Social media data analysis and application research in the Era of digital economy

Zhao Xuanqi，Dongbei University of Finance and Economics,

Dalian City, Liaoning Province,116023,China

Abstract: With the rapid development of digital economy and the wide application of social media, social media data has become an important resource in the digital economy era. Social media data provides new opportunities and challenges with the characteristics of large data volume, fast update speed, diverse types, high authenticity and timeliness for the development of digital economy. Social media data is of great value in insight into users' behavior preferences, mining hot social public opinions, and analyzing industry dynamics. Rational development and utilization of social media data is of great significance to assisting precision marketing, serving public governance decision-making, and guiding industrial transformation and upgrading. This paper systematically discusses the analysis and application of social media data in the era of digital economy from three aspects: the characteristics of social media data, the value of digital economy development, and the analysis technology and methods, in order to provide reference for related research.

Key words: digital economy; social media; data analytics; big data; applied research

Introduction

Introduction to digital economy is digital knowledge and information as the key factor of production, digital technology innovation as the core driving force, with modern information network as an important carrier, through the digital technology and the real economy depth fusion, constantly improve the level of economic and social digital, network, intelligent of a series of economic activities. As well as an important platform for daily communication and information sharing, social media provides massive, diverse and dynamic data resources for the development of the digital economy. In the digital economy era, social media data has become an important data source for corporate marketing, government governance, and academic research. How to effectively collect, store, analyze and use social media data is of great significance to promoting the development of the digital economy.

1. Characteristics of the social media data

1.1 Large amount of data and fast update speed

Social media users are huge and growing fast. Take Weibo as an example, as of December 2021, the monthly active users of Weibo reached 573 million, and the daily active users exceeded 250 million, generating billions of interactive information every day. According to sina Weibo's 2020 annual report, it generates an average of about 875 original microblogs per second, forwarding more than 1 billion times per day. Social media data is measured in PB and

even EB and is updated all the time. The large scale and fast update of the social media data exceeds the ability of the traditional data processing mode, and puts forward higher requirements for the data analysis technology.

1.2 Diversification of data types

Social media platforms cover a variety of data types, including text, pictures, video, and audio. Unlike traditional structured data, more than 80% of social media data is unstructured and semi-structured data. For example, Weibo data includes 140 characters of text, emoticons, pictures, short videos, live broadcasts, etc., and wechat data covers public account articles, circle of friends information, small programs, etc. Diversified data types bring challenges to the unified collection, storage and analysis of data. It is necessary to comprehensively use text mining, image recognition, audio and video analysis and other technologies to realize the fusion processing and correlation analysis of multi-modal data.

1.3 High authenticity and timeliness of the data

Social media provides a platform for users to release information, express their opinions and share their lives anytime and anywhere. Compared with traditional survey data, social media data can more truly reflect users' thoughts, emotions and behaviors, and is less affected by the outside world. For example, users' evaluation of products on shopping websites is often influenced by the scoring mechanism, while evaluations on social media are more direct and realistic. Social media information spreads fast, and hot events are fleeting, reflecting the good timeliness

characteristics of social media data. With the help of social media data, we can quickly gain insight into user needs, grasp the market opportunities, and provide strong support for business decisions[1].

2. Value of social media data for the development of the digital economy

Social media data is of great value and practical significance to the development of the digital economy in terms of improving user insight, optimizing resource allocation, and driving industrial innovation.

2.1 Insight into user behavior preferences, and help precision marketing

Social media data truly records the interactive behavior and content preferences of massive users on social platforms, and is an important basis for enterprises to carry out precision marketing. Using social media data, users' interests, interactive social interaction, media use and other behavioral characteristics can be described from multiple dimensions, and user portraits can be constructed to gain insight into the needs of target customers. On this basis, enterprises can provide personalized advertisements for different groups of people, carry out social marketing, and achieve accurate and [8] touch of advertising. Through the deep mining of social media data, Gome's "Pinshopping" social e-commerce platform understands the characteristics of users of different ages, regions and occupations, matches appropriate products and pushes corresponding promotional information, achieving a significant improvement of marketing effect.

2.2 Explore hot social public opinions and serve public governance decisions

Social media is an important channel for netizens to express their demands and participate in discussions. Paying attention to social media data helps the government to timely understand public opinion and formulate public policies scientifically. Using text mining, emotion analysis and other technologies, we can find hot topics and emotional tendencies in social media, monitor the public opinion trend of major emergencies, and warn negative public opinions. The "COVID-19 public opinion Big Data Analysis platform" developed by Tsinghua University has formed a daily epidemic public opinion through intelligent analysis of the data of Weibo and wechat platforms, providing data support for relevant departments to make decisions and achieved good results[2].

2.3 Analyze industrial dynamic changes and guide industrial transformation and upgrading

Social media is an important position for enterprises to release new product information and users to share their use experience. Social media data truly reflects the industry development trend, competitive pattern and user evaluation, and is an indispensable data source for industry analysis. Using social network analysis, potential semantic analysis and other methods, we can find the key subjects of the industrial chain, grasp the evolution trend of technology, gain insight into the advantages and disadvantages of products, and provide decision-making reference for industrial layout, technological innovation and product improvement. With the help of social media data, the new energy vehicle industry can timely grasp users' evaluation and feedback on model design, performance experience, charging

convenience and other aspects, constantly optimize products, improve services, and lead the healthy development of the industry. The big data report of social e-commerce released by Shenzhen Social E-commerce Industry Association, based on the platform data of TikTok, Kuaishou and Xiaohongshu, analyzes the development status, platform characteristics and category distribution of the social e-commerce industry, which provides useful inspiration for the strategic planning and business expansion of social e-commerce enterprises.

