Модуль 3 Использование методов машинного обучения в анализе временных рядов.

**Введение:**

Модуль дает представление об особенностях использования современных методов машинного обучения в анализе временных рядов. В том числе обсуждаются релевантные задачи и особенности подходов, которые связаны со спецификой изучаемой модальности данных. Отдельно обсуждаются вопросы глубокого обучения нейронных сетей в анализе временных рядов.

Отметим, что в анализе временных рядов глубокое обучение нейронных сетей далеко не всегда является необходимым. Такой подход чаще оправдан в случае больших объемов данных много-переменных временных рядов, чьи свойства не позволяют быть легко формализованными.

Упор в модуле сделан на решения дискретных задач, таких как задачи классификации и обнаружения аномалий. Однако следует отметить, что подходы могут быть с лёгкостью транслированы и на решения других задач анализа ВР.

В ходе изучения модуля будут изучены такие вопросы, как:

* Типы и примеры задач классификации ВР
* Особенности методов классификации ВР на основе сырых данных
* Особенности выбора расстояний и других метрик для классификации ВР
* Особенности выбора признаков ВР и методов классификации на их основе, а также особенности использования комбинаций изученных методов
* Типы аномалий во временных рядах и методы работы с ними.
* Особенности использования глубоких нейронных сетей в анализе ВР, в том числе рекуррентных сетей и других классов сетей.

На практических занятия будет произведено знакомство с методами настройки и использования:

* моделей классификации ВР в фреймворке SKTime
* методов обнаружения аномалий
* методов работы с глубокими нейронными сетями в фреймворке TSAI

В результате прохождения модуля слушатель будет:

* Знать специфику, разновидности, и возможности использования методов машинного обучения в анализе временных рядов
* Уметь провести выбор методов анализа временных рядов с использованием машинного обучения для решения той или иной задачи.
* Владеть навыками использования методов машинного обучения в анализе временных рядов

Модуль состоит из 2 лекций и 3 практик.

Лекции

1. Классификация и обнаружение аномалий
2. Глубокое обучение НС во ВР

Практики

1. Классификация ВР
2. Обнаружение аномалий во ВР
3. Введение в TSAI и классификация ВР при помощи НС

**В ходе изучения модуля будут изучены такие понятия как:**

**Классификация ВР –** задача оценки соответствия некоторых сегментов ВР заданным дискретным категориям – т.н. классам. То есть оценка схожести некоторым множествам примеров. Схожесть может быть разной, например, могут быть следующие варианты:

* + схожесть во временном поведении .
  + схожесть в частотном поведении (налчие определннных компонент в спектре) .
    - А также время-частотное поведение.
  + схожесть по форме (патерн, без привязки ко времени).
    - Схожесть по частоте повторяемости, паттерн – шаблон поведения.
  + схожесть поведения компонет (монотонный тренд, характер сезонности и т.д.).

**Методы на основе сырых данных** подразумевают поиск "схожести" участков ВР для каждого класса как таковых**.** При этом "схожесть" определяется как значение определенной метрики, например, расстояние. Достоинствами данного подхода являются потенциально высокая точность. Однако, для работы подход требует достаточно много вычислительных ресурсов и не всегда позволяет "правильно" учесть всю содержащуюся в рядах информацию. Также отметим, что результаты работы для данного подхода не всегда являются интерпретируемыми, что может представлять отдельную проблему.

**Эластичные меры расстояний** являются модификациями традиционных подходов, устойчивыми к небольшим изменениям в данных. Наиболее универсальным подходом можно считать т.н. Расстояние динамической трансформации (деформации) по времени **(Dynamic Time Warping Distance, DTW).** Алгоритм DTW, как следует из названия, пытается деформировать ось времени одного из сигналов(сжатие/растяжение для каждых двух точек). Деформация производится таким образом, чтобы найти минимальное расстояние между двумя точками. Метод используется например в алгоритме **kNN-DTW.** Алгоритм DTW подходит для классификаторов по наличию шаблона.

В противоположность к подходам на полных сегментах данных ("whole series") авторы выделяют **интервальные оценки** ("intervals" ) - то есть оценки по одному или несколькими частям сегмента временного ряда. Могут быть выбраны как одно или совокупность интервалов. Интервалы могут быть выбраны: в соответствии с каким либо критерием, случайно или например на заданной сетке.

**Шейплет** **(Shapelet)** - сравнительно короткий участком (паттерн) временного ряда, который в наибольшей степени позволяет отличить один класс от другого. Шейплеты можно использовать для обнаружения «фазово-независимого локализованного сходства между сериями одного и того же класса». Шейплеты подходят для классификаторов по наличию шаблона или паттерна поведения сегмента.

**Алгоритмы на основе словарей –** это алгоритмы представления ВР или их частей в виде гистограмм некоторого описания в грубых дискретных шагах и квантах значений ВР.

