosting, random forest, logistic regression, the method of *k*-nearest neighbors and a neural network. Depending on the accuracy of predictions of artificial intelligence models, the most effective method for assessing water quality is selected. The accuracy of model predictions is calculated using quality metrics (accuracy, precision, recall, f1).

Keywords: machine learning models, water quality, neural networks, forecast, accuracy of predictions, quality metrics

Введение

На сегодняшний день, создание программ с применением искусственного интеллекта, предназначенных для решения сложных задач является весьма актуальным. В работе, авторами предпринята попытка оценить качество природных вод с применением разных методов искусственного интеллекта: градиентный бустинг (eXtreme Gradient Boosting), случайный лес (Random Forest), логистическая регрессия (Logistic Regression), метод *k*-ближайших соседей (k Nearest Neighbors) и нейронная

сеть. Ранее оценка качества вод проводилась расчетными методами по разным методикам [1-8].

Методы искусственного интеллекта

Для реализации методов искусственного интеллекта использовали язык Python и импорт необходимых библиотек (scikit-learn и tensorflow). Точность предсказаний моделей оценивали метриками: accuracy, precision, recall, f1 и confusion matrix, которые подробно описаны в [9].

В качестве тестируемой выборки, выбрана главная река Кемеровской области - р.Томь. Фрагмент выборки представлен в таблице 1.

Таблица 1

Фрагмент выборки. Гидрохимический состав реки Томь

Ингредиенты	Концентрация ингредиентов, мг/л						
Аммоний	0,14	0,09	0,03	0,06	0,06	0,13	0,12
Нитриты	0,86	0,49	0,73	0,44	1,46	0,20	0,25
Нитраты	0,007	0,010	0,005	0,006	0,008	0,007	0,007
БПК5	1,30	1,20	2,31	1,80	1,20	2,95	2,10
Взвешенные вещества	2,00	7,60	13,60	8,40	9,20	7,90	4,20
Железо	0,21	0,18	0,04	0,27	0,22	0,11	0,15
Магний	10,33	4,60	0,20	5,20	5,00	5,60	3,50
Нефтепродукты	0,03	0,12	1,31	0,16	0,16	0,53	0,14
ХПК	6,00	3,00	12,10	4,50	9,10	22,40	11,30
Хлориды	9,01	1,00	10,90	0,90	2,90	3,60	0,90
Марганец	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Медь	0,006	0,001	0,006	0,006	0,006	0,018	0,006
Фенолы	0,003	0,002	0,004	0,003	0,002	0,001	0,002
Класс качества	4	3	4	3	5	3	2

Градиентный бустинг (eXtreme Gradient Boosting). Идея градиентного бустинга состоит в построении ансамбля последовательно уточняющих друг друга элементарных моделей. К особенностям реализации бустингового алгоритма в **xgboost** можно отнести использование помимо первой еще и второй производной от функции потерь, наличие встроенной регуляризации, а также возможность задавать пользовательские функции потерь и метрики качества [10]. Метрики качества, полученные при анализе водного объекта: accuracy - 0.80, precision - 0.67, recall - 0.82, F1- 0.70

Случайный лес (Random Forest)

Этот алгоритм составлен из множества решающих деревьев. Качество работы модели проверено на тестовой выборке. Лучший результат работы модели достигается при применении 15 деревьев (n_estimators). Метрики качества, полученные при анализе водного объекта: accuracy - 0.71, precision - 0.64, recall - 0.64, F1- 0.59

Логистическая регрессия (LogisticRegression) применяется для прогнозирования вероятности возникновения некоторого события по значениям множества признаков [11].

Метрики качества, полученные при анализе водного объекта: accuracy - 0.56, precision - 0.47, recall - 0.37, F1- 0.36

Метод k-ближайших соседей (k Nearest Neighbors (kNN)) - это метрический алгоритм для автоматической классификации объектов. Для классификации объект присваивается тому классу, который является наиболее распространённым среди k соседей данного элемента, классы которых уже известны [12]. Алгоритм был применен

к выборкам с большим количеством атрибутов (многомерным). Для этого определили функцию расстояния; классический вариант такой функции — евклидова метрика (metric=euclidean).

Метрики качества, полученные при анализе водного объекта: accuracy - 0.71, precision - 0.61, recall - 0.52, F1 - 0.53

Нейронная сеть (Neural Network)

Экспериментальным путем, подобрана оптимальная архитектура нейронной сети и ее параметры: три полносвязных (Dense) слоя, на первых двух слоях по 64 нейрона и функция активации – relu, на последнем - 6 нейронов и функция активации - softmax. В результате, получена достаточно высокая точность на обучающей (99,91%) и проверочной выборке (97,46%) рис.1.

