Harnessing the power of the general public for crowdsourced business intelligence: a survey

Bin Guo, Yan Liu, Yi Ouyang, Vincent W. Zheng, Daqing Zhang, Zhiwen Yu

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 7, 2019, accepted February 3, 2019, date of publication February 22, 2019, date of current version March 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901027

Harnessing the Power of the General Public for Crowdsourced Business Intelligence: A Survey

BIN GUO[1], (Senior Member, IEEE), YAN LIU[1], YI OUYANG[1], VINCENT W. ZHENG[2], DAQING ZHANG[3], (Fellow, IEEE), AND ZHIWEN YU[1], (Senior Member, IEEE)

[1]School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China
[2]Advanced Digital Sciences Center, Singapore 138632
[3]Département Réseaux et Services Multimédia Mobiles, Institut Mines-Télécom/Télécom SudParis, 91011 Évry, France

Corresponding author: Bin Guo (guobin.keio@gmail.com)

This work was supported in part by the National Key R&D Program of China under Grant 2017YFB1001800, and in part by the National Natural Science Foundation of China under Grant 61772428 and Grant 61725205.

ABSTRACT Crowdsourced business intelligence (CrowdBI), which leverages the crowdsourced user-generated data to extract useful knowledge about business and create marketing intelligence to excel in the business environment, has become a surging research topic in recent years. Compared with the traditional business intelligence that is based on the firm-owned data and survey data, CrowdBI faces numerous unique issues, such as customer behavior analysis, brand tracking, and product improvement, demand forecasting and trend analysis, competitive intelligence, business popularity analysis and site recommendation, and urban commercial analysis. This paper first characterizes the concept model and unique features and presents a generic framework for CrowdBI. It also investigates novel application areas as well as the key challenges and techniques of CrowdBI. Furthermore, we make discussions about the future research directions of CrowdBI.

INDEX TERMS Crowdsourced business intelligence, consumer behaviors, competitive intelligence, crowd intelligence, commercial site recommendation, brand trending.

I. INTRODUCTION

Business intelligence (BI) [1], termed early in [2], is typically performed by enterprises for data analysis of business information and support a wide range of business decisions. In traditional business intelligence, business data mainly includes internal data (e.g., firm-owned data) and external data (e.g., survey data), which can be processed by BI technologies to help develop new strategic business opportunities [3]. Common BI technologies include online analytical processing, mobile analytics, text mining, data analysis, predictive analytics, and so on.

In recent years, with the development of social media technology, the amount of user-generated data from various online social media has grown explosively, such as social networking sites (e.g., Facebook,[1] Instagram[2] and Twitter[3]), electronic commerce websites (e.g., Amazon[4] and Taobao[5]), and review websites (e.g., Yelp[6] and Dianping[7]). In general, user-generated data from different sources can be regarded as crowdsourced data, because the data is collected with the help of a large group of people. The large-scale *crowdsourced business data* (e.g., company dynamics, product and customer information) offers a unique opportunity for business intelligence. Furthermore, by analyzing customer-contributed data (e.g., clickstream data logs and reviews), we can discover customers' purchasing patterns and preferences, understanding their particular needs, and identifying new business opportunities. For example, Leskovec [4] presents that user-generated content in the form of blog posts, comments, and tweets establishes a connection between companies and consumers. Dey *et al.* [5] suggests that we can distill competitive intelligence among companies from user-contributed comments and ratings, which can help the associated companies to improve their products or adjust their business strategies.

Crowdsourced Business Intelligence (CrowdBI) is an emerging research area that builds on BI and social media.

The associate editor coordinating the review of this manuscript and approving it for publication was Hao Ji.

[1]https://www.facebook.com/
[2]https://www.instagram.com/
[3]https://twitter.com/
[4]https://www.amazon.com/
[5]https://www.taobao.com/
[6]https://www.yelp.com/
[7]https://www.dianping.com/

More specifically, it is an interdisciplinary research field that relies on the rich knowledge from marketing, computer science, and management science. A formal definition of CrowdBI is as follows: CrowdBI is data-driven, which leverages heterogeneous user-contributed data to facilitate the sensing of business dynamics, understanding of customer behaviors, and making intelligent decisions regarding business development. In particular, companies are expected to harness user-generated data to extract entities and themes, understand consumer preferences, visualize relationships and create their marketing intelligence in the business environment. For example, a marketing intelligence report can include market information about the popularity of competitors' products and services, consumer sentiments and opinions on their products and services, promotional information, and competitor dynamics [5]. Compared with traditional word of mouth, online product reviews provide a more publicly accessible information source to understand consumer perceptions and preferences, considering that such data were difficult to collect on a large scale in the offline world [6]. In addition, unlike traditional BI that analyzes a small amount of data by existing analytical tools, CrowdBI requires advanced technologies to process multi-modal user-contributed data (e.g., spatial-temporal analysis, pervasive sensing, machine learning, etc.) to obtain diversified information about business entities.

In general, based on the concept of CrowdBI, the research about CrowdBI mainly consists of three aspects: *crowdsourced data*, *advanced data processing techniques*, and *business applications*. On the one hand, crowdsourced data offers many opportunities for CrowdBI to develop various types of available applications for customers and companies. On the other hand, crowdsourced data presents numerous technical challenges to process and understand the large amount of data. This paper aims to give an overview of CrowdBI from different perspectives. In particular, we have made the following contributions:

1) Characterizing the concepts of CrowdBI, including the definition of CrowdBI, and a generic framework of CrowdBI.
2) Reviewing the major applications of CrowdBI, including customer behavior analysis, brand tracking and product improvement, demand forecasting and trend analysis, competitive intelligence, business popularity analysis and site recommendation, and urban commercial analysis.
3) Investigating the research challenges and key techniques of CrowdBI, including fine-grained data collection, crowdsourced data processing, and crowdsourced business knowledge mining.
4) Discussing CrowdBI for new economy, including mobile and IoT products, omni-channel retailing, and shared economy.

The remainder of the paper is organized as follows. In Sections II, we characterize the unique features and concepts of CrowdBI. Section III classifies various novel applications of CrowdBI, followed by the challenges and key techniques discussed in Section IV. Our insights and future research directions are discussed in Section V. Finally, we conclude the paper in Section VI.

II. CHARACTERIZING CROWDSOURCED BUSINESS INTELLIGENCE

In this section, we first introduce the definition of crowdsourced business intelligence (CrowdBI) and its concept model. We then present a generic framework of CrowdBI.

A. THE DEFINITION OF CROWDSOURCED BUSINESS INTELLIGENCE

We compare the differences between CrowdBI and closely-related concepts.

- ***Business intelligence (BI)***. BI is "*a set of methodologies, processes, architectures, and technologies that transform raw data into meaningful and useful information used to enable more effective strategic, tactical, and operational insights and decision-making*" [7]. In general, BI consists of three components: *the data*, *technologies*, and *applications,* which is typically performed by enterprises to present complex internal and competitive information to planners and decision makers [1].
- ***Crowdsourcing***. Crowdsourcing represents *the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call* [8]. Generally, most crowdsourcing tasks are viewed as participative activities based on online platforms, such as natural language understanding, image labeling, speech transcription, software development, and so on. In recent years, crowdsourcing is becoming a popular way to collect a large amount of data with the help of a large group of people, known as *crowdsourcing data collection*.
- ***Mobile crowdsourcing (MCS)***. MCS is a term that *describes crowdsourcing activities that are processed on smartphones or other mobile devices* [9]. Different from traditional crowdsourcing which usually obtains online data, with the aid of mobile devices, MCS can be used to collect spatio-temporal data in the real world. For example, people can capture pictures about the queuing status of a popular restaurant using smartphones and share the information in an MCS platform. A variety of novel MCS applications appear are enabled, including urban dynamics mining, public safety, traffic planning, and so on [10].
- ***Crowdsourced business intelligence (CrowdBI)***. CrowdBI is a new form of BI and it is presented in a data-driven manner. The data are collected from various crowdsourced data sources, including both online social media in the cyber space and pervasive sensing data in the physical space. All crowd-contributed business data are analyzed by a series of advanced machine learning

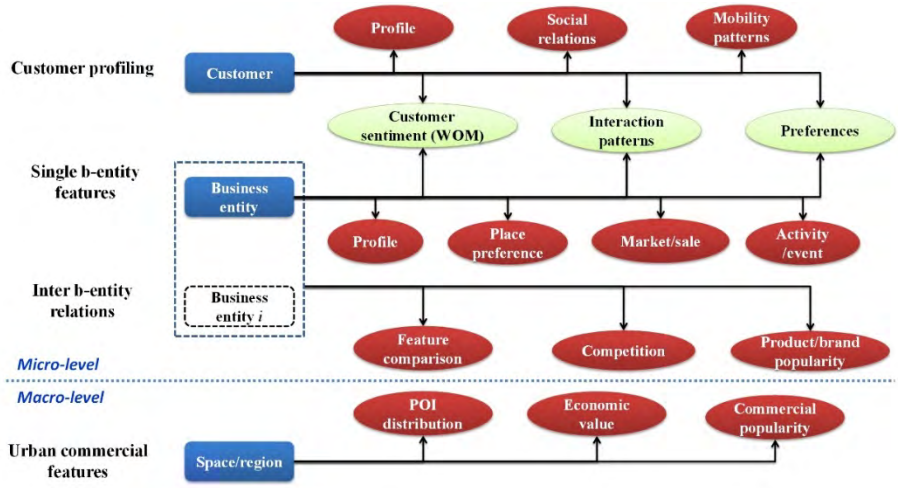


FIGURE 1. The concept model of CrowdBI.

and data mining technologies to present valuable information about customers, products, brands, enterprises, and so on. The learned information can be further used to support a wide range of business decisions, such as product pricing, brand promotion, customer understanding, and product improvement. Unlike traditional BI that analyzes a small amount of data by existing analytical tools, CrowdBI requires advanced technologies to process heterogeneous crowdsourced data to extract diversified information about business.

In general, CrowdBI has some unique characteristics in comparison with other concepts. First, compared with crowdsourcing and MCS which focus on data collection, CrowdBI pays more attention to process crowdsourced data to distill implicit information for business decision making and novel application development. Moreover, compared with traditional BI, CrowdBI can easily obtain a large amount of business data from various data sources by crowdsourcing. By analyzing different types of user-generated data (e.g., text, pictures), CrowdBI can obtain more effective and valuable information, such as extracting entities from users' reviews to understand consumer preferences, visualizing competitive relationships, and creating marketing strategies.

B. THE DATA-CENTRIC PERSPECTIVE OF CROWDBI

1) Data sources

The essential to CrowdBI is multi-source crowdsourced data, as depicted below.

- ***Online social media*** refers to the platforms or websites that people can interact and communicate on it, e.g., Facebook and Twitter. Social media data contains a lot of user-generated content, such as text posts or comments, images or videos, and users' interactions. By mining social media data, we can obtain users' profile and social relations, learn users' preferences, and so on.
- ***LBSNs*** are social networks that use GPS features to locate people and that let people share their location and other content from their mobile devices, such as Foursquare.[8] Check-in data in LBSNs reflect users' trajectories in the real world. In addition, online data and offline data of users can be connected in LBSNs.
- ***Consumer product reviews*** are now widely recognized to have a significant impact on consumer buying decisions. Those online reviews on e-commerce websites present a wealth of information on the products and services, and if properly utilized, can provide vendors highly valuable network intelligence and social intelligence to facilitate the improvement of their business.
- ***Discussion forums*** are online discussion platforms where people can talk about some topics and exchange opinions by posted messages. Different from other social networks, online discussion forums provide a collaborative environment for members to reflect on organization initiatives, express opinions, share thoughts and resources, and engage in community-wide discussions.
- ***Web queries*** contain many entries searched by the search engine, and a user can enter into a web search engine to satisfy his or her information needs. Therefore, the information about users' online behavior can be obtained from query logs which record what users are searching for on the Web.
- ***Pervasive sensors*** can automatically collect the valuable information of users, objects and the environment, such as RFID, WiFi, mobile and wearable devices, and so on. In recent years, pervasive sensors are considered as the most exciting and fastest growing technology in terms of scope of application in the next generation of BI [11]. For example, RFID is used to identify and track tags attached to objects, and WiFi signals are leveraged to recognize shopper's behaviors in retail stores.

[8] https://foursquare.com/

2) CROWDBI KNOWLEDGE GRAPH

The concept model of CrowdBI is shown in Fig. 1, including *business entities*, *customers*, and *multi-dimensional business knowledge*.

- ***Business entities***. In a broad sense, a business entity is defined as "*an entity that is formed and administered as per corporate law in order to engage in business activities*".[9] Business entities can be either physical or virtual. Physical entities contain the products, retails, chain stores, shopping malls, restaurants, and so on, while virtual entities contain brands, e-products, mobile apps, box office, and so on.
- ***Customers***. A customer is an individual that purchases the goods or services obtained from a business entity via a financial transaction. Attracting more and more customers is the primary goal of most businesses to increase sales. Traditional business entities generally compete through advertisement or lowered price to attract more customers, while in CrowdBI, business entities could take advantage of the big data of customers to understand customer demands and preferences.
- ***Multi-dimensional business knowledge***. We categorize the knowledge that can be learned from the crowdsourced data into two levels: *macro-level* and *micro-level*. The macro-level knowledge refers to the knowledge mined from the urban-level data (e.g., POIs), and it mainly consists of *urban commercial features*. The micro-level knowledge is finer-grained data about business entities and customers, which consists of *single entity feature learning*, *inter-entity relation mining* and *customer features*. We characterize them in detail below.
  (a) *Urban commercial features* are extracted from urban-level data, typical examples of which include *POI distribution*, *economic value*, and *popularity of different places in commerce*. POI distribution characterizes the popularity of different areas. The economic value of an area represents the income level and consumption level of urban residents.
  (b) *Single-business-entity features* refer to some unique information or intrinsic natures of a business entity, such as the *product profile*, *brand value*, *place preference*, *market/sale*, *business activity/event*, *customer satisfaction* (e.g., customer sentiment and word-of-mouth), *suggestion*, and so on. By analyzing these features, CrowdBI can provide valuable advices for attracting more customers.
  (c) *Inter-business-entity relations* refer to the relations of interaction among different business entities, such as *brand/product comparison*, *competitiveness*, *complementarity*, etc. For example, the promotions of its competitors could affect the sales of a brand to some extent.
  (d) *Customer features* contain valuable information of customers in business, include *customer profile*, *customer preferences*, *mobility patterns*, *social relationship*, *interaction patterns*, and so on.

Overall, CrowdBI covers a wide range of areas and has a lot of applications for business entities and customers. In this paper, we give a review of some key techniques and widespread applications of CrowdBI.

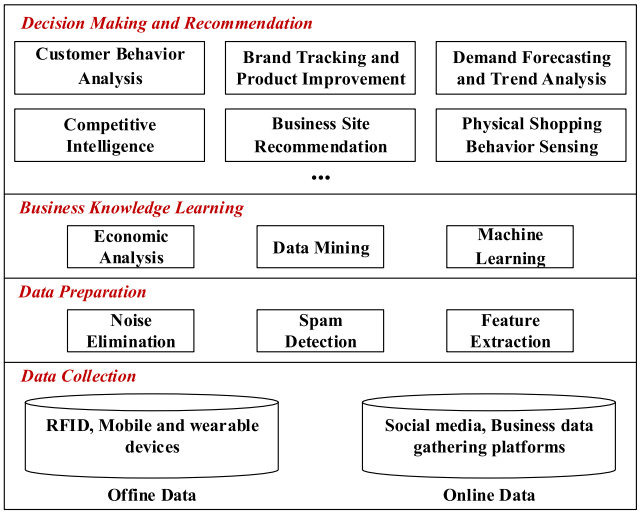


FIGURE 2. The generic architecture of CrowdBI.

