

ISSN: 2587-6309
DOI: 10.47501/ITNOU

ИТНОУ

2021. № 2 (18)



**информационные
технологии
в науке, образовании
и управлении**

Свидетельство о регистрации: ПИ № ФС 77 – 68753

Учредитель: Глориозов Евгений Леонидович

Правообладатель: ООО «Институт информационных технологий»

Главный редактор

Глориозов Е.Л., профессор, доктор технических наук

Заместители главного редактора

Бородин В.А., доктор технических наук, член-корреспондент РАН

Журавлёв В.З., руководитель издательства, ответственный редактор

Топорков В.В., доктор технических наук, профессор, МЭИ

Члены редколлегии

Горбунов В.Г., начальник Специального конструкторского бюро, ФГУП ЭЗАН

Кравец А.Г., доктор технических наук, профессор, ВолГУ

Никонов В.Г., доктор технических наук, профессор, член президиума РАЕН

Подиновский В.В., доктор технических наук, профессор, ВШЭ

Рудакова Г.М., кандидат физико-математических наук, ИВМ СО РАН, СибГАУ

Цыганов В.В., доктор технических наук, профессор, ИПУ РАН

Черемисина Е.Н., доктор технических наук, профессор, Международный университет "Дубна"

Шабров О.Ф., доктор политических наук, профессор, РАНХиГС

Шабалина О.А., доктор технических наук, доцент, ВолГУ

Все права на материалы, опубликованные в журнале ИТНОУ, принадлежат Издательству. Не разрешается использование публикаций в журнале в коммерческих целях (ст. 1304 ГК РФ). При использовании материалов в научных и образовательных целях ссылка на источник обязательна.

Пример ссылки на публикацию в журнале:

Фамилия И.О. Название статьи // ИТНОУ: Информационные технологии в науке, образовании и управлении. 2021. № 2. С. XX-XX. –DOI: 10.47501/ITNOU.2021.2.XX-XX где XX - номера страниц статьи

СОДЕРЖАНИЕ

МЕТОДЫ СТРАТЕГИЧЕСКОГО УПРАВЛЕНИЯ ТРАНСПОРТНОЙ ИНФРАСТРУКТУРОЙ СИБИРИ, ДАЛЬНЕГО ВОСТОКА И РОССИЙСКОЙ АРКТИКИ <i>Цыганов В.В.</i>	3
ИННОВАЦИОННЫЕ РЕШЕНИЯ И РАЗРАБОТКИ АО «АТЛАНТИКТРАНСГАЗСИСТЕМА» В ОБЛАСТИ АВТОМАТИЗАЦИИ ДЛЯ ГАЗОВОЙ И НЕФТЕГАЗОВОЙ ПРОМЫШЛЕННОСТИ <i>Бернер Л.И., Рощин А.В., Ковалёв А.А., Зельдин Ю.М., Лавров С.А.</i>	8
МАШИННОЕ ОБУЧЕНИЕ ДЛЯ РЕШЕНИЯ ЗАДАЧ РАСПОЗНАВАНИЯ ОБРАЗОВ <i>Киселева Т.В., Маслова Е.В., Бычков А.Г.</i>	19
ЦИАНОБАКТЕРИИ В ДЕТОКСИКАЦИИ ВОДНОЙ СРЕДЫ ОТ РЯДА ЗАГРЯЗНИТЕЛЕЙ <i>Саванина Я.В.</i>	25
ТЕХНОЛОГИЯ СЦЕНАРНОГО МОДЕЛИРОВАНИЯ ЭВОЛЮЦИИ КРУПНОМАСШТАБНОГО ТРАНСПОРТНОГО КОМПЛЕКСА <i>Цыганов В.В., Савушкин С.А., Бородин В.А.</i>	31
МЕТОДЫ ОБОСНОВАНИЯ ПРОЕКТА ПРИМОРСКОГО ТРАНСПОРТНО - ЛОГИСТИЧЕСКОГО КЛАСТЕРА <i>Цыганов В.В., Савушкин С.А., Горбунов В.Г.</i>	37
МОДЕЛИРОВАНИЕ ДВУХКАНАЛЬНОГО КОГНИТИВНОГО МЕХАНИЗМА УПРАВЛЕНИЯ НА ТРАНСПОРТЕ <i>Еналеев А.К.</i>	42
СТАТИСТИЧЕСКИЙ АНАЛИЗ ЗАБОЛЕВАЕМОСТИ МЕЛАНОМОЙ КОЖИ С УЧЁТОМ ФЛУКТУАЦИЙ: МЕТОДИКА, РЕЗУЛЬТАТЫ АПРОБАЦИИ <i>Неретин Е.Ю., Акулов В.А., Минаев Ю.Л.</i>	47

