

2020. – Вып. 2 (30). – DOI: 10.15393/j5.art.2020.5689.

2. Иванова И.Ю. Реализация Концепции развития математического образования в деятельности образовательных организаций: монография / И.Ю.Иванова, А.Д.Нахман. – «Инновации в образовании». Специальный выпуск. – Издательская платформа Российской академии естествознания. – 2016. – 84 с.

УДК 004.93

И.И. Буторин, маг.; А.Н. Колесенков, проф.
(РГРТУ, г. Рязань, Россия)

МЕТОД ГЛАВНЫХ КОМПОНЕНТОВ (PCA): ПРЕИМУЩЕСТВА, ОГРАНИЧЕНИЯ И ПРИМЕНЕНИЕ В ОБРАБОТКЕ ИЗОБРАЖЕНИЙ

Введение. В докладе рассмотрены основные концепции работы метода PCA, его плюсы, минусы и области применения. Метод главных компонентов представляет собой технику уменьшения размерности данных путем преобразования первоначального набора признаков в новый набор, содержащий меньшее число независимых переменных, имеющих главными компонентами. Основополагающая идея заключается в выявлении линейных комбинаций исходных признаков, которые сохраняют основную часть изменчивости в данных.

Этапы работы алгоритма PCA:

1. Стандартизация: Исходные данные стандартизируются для обеспечения одинакового масштаба признаков.

2. Расчет матрицы ковариации: вычисляется ковариационная матрица для оценки взаимосвязей между признаками.

3. Вычисление собственных векторов и собственных значений: собственные значения и соответствующие им собственные векторы ковариационной матрицы используются для определения главных компонент.

4. Матрица главных компонентов: главные компоненты выбираются на основе их собственных значений, с наибольшими значениями соответствующими наибольшей доле вариации в данных.

5. Трансформация данных: путем отображения исходных данных на основные компоненты создается новый набор данных с уменьшенной размерностью.

В ходе исследования данного алгоритма для обработки изображений были выделены основные недостатки и преимущества.

PCA может быть использован для уменьшения размерности пу-

тем преобразования многомерных пиксельных данных в более компактное представление. Это позволяет снизить объем информации, сохраняя при этом ключевые характеристики, что упрощает обработку и анализ данных в будущем. Также этот метод также может помочь в выявлении скрытых закономерностей в изображениях и улучшении эффективности работы моделей машинного обучения, использующих изображения в качестве входных данных [1].

Однако важно отметить, что РСА имеет свои ограничения, такие как чувствительность к выбросам и возможную потерю информации при уменьшении размерности, что потенциально может отразиться на качестве обработки изображений.

На этапе разработки программного обеспечения появилась такая проблема, как выбор оптимального количества компонентов, для сохранения нужного количества информации. К сожалению, этот критерий изменяется в зависимости от самого изображения, но для автоматизированного поиска значения можно использовать такие методы, как: перекрестная проверка, информационный критерий, метод локтя и сохранение фиксированного объема информации [2].

Также сильно сказывается повышение сложности обработки в зависимости от того, какое разрешение имеет изображение и какого формата используется цветопередача. Монохромное изображение представляется набором пикселей, где каждое кодируется одним значением яркости, что соответствует оттенкам серого. Отличие RGB-модели в том, что каждый пиксель кодируется тремя каналами: красным, зеленым и синим. Из-за большего количества переменных в цветном формате больше вероятность появления различного рода дефектов.

Анализ факторов с использованием метода главных компонентов представляет собой эффективное средство для обработки данных, обнаружения скрытых закономерностей и применения в разнообразных сферах. Глубокое понимание его основ и умение применять в реальной практике способны значительно расширить аналитические возможности исследователя [3].

Язык программирования Golang (Go) подходит для написания backend-части веб-сайта ввиду таких качеств, как:

1. Высокая производительность. Основное достоинство языка программирования заключается в его эффективности. Прямая компиляция в машинный код приводит к высокой скорости работы приложений. Это критично для платформ, анализирующих космические снимки, ввиду их работы с обширными массивами данных и потребностей в продвинутом аналитическом аппарате. Способность к мгновенной обработке снимков и обеспечение доступа к результатам в режиме реального времени существенно усиливает производительность таких

веб-сервисов.

