A Review on Big Data: Recommendation Systems

Prarthana Rao H M,
M.Tech, IV sem
Department of Computer Science,
JNNCE, Shimoga,

Sathyanarayana S
Assistant Professor
Department of Computer Science,
JNNCE, Shimoga

Abstract- Recommendation is the key to success for discovering and retrieval of content in this era of huge data. It is the proven use case of BigData. This paper presents an overview of the field of recommender systems. The term 'Big Data' describes innovative techniques and technologies to capture, store, distribute, manage and analyze petabyte or larger-sized datasets with high-velocity and different structures. Improve the quality of recommendations for users. Challenges faced by these recommendation systems. So less time an algorithm spends searching for neighbors, the more scalable it will be and worse its quality.

Keywords---BigData, Hadoop, MapReduce, Recommendation system

I. INTRODUCTION

We have entered a new phase of the information era, one in which organizations query and analyze huge volumes of diverse data in real-time or near real-time to improve outcomes for their most critical business processes. The amount of information in the world is increasing far more quickly than our ability to process it. Every one of us is unique, but at the same time there's a lot that are damned similar to each one of us, exhibiting the same behavior, interacting with the same people, liking the same things. But it's not necessarily a bad thing; we can enjoy the benefits of the so called collective intelligence, which is embedded in a lot of applications we use on a daily basis. User need recommendations they can trust to help them find items they like.

II. BIGDATA,BIG ANALYSIS

With the increasing production of information from various initiatives, there is also the need to transform a large volume of unstructured data into useful information for society. All this information should be easily accessible and made available in a meaningful and effective way in order to achieve semantic interoperability in electronic services. The concept of variety often underlies use of the term big data. Big Data refers to technologies and initiatives that involve data that is too diverse, fast-changing or massive for conventional technologies, skills and infra- structure to address efficiently. It's about the ability to make better decisions and take meaningful actions at the right time.

Innovations in technology and greater affordability of digital devices have presided over today's Age of Big Data, an umbrella term for the explosion in the quantity and diversity of high frequency digital data. The world is filled with data. Several other authors often refer to the "Three V's" of big data [6]: volume, variety, and velocity, originally discussed by Laney in 2001, to distinguish big data. Volume refers to the actual size of the dataset(s) analyzed, variety to the various types of datasets possibly combined to produce new insights, and velocity to the frequency with which data is recorded and/or analyzed for action. When making an attempt to understand the concept of Big Data, the words "MapReduce" and "Hadoop" cannot be avoided [5].

III. MAPREDUCE

Mapreduce is a huge hit. Google faced with making sense of the largest collection of data in the world and took on this challenge [5,4]. The result was MapReduce: a software framework that breaks big problems into small, manageable tasks and then distributes them to multiple servers. Google makes very heavy use of MapReduce internally, and the Apache Software Foundation turned to MapReduce to form the foundation of its Hadoop implementation. MapReduce can work with raw data that's stored in disk files, in relational databases, or both. The data may be structured or unstructured, and is commonly made up of text, binary, or multi-line records. The most common MapReduce usage pattern employs a distributed file system known as Hadoop Distributed File System (HDFS).

Data is stored on local disk and processing is done locally on the computer with the data. The Hadoop Distributed File System (HDFS) is based on the Google File System (GFS) [1] and provides a distributed file system that is designed to run on commodity hardware. HDFS holds very large amount of data and provides easier access. To store such huge data, the files are stored across multiple machines. These files are stored in redundant fashion to rescue the system from possible data losses in case of failure. HDFS also makes applications available to parallel processing.

At its core, MapReduce is composed of two major processing steps: Map and Reduce.

1. The Map task takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key-value pairs).
2. The Reduce task takes the output from the Map as an input and combines those data tuples (key-value pairs) into a smaller set of tuples. The reduce task is always performed after the map job.