C. A GENERIC FRAMEWORK

We present a generic framework of CrowdBI as shown in Fig. 2, which consists of four layers: *data collection*, *data preparation*, *business knowledge learning*, *decision making* and *recommendation*.

- ***Data collection***. We collect crowdsourced data for business intelligence analysis which contains offline data and online data. Generally, the crowdsourced data is obtained in two ways, including *passive and active crowdsourcing*. For example, most *offline sensing data* in the physical world (e.g., mobile sensing data) is obtained by the active crowdsourcing, since users are recruited to collect spatio-temporal data via smartphones or other mobile devices. For *online data* in the virtual world (e.g., data from social media), it is often acquired by passive crowdsourcing since users generate a lot of social data in daily life.
- ***Data preparation***. It is responsible for cleaning and extracting relevant content from different data sources.
- ***Business knowledge learning***. It applies different theories and methods (e.g., economic analysis, data mining, and machine learning) for business knowledge learning.
- ***Decision making and recommendation***. It makes use of the learned business knowledge in decision making and recommendation applications.

III. MAJOR APPLICATION AREAS

In this section, we develop a taxonomy of the major application areas of CrowdBI, including *customer behavior*

[9] https://en.wikipedia.org/wiki/List_of_business_entities

TABLE 1. Customer behavior analysis.

Purpose	Feature	Description
Customer behavior and local business analysis [13]	Geo-Location density	The number of nearby restaurants and population size.
	Geo-Heterogeneity	Entropy measurement to assess the level of spatial heterogeneity of an area; the frequency of restaurant types in the area.
	Geo-Competitiveness	The proportion of nearby restaurants of the same cuisine type with the total number of restaurants within this area.
	Mobility-Mobile Density	To assess the general popularity of an area, measure the total number of check-ins collected among the neighborhood of restaurants.
	Social stability	If the area can maintain current consumers for a long period of time. Use consumers' consecutive check-in behaviors to assess the stability of current consumers staying in the same place.
	Incoming mobility	Whether it can attract consumers from its neighborhoods.
	Traffic efficiency	Measure the dynamic traffic conditions, traffic accidents, traffic jams, bus delays.
	Street closure	Road construction or special city events, number and length.
	Restaurant-specific features	Price level, rating, and comment reviews.
Customer preferences toward different hotel characteristics [15]	Transaction features	Transaction price (per room per night), and number of rooms sold (per night).
	Service-based features	Hotel class, and hotel amenities.
	Review-based features	Number of customer reviews, overall reviewer rating, disclosure of reviewer identity information, subjectivity and readability of reviews.
	Location-based features	Near the beach, near downtown, external amenities (number of restaurants/shopping destinations), near public transportation, near the interstate highway, near the lake/river, and the city annual crime rate.
Using crowdsourced data to evaluate the weight of product features to the product price [16]	Product features	The set of nouns that appear in the reviews, which reflect the quality level of the product.
	User evaluations	Consumers' opinions to the quality of a product characteristic (e.g., bad, good, amazing).

analysis, *brand tracking* and *product improvement*, *demand forecasting and trend analysis*, *competitive intelligence*, *business popularity analysis and site recommendation*, and *urban commercial analysis*.

A. CUSTOMER BEHAVIOR ANALYSIS

The customer behavior analysis in CrowdBI illustrates how to understand customers' behaviors by incorporating multi-source crowdsourced data. A summary of some recent customer behavior analysis is given in Table 1.

1) ECONOMIC ANALYSIS

Econometrics is the application of statistical methods to economic data and is described as the branch of economics that aims to give empirical content to economic relations [12]. Zhang *et al.* [13] use the econometrics model to understand the impact of urban infrastructure changes (e.g., transportation, street facilities, and neighborhood walkability) to local business. Goel *et al.* [14] apply econometrics to understand user search behavior, and use search query volume to forecast the item demand, such as the opening weekend box-office revenue for feature films, and the rank of songs on the Billboard Hot 100 chart.

2) PRODUCT FEATURE ANALYSIS

Products can be characterized by its features (e.g., price and location are features of restaurants). Ghose *et al.* [15] study customer preferences toward different hotel characteristics by mining user-generated reviews and opinions about hotels. Moreover, they are able to estimate consumer preferences toward different hotel characteristics and recommend the most cost-effective hotels. Archak *et al.* [16] use crowdsourced consumer reviews to evaluate the weight of product features to the product price. Using the econometrics method (i.e., the hedonic regression method [17]), they can estimate the weight of each product feature, the implicit evaluation score that customers assign to each feature, and how these evaluations affect the revenue of a given product.

B. BRAND TRACKING AND PRODUCT IMPROVEMENT

Brand-related user posts on social media are increasing rapidly. Users express their opinions about brands by sharing multimodal posts (i.e., texts, images). Brand tracking from social media has attracted much research attention in recent years, the objective of which is to gather brand-related data from online social media streams. Furthermore, the collection of online product reviews has become an important information source for purchase decisions. Measuring such aggregated consumer ''opinions'' is critical to enterprises, because it can facilitate their planning and decision-making pertaining to product improvement, new product development, pricing, positioning, and advertising

1) BRAND TRACKING

The data gathering problem of brand-related user posts is particularly challenging due to the short and conversational nature of posts, the huge data volume, and the increasing heterogeneous multimedia content in social media streams [18]. Moreover, as the focused topic often shifts quickly in social media, the traditional keyword-based approach to gather data with respect to a target brand is grossly inadequate. Gao *et al.* [18] propose a multi-faceted brand tracking method that gathers relevant crowdsourced data based on evolving keywords, social factors, and visual contents. During a critical period, such as a product release, it would be useful if a brand manager could be provided with recommendations about who will talk about their product in social media. Wu *et al.* [19] develop a recommendation system to learn historical user posts in social media, and predict the potential customers that the brands need to follow when they release new products. Ibrahim *et al.* [20] explore how the online brand community will influence the company retail businesses by using the data from Amazon and Twitter. They find that social media is a good platform to explore the trends of marketplace, and businesses could improve their brand images by engaging with customers in social media. Mazloom *et al.* [21] identify which aspects of posts will determine the popularity of brand-related posts from Instagram. They consider several brand-related cues, including factual information, sentiment, vividness and entertainment parameters. To help companies monitor their brands, Liao *et al.* [22] extract representative aspects and posts from crowdsourced social media data. They extract the aspects by maximizing the representativeness, which accounts for both the coverage of aspects and the impact of posts. Recognizing the business venue (e.g. cafe shops, local restaurants) in an image can help many applications for personalization and location-based services/marketing. Chen *et al.* [23] propose a multimodal mining approach to recognize the business venues from multi-source crowdsourced data, including images from Instagram and reviews from Yelp.

2) PRODUCT IMPROVEMENT

Customers may discuss the pros and cons, strengths and weaknesses of products in social media, these online reviews generated by customers can benefit the product designers. Therefore, for manufacturers and enterprises, it is important to identify helpful online reviews and learn from these reviews for new product development. Liu *et al.* [24] extract four categories of features (i.e., linguistic features, product features, features extracted based on information quality and features extracted using information theory) to identify helpful reviews. Decker and Trusov [25] use three regression approaches to measure the aggregated consumer preferences about mobile phones from online product reviews. To facilitate product design and improvement, Xiao *et al.* [26] propose an econometric preference measurement model to extract aggregated consumer preferences from online product reviews. Moreover, they extend the Kano model [27] to categorize customer requirements based on the aggregate consumer preferences. Qi *et al.* [28] identify the helpfulness of online reviews by using the crowdsourced data from JD.com.[10] They then apply the Kano model to analyze online reviews to develop appropriate product improvement strategies. Xie *et al.* [29] identify the business value of consumer reviews and management responses to hotel performance. Based on econometrics model, they present a panel data analysis of online consumer reviews and management responses of 843 hotels from a hotel review website (TripAdviser.com.[11]) The results show that the overall rating, attribute ratings of purchase value, location and cleanliness, variation and volume of consumer reviews, and the number of management responses are significantly associated with hotel performance. Xiang *et al.* [30] explore and demonstrate the utility of big data analytics to understand the relationship between consumers experience and satisfaction using the data from Expedia.com.[12]

C. DEMAND FORECASTING AND TREND ANALYSIS

Forecasting demands or sales from relevant indicators is an important task in marketing and business. Traditionally business owners make decisions based on experience, which could be limited or even biased. We discuss approaches that use crowdsourced data for demand forecasting and trend analysis.

1) DEMAND FORECASTING

Yu *et al.* [31] apply the regression model for sales prediction by considering the quality factors of online reviews of products. Wu and Brynjolfsson [32] utilize a seasonal autoregressive method to model the relationship between queries submitted to Google's search engine and housing market indicators, and predict housing prices and sales. Through analyzing the sales rank values and posts for books, Gruhl *et al.* [33] find that there exists strong correlation between the volume of reviews and sales spikes. Liu *et al.* [34] and Yu *et al.* [35] analyze sentiment from blogs by using the Probabilistic Latent

[10] https://www.jd.com/
[11] https://www.tripadvisor.com/
[12] https://www.expedia.com/

Semantic Analysis (PLSA) method, and present an autoregressive sentiment-aware model to predict product sales.

2) ECONOMIC IMPACT ANALYSIS

Based on economic models (e.g., econometrics), Ghose and Ipeirotis [36] analyze the impact of reviews on economic outcomes (e.g., product sales) and examine how different factors affect social outcomes such as their perceived usefulness. The econometric analysis reveals that the extent of subjectivity, informativeness, readability, and linguistic correctness in reviews matters in influencing sales and perceived usefulness. Yu *et al.* [35] investigate the impact of reviews for product sales performance. They find that both the sentiments expressed in the reviews and the quality of the reviews have a significant impact on the future sales performance of products. Zheludev *et al.* [37] use sentiment analysis and information theory (mutual information between time series) to quantify and validate which assets could qualify for trading from social media analysis. Mishne and Glance [38] find that the volume and sentiment of discussion about a product (e.g., a movie) in weblogs is correlated with the product's financial performance.

3) TRENDING ESTIMATION

For business managers, one of their biggest common concerns is whether their business will thrive or fail in the future. Knowing about long-term trends, business managers can make decisions in advance. Lian *et al.* [39] study the potential indicators for the long-term survival of a physical store. Specifically, they study four types of factors by using the data from Dianping and Yelp, including geography, user mobility, user rating, and review text. They find that the location and nearby places play an important role in the success of the shop, and usually less competitiveness and more heterogeneity is better. Based on the four types of factors, they use three prediction methods to estimate the trend, including Logistic regression, gradient boosted decision tree (GBDT), supported vector machine (SVM). In order to improve product sales, it is very important for companies to make rapid and correct response to the hot trends, for example, there may exist the implicit relation between air pollution and mouth mask. Wang *et al.* [40] aim to learn commercial intents from microblogs by using a graph-based product-trend association model. Arias *et al.* [41] study whether the indicator extracted from Twitter can improve the forecasting of social, economic, or commercial indicators. They find that the addition of crowdsourced Twitter data can improve the prediction of the trend of volatility indices in the stock market.

D. COMPETITIVE INTELLIGENCE

The social comparison theory [42] suggests that competitions are ubiquitous. Detecting and monitoring competitors becomes a critical issue for a company to make marketing strategies. On the one hand, companies have to keep track of its competitors' status. On the other hand, consumers typically try to compare similar products before purchase. Competitive intelligence is defined by Kahaner [43] as "*a process of monitoring the competitive environment, with a goal to provide actionable intelligence that will provide a competitive edge to the organization*." The main goal of competitive intelligence is to monitor a firm's external environment for information that is relevant to its decision-making process [44].

Competitive intelligence allows a company to identify its competitors' strengths, weaknesses, strategies and other areas, and in turn to help the company improve its strategic decisions against its competitors. Competitive intelligence offers a number of benefits such as creating new growth opportunities, minimizing the impact of surprises, enabling faster responses to changes in the market place, improving the quality of strategic planning processes, identifying potential vulnerabilities, providing early warning or alert for competitive threats [5], [45]. However, the functions of competitive intelligence are often greatly restricted by the lack of sufficient information sources about the competitors. With the emergence of Web 2.0, the large numbers of customer-generated product reviews often contain information about competitors and have become a new source to mine competitive intelligence. A summary of the recent competitive intelligence studies is given in Table 2.

1) COMPETITOR IDENTIFICATION

The objective of competitor identification is to mine competitive relationships and identify competitors of a company. Xu *et al.* [46] develop a graphical model to extract and visualize comparative relations between products from customer reviews on Amazon. Netzer *et al.* [47] convert the user-generated content to market structures and competitive landscape insights. Given a focal company, these studies focus on extracting the competitors based on their direct co-occurrence with the focal one in social media data. But the co-occurrence hypothesis is too restricted to suit various application domains; besides, it usually does not hold for small marketplaces, e.g., a few reviews can be found to cover two or more peer. Instead of co-occurring analysis, Bao *et al.* [48] propose to compute the competitiveness based on the feature overlapping between items. Then, competitors are ranked according to the combination of several metrics including mutual information, match count, and candidate confidence. Yang *et al.* [49], [50] study the problem of mining competitive relationships from two complementary data sources (i.e., Twitter and patent records), and they propose the Topical Factor Graph Model (TFGM) to bridge the two networks and learn a semi-supervised learning model to classify the relationships between entities.

2) COMPARATIVE ANALYSIS

Competitor identification is the foundation of comparative analysis, assume the competitors are identified or already known, one of the basic steps for comparative analysis is comparison sentence identification from crowdsourced customer reviews. Identifying comparative sentences is useful in practice because direct comparisons are perhaps one

TABLE 2. Competitive intelligence.

Work	Problem and method	Dataset
Xu et al. [46]	*Problem*: extracting and visualizing comparative relations between products from customer reviews; *Method*: conditional random field	1347 customer reviews of 33 types of mobile phones from Amazon.
Netzer et al. [47]	*Problem*: exploring the market structure and the brand-associative network derived from online discussions; *Method*: text-mining and network analysis approaches	868,174 consumer messages (nearly 6 million sentences) on sedan cars from Edmunds.com[13] , and more than 670,000 messages (more than 5 million sentences) on diabetes drug from diabetes drug forums.
Bao et al. [48]	*Problem*: mining competitors of a given entity from the web; *Method*: text mining, linear regression, opinion summarization	10 entity queries collected as training data by using Google API[14], 70 entities (e.g., company, product, and football club) as test data.
Yang et al. [49]	*Problem*: mining competitive relationships by learning across heterogeneous networks; *Method*: topical factor graph model	3,770,411 patents from USPTO[15], 1,033,750 tweets from Twitter.
Jindal et al. [52]	*Problem*: identifying comparative sentences from the texts and extracting comparative relations from the identified comparative sentences; *Method*: class sequential rules, naïve Bayesian Classification, label sequential rules	Customer reviews, forum discussions and random news articles from disparate web sources.
Zhang et al. [54]	*Problem*: monitoring temporal evolution of market competition; *Method*: dynamic topic model	About 10 million of tweets and 8 million of associated images from Twitter of the 23 brands in the two categories of luxury and beer.
Zheng et al. [55]	*Problem*: estimating important competitive measures; *Method*: Limited Information Dirichlet model	50,000 panelists' online visiting and purchasing activities from comScore Networks[16].
Zhang et al. [57]	*Problem*: exploring competitive information in product reviews; *Method*: network analysis, text sentiment mining techniques	Customer reviews and ratings of digital cameras from Amazon.
He et al. [58]	*Problem*: performing the competitive analysis of the three largest pizza chains; *Method*: text mining	Text messages posted on Facebook and Twitter sites of the three largest pizza chains: Pizza Hut, Domino's Pizza and Papa John's Pizza.
He et al. [45]	*Problem*: performing the competitive analysis of two largest retail chains in the world: Walmart and Costco *Method*: sentiment analysis	Tweets from Twitter about Walmart and Costco.
Kim et al. [59]	*Problem*: comparing the social media opinion and the shipment gap between two rival smart phones: iPhone6 and GalaxyS5; *Method*: Opinion mining and sentiment analysis	229,948 tweets mentioning the iPhone6 or the GalaxyS5 from Twitter.
Wei et al. [60]	*Problem*: measuring the competitiveness degree between peers; *Method*: a bipartite graph model	A total of 20 focal entities in various fields, approximately 10^9 query logs collected from Google AdWords about the focal entities and their competitors.

of the most convincing ways of evaluation. Comparative sentence mining has been intensively studied by Jindal and Liu [51], [52]. Two new types of rules, class sequential rules (CSRs) and label sequential rules (LSRs) are proposed for sentence identification and relation extraction respectively.