Сведения об авторах

Бернер Леонид Исаакович

доктор техн. наук, профессор
генеральный директор
АО «АтлантикТрансгазСистема»
Москва, Россия
Эл. почта: berner@atgs.ru

Рошин Алексей Владиславович

канд. техн. наук
первый заместитель генерального директора
по производству
АО «АтлантикТрансгазСистема»
Москва, Россия
Эл. почта: roschin@atgs.ru

Ковалёв Андрей Александрович

канд. техн. наук
заместитель генерального директора по развитию
АО «АтлантикТрансгазСистема»
Москва, Россия
Эл. почта: kovalev@atgs.ru

Зельдин Юрий Маркович

канд. техн. наук
заведующий отд. ПУС
АО «АтлантикТрансгазСистема»
Москва, Россия
Эл. почта: zeldin@atgs.ru

Лавров Сергей Анатольевич

канд. техн. наук
заведующий отделом АСУТП
АО «АтлантикТрансгазСистема»
Москва, Россия
Эл. почта: lavrov@atgs.ru

Information about authors

Berner Leonid

Doctor of Science (Tech.), Professor
General Director
«AtlanticTransgasSystem» JSC
Moscow, Russian Federation
E-mail: berner@atgs.ru

Roshchin Alexey

PhD Tech.
First Deputy General Director, ATGS, «Atlantic-
TransgasSystem» JSC
Moscow, Russian Federation
E-mail: rav@atgs.ru

Kovalev Andrey

PhD Tech.
Deputy General Director for Development
«AtlanticTransgasSystem» JSC
Moscow, Russian Federation
E-mail: kovalev@atgs.ru

Zeldin Yuriy

PhD Tech.
Head of ICS Department
«AtlanticTransgasSystem» JSC
Moscow, Russian Federation
E-mail: zeldin@atgs.ru

Lavrov Sergey

PhD Tech.
Head of SCADA Department
«AtlanticTransgasSystem» JSC
Moscow, Russian Federation
E-mail: lavrov@atgs.ru

УДК 004.4

ГРНТИ 28.23.15

DOI: 10.47501/ITNOU.2021.2.19-24

Т.В. Киселева, Е.В. Маслова, А.Г. Бычков

Сибирский государственный индустриальный университет

МАШИННОЕ ОБУЧЕНИЕ ДЛЯ РЕШЕНИЯ ЗАДАЧ РАСПОЗНАВАНИЯ ОБРАЗОВ

В статье рассмотрен вопрос актуальности применения машинного обучения для решения задач из различных сфер деятельности, в частности, для решения задачи распознавания образов. Описан метод обучения с учителем. Приведен в качестве примера метод *k*-ближайших соседей, а также рассмотрен способ оценки точности работы алгоритма в случае бинарной и мультиклассовой классификации. Приведена базовая структура нейрона.

Ключевые слова: машинное обучение, распознавание образов, нейронные сети, метод *k*-ближайших соседей, обучение с учителем.

Kiseleva T., Maslova E., Bychkov A.
Siberian State Industrial University

MACHINE LEARNING TO SOLVE THE PATTERN RECOGNITION PROBLEM

The article considers the issue of the relevance of using machine learning for solving problems from various fields of activity, in particular, for solving the problem of pattern recognition. The method of teaching with a teacher is described. The method of k-nearest neighbors is given as an example, and a method for assessing the accuracy of the algorithm in the case of binary and multiclass classification is considered. The basic structure of a neuron is shown.

Keywords: machine learning, pattern recognition, neural networks, k-nearest neighbors method, supervised learning.

В случае современного внедрения информационных систем объем данных, с помощью которых они обрабатываются, увеличивается. Для упрощения работы с ними необходимы инструменты, обладающие большой скоростью сбора, обработки и анализа. Такие средства называют *Big Data*.

В основе этой технологии лежит машинное обучение (англ. *machine learning, ML*) – класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а на основе поиска закономерностей в исходных данных. То есть для решения устанавливается зависимость полученных результатов от исходного набора признаков и их значений.

