

TRABAJO DE FIN DE GRADO



ugr

Universidad
de Granada



TimeSeriesAnalysis: Sistema integral para análisis y predicción de series temporales

Tutor de proyecto

José Manuel Benítez Sánchez

Autor

Javier Moreno Vega

June 29, 2017

Motivación

Serie temporal. Definición

Secuencia de datos medidos en determinados momentos y ordenados cronológicamente, pudiendo estar estos datos espaciados a intervalos iguales o desiguales.

- Tipo de dato de enorme importancia en múltiples campos.
- Diversos tipos de métodos para analizarlas y predecir valores futuros.
- A falta de una biblioteca que englobe estos métodos y plataforma desde la que aplicarlos...

Motivación

Serie temporal. Definición

Secuencia de datos medidos en determinados momentos y ordenados cronológicamente, pudiendo estar estos datos espaciados a intervalos iguales o desiguales.

- Tipo de dato de enorme importancia en múltiples campos.
- Diversos tipos de métodos para analizarlas y predecir valores futuros.
- A falta de una biblioteca que englobe estos métodos y plataforma desde la que aplicarlos...
- [TimeSeriesAnalysis](#)

Objetivos

Sistema clasificador

- Usando una serie temporal almacenada
- A partir del conocimiento obtenido de todas las series temporales almacenadas
- Calcula la complejidad de la serie temporal
- La clasifica comparando su complejidad con la de las demás

Objetivos (II)

Biblioteca de métodos de análisis

- Incluye métodos de análisis (medidas de complejidad), predicción y transformación de series temporales
- Además se han añadido métodos para clustering
- Y funciones básicas para trabajar con datos y series temporales
- El sistema clasificador hace uso de esta biblioteca

Objetivos (II)

Biblioteca de métodos de análisis

- Incluye métodos de análisis (medidas de complejidad), predicción y transformación de series temporales
- Además se han añadido métodos para clustering
- Y funciones básicas para trabajar con datos y series temporales
- El sistema clasificador hace uso de esta biblioteca

Plataforma web

- Aplicación web de fácil uso para el usuario
- Funciona junto a la biblioteca y la base de datos de series temporales

Metodología

SCRUM

Metodología de desarrollo ágil caracterizada por tener una estrategia de desarrollo incremental, ejecución completa del producto por intervalos, equipos de desarrollo auto organizados y solapamiento de las diferentes fases del desarrollo.

Metodología

SCRUM

Metodología de desarrollo ágil caracterizada por tener una estrategia de desarrollo incremental, ejecución completa del producto por intervalos, equipos de desarrollo auto organizados y solapamiento de las diferentes fases del desarrollo.

MVC

Arquitectura de software que divide el desarrollo de un sistema en tres módulos o partes principales, separando la interfaz de usuario (vista) de la lógica (controlador) y los datos (modelo).

Herramientas

- Git
- Docker
- RStudio
- PhpStorm
- PyCharm
- Youtrack
- Teamcity
- Visual Paradigm
- InfluxDB
- Apache
- PhpMyAdmin

Herramientas (II)

- Paquetes R: R6, parallel, Rcpp, bigmemory, testthat, ...
- Paquetes PHP (Composer): influxdb-php, slim, phpunit, ...
- Paquetes NPM: angular, bootstrap, highcharts, karma, ...

Medidas de complejidad

Definición

Cálculo que se aplica sobre un conjunto de datos, en este caso series temporales, y devuelve como resultado el grado de dificultad de los datos de esta para analizarlos.

- Los resultados de complejidad son usados para la clasificación
- Las medidas de complejidad que han sido implementadas: Kolmogorov, Lempel-Ziv, Aproximation-Entropy, Sample Entropy, Permutation Entropy, Shannon Entropy, ChaoShen Entropy, Dirichlet Entropy, MillerMadow Entropy, Shrink Entropy.

