Саратовский национальный исследовательский государственный университет им. Н. Г. Чернышевского

Факультет компьютерных наук и информационных технологий

**Машинное обучение**

Выполнил: студент 1 курса 111 группы

Долесов Валентин Пантелеевич

Проверил: кандидат физико-математических наук,

доцент Машников В. В.

Саратов, 2023

**Содержание**

Введение 3

Глава I. Типы машинного обучения 5

Глава II. Методы машинного обучения 7

Глава III. Применение машинного обучения в разных областях 12

Глава IV. Проблемы машинного обучения 15

Глава V. Перспективы машинного обучения в будущем 16

Заключение 17

Список литературы и источников 18

**Введение**

Машинное обучение – это подход к решению задач, который заключается в разработке и использовании алгоритмов, позволяющих компьютеру самостоятельно извлекать знания из данных, обучаться на основе этих знаний и принимать решения без явного программирования.

История машинного обучения начинается еще в 1950-х годах, когда появились первые искусственные нейронные сети. Однако тогда компьютеры были далеки от того, чтобы обрабатывать большие объемы данных, поэтому разработка алгоритмов машинного обучения не получила широкого распространения.

В 1960-х годах возникли первые системы распознавания речи и образов, но они были очень ограничены в своих возможностях. В 1970-х годах появились первые экспертные системы, которые использовали знания экспертов для принятия решений. Однако такие системы были очень сложными и требовали больших затрат на разработку.

В 1980-х годах стали появляться новые методы машинного обучения, например метод опорных векторов и деревья решений. Они позволяли обрабатывать большие объемы данных и делать более точные прогнозы.

В 1990-х годах с развитием интернета и электронной коммерции стала возрастать потребность в алгоритмах машинного обучения для анализа данных и принятия решений. В этот период появилось множество новых методов машинного обучения, например нейронные сети глубокого обучения и алгоритмы кластеризации.

Сегодня машинное обучение используется во многих областях, таких как медицина, финансы, транспорт и многие другие. Оно позволяет компьютерам извлекать знания из данных и делать прогнозы, которые помогают людям принимать более эффективные решения.

Эта тема является **актуальной**, т. к. в современном мире данные являются ключевым ресурсом для бизнеса и государственных организаций. Машинное обучение позволяет автоматизировать процесс анализа данных и принятия решений на их основе. Кроме того, оно может помочь в решении сложных задач, которые требуют высокой точности и скорости выполнения.

**Цель** реферата: изучение принципов и методологии развития и применения алгоритмов машинного обучения для решения различных задач в науке, технике и других областях жизни.

Для реализации данной цели были поставлены следующие **задачи:**

1. Изучение и систематизация базовых понятий и методологии машинного обучения.
2. Исследование применения машинного обучения в различных областях.
3. Оценка проблем и перспектив развития машинного обучения.

**Материалами** для реферата послужили интернет-ресурсы и учебные пособия по машинному обучению.

**Структура работы** представлена введением, пятью главами, заключением и списком использованной литературы.

**Глава I. Типы машинного обучения**

Существуют три основных типа машинного обучения:

1. **Обучение с учителем.**

Машинное обучение с учителем – это метод обработки данных, при котором компьютер получает обучающую выборку, состоящую из пар "входные данные - правильный ответ". На основе этой выборки компьютер строит математическую модель, которая может использоваться для решения задач, связанных с анализом данных, прогнозированием и классификацией.

Примеры задач, которые можно решить с помощью машинного обучения с учителем, включают в себя: классификацию текстов (например, определение тональности отзыва), прогнозирование цен на акции, определение спама в электронной почте и т.д.

Для успешного обучения модели необходимо правильно подготовить данные, провести их предобработку и выбрать подходящие методы оценки качества модели. Также для применения такого типа обучения необходимо иметь знания в области математики, статистики и программирования.

1. **Обучение без учителя.**

Машинное обучение без учителя – это метод обработки данных, при котором компьютер получает только входные данные без правильных ответов. На основе этих данных компьютер ищет закономерности и шаблоны, которые помогут ему делать выводы и группировать данные.

Например, если компьютеру предоставить данные о покупках в магазине, то он может сгруппировать их по сходству и выделить несколько категорий покупателей, которые покупают похожие товары.

Такой вид обучения может использоваться для анализа текстовых данных, медицинских данных, финансовых отчетов и других типов информации.

