

МОДИФИКАЦИЯ АЛГОРИТМА: МЕТОД БЛИЖАЙШЕГО СОСЕДА

***Аннотация.** Статья посвящена описанию алгоритма, который широко используется в машинном обучении для классификации и регрессии. Он основан на принципе, что объекты, которые находятся друг к другу в пространстве признаков, имеют схожие характеристики.*

N.A. Zhilyak, D. Korshikova

Belarusian State Technological University
Minsk, Belarus

MODIFICATION OF THE ALGORITHM: NEAREST NEIGHBOR METHOD

***Abstract.** The article is dedicated to describing an algorithm that is widely used in machine learning for classification and regression. It is based on the principle that objects that are close to each other in feature space share similar characteristics.*

Целью работы является изучение метода ближайшего соседа, определение основных принципов работы метода, определение преимуществ и недостатков метода, модификация метода и проведение необходимых испытаний.

Современные технологии все чаще требуют точного и эффективного анализа данных и метод ближайшего соседа становится важным инструментом в этой области. Этот алгоритм находит широкое применение в задачах классификации и регрессии, включая области, такие как диагностика, финансовый анализ, системы рекомендаций и распознавание образов. Метод ближайшего соседа позволяет быстро находить похожие объекты, что делает его полезным для создания интуитивно понятных интерфейсов, автоматизированного принятия решений и адаптивных систем, реагирующих на поведение пользователей.

Однако, несмотря на свою простоту и эффективность, метод ближайшего соседа сталкивается с определенными проблемами. Одной из основных трудностей является высокая вычислительная сложность, особенно при больших объемах данных, поскольку для каждого нового объекта необходимо вычислять расстояния до всех объектов

обучающей выборки. Это может привести к значительным задержкам в обработке, особенно в реальном времени. Кроме того, алгоритм чувствителен к шумам и выбросам в данных, что может негативно сказаться на точности предсказаний.

Также существует необходимость в тщательном выборе параметра k , который определяет количество ближайших соседей. Неправильный выбор может привести к переобучению или недообучению модели, что делает адаптацию метода к различным условиям эксплуатации актуальной задачей. Все это подчеркивает важность разработки новых подходов и оптимизаций для метода ближайшего соседа, направленных на преодоление существующих ограничений и повышение его эффективности в современных приложениях.

Метод ближайшего соседа является универсальным инструментом, который находит применение в самых разных сферах, улучшая процессы анализа и принятия решений.

Метод ближайшего соседа работает по следующим основным принципам.

Начинается с выбора параметра K , который представляет собой количество ближайших соседей, учитываемых при классификации или регрессии. Обычно значение K выбирается нечетным, что позволяет избежать ситуаций, когда происходит равное голосование между классами.

Затем происходит расчет расстояний. Для каждого нового объекта или точки вычисляется расстояние до всех объектов в обучающем наборе данных. Расстояния могут быть рассчитаны с использованием различных метрик, таких как евклидово расстояние, манхэттенское расстояние или расстояние Минковского, в зависимости от специфики задачи.

На этапе классификации алгоритм находит K -ближайших соседей и определяет класс нового объекта на основе большинства классов его соседей. Например, если среди K соседей 3 объекта относятся к классу А и 2 – к классу В, новый объект будет отнесен к классу А.

В случае регрессии предсказание выполняется путем усреднения значений целевой переменной K ближайших соседей. Таким образом, метод ближайшего соседа использует простые, но эффективные принципы для анализа данных и принятия решений, что делает его популярным инструментом в различных областях.

Метод ближайшего соседа обладает рядом преимуществ и недостатков, которые следует учитывать при его использовании.

Преимущества метода.

Одним из основных достоинств метода является его простота. Алгоритм легко понять и реализовать, не требуя сложной настройки или обучения модели. Это делает его отличным выбором для начинающих специалистов и тех, кто только начинает изучать машинное обучение.

Кроме того, метод отличается гибкостью: он может использоваться как для классификации, так и для регрессии. Также метод хорошо справляется с многомерными данными, что расширяет его область применения.

Еще одним преимуществом является отсутствие предположений о распределении данных. Это делает метод универсальным инструментом для работы с различными типами наборов данных, в отличие от некоторых других алгоритмов, которые требуют строгих предпосылок.

Недостатки метода.

Среди недостатков метода стоит отметить выбор параметра K . Оптимальное значение этого параметра может значительно влиять на производительность алгоритма. Слишком малое значение K может привести к чувствительности к шуму, в то время как слишком большое значение может скрыть локальные структуры данных и уменьшить точность предсказаний.

Другой важный недостаток – вычислительная сложность. Для больших наборов данных метода может оказаться неэффективным, так как требует вычисления расстояний до всех обучающих объектов. Это может потребовать значительных временных и вычислительных ресурсов.

Наконец, метод чувствителен к шкалированию признаков. Если данные имеют разные масштабы, это может негативно сказаться на результатах. Поэтому перед применением метода важно нормализовать или стандартизировать данные, чтобы обеспечить корректность вычислений.

Модификация метода должны значительно повысить как точность, так и читаемость реализации классификации метода, делая её более эффективной и удобной для анализа.

Во-первых, было увеличено значение K с 3 до 5, что позволяет достичь лучшей стабильности классификации. Это изменение помогает снизить влияние шумов и выбросов, обеспечивая более надежные результаты.

Во-вторых, была внедрена стратифицированная выборка, что обеспечивает пропорциональное представительство классов как в

обучающих, так и в тестовых наборах. Это позволяет сохранить соотношение классов и улучшает обобщающую способность модели.

Кроме того, были произведены визуальные улучшения. Размер точек рассеяния был увеличен для лучшей видимости, а также добавлены цвета границ точек, чтобы обозначить их более четко. Заголовки и подписи осей были улучшены с увеличением размера шрифта, что делает графики более читаемыми. Включение цветовой шкалы помогает пояснить, что представляют собой используемые цвета, улучшая понимание визуализируемых данных.

Сделанные улучшения включают несколько важных изменений, направленных на повышение стабильности и читаемости реализации классификации метода.

Метод ближайшего соседа – это мощный инструмент в арсенале методов машинного обучения. Его простота и универсальность делают его подходящим для разнообразных задач, однако важно помнить о его ограничениях. Правильный выбор параметров и предварительная обработка данных могут значительно повысить эффективность этого метода.

Список использованных источников

4. Shakhbanov.org: Подробное описание метода k ближайших соседей (k-NN), включая алгоритм обучения и примеры применения.

5. DeepMachineLearning.ru: Описание метода K ближайших соседей с акцентом на идею метода и его применение для классификации и регрессии.

6. Publications.hse.ru: Сравнительный анализ структур данных для приближенного поиска ближайшего соседа.

УДК 005.591.6

Е.Е. Косенко
ЮФУ
Таганрог, Россия

СОВЕРШЕНСТВОВАНИЕ МЕХАНИЗМОВ РАСПРЕДЕЛЕНИЯ МАТЕРИАЛЬНЫХ ПОТОКОВ В УСЛОВИЯХ ДИНАМИЧНОЙ СРЕДЫ

Аннотация. Исследование посвящено преодолению ограничений классических транспортных моделей в условиях современной сложной логистики.