3) MARKET ATTENTION ANALYSIS

He *et al.* [53] propose a business-driven social media competitive analytics tool named VOZIQ, which compares social media comments from different companies within the same industry, and compares the share of different social media platforms in mentions of companies in the same industry.

4) COMPETITIVE DOMAIN MINING

Competitive domain mining aims to extract the competitive domains of the given entity and its competitors. For example, for a market of interest (e.g., luxury goods), it is important to detect the latent topics (e.g., bags, clothes, luxurious) that are competitively shared by multiple brands, and track the temporal evolution of the brand's competition over the shared topics. Zhang *et al.* [54] propose a dynamic topic model to monitor temporal evolution of market competition by combining tweets and the associated images. Zheng *et al.* [55] develop the Limited Information Dirichlet model [56] to identify key competitive measures (e.g., penetration, market share, share of wallet). Zhang *et al.* [57] construct product comparison networks from consumer product reviews for competitive analysis and comparison among products. He *et al.* [58] focus on a case study of competitive analysis among the three largest pizza chains (e.g., Pizza Hut, Domino's Pizza and Papa John's Pizza) by using data from Facebook and Twitter, and the results reveal the business value of comparing social media content. To understand the

trends of popularity and customer engagement, He *et al.* [45] propose a framework for social media competitive intelligence to enhance business value and market intelligence. They conduct a case study to two largest retail chains in the world: Walmart and Costco using Twitter data, and analyze the social media volume and sentiment trend analysis for these two retail chains. A similar study is reported in [59], which compares two competing smartphone manufacturers (i.e., iPhone 6 and Samsung Galaxy S5) using Twitter data. Competitiveness degree analysis is a focal point of business strategy and competitive intelligence, which aims to help managers closely monitor to what extent their rivals are competing with them. Wei *et al.* [60] propose a bipartite graph model to measure the competitiveness degree between peers from query logs.

E. BUSINESS POPULARITY ANALYSIS AND SITE RECOMMENDATION

Location is a crucial factor of retail success, as 94% of retail sales are still transacted in physical stores [61]. Choosing a good location when opening a new store is crucial for the future success of a business. To increase the chance of success for their stores, business owners require not only the knowledge of where their potential customers are, but also their surrounding competitors and complementary businesses. The problem of identifying the optimal location for a new retail store has been the focus of past research, especially in the field of land economy [62]. Traditional approaches to the problem have factored in demographics, revenue and aggregated human flow statistics from nearby or remote areas. However, the acquisition of relevant data is usually expensive. Traditionally, business and property owners conduct surveys to assess the value of store locations. Such surveys, however, are costly and do not scale up well. With fast changing environments (neighborhood rental, local population size, composition) and emergence of new business locations, one also needs to continuously reevaluate the value of store locations. In the era of social media and mobile apps, we have an abundance of online user-generated data, which capture both activities of users in social media as well as offline activities at physical locations. This section presents the usage of crowdsourced data for business site selection, which generally consists of two types of approaches: *model-based* and *data-mining-based* approaches.

1) MODEL-DRIVEN APPROACH

Early studies are based on dedicated models for store placement. The spatial interactions among different categories of stores are quantified using network analysis in [63], and the weighted links reveal the spatial attraction or repulsion between retail stores. Based on the quantification of store interactions in the network, a retail quality index is devised to detect appropriate locations for retail stores. Porta *et al.* [64], [65] investigate the relationship between street centrality and retail store density in the cities of Bologna and Barcelona respectively. In general, a central location has the potential to attract more customers and sustain higher densities of retail stores and service activities. A multiple centrality assessment model is used to characterize the street centrality, which is composed of multiple measures such as closeness, betweenness, and straightness. Kernel density estimation is then used to transform datasets of centrality and activities to one scale unit for analysis of correlation between them. Finally, they verify that retail stores and service activities in Bologna tend to concentrate in areas with better centralities [64]. They further study the street centrality and their correlations with different types of economic activities by using the Multiple Centrality Assessment (MCA) model [65]. They demonstrate that pure spatial organization can be indicative of the quality of retail and economic activities.

Huff [66] reports that the probability that a consumer patronizes a certain shopping area is directly related to its size, and inversely related to its distance from the consumer. It further proposes a probabilistic model to determine the retail trade areas of the given shopping center. Li and Liu [67] present a modified Huff model to estimate the potential sales of individual Kmart and Wal-Mart stores in the Greater Cincinnati Area. Compared to traditional interaction models, the modified Huff model integrates the traditional Huff model and the competing destination model (CDM), which is capable of more accurately examining the impact of competition and agglomeration at the same time, and estimating individual store sales. Sevtsuk *et al.* [68] analyze location patterns of retail and food establishments. It tests five hypotheses about retail locations found in previous literature using an economic model, and estimates the impacts of different location characteristics for store placement. More specifically, they analyze the spatial distributions and location patterns of retail and food establishments in dense urban environments (Cambridge and Somerville, Massachusetts). The results indicate that both micro-location characteristics and clustering between stores are important to explain the retail landscape, including other neighboring retailers, land use, urban form, and transit characteristics around each building. Roig-Tierno *et al.* [69] present a methodology for retail site location decision, which takes both geographic information systems (GIS) and the analytical hierarchy process (AHP) into consideration. Specifically, they first analyze the factors that influence the success of a supermarket, including establishment, location, demographic factors, and competition. The AHP method is then used to rank possible sites based on the combined results of geo-demand and geo-competition analyses.

2) DATA-DRIVEN APPROACH

Traditional site selection mainly relies on domain expert knowledge or traditional data, such as demographic studies

[13] https://www.edmunds.com/
[14] https://developers.google.com/apis-explorer/#p/
[15] https://www.uspto.gov/
[16] https://www.comscore.com/

or consumer surveys, which are very expensive and time-consuming to obtain. Recently, with the emergence of large-scale user-generated data (e.g., check-in data, rating data), there is a potential to leverage these data to analyze and mine users' preferences for business popularity and site recommendation. Karamshuk *et al.* [70] demonstrate the power of geographic and user mobility features in predicting the best placement of retail stores based on check-in data in Foursquare[17] in New York. The geographic features are characterized according to the types and density of entities in the area of interest, including density, neighborhood entropy, competitiveness, and quality by Jensen. The mobility features are formulated to measure the popularity of an area in terms of check-in patterns, including area popularity, transition density, incoming flow, and transition quality. Except for user mobility factors, user-generated reviews can also construct predictive features for analyzing users' preferences and assessing the attractiveness of candidate locations. In [71], three types of features are incorporated into a regression model to predict the number of check-ins at a candidate location, including review-based market attractiveness features, review-based market competitiveness features and geographic features. Lin *et al.* [72] analyze the popularity of a business location using Facebook data in Singapore. They first extract a set of relevant features from several key factors, such as business categories, locations, and neighboring businesses. A robust model based on gradient boosting machine (GBM) is then developed to estimate the popularity of a given target location. A summary of the representative features for retail placement in related works is given in Table 3.

TABLE 3. Representative features for retail placement.

Category	Features Used in Related Works
Establishment	*The characteristics of the property itself* [69]: sales floor area, parking, number of departments, number of checkouts.
Demographics	*Profile of the clients living in the trade area* [69]: potential markets, socio-demographic characteristics, growth in the area, seasonality.
Geography	*The information of the environment around the place* [69], [70], [77]: neighbors' density, neighbors' entropy, competitiveness, complementary, accessibility by car, accessibility by foot, visibility, volume of passing trade, square distance from the customer point to the place.
Mobility	*The information of mobile users in the palace* [70]: area popularity, transition density, incoming flow, transition quality.
Social-geography	The social connections between customer and the place [77].
Review	Review-based market attractiveness and competitiveness [71].

Ranking areas by popularity of a business category is an essential task for business site selection. Yu *et al.* [73] propose a novel approach to estimate the popularity of a business category by exploiting user-generated contents from LBSNs (e.g., Foursquare). Eravci *et al.* [74] introduce a new problem of identifying neighborhoods with a potential of success in a line of business. Specifically, they use the similarities of the neighborhoods to identify specific neighborhoods as candidates for investment for a new business opportunity. Two methods are proposed to solve the problem, 1) a probabilistic approach based on Bayesian inference to select location, and 2) an adaptation of collaborative filtering in view of the similarity of neighborhoods. Xu *et al.* [75] propose a Demand Distribution Driven Store Placement (D3SP) framework for store location selection via mining search query logs of Baidu Maps.[18] D3SP first identifies the spatial-temporal distributions of customer demands from map query data, and then determines candidate locations via clustering places that demands exceed supplies. The business store placement problem is solved by supervised regression models to predict the number of customers, and learn-to-rank models to rank the candidate locations directly.

The volume of customers for places is the main factor to analyze the shop popularity, which can be used for site selection and advertisement recommendation. Hsieh *et al.* [76] develop a method, called Potential Customer Estimator (PCE), to estimate the potential customers of a given location and any time of interest in an urban area. It considers existing stores as one kind of sensors to obtain the information of customers (e.g., the historical check-in records), which can be exploited to estimate potential customers. Different from most approaches of customer volume predication that mainly use supervised or unsupervised leaning in view of partial information, Wang *et al.* [77] propose a Geographical Regression and Non-negative Matrix Factorization method (GR-NMF), which can jointly model implicit footfall knowledge and explicit geographical knowledge via an integrated probabilistic framework.

F. PHYSICAL SHOPPING BEHAVIOR SENSING

Analyzing shopper's behaviors in brick-and-mortar retail stores and shopping malls can provide crucial insights in a variety of aspects, such as browsing habits, shopping interests, store visiting frequency, companions, etc. Customer shopping behavior analysis is of importance to improve retailer profits and customer experience. Such tracking and analysis of a shopper in a retail store is referred as *physical shopping behavior analytics*. Compared to online shopping, in-store shopping lacks of effective approaches to identify comprehensive customer behaviors, such as picking up items, trying-on clothes, and price comparisons. Mining of such data could provide deep and comprehensive insights of customer interests, experiences, and expectations, which are important to improve service quality and profits of retailers.

1) WHY NOT VIDEOS?

Traditional methods to analyze customer behaviors is based on video monitoring. These methods deploy plentiful

[17] https://foursquare.com/

[18] http://map.baidu.com/

TABLE 4. In-store behavior analysis.

Project (Device)	Activity (Sensor)	Description
Shopper-device-based systems: **ThirdEye** (Google Glass) [82]	Walking detection (accelerometer)	Walk detection uses the step counter to detect steps.
	Dwelling detection (inertial sensors)	Dwell detection scheme is based on the observation that while dwelling, the net physical displacement would be small even though the user may be moving continuously.
	Gazing detection (inertial sensor and camera)	Gaze detection uses inertial sensing to turn on the video and then uses vision-based techniques to complement inertial sensing, to achieve both energy efficiency and a gaze detection accuracy.
	Reaching out detection (camera)	Reaching out detection is accomplished by hand detection using video, because shoppers are typically looking at an item while reaching out for it.
Shopper-device-free systems: **ShopMiner** (RFID) [84]	Discovering popular category (passive RFID tags)	Popular category represents the products frequently viewed by customers, which can be discovered by distinguishing the popular items that show remarkable phase changing from the temporal phase trend.
	Identifying hot items	Hot items are the products frequently picked up or turned over by customers, which can be identified by exploiting the phase changing caused by customer behaviors.
	Excavating correlated items	Correlated items are the products that are frequently matched with or tried on together. ShopMiner explores the spatial-temporal correlation of phase trends to discover those correlated items.
Combination of in-store deployed devices and human wearables: **IoT+Small** (personal wearable-devices and store-deployed sensors) [91]	Identifying "Item Picking" Gesture (smartwatch's accelerometer and gyroscope)	Picking action should be distinguishable from other similar actions or gestures such as putting items back/putting items aside, pushing/pulling a door etc. The decision tree algorithm is used to classified different actions.
	Identifying "Shelf-level Location of Item Picked" (smartwatch's camera)	Identifying the picking action when a person picked an item from shelves which are at different heights.

cameras in stores and collect shopping data by recognizing customer activity from video, then they capture physical shopping behaviors through video analyses and image matching technology [78]. However, there are at least the following four limitations for video-based methods. First, they demand good light and unobstructed line of sight to keep good accuracy. Second, image feature databases of customer behaviors are required for recognizing customer actions, which will take a high cost to gather before system deployment. Third, most of the analysis is done offline and need lots of computing resources. Even so, the accuracy of recognizing actions and target object is hardly satisfactory. Finally, installing cameras in stores may violate customer privacy in some cases.

2) COARSE-GRAINED SHOPPING ACTIVITIES

Non-in-Store Behaviors: You *et al.* [79] propose a phone-based system to sense physical shopping activities while tracking shopping time at physical stores. The system trains a trajectory classifier to label a given motif group as shopping or non-shopping. Lee *et al.* [80] study customer malling behavior patterns in shopping mall over different activities, such as eating, shopping, reading, resting, and so on. They propose a computational framework, named MallingSense, to understand customers' behaviors in shopping malls. Banerjee *et al.* [81] study the problem of persona-based shopping recommendation for physical retail stores. A dynamic programming-based recommendation system is proposed to make Top-k recommendations to the shopper for maximizing her value for money, which considers about shopper personality as well as time and budget constraints.

3) FINE-GRAINED SHOPPING ACTIVITIES: IN-STORE BEHAVIORS

In general, in-store behaviors is categorized into three types: *shopper-device-based*, *shopper-device-free*, and *a combination of them*. A summary of the in-store behavior analysis is given in Table 4.

a: SHOPPER-DEVICE-BASED SYSTEMS

Some shopper-carried devices are employed to sense physical shopping behaviors, such as smartphones, Google glasses, etc. ThirdEy [82] uses shoppers' smart glasses to obtain images, inertial sensors, and WiFi data, to track the physical browsing of shoppers, without requiring any input from either the store owners or from the shoppers themselves. Particularly, it can automatically identify shopper behavior into one of the four categories: *walking*, *dwelling*, *gazing*, and *reaching out*. IRIS [83] investigates the possibility of using a combination of a smartphone and a smartwatch, carried by a shopper, to get insights into the shopper's behavior inside a retail store. More specifically, IRIS uses standard locomotive and gestural micro-activities as building blocks to define novel composite features, which help classify different facets of a shopper's interaction with individual items, and attributes of the overall shopping episode or the store.

b: SHOPPER-DEVICE-FREE SYSTEMS

It refers to the systems where devices are deployed in the retail stores (e.g., WiFi, RFID) while not relying on shoppers' devices. Shangguan *et al.* [84] show that backscatter

TABLE 5. Urban commercial analysis.