1. **Обучение с подкреплением.**

Машинное обучение с подкреплением – это метод обработки данных, при котором компьютер самостоятельно учится принимать решения на основе опыта и награды.

Для этого компьютеру предоставляется среда, в которой он действует. Каждое принятое решение оценивается наградой или штрафом, которые определяются заранее.

На основе этих наград компьютер учится выбирать наиболее оптимальные действия в данной ситуации. Он анализирует свой опыт и корректирует свои действия, чтобы достичь наилучшего результата.

Машинное обучение с подкреплением может использоваться для создания автоматических систем управления, робототехники, игр и других приложений, где необходимо принимать решения на основе опыта и награды.

**Глава II. Методы машинного обучения**

**Методы машинного обучения включают в себя:**

1. **Деревья решений.**

**Дерево решений (Decision tree) – это тип машинного обучения, который используется для решения задач классификации и регрессии. Оно представляет собой структуру в виде дерева, в которой каждый узел – это тест на один из признаков, а каждая ветвь по направлению от узла к другому узлу отражает возможный результат этого теста.**

**Процесс построения дерева решений включает в себя:**

**1. Выбор признака: алгоритм выбирает признак, который будет помещен в корневой узел дерева.**

**2. Разбиение данных: на основе выбранного признака данные разбиваются на две или более группы, которые представляют собой ветви дерева.**

**3. Повторение: процесс повторяется для каждой ветви, пока все данные не будут отнесены к какому-то классу или пока не будет достигнута заданная глубина.**

***Преимущества:* Легкость в интерпретации, возможность обработки разных видов данных, быстрое обучение и предсказание.**

***Недостатки:* Склонность к переобучению, неспособность работать с пропущенными данными или данными с выбросами.**

1. **Нейронные сети.**

**Нейронные сети (Neural Networks) – это тип машинного обучения, моделирующий работу мозга по принципу связей нейронов. Они соединены друг с другом через веса связей, на вход сети подаются данные, а выходным является результат классификации или регрессии.**

**Процесс обучения нейронной сети включает в себя:**

**1. Инициализацию весов: веса связей случайным образом задаются в начале и обновляются в процессе обучения.**

**2. Прямое распространение: данные, передаваемые на вход нейронной сети, проходят через нейроны и с помощью весов связей передаются в следующие слои.**

**3. Обучение: при обучении нейронной сети происходит обновление весов связей на основе разности между предсказанным и ожидаемым результатами.**

**4. Обратное распространение: веса связей обновляются с помощью расчета градиента функции потерь.**

***Преимущества:* Способность работать с большими наборами данных, распознавать сложные закономерности в данных, возможность обобщения и распознавания шаблонов, состоятельность в работе с разными типами данных.**

***Недостатки:* Сложности в интерпретации полученных результатов, долгое время обучения, необходимость большого количества данных для обучения, склонность к переобучению.**

1. **Метод опорных векторов.**

**Метод опорных векторов (Support Vector Machines, SVM) – это метод обучения с учителем для задач классификации или регрессии.**

**Идея метода заключается в том, что гиперплоскость должна иметь максимальное расстояние до ближайших точек каждого класса, которые называются опорными векторами, т.е. она должна максимально разделять точки двух классов.**

**Процесс обучения SVM включает в себя:**

**1. Выбор ядра: ядро определяет, как точки будут отображаться в пространство большей размерности, где они линейно отделимы.**

**2. Оптимизация между классами: находится оптимальная гиперплоскость, которая максимально разделяет точки каждого класса.**

**3. Оптимизация внутри класса: находятся опорные векторы, которые принадлежат разделяющей гиперплоскости и максимизируют расстояние до гиперплоскости.**

***Преимущества:* Хорошая обобщающая способность, возможность выбора различных ядер для преобразования данных, способность работать с большими размерностями данных.**

***Недостатки:* Может стать непрактичным для больших объемов данных из-за больших вычислительных затрат, может столкнуться с проблемами в многоклассовой классификации.**

1. **Алгоритмы кластеризации.**

**Алгоритмы кластеризации в машинном обучении используются для разделения набора данных на группы или кластеры в соответствии с схожестью объектов.**

