



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 5, May 2023



Impact Factor: 8.423

An AI Approach to Detect Abrupt Concept Drift in Disease Diagnosis System

Lakshmi H P¹, Manjunath J M², Sandesh V C³, Sanjana B M⁴, Dr. Abdul Razak M S⁵,
Prof. Madhu Hiremath⁶

U.G. Student Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology,
Davanagere, Karnataka, India Davanagere, Karnataka, India¹

U.G. Student Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology,
Davanagere, Karnataka, India²

U.G. Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology,
Davanagere, Karnataka, India³

U.G. Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology,
Davanagere, Karnataka, India⁴

Assistant Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology,
Davanagere, Karnataka, India⁵

Assistant Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology,
Davanagere, Karnataka, India⁶

ABSTRACT: Concept drift techniques aim at learning patterns from data streams that may change over time. Although such behavior is not usually expected in controlled environments, real-world scenarios can face changes in the data, such as new classes, clusters and features. Common concept drift domains include recommendation systems, energy consumption, artificial intelligence systems with dynamic environment interaction. We considered a review of different types of drifts and approaches to handling such changes in the data.

KEYWORDS: Artificial Intelligence, Concept drift, Disease diagnosis system

I. INTRODUCTION

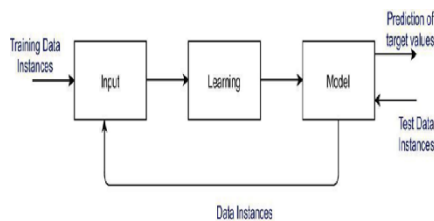
In the digital world, the data generated from various applications and grow at a rapid rate. The world is changing with the growth of technologies. Any sequence of data with the timestamp is known as data stream. The data stream contains a huge volume of data with varying velocities. Digital data generated from different sources have the following characteristics:

- Data streams are unbounded in size.
- Data arrives at high rate and varying velocity.
- Data may evolve over time.
- Data processing has constraints of memory.

The data stream is not finite, i.e., we do not know how much data in the sequence are left. Nowadays, with various advancements in technology, different applications generate vast amounts of data at a very high speed. Over time the generated data must be analyzed near real-time as one hardly has a chance to look at the data before it passes. In a data stream the problem of memory constraint must be resolved because real-time data requires storage for its processing. The data is stored for its processing. The data is stored on a traditional hard disk in traditional systems, whereas the data streams mostly need to be processed in the main memory. The data stream exhibits non-stationary distribution because the data may change over time. Shows the various aspects of traditional data mining and data stream mining. The order

of observation makes over the data in a stream that has significant learning. So, the changeover data distribution needs to be detected. This detection is one of the main challenges in data stream mining. In the data stream the data points are generated sequentially and independently under some probability distribution. Data stream has several requirements as follows:

- Process one example at most once.
- There is a memory restriction for learning of an example.
- Required to process in a limited amount of time.
- Ready to predict the class label whenever required.



Concept drift: The data stream is continuous time series data of infinite length. The online methods are required to learn those data streams. Streaming data generally encounters the limitation of space and time compared with batch processing, the real-time non-stationary data stream distribution has only one scan.

Types of drift:

- Sudden drift
- Gradual drift
- Incremental drift
- Recurrent drift

II. RELATED WORK

A survey on concept drift adaption [1]. Authors: Joao gama, Indre Zliobaite, Albert Bifet, Mykola Pechenizkiy, Abdelhamid Bouchachia. Concept drift primarily refers to an online supervised learning scenario when the relation between input data and the target variable changes over time. Assuming a general knowledge of supervised learning in this article, we characterize adaptive learning processes, categorize existing strategies for handling concept drift, overview the most representative, distinct and popular techniques and algorithms, discuss evaluation methodology of adaptive algorithms, and present a set of illustrative applications. The survey covers the different facets of concept drift in an integrated way to reflect on the existing scattered state of the art. Thus it aims at providing a comprehensive introduction to the concept drift adaption for researchers, industry analysts and practitioners.

