

**Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Сибирский государственный индустриальный университет»
ООО «Объединённая компания Сибшахтострой»
АНО «Научно-образовательный центр «Кузбасс»
ООО «Научно-исследовательский центр систем управления»
Кузбасский научный центр СО АИН им. А.М. Прохорова**

**СИСТЕМЫ АВТОМАТИЗАЦИИ
(в образовании, науке и производстве)
AS' 2021**

**ТРУДЫ XIII ВСЕРОССИЙСКОЙ НАУЧНО-ПРАКТИЧЕСКОЙ
КОНФЕРЕНЦИИ**

(с международным участием)

2 – 3 декабря 2021 г.

**Новокузнецк
2021**

УДК 658.011.56
С 409

Редакционная коллегия

д.т.н., профессор Н.А. Козырев, д.т.н., профессор С.М. Кулаков,
д.т.н., профессор Л.П. Мышляев, к.т.н. О.В. Михайлова,
к.т.н., доцент В.А. Кубарев

С 409 Системы автоматизации (в образовании, науке и производстве) :
AS'2021 : труды XIII Всероссийской научно-практической
конференции (с международным участием) 2 – 3 декабря 2021 г. /
Министерство науки и высшего образования Российской Федерации ,
Сибирский государственный индустриальный университет ; под общей
ред. : С. М. Кулакова, Л. П. Мышляева. – Новокузнецк : Издательский
центр СибГИУ, 2021. - 417 с. : ил.

ISBN 978-5-7806-0583-6

Труды конференции посвящены научным и практическим вопро-сам автоматизации управления технологическими процессами и пред-приятиями, социально-экономическими системами, образованием и ис-следованиями. Представлены результаты исследования, разработки и внедрения методического, математического, программного, техниче-ского и организационного обеспечения систем автоматизации и инфор-мационно-управляющих систем в различных сферах деятельности.

Сборник трудов ориентирован на широкий круг исследователей, научных работников, инженерно-технический персонал предприятий и научно-исследовательских лабораторий, преподавателей вузов, аспиран-тов и студентов.

УДК 658.011.56

ОРГАНИЗАТОРЫ И ПАРТНЕРЫ КОНФЕРЕНЦИИ

ФГБОУ ВО «Сибирский государственный индустриальный университет»
ООО «Объединённая компания «Сибшахтострой» (г. Новокузнецк)
ООО «Научно-исследовательский центр систем управления» (г. Новокузнецк)
ООО «ЕвразТехника» (г. Новокузнецк)
АНО «Научно-образовательный центр «Кузбасс» (г. Кемерово)
Журналы: «Известия вузов. Чёрная металлургия» (г. Москва, Новокузнецк),
«Вестник СибГИУ» (г. Новокузнецк)

© Сибирский государственный
индустриальный университет, 2021

По таблице видно, как мелкие зерна сильно влияют на статистику. У выделенного зерна площадь $184 \mu\text{m}^2$, но из-за двух зёрен с площадью $5 \mu\text{m}^2$, средняя площадь снизилась до $64,7 \mu\text{m}^2$. Можно для таких случаев ввести в программу настройку минимальной площади зёрен, учитываемых в статистике.

Скорость обработки данных с использованием графического процессора увеличилась приблизительно в 80 – 100 раз по сравнению с центральным процессором.

Заключение

В общих чертах был изучен метод дифракции отражённых электронов (EBSD), рассмотрен фреймворк для работы с вычислительными мощностями графического процессора (OpenCL) и представлены особенности работы с ним.

В процессе работы создано программное обеспечение, позволяющее производить анализ и частично ускоренную обработку EBSD данных.

Библиографический список

1. В.Н. Варюхин, Е.Г. Пашинская, А.В. Завдоев, В.В. Бруховецкий Возможности метода дифракции обратнорассеянных электронов для анализа структуры деформированных материалов. Национальная академия наук Украины, Донецкий физико-технический институт им. А.А. Галкина 2014. 101с.
2. Open standard for parallel programming of heterogeneous systems [Электронный ресурс] – URL: <https://www.khronos.org/opencv/> (дата обращения: 10.10.2021).
3. Generating Orientation Maps to Present EBSD Data [Электронный ресурс] – URL: <https://www.azom.com/article.aspx?ArticleID=11775> (дата обращения: 10.10.2021).