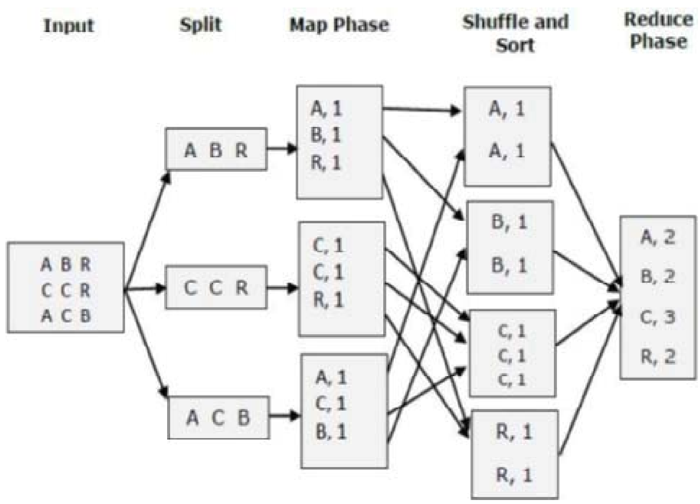

Figure 1 . Overall MapReduce wordcount process

IV. RECOMMENDATION SYSTEM

Proven Big Data Use Case is the Real-Time Recommendation System. Now days on the web, vast amount of information is present, so it is difficult to user to find relevant information. Big Data has concern for the Data that is used for Recommendation and Filtering systems. Earlier Benchmarking was set as a standard for Big Data [11]. Benchmarking often needs an educated decision maker at the end to explain why to benchmark, benchmarking products often are not scalable. Bigger the data, leads to the requirement of educated decision makers. Recommendations narrow a complex decision to just a few recommendations. Recommendation systems can solve this problem. Recommendation systems have impacted or even redefined our lives in many ways.

Big Data allowed us to do recommendations on a new scale that we did not see before [8, 3, 10]. The most well-known example is how the Google search algorithm trumped Altavista by recommending the best websites to view. Another well-known example is the recommendation from Amazon based on the reading behavior from other readers. Both of those systems are based on algorithms that “learn” from past data. A recommendation system outdoes benchmarking because it does not need an analyst at the end. A recommendation system suggests a few data points out of a large pool of data. Logically, the world of startups is filled with companies doing recommendation products in one way or the other.

Recommendation engines aim to predict user preferences based on historical activity and implicit or explicit feedback. They enable applications to present the most relevant information to users. In an effort to strengthen the relevance of the recommendations that is, to increase the likelihood that a customer would take action we realized the need to augment a standard recommendation engine with contextual information. This information could help us match the recommendation to tendencies inferred from historical activity. The real-time recommendation system, in particular, offers potentially high value across virtually all industry sectors. It also has relatively sophisticated infrastructure requirements. Real-time recommendation systems were pioneered by early Internet companies, such as Amazon, Google, and Netflix.

Recommender systems use the opinions of a community of users to help individuals in that community more effectively identify content of interest from a potentially overwhelming set of choices. One of the most successful technologies for recommender systems, called collaborative filtering, has been developed and improved over the past decade to the point where a wide variety of algorithms exist for generating recommendations. Each algorithmic approach has adherents who claim it to be superior for some purpose.

V. HOW RECOMMENDATION SYSTEM WORKS?

A Recommendation system works in well-defined, logical phases which are data collection, ratings, and filtering [10].
Ratings: Ratings are important in the sense that they tell you what a user feels about a product. User’s feelings about a product can be reflected to an extent in the actions he or she takes such as likes, adding to shopping cart, purchasing or just clicking. Recommendation systems can assign implicit ratings based on user actions.
Filtering: Filtering means filtering products based on ratings and other user data. Recommendation systems use three types of filtering: collaborative, user-based and a hybrid approach. In collaborative filtering, a comparison of user’s choices is done and recommendations given. In user-based filtering, the user’s browsing history, likes, purchases and ratings are taken into account before providing recommendations. Many companies also use a hybrid approach. Netflix is known to use a hybrid approach. Collaborative filtering approaches build a model from a user’s past behavior (items previously purchased or selected and/or numerical ratings given to those items) as well as similar decisions made by other users; then use that model to predict items (or ratings for items) that the user may be interested in [8].

VI. THE AMAZON USE CASE

In the field of recommender system development has grown from a couple of dozen researchers in mid-1990s to hundreds of researchers today-working for universities, the large online retailers, and dozens of other companies whose sole focus is on these types of systems. Amazon has been in certain ways a pioneer of ecommerce but more important than that accolade is how it is driving its revenue up by providing more and more effective recommendations. Buying can be both impulsive and planned and Amazon is smartly tapping into the impulsive shopper’s mind by providing relevant and useful product recommendations. For that, it is relentlessly working on making its Recommendation engine more powerful. Shopping has a connection with psychology. Shoppers buy for instant gratification, instant mood uplift, social esteem and reasons not even known to them clearly. Amazon is smart enough to take these factors into account.

When Amazon recommends a product on its site, it is clearly not a coincidence [13]. At root, the retail giant’s recommendation system is based on a number of simple elements: what user has bought in past, which items they have in their virtual shopping cart, items they’ve rated and liked, and what other customers have viewed and purchased. The company also doles out recommendations to users via email. The website recommendation process is more automated.

Figure 2. Example of ratings in Amazon

Unlike Facebook which also relies a lot on big data which knows a lot of details about its subscribers, all Amazon knows about its customers are the spending patterns. Amazon has been cashing on this knowledge smartly in an attempt to get more out of your pockets. It is a difficult job to analyze spending patterns, likes, product preferences and provide effective recommendations just on that basis. And now, Amazon is trying to make available its tools and technologies that use big data and Recommendation systems so effectively for sale to other corporations that use big data. So, Amazon's product ads will start to appear more frequently on other websites as well and that is going to drive up sales.