Рисунок 1. Оценка качества обучения нейронной сети

Метрики качества, полученные при анализе водного объекта: accuracy - 0.85, precision - 0.72, recall - 0.85, F1 - 0.75. Точность предсказаний моделей и результаты их расчетов представлены в таблицах 2-3.

Таблица 2.

Точность предсказаний моделей.

Модели	Accuracy, %	F1, %
Neural Network	85	75
Gradient Boosting	80	70
Random Forest	71	59
LogisticRegression	56	36
k Nearest Neighbors	71	53

Таблица 3.

Результат работы всех моделей.

Класс качества	Правильные ответы	Neural Network	k Nearest Neighbors	Logistic Regression	Random Forest	Gradient Boosting
1	1	1	0	0	1	1
2	329	272	276	144	286	301
3	527	409	283	154	366	404
4	934	852	826	840	760	799
5	229	178	58	5	40	113
6	44	36	26	25	15	40

Заключение

В статье представлен результат работы методов искусственного интеллекта для оценки качества воды: градиентного бустинга, случайного леса, логистической регрессии, метода k-ближайших соседей и нейронной сети. Самая высокая точность (85%) предсказания класса качества воды была достигнута при применении модели нейронной сети. Таким образом, наиболее эффективный метод искусственного интеллекта для оценки качества воды – нейронные сети.

1. Счастливцев Е.Л., Юкина Н.И., Харлампенков И.Е. Информационно-аналитическая система геоэкологического мониторинга водных ресурсов угледобывающего региона // Вестник КузГТУ. Кемерово. 2016. № 2 (114). С. 157-164.
2. Счастливцев Е.Л. «Техногенное воздействие угледобывающих предприятий на окружающую среду (на примере Кузбасса)». Автореферат дис. д.т.н., Барнаул, 2006г.
3. Потапов В.П., Мазикин В.П., Счастливцев Е.Л., Вашлаева Н.Ю. Геоэкология угледобывающих районов Кузбасса. - Новосибирск: Наука, 2005. 660с.
4. Ковалев В.А. Моделирование геоэкологических систем угледобывающих районов / В.А. Ковалев, В.П. Потапов, Е.Л. Счастливцев, Ю.И. Шокин; отв. ред. А.М. Федотов; Рос. акад. наук, Сиб. отд-ние, ин-т выч. технологий; Ин-т водн. и экол. проблем; М-во образования и науки, КузГТУ. – Новосибирск: Изд-во СО РАН, 2015. 298 с.
5. Логов А.Б., Опарин В.Н., Потапов В.П., Счастливцев Е.Л., Юкина Н.И. Энтропийный метод анализа состава техногенных вод горнодобывающего региона // Физико-технические проблемы разработки полезных ископаемых. Издательство: Издательство Сибирского отделения РАН (Новосибирск) ISSN: 0015-3273, 2015, № 1, С.168-179.
6. Логов А.Б., Замараев Р.Ю., Логов А.А. Анализ состояния систем уникальных объектов. – Вычислительные технологии, Том 10. №5, 2005. С.49-53.
7. Логов А.Б., Замараев Р.Ю., Логов А.А. Алгоритмы энтропийного метода анализа для отображения свойств объекта в фазовом пространстве // Вычислительные технологии, 2005. т. 10. - №6. С.75-81.
8. РД 52.24.643-2002. Руководящий документ. Методические указания. Метод комплексной оценки степени загрязненности поверхностных вод по гидрохимическим показателям" (утв. и введен в действие Росгидрометом 03.12.2002).
9. Метрики в задачах машинного обучения. <https://habr.com/ru/company/ods/blog/328372> (дата обращения 08.11.2019).
10. Xgboost. Теория и практика. <http://biostatr.blogspot.com/2016/08/xgboost.html> (дата обращения 08.11.2019).
11. Логистическая регрессия. <https://ru.wikipedia.org> (дата обращения 08.11.2019).
12. Метод k-ближайших соседей. <https://ru.wikipedia.org> (дата обращения 08.11.2019).

Щербаков Д.В., Козлов В.А.

Теория игр в искусственном интеллекте

АГПУ

(Россия, Армавир)

doi: 10.18411/sr-10-06-2020-12

idsp: sciencerrussia-10-06-2020-12

Аннотация

В данной статье представлено применение теории игр в искусственном интеллекте при работе с играми, содержащими полную и неполную информацию. В качестве примера рассматриваются генеративно-состязательная сеть и программа Libratus.

Ключевые слова: теория игр, искусственный интеллект, генеративно-состязательная сеть, Libratus, полная информация, неполная информация.

Теория игр в GAN

Теория игр, с точки зрения ИИ, в основном помогает в принятии решений. Это не очень сложно, учитывая тот факт, что «Рациональность» является основой теории