Análisis del conjunto de series temporales

- Dependiente de los métodos implementados en la biblioteca
- El resultado obtenido en este análisis es un sistema clasificador
- Los últimos resultados se ejecutaron sobre un conjunto de 60000 series temporales

Medidas de complejidad

- Sobre todas las series temporales se aplican todas las medidas de complejidad

Análisis del conjunto de series temporales

- Dependiente de los métodos implementados en la biblioteca
- El resultado obtenido en este análisis es un sistema clasificador
- Los últimos resultados se ejecutaron sobre un conjunto de 60000 series temporales

Medidas de complejidad

- Sobre todas las series temporales se aplican todas las medidas de complejidad
- Esta matriz de 60000x10 se almacena en la base de datos, cada serie temporal con sus resultados de complejidad

Análisis del conjunto de series temporales

- Dependiente de los métodos implementados en la biblioteca
- El resultado obtenido en este análisis es un sistema clasificador
- Los últimos resultados se ejecutaron sobre un conjunto de 60000 series temporales

Medidas de complejidad

- Sobre todas las series temporales se aplican todas las medidas de complejidad
- Esta matriz de 60000x10 se almacena en la base de datos, cada serie temporal con sus resultados de complejidad
- Se almacenan para una ejecución de los experimentos más rápida, las medidas de complejidad de cada serie temporal solo son calculadas una vez.

Análisis del conjunto de series temporales (II)

Clustering

- A la matriz de medidas de complejidad se le aplica Clustering

Análisis del conjunto de series temporales (II)

Clustering

- A la matriz de medidas de complejidad se le aplica Clustering
- Los métodos de clustering usados han sido: KMeans y CMeans

Análisis del conjunto de series temporales (II)

Clustering

- A la matriz de medidas de complejidad se le aplica Clustering
- Los métodos de clustering usados han sido: KMeans y CMeans
- Son métodos no-jerárquicos

Análisis del conjunto de series temporales (II)

Clustering

- A la matriz de medidas de complejidad se le aplica Clustering
- Los métodos de clustering usados han sido: KMeans y CMeans
- Son métodos no-jerárquicos
- Como método para el cálculo de centros se han implementado el método de Chiu

Análisis del conjunto de series temporales (II)

Clustering

- A la matriz de medidas de complejidad se le aplica Clustering
- Los métodos de clustering usados han sido: KMeans y CMeans
- Son métodos no-jerárquicos
- Como método para el cálculo de centros se han implementado el método de Chiu
- Para 60000 series temporales se obtuvieron 129 centros

Análisis del conjunto de series temporales (II)

Clustering

- A la matriz de medidas de complejidad se le aplica Clustering
- Los métodos de clustering usados han sido: KMeans y CMeans
- Son métodos no-jerárquicos
- Como método para el cálculo de centros se han implementado el método de Chiu
- Para 60000 series temporales se obtuvieron 129 centros
- Se seleccionó el método KMeans, ya que es el que generaba mejores resultados, en las comparaciones del experimento

Análisis del conjunto de series temporales (II)

Clustering

- A la matriz de medidas de complejidad se le aplica Clustering
- Los métodos de clustering usados han sido: KMeans y CMeans
- Son métodos no-jerárquicos
- Como método para el cálculo de centros se han implementado el método de Chiu
- Para 60000 series temporales se obtuvieron 129 centros
- Se seleccionó el método KMeans, ya que es el que generaba mejores resultados, en las comparaciones del experimento
- Al igual que con las medidas de complejidad los resultados de clustering se almacenan en la base de datos, el número de grupo con su centro

Análisis del conjunto de series temporales (III)

Clasificación

- Sobre cada grupo obtenido se aplican todos los métodos de predicción a cada serie temporal, del grupo. Se selecciona el método con menor error.

Análisis del conjunto de series temporales (III)

Clasificación

- Sobre cada grupo obtenido se aplican todos los métodos de predicción a cada serie temporal, del grupo. Se selecciona el método con menor error.
- El método de predicción para cada centro se guarda en base de datos.

