CONCEPT DRIFT DETECTION USING QUADTREE BASED SPATIAL MAPPING OF DATA IN FAKE REVIEW DETECTION DOMAIN

Devni Indula Heraliyawala

A dissertation submitted in partial fulfilment of the requirement for

Master of Science degree in Big Data Analytics

Department of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

Robert Gordon University Aberdeen

2023

Abstract

Fake reviews have become a significant issue in online platforms that rely on user feedback to evaluate products. To address this issue, several machine learning methods for fake review detection have been developed. Unfortunately, these models frequently suffer from concept drift, which refers to the phenomena of statistical features of data changing with time, resulting in model performance loss.

In this paper, we implemented a new approach proposed by (Amador Coelho, Bambirra Torres and Leite de Castro 2023) for detecting concept drift in the context of fake review detection. The proposed method employs a quadtree-based spatial data mapping to identify places where concept drift is likely to occur. We analyse our approach using publicly accessible datasets of Yelp reviews (Abid, M. 2019) and Fake Reviews (Salminen, J. 2023) with three feature extraction techniques against five classifiers for our analysis. The results demonstrate that the proposed method competes well in terms of accuracy, precision, and recall with state-of-the-art methods for concept drift detection. Our findings indicate that our approach can assist to improve the effectiveness of machine learning models for detecting fake reviews, resulting in more reliable and trustworthy feedback on online platforms.

Keywords: concept drift detection, fake review detection, machine learning, feature extraction online reviews, quad tree, spatial mapping