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ОБРАЗОВ

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

*Сибирский государственный индустриальный университет,
Новокузнецк, Россия, kis@siu.sibsiu.ru*

Аннотация. В докладе 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.

Abstract. 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 современных условиях широкого применения информационных систем объем данных, с которыми они работают, расширяется в геометрической прогрессии и для эффективной работы с ними необходимы инструменты, которые могли бы собирать и обра-

батывать эти данные с большой скоростью, анализировать их, находя различные закономерности для облегчения их классификации и прогнозирования. Такие средства все чаще называют Big Data.

В основе технологии Big Data 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 применяется не конкретная четкая формула, устанавливается зависимость полученных результатов от исходного набора признаков и их значений.

В последнее время одной из наиболее популярных задач для решения с применением методов машинного обучения являются задачи распознавания изображений. Для этого используются искусственные нейронные сети, но прежде, чем распознавать то или иное изображение, нейронная сеть должна быть обучена на обучающих данных.

Каким образом решается любая задача машинного обучения? Задается некоторый набор объектов, который переводится в цифровой вектор. Далее подбирается метод машинного обучения для его обработки, и на выходе получается модель, которая сможет 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 распространенных способов организации работы по распознаванию образов.

Для распознавания образов можно использовать разные алгоритмы машинного обучения. К примеру, для классификации изображений может быть использован метод ближайших соседей. Если перевести каждое изображение в многомерное пространство, то разбитые по признакам однородные изображения будут располагаться относительно близко (рисунок 1).

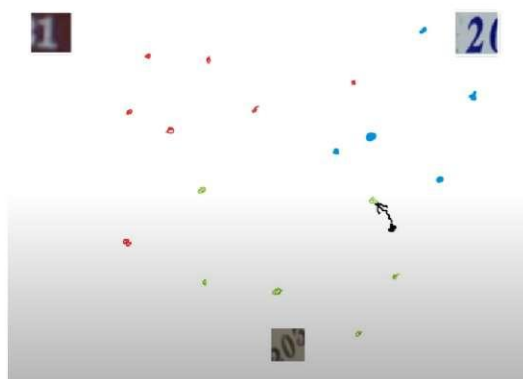


Рисунок 1 – Распределение однотипных изображений в многомерном пространстве для метода ближайшего соседа

Предполагается, 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. Поэтому можно найти ближайшую точку по отношению к новой и сделать предположение, к какому классу принадлежит новое изображение.

Следует уточнить, что такое «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 $L2$ – стандартное Евклидово расстояние (длина многомерного отрезка), а $L1$ – сумма разниц между координатами.

В этом случае важна точность на обучающей и контрольной выборках. Точность любого алгоритма определяется по тому, насколько он хорошо работает на новых, 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 .

Выбираются несколько ближайших соседей, и производится расчёт того, какой класс выражен сильнее всего из данной выборки. Для каждой задачи параметр k подбирается индивидуально. В зависимости от этого параметра может возникать ситуация недообучения или переобучения (*underfitting* или *overfitting*).

Недообученная модель не даёт высокой точности на реальных данных из-за того, что она недостаточно сложна. В случае переобучения граница между классами слишком сложна, она не генерализована, вместо обобщения модель запоминает случайные шумы, модель находит закономерности, которых нет, что также приводит к снижению точности на новых данных. Поэтому вопрос выбора гиперпараметров является одним из основных на практике, для его решения можно привлекать экспертов.

Для оценки точности работы алгоритма можно разбить исходный набор 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.

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

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

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 $\text{precision} = 0,8$, 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.

Такой подход лучше работает в несбалансированных наборах данных. Управляя тем, какой параметр более важен, можно приоритезировать конкретный тип ошибки, уменьшая в первую очередь его.

В случае мультиклассовой классификации используются те же способы, но с небольшими изменениями.

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.

