

Chapter 14

Intelligent Big Data Analytics: Adaptive E-Commerce Website Ranking Using Apriori Hadoop – BDAS-Based Cloud Framework

Dheeraj Malhotra
University of Kota, India

Om Prakash Rishi
University of Kota, India

Neha Verma
I. K. Gujral Punjab Technical University, India

Jatinder Singh
DAV University, India

ABSTRACT

With the explosive increase in regular E Commerce users, online commerce companies must have more customer friendly websites to satisfy the personalized requirements of online customer to progress their market share over competition; Different individuals have different purchase requirements at different time intervals and hence novel approaches are often required to be deployed by online retailers in order to identify the latest purchase requirements of customer. This research work proposes a novel MR apriori algorithm and system design of a tool called IMSS-SE, which can be used to blend benefits of Apriori-based Map Reduce framework with Intelligent technologies for B2C E-commerce in order to assist the online user to easily search and rank various E Commerce websites which can satisfy his personalized online purchase requirement. An extensive experimental evaluation shows that proposed system can better satisfy the personalized search requirements of E Commerce users than generic search engines.

INTRODUCTION

In a short span, less than 10 years, the shopping process has become modified enormously. This is largely due to the magnificent growth in web based shopping portals and websites, Modern customers prefer to shop online because of busy life style, easy availability of Internet, high computer literacy rate. Other attractive offers like easy exchange, cash back, cash on delivery and feedback availability like reliable offers are frequently available. However searching for an appropriate E-Commerce websites which best

DOI: 10.4018/978-1-5225-2599-8.ch014

suits the customer requirements are still not so easy. Most of the online users are dependent on search engines to search for E-Commerce web site. When the same query is searched by different users, even a state of art search engine returns the same result, irrespective of the user submitting query. The search engines tends to return the results by interpreting the customer's query in all possible ways, the situation gets worse, if the query is incomplete or ambiguous. For example, for the incomplete search query *Orange*, some users may be interested in links to buy a new postpaid connection of *Orange* company, while others may be interested in searching documents for a fruit. Hence adaptive search system is required, which may intermediately modify the search query by keeping track of customer's preferences over a period of time and return results in correct order of ranking of output links to best match a customer's requirements.

Online generated data is explosively increasing on daily basis in the scale of many Giga Bytes. This is due to increased web traffic. For example to purchase an item online, user explores many links to search for good E-Commerce website which provides branded quality material at best possible discounted price. As a result, many shopping portals getting bulky data on daily basis like *Wal-Mart*, which handles more than 1 million customer transaction logs per hour, resulting into PB of data generated each day. This inordinate online generated data may be called as '*Big Data*' with emphasis on high values of 5 V's i.e. Volume, Variety, Velocity, Veracity and Value. Big Data is a term applied to those data sets having size, range of data sources and speeds of in/out beyond the capabilities of traditional relational databases to process and manages. This '*Big Data*' contains useful patterns which are never explored and advanced analytics is required to unhide these patterns. These extracted patterns are helpful for E -Commerce websites which they may apply in useful decision making process like market basket analysis to increase sales by exploring customer purchase patterns, improved inventory management to avoid situations like out of stock or overstock. This may be done by identifying unexpected sales trends from various sources like social media, previous transactions etc. For example a company named *Tesco* (British multinational grocery, United Kingdom) has implemented effective strategies to promote Market Basket Analysis i.e. Association Mining by Loyalty Card Program, under this program, Tesco mines shopping habits of millions of families which helps company in taking various important decisions like which items to be put at sale, BOGO (Buy One, Get One) offers and assortment of number of items to be offered as a discounted package etc.

Big Data analysis for market basket analysis on E-Commerce website data can be easily accomplished by employing Apriori-Hadoop-BDAS (Berkley Data Analysis Stack) framework, which is a popular, scalable and robust open source platform for processing Big Data. Hadoop can be used for writing efficient applications to process huge amount of data. Hadoop cluster has many parallel machines which can easily store and process huge data sets. Different clients may submit their jobs to distributed Hadoop cluster from distant locations. MapReduce programming model may be used to process data in Hadoop cluster with the help of two functions known as Map and Reduce to process Big Data in (Key, Value) pair format. Hadoop and Map Reduce based cloud framework may be used for deployment of Big Data based personalized E-Commerce website ranking system.

The overall objective of this chapter is to assist the customer while carrying out online purchase transactions to buy authentic and rationally priced products and E-Commerce website organization to optimize the structure of its website as to take advantage over its competitor E-Commerce websites.

16 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/intelligent-big-data-analytics/183290

Related Content

Towards Scalable Certificate Status Validation in Mobile Ad Hoc Networks

Mohammad Masdari, Javad Pashaei Barbinand Moazam Bidaki (2013). *International Journal of Mobile Computing and Multimedia Communications* (pp. 35-48).

www.irma-international.org/article/towards-scalable-certificate-status-validation-in-mobile-ad-hoc-networks/103968

Combined Queue Management and Scheduling Mechanism to Improve Intra-User Multi-Flow QoS in a Beyond 3,5G Network

Amine Berqia, Mohamed Haniniand Abdelkrim Haqiq (2012). *International Journal of Mobile Computing and Multimedia Communications* (pp. 57-68).

www.irma-international.org/article/combined-queue-management-scheduling-mechanism/63051

Combining Static Code Analysis and Machine Learning for Automatic Detection of Security Vulnerabilities in Mobile Apps

Marco Pistoia, Omer Trippand David Lubensky (2017). *Mobile Application Development, Usability, and Security* (pp. 68-94).

www.irma-international.org/chapter/combining-static-code-analysis-and-machine-learning-for-automatic-detection-of-security-vulnerabilities-in-mobile-apps/169677

Improving Energy Efficiency and Throughput in Heterogeneous Mobile Ad Hoc Networks

Manu.J. Pillaiand M. P. Sebastian (2009). *International Journal of Mobile Computing and Multimedia Communications* (pp. 48-60).

www.irma-international.org/article/improving-energy-efficiency-throughput-heterogeneous/4069

Special Features of Mobile Advertising and Their Utilization

Jari Saloand Jaana Tähtinen (2009). *Mobile Computing: Concepts, Methodologies, Tools, and Applications* (pp. 1878-1885).

www.irma-international.org/chapter/special-features-mobile-advertising-their/26633