Advanced Science and Technology Letters
Vol.94 (Networking and Communication 2015), pp.36-41
http://dx.doi.org/10.14257/astl.2015.94.09

Customized Efficient Collection of Big Data for Advertising Services

Jun-Soo Yun[1], Jin-Tae Park[1], Hyun-Seo Hwang[1], Il-Young Moon[1]

[1] 1600 Chungjeol-ro, Byongcheon-myon, Dongnam-gu, Cheonan-si, Chungcheongnam-do, 330-861 Republic of Korea
{yuntn55, wlsxo05, smilebear1, iymoon}@koreatech.ac.kr

Abstract. Because of the advent of smartphones and Internet development, large quantities of data are communicated indiscriminately, thereby generating rapid increases in data traffic. Particularly in the context of social commerce, with its enormous amounts of information, the user finds it difficult to locate information needed. Even for a company that provides this service, it can be difficult to ascertain the information that is needed by the user. Thus, other companies have arisen to provide the service of helping those companies extract the necessary information from the user data in order to meet the user's needs. In this paper, we describe an efficient method of collecting big data for use by users of customized advertising services.

Keywords: Big Data, Social Commerce, Data Traffic, Smartphone

1 Introduction

The birth of the Internet inaugurated the modern explosion of primary data. The Internet provides the framework for the world's information exchange; within this framework, information is replicated and can be indiscriminately mass produced, providing secondary data. This data expansion has created a surge in customer-related data associated with communication and trade channels, and these data play a role in data processing and analysis by companies.

Now, we are experiencing an explosion of secondary data [1]. The cause of this new data explosion is the generation of data associated with the diversification and communication of the sources of primary data. Social networks and social commerce arose, beginning with Facebook, because the heart of the Internet is big data, and mass-produced data about relationships between users enabled applications such as Twitter [2]. As of 2011, the amount of digital information that had been generated worldwide was 1.8 zettabytes. 1.8 zettabyte's amount to be issued three statements on Twitter every minute without rest all of the people Republic of Korea 18 million years. In this way, a great deal of information and inestimable numbers of data are produced, diffusing into the digital economy [3], [4]. Furthermore, because of existing data management and analysis methods, unaffordably immense quantities of data are generated. Therefore, as the quantity of data increases rapidly, not only are techniques

ISSN: 2287-1233 ASTL

for classifying data required, but also needed are services that can easily find data for users on the Internet [5].

In developed countries such as the United States and Europe, with their big data companies, interest is already rising and investments are increasing in order to obtain useful information of value to society and humanity. Companies can analyze user preferences, making it possible to target their marketing; thus, it is necessary to improve the quality of service [6]. Therefore, in this paper, we describe how to efficiently collect the myriad data needed for big data analysis.

2 Related Research

2.1 Amazon

Amazon.com is an international e-commerce company headquartered in the United States (in Seattle, Washington); it is the largest online shopping brokerage company in the world. Amazon started as an online bookstore and advanced to various other fields through a business expansion in 1997. For big data analysis in Amazon, cloud computing technology is used. Cloud computing integrates IT resources with a very large cloud server to provide computer resources to individuals, who only require online access to an Amazon cloud drive account to utilize it. Service is provided only to registered devices, and it is possible to register up to eight devices, regardless of the type of device. A web browser can connect to any device, but it is also possible to be connected to the storage devices via Android applications. Amazon has maintained the highest level of customer satisfaction in the e-commerce industry through its thoroughly customer-centric business development.

Amazon Web Services (AWS) allows any company to save information and to run software on Amazon's computer. Companies pay only for the capacity used when using cloud services. This system is flexible enough to deliver reliable, fast response, even with the tremendous variation in traffic that exists.

2.2 Google

Google is a U.S. multinational company whose main business areas are cloud computing and advertising. Google's big data analysis service is called the "Google Cloud Dataflow" (GCD). GCD resides in Google's cloud platform and is linked to the available big data analysis services, and it can process a large quantity of data in either "batch mode" or "streaming mode" format. The basic concept of the GCD is that it handles the big data optimization, distribution, and scheduling for the user, who remains in charge of all the peripheral functions such as monitoring the GCD, in exchange for being able to concentrate on big data analysis applications.

2.3 eBay

eBay is a multinational company in the United States that provides C2C services via the Internet. eBay can take advantage of predictive analytics using SAP HANA, thereby expanding as a seller of business and reducing the burden of its business analysts, who are working to drive the growth of the eBay marketplace. SAP HANA is a platform for real-time analysis and applications, a next-generation solution for real-time business. Data can be queried and analyzed by all users, enabling support for business insights that require a large quantity of data containing detailed information starting from the moment a business transaction occurs. SAP HANA may be 500 index the best model and analyzed automatically select, with 100% accuracy and 97% reliability, and to determine the actual positive responders. Thanks to eBay analysts, their automated system can reliably detect the best time to invest in a strategic business, so that eBay can allow sellers to focus on the items that buyers need.

