

Title: Data-driven model

URL: https://en.wikipedia.org/wiki/Data-driven_model

PageID: 73682940

Categories: Category:Data analysis, Category:Machine learning, Category:Models of computation, Category:Statistical models

Source: Wikipedia (CC BY-SA 4.0). Content may require attribution.

Data-driven models are a class of computational models that primarily rely on historical data collected throughout a system's or process' lifetime to establish relationships between input, internal, and output variables. Commonly found in numerous articles and publications, data-driven models have evolved from earlier statistical models , overcoming limitations posed by strict assumptions about probability distributions. These models have gained prominence across various fields, particularly in the era of big data , artificial intelligence , and machine learning , where they offer valuable insights and predictions based on the available data.

Background

These models have evolved from earlier statistical models , which were based on certain assumptions about probability distributions that often proved to be overly restrictive. The emergence of data-driven models in the 1950s and 1960s coincided with the development of digital computers , advancements in artificial intelligence research, and the introduction of new approaches in non-behavioural modelling, such as pattern recognition and automatic classification.

Key Concepts

Data-driven models encompass a wide range of techniques and methodologies that aim to intelligently process and analyse large datasets. Examples include fuzzy logic , fuzzy and rough sets for handling uncertainty, neural networks for approximating functions, global optimization and evolutionary computing , statistical learning theory , and Bayesian methods . These models have found applications in various fields, including economics, customer relations management, financial services, medicine, and the military, among others.

Machine learning , a subfield of artificial intelligence , is closely related to data-driven modelling as it also focuses on using historical data to create models that can make predictions and identify patterns. In fact, many data-driven models incorporate machine learning techniques, such as regression , classification , and clustering algorithms , to process and analyse data.

In recent years, the concept of data-driven models has gained considerable attention in the field of water resources, with numerous applications, academic courses, and scientific publications using the term as a generalization for models that rely on data rather than physics. This classification has been featured in various publications and has even spurred the development of hybrid models in the past decade. Hybrid models attempt to quantify the degree of physically based information used in hydrological models and determine whether the process of building the model is primarily driven by physics or purely data-based. As a result, data-driven models have become an essential topic of discussion and exploration within water resources management and research.

The term "data-driven modelling" (DDM) refers to the overarching paradigm of using historical data in conjunction with advanced computational techniques, including machine learning and artificial intelligence, to create models that can reveal underlying trends, patterns, and, in some cases, make predictions. Data-driven models can be built with or without detailed knowledge of the underlying processes governing the system behavior, which makes them particularly useful when such knowledge is missing or fragmented.

References