Análisis del conjunto de series temporales (III)

Clasificación

- Sobre cada grupo obtenido se aplican todos los métodos de predicción a cada serie temporal, del grupo. Se selecciona el método con menor error.
- El método de predicción para cada centro se guarda en base de datos.
- Para clasificar una nueva serie temporal se calculan sus medidas de complejidad y con una función de distancia a todos los centros se selecciona el de menor distancia

Análisis del conjunto de series temporales (III)

Clasificación

- Sobre cada grupo obtenido se aplican todos los métodos de predicción a cada serie temporal, del grupo. Se selecciona el método con menor error.
- El método de predicción para cada centro se guarda en base de datos.
- Para clasificar una nueva serie temporal se calculan sus medidas de complejidad y con una función de distancia a todos los centros se selecciona el de menor distancia
- Se devuelve el grupo y el método de predicción que se le asignó.

Análisis del conjunto de series temporales (IV)

Sistema Clasificador: Características

- Usa la API para obtener las series temporales y, obtener y almacenar los resultados

Análisis del conjunto de series temporales (IV)

Sistema Clasificador: Características

- Usa la API para obtener las series temporales y, obtener y almacenar los resultados
- Se han desarrollado scripts en R específicos para el análisis

Análisis del conjunto de series temporales (IV)

Sistema Clasificador: Características

- Usa la API para obtener las series temporales y, obtener y almacenar los resultados
- Se han desarrollado scripts en R específicos para el análisis
- En algunas clases R ha sido necesario implementarlas en C++ debido a los altos requisitos de computación (Ejemplo: Cálculo de centros Chiu)

Análisis del conjunto de series temporales (IV)

Sistema Clasificador: Características

- Usa la API para obtener las series temporales y, obtener y almacenar los resultados
- Se han desarrollado scripts en R específicos para el análisis
- En algunas clases R ha sido necesario implementarlas en C++ debido a los altos requisitos de computación (Ejemplo: Cálculo de centros Chiu)
- Se han desarrollado varios paquetes R

Análisis del conjunto de series temporales (IV)

Sistema Clasificador: Características

- Usa la API para obtener las series temporales y, obtener y almacenar los resultados
- Se han desarrollado scripts en R específicos para el análisis
- En algunas clases R ha sido necesario implementarlas en C++ debido a los altos requisitos de computación (Ejemplo: Cálculo de centros Chiu)
- Se han desarrollado varios paquetes R
- Los paquetes R más importantes para el sistema clasificador son:

Análisis del conjunto de series temporales (IV)

Sistema Clasificador: Características

- Usa la API para obtener las series temporales y, obtener y almacenar los resultados
- Se han desarrollado scripts en R específicos para el análisis
- En algunas clases R ha sido necesario implementarlas en C++ debido a los altos requisitos de computación (Ejemplo: Cálculo de centros Chiu)
- Se han desarrollado varios paquetes R
- Los paquetes R más importantes para el sistema clasificador son:
 - Clustering

Análisis del conjunto de series temporales (IV)

Sistema Clasificador: Características

- Usa la API para obtener las series temporales y, obtener y almacenar los resultados
- Se han desarrollado scripts en R específicos para el análisis
- En algunas clases R ha sido necesario implementarlas en C++ debido a los altos requisitos de computación (Ejemplo: Cálculo de centros Chiu)
- Se han desarrollado varios paquetes R
- Los paquetes R más importantes para el sistema clasificador son:
 - Clustering
 - [TimeSeriesDatabase](#)

Análisis del conjunto de series temporales (IV)

Sistema Clasificador: Características

- Usa la API para obtener las series temporales y, obtener y almacenar los resultados
- Se han desarrollado scripts en R específicos para el análisis
- En algunas clases R ha sido necesario implementarlas en C++ debido a los altos requisitos de computación (Ejemplo: Cálculo de centros Chiu)
- Se han desarrollado varios paquetes R
- Los paquetes R más importantes para el sistema clasificador son:
 - Clustering
 - TimeSeriesDatabase
 - TimeSeriesComplexity