3 Main Discourse

In order to efficiently collect big data, we have developed a mobile social advertising platform that sends advertisements and coupons to the surrounding commercial area, based on information about users' locations. Users can use the social advertising platform to collect data for multiple users, analyze the results, and provide customized service based on their findings. In order to collect the data, tables in the database store interests of each user, allowing views that are designed to be configured (see Fig. 1). By configuring the tables in the database, it is possible to instruct it to store only the data that are of interest, thereby shortening the time needed for further analysis as well as reducing the space needed to store the data.

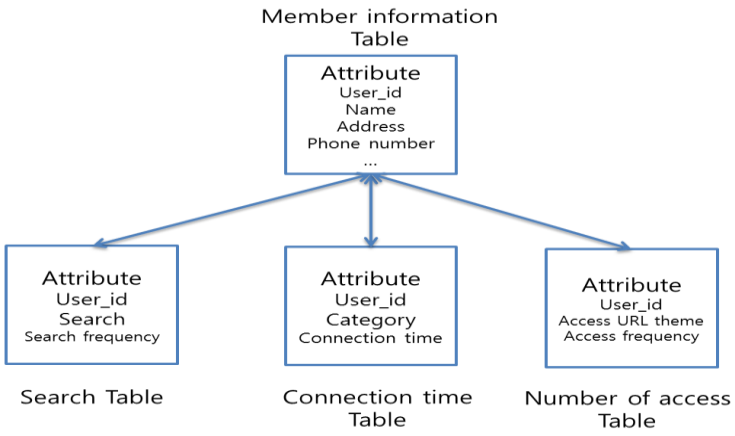

Fig. 1. Database design.

3.1 Collection of User Queries

One way to learn the simplest interest of the user is to collect and analyze user search queries. By using the search keywords collected, it is possible to simplify the data mining using big data analysis techniques. Data mining of the stored data can be performed on a large scale, systematically and automatically finding the statistical rules and search query patterns of each user by searching the common words stored in the database. The extracted words, in themselves, reflect the interests of the user. Additionally, by measuring the frequency of search queries, a user's interests and conclusions can be determined with high confidence.

3.2 Collection of User Connection Durations

By collecting the lengths of time users are connected to their subjects of interest in social commerce, it is possible to gain an understanding of their interests. In using the ASP, they also record the start time and the disconnection time for the session in which they accessed areas of interest. Subtracting the start time from the end time, it is possible to measure the time the user spent connected to that category. Only by measuring the connection time is it possible to determine the users' interest. If a category attracts long stays, it means that users are finding more things of interest there.

3.3 Collection of User Access Counts

By analyzing the number of accesses in the general access log, it is possible to measure how many users are accessing a page. Thus, by saving the user and the page

that was accessed, it is possible to ascertain a user's interest by extracting the theme of the stored page. In this manner, a session can be analyzed according to the number of accesses. However, if the page theme is recorded for the session, it is not necessary to save the actual database or the page URL; the subject matter has been registered in the session, along with the number of connections, allowing us to sort these topics according to their frequency of access. The more frequently a subject is accessed, the more interest the user has demonstrated in that subject.

4 Conclusion

Currently, as greater and greater amounts of information are exchanged over the Internet, a large quantity of data is rapidly enough never to understand the development of smart phones and wireless Internet. This allows users to obtain information through the Internet, though from the company's perspective, it is a difficult challenge to find the data necessary to provide the information, i.e., to provide a service that matches the user with the information requested. The surge of social commerce in particular points out that "hard to use" is out by complex advertising. Therefore, a need arose to provide services to meet users' needs, and hence the term "targeted marketing" appeared. In order to fulfill this demand to provide customized services to analyze users' preferences, it is necessary to have an accurate understanding of their interests. In this paper, we have proposed a scheme for collecting only the user data necessary to provide a specific service. The aim is to collect the data that allow us to grasp the users' exact interests by analysis of connection counts, connection times, and most frequently accessed topic (search query or web page or category). By collecting only those data that match the purpose, the phenomenon of data overload is avoided, and it is expected that more accurate analysis can be achieved.

In order to improve the accuracy of intensive analysis, we intend to investigate additional types of collection methods. As data traffic is reduced, it is expected that user friendliness will increase.

Acknowledgments. This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE) and the Korea Institute for Advancement of Technology (KIAT) through the Research and Development for Regional Industry.

References

1. Y. H. Yong.: The advent of an era of big data, Korea Institute of Information Technology. vol. 10, no. 3, pp. 39-43, (2012)
2. Y. I. Yun, Svetlana Kim.: Big Data and Cloud Era, Korea Information and Communications Society. vol. 30, no. 4, pp. 3-6, (2013)
3. S. H. Lee.: Utilization of big data, Korea Institute of Information Technology. vol. 10, no. 3, pp. 51-54, (2012)
4. S. H. Lee.: A Big Data Example of Support Public Services, Korea Institute of Information Technology, Proceedings of KIIT Summer Conference, (2013)

5. C. W. Ahn, S. K. Hwang.: Big Data Technologies and Main Issues, Journal of Information Department, vol. 30, no. 6, pp. 10-17, (2012)
6. S. W. Choi, H. Y. Kim, Y. K. Kim.: Research Trend of Bigdata Technology and Analysis Technique, Korea Information Science Society, vol. 39, no. 2, pp. 194-196, (2012)