**Подход на основе признаков -** предполагает, что модель строится на выделение признаков и принятии решений по ним. Достоинствами подхода являются низкая избыточная и более высокая интерпретируемость классификатора. Однако не всегда признаковое описание достаточно в рамках решаемой задачи.

**Признаки** - это определенные регулярные характеристиках для каждой последовательности. Признаки:

* + Регулярны для набора данных.
  + Отражают класс (регулярны для класса).
  + Позволяют различать классы как можно более четко

Признаки могут быть выделены как для всего сегмента ВР, так и для интервалов в нем. Могут быть как точечными, так и векторными.

**По группам признаки** можно разделить на:

* Описательные статистики
* Временные параметры
* Спектральные параметры (а также Время-частотные, псевдо спектральные)
* Экзогенные факторы, а также некоторые критерии (например наличие паттерна в данных) также могут быть выбраны как признаки.

Подход на основе признаков предполагает:

* Выделение репрезентативного и не избыточного пространства признаков
* Построение классификатора на их основе

Для многопеременных ВР могут быть несколько подходов к выделению признаков:

* объединение признаков нескольких составляющих в один вектор;
* ансамблирование результатов классификации по каждой составляющей;
* использование специальных методов,

Есть готовые пространства признаков и готовые алгоритмы на их основе. К таковым относят:

* **Лес временных рядов (TSF) –** построение случайного леса по специальному критерию для деревьев, сформированных из временных признаков для случайных интервалов внутри ВР.
* **Canonical Interval Forest (CIF)** используетоснову TSF и набора признаков Catch22 с расширением для построения классификатора.
* **Спектральный ансамбль со случайными интервалами (RISE) –** случайный лес, деревья которого представляют признаковое описаниевсего сегмента и интервалов внутри, при этом для каждого участка ВР выделяются векторные признаки имеющие спектральную интерпретацию.
* **И другие.** Часто это набор признаков (например tsfel) и алгоритм на подобие случайного леса или регрессии над ними для принятия решений.

**Комбинации подходов –** использование нескольких гетерогенных подходов к классификации ВР. Используется если одним подходом ВР не описать. методы на основе данных и на основе признаков могут быть объединены в ансамбли таким образом, чтобы учесть преимущества и недостатки тех и других подходов. Пример таких методов **HIVE-COTE.**

**HIVE-COTE -** Метод ансамбля с иерархическим коллективным голосованием на основе преобразований; мета-ансамбль, построенный на классификаторах (The Hierarchical Vote Collective of Transformation-based Ensembles with Collective of Transformation-based Ensembles): шейплеты (паттерны), BOSS (словари – частота паттернов), Лес временных рядов (временное Поведение), RISE (частотное поведение).

**ROCKET** (RandOm Convolutional KErnel Transform) Классификатор - представляющий набор случайных светрточных ядер и признаковое описание над каждым из результатов таких сверток. Как правило ядер достаточно много. Модификации алгоритма типа mini-rocket позволяют упростить модель.

**Задача обнаружения аномалий в ВР -** обнаружение резких (относительно кратковременных) изменений значений данных относительно некоторых стандартных, например предшествующих значений ряда. Такие значения могут иметь различную интерпретацию в том числе новизна в данных, выбросы, точки перегиба или другие.

**Корневыми причинами** аномалий могут быть, например, неисправности, мошенничество, появлене новых факторов или другие нерегулярные события.

Аномалии могут быть как

* **Локальными (**не превышают общий диапазон значений - **Контекстные**.)
* **Глобальными**
* **Точечными** .
* **Коллективными (**продолжительными**)**.

По характеру могут быть:

* **точечный скачок**
* **изменение дисперсии**
* **сдвиг уровня**
* изменение огибающей/сезонности/друой составляющей
* **кратковременное отсутствие паттернов или изменение**  поведения

**Методы обнаружения аномалий:**

* **контролируемые методы обнаружения аномалий** (класс нормы и аномалий известны)
* **полу-контролируемые методы обнаружения аномалий (**класс нормы известен**)**
* **неконтролируемые методы обнаружения аномалий (**по внешнему критерию**)**

**Глубокие нейронные сети** используются в случаях сложных ВР с не формулируемыми зависимостями при достаточно больших объемах данных. Часто это многопеременны ВР, где также сложно сформулировать связь компонент. При этом подходе:

* + - Нет необходимости в статистической гипотезе или конкретной форме модели.
    - Может аппроксимировать любую функцию с пропущенными данными, аномалиями и другими нерегулярными шаблонами.
    - Позволяют работать с огромными многомерными рядами данных со сложной взаимосвязью поведения между данными.
    - Автоматически извлекают и обрабатывают сложные признаки и отношения между ними.
    - Для сложных и многомерных данных в большом количестве модели, управляемые данными, могут обеспечить лучшую производительность.
  + **НО**
    - Требуется тщательная настройка гиперпараметров.
    - Часто требуется ансамбль сетей для получения высокой точности.
    - Тяжело перетренировать.
    - Трудно достичь сопоставимой точности с подходом на основе модели для относительно простых рядов.
  + Выбор конкретных методов зависит от поставленной задачи.

