

Les séries temporelles

Wahiba BOUZYAN

February 14, 2024

- ① Introduction
- ② Caractéristiques des séries temporelles
 - Motifs saisonniers, tendances, cycles, variations aléatoires
 - Décomposition des séries temporelles
- ③ Méthodes d'analyse
 - Décomposition des séries temporelles
 - Modèles ARIMA
 - Modèles SARIMA
 - Modèles de réseaux de neurones
- ④ Prévion et Prédiction
- ⑤ Application pratique

Introduction

Définition

Les séries temporelles sont des ensembles de données chronologiques, offrant une vue dynamique de phénomènes enregistrés à des intervalles réguliers, tels que les ventes mensuelles ou les températures quotidiennes.

Importance dans divers domaines

Les séries temporelles sont cruciales en finance, météorologie, santé, etc., permettant de déceler des tendances et des cycles. Leur analyse informe la prise de décision, améliorant l'efficacité opérationnelle et la planification stratégique.

Motifs saisonniers, tendances, cycles, variations aléatoires

- **saisonniers** : Schémas récurrents liés à des périodes spécifiques, comme les fluctuations de ventes pendant les vacances.
- **Tendances** : Mouvements généraux à long terme indiquant une augmentation, une diminution ou une stabilité des valeurs.
- **Cycles** : Fluctuations régulières, par exemple, les cycles économiques influençant les performances financières.
- **Variations aléatoires** : Fluctuations imprévisibles représentant le bruit dans la série temporelle.

Décomposition des séries temporelles

Composantes saisonnières

Schémas récurrents liés à des périodes spécifiques, tels que les saisons, mettant en évidence les variations cycliques prévisibles.

Composantes tendancielle

Mouvements généraux à long terme, indiquant l'évolution générale de la série temporelle sur une période prolongée.

Composantes résiduelles

Variations restantes représentant le bruit ou l'irrégularité non expliqué par les composantes saisonnières et tendancielle, pouvant résulter de facteurs imprévisibles.

Décomposition des séries temporelles

La décomposition divise une série temporelle en parties comme les tendances, les motifs saisonniers et les variations aléatoires. Cela aide à mieux comprendre les influences temporelles et à repérer les tendances et les motifs.

Modèles ARIMA

Les modèles ARIMA combinent trois éléments pour prédire les séries temporelles : la dépendance temporelle, la gestion des tendances et la prise en compte des variations aléatoires. Ils sont utiles pour modéliser des données complexes.

Modèles SARIMA

Les modèles SARIMA étendent les ARIMA en incluant des composantes saisonnières. Ils sont adaptés aux données avec des variations régulières au fil du temps, améliorant ainsi les prédictions pour ces cas.

Modèles de réseaux de neurones

Les réseaux de neurones, comme les LSTM, utilisent l'apprentissage profond pour capturer des relations complexes dans les séries temporelles. Ils sont utiles lorsque les modèles traditionnels rencontrent des difficultés.

- **Objectif principal de l'analyse des séries temporelles:**
L'objectif principal de l'analyse des séries temporelles est de prévoir et comprendre les tendances, les modèles et les variations temporelles dans les données. Cela aide à anticiper les évolutions futures et à prendre des décisions éclairées.
- **Utilisation des modèles pour la prédiction des valeurs futures:** Les modèles développés dans l'analyse des séries temporelles sont utilisés pour anticiper les valeurs à venir. En se basant sur les tendances historiques et les motifs identifiés, ces modèles permettent de faire des prédictions, ce qui est essentiel pour la planification et la prise de décision dans de nombreux domaines.

Explication

Nous allons utiliser le jeu de données disponible sur:

<https://zenodo.org/records/3893504>) Ce jeu de données a été utilisé dans la compétition de prévision KDD Cup 2018. Il contient des séries temporelles longues et horaires représentant les niveaux de qualité de l'air dans 59 stations réparties dans deux villes : Beijing (35 stations) et Londres (24 stations), couvrant la période du 01/01/2017 au 31/03/2018. La qualité de l'air est mesurée à l'aide de plusieurs indicateurs tels que PM2.5, PM10, NO2, CO, O3 et SO2.

Ce fichier comprend 270 séries temporelles horaires, classées en fonction de la ville, du nom de la station et de la mesure de la qualité de l'air. Les prévisions horaires sont attendues pour la période allant du 01/05/2018 au 31/05/2018.

Explication

Modèles de réseaux de neurones

L'application pratique de ce jeu de données réside dans la possibilité d'utiliser l'analyse des séries temporelles pour anticiper les niveaux futurs de pollution atmosphérique. En se basant sur les données historiques fournies, les modèles de séries temporelles peuvent être développés pour prédire les concentrations de polluants atmosphériques à des heures spécifiques dans différentes stations. Cela peut avoir d'importantes implications en matière de gestion de la qualité de l'air, permettant aux autorités de prendre des mesures préventives, d'informer le public et de mettre en œuvre des politiques environnementales basées sur des prévisions précises.

Code source

Le code source se trouve dans ce lien:

https://github.com/Wahiba275/KDD_Cup_Dataset_with_Missing_values-.git