Toward to reducing de_duplicate data in Big Data Based on Deep Learning methods and parallel programming

PhD Student: Fatemeh Nasri Shandiz

University of Guadalajara (CUCEA)

Department of Information Technology

Guadalajara, México

Key words: de-duplicate data, reduce data, massive data processing, parallel programing, Deep learning, and Image processing.

Abstract:

Now a day by fast speed of information technology in every aspect of the life and with growing number of users throughout networks, every second produce so many information and data. A considerable portion of these data are duplicate or expired data. Maintenance massive duplicated data will be costly and against Green computing goal since process massive data consume immense of energy. Besides the colossal expenditure of data storage, calculated based on the usage space, enterprise companies are always interested to pay less for the storage of data. Regarding these significant facts it is important to reduce de-duplicate data in massive data of data storages. Since processing of massive data is in parallel this work will look for a parallel optimize algorithm to recognize and reduce de-duplicate data in data storages. In light of the proposed idea, data transfer rates, server hosting costs, and data storage cost can be substantial reduced. Also access to correct data will be faster and easier. In this research focus on one type of data which has more rank of being duplicate through data centers. Hence images