An overview of concept drift applications [2]. Authors: Indre Zliobaite, Mykola Pechenizkiy, Joao Gama. This chapter provides an application-oriented view towards concept drift research, with a focus on supervised learning tasks. First, we overview and categorize application tasks for which the problem of concept drift is particularly relevant. Then we construct a reference framework for positioning application tasks within a spectrum of problems related to concept drift. Finally, we discuss some promising research directions from the application perspective, and present recommendations for application driven concept drift research and development.

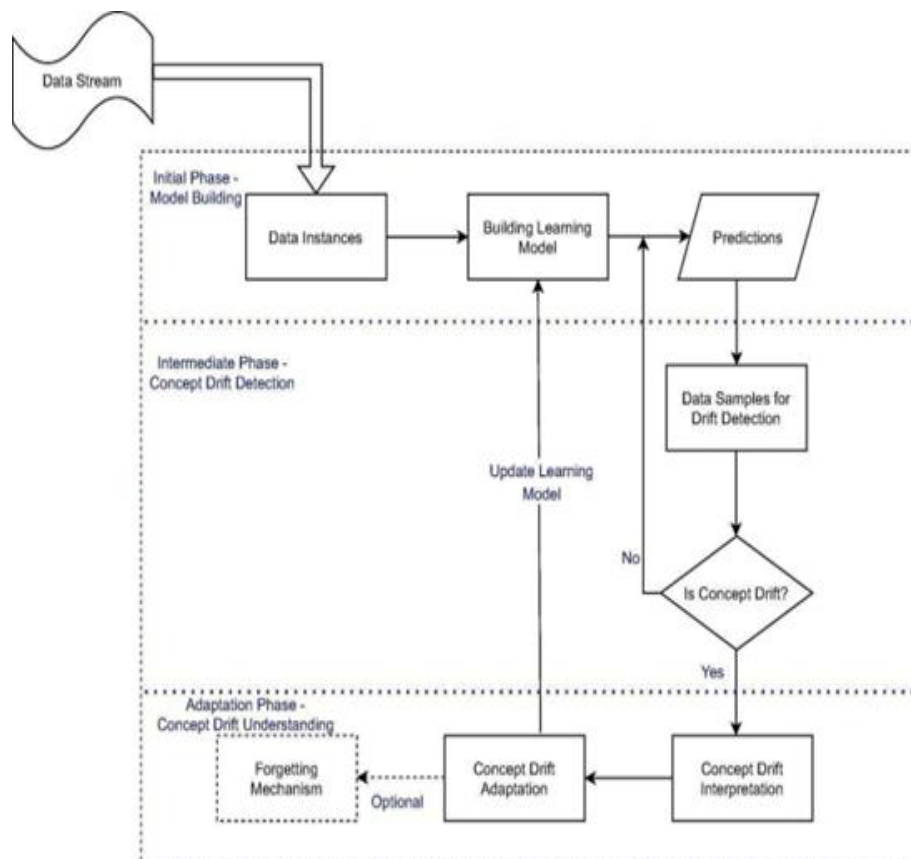
Soft clustering for enhancing the diagnosis over machine learning algorithms [3]. Authors: Theyazn H. HAlldhyani, Ali Saleh Alshebami, Mohammed Y. Alzahrani. The objective of the present research work is presented to improve the surveillance detection system for chronic disease, which is used for the protection of people's lives. For this purpose, the proposed system has been developed to enhance the detection of chronic disease by using machine learning algorithms. The standard data related to chronic disease have been collected from various worldwide resources.

The immediate effect of covid-19 policies on social distancing in the United States. Public Health Reports [5]. Authors: Abouk R, Heydari B. Although anecdotal evidence indicates the effectiveness of coronavirus disease 2019 social-distancing policies in the United States during the early stage of the pandemic.

Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda [4]. Authors: Yogesh Kumar, Apeksha Koul, Muhammad Fazal Ijaz. This article covers the comprehensive survey based on artificial intelligence techniques to diagnose numerous diseases such as Alzheimer, cancer, diabetes, chronic heart disease, tuberculosis, stroke and cerebrovascular, hypertension, skin, and liver disease. We conducted an extensive survey including the used medical imaging dataset and their feature extraction and classification process for predictions. Based on the study of different articles on disease diagnosis, the results are also compared using various quality parameters such as prediction rate, accuracy, sensitivity, specificity, the area under curve precision, recall and F1-score.