Сейчас машинное обучение повсеместно применяется для решения задачи распознавания изображений с использованием искусственных нейросетей. Любую нейросеть необходимо обучить на обучающих данных перед применением.

Как решается задача машинного обучения? На входе есть некоторый набор объектов, который переводится в цифровой вектор. Далее подбирается метод машинного обучения для его обработки, и на выходе получается модель, которая должна предсказывать результат на новых данных, которых не было в исходном наборе. Оценка качества обучения модели производится по тому, насколько успешно

будут распознаны ранее неизвестные объекты. Такое обучение называется «обучение с учителем» или «*supervised learning*» и является одним из самых популярных способов организации работы по распознаванию образов.

Для распознавания образов используются различные алгоритмы машинного обучения. Например для классификации изображений возможно использование метода ближайших соседей. Если перевести каждое изображение в многомерное пространство, то разбитые по признакам однородные изображения будут располагаться относительно близко (рисунок 1).

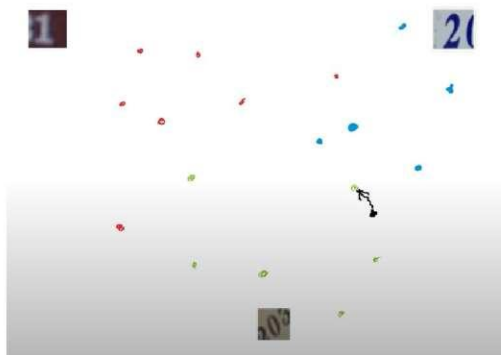


Рис. 1 Распределение однотипных изображений в многомерном пространстве для метода ближайшего соседа

Использование метода строится на предположении, что новая точка при появлении в этом пространстве будет располагаться близко к другим точкам того же типа. То есть, если на новом изображении будет ноль, то оно будет располагаться в этом пространстве близко к другим подобным изображениям. Следовательно, можно найти ближайшую точку по отношению к новой и выдвинуть предположение о классе, к которому принадлежит новое изображение.

Уточним, что означает в данном случае «ближайший». Для определения расстояния между точками обычно используются следующие формулы:

$$L_2 = \sqrt{\sum_i (v_i - u_i)^2}, \quad (1)$$

$$L_1 = \sum_i |v_i - u_i|, \quad (2)$$

где L_2 – стандартное Евклидово расстояние (длина многомерного отрезка), v_i, u_i – координаты двух точек, L_1 – сумма разниц между координатами.

Здесь крайне важна точность на обучающей и контрольной выборках. Точность алгоритмов определяется по тому, насколько он хорошо работает на ранее неизвестных данных. Поэтому была разработана модификация метода ближайшего соседа в виде метода k -ближайших соседей. Этот метод показывает части себя чуть хуже на обучающей выборке, но лучше на контрольной. Идея состоит в том, чтобы использовать не одного ближайшего соседа, а нескольких, число которых определяется гиперпараметром k .

Производится выбор нескольких ближайших соседей, и рассчитывается, какой класс выражен сильнее из этой выборки. Для каждой задачи параметр k подбирается индивидуально.

Для оценки точности работы алгоритма исходный набор разбивается на три части – тренировочный, валидационный и тестовый. Валидационный набор данных используется как средство подбора гиперпараметра k . После работы тренировочного и валидационного наборов модель проверяется на тестовом наборе данных.

Для оценки точности используются различные метрики. Выбор каждого конкретного метода оценки точности зависит от специфики задачи.

В случае бинарной классификации, когда всего два класса (пример: есть дефект или нет дефекта), можно использовать простейший вариант с расчётом точности по формуле (3).

$$Accuracy = \frac{correct}{total}. \quad (3)$$

$Accuracy$ – точность, $correct$ – правильные варианты ответов, $total$ – все попытки отнести объект к одному из двух классов. У этого подхода есть два недостатка.

Первый заключается в том, что с этим подходом тяжело работать, когда наборы данных не сбалансированы. В случае с наличием дефектов может возникнуть ситуация, когда дефекты редки в исходном наборе данных, например, их всего 1%. Тогда самый

простейший классификатор будет точен на 99%. Но при этом на практике количество брака будет намного больше.

