LITHUANIAN COMPUTER SOCIETY
VILNIUS UNIVERSITY INSTITUTE OF DATA SCIENCE AND DIGITAL TECHNOLOGIES
LITHUANIAN ACADEMY OF SCIENCES

14th Conference on

DATA ANALYSIS METHODS
for Software Systems

November 30 – December 2, 2023

Druskininkai, Lithuania, Hotel "Europa Royale"
https://www.mii.lt/DAMSS

VILNIUS UNIVERSITY PRESS
Vilnius, 2023

Co-Chairmen:

Prof. Gintautas Dzemyda (Vilnius University, Lithuanian Academy of Sciences)
Dr. Saulius Maskeliūnas (Lithuanian Computer Society)

Programme Committee:

Dr. Jolita Bernatavičienė (Lithuania)
Prof. Juris Borzovs (Latvia)
Prof. Robertas Damaševičius (Lithuania)
Prof. Janis Grundspenkis (Latvia)
Prof. Janusz Kacprzyk (Poland)
Prof. Ignacy Kaliszewski (Poland)
Prof. Bożena Kostek (Poland)
Prof. Tomas Krilavičius (Lithuania)
Prof. Olga Kurasova (Lithuania)
Assoc. Prof. Tatiana Tchemisova (Portugal)
Prof. Julius Žilinskas (Lithuania)

Organizing Committee:

Dr. Jolita Bernatavičienė
Prof. Olga Kurasova
Assoc. Prof. Viktor Medvedev
Laima Paliulionienė
Assoc. Prof. Martynas Sabaliauskas
Prof. Povilas Treigys

Contacts:

Dr. Jolita Bernatavičienė
jolita.bernataviciene@mif.vu.lt
Prof. Olga Kurasova
olga.kurasova@mif.vu.lt
Tel. +370 5 2109315

https://doi.org/10.15388/DAMSS.14.2023
ISBN 978-609-07-0985-6 (digital PDF)

What Is a Concept Drift, and Does It Affect Machine Learning Performance?

Dalia Breskuvienė, Gintautas Dzemyda

Institute of Data Science and Digital Technologies
Vilnius University

dalia.breskuviene@mif.vu.lt

Traditional machine learning algorithms are built expecting that data for training and testing has roughly the same proportion of classes. Additionally, data and target distribution are expected to not change over time. However, this is not the case in many real-world situations, such as fraud detection, medical diagnosis, natural disaster prediction, or manufacturing quality control.

A phenomenon where the statistical properties of the target variable or data distribution change over time is called Concept Drift, meaning that models trained on historical data become less effective. We aim to comprehensively explore the concept drift phenomenon, its manifestations, and its impact on machine learning performance. This study describes concept drift and discusses its significance in real-world fraud detection systems. It delves into the causes and drivers of concept drift, including evolving user preferences, environmental changes, and intrinsic data dynamics.

We provide a comprehensive overview of concept drift, highlighting its importance and challenges in machine learning. It emphasizes the need for robust and adaptive models to address concept drift in dynamic systems and discusses the implications of this phenomenon in the broader field of artificial intelligence. Understanding and effectively managing concept drift is crucial for successfully deploying machine learning systems in evolving and complex domains. Our review focuses on the state-of-the-art techniques and approaches to detect and adapt to concept drift, focusing on incremental, transfer, and diversity ensemble learning.

To quantify the effect of concept drift on machine learning performance, we present experimental results based on benchmark datasets and real-world case studies in the literature. These results reveal how concept drift can degrade model performance, increase false positives, and lead to unexpected consequences in dynamic environments.