Work	Feature/Factor	Description
Yang et al. [94]	Visitors	The numbers of visitors to commercial districts reflect hot degrees of such districts to some extent.
	Population and Purchasing Power	The total number of the houses reflects the population in the surrounding region while the average house price reflects the purchasing power to some extent.
	Region Functions	The distribution of different categories of POIs within a range of 5 kilometers to the center of each commercial district.
	Rating from Customers	The crowdsourced data of consumers rating on the services available at dianping.com.
	Commercial activeness	The number of online comments on commercial entities, available at dianping.com, as an indicator of commercial activeness.
Georgiev et al. [101]	Distance	The geographic distance in meters between the target venue and the closest event-related venue.
	Nearby Place Entropy	The heterogeneity of a neighborhood in terms of the specific types of places located inside.
	Jensen Quality	The spatial distribution of places with respect to their ability to attract other venues of certain types.
	Entertainment Flow	The mean empirical probability of observing such transitions in the area around a target venue.
	Social Area	The sociability of a neighborhood by counting the pairs of friends that have visited the area.

signals of passive RFID tags can be exploited to detect and record customer behaviors, for example which items of clothes they pay attention to, and which items of clothes they usually match with. They design a framework called ShopMiner, which can harness the unique spatial-temporal correlations of time-series phase readings of RFID tags to detect comprehensive shopping behaviors (e.g., look at, pick out, or turn over desired items). In addition, it supports three basic behavior mining functionalities essential to retailers, namely discovering popular categories, identifying hot items, and excavating correlated items. Han *et al.* [85] present an RFID-based customer behavior identification system called CBID. It can detect and track tag movements and further infer corresponding customer behaviors (e.g., picking up an item). Specifically, CBID includes a Doppler-effect-based protocol to detect tag movements. TagBooth [86] detects items' motion using RFID devices and further discovers customers' behaviors, such as picking up and toggling clothes on hangers. It first exploits the motion of tagged commodities by leveraging physical-layer information, like phase and RSS, and then designs a comprehensive solution to recognize customers' actions. Pradhan *et al.* [87] also introduce an RFID-based smart shopping system, KONARK, which helps user's checkout items faster and track purchases in real-time. Wang and Yang [88] propose a sensor-based smart shopping cart system called 3S-cart by using the context-aware ability of sensors to detect movements and actions from the customers. It exploits a force-sensitive resistor to check if the customer is holding the cart's handle by detecting the change of its voltage, and an accelerometer to derive the cart's moving direction. Particularly, they propose applications to demonstrate its practicability. For example, in the sales-promotion application, each cart checks if its customer has interest in some products and shows sales information at once to increase the purchasing desire.

WiFi CSI sensing is a non-intrusive, device-free, and privacy-preserving form of sensing shopper's behavior for accurate physical analytics. Zeng et al. [89] propose a WiFi CSI-based method for sensing the shopper's behaviors (e.g., location and movement, interaction with items) in a retail store. The experiment results indicate that their proposed method can achieve around 90% accuracy to classify different states of the shopper during a typical in-store visit, such as near the entrance or inside the store. PreFi [90] is a customer's product preference analysis system. It can recognize what the customer is doing based on WiFi devices, such as walk inside, walk away, taking item, and put the item back.

c: COMBINATION OF IN-STORE DEPLOYED DEVICES AND HUMAN WEARABLES

Radhakrishnan *et al.* [91] propose an architecture where low-cost BLE beacons and embedded sensors are mounted on product shelves, and their data is fused with sensor readings from a smartwatch worn by a shopper to create individualized services for in-store shoppers. This system can recognize some shopping activities, including picking item from a shelf, putting it in a trolley and interacting with shelves (opening and closing) in a store. AudioSense [92] monitors user-item interactions inside a store for precisely customized promotions. It automatically detects the items that a shopper touches or picks by localizing the inaudible sound signals emitted by a smartwatch worn by the shopper.

G. URBAN COMMERCIAL ANALYSIS

Urban commercial analysis extends business analysis/mining from a store-level to the city-level. Urban researchers and planners are often interested in understanding how economic activities are distributed in urban regions, and what forces influence their special patterns. A summary of the urban commercial analysis is given in Table 5.

1) COMMERCIAL AREA ANALYSIS

Urban commercial areas are key functional areas in cities with high density of shopping malls, restaurants, entertainment services, and other commercial entities. Qu and Zhang [93] illustrate how user-generated mobile location data (e.g., Foursquare check-ins) can be used in commercial area analysis. They present the analytic method within a trade area analysis framework to model customer mobility, create customer profiles and preferences, and examine interactions between customers and stores. Yang *et al.* [94] aim at revealing how commercial hotness of urban commercial areas is shaped by social contexts of surrounding areas to render predictive business planning. The cyber- and physical-data are combined to infer the association between the heat map of business entities and various social contexts, including human mobility patterns reflected in taxi GPS traces, urban planning issues reflected in POIs, and configuration of ecosystem.

2) URBAN ECONOMIC ACTIVITIES

Urban economic activities and customer behaviors play an important role in urban commercial analysis. Ravulaparthy and Goulias [95] introduce the link-based network centrality indices (e.g., closeness and accessibility) to represent location properties. These centrality indices exhibit unique geometric properties delineating network regions and critical locations to guide urban planning. In addition, they use Latent Class Cluster Analysis (LCCA) to classify the region into highly central and least central places and further study the spatial distribution and composition of economic activities within these clusters. Anagnostopoulos *et al.* [96] propose a methodology for performing targeted outdoor advertising by leveraging the data from Twitter. They use Twitter data to gather information about users' degree of interest in given advertising categories and about the routes that they follow, which is then used to estimate the most promising areas for ad placement in the city. Çelikten *et al.* [97] analyze data from LBSNs with the goal of understanding how different locations within a city are associated with different kinds of activities (e.g., shopping centers, dining venues). Specifically, their analysis makes use of a probabilistic model to reveal how venues are distributed in cities in terms of several features, including the exact location of an activity, the users who participate in the activity, as well as the time of the day and day of week the activity takes place. The Sparse Additive Generative model (SAGE) is proposed by Eisentein *et al.* [98] to identify significant features for distinguishing regions. The central idea of SAGE is that each class label or latent topic is endowed with a model of the deviation in log-frequency from a constant background distribution.

Sarwat *et al.* [99] present PLUTUS, a framework that assists venue (e.g., restaurant, gym, shopping mall) owners in growing their business by recommending potential customers. Three main aspects are considered in PLUTUS to recommend the best set of customers: social aspect, spatial aspect, and user opinions. Agryzkov *et al.* [100] propose an adapted PageRank algorithm for ranking the nodes in a network, which can be used to understand and visualize certain commercial activities (e.g., restaurants and bars, shops, banks and supermarkets) of a city.

3) RISE AND FALL PATTERNS

The rise and fall of business entities and their impact factors are also studied recently. Georgiev *et al.* [101] provide an approach to modeling the impact of the Olympic Games on local retailers by analyzing a dataset mined from a large location-based social service, Foursquare. The results suggest that the venue popularity rankings from subsequent time periods see their lowest agreement around the Olympic period and at places that are close to where the event itself and live broadcasts are held. Volatile Point-of-Interests (vPOIs) refer to those small businesses which appear and disappear quickly in cities. The prediction task for the rise and fall of vPOIs is valuable for both shopkeepers and administrators by supporting a variety of applications in urban economics. Lu *et al.* [102] propose DC-CRF, a dynamic continuous CRF model which can predict the prosperity of vPOIs over time. Specifically, they first aggregate vPOIs prosperities at focal areas in view of the data sparsity and skewness of the individual vPOIs. Then the dynamic-continuous CRF (DC-CRF) model is developed to integrate the association between input and output as well as the correlations between outputs from temporal, spatial and contextual perspectives.

IV. RESEARCH CHALLENGES AND KEY TECHNIQUES

Beyond the well-studied problems such as text mining and sentiment analysis, CrowdBI has the following particular issues to be addressed, as discussed below.

A. FINE-GRAINED DATA COLLECTION

There are numerous issues regarding data collection from different data sources.

1) PERVASIVE SENSING

For fine-grained human offline shopping behavior sensing, there are quite a few issues to be addressed. The first challenge is about how to collect high quality data. Zhou *et al.* [103] aim to detect and record customer shopping behaviors when they browse physical stores. By exploiting the backscatter signals of passive RFID tags, they detect which garments customers pay attention to, and which garments users usually pair up. However, the coverage of multiple readers may overlap, and thus collision becomes an important issue in a multi-tag-multi-reader system. There are two types of collisions: tag-to-tag collision and reader-to-reader collision. They adopt retransmission to resolve tag-to-tag collision, and use reader scheduling to minimize reader-to-reader collision.

The second challenge refers to energy saving. For example, Rallapalli *et al.* [82] study the problem of tracking physical browsing by users in retail stores. They track physical browsing by using a combination of a first-person vision enabled by smart glasses, and inertial sensing using both the glasses and a smartphone. However, continuous video is extremely

expensive in terms of energy consumption, they rely on inertial sensors as much as possible and to trigger vision only when necessary. Radhakrishnan *et al.* [91] aim to use infrastructure sensors, and wearable/mobile devices to detect customers' in-store behaviors in an energy-efficient manner. They propose an architecture where low-cost BLE beacons and embedded sensors are mounted on product shelves, and their data is fused with sensor readings from a smartwatch worn by a shopper.

2) DATA GATHERING IN SOCIAL MEDIA

Besides offline sensing, there is a pressing need for business-related data gathering from online social media. There are several challenges regarding data gathering [18]. First, social media microblogs are usually short and conversational in nature, and thus the contents and vocabularies used in the microblogs usually change rapidly. There is a need to track relevant microblogs using evolving keywords. Second, the amount of social media for a popular entity may be huge. The traditional keyword-based data crawling methods are limited in the coverage of all relevant data, and a multi-faceted approach is needed to ensure wide coverage of data from online social media streams.

Existing works mainly focus on query expansion techniques for data gathering. For example, Massoudi *et al.* [104] propose a topic expansion method to gather relevant data, and they use query expansion to generate dynamic topics for the target. The data gathering using brand-related keywords usually contain a lot of noise, as the presence of brand names does not necessarily guarantee the relevance of posts. To address this problem, Gao *et al.* [18] leverage image content to find relevant microblogs that have high relevance in texts, social factors, and visual contents. They apply a hypergraph method to eliminate noise in data collection. Gao *et al.* [105] aim to filter noise by considering the multimedia content and social nature of brand-related data. They develop a microblog filtering method based on a discriminative multi-view embedding method. With the learned embedding, they use the classification algorithm (e.g., SVM) to filter microblogs. Qi *et al.* [106] propose a microblog brand identification framework. They first train visual/textual content of microblogs to determine the relevant degree between microblogs and the predefined brands. Then, they construct a microblog similarity graph and propose a graph mining model to filter out irrelevant microblogs. Chen *et al.* [107] propose to jointly employ the keywords, candidate topics and popular topics for data gathering. They also devise a new ranking scheme by combining the relationship between the users and tweets to support real-time search in microblogging systems.

3) DATA GATHERING PLATFORMS

There have been several existing social media monitoring platforms online, as shown in Table 6. *Trackur*[19] provides social media monitoring, and supports the mainstream social media and news, such as Twitter, Facebook, and Google+. Trackur supports the analysis of trends, keyword discovery, automated sentiment analysis and influence scoring. *Brandseye*[20] provides social media data monitoring and sentiment analysis for specific targets. *Rankur*[21] is another social media monitoring and online reputation management service provider. It is able to monitor online reviews, blogs, news, forums and social networks, and also can identify community leaders and customers' behaviors. *SocialMention*[22] provides the service of social media search and analysis. It is able to track and measure the content about a specific target, such as company, product or any other topic. *Google Alerts*[23] targets on content change detection and notification on social media platforms. New results, such as web pages, blogs and articles, can be sent to users through emails.

TABLE 6. Business data gathering platforms.

Platform	Description
Tackur	It provides social media monitoring, supporting all social media and mainstream news.
Brandseye	It provides social media data monitoring and sentiment analysis for specific targets.
Rankur	It is a social media monitoring and online reputation management service provider.
Social-Mention	It provides the service of social media search and analysis.
Google Alerts	It targets on content change detection and notification on social media platforms.
Gigwalk	It collects data of business performance in stores.
EasyShift	It takes photos of products, check prices, and review promotions by thousands of smartphone-armed shoppers.
Twentify	It enables companies to mobilize an on-demand workforce of smartphone users to collect data, and provides valuable insights and empowers better business decisions.

4) MOBILE CROWDSOURCING PLATFORMS

There are also many crowdsourcing platforms that harness the power of citizens for business data collection, as shown in Table 6. *Gigwalk*[24] collects data of business performance in stores. It helps brands manage channel execution at scale to collect data and get execution issues fixed by the crowd of Gigwalkers. *EasyShift*[25] takes photos of products, check prices, and review promotions by thousands of smartphone-armed shoppers. The shoppers can check in-store displays, assess competitive dynamics (record how competitor brands are stocked and displayed), report product out-of-stock and pricing information, and monitor promotions and launches, which enables brands to greatly improve the in-store execution and the shopper experience. *Twentify*[26] enables

[19] http://www.trackur.com/

[20] https://www.brandseye.com/

[21] https://rankur.com/

[22] http://socialmention.com/

[23] https://www.google.com/alerts

[24] http://www.gigwalk.com/

[25] http://www.easyshiftapp.com/

[26] http://www.twentify.com/

companies to mobilize an on-demand workforce of smartphone users to collect data, and provides valuable insights and empowers better business decisions. It provides an effective and efficient mobile insights platform allowing businesses of every size to reach hundreds of thousands of consumers and shoppers for their market research and field audit needs.

5) PRIVACY OF CROWDSOURCED DATA

Although the crowdsourced user-generated data can be exploited to provide tremendous benefits for users, the sharing of personal data (e.g., user location) with third-parties or the greater public can raise significant concerns about security and user privacy. To motivate user participation, many researchers explore techniques to protect information security in data collection, so that participants can conveniently and safely share high-quality data. Anonymizing user data is a popular method used in data publishing to protect user privacy. In addition, uploading the processed data rather than the original raw data is also a possible way. Besides some methodology efforts for input data, some systematic studies are also needed. For example, a privacy-aware architecture should be provided to support the development of crowdsourced applications.

B. CROWDSOURCED DATA PROCESSING

CrowdBI collects and processes crowdsourced data for business intelligence. The data from the uncontrolled crowd suffers from numerous issues, such as noise, low quality, data reliability, and so on.

1) NOISE ELIMINATION IN SENSING

For sensing data, there are many issues, such as noise in captured features due to multi-path and blockage, and a smaller number of reads or no reading of tags due to random back-off and collision [87]. Pang *et al.* [90] apply Principal Component Analysis to smooth the preprocessed CSI values in removing the bursty and impulse noises. Many factors may influence the reader's ability to obtain accurate Doppler estimates, such as the multiple effects in indoor environments, antenna switching, and frequency hopping. Han *et al.* [85] propose a Doppler frequency-based solution to detect tag movements, and employ a phase-based Doppler frequency estimation method to tackle the inaccurate measurements incurred by the current reader hardware.