**Существует множество алгоритмов кластеризации, относящихся к разным типам.**

**Например, алгоритм k-means является наиболее распространенным. Он основывается на минимизации среднеквадратичного отклонения между точками кластера и центром кластера. На первом шаге алгоритма выбираются случайным образом k центров кластеров, а затем точки группируются в ближайшие кластеры на основе расстояния до центров. Центры переставляются до тех пор, пока минимум функции ошибки не будет достигнут.**

**Алгоритм иерархической кластеризации использует другой подход, когда кластеры создаются последовательно путем слияния наиболее близких кластеров, начиная с наименьших. Этот алгоритм может быть иерархическим и восходящим, где каждый кластер начинает с одного объекта и последовательно дополняется элементами более крупных кластеров, или же он может быть иерархическим и нисходящим, где все объекты начинаются на высшем уровне и последовательно сливаются в меньшие кластеры.**

**Другие типы алгоритмов кластеризации включают методы плотности, которые определяют кластеры на основе плотности точек, или же методы моделей распределения, которые стараются описать данные с помощью математических моделей.**

**В зависимости от данных, которые требуется кластеризовать, и задач, которые требуется решить, каждый из этих типов алгоритмов может быть более эффективен в сравнении с другими.**

1. **Байесовский классификатор.**

**Байесовский классификатор — это метод классификации в машинном обучении, который основывается на теории вероятности и статистике.**

**Байесовский классификатор основан на так называемой теореме Байеса, которая позволяет пересчитывать вероятность принадлежности объектов к определенным классам на основе изменившейся информации об этих объектах.**

**Суть метода заключается в следующем: для каждого класса определяются вероятности каждого из возможных значений свойств объекта на базе набора тренировочных данных. Затем по формуле теоремы Байеса находятся вероятности принадлежности новых объектов к каждому из классов. В конечном итоге объект относится к тому классу, для которого вероятность принадлежности максимальна.**

***Преимущества:* Простота, быстрота, высокая точности на больших наборах данных, возможность учитывания потенциально сложных зависимости между свойствами объекта и классами.**

***Недостатки:* Не справляется с определенными типами задач, например, когда объекты имеют сложную внутреннюю структуру или многие свойства коррелируют друг с другом, неточность.**

**Глава III. Применение машинного обучения в разных областях**

Машинное обучение применяется в следующих областях:

1. **Медицина**

Машинное обучение может использоваться для определения диагнозов, прогнозирования эффективности лечения, выявления рисков развития заболеваний, а также для работы с медицинскими изображениями и сигналами.

Одним из наиболее распространенных примеров применения машинного обучения в медицине является диагностирование заболеваний, например, рака.

Также машинное обучение применяется для выявления факторов риска заболеваний, а также для разработки новых лекарств и методов лечения.

Помимо этого, машинное обучение может быть использовано для обработки и анализа медицинских изображений, таких как снимки МРТ или рентгеновские снимки.

1. **Финансы**

Машинное обучение может использоваться для прогнозирования цен на акции и другие финансовые инструменты, оценки кредитных рисков, автоматизации торговых операций и определения оптимальных портфелей инвестиций.

Ярким примером применения машинного обучения в финансах является прогнозирование цен на акции и определение оптимальных портфелей инвестиций.

Машинное обучение может использоваться для оценки кредитных рисков.

Кроме того, машинное обучение может быть использовано для автоматизации торговых операций.

1. **Транспорт**

Машинное обучение применяется в транспортной сфере для улучшения эффективности и безопасности транспортных средств, оптимизации грузоперевозок и пассажирских перевозок, а также для управления инфраструктурой.

Один из наиболее заметных примеров – это автоматизированные транспортные системы, такие как автоматические поезда и автобусы.

В области логистики и перевозок машинное обучение может использоваться для оптимизации маршрутов и расписаний, что позволяет сократить затраты на топливо и время, а также улучшить удобство и качество обслуживания клиентов.

Возможности машинного обучения можно использовать и для управления инфраструктурой, включая светофоры, дорожные знаки и другие элементы дорожной инфраструктуры.

1. **Интернет-маркетинг**

Машинное обучение играет важную роль в развитии интернет-маркетинга. Оно может использоваться для оптимизации рекламных кампаний, предсказания поведения покупателей, персонализации контента и повышения конверсионного коэффициента.

Например, машинное обучение используется для оптимизации рекламных кампаний в поисковых системах.

Также машинное обучение может быть использовано для персонализации контента и рекламы.