**Среди глубоких НС доступны 4 типа:**

* Рекуррентные сети (актуальны для не слишком сложных данных)
* Сверточные сети (Самый популярный подход в применении нейронных сетей к временным рядам для задач классификации - ​​это сверточные сети)
* Полносвязные сети (не популярны)
* Сети трансформеры (для очень больших объемов данных)

**Рекуррентные нейронные сети** это тип сетей, основанный на переиспользовании одних и тех же весовых параметров, но изменяющегося вектора состояний для последовательных участков данных.

Классический подход использования рекуррентных сетей (**vanilla RNN**) страдает от проблема взрыв веса (или градиента) и вымывание градиента в силу переиспользования параметров, особенно при длинном контексте. Также следствием высокой вероятности переобучения является не возможность сделать рекуррентные сети слишком глубокими. Для решения этой проблемы используют:

* усечение обучение (усеченное обратное распространение).
* двунаправленное обучение (если возможно).
* использование продвинутых архитектур рекуррентных сетей (LSTM, GRU и др.).
* регуляризация обучения (батч-норма, L1, L2,ограничение градиентные (или его нормы), функции активации с самонормализацией и т. д.).

Особенностью **LSTM** ячейки является раздельный учет скрытых состояний

* кратковременного контекста (например так мы можем учесть сезонность),
* и долговременного контекста (так мы сможем учесть тренд, Долговременный контекст – это контекст с долгим затуханием).

Идея **GRU** ячейки объединить оба скрытых состояния вместе и передавать их одним параметром. За счет этого общее число параметров будет меньше.

**В вероятностном обучении**  мы пытаемся не само значение прогноза, но его среднее ожидание и его дисперсию, подчиняющиеся распределению. Итоговое предсказение поржадается из этого распредления. Другими словами, мы пытаемся построить какое-то распределение и минимизировать сумму его значений.

Пример такого подхода это сеть Deep AR. Сеть использует итеративного обучения (предсказание шаг-за-шагом) в сочетании с lstm-ячейками и так называемым вероятностным обучением с учителем

**Полностью связанная сеть** типа многослойный персептрон - подход не пользуется популярностью в анализе временных рядов из-за нескольких **недостатки**:

* Отсутствие учета связности между сегментами  -
* Отсутствие учета казуальности (не направлена)
* Огромное количество параметров  .

Другие примеры подхода с полносвязными сетями это **Force-back FCN** послойный подход, которые предполагает, что для каждого следующего слоя должны быть предсказаны как следующие, так и предыдущие значения. И сети **нелинейной сети авторегрессии, NAR**  - расширение понятия AR на MLP архитектуру.

**Сверточные сети** представляют аналог двух-мерных сверточных сетей для 1d случая. Популярны как простые архитектуры, так и использование таких приемов, как расширенная свертка и внимание.

**Внимание** - механизм улучшения долгосрочного обучения зависимостям в нейронной сети. Механизм внимания дает несколько ключевых преимуществ

* позволяют реагировать на любые значимые редкие события
* работать как регуляризация, чтобы избежать исчезновения градиента и взрыва при проходе по данным.

Чаще используется вариант **самовнимания** **-** основная идея — это учиться только из входных данных. В более продвинутых вариантах используется с маскировкой для эффекта казуальности.

**Блок трансформаторов** состоит из: слой самовнимания, нормализация слоя, слой прямой связи (один MLP применяется независимо к каждому вектору), и еще одна нормализация слоя. Также добавляются остаточные соединения. Блок позволяет не содержит сверток или рекуррентных ячеек.

**Архитектуры трансформеры** позволяет легко учитывать долговременный контекст, однако для кратковременного контекста их обучать сложно. Как правило используются для очень больших массивов данных.

**Заключение:**

Модуль «Использование методов машинного обучения в анализе временных рядов» раскрывает особенности использования методов машинного обучения в анализе временных рядов. В том числе показаны основные особенности данных подходов в отношении этой модальности данных. Напомним, что эти особенности связаны с казуальностью и заданным порядком значений ВР, имеющим четкую интерпретацию. Именно это свойство и дает новую информацию, использование которой позволяет ввести те техники, примеры которых показаны в модуле.

Рассмотренные в модуле примеры задач в настоящее время имеют высокую степень важности, а методы для их решения актуальны и составляют т.н. State-of-the-art (то есть современный уровень решения). Однако, стоит понимать, что область знаний развивается достаточно быстро. Поэтому специализируясь на данной области нужно всегда поддерживать уровень знаний акутальным.