2) HELPFUL DATA SELECTION

Online reviews play a crucial role in today's electronic commerce. Helpfulness is widely used to measure the performance of online reviews [28]. Ghose and Ipeirotis [36] find that the writing style of reviews plays an important role in determining the helpfulness of reviews. Specifically, they identify helpful reviews that have high predictive power at different level, including lexical, grammatical, semantic, and stylistic levels. Forman *et al.* [108] evaluate the helpfulness of online reviews for e-commerce. They characterize the helpfulness of reviews from two dimensions: reviewer expertise and attractiveness. Zhu *et al.* [109] find that the helpfulness of reviews on Yelp is mainly related to the central and peripheral cues of the argument. Besides the review content, the reviewer should also be considered. Ku *et al.* [110] find that four variables can discriminate reputable reviewers from others: 1) trust intensity, 2) average trust intensity of trustors, 3) degree of review focus in the target category, and 4) average product rating in the target category. Based on the analysis of three factors (i.e., the reviewer's expertise, the writing style of the review, and the timeliness of the review), Liu *et al.* [111] present a non-linear regression model to predict helpfulness, and demonstrate that the proposed approach is highly effective. Reference [28] identifies five types of features for review helpfulness identification, including linguistic features, features based on information quality, features based on information theory, reviewer features, and metadata features. Liao *et al.* [22] aim to extract representative posts and aspects from users' posts about the brands in social media. Specifically, they formulate the task as a submodular optimization problem, and apply the greedy-based algorithm to select representative posts. A summary of the business review selection is given in Table 7.

TABLE 7. Business review selection.

Feature	Description
Linguistics [24], [36], [111]	The length, volume, writing style of the sentences.
Information quality [22], [24]	The information coverage and information accuracy of reviews.
Information theory [24]	The information gain of reviews.
Reviewer [28], [108] [109], [110], [111],	The expertise and activeness of reviewers.
Metadata [28], [111]	The time of the review, descriptions of the review text.

3) SPAM DETECTION

Spam detection aims to detect the malicious manipulation of user generated data. Review spams in website are designed to give unfair view of some products so as to influence the consumers' perception of the products and inflate the product's reputation. A product or store with positive ratings and a high proportion of positive reviews will attract more customers and larger amount of business, while a couple of negative reviews/ratings could substantially harm the reputation, leading to financial losses. Since there is no rule governing online reviews and ratings, some product providers or retailers are leveraging such public media to defame competitors and promote themselves unfairly, or even to cover the truth disclosed by genuine reviews. Due to the spam reviews, customers can be misled to buy low-quality products, while decent stores can be defamed by malicious reviews.

Therefore, it is crucial to detect spam reviews or untruthful reviews. Jindal and Liu [112] study the problem of product review spammer detection, and they identify three categories of spams: fake reviews (also called untruthful opinions), reviews on brand only, and non-reviews. Xie *et al.* [113]

observe that the normal reviewers' arrival pattern is stable and uncorrelated to their rating pattern temporally. Nevertheless, the spam attacks are usually bursty and either positively or negatively correlated to the rating. Thus, they propose a hierarchical algorithm to detect such attacks via unusually correlated temporal patterns. Lim *et al.* [114] use reviewers' behaviors as indicators of spamming. The reviewers' behaviors include multiple similar rating on a single product, similarity between reviews written by a single reviewer, etc. Mukherjee *et al.* [115] study the problem of group spamming, where a group of spammers write spam reviews together on a few target stores. They use frequent pattern mining and group behavior analysis to detect group spamming. Wang *et al.* [116] introduce a heterogeneous graph model to discover the reinforcement relations of reviewers' trustiness, reviews' honesty, and stores' reliability, which are used to discover suspicious spammers. Mukherjee *et al.* [117] propose an unsupervised Bayesian framework to exploit observed behavioral footprints to detect spammers (e.g., fake reviewers). The intuition is that opinion spammers have different behavioral distributions than non-spammers. KC and Mukherjee [118] discover various temporal patterns and their relationships with the rate at which fake reviews are posted on Yelp. They also leverage the discovered temporal patterns in deception detection. Li *et al.* [119] discover that reviewers' posting rates follow a bimodal distribution pattern. Multiple spammers tend to collectively and actively post reviews to the same set of products within a short time frame, the so-called co-bursting. Based on these findings, they propose the Coupled Hidden Markov Model (CHMM) to detect spams.

4) DATA AGGREGATION

One major challenge in data aggregation is the data variance, e.g., opinion heterogeneity expressed in each review that may be influenced by their personal characteristics. Xiao *et al.* [26] propose an econometric model, referred to as the modified ordered choice model (MOCM), to measure aggregated consumer preferences from online product reviews. This model considers the heteroscedasticity of reviewers' rating variance and allows reviewers to assign rating scores according to their own thresholds. Kamar *et al.* [120] use probabilistic graphical models to model the learning task and human bias. Reviews vary in quality; thus, it is desirable to differentiate the reviews in terms of quality in order to achieve better prediction performance. Yu *et al.* [35] use the writing style (e.g., part of speech) of a review [121] to help predict its quality.

Another major challenge in data aggregation is *truth discovery*. Crowdsourced data may provide conflict information for the same object, the task of truth discovery is to detect trustworthy information by identifying reliable sources [122], [123]. To validate the trustworthiness of crowdsourced data, reputation and trust modeling are important [124]. Various truth discovery methods have been developed based on the following general principle: the users who provide trustworthy answers are more reliable, and the answers from reliable users are more trustworthy. This principle tightly combines the process of user reliability estimation with that of information trustworthiness inference. Hung *et al.* [125] aim to estimate the user reliability and infer trustworthy answers from noisy user-generated data. They propose a novel Bayesian nonparametric model to aggregate the partial-agreement answers in a generic way. Sheng *et al.* [126] study the aggregation of crowdsourced data based on two ideas, majority voting and paring. They find that majority voting strategies work well under the situation where the certainty level is high. While the pairing strategies are more preferable under the situation where the certainty level is low.

5) DATA HETEROGENEITY

CrowdBI collects and processes crowdsourced data from multiple sources, such as offline sensing data in the physical world and online data in the virtual world (e.g., data from social media). Obviously, the heterogeneity of multi-source data will bring new opportunities to develop new techniques to deal with a lot of data from different dimensional sources. Generally, in order to integrate the information from heterogeneous data sources to attain a comprehensive understanding, most works extract different types of features from heterogeneous data based on empirical and expert knowledge. In addition, there are some studies aiming to develop the system that can directly mine information from raw data without any manual feature engineering.

C. CROWDSOURCED BUSINESS KNOWLEDGE MINING

In CrowdBI, crowdsourced business knowledge mining aims to explore and mine business knowledge from crowdsourced data, including *customer behavior*, *product aspects*, *customer requirement* and *competitive relationship*. A summary of the representative works on business knowledge mining is given in Table 8.

TABLE 8. Representative works on business knowledge mining.

Knowledge	Method
Customer behavior	Econometrics model [13], [14]
Product aspects	Topic model [71], [128], ranking algorithm [127]
Customer requirement	Conjoint analysis and Kano model [28], [26], knowledge-based method and probabilistic model [129]
Market structure	Conditional random field [47]
Competitive relationship	Conditional random field [130], class sequential rules mining and naïve Bayesian Classification [51], label sequential rules [52]

1) CUSTOMER BEHAVIOR UNDERSTANDING

The customer behavior analysis aims to understand and mine customers' behaviors by incorporating multi-source crowdsourced data. Zhang *et al.* [13] extract traffic and human mobility features (e.g., static spatial features, human

mobility features) from crowdsourced data. Then, they use econometrics model to analyze the impact of these extracted features to local business. Goel *et al.* [14] extract features from user search behavior and apply econometrics model to forecast the item demand.

2) ASPECT MINING

Generally, given a collection of reviews about products, aspect-based mining approaches can be applied to extract the aspects of a product (e.g., price and service), as well as inferring sentiment polarity for each aspect. Wang *et al.* [71] aim to understand the preferences of customers about the restaurant. Specifically, they adopt an LDA-based topic model to identify aspects from restaurant reviews. Zha *et al.* [127] propose a product aspect ranking framework to identify the important aspects of products from online consumer reviews. They identify the important aspects based on two observations: 1) the important aspects are usually commented on by a large number of consumers, and 2) consumer opinions on the important aspects greatly influence their overall opinions on the product. Mukherjee and Liu [128] aim to discover aspects in hotel reviews, they propose two statistical models to jointly model both aspects and aspect specific sentiments.

3) CUSTOMER REQUIREMENT MINING

Requirement measurement is fundamental to product positioning and strategic marketing. To facilitate product improvement, Qi *et al.* [28] propose an automatic filtering model to predict the helpfulness of online reviews from the perspective of the product designer. Specifically, they apply the Kano model [27] to analyze the online reviews to develop appropriate product improvement strategies. Xiao *et al.* [26] extract consumer preferences from online product reviews, then categorize customer requirements based on the extracted consumer preferences. Furthermore, they extend the Kano model and propose a marginal effect-based Kano model to estimate the importance of customer requirements. Tutubalina [129] aim to the extract problem phrases from user reviews about products. they propose knowledge-based methods and probabilistic models to classify users' phrases and extract latent problem indicators, aspects and related sentiments from online reviews.

4) COMPETITIVE RELATIONSHIP MINING

To understand the competitors' products, Netzer *et al.* [47] explore the market structure derived from online discussions to understand the co-mention of brands within a sentence or paragraph, and extract the relationship between the brands. Lafferty *et al.* [130] use conditional random field (CRF) approach to extract comparative relation among multiple entities. Jindal and Liu [51] use integrated pattern discovery and supervised learning approach to identify comparative sentences in text documents. Jindal and Liu [52] identify the comparative sentences and extract the comparative relations, which are classified into three types: ''non-equal gradable'', ''same'', and ''superlative''.

D. INTEGRATING ECONOMIC AND DATA MINING MODELS

As an interdisciplinary research field, CrowdBI should study the combination of economic models and machine learning models, which will generate added value by taking advantage of their complementary features.

1) CUSTOMER REQUIREMENT ANALYSIS

The objective of customer requirement analysis is to determine what combination of a limited number of attributes is influential on respondent choice or decision making. Conjoint analysis [131] is a survey-based statistical technique used in market research that helps determine how people value different attributes (e.g., features, functions, benefits) that make up an individual product or service. It depends strongly on survey data and its data collection process is time consuming and costly.

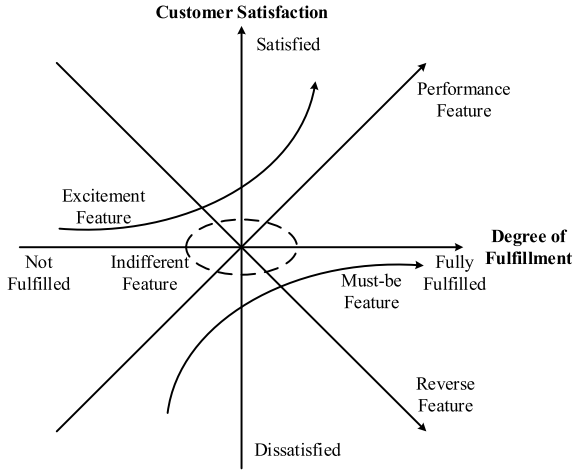


FIGURE 3. Kano model with crowdsourced big data.

The other method used to evaluate the impact of product attributes on customer satisfaction is *Kano model* [26]. It classifies product features into different categories, according to the degree of fulfillment of the product features and their effect on customer satisfaction. The Kano model with crowdsourced big data is shown in Fig. 3.

- *Must-be features.* These product features are taken for granted when fulfilled.
- *Performance features.* If the fulfillment is positively associated with customer satisfaction, the product features are considered as performance features.
- *Excitement features.* These product features are the opposites of must-be features. Excitement features offer satisfaction when fulfilled but do not result in dissatisfaction if not fulfilled.
- *Indifferent features.* When the degree of fulfillment of a product feature is either not associated or only marginally associated with customer satisfaction/dissatisfaction, this feature is referred to as an indifferent feature.

- *Reverse features.* It implies that when the degree of fulfillment increases, customers become more dissatisfied.

2) TIME SERIES ANALYSIS

Is a well-established field with a large body of the literature [132], [133]. Its goal is to reveal the underlying forces and structures, and monitor or forecast the time series data based on the observed data. It has been widely applied in a large variety of areas in business, such as economic forecasting, sales forecasting, stock market analysis, etc. [35]. State-of-the art time series analysis methods, such as ARIMA (Autoregressive Integrated Moving Average models), Markov chain, wavelet transform based prediction [134] and RNN (recurrent neural networks) [135] are widely applied in evolution analysis. There are also methods that explore the connections among a group of time series [136], combine matrix factorization with time series analysis for time-evolving forecasting [137], etc.

Econometrics is widely used for panel analysis and time-series analysis. Econometrics is a well-established statistics technique to test hypotheses and to predict future changes. It aims to identify and quantify the causal effects of different features on the economic outcome, i.e., the so-called identification. Instrumental variable regression is a general way in econometrics to rule out of the reverse causal explanations and unobservable variables that can cause both the performance outcome and features. For example, Zhang *et al.* [13] extract traffic and human mobility features from Manhattan restaurants data and study how static and dynamic factors affect the economic outcome of local businesses in the city. Hedonic regression [16] is also commonly used in econometrics to identify the weight of individual features in determining the overall price of a product [17]. Specifically, Hedonic models are designed to estimate the value that different product aspects (e.g., product price) contribute to a consumer's utility.

3) RELATIONS BETWEEN ECONOMIC MODEL AND DATA MINING MODEL

Economic models usually focus on one or a few coefficients of interest, which often represent the causal effect of a particular policy or policies [138]. Economists often focus on assessing the results of a specific policy or testing theories that predict a particular causal relationship. This approach is basically different from some of the data mining methods that have become popular for large-data applications in computer science. Data mining methods focus on prediction, which uses data-driven model selection to identify the most meaningful predictive variables. It pays less attention to statistical uncertainty and standard errors. Economic models, however, tend to place a high degree of importance on the identification of causal effects and on statistical inference to assess the significance of these effects. Having a model with an overall high degree of predictive fit is often viewed as secondary to finding a specification that cleanly identifies a causal effect.

The two approaches are not necessarily in competition, and should be combined for business intelligence. For instance, if only a subset of control variables is truly predictive, an automated model-selection approach may be helpful to identify the relevant ones. Data mining methods may also be useful if there are important interaction effects so that one cares about predicting effects for specific individuals rather than an average effect for the population. Economic theory also plays a crucial role in the analysis of large data sets, in large part because the complexity of many new data sets calls for simpler organizing frameworks. Economic models are useful for this purpose. The development of business depends on the interplay between the predictive modeling and the interplays of the auction. Therefore, making decisions about how to run the market requires a sophisticated understanding of both big data predictive modeling and economic theory. We should investigate the combined usage of economic models and data mining models. The relationships between economic and data mining methods are presented in Fig. 4.

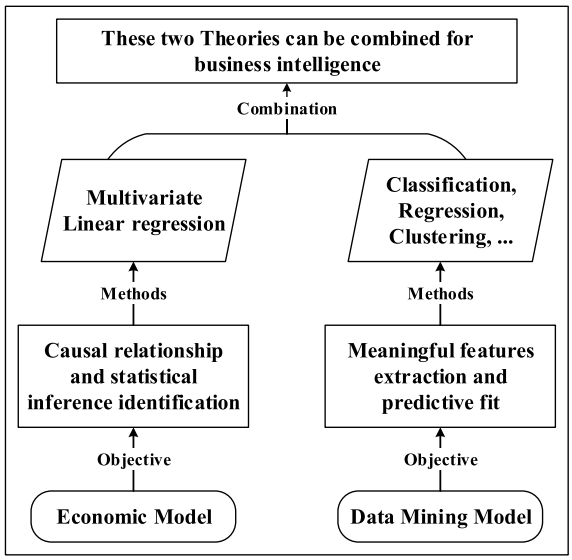


FIGURE 4. Relationships between economic and data mining methods.