Наконец, машинное обучение может помочь интернет-маркетологам в анализе больших объемов данных, связанных с поведением пользователей, и предоставлять ценные инсайты для улучшения рекламных кампаний и контента.

1. **Разработка игр**

Машинное обучение в разработке игр сегодня является широко распространенной технологией. Оно позволяет разработчикам создавать более умных и сложных компьютерных персонажей, улучшать игровую механику, повышать интерактивность и эмоциональную привлекательность игры.

Один из наиболее распространенных подходов к использованию машинного обучения в разработке игр – это создание интеллектуальных агентов, которые способны самостоятельно обучаться новым поведениям на основе опыта.

Машинное обучение также может быть использовано для создания интеллектуальных систем распределения ресурсов, таких как энергия, кошельки или ресурсы. Системы могут выполнять быстрый анализ игровых сценариев и предлагать оптимальные стратегии для игроков.

В этой сфере машинное обучение чаще всего используется для улучшения игровой механики. Оно позволяет создавать более стильные, интерактивные и увлекательные игры, которые привлекают широкую аудиторию игроков.

**Глава IV. Проблемы машинного обучения**

Несмотря на большой потенциал, машинное обучение имеет свои недостатки:

1. Необходимость большого количества данных. Для того, чтобы модель машинного обучения работала правильно, требуется большое количество данных для обучения, что может быть затруднительным для компаний, которые только начинают внедрять машинное обучение.
2. Алгоритмическая прозрачность. Некоторые алгоритмы машинного обучения могут быть очень сложными и неясными, что может привести к трудностям при поиске ошибок и устранении неполадок.
3. Ограниченность в обучении. Многие модели машинного обучения создаются с использованием предопределенных алгоритмов и ограниченны набором характеристик. Это означает, что модели могут быть недостаточно гибкими для решения новых задач и не могут подстраиваться под изменяющуюся среду.
4. Риск обобщения. Машинное обучение может продемонстрировать невероятные результаты при обучении на наборе данных, но может дать неправильный результат в новой среде или при работе с другими наборами данных.
5. Проблема интеграции. Для того, чтобы машинное обучение полностью внедрить в бизнес, необходимо наладить интеграцию между ключевыми элементами, такими как процессы, базы данных и человеческие ресурсы, что может быть трудно и затратно.

**Глава V. Перспективы машинного обучения в будущем**

Искусственный интеллект является одной из наиболее перспективных технологий на современном этапе. Он используется для решения многих задач, от распознавания образов до автоматизированного управления транспортным потоком. Машинное обучение является одним из основных его направлений.

Одной из основных причин интереса к машинному обучению является его потенциал улучшения качества жизни человека. Например, машинное обучение может помочь в оптимизации процессов в городском планировании, улучшении услуг в области образования и экономики.

Процессы производства также могут быть оптимизированы с помощью машинного обучения. Предсказательные модели машинного обучения позволяют сократить количество ошибок и улучшить качество продукции. Кроме того, машинное обучение может помочь в оптимизации логистических процессов, управлении запасами и решении других задач, связанных с производством.

Таким образом, использование машинного обучения становится все более популярным в различных отраслях экономики и находит все большее применение в решении различных задач. В ближайшем будущем ожидается ускорение развития технологий машинного обучения и расширение областей их применения.

**Заключение**

В результате исследования темы были выявлены основные идеи и принципы данной технологии.

Внедряя машинное обучение в различные сферы деятельности, можно ожидать повышения производительности, снижения затрат и сокращения рисков при принятии решений. Однако, при использовании данной технологии нужно учитывать экономические, социальные и этические аспекты ее применения.

Таким образом, машинное обучение – это перспективная технология, которая должна быть использована с учетом конкретных задач и контекста их решения. Нам предстоит продолжать исследования и повышать свою компетенцию в данной области, чтобы использовать машинное обучение максимально эффективно и безопасно в будущем.

**Список использованных источников**

1. *Бушманов А., Смоляков А.*  Прикладное машинное обучение, 2018 г.
2. *Кашницкий Ю.* Машинное обучение и анализ данных, 2019 г.
3. *Бенджио И., Гудфеллоу Я., Курвилль А.* Глубокое обучение, 2018 г.
4. https://www.kaggle.com/
5. https://www.coursera.org/browse/data-science/machine-learning
6. https://medium.com/tag/machine-learning