2. Простота и читаемость кода. Язык программирования Go преимущественно отличается своей простотой и сжатостью синтаксиса, что делает его доступным для программистов различного уровня квалификации. Ясный и читаемый код способствует легкому сопровождению и развитию проектов. Это особенно важно для разнообразных команд разработки, в которых критично, чтобы новые члены быстро адаптировались и понимали архитектуру и логику приложений. Простота Go позволяет разработчикам уделять больше внимания разработке функционала, например, в области обработки изображений, минимизируя при этом время на освоение сложностей языка программирования.

3. Поддержка параллелизма. Go обеспечивает высокоэффективные механизмы для реализации параллельной обработки, включая горутины и каналы, что способствует разработке высокопроизводительных многопоточных приложений, оптимизирующих использование ресурсов машин. В контексте веб-сервиса по исправлению аэрокосмических снимков это дает преимущество в виде одновременной обработки большого количества картинок, существенно сокращая время, необходимое для выполнения операций. Таким образом, система может асинхронно анализировать снимки, отправленные пользователями, обеспечивая бесперебойную работу без ожидания обработки.

4. Кроссплатформенность. Go обладает функциональностью для кроссплатформенной разработки, обеспечивая совместимость написанного кода с множеством операционных систем. Этот аспект упрощает задачу разработчиков, позволяя им создавать и деплоить веб-приложения на разнообразных серверах или в облачных инфраструктурах без дополнительных модификаций кода. Такая возможность предоставляет удобство в масштабировании и интеграции с различными системами, делая разработку более эффективной и гибкой.

Для создания, развертывания и управления контейнерами, обеспечивающими легкость и изоляцию запуска приложений в сопровождении всех необходимых зависимостей, прекрасно подходит Docker. Использование Docker значительно улучшает переносимость приложений и облегчает их деплоймент на многообразие систем [4].

В отличие от своих немногих конкурентов он имеет ряд преимуществ:

1. Изоляция приложений: каждое приложение и его зависимости упаковываются в контейнер, что обеспечивает непрерывную изоляцию.

2. Легковесность: контейнеры используют меньше ресурсов, поскольку они разделяют ядро операционной системы, в отличие от тра-

диционных виртуальных машин, которым требуется полнофункциональная операционная система.

В качестве базы данных идеально подходит PostgreSQL ввиду таких качеств, как:

1. Гибкость: PostgreSQL обеспечивает возможность разработки собственных типов данных, создания пользовательских функций и определения уникальных операторов, предоставляя тем самым инструментарий для тонкой настройки базы данных в соответствии с уникальными нуждами заказчиков.

2. Управление геоданными: применяя модуль PostGIS, PostgreSQL способен эффективно управлять геопространственной информацией, обеспечивая высокую производительность при работе с картографическими и аэрокосмическими снимками.

3. Надежность: PostgreSQL обеспечивает высокий уровень защиты данных и поддерживает транзакционность, что является ключевым аспектом для систем, где требуется гарантированная надежность.

4. Масштабируемость: система эффективно управляет значительными нагрузками данных, решая вопросы с высокой сложностью, что обеспечивает её применение в крупномасштабных инициативах.

Заключение. Метод главных компонент – это мощный инструмент для анализа данных, который помогает исследователям и практикам извлекать полезную информацию из сложных многомерных наборов данных. Понимание его принципов и возможностей позволяет эффективно применять его в различных областях в нашем случае для обработки космических изображений. Используя различные web-технологии, мы можем создать информационную систему, которая будет обрабатывать данные.

ЛИТЕРАТУРА

1. Дюма, С. Метод главных компонент и его применение / С. Дюма // Труды Института прикладной астрономии РАН. – 2011. – № 22. – С. 31-41.

2. Лебедев, К.Н. Проблемы факторного анализа, основанного на методах детерминированного факторного анализа (проблемы науки «экономический анализ») / К.Н. Лебедев // ЭТАП: экономическая теория, анализ, практика. – 2012. – № 3. – С. 4-13.

3. Сычев, А.С. Комплексирование изображений по методу главных компонент с адаптацией к мощности шума / А.С. Сычев, И.С. Холопов // Вестник Рязанского государственного радиотехнического университета. – 2020. – № 71. – С. 3-14.

4. Разработка картографических веб-приложений на основе геоинформационных технологий / Н.В. Акинина, А.В. Курагин, А.Н. Колесенков, Б.В. Костров // Телекоммуникации. – 2023. – № 2. – С. 23-31.