V. CROWDBI FOR NEW ECONOMY

Besides traditional markets and products, the development of IT technology is driving new forms of economy. This leads to new challenges and uncertainty to retailers, customers, as well as the managers. However, the benefit is that new economy facilitates data collection from consumers. Therefore, CrowdBI will act an important role in the future of economy based on the large amount of crowd-contributed data.

A. MOBILE AND IoT PRODUCTS

With the development of mobile devices and Internet of Things, we study new type of products and their new features, including mobile apps and IoT products.

1) MOBILE APPS

Mobile apps are now being used worldwide to accomplish a variety of things, such as accessing social networks,

reading e-books, playing games, and many other meaningful aspects of our lives. The increasing availability of app-level data presents exciting opportunities for business analytics. Mobile-related business analytics such as estimating demand for apps can help developers design apps that are more likely to engage users. Leveraging mobile app analytics can optimize app content and navigation, improved app merchandizing such as placement of ads and product links within apps, and so on.

Competition in a mobile apps market plays an irreplaceable role in natural selection. Guo *et al.* [139] propose a mobile app competitor analysis system, which can conduct competitor analysis based on processing online text content automatically. Through analyzing users' reviews, the system then recommends some operational strategies for developers. Analyzing the demands of users is also important for mobile app business. For example, Ghose and Han [140] present the first formal model of app-level demand estimation that is based on the app downloading behavior of users. To quantify the competitiveness about mobile apps between different platforms, they build a structural econometric model and estimate consumer preferences towards different mobile app characteristics.

There are some important features needed to be considered when analyzing the mobile app, such as labels of apps, time series of apps, user reviews, etc. The labels of apps play an important role in app stores when users search different categories of apps. To reduce the workload to label apps, Liu *et al.* [141] develop a framework that automatically labels apps with a richer and more detailed categorization. Wang *et al.* [137] propose the evolutionary hierarchical competition model (EHCM) to model the time-evolving hierarchical competition among products, and it provides more accurate product sales forecasting. Zhu *et al.* [142] propose a sequential approach based on hidden Markov model (HMM) for modeling the popularity information of mobile Apps toward mobile App services. User review is a crucial component in app markets which reflects users' preference. Fu *et al.* [143] propose WisCom, a multi-level system that can analyze tens of millions user reviews at various granularities. They provide a dynamic view of how users' opinions evolve over time, and identify reasons why users like or dislike a given app. Di Sorbo *et al.* [144] propose the SURF (Summarizer of User Reviews Feedback) approach to condense the enormous amount of information that developers of popular apps have to manage due to user feedback received on a daily basis. Li *et al.* [145] present an empirical analysis of a very large collection of app management activities of smartphone users. They identify useful patterns from the app management activities that can accurately predict the user preferences of an app, such as the number of users/devices that have downloaded the app, or the number of most recent activities.

2) IoT PRODUCTS

The phrase ''Internet of Things'' has arisen to reflect the growing number of smart, connected products and highlight the new opportunities they can represent. For the IoT products in BI, they not only provide novel functionalities, but also may establish new business models, markets, or distribution channels, strengthen relationships with consumers, and add smart remote services. While IoT industry is still in its infancy, it has already become rather obvious that those IoT products have the potential to revolutionize their respective markets and it is widely expected that their era is set to start in the near future.

At present, many IoT products for some applications have already been overcome, such as IoT devices or products in smart homes. However, the upcoming market dissemination of smart products poses new managerial challenges that may be addressed by researchers from economics, business administration, and other non-technical fields. On the one hand, smarts products in business are forcing companies to redefine their industries and rethink nearly everything they do, beginning with their strategies [146]. On the other hand, smart products require functions within manufacturing firms to collaborate in new ways. In general, IoT products with unique functions will accelerate the development of BI and the evolvement of firms' structures.

Dawid *et al.* [147] first analyze and overview the technical potential and future trends of smart consumer products. For example, from a technical perspective, smart products can be characterized and improved along the dimensions including input, output, human-computer interfaces, interoperability, integration and resource-efficiency. IoT products are different from traditional products, and they require companies to build and support an entirely new technology infrastructure [146]. For example, the product should be made up of multiple layers, including new product hardware, embedded software, connectivity, and a product cloud. Based on the infrastructure, the smart products in business can have functions, including monitoring, control, optimization, and autonomy. In general, smart, connected products will have a transformative effect on the industry structure.

There are still numerous challenges to be tackled regarding smart products. The high degree of innovativeness of smart products may make the added value unclear to consumers. It limits the applicability of traditional approaches for eliciting consumer preferences or for estimating the willingness to pay. Furthermore, smart products may fundamentally change the way products and services are brought to the market, thus, they may impact the ''rules of engagement'' in existing markets or even create new markets and/or allow for novel business models. The analysis of innovation diffusion mechanisms for smart products differs from that of traditional products in several respects, such as a variety of data.

B. OMNI-CHANNEL RETAILING

In the past, *brick-and-mortar retail* stores were unique in allowing consumers to touch and feel merchandise and provide instant gratification. With the development of the Internet, *online retails* try to woo shoppers with wide product selection, low prices, and content such as product reviews

and ratings. As the retailing industry evolves toward a seamless *omni-channel retailing* experience, the distinctions between physical and online will vanish, turning the world into a showroom without walls [148], [149]. In general, the changes are driven by new technologies, such as smart mobile devices (e.g., smartphones and tablets) and related software (e.g., apps, mobile payments, e-coupons, digital flyers, location-based services). Especially, there are some new in-store technologies which accelerate the development of omni-channel retailing, such as virtual screens and aisles, virtual mirrors in fitting rooms, digital signage, and intelligent self-service kiosks. From the perspective of customers, they expect uniform and integrated service/experience, regardless of the channel they use. In addition, they are willing to move seamlessly between different channels (e.g., traditional store, online, and mobile devices) depending on their preferences and context. Therefore, the omni-channel retailing experience is able to break down old barriers. A summary of different retailing manners is given in Table 9.

TABLE 9. Different retailing manners.

Work	Data/Feature	Description
Brick-and-mortar retailing	Buying products in physical retail spaces; Touching and feeling merchandise.	Providing information, services and instant gratification.
Online retailing	Using online data and analytics to better understand customer needs and values.	Providing low prices and neatly curated content; Converting "experience goods" to "search goods".
Omni-channel retailing	Integrating multiple channels; Using new technologies to obtain customer behaviors; Utilizing heterogeneous data to understand customer preference.	Turning the world into a showroom without walls; Combing the offline and online world; Providing seamless checkout.

1) MULTI-SOURCE DATA

Different from traditional e-commerce, retailing should combine the characterization of both offline and online world. Specifically, omni-channel retailing is an explosion of new data from social, mobile and local channels. This provides an unprecedented opportunity to understand not just customer transactions but also customer interactions such as visits to the store, likes on Facebook, searchers on websites, and check-ins at nearby establishments. Recently, mashups of data from multiple sources will give savvy retailers an ability to do predictive analytics to make location- and time-specific offers and recommendations to each of their potential and existing customers [148]. Cao and Li [150] focus on how the integration of multiple channels impacts performance, using secondary data on the use of multiple channels in the US retail industry.

2) SENSING

The future of retail is a seamless integration between the commercial and personal space. Seamless checkout is the key technology to realize this vision. For example, Amazon Go[27] is one of the most recent attempts to improve the checkout process, by allowing users to come in the stores, pick up items, and head out. It uses vision-based technology combined with sensor fusion to automate the checkout process. Besides vision-based techniques, there are some other sensors which can supplement some poor characterization of computer vision (e.g., limited retailer insight and personalization). There is an option to build a customized energy-efficient RFID reader which is solely focused toward retail scenario to help in faster decoding of RFID tags utilizing a few recent advances. For example, Pradhan *et al.* [87] propose an RFID-based smart shopping system that can track purchases in real-time.

C. SHARED ECONOMY

Shared economy platforms have become extremely popular in the past few years, and they have changed the way in which we commute, travel, and borrow among many other activities. Contrary to the traditional market model, which is based on ownership, *the sharing economy is built on using and sharing of products and services among others.* The rapid growth of the sharing economy, exemplified by ridesharing platforms Uber,[28] Lyft,[29] and Didi,[30] bike sharing platforms Mobike[31] and Ofo,[32] mobile crowdsourcing platforms TaskRabbit[33] and CrowdFlower,[34] as well as home-sharing platforms Airbnb,[35] is changing the patterns of ownership and consumption of goods and services. At present, the sharing economy has begun to permeate nearly every sector of the economy. According to [151], it was already worth US$15 billion in 2013, and was projected to grow to US$335 billion by 2025. In a sharing economy, consumers exchange services in a peer-to-peer fashion, through matching markets facilitated by social networks and online applications. The various benefits provided to consumers, such as convenience, cost savings, and new social interactions, have fueled the sharing economy's rapid growth [152].

1) IMPACT ANALYSIS

A fundamental question for the sharing economy is how to quantify the impact of the platforms on incumbent firms. Quantifying such impact is important for several reasons such as (1) helping municipalities and regulators define

[27] https://www.amazon.com/b?node=16008589011
[28] https://www.uber.com.cn/
[29] https://www.lyft.com/
[30] http://www.didiglobal.com/#/
[31] https://mobike.com/cn/
[32] http://www.ofo.so/#/
[33] https://www.taskrabbit.com/
[34] https://www.figure-eight.com/
[35] https://zh.airbnb.com/

appropriate laws to better regulate the sharing economy and (2) informing incumbent firms about new competitors that they should (or should not) pay attention to when developing their marketing strategies. Airbnb is an online community marketplace facilitating short-term rentals ranging from shared accommodations to entire homes, which is usually studied as one of typical cases to analyze the impact of shared economy. Zervas *et al.* examine the impact that the rise of the sharing economy has on the hospitality industry, such as hotel room revenue [153], [154]. Specifically, they identify Airbnb's impact by exploiting significant spatio-temporal variation in the patterns of adoption across city-level markets. By examining data from about 14,000 Airbnb listings and monthly revenue of approximately 3000 hotels in Texas, they find that the entry of Airbnb in Texas negatively affected hotel room revenue. Because of such negative externalities, the sharing economy and its regulation have become highly popular policy topics. Quattrone *et al.* [155] propose to perform the socio-economic analysis of Airbnb adoption to envision regulations. After crawling Airbnb data for the entire city of London, they find out where and when Airbnb listings are offered. Then the socio-economic conditions of the areas are determined which benefit from the hospitality platform.

Ridesharing is another typical service in shared economy, which enable consumers to request rides from other people who own private able consumers to request rides from other people who own private. Despite its growing popularity, there are some studies that examine the factors affecting user participation in the sharing economy. Kooti *et al.* [156] examine large-scale Uber data covering 59 million rides. It evaluates the impact of dynamic pricing (i.e., surge pricing) and income on both rider and driver behavior. They find that surge pricing does not favor more affluent riders, but mostly affects younger riders. Drivers with many surge rides receive lower ratings, on average, suggesting the riders' dislike of surge pricing. Furthermore, in order to understand the impact of surge pricing on passengers and drivers, Chen *et al.* [157] present the in-depth investigation on the Uber platform across locations in New York City and San Francisco. By analyzing the movements and actions of Uber drivers, they observe that surge prices have a small, positive effect on vehicle supply, and a large, negative impact on passenger demand. Chen and Sheldon [158] study how driver-partners on the Uber platform respond to surge pricing. They find that Uber partners drive more at high surge times, and that surge pricing significantly increases the supply of rides on the Uber system. This suggests that surge pricing significantly increases the number of trips that occur, and boosts the overall efficiency of the Uber system. Based on the vast amount of accurate data obtained by ridesharing services, Guo *et al.* [159] analyze the nature of the demand and dynamic pricing mechanisms that match the supply with demand. In demand analysis, they discuss its general characteristics, passenger grouping and demand clustering; in dynamic pricing analysis, they discuss the pattern and determination of dynamic pricing multipliers. With the emergence of ride-on-demand services, there may be some competition among different ridesharing platforms. For example, a battle between two Chinese taxi booking mobile apps, namely, Didi and Kuaidadi, had recently occurred in early 2014. Leng *et al.* [160] study the debates on social justice, equity, and improvements of taxi service under the development of taxi booking mobile apps: Didi and Kuaidi. They found that productively and critically employing big data can help address long-standing questions of social justice, equity, and many other concerns.

VI. CONCLUSIONS

This paper introduces a new business intelligence, named Crowdsourced Business Intelligence (CrowdBI). First, we illustrate the definition and concept model of CrowdBI. The generic framework of CrowdBI is presented, consisting of data collection, data preparation, business knowledge learning, decision making and recommendation. Second, we develop a taxonomy of the major application areas of CrowdBI, including customer behavior analysis, brand tracking and product improvement, demand forecasting and trend analysis, competitive intelligence, business popularity analysis and site recommendation, and urban commercial analysis. Third, we investigate the research challenges and key techniques, such as fine-grained data collection, crowdsourced data processing, crowdsourced business knowledge mining, and integrating economic and data mining models. Finally, we discuss the new economy in CrowdBI, including mobile and IoT products, omni-channel retailing, and shared economy.

REFERENCES

[1] S. Negash, "Business intelligence," *Commun. Assoc. Inf. Syst.*, vol. 13, Feb. 2004, Art. no. 15.
[2] H. P. Luhn, "A business intelligence system," *IBM J. Res. Develop.*, vol. 2, no. 4, pp. 314–319, Oct. 1958.
[3] N. Dedić and C. Stanier, "Measuring the success of changes to existing business intelligence solutions to improve business intelligence reporting," in *Research and Practical Issues of Enterprise Information Systems*. Cham, Switzerland: Springer, 2016, pp. 225–236.
[4] J. Leskovec, "Social media analytics: Tracking, modeling and predicting the flow of information through networks," in *Proc. 20th Int. Conf. Companion World Wide Web*, 2011, pp. 277–278.
[5] L. Dey, S. M. Haque, A. Khurdiya, and G. Shroff, "Acquiring competitive intelligence from social media," in *Proc. Joint Workshop Multilingual OCR Anal. Noisy Unstructured Text Data*, 2011, p. 3.
[6] H. Schoen, D. Gayo-Avello, T. M. Panagiotis, M. Eni, S. Markus, and G. Peter, "The power of prediction with social media," *Internet Res.*, vol. 23, no. 5, pp. 528–543, 2013.
[7] B. Evelson and N. Norman, "Topic overview: Business intelligence," Forrester Research, Cambridge, MA, USA, Tech. Rep., 2008.
[8] J. Howe and M. Robinson, "Crowdsourcing: A definition," in *Crowdsourcing: Tracking the Rise of the Amateur*. Weblog, 2006.
[9] B. Guo, Y. Liu, L. Wang, V. O. Li, C. K. Jacqueline, and Z. Yu, "Task allocation in spatial crowdsourcing: Current state and future directions," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1749–1764, Jun. 2018.
[10] B. Guo *et al.*, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm," *ACM Comput. Surv.*, vol. 48, no. 1, p. 7, 2015.
[11] H. J. Watson and B. H. Wixom, "The current state of business intelligence," *Computer*, vol. 40, no. 9, pp. 96–99, 2007.
[12] *Regional Economic Prospects: Analysis and Projections to the Year 2000*, Cambridge Econ., Cambridge, U.K., 1987.