VII.CHALLENGES IN RECOMMENDATION ENGINES

Cold Start Problem: The heart of a recommendation system is that a computer learns from data [9]. One of the biggest challenges can be that there is not sufficient historical data at the start.

No Surprises: If there were sufficient data and if recommendation engines, executed badly there might be no surprises [11]. Evaluating recommender systems and their algorithms is inherently difficult for several reasons. First, different algorithms may be better or worse on different data sets. The second reason that evaluation is difficult is that the goals for which an evaluation is performed may differ. The two main industries that at this moment benefit strongly from recommendation engines are the retail industry and the media industry because both have a lot of data in the long tail, and both have a lot of data to overcome the cold-start problem.

Sentiment analysis: Another challenge relates to sentiment analysis (or opinion mining). The term refers to "the computational study of opinions, sentiments and emotions expressed in text" that aims at "translating the vagaries of human emotion into hard data." Scraping blogs and other social media content has become a common undertaking of corporations and academic researchers. Sentiment analysis aims at finding out and quantifying whether, how many, and how strongly people are happy vs. unhappy, pessimistic vs. optimistic, what they like or dislike, support or reject, fear or look forward to something and any "shades of grey" in between.

Overall, the fundamental challenge is getting to the true intent of a statement, in terms of polarity, intensity, etc. Many obstacles may impede this, from the use of slang, local dialect, sarcasm, hyperboles, and irony, to the absence of any key words [1]. These are, in a sense, technical or measurement challenges that become easier to handle as the degree of sophistication of sentiment analysis algorithms improves. But the conceptualization and classification phase of any analysis which is to be conducted is non-trivial. This implies deciding, for instance, whether what matters is frequency or presence of key word(s). Thus, the human analyst's input is critical. Classification is one of the most central and generic of all our conceptual exercises. Without classification, there could be no advanced conceptualization, reasoning, language, data analysis, or, for that matter, social science research.

VIII.CONCLUSIONS

Recommender systems are a powerful new technology for extracting additional values for a business from its user data-bases. These systems allow users to select items of their interests. Also, acts a supporter to increase the business sales. Recommender systems are being dumped by large data sets day-to-day of the user's data available on web. So, a technique is required to improvise the scalability of these recommendation systems. This paper, reviewed on various limitations of the current recommendation methods. Thus we need to create technologies that can help us shift through all available information to find that which is most valuable to us. As more and more information is being created and collected, and analytic capabilities continue to advance, companies can generate real value through recommendation systems. Hence, doing it correctly will help to maximize that value and protect the all-important corporate brand and reputation.

REFERENCES

[1] Abhinandan Das, Mayur Datar, Ashutosh Garg, "Google News Personalization: Scalable Online Collaborative Filtering", Industrial Practice and Experience, May 8-12, 2007.

[2] Archana.T.Mulikl, S.Z.Gawali, "Recommendation system: Online movie store", International Journal of Application or Innovation in Engineering & Management (IJAIEM) , Volume 2, Issue 6, June 2013.

[3] G. Adomavicius, and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and possible Extensions", IEEE Transactions on Knowledge And Data Engineering, Vol 17(6), June 2005.

[4] Diana MacLean , "A Very Brief Introduction to MapReduce", CS448G, 2011

[5] Fatima EL Jamiy, Abderrahmane Daif, Mohamed Azouazi, Abdelaziz Marzak, "The potential and challenges of Big data - Recommendation systems next level application".

[6] Harshawardhan S. Bhosale1, Prof. Devendra P. Gadekar, "A Review Paper on Big Data and Hadoop", International Journal of Scientific and Research Publications, Volume 4, Issue 10, October 2014.

[7] Intel Reference Architecture Big Data Analytics, "Predictive Analytics and Interactive Queries on Big Data", White Paper.

[8] Jai Prakash Verma, Bankim Patel, Atul Patel, "Big Data Analysis: Recommendation System with Hadoop Framework", IEEE International Conference on Computational Intelligence & Communication Technology, 2015.

[9] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, And John T. Riedl, "Evaluating Collaborative Filtering Recommender Systems", ACM Transactions on Information Systems, Vol. 22, No. 1, Pages 5–53 January 2004.

[10] Kaushik.Pal, "How Big Data is used in Recommendation Systems to change our lives", kdnuggets.com

[11] Lutz Finger, "Recommendation Engines: The Reason Why We Love, Big Data", forbes.com.

[12] Zhi-Dan Zhao, Ming-Sheng Shang, "User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop", Third International Conference on Knowledge Discovery and Data Mining, 2010.

[13] https://www.tutorialspoint.com/mapreduce-parallel&distributed programming model.