3. Technology and methods of social media data analysis in the era of digital economy

3.1 Data collection

Social media data collection is the basis and premise of data analysis. In the face of massive, heterogeneous and dynamically changing social media data, how to obtain data efficiently, comprehensively and continuously is a key issue. At present, there are three main data collection methods: API interface call, web page crawler and manual annotation.

API is a data access interface for social media platforms for developers. Mainstream social media platforms, such as Weibo, wechat, TikTok, etc., all provide the corresponding data API, such as Weibo OpenAPI, wechat public platform API, etc. By calling the relevant API, structured data such as user information, social relations, and published content can be easily obtained. Compared with other acquisition methods, API acquisition data format standard, high quality and easy subsequent processing. However, the API generally limits the data access volume and frequency, and can only

obtain part of the data, which is difficult to meet the needs of full data collection[3].

Web crawler is a way to capture web page data by simulating the user's browser behavior. Compared with API collection, crawlers can obtain more comprehensive data, including user dynamics, comments, user homepage and other page data. Using Python, Java and other languages, customized crawler programs can be written to collect data according to research needs. However, the crawler collection also faces some challenges. On the one hand, crawlers are often restricted by anti-climbing, and need to take IP agent, verification code identification and other measures. On the other hand, the data collected by crawler is often not standard format, mixed with a lot of redundant information, and the data is noisy, which requires a lot of energy into data cleaning. In the process of crawler collection, attention should also be paid to abide by laws and regulations and platform agreements to avoid infringing on the rights and interests of others.

For the collection of unstructured data such as pictures and videos, manual annotation is usually needed. Manual annotation can make up for the limitations of algorithm identification and obtain high-quality training data. However, manual labeling is expensive and inefficient, making it difficult to cope with massive social media data. As an emerging labeling method, crowdsourcing labeling can significantly improve the labeling efficiency and reduce the cost of annotation by distributing tasks to online crowdsourcing workers.

3.2 Data storage

Social media data, with its large volume, diverse types and fast update, brings great challenges to data storage. On the one hand, social media data is growing in the order of TB and PB every day, which puts forward high requirements on the capacity and expansibility of the storage system. On the other hand, social media data covers various types of text, pictures, video, audio, and most of them are unstructured and semi-structured data, which is difficult to effectively store and manage with traditional relational databases.

In recent years, non-relational database (NoSQL) has been widely used in the field of social media data storage with its flexible data model and scalable architecture. Among them, HBase and MongoDB are two typical representatives. HBase is a distributed, column-oriented NoSQL database built on top of Hadoop with high reliability, high performance, and scalability. HBase uses a type K-V data model to organize the data in a family of columns (Column Family). Data is indexed via Row Key to support fast random read and write. HBase can easily support PB level data storage by adding a linear extension of nodes. HBase is ideal for storing huge amounts of social data, such as user behavior logs and message streams.

MongoDB is a document NoSQL database that stores data in a BSON format like JSON. Compared to relational databases, MongoDB does not need a predefined table structure, and fields can be dynamically added and deleted. This "no-mode (Schema-less)" data model is ideal for storing changing social media data. MongoDB Support for secondary index, full-text search, geographic location index and other rich query functions, can meet the diversified data access needs. MongoDB Provide native sharding (Sharding) function to achieve load balancing

and horizontal expansion through shard cluster.

3.3 Data processing

Although social media data contains rich value, it has uneven quality, serious noise and redundancy, which brings great difficulties to data analysis. In order to extract valuable information from the massive and messy raw data, a system science method is needed to process and transform the data. Data cleaning, data fusion and data analysis are the three core links of social media data processing[4].

Data cleaning is the process of detecting and correcting identifiable errors in the data file. Social media data is full of errors, and common problems include: incomplete data, data duplication, data inconsistency, data format errors, etc. For example, the personal information provided by users is often incomplete; and advertising; the format of social data from different sources is different, and it is difficult to compare directly compared. If these "dirty data" go directly into the analysis without processing, it will seriously affect the accuracy and reliability of the analysis results.

There are two main methods of data cleaning: manual identification and algorithm identification. With expert experience, we set identification rules and write SQL statements or regular expressions to identify all kinds of errors from the data. This way requires high field knowledge and heavy workload, so it is difficult to deal with massive data. Algorithm recognition makes full use of machine learning and data mining technology to automatically discover and correct dirty data. Common algorithms include statistical analysis, pattern recognition, clustering, and so on[5].

epilogue

As a "barometer" of mapping social and economic activities, social media data is of great significance to the development of digital economy. Based on the development needs of the digital economy, we should strengthen the development and utilization of social media data resources, improve the data analysis and processing capacity, and accelerate the construction of new advantages in the digital economy. At the same time, data security and personal privacy protection should be strengthened to maintain the health and order of the digital ecology. In the digital era, let us work together to empower the construction of the digital China with data, and light up the new future of the digital economy with wisdom.

Reference

[1] Wang Ruining. Exploration on the role and strategy of social media marketing in the era of digital Economy [J]. E-Commerce Review, 2024,13-25.

[2] Pang Xiaofei. Research on the influence of social media on the trust mechanism of social services in the context of digital economy [J]. 2023(6):53-55.

[3] Wu Jinhua, Shi Yanqing, is the Qin. The "death" of personal digital memory world: the influence factor of social media users [J]. Library Forum, 2024,44 (1): 116-123.

[4] Dong Qingling. Artificial Intelligence and Digital Diplomacy [J]. Chinese Social Science Digest, 2023 (8): 111-112.

[5] Wang Qianqian. Exploration of youth social changes in the context of Digital Technology and

Social Media —— MBTI-based analysis [J]. College Counselor Academic Journal, 2025,17(02):68-74