Второй недостаток состоит в том, что ошибки могут быть разного рода. То есть, «стоимость» того, что дефект был пропущен, для предприятия будет гораздо выше, чем при ложном срабатывании.

Поэтому в таких ситуациях используют метрики *Precision* и *Recall*.

$$Precision = \frac{TP}{TP + FP}, \quad (4)$$

$$Recall = \frac{TP}{TN + FN}, \quad (5)$$

где *TP* (*True positives*) – это те объекты, которые действительно являются дефектами и которые были выделены моделью как дефекты; *FN* (*False negatives*) – объекты, которые являются дефектами, но которые модель не обнаружила; *TN* (*True negatives*) – объекты, в которых нет проблем и которые модель проигнорировала. *FP* (*False positives*) – объекты, в которых нет проблем, но которые модель отметила как дефектные.

Метрика *Precision* (формула (4)) показывает, насколько выбранные моделью объекты релевантны условию поиска. То есть, если *precision* = 0,8, то это значит, что из 100 найденных моделью объектов только у 80 дефект действительно есть. Метрика *Recall* (формула(5)) показывает, сколько релевантных объектов было выбрано, то есть, сколько было выбрано объектов с дефектом по сравнению со всеми дефектными объектами в целом.

В случае мультиклассовой классификации используются те же способы, они модифицируются.

Точность *Accuracy* рассчитывается также по формуле (3). Для расчёта *Precision* и *Recall* строится матрица ошибок, пример которой приведен на рисунке 2. По горизонтали расположены фактические значения распознаваемых объектов, по вертикали то, что предсказывает алгоритм.

С помощью такой матрицы можно считать *Precision* и *Recall* как для любого класса.

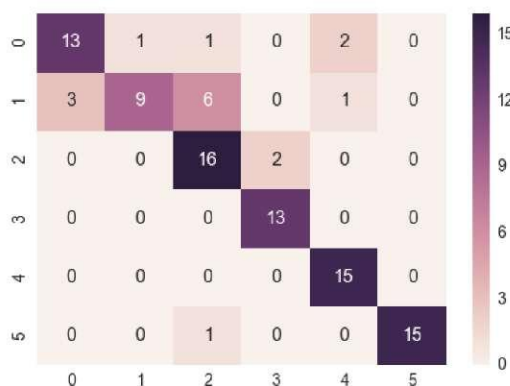


Рис. 2 Матрица ошибок.

Общий подход к машинному обучению с учителем при использовании разделения исходного набора данных на три (тренировочный, валидационный и тестовый) заключается в следующем: сначала производится обучение на тренировочном наборе. Если ошибка большая, это означает, что модель недообучена. Следует уточнить параметры модели, либо использовать другой подход. Если погрешность небольшая, то происходит проверка на валидационном наборе. Если же ошибка большая, значит, что модель переобучена, нужно использовать большее количество исходных данных, либо добавлять дополнительные ограничения. Далее следует проверка на тестовом наборе данных. Если в этом случае получена большая погрешность, то проверке подвергаются исходные данные. Если тренировочный и тестовый наборы отличаются, то следует поменять наборы, так как тренировочный и тестовый наборы данных должны быть близки друг к другу. Если же ошибка на всех трёх проверках была приемлемой в рамках конкретной задачи, то можно такую модель использовать по назначению.

При распознавании образов используются нейронные сети. Далее рассмотрим их подробнее. На рисунке 3 приведена базовая схема нейрона.

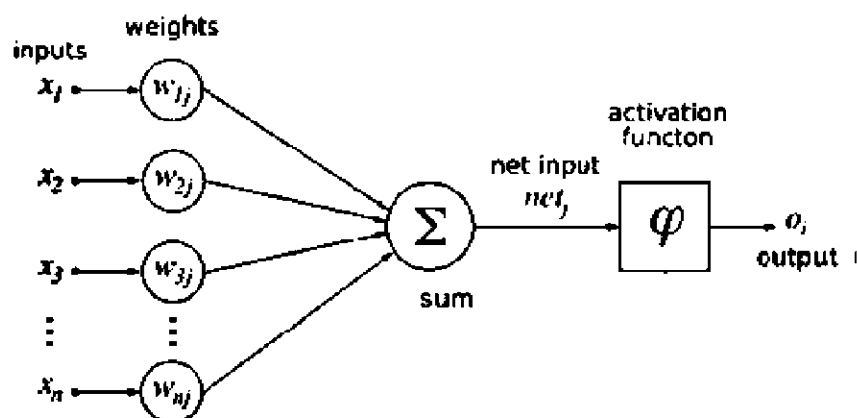


Рис. 3 Пример структуры нейрона.

