

Bachelor-/Masterarbeit

Evaluation of Attribution Methods in Time Series Forecasting: Concepts, Methods, and Strategies



Ausgangslage

Moderne Deep-Learning-Modelle erzielen in der Zeitreihenvorhersage hohe Prognosegenauigkeiten, sind jedoch häufig schwer interpretierbar. Zur Erklärung von Vorhersagen werden daher Attributionsmethoden wie SHAP, Integrated Gradients oder TimeSHAP eingesetzt, die die Relevanz einzelner Eingangsvariablen und Zeitpunkte bestimmen. Trotz ihrer zunehmenden Verbreitung ist bislang allerdings meist unklar, wie zuverlässig diese Verfahren wirklich sind und wie ihre Qualität bewertet werden kann.

Problemstellung

In den letzten Jahren wurden zahlreiche Attributionsmethoden für Deep-Learning-Modelle entwickelt. Diese unterscheiden sich hinsichtlich ihrer theoretischen Grundlagen, ihrer Anwendbarkeit auf Zeitreihendaten sowie ihrer Aussagekraft. Gleichzeitig existieren verschiedene Ansätze zur Bewertung der Qualität solcher Erklärungen. Der aktuelle Forschungsstand ist jedoch stark fragmentiert, sodass eine strukturierte Übersicht über bestehende Methoden und Evaluationsstrategien fehlt.

Vorgehensweise und erwartete Ergebnisse

Ziel der Arbeit ist die Durchführung einer systematischen Literaturrecherche zu Attributionsmethoden im Bereich der Zeitreihenvorhersage. Hierzu sollen relevante wissenschaftliche Veröffentlichungen identifiziert, analysiert und strukturiert eingeordnet werden.

Es sollen insbesondere folgende Fragestellungen untersucht werden:

- Welche Attributionsmethoden werden für die Erklärung von Zeitreihenvorhersagen eingesetzt?
- Wie werden die erstellten Attributionen evaluiert?
- Welche Ground-Truth-Ansätze gibt es?
- Welche Benchmark-Datensätze werden genutzt?
- Welche Probleme bestehender Evaluationen werden diskutiert?

Als Ergebnis der Arbeit soll eine strukturierte Übersicht über den Stand der Forschung entstehen. Darüber hinaus sollen bestehende Evaluationsansätze systematisch verglichen und zukünftige Forschungsrichtungen identifiziert werden.

Ansprechpartner

Jan Voets | E-Mail: voets@uni-wuppertal.de

Bachelor-/Masterarbeit

Evaluation of Attribution Methods in Time Series Forecasting: Concepts, Methods, and Strategies



Initial Situation

Modern deep learning models achieve high forecasting accuracy in time series prediction, but they are often difficult to interpret. To explain their predictions, attribution methods such as SHAP, Integrated Gradients, and TimeSHAP are used to determine the relevance of individual input variables and time steps. Despite their growing popularity, it is still largely unclear how reliable these methods actually are and how their quality can be evaluated.

Problem Definition

In recent years, numerous attribution methods for deep learning models have been developed. These methods differ in terms of their theoretical foundations, their applicability to time series data, and their explanatory power. At the same time, various approaches for evaluating the quality of such explanations have emerged. However, the current state of research remains highly fragmented, and a structured overview of existing attribution methods and evaluation strategies is still lacking.

Methods and Expected Results

The aim of this work is to conduct a systematic literature review on attribution methods in the field of time series forecasting. To this end, relevant scientific publications will be identified, analyzed, and systematically classified.

In particular, the following research questions will be investigated:

- Which attribution methods are used to explain time series forecasting models?
- How are the generated attributions evaluated?
- What ground-truth approaches exist for assessing attribution quality?
- Which benchmark datasets are used in the literature?
- What limitations and challenges of existing evaluation approaches are discussed?

The outcome of the work should be a structured overview of the current state of research. In addition, existing evaluation approaches should be systematically compared and future research directions identified.

Contact Person

Jan Voets | E-Mail: voets@uni-wuppertal.de