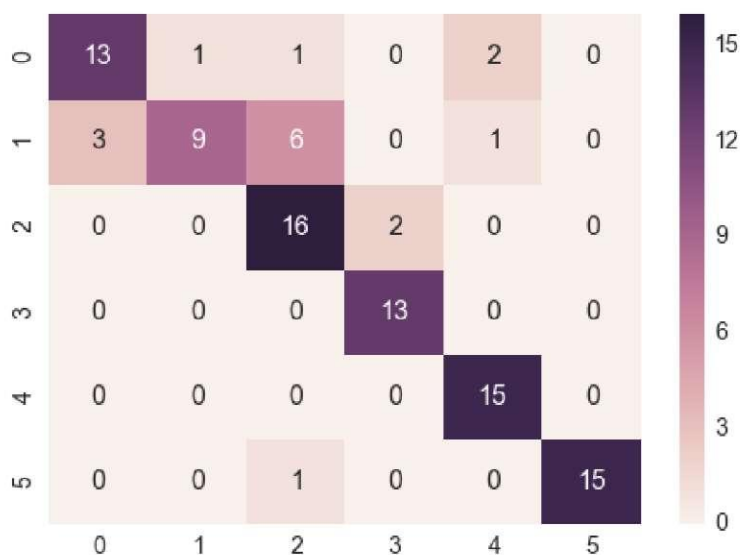


Рисунок 2 – Матрица ошибок

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

Общий подход к машинному обучению с учителем при использовании разделения исходного набора данных на три (тренировочный, валидационный и тестовый) выразил корейский программист и автор книг по машинному обучению Эндрю Бин (Andrew Ng). Сначала производится обучение на тренировочном наборе. Если ошибка большая, то это недообученная модель. Следует уточнить параметры модели, либо использовать другой подход. Если погрешность небольшая, то происходит проверка на валидационном наборе. Если ошибка большая, значит, модель переобучена. В данном случае 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, как уже было сказано выше, используются нейронные сети. Рассмотрим их подробнее. Базовая схема нейрона приведена на рисунке 3.

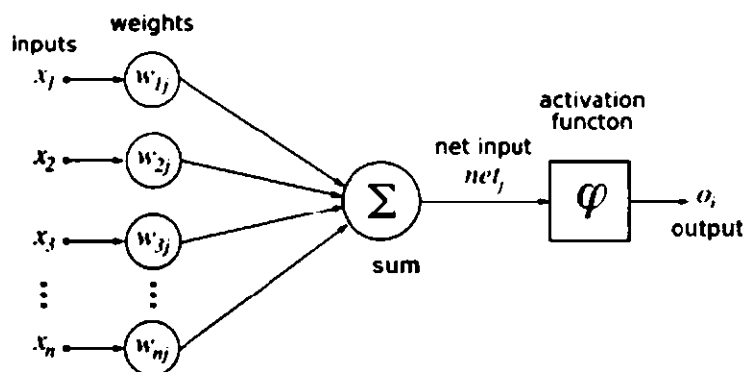


Рисунок 3 – Пример структуры нейрона

На вход нейрона подаются некоторые числа от x_1 до x_n . Важно отметить, что нейронные сети всегда работают только с числами. Каждое из этих чисел умножается на некий соответствующий ему вес w_{ij} . Далее полученные значения суммируются. После чего к этой сумме применяется функция активации, которая преобразует её в выход o_i . Этот выход может пойти 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, рассмотрена актуальная задача применения машинного обучения, приведено определение машинного обучения в целом и «обучения с учителем» в частности. Дано понятие точности, описаны алгоритмы расчёта точности для бинарного и мультиклассового классификатора. Описана базовая методика повышения точности с разделением исходного набора данных на тренировочный, валидационный и тестовый. Приведен пример базовой структуры нейрона.

Библиографический список

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. – 77-82 с.
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. – 507-513 с.
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. – 89-93 с.
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. – 118-124 с.
5. Комашинский В.И. Нейронные сети и их применение в системах управления и связи / В.И. Комашинский. – М.: Горячая линия-Телеком, 2002 – 94 с.
6. Червяков Н.И. Применение нейронных сетей для задач прогнозирования и проблемы идентификации моделей прогнозирования / Н.И. Червяков // Нейрокомпьютеры: разработка и применение, 2003. – № 10. – 11-14 с.
7. Круглов В.В. Искусственные нейронные сети. Теория и практика / В.В. Круглов. – М.: Горячая линия-Телеком, 2001 – 382 с.