На вход нейрона подаются некоторые числа от x_1 до x_n . Следует сказать, что нейронные сети работают только с числами. Каждое из этих чисел умножается на соответствующий ему вес w_{ij} , эти значения после этого суммируются, и к сумме применяется функция активации, которая преобразует её в выход o_i . Этот выход передается на входы другим нейронам. Так же, как и природный нейрон получает на вход разные электрические сигналы, а выдаёт только один, нейрон в машинном обучении получает на вход несколько разных чисел, а выдаёт одно число.

Таким образом, выше приведен обзор актуальности применения машинного обучения для распознавания изображений.

Литература:

1. Киселева Т.В. Прогнозирование реализаций показателей, характеризующих здоровье населения с помощью нейронных сетей на примере травматизма / Т.В. Киселева, Н.Н. Максимова, А.В. Дейли // Сборник докладов Международной конференции «Инноватика, 2009». – М: Энергоатомиздат, 2009. С. 77-82.

2. Киселева Т.В. Применение классических и нейросетевого методов расчета вероятности прогноза в многовариантном прогнозаторе / Т.В. Киселева, А.С. Дружилов // Сборник докладов Российской конференции «Системы автоматизации в образовании, науке и производстве». – Новокузнецк, 2011. С. 507-513.

3. Киселева Т.В. Прогнозирование показателей травматизма с помощью нейронных сетей / Т.В. Киселева, Н.Н. Максимова, В.В. Трофимов // Сборник трудов Международной конференции по теории активных систем. – М.: 2009. С. 89-93.

4. Киселева Т.В. Многовариантное прогнозирование потребности населения города Новокузнецка в лекарственных препаратах на базе нейросетевых технологий / Т.В. Киселева, А.С. Дружилов, А.В. Дейли // Сборник докладов Международной научно-технической конференции «Инноватика, 2010». – М: Энергоатомиздат, 2010. с. 118-124.

5. Комашинский В.И. Нейронные сети и их применение в системах управления и связи / В.И. Комашинский. – М.: Горячая линия-Телеком, С. 2002 – 94.

6. Червяков Н.И. Применение нейронных сетей для задач прогнозирования и проблемы идентификации моделей прогнозирования / Н.И. Червяков // Нейрокомпьютеры: разработка и применение, 2003. № 10. С.11-14.

7. Круглов В.В. Искусственные нейронные сети. Теория и практика / В.В. Круглов. – М.: Горячая линия-Телеком, 2001. С. 382.

Сведения об авторах

Тамара Васильевна Киселева

доктор техн. наук, профессор

кафедра прикладных информационных технологий

ФГБОУ ВО «Сибирский государственный индустриальный университет»

г. Новокузнецк, Кемеровская область, Россия

Эл. почта: kis@siu.sibsiu.ru

Елена Владимировна Маслова

кандидат техн. наук

кафедра прикладной математики и информатики

ФГБОУ ВО «Сибирский государственный индустриальный университет»

г. Новокузнецк, Кемеровская область, Россия

Эл. почта: elenamaslova1805@yandex.ru

Александр Григорьевич Бычков

аспирант

кафедра прикладных информационных технологий

ФГБОУ ВО «Сибирский государственный индустриальный университет»

г. Новокузнецк, Кемеровская область, Россия

Эл. почта: aleksds1@yandex.ru

Information about authors

Tamara Vasil'yevna Kiseleva

Doctor of Science (Tech.), Professor

Department of Applied Information Technologies

Siberian State Industrial University

Novokuznetsk, Kemerovo Region, Russia

E-mail: kis@siu.sibsiu.ru

Elena Vladimirovna Maslova

PhD (Tech.)

Department of Applied Mathematics and Informatics

Siberian State Industrial University

Novokuznetsk, Kemerovo Region, Russia

E-mail: elenamaslova1805@yandex.ru

Alexander Grigorievich Bychkov

postgraduate student

Department of Applied Information Technologies

Siberian State Industrial University

Novokuznetsk, Kemerovo Region, Russia

E-mail: aleksds1@yandex.ru