

КЛАСТЕРИЗАЦІЯ І КЛАСИФІКАЦІЯ ЧАСОВИХ РЯДІВ

У сучасному світі обсяги даних постійно збільшуються, що потребує використання нових методів та підходів до їх аналізу. Часові ряди є важливим типом даних, який описує динаміку певних явищ в часі. Однак, аналіз часових рядів може бути складним завданням через їх високу динамічність та складні взаємозв'язки між даними.

Кластеризація і класифікація часових рядів є актуальними напрямками досліджень у багатьох галузях, таких як фінанси, медицина, промисловість, соціальні мережі, трейдинг тощо. Кластеризація та класифікація є важливими завданнями аналізу часових рядів, оскільки дозволяють групувати подібні ряди та класифікувати їх за певними критеріями.

Існує багато методів для кластеризації та класифікації часових рядів. У цій роботі зроблено короткий огляд найпопулярніших і найефективніших, за різними оцінками, з них.

Метод головних компонент (РСА) використовується для зменшення кількості змінних у даних шляхом перетворення даних на нові, які є лінійними комбінаціями початкових змінних. Він може бути використаний для кластеризації часових рядів за допомогою розрахунку власних векторів та значень, що відповідають найбільш значимим змінним.

Також досить популярним методом є стандартний кластерний аналіз. Кластерний аналіз використовується для групування часових рядів за схожістю. Метод кластерного аналізу використовується для визначення кількох кластерів, кожен з яких складається з часових рядів, які схожі між собою. Можуть бути використані різні метрики подібності, такі як Евклідова відстань, косинусна відстань, Мангеттенська відстань та інші.

Для аналізу та прогнозу часових рядів часто використовуються методи машинного навчання. Одним з найбільш ефективних методів є використання рекурентних нейронних мереж, зокрема, довго-короткочасної пам'яті (LSTM).

Нейромережа LSTM є спеціальним типом рекурентних нейронних мереж, який може ефективно працювати з даними, які мають довготривалу залежність. Це робить його досить ефективним для аналізу та передбачення часових рядів.

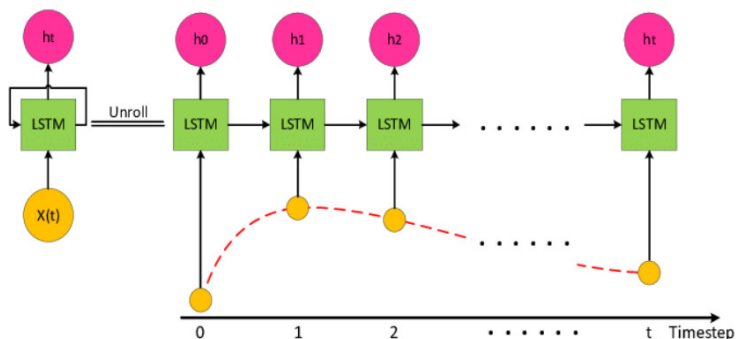


Рис.1. Схема використання LSTM при аналізі часових рядів

У LSTM мережі кожен нейрон має внутрішні стани, які зберігають інформацію з попередніх часових кроків. Це дозволяє мережі зберігати інформацію про довготривалу залежність між даними в часі. При використанні LSTM для аналізу часових рядів дані спочатку розбиваються на послідовності, що містять фіксовану кількість кроків. Кожна послідовність використовується для тренування мережі, яка використовується для передбачення наступних значень у часовому ряді.

Проведений порівняльний аналіз методів кластеризації і класифікації часових рядів показав, що універсальних методів не існує, кожен має свої переваги і недоліки, тому для конкретного набору даних необхідно використовувати різні підходи, методи та моделі машинного навчання.

ВИКОРИСТАНІ ДЖЕРЕЛА

1. *How to Develop LSTM Models for Time Series Forecasting* [Електронний ресурс] – Режим доступу: [www/ URL: https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-seriesforecasting/](http://www.URL:https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-seriesforecasting/) – Загол. з екрану