III. METHODOLOGY

The general block diagram of the concept drift detection method. It is divided into three phases. In the initial phase, initial data stream are used to build the learning model which predicts the target values. In the next phase, concept drift is identified in the data samples. If there is no drift in the data samples, further prediction of current data instances is performed. When drift is detected, the interpretation of concept drift is occurred, and successively concept drift adaptation and the forgetting mechanism are performed in the last phase. Here in the adaptation process, it updates the learning model using current data instances. In this way, concept drift detection is performed in the streaming environment.



There are many ways to monitor concept drift, as given below:

- Concept drift detection is done by checking the probability distribution of data since it may change over time.
- One can analyze whether the concept drift has happened. It can be done by monitoring and tracking the similarities or dissimilarities between different symptoms shown by a person.
- Concept drift leads to degrading the accuracy of the classifier. So, it can be one of the measures while detecting the concept drift in a given data stream.
- Recall, Precision, F-measure, ROC and AUC are some of the accuracy measures of a classification model, Generally, these measures are used by many authors.



- Timestamp can be used with either a single sample or block of samples. Timestamp can be one of the given attributes to know the occurrence of concept drift.

IV. EXPERIMENTAL RESULTS

The proposed system uses a simple AI model and data collected by a period to use it is the AI knowledge and using that knowledge to predict the concept drift occurred in the covid-19 disease. The system is showing a graph of covid-19 cases occurring daily and if there are sudden changes due to any external factors, it will identify that some drift is going to happen which can be viewed through the graph.

- Total number of processed samples: 16,336
- Number of data samples used for pre-training: 3,267
- Number of data samples used for testing: 13,067
- Method used: Prequential Evaluation
- Algorithm used: Naïve Bays and Adaptive Random Forest
- Mean performance:
 - ARF – Accuracy: 0.9858
 - NB – Accuracy: 0.1784



V. CONCLUSION

In conclusion, the methodology presented for handling concept drift using evaluation prequential on Adaptive Random Forest Algorithm is an effective approach to detecting and adapting to changes in the data distribution. The methodology includes preprocessing the data, training the models on the first chunk of data and continuously evaluating their performance using evaluation prequential. When a concept drift is detected, the models are updated with the new data and their performance is evaluated again. If the performance degrades significantly the models are re-grown with the new data. As the COVID-19 pandemic continues to evolve, detecting concept drift in disease patterns and transmission dynamics becomes increasingly important for public health decision making. An AI approach to detect concept drift can help health authorities and policymakers to better understand the changing patterns of the disease and respond appropriately.

REFERENCES

- [1]. Theyazn H. H Aldhyani, Ali Saleh Alshebami, and Mohammed Y . Alzahrani, “Soft Clustering for Enchancing the Diagnosis of Chronic Disease over Machine Learning Algorithms”. *Hindawi, J Healthc Eng. 2020*
- [2]. Manzoor Hashmani, Muslim Jameel Syed, Mobashar Rehman, Atsushi Inoue, “Concept Drift Evolution In Machine Learning Approaches: A Systematic Literature Review”, Volume 13(2021)-*Edizione 1*(January 2020)pp. 1-16, (2021)
- [3]. Abouk, R., &Heydari, B., “The immediate effect of COVID-19 policies on social distancing behavior in the United States”. *Public Health Reports*, 136(2): 245-252, (2021)
- [4]. Naresh Kuma, Nripendra Narayan Das, Kamali Gupta, Jatin Bindra, “Automated Disease Diagnosis Using Machine Learning Models”, *Journal of Healthcare Engineering* (12):1-13,(2021)
- [5]. Yogesh Kumar, Apeksha Koul, and Muhammad Fazal Ijaz, “Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda”, *J Ambient IntellHumanizComput*, 1-28,(2022 Jan 13)
- [6]. Md Manjurul Ahsan, Shahana Akter Luna, Zahed Siddique, “Machine-Learning-Based Disease Diagnosis : A Comprehensive Review”, *Healthcare* 10(3), 541, (2022)
- [7]. Ali Kashif Bashir, “Online Disease Identification and Diagnosis and Treatment Based on Machine Learning Technology”, (2022)



SJIF Scientific Journal Impact Factor

Impact Factor: 8.423



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details