

Masterarbeit***Vergleich datenbasierter Signalgenerierung mittels Machine Learning und Time Series Decomposition***

Name, Vorname: Fischer, Felix

Studiengang: Informatik

Matrikelnummer: 4682953

Durch die Simulation von Signalen, bspw. Sensorwerten, können Einsatzszenarien für zeitabhängige Werteänderungen modelliert und Experimente durchgeführt werden. Mit dem Ziel, blinden Menschen die Überwachung von Live-Daten zu ermöglichen, werden im MONITOR-Projekt Konzepte der Datenpräsentation entwickelt und in simulierten Anwendungsszenarien evaluiert. Für letzteres existiert ein Anwendungsprototyp, der basierend auf den Kernaspekten der Zeitreihenzerlegung (Time Series Decomposition) nutzerdefinierte Daten sendet, um Signale zu simulieren. Dabei werden u.a. die Zeitreihenkriterien Trend, Zyklus und Abweichungen von Nutzenden festgelegt und dementsprechenden die zu sendenden Daten generiert. Zur Simulation von realen Daten ist eine manuelle Vorabanalyse notwendig, um die erforderlichen Eingaben in der Anwendung zu tätigen. Eine automatisierte Analyse existierender Daten ist daher wünschenswert, um den Prozess der Datengenerierung effizienter zu gestalten.

Die Verwendung künstlich generierter (synthetischer) Daten hat einige Vorteile gegenüber der Verwendung realer Daten. So ist beispielsweise neben datenschutzrechtlichen Aspekten das Sammeln großer Datenmengen unter Umständen sehr aufwändig. Synthetische Daten basieren oft auf einem Beispieldatensatz und weisen die gleichen Charakteristiken auf, wie die realen Daten. In verwandten Arbeiten erscheint der Einsatz von Machine Learning Algorithmen sehr vielversprechend bei der Erzeugung synthetischer Daten. Zur Erstellung von Signalen wird hingegen häufig der Ansatz des Time Series Decomposition verwendet.

Ziel dieser Arbeit ist es herauszufinden, welcher Ansatz zur datenbasierten Generierung von Signalen geeigneter ist: Machine Learning oder Time Series Decomposition.

Im Rahmen dieser Arbeit werden folgende Schwerpunkte dieses Forschungsfeldes adressiert:

1. Recherche und Analyse bestehender Literatur und aktuellen Möglichkeiten, um Daten basierend auf existierenden Datensätzen zu erzeugen, mit besonderem Augenmerk auf
 - a. Time Series Decomposition und
 - b. Machine Learning Algorithmen
2. Ermittlung von Anforderungen an eine Software, die Daten auf Grundlage eines bestehenden Datensatzes simuliert, sowohl technisch als auch benutzendenzentriert
3. Konzeption einer Lösung, um Daten mittels Machine Learning zu generieren
4. Erweiterung des bestehenden Prototyps zur Simulation von Sensoren um einen datenbasierten Ansatz zum Erstellen der Berechnungsvorschrift für zu sendenden Daten basierend auf
 - a. Time Series Decomposition und
 - b. Machine Learning.
5. Vergleich der beiden Ansätze hinsichtlich ausgewählter Aspekte der Datengenerierung sowie Benutzungsfreundlichkeit.

Relevante Literatur:

- [1] Kothare, A., Chaube, S., Moharir, Y., Bajodia, G., & Dongre, S. (2021, November). SynGen: Synthetic Data Generation. In *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)* (pp. 1-4). IEEE.
- [2] Gujar, S., Shah, T., Honawale, D., Bhosale, V., Khan, F., Verma, D., & Ranjan, R. (2022, June). GenEthos: A Synthetic Data Generation System with Bias Detection And Mitigation. In *2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)* (pp. 1-6). IEEE.
- [3] Fan, K., Xu, B., Zhang, M., Nan, M., & Huang, J. (2018, October). A New Method for Time Series Signal Decomposition. In *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 1-9). IEEE.
- [4] Hendarhewa, D. N., Perera, A. A. D. S. N. A., De Meraal, T. D. N., Indikadulle, I. K. P. H., Perera, S. D., & Rathnayake, H. M. S. C. (2021, October). Comparative Analysis between Supervised Machine Learning and Time Series Models for stock price prediction. In *2021 International Conference on Data Analytics for Business and Industry (ICDABI)* (pp. 354-358). IEEE.

Fachbetreuende:

Verantwortlicher Hochschullehrender:

Institut:

Beginn am:

Einzureichen am:

Emma Franziska Etzold, M.Sc.

Prof. Dr. Gerhard Weber

Institut für Angewandte Informatik, Prof. MCI

31.07.2023

01.01.2024

Gerhard Weber

Digital unterschrieben von
Gerhard Weber
Datum: 2023.07.20 15:32:57
+02'00'

Unterschrift des verantwortlichen Hochschullehrenden