[13] Y. Zhang, B. Li, and J. Hong, "Understanding user economic behavior in the city using large-scale geotagged and crowdsourced data," in *Proc. 25th Int. Conf. World Wide Web*, Apr. 2016, pp. 205–214.

[14] S. Goel, J. M. Hofman, S. Lahaie, D. M. Pennock, and D. J. Watts, "Predicting consumer behavior with Web search," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 41, pp. 17486–17490, 2010.

[15] A. Ghose, P. G. Ipeirotis, and B. Li, "Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content," *Marketing Sci.*, vol. 31, no. 3, pp. 493–520, 2012.

[16] N. Archak, A. Ghose, and P. G. Ipeirotis, "Show me the money!: Deriving the pricing power of product features by mining consumer reviews," in *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2007, pp. 56–65.

[17] S. Rosen, "Hedonic prices and implicit markets: Product differentiation in pure competition," *J. Political Economy*, vol. 82, no. 1, pp. 34–55, 1974.

[18] Y. Gao, F. Wang, H. Luan, and T.-S. Chua, "Brand data gathering from live social media streams," in *Proc. Int. Conf. Multimedia Retr.*, 2014, p. 169.

[19] S. Wu, W. Rand, and L. Raschid, "Recommendations in social media for brand monitoring," in *Proc. 5th ACM Conf. Recommender Syst.*, 2011, pp. 345–348.

[20] N. F. Ibrahim, X. Wang, and H. Bourne, "Exploring the effect of user engagement in online brand communities: Evidence from Twitter," *Comput. Hum. Behav.*, vol. 72, pp. 321–338, Jul. 2017.

[21] M. Mazloom, R. Rietveld, S. Rudinac, M. Worring, and W. Van Dolen, "Multimodal popularity prediction of brand-related social media posts," in *Proc. ACM Multimedia Conf.*, 2016, pp. 197–201.

[22] L. Liao, X. He, Z. Ren, L. Nie, H. Xu, and T. S. Chua, "Representativeness-aware aspect analysis for brand monitoring in social media," im *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 310–316.

[23] B.-C. Chen, Y.-Y. Chen, F. Chen, and D. Joshi, "Business-aware visual concept discovery from social media for multimodal business venue recognition," in *Proc. AAAI*, 2016, pp. 101–107.

[24] Y. Liu, J. Jin, P. Ji, J. A. Harding, and R. Y. Fung, "Identifying helpful online reviews: A product designer's perspective," *Comput.-Aided Des.*, vol. 45, no. 2, pp. 180–194, 2013.

[25] R. Decker and M. Trusov, "Estimating aggregate consumer preferences from online product reviews," *Int. J. Res. Marketing*, vol. 27, no. 4, pp. 293–307, 2010.

[26] S. Xiao, C.-P. Wei, and M. Dong, "Crowd intelligence: Analyzing online product reviews for preference measurement," *Inf. Manage.*, vol. 53, no. 2, pp. 169–182, 2016.

[27] N. Kano, S. Nobuhiko, T. Fumio, and T. Shinichi, "Attractive quality and must-be quality," *J. Jpn. Soc. Service Qual. Control*, vol. 14, no. 2, pp. 39–48, Apr. 1984.

[28] J. Qi, Z. Zhang, S. Jeon, and Y. Zhou, "Mining customer requirements from online reviews: A product improvement perspective," *Inf. Manage.*, vol. 53, no. 8, pp. 951–963, 2016.

[29] K. L. Xie, Z. Zhang, and Z. Zhang, "The business value of online consumer reviews and management response to hotel performance," *Int. J. Hospitality Manage.*, vol. 43, pp. 1–12, Oct. 2014.

[30] Z. Xiang, Z. Schwartz, J. H. Gerdes, Jr., and M. Uysal, "What can big data and text analytics tell us about hotel guest experience and satisfaction?" *Int. J. Hospitality Manage.*, vol. 44, pp. 120–130, Jan. 2015.

[31] X. Yu, Y. Liu, X. Huang, and A. An, "A quality-aware model for sales prediction using reviews," in *Proc. 19th Int. Conf. World Wide Web*, 2010, pp. 1217–1218.

[32] L. Wu and E. Brynjolfsson, "The future of prediction: How Google searches foreshadow housing prices and sales," in *Economic Analysis of the Digital Economy*. Chicago, IL, USA: Univ. of Chicago Press, 2015, pp. 89–118.

[33] D. Gruhl, R. Guha, R. Kumar, J. Novak, and A. Tomkins, "The predictive power of online chatter," *Proc. 11th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2005, pp. 78–87.

[34] Y. Liu, X. Huang, A. An, and X. Yu, "ARSA: A sentiment-aware model for predicting sales performance using blogs," in *Proc. ACM 30th Annu. Int. SIGIR Conf. Res. Develop. Inf. Retr.*, 2007, pp. 607–614.

[35] X. Yu, Y. Liu, X. Huang, and A. An, "Mining online reviews for predicting sales performance: A case study in the movie domain," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 4, pp. 720–734, 2012.

[36] A. Ghose and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 10, pp. 1498–1512, Oct. 2011.

[37] I. Zheludev, R. Smith, and T. Aste, "When can social media lead financial markets?" *Sci. Rep.*, vol. 4, p. 4213, Feb. 2014.

[38] G. Mishne and N. S. Glance, "Predicting movie sales from blogger sentiment," in *Proc. AAAI Spring Symp., Comput. Approaches Analyzing Weblogs*, 2006, pp. 155–158.

[39] J. Lian, F. Zhang, X. Xie, and G. Sun, "Restaurant survival analysis with heterogeneous information," in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 993–1002.

[40] J. Wang, W. X. Zhao, H. Wei, H. Yan, and X. Li, "Mining new business opportunities: Identifying trend related products by leveraging commercial intents from microblogs," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2013, pp. 1337–1347.

[41] M. Arias, A. Arratia, and R. Xuriguera, "Forecasting with twitter data," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 1, p. 8, 2013.

[42] L. Festinger, "A theory of social comparison processes," *Human Rel.*, vol. 7, no. 2, pp. 117–140, 1954.

[43] L. Kahaner, *Competitive Intelligence: How to Gather Analyze and Use Information to Move Your Business to the Top*. New York, NY, USA: Simon and Schuster, 1997.

[44] H. Chen, M. Chau, and D. Zeng, "CI spider: A tool for competitive intelligence on the Web," *Decision Support Syst.*, vol. 34, no. 1, pp. 1–17, 2002.

[45] W. He *et al.*, "Gaining competitive intelligence from social media data: Evidence from two largest retail chains in the world," *Ind. Manage. Data Syst.*, vol. 115, no. 9, pp. 1622–1636, 2015.

[46] K. Xu, S. S. Liao, J. Li, and Y. Song, "Mining comparative opinions from customer reviews for competitive intelligence," *Decis. Support Syst.*, vol. 50, no. 4, pp. 743–754, 2011.

[47] O. Netzer, R. Feldman, J. Goldenberg, and M. Fresko, "Mine your own business: Market-structure surveillance through text mining," *Marketing Sci.*, vol. 31, no. 3, pp. 521–543, 2012.

[48] S. Bao, R. Li, Y. Yu, and Y. Cao, "Competitor mining with the Web," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 10, pp. 1297–1310, Oct. 2008.

[49] Y. Yang *et al.*, "Mining competitive relationships by learning across heterogeneous networks," in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage.*, 2012, pp. 1432–1441.

[50] Y. Yang, J. Tang, and J. Li, "Learning to infer competitive relationships in heterogeneous networks," *ACM Trans. Knowl. Discovery Data*, vol. 12, no. 1, p. 12, 2018.

[51] N. Jindal and B. Liu, "Identifying comparative sentences in text documents," in *Proc. 29th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2006, pp. 244–251.

[52] N. Jindal and B. Liu, "Mining comparative sentences and relations," in *Proc. AAAI*, vol. 22, no. 13311336, p. 9, 2006.

[53] W. He, H. Wu, G. Yan, V. Akula, and J. Shen, "A novel social media competitive analytics framework with sentiment benchmarks," *Inf. Manage.*, vol. 52, no. 7, pp. 801–812, 2015.

[54] H. Zhang, G. Kim, and E. P. Xing, "Dynamic topic modeling for monitoring market competition from online text and image data," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 1425–1434.

[55] Z. Zheng, P. Fader, and B. Padmanabhan, "From business intelligence to competitive intelligence: Inferring competitive measures using augmented site-centric data," *Inf. Syst. Res.*, vol. 23, pp. 698–720, Sep. 2012.

[56] B. Sharp, *How Brands Grow: What Marketers Don't Know*. Oxford, U.K.: Oxford Univ. Press, 2010.

[57] Z. Zhang, C. Guo, and P. Goes, "Product comparison networks for competitive analysis of online word-of-mouth," *ACM Trans. Manage. Inf. Syst.*, vol. 3, no. 4, p. 20, 2013.

[58] W. He, S. Zha, and L. Li, "Social media competitive analysis and text mining: A case study in the pizza industry," *Int. J. Inf. Manage.*, vol. 33, no. 3, pp. 464–472, 2013.

[59] Y. Kim, R. Dwivedi, J. Zhang, and S. R. Jeong, "Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5," *Online Inf. Rev.*, vol. 40, no. 1, pp. 42–61, 2016.

[60] Q. Wei, D. Qiao, J. Zhang, G. Chen, and X. Guo, "A novel bipartite graph based competitiveness degree analysis from query logs," *ACM Trans. Knowl. Discovery Data)*, vol. 11, no. 2, p. 21, 2016.

[61] B. Thau. (2015). *How Big Data Helps Chains Like Starbucks Pick Store Locations—An (Unsung) Key to Retail Success*. [Online]. Available: http://onforb.es/1iijr2o

[62] A. Athiyaman, "Location decision making: The case of retail service development in a closed population," *Acad. Marketing Stud. J.*, vol. 15, no. 1, p. 87, 2011.
[63] P. Jensen, "Network-based predictions of retail store commercial categories and optimal locations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 3, 2006, Art. no. 035101.
[64] S. Porta *et al.*, "Street centrality and densities of retail and services in Bologna, Italy," *Environ. Planning B, Planning Des.*, vol. 36, no. 3, pp. 450–465, 2009.
[65] S. Porta *et al.*, "Street centrality and the location of economic activities in Barcelona," *Urban Stud.*, vol. 49, no. 7, pp. 1471–1488, 2012.
[66] D. L. Huff, "A probabilistic analysis of shopping center trade areas," *Land Econ.*, vol. 39, no. 1, pp. 81–90, 1963.
[67] Y. Li and L. Liu, "Assessing the impact of retail location on store performance: A comparison of Wal-Mart and Kmart stores in Cincinnati," *Appl. Geogr.*, vol. 32, no. 2, pp. 591–600, 2012.
[68] A. Sevtsuk, "Location and agglomeration: The distribution of retail and food businesses in dense urban environments," *J. Planning Educ. Res.*, vol. 34, no. 4, pp. 374–393, 2014.
[69] N. Roig-Tierno, A. Baviera-Puig, J. Buitrago-Vera, and F. Mas-Verdu, "The retail site location decision process using GIS and the analytical hierarchy process," *Appl. Geogr.*, vol. 40, pp. 191–198, 2013.
[70] D. Karamshuk, A. Noulas, S. Scellato, V. Nicosia, and C. Mascolo, "Geo-spotting: Mining online location-based services for optimal retail store placement," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 793–801.
[71] F. Wang, L. Chen, and W. Pan, "Where to place your next restaurant?: Optimal restaurant placement via leveraging user-generated reviews," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, 2016, pp. 2371–2376.
[72] J. Lin, R. Oentaryo, E. P. Lim, C. Vu, A. Vu, and A. Kwee, "Where is the goldmine?: Finding promising business locations through Facebook data analytics," in *Proc. 27th ACM Conf. Hypertext Social Media*, 2016, pp. 93–102.
[73] Z. Yu, D. Zhang, and D. Yang, "Where is the largest market: Ranking areas by popularity from location based social networks," in *Proc. IEEE 10th Int. Conf. Autonomic Trusted Comput. (UIC/ATC)*, 2013, pp. 157–162.
[74] B. Eravci, N. Bulut, C. Etemoglu, and H. Ferhatosmanoğlu, "Location recommendations for new businesses using check-in data," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2016, pp. 1110–1117.
[75] M. Xu, T. Wang, Z. Wu, J. Zhou, J. Li, and H. Wu. (2016). "Store location selection via mining search query logs of Baidu maps." [Online]. Available: https://arxiv.org/abs/1606.03662
[76] H.-P. Hsieh, C.-T. Li, and S.-D. Lin, "Estimating potential customers anywhere and anytime based on location-based social networks," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, 2015, pp. 576–592.
[77] J. Wang, Y. Lin, J. Wu, Z. Wang, and Z. Xiong, "Coupling implicit and explicit knowledge for customer volume prediction," in *Proc. AAAI*, 2017, pp. 1569–1575.
[78] M. C. Popa, L. J. Rothkrantz, Z. Yang, P. Wiggers, R. Braspenning, and C. Shan, "Analysis of shopping behavior based on surveillance system," in *Proc. SMC*, 2010, pp. 2512–2519.
[79] C. W. You, C. C. Wei, Y. L. Chen, H. H. Chu, and M. S. Chen, "Using mobile phones to monitor shopping time at physical stores," *IEEE Pervasive Comput.*, vol. 10, no. 2, pp. 37–43, Apr. 2011.
[80] S. Lee, C. Min, C. Yoo, and J. Song, "Understanding customer malling behavior in an urban shopping mall using smartphones," in *Proc. ACM Conf. Pervasive Ubiquitous Comput. Adjunct Publication*, 2013, pp. 901–910.
[81] J. Banerjee, G. Raravi, M. Gupta, S. K. Ernala, S. Kunde, and K. Dasgupta, "CAPReS: Context aware persona based recommendation for shoppers," in *Proc. AAAI*, 2016, pp. 680–686.
[82] S. Rallapalli, A. Ganesan, K. Chintalapudi, V. N. Padmanabhan, and L. Qiu, "Enabling physical analytics in retail stores using smart glasses," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 115–126.
[83] M. Radhakrishnan, S. Eswaran, A. Misra, D. Chander, and K. Dasgupta, "IRIS: Tapping wearable sensing to capture in-store retail insights on shoppers," Mar. 2016, pp. 1–8.
[84] L. Shangguan, Z. Zhou, X. Zheng, L. Yang, Y. Liu, and J. Han, "ShopMiner: Mining customer shopping behavior in physical clothing stores with COTS RFID devices," *Proc. 13th ACM Conf. Embedded Netw. Sensor Syst.*, 2015, pp. 113–125.
[85] J. Han *et al.*, "Cbid: A customer behavior identification system using passive tags," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2885–2898, Oct. 2016.
[86] T. Liu, L. Yang, X.-Y. Li, H. Huang, and Y. Liu, "TagBooth: Deep shopping data acquisition powered by RFID tags," in *Proc. IEEE Conf. Comput. Commun.*, Apr./May 2015, pp. 1670–1678.
[87] S. Pradhan, E. Chai, K. Sundaresan, S. Rangarajan, and L. Qiu, "Konark: A RFID based system for enhancing in-store shopping experience," in *Proc. 4th Int. Workshop Phys. Anal.*, Jun. 2017, pp. 19–24.
[88] Y.-C. Wang and C.-C. Yang, "3S-cart: A lightweight, interactive sensor-based cart for smart shopping in supermarkets," *IEEE Sensors J.*, vol. 16, no. 17, pp. 6774–6781, Sep. 2016.
[89] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Analyzing shopper's behavior through WiFi signals," in *Proc. 2nd Workshop Phys. Anal.*, 2015, pp. 13–18.
[90] N. Pang, D. Zhu, K. Xue, W. Rong, Y. Liu, and C. Ou, "Analyzing customer's product preference using wireless signals," in *Proc. Int. Conf. Knowl. Sci., Eng. Manage.* Cham, Switzerland: Springer, 2017, pp. 139–148.
[91] M. Radhakrishnan, S. Sen, V. S, A. Misra, and R. Balan, "IoT+Small data: transforming in-store shopping analytics & services," in *Proc. 8th Int. Conf. Commun. Syst. Netw.*, Jan. 2016, pp. 1–6.
[92] A. Sharma and Y. Lee, "AudioSense: Sound-based shopper behavior analysis system," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput., Int. Symp. Wearable Comput.*, 2017, pp. 488–493.
[93] Y. Qu and J. Zhang, "Trade area analysis using user generated mobile location data," in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 1053–1064.
[94] S. Yang *et al.*, "Predicting commercial activeness over urban big data," in *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, p. 119, 2017.
[95] S. Ravulaparthy and K. Goulias, "Characterizing the composition of economic activities in central locations: Graph-theoretic approach to urban network analysis," *Transp. Res. Rec., J. Transp. Res. Board*, 2014, pp. 95–104.
[96] A. Anagnostopoulos, F. Petroni, and M. Sorella, "Targeted interest-driven advertising in cities using Twitter," *Data Mining Knowl. Discovery*, vol. 32, no. 3, pp. 737–763, 2018.
[97] E. Çelikten, G. Le Falher, and M. Mathioudakis, "Modeling urban behavior by mining geotagged social data," *IEEE Trans. Big Data*, vol. 3, no. 2, pp. 220–233, 2017.
[98] J. Eisenstein, A. Ahmed, and E. P. Xing, "Sparse additive generative models of text," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 1041–1048.
[99] M. Sarwat, A. Eldawy, M. F. Mokbel, and J. Riedl, "PLUTUS: Leveraging location-based social networks to recommend potential customers to venues," in *Proc. IEEE 14th Int. Conf. Mobile Data Manage.*, Jun. 2013, pp. 26–35.
[100] T. Agryzkov, J. L. Oliver, L. Tortosa, and J. F. Vicent, "Analyzing the commercial activities of a street network by ranking their nodes: A case study in Murcia, Spain," *Int. J. Geograph. Inf. Sci.*, vol. 28, no. 3, pp. 479–495, 2014.
[101] P. I. Georgiev, A. Noulas, and C. Mascolo, "Where businesses thrive: Predicting the impact of the olympic games on local retailers through location-based services data," in *Proc. Int. AAAI Conf. Web Social Media*, 2014, pp. 151–160.
[102] X. Lu, Z. Yu, C. Liu, Y. Liu, H. Xiong, and B. Guo, "Forecasting the rise and fall of volatile point-of-interests," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 1307–1312.
[103] Z. Zhou, L. Shangguan, X. Zheng, L. Yang, and Y. Liu, "Design and implementation of an RFID-based customer shopping behavior mining system," *IEEE/ACM Trans. Netw.*, vol. 25, no. 4, pp. 2405–2418, Apr. 2017.
[104] K. Massoudi, M. Tsagkias, M. De Rijke, and W. Weerkamp, "Incorporating query expansion and quality indicators in searching microblog posts," in *Proc. Eur. Conf. Inf.* Berlin, Germany: Springer, 2011, pp. 362–367.
[105] Y. Gao, Y. Zhen, H. Li, and T.-S. Chua, "Filtering of brand-related microblogs using social-smooth multiview embedding," *IEEE Trans. Multimedia*, vol. 18, no. 10, pp. 2115–2126, Oct. 2016.
[106] S. Qi, F. Wang, X. Wang, J. Wei, and H. Zhao, "Live multimedia brand-related data identification in microblog," *Neurocomputing*, vol. 158, pp. 225–233, Jun. 2015.