Sistema: Despliegue en Cloud

- Con el fin de conseguir una aplicación distribuida se ha usado el software Docker

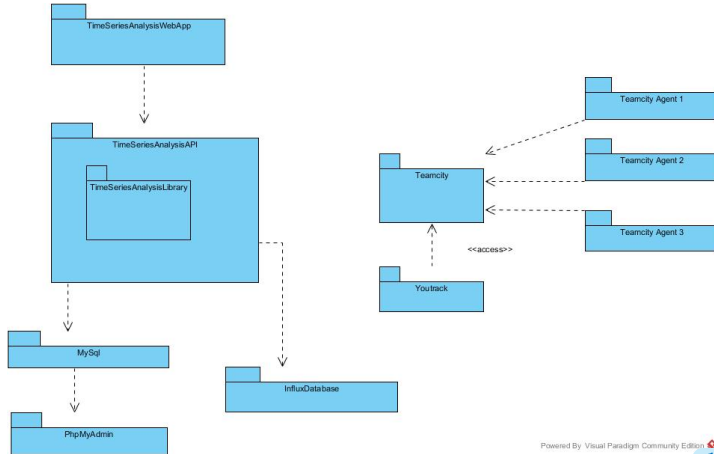
Sistema: Despliegue en Cloud

- Con el fin de conseguir una aplicación distribuida se ha usado el software Docker
- Se ha desarrollado un script en Python para gestionar todos los contenedores Docker (DockersProject)

Sistema: Despliegue en Cloud

- Con el fin de conseguir una aplicación distribuida se ha usado el software Docker
- Se ha desarrollado un script en Python para gestionar todos los contenedores Docker (DockersProject)
- Los contenedores usados han sido:
 - TimeSeriesAnalysisWebApp
 - TimeSeriesAnalysisAPI
 - MySQL
 - PhpMyAdmin
 - InfluxDB
 - TeamCity
 - YouTrack
 - TeamCity Agent 1-3

Sistema: Despliegue en Cloud (II)



Conclusiones

El proyecto desarrollado se compone de tres módulos principales:

- Biblioteca de análisis (TimeSeriesAnalysisLibrary)

Conclusiones

El proyecto desarrollado se compone de tres módulos principales:

- Biblioteca de análisis (TimeSeriesAnalysisLibrary)
- API (TimeSeriesAnalysisAPI)

Conclusiones

El proyecto desarrollado se compone de tres módulos principales:

- Biblioteca de análisis (TimeSeriesAnalysisLibrary)
- API (TimeSeriesAnalysisAPI)
- Plataforma web (TimeSeriesAnalysisWebApp)

Conclusiones (I)

TimeSeriesAnalysisLibrary

Módulo formado por varios paquetes R que incluyen todos los métodos de análisis utilizados y la funcionalidad que ha sido necesaria desarrollar para trabajar con series temporales.

Conclusiones (I)

TimeSeriesAnalysisLibrary

Módulo formado por varios paquetes R que incluyen todos los métodos de análisis utilizados y la funcionalidad que ha sido necesaria desarrollar para trabajar con series temporales.

TimeSeriesAnalysisAPI

Este módulo contiene toda la funcionalidad necesaria para una API, escrita en PHP.

Conclusiones (I)

TimeSeriesAnalysisLibrary

Módulo formado por varios paquetes R que incluyen todos los métodos de análisis utilizados y la funcionalidad que ha sido necesaria desarrollar para trabajar con series temporales.

TimeSeriesAnalysisAPI

Este módulo contiene toda la funcionalidad necesaria para una API, escrita en PHP.

TimeSeriesAnalysisWebApp

Plataforma web escrita en javascript usando el framework Angular.

Conclusiones (II)

¿Preguntas?