Big Data Based Logistics Data Mining Platform: Architecture and Implementation

Fei Gao, School of Economics and Management, Beijing Jiaotong University, Beijing, China
Qilan Zhao, School of Economics and Management, Beijing Jiaotong University, Beijing, China

ABSTRACT

With the development of intelligent logistics, enormous amount of logistics data are becoming one of the sources of big data. Building the logistics information platform with big data mining and analysis capabilities to make full use of the huge logistics data is the inexorable trend for intelligent logistics. This paper studied the characteristics of the logistics big data, then, a big data based logistics data mining platform is designed and implemented by utilizing big data processing and storage techniques. The architecture and functions of the platform will be described in detail. This paper also studied the mining steps and requirements for logistics data mining, which is significant for practical applications.

Keywords: Architecture, Big Data, Data Mining, Implementation, Logistic, Platform

1. INTRODUCTION

Logistics is the flow process of goods (both real and virtual) from supplier to recipient. Generally speaking, it includes the following aspects: transportation, storage, loading and unloading, packing, distribution processing, information processing, etc. Logistics management is helpful for optimizing the relationship of the supply chain, reducing the costs and improving competitiveness of the enterprise (Cambra-Fierro, J., 2009). Therefore, it is always an important subject studied by the government and enterprises. So far many measures have been taken to improve the efficiency of logistics management. One of these measures is logistics information platform (LIP), which provides timely and effective logistics information for suppliers, commodity producers and consumer. Recently, Intelligent Logistics (IL) has been proposed to make the logistic information system to be more intelligent and more efficient by utilizing the Internet of Things (IOT) and intelligent information processing technologies (Waller, M. A., & Fawcett, S. E. 2013).

On a large regulated logistic information platform, amounts of data are collected and updated every day. With the development of
IL, logistics information has become one of the sources of big data. However, rough logistics information, such as freight route, transportation providers as well as simple statistics of merchandise sales, cannot provide the underlying information behind the logistics information (Fu, P., & Gu, X., 2012). For example, as the statistics of logistics information platform of Sichuan Province shows, its ‘total express delivery in May was 26,213,800, with an increase of 48.38% compared with last year, the provincial highway freight volume is 124 million tons, with an increase of 8.3%.’ In fact, what we more concerned is a deeper interpretation of the data, for example, where exists a market opportunity, which product customers prefer to buy, what is the developing trend of the market. Current logistic information platform is facing an embarrassing situation that we have enormous data, but we don’t know how to use them (Chen, H., Chiang, R. H., & Storey, V. C., 2012).

For logistics management, logistics data mining platform with big data processing capacity will be an important innovation to make full use of logistics information to promote the development of logistics (Wang, Z.T., 2015). This paper aims at proposing and designing a big data based logistics data mining platform to extract the knowledge and rules underlying logistic big data to provide effective decision references for users. Start at studying the characteristics of logistics big data, we proposed a logistic data mining platform combined with big data processing techniques. The platform’s architecture, functions will be presented in detailed. In addition, the data mining process and requirements will also be illustrated in the paper.

2. CHARACTERISTICS OF LOGISTICS BIG DATA

Big data technologies are developed when it is hard to the deal with the data with too huge volume and complexity for conventional data processing technologies. Currently, there is no a unified definition for big data. In general, big data is referred as the data with huge volume, high variety, changing velocity. Nowadays the global economy’s continuing growth and the development of network technique make logistics information becomes one of the sources of the big data, such as commodity price fluctuations, information of commodities, customers’ reviews and customer behaviors of online shopping sites. Combined with the HACE theory in (Wu X. 2014), the characteristics of logistic data can be concluded as following:

In the perspective of data, logistics big data has the characteristics of huge volume, heterogeneous attributes and diverse dimensionality. The ‘huge’ here may not mean the logistic big data occupy a huge storage space, such as TB level. Most of the big data are automatically generated by machines, but most logistics big data are generated by human’s production activities and most of them are organized as structured data or semi-structured data. In many actual applications, the ‘huge’ of big data here means extremely larger amounts of data entries. For example, the daily site visits of America’s Amazonas website is about 15 million and Taobao’s daily site visits is even more striking: about 100million hits. To analysis these logistic data, it doesn’t need very large storage space for the sales record of these websites, but it is impossible without big data processing techniques.

Heterogeneous attributes mean an object can be characterized by different data formats. For example, to describe a freight line, some data are presented in HTML web pages, some data are structured tables and some are images of maps. Diverse dimensionality means that an object is characterized the same format, but with different data entries. For instance, even if freight lines are described by structured tables, some of them may include providers’ name, departure, destination, start time, end time and prices, some others may be presented by a part of these entries, or may be more than them. Some logistic platforms provide detailed information, while others are rough and incomplete. When aggregating data with different attributes and dimensions, it is a prerequisite to transform them into the same patterns for further data processing.
Device-to-Device Communications
www.igi-global.com/chapter/device-to-device-communications/136561?camid=4v1a

True Concurrency Semantics: Towards a Verification of Timed Systems
Souad Guellati, Ilham Kitouni, Riad Matmat and Djamel-Eddine Saidouni (2014).
www.igi-global.com/article/true-concurrency-semantics/121727?camid=4v1a

Buffer Managed Multiple Replication Strategy Using Knapsack Policy for Intermittently Connected Mobile Networks
www.igi-global.com/chapter/buffer-managed-multiple-replication-strategy-using-knapsack-policy-for-intermittently-connected-mobile-networks/105669?camid=4v1a