[107] C. Chen, F. Li, B. C. Ooi, and S. Wu, "Ti: An efficient indexing mechanism for real-time search on tweets," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2011, pp. 649–660.

[108] C. Forman, A. Ghose, and B. Wiesenfeld, "Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets," *Inf. Syst. Res.*, vol. 19, no. 3, pp. 291–313, 2008.

[109] L. Zhu, G. Yin, and W. He, "Is this opinion leader's review useful? Peripheral cues for online review helpfulness," *J. Electron. Commerce Res.*, vol. 15, no. 4, p. 267, 2014.

[110] Y.-C. Ku, C.-P. Wei, and H.-W. Hsiao, "To whom should I listen? Finding reputable reviewers in opinion-sharing communities," *Decis. Support Syst.*, vol. 53, no. 3, pp. 534–542, 2012.

[111] Y. Liu, X. Huang, A. An, and X. Yu, "Modeling and predicting the helpfulness of online reviews," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Dec. 2008, pp. 443–452.

[112] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proc. Int. Conf. Web Search Data Mining*, 2008, pp. 219–230.

[113] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 823–831.

[114] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage.*, 2010, pp. 939–948.

[115] A. Mukherjee, B. Liu, J. Wang, N. Glance, and N. Jindal, "Detecting group review spam," in *Proc. 20th Int. Conf. Companion World Wide Web*, 2011, pp. 93–94.

[116] G. Wang, S. Xie, B. Liu, and P. S. Yu, "Identify online store review spammers via social review graph," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 4, p. 61, 2012.

[117] A. Mukherjee *et al.*, "Spotting opinion spammers using behavioral footprints," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 632–640.

[118] S. Kc and A. Mukherjee, "On the temporal dynamics of opinion spamming: Case studies on yelp," in *Proc. 25th Int. Conf. World Wide Web*, 2016, pp. 369–379.

[119] H. Li *et al.*, "Bimodal distribution and co-bursting in review spam detection," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 1063–1072.

[120] E. Kamar, A. Kapoor, and E. Horvitz, "Identifying and accounting for task-dependent bias in crowdsourcing," in *Proc. 3rd AAAI Conf. Hum. Comput. Crowdsourcing*, 2015, pp. 92–101.

[121] Z. Zhang and B. Varadarajan, "Utility scoring of product reviews," in *Proc. 15th ACM Int. Conf. Inf. Knowl. Manage.*, 2006, pp. 51–57.

[122] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han, "Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2014, pp. 1187–1198.

[123] J. Gao, Q. Li, B. Zhao, W. Fan, and J. Han, "Truth discovery and crowdsourcing aggregation: A unified perspective," in *Proc. VLDB Endowment*, vol. 8, no. 12, pp. 2048–2049, 2015.

[124] Y. Li *et al.*, "Reliable medical diagnosis from crowdsourcing: Discover trustworthy answers from non-experts," in *Proc. 10th ACM Int. Conf. Web Search Data Mining*, 2017, pp. 253–261.

[125] N. Q. V. Hung, H. H. Viet, N. T. Tam, M. Weidlich, H. Yin, and X. Zhou, "Computing crowd consensus with partial agreement," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 1, pp. 1–14, Jan. 2018.

[126] V. Sheng, J. Zhang, B. Gu, and X. Wu, "Majority voting and pairing with multiple noisy labeling," *IEEE Trans. Knowl. Data Eng.*, to be published. doi: 10.1109/TKDE.2017.2659740.

[127] Z.-J. Zha, J. Yu, J. Tang, M. Wang, and T.-S. Chua, "Product aspect ranking and its applications," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1211–1224, May 2014.

[128] A. Mukherjee and B. Liu, "Aspect extraction through semi-supervised modeling," in *Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Long Papers*, 2012, pp. 339–348.

[129] E. Tutubalina, "Mining complaints to improve a product: A study about problem phrase extraction from user reviews," in *Proc. 9th ACM Int. Conf. Web Search Data Mining*, 2016, p. 699.

[130] J. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. Mach. Learn.* San Francisco, CA, USA: Morgan Kaufmann, 2001, pp. 282–289.

[131] P. E. Green and V. R. Rao, "Conjoint measurement-for quantifying judgmental data," *J. Marketing Res.*, vol. 3, no. 3, pp. 355–363, 1971.

[132] W. Enders, *Applied Econometric Time Series*. Hoboken, NJ, USA: Wiley, 2008.

[133] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ, USA: Wiley, 2015.

[134] O. Renaud, J.-L. Starck, and F. Murtagh, "Wavelet-based combined signal filtering and prediction," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 6, pp. 1241–1251, Dec. 2005.

[135] R. Chandra and M. Zhang, "Cooperative coevolution of Elman recurrent neural networks for chaotic time series prediction," *Neurocomputing*, vol. 86, pp. 116–123, Jun. 2012.

[136] Y. Matsubara, Y. Sakurai, C. Faloutsos, T. Iwata, and M. Yoshikawa, "Fast mining and forecasting of complex time-stamped events," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 271–279.

[137] Y. Wang, N. J. Yuan, Y. Sun, C. Qin, and X. Xie, "App download forecasting: An evolutionary hierarchical competition approach," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 2978–2984.

[138] L. Einav and J. Levin, "Economics in the age of big data," *Science*, vol. 346, no. 6210, 2014, Art. no. 1243089.

[139] L. Guo, R. Sharma, L. Yin, R. Lu, and K. Rong, "Automated competitor analysis using big data analytics: Evidence from the fitness mobile app business," *Bus. Process Manage. J.*, vol. 23, no. 3, pp. 735–762, 2017.

[140] A. Ghose and S. P. Han, "Estimating demand for mobile applications in the new economy," *Manage. Sci.*, vol. 60, no. 6, pp. 1470–1488, 2014.

[141] X. Liu, H. H. Song, M. Baldi, and P. N. Tan, "Macro-scale mobile app market analysis using customized hierarchical categorization," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2016, pp. 1–9.

[142] H. Zhu, C. Liu, Y. Ge, H. Xiong, and E. Chen, "Popularity modeling for mobile apps: A sequential approach," *IEEE Trans. Cybern.*, vol. 45, no. 7, pp. 1303–1314, Jul. 2015.

[143] B. Fu, J. Lin, L. Li, C. Faloutsos, J. Hong, and N. Sadeh, "Why people hate your app: Making sense of user feedback in a mobile app store," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 1276–1284.

[144] A. Di Sorbo *et al.*, "What would users change in my app? Summarizing app reviews for recommending software changes," in *Proc. 24th SIGSOFT Int. Symp. Found. Softw. Eng.*, 2016, pp. 499–510.

[145] H. Li *et al.*, "Voting with their feet: Inferring user preferences from app management activities," in *Proc. 25th Int. Conf. World Wide Web*, 2016, pp. 1351–1362.

[146] M. E. Porter and J. E. Heppelmann, "How smart, connected products are transforming competition," *Harvard Bus. Rev.*, vol. 92, no. 11, pp. 64–88, 2014.

[147] H. Dawid *et al.*, "Management science in the era of smart consumer products: challenges and research perspectives," *Central Eur. J. Oper. Res.*, vol. 25, no. 1, pp. 203–230, 2017.

[148] E. Brynjolfsson, Y. J. Hu, and M. S. Rahman, *Competing in the Age of Omnichannel Retailing*. Cambridge, MA, USA: MIT Press, 2013.

[149] P. C. Verhoef, P. K. Kannan, and J. J. Inman, "From multi-channel retailing to omni-channel retailing: Introduction to the special issue on multi-channel retailing," *J. Retailing*, vol. 91, no. 2, pp. 174–181, 2015.

[150] L. Cao and L. Li, "The impact of cross-channel integration on retailers' sales growth," *J. Retailing*, vol. 91, no. 2, pp. 198–216, 2015.

[151] C. Narasimhan *et al.*, "Sharing economy: Review of current research and future directions," *Customer Needs Solutions*, vol. 5, nos. 1–2, pp. 93–106, 2018.

[152] J. Hamari, M. Sjöklint, and A. Ukkonen, "The sharing economy: Why people participate in collaborative consumption," *J. Assoc. Inf. Sci. Technol.*, vol. 67, no. 9, pp. 2047–2059, 2016.

[153] G. Zervas, D. Proserpio, and J. W. Byers, "The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry," *J. Marketing Res.*, vol. 54, no. 5, pp. 687–705, 2017.

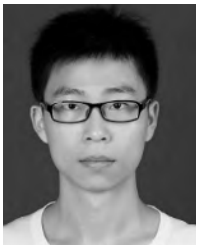
[154] G. Zervas, D. Proserpio, and J. W. Byers, "The impact of the sharing economy on the hotel industry: Evidence from Airbnb's entry into the Texas market," in *Proc. 16th ACM Conf. Econ. Comput.*, 2015, p. 637.

[155] G. Quattrone, D. Proserpio, D. Quercia, L. Capra, and M. Musolesi, "Who benefits from the sharing economy of Airbnb?" in *Proc. 25th Int. Conf. World Wide Web*. Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2016, pp. 1385–1394,.

[156] F. Kooti, M. Grbovic, L. M. Aiello, N. Djuric, V. Radosavljevic, and K. Lerman, "Analyzing Uber's ride-sharing economy," in *Proc. 26th Int. Conf. World Wide Web Companion*. Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2017, pp. 574–582.
[157] L. Chen, A. Mislove, and C. Wilson, "Peeking beneath the hood of Uber," in *Proc. Internet Meas. Conf.*, 2015, pp. 495–508.
[158] M. K. Chen and M. Sheldon, "Dynamic pricing in a labor market: Surge pricing and flexible work on the Uber platform," in *Proc. EC*, 2016, p. 455.
[159] S. Guo, Y. Liu, K. Xu, and D. M. Chiu, "Understanding ride-on-demand service: Demand and dynamic pricing," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom)*, Mar. 2017, pp. 509–514.
[160] B. Leng, H. Du, J. Wang, Z. Xiong, and L. Li, "Analysis of taxi drivers' behaviors within a battle between two taxi apps," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 296–300, Jan. 2015.

BIN GUO received the Ph.D. degree in computer science from Keio University, Japan, in 2009. He was a Postdoctoral Researcher with the Institut Télécom SudParis, France. He is currently a Professor with Northwestern Polytechnical University, China. His research interests include ubiquitous computing, mobile crowd sensing, and human–computer interaction.

YAN LIU received the B.E. degree from Northwestern Polytechnical University, Xi'an, China, where she is currently pursuing the Ph.D. degree with the School of Computer Science. Her research interests include mobile crowd sensing and spatial crowdsouring.



YI OUYANG received the B.E. degree in computer science and technology from the Xi'an University of Technology, Xi'an, China, in 2015. He is currently pursuing the Ph.D. degree in computer science with Northwestern Polytechnical University, Xi'an. His research interests include data mining and social media mining.

VINCENT W. ZHENG received the B.S. degree in computer science from the University of Science and Technology of China, Hefei, China, and the Ph.D. degree in computer science from The Hong Kong University of Science and Technology, Hong Kong, in 2011. He is currently a Research Scientist with the Advanced Digital Sciences Center, Singapore. His current research interests include Web information extraction and mobile computing.

DAQING ZHANG received the Ph.D. degree from the University of Rome "La Sapienza" and the University of L'Aquila, Rome, Italy, in 1996. He is currently a Full Professor with the Institut Mines-Télécom, Télécom SudParis, Evry, France. His research interests include context-aware computing, urban computing, mobile computing, big data analytics, and pervasive elderly care.

ZHIWEN YU received the Ph.D. degree from Northwestern Polytechnical University, Xi'an, China, where he is currently a Professor and the Vice-Dean of the School of Computer Science, Northwestern Polytechnical University. His research interests cover ubiquitous computing and human–computer interaction.