

Chapter 15

Recommender System in the Context of Big Data: Implementing SVD-Based Recommender System using Apache Hadoop and Spark

Khadija Ateya Almohsen
Ahlia University, Bahrain

Huda Kadhim Al-Jobori
Ahlia University, Bahrain

ABSTRACT

The increasing usage of e-commerce website has led to the emergence of Recommender System (RS) with the aim of personalizing the web content for each user. One of the successful techniques of RSs is Collaborative Filtering (CF) which makes recommendations for users based on what other like-minded users had preferred. However, as the world enters the Big Data era, CF has faced some challenges such as: scalability, sparsity and cold start. Thus, new approaches that overcome the existing problems have been studied such as Singular Value Decomposition (SVD). This chapter surveys the literature of RSs, reviews the current state of RSs with the main concerns surrounding them due to Big Data, investigates thoroughly SVD and provides an implementation to it using Apache Hadoop and Spark. This is intended to validate the applicability of, existing contributions to the field of, SVD-based RSs as well as validated the effectiveness of Hadoop and Spark in developing large-scale systems. The results proved the scalability of SVD-based RS and its applicability to Big Data.

INTRODUCTION

Advances in technology, the wide spread of its usage and the connectivity of everything to the Internet have led the world to experience unusual rate of generating and storing data; resulting in what is being called Big Data phenomenon. As a consequence of this emerging fluid of data, normal tasks and activities become challenges. For instance, browsing the web and searching for interesting information or products

DOI: 10.4018/978-1-5225-0182-4.ch015

is a routine and common task. However, the massive amount of data on the web is expanding the noise there making it harder and more time consuming to choose the interesting pieces of information from all this noise (Villa, 2012; Schelter & Owen, 2012).

Likewise, the currently available systems, technologies and tools show their limitation in processing and managing this massive amount of data. This leads to the invention of new technologies, such as Map Reduce of Google, Hadoop of Yahoo! And Spark from University of California, Berkely. These technologies are just like the telescopes which allow us to comprehend the universe (Schönberger & Cukier, 2013). With this in mind, existing systems have been adapted to meet Big Data by using the newly invented tools and technologies. One of these systems is recommender system and it is the one under study in this chapter.

Recommender systems have been implemented long time ago by several Internet giants; like Amazon.com, Facebook and Google. These systems suggest new items that might be of interest to the user by analyzing user's profiles, their activities on the websites as well as their purchase history; if applicable. However, Big Data increases the cognitive load on the user, posing more challenges on recommender systems as it should provide many recommendations of high quality by analyzing huge data of customers and products. In another word, high quality, scalability and performance become concerns. (Berkeley School: Lecture12 -Analyzing Big Data with Twitter: Recommender Systems, 2012; Chiky, et al, 2012; Thangavel, et al, 2013). This encourages more research work on recommendation algorithms and the use of new tools and frameworks like Apache Hadoop and Spark, i.e. Big Data tools, in the development of scalable systems as well as preventing the computational cost from going up while processing vast amount of data (Schelter & Owen, 2012; Zhao & Shang, 2010).

This chapter provides a comprehensive and self-contained description of this research area. Such a work will constitute a milestone for studies on Big Data; since it will provide review of key references which will be useful in the search for research topics dealing with Big Data. To achieve such goal, this chapter will review the literature on Big Data and recommendation engines. In addition, survey the promising approaches of recommender systems that are expected to perform well while handling Big Data; such as Singular Value Decomposition. Furthermore, assess the applicability and viability of Big Data technologies (i.e. Apache Hadoop and Spark) to the field of recommendation system as well as develop a scalable recommender system that can handle large volume of data.

BACKGROUND

Recommender System Overview

Recommender system's (RS) main mission is to find the taste of a person and automatically suggests, new, relevant content for him/her. These suggestions aid to decision-making; for example: which item to buy, which music to listen to, or which news to read. This is achieved by finding patterns in people opinions, even though their opinions vary. These patterns are useful in predicting what a user might like or dislike. For illustration, people like something which is similar to what they liked before, or they like what others of similar taste and opinion seem to like (Owen et al., 2012; Ricci et al., 2011).

Most of recommender systems aim to provide a personalize websites by suggesting different items to different users. However, there are some recommendations which are non-personalized such as: Top

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/recommender-system-in-the-context-of-big-data/157695

Related Content

Knowledge Discovery and Big Data Analytics: Issues, Challenges, and Opportunities

Vinoth Kumar Jambulingam and V. Santhi (2017). *Web Semantics for Textual and Visual Information Retrieval* (pp. 144-164).

www.irma-international.org/chapter/knowledge-discovery-and-big-data-analytics/178371

Semantic Integrating for Intelligent Cloud Data Mining Platform and Cloud Based Business Intelligence for Optimization of Mobile Social Networks

Gebeyehu Belay Gebremeskel, Zhongshi He and Xuan Jing (2013). *Data Mining in Dynamic Social Networks and Fuzzy Systems* (pp. 173-211).

www.irma-international.org/chapter/semantic-integrating-intelligent-cloud-data/77528

A Survey of Open Source Tools for Business Intelligence

Christian Thomsen and Torben Bach Pedersen (2009). *International Journal of Data Warehousing and Mining* (pp. 56-75).

www.irma-international.org/article/survey-open-source-tools-business/3896

Hierarchical Hybrid Neural Networks With Multi-Head Attention for Document Classification

Weihao Huang, Jiaojiao Chen, Qianhua Cai, Xuejie Liu, Yudong Zhang and Xiaohui Hu (2022). *International Journal of Data Warehousing and Mining* (pp. 1-16).

www.irma-international.org/article/hierarchical-hybrid-neural-networks-with-multi-head-attention-for-document-classification/303673

Analysis of Genomic Data in a Cloud Computing Environment

Philip Groth, Gerhard Reuter and Sebastian Thieme (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 1129-1158).

www.irma-international.org/chapter/analysis-of-genomic-data-in-a-cloud-computing-environment/150209