РАЗРАБОТКА АВТОМАТИЧЕСКОЙ СИСТЕМЫ НАВЕДЕНИЯ НА КОСМИЧЕСКИЕ ОБЪЕКТЫ	331
Казанцев М.Е., Попов А.С., Саламатин А.С.	
ОБНАРУЖЕНИЕ СУЩЕСТВЕННЫХ ПЕРЕМЕННЫХ МЕТОДАМИ НЕПАРАМЕТРИЧЕСКОЙ СТАТИСТИКИ	334
Караванов А.В., Кириченко В.Н., Михов Е.Д.	
ИССЛЕДОВАНИЕ АЛЬТЕРНАТИВНЫХ СПОСОБОВ ПОСТРОЕНИЯ API НА ПРИМЕРЕ СОЗДАНИЯ ИНФОРМАЦИОННОЙ СИСТЕМЫ УПРАВЛЕНИЯ НОМЕНКЛАТУРОЙ ТОВАРОВ С ПОМОЩЬЮ GraphQL	339
Стрелков А.В., Истомин А.С.	
РАЗРАБОТКА ПРОГРАММЫ ДЛЯ ВОССТАНОВЛЕНИЯ МИКРОСТРУКТУРЫ ИСХОДНОГО АУСТЕНИТА ЗЕРНА С ИСПОЛЬЗОВАНИЕМ ДАННЫХ EBSD	343
Демьяненко Е.О., Истомин А.С., Карабаналов М.С., Корниенко О.Ю.	
ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ОБРАЗОВ	349
Киселева Т.В., Маслова Е.В., Бычков А.Г.	
МОДЕЛЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА ОСНОВЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ЦИФРОВОГО РАСПОЗНАВАНИЯ РУКОПИСНЫХ ЦИФР	354
Стародубов А.Н. ^{1,2} , Пылов П.А. ²	
РЕАКТОРЫ НА БЫСТРЫХ НЕЙТРОНАХ. ПЕРСПЕКТИВЫ РАЗВИТИЯ И ИХ РОЛЬ В СТАНОВЛЕНИИ АТОМНОЙ ЭНЕРГЕТИКИ	358
Гусев С.С.	
АНАЛИЗ И ОБРАБОТКА ДАННЫХ В СИСТЕМЕ «МЕДИАЛОГИЯ ИНЦИДЕНТ»	366
Конюхова Е.С.	
ОЦЕНИВАНИЕ ХИМИЧЕСКОГО СОСТАВА СТАЛИ В АГРЕГАТЕ КОВШ-ПЕЧЬ	369
Гизатулин Р.А.	
ОСОБЕННОСТИ СИТУАЦИОННО-НОРМАТИВНОГО МОДЕЛИРОВАНИЯ ПРОИЗВОДИТЕЛЬНОСТИ И НОРМ ВЫРОБОТКИ (НА ПРИМЕРЕ ОТДЕЛЕНИЯ МЕДНЕНИЯ МЕТИЗНОГО ПРОИЗВОДСТВА)	374
Кулаков С.М., Мусатова А.И.	
ОБЗОР ПРОГРАММ КЛАССА SIEM ДЛЯ РЕШЕНИЯ ЗАДАЧ ПРОЦЕССА УПРАВЛЕНИЯ ИНЦИДЕНТАМИ ЗАЩИТЫ ИНФОРМАЦИИ В БАНКОВСКОЙ СФЕРЕ ДЕЯТЕЛЬНОСТИ	379
Барышникова К.В.	
ВЫБОР ТАКТИКИ ВЕДЕНИЯ БЕРЕМЕННОСТИ НА ОСНОВЕ ПРОГНОЗИРОВАНИЯ РАЗВИТИЯ ВНУТРИУТРОБНОЙ ИНФЕКЦИИ У НОВОРОЖДЁННОГО	384
Власенко А.Е., Григорьева Е.Ю., Ренге Л.В., Лихачева В.В.	