

Received January 3, 2020, accepted February 5, 2020, date of publication February 19, 2020, date of current version February 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975004

A Survey of Online Data-Driven Proactive 5G Network Optimisation Using Machine Learning

BO MA¹, WEISI GUO^{2,3,4}, (Senior Member, IEEE), AND JIE ZHANG¹

¹Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield S1 3JD, U.K.

²School of Engineering, The University of Warwick, Coventry CV4 7AL, U.K.

³The Alan Turing Institute, London NW1 2DB, U.K.

⁴Human Machine Intelligence Centre for Autonomous and Cyber-Physical Systems, Cranfield University, Bedford MK43 0AL, U.K.

Corresponding author: Jie Zhang (jie.zhang@sheffield.ac.uk)

This work was supported in part by the European Commission Horizon 2020 Project Data Aware Wireless Network for Internet-of-Everything under Grant 778305, and in part by the Department of Science and Technology of Guangdong Province under Agreement 2018A050501008.

ABSTRACT In the fifth-generation (5G) mobile networks, proactive network optimisation plays an important role in meeting the exponential traffic growth, more stringent service requirements, and to reduce capital and operational expenditure. Proactive network optimisation is widely acknowledged as one of the most promising ways to transform the 5G network based on big data analysis and cloud-fog-edge computing, but there are many challenges. Proactive algorithms will require accurate forecasting of highly contextualised traffic demand and quantifying the uncertainty to drive decision making with performance guarantees. Context in Cyber-Physical-Social Systems (CPSS) is often challenging to uncover, unfolds over time, and even more difficult to quantify and integrate into decision making. The first part of the review focuses on mining and inferring CPSS context from heterogeneous data sources, such as online user-generated-content. It will examine the state-of-the-art methods currently employed to infer location, social behaviour, and traffic demand through a cloud-edge computing framework; combining them to form the input to proactive algorithms. The second part of the review focuses on exploiting and integrating the demand knowledge for a range of proactive optimisation techniques, including the key aspects of load balancing, mobile edge caching, and interference management. In both parts, appropriate state-of-the-art machine learning techniques (including probabilistic uncertainty cascades in proactive optimisation), complexity-performance trade-offs, and demonstrative examples are presented to inspire readers. This survey couples the potential of online big data analytics, cloud-edge computing, statistical machine learning, and proactive network optimisation in a common cross-layer wireless framework. The wider impact of this survey includes better cross-fertilising the academic fields of data analytics, mobile edge computing, AI, CPSS, and wireless communications, as well as informing the industry of the promising potentials in this area.

INDEX TERMS Online data, data analytics, proactive network optimisation, 5G.

I. INTRODUCTION

The 5G mobile network is the foundation of the future Cyber-Physical-Social Systems (CPSS) by supporting three highly heterogeneous services, enhanced mobile broadband (eMBB), ultra-reliable and low latency communications (uRLLC), and massive machine type communications (mMTC). 5G and beyond 5G services need to support an 600x to 2500x capacity increase [1], sub 1ms round-trip latency [1], and 10,000 or more low-rate devices per cell

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

site [2]. Such significant improvements translate to a sharp rise in the operational expenditure (OPEX) and optimisation complexity ($\approx 60\times$ increase [3]), which is not desirable. As a result, the expenditure of some leading mobile operators may exceed revenues if no effective action taken by the end of this decade [4]. Accordingly, there are widespread concerns that the ambitious quality of the 5G heterogeneous network will be considerably disadvantaged by the OPEX hike.

To alleviate the threat of complexity and OPEX explosion, proactive optimisation has the potential to transform the trade-off between performance and revenue in a fundamental way. One example in 3GPP of successful proactive

optimisation is the video context-aware scheduling, based on the user-side attention information, which is a promising implementation in CPSS [5] associated with the cloud-edge computing techniques. Data-driven proactive optimisation aims to closely couple network resource allocation dynamics in cyber and physical spaces with predicted consumer dynamics in social spaces, allowing it to respond before a negative situation arises (e.g. signal outage or network congestion) and leads to the negative consumer experience. For example, the 3GPP Release 15, System Architecture for the 5G System, includes building user device mobility pattern for Access and Mobility Management Function to benefit the priority in radio resource management [6]. The transformation from passive/reactive to proactive optimisation requires several new network elements that are previously neither seen in 1G/2G/3G/4G nor current 5G standards.

These are:

- Data mining and analytics to infer consumer demand, context, and experience. They require the fusion of qualitative and quantitative analytics in cyber, physical and social spaces;
- Coupling between forecasting and network optimisation functions. This ideally would use the posterior probability of forecasts to inform risks in optimisation;

Current research efforts encounter the bottleneck of both collecting the data for retrieving consumer demand and behaviour context, as well as forecasting at a sufficiently high resolution across slices with uncertainty quantification. Whilst an increasing amount of online information (from the Internet) is generated and accessed, the accuracy is sparse, and reliability is subject to varying platform bias. So while online-data analytics is promising to break the bottleneck for proactive optimisation, much evidence-based work is needed to show the envelope of its applicability. This survey represents an attempt to couple the potential of big data analytics of online information and proactive network optimisation with machine learning, and show its potentially useful areas using demonstration examples.

A. FROM REACTIVE TO PROACTIVE NETWORK OPTIMISATION

Radio resource management (RRM) and network deployment are the primary focus areas of the survey, and there are many underlying optimisation functions include scheduling, mobile edge caching, backhaul optimisation, interference management, load balancing, as well as many aspects of coverage and capacity optimisation. From a historical perspective, RRM and network optimisation have moved from engineering expertise based (e.g. human-expertise driven manual configuration in the late 1990s) to reactive numerical optimisation (e.g. expertise-driven numerical functions with parameter inference in post-2010). With increased complexity and the need for real-time analytics that is personal to consumers, it now needs to evolve into big-data-driven proactive self-optimisation. We will first briefly review the historic

development of optimisation before diving into the enabling technologies for proactive optimisation.

1) REACTIVE NETWORK OPTIMISATION

In the early days of the 2G network, radio engineers monitored the network statistics and tuned the network to improve key performance indicators (KPI). Engineers used their field knowledge and previous experience (e.g. drive testing) to diagnose the origin of problems [7]. However, it took a long time for engineers to manually detect and diagnose the problem, and the network might need several hours from the occurrence of a problem to network recovery. For 3G optimisation, researchers and operators tried to reduce the human-machine interaction. For example, the 3GPP proposed Minimisation of Drive Testing (MDT) in [10] and designed Markov Decision Tree-based optimisation to maximise traffic offload in Wide-band Code Division Multiple Access (WCDMA) [8]. However, each optimisation algorithm typically still required frequent configuration by engineers and was not personal to individual consumers, but more towards a service area (e.g. city council or a shopping mall) or a service genre (e.g. maximum rate or proportional fair). Besides in the automatic examples, the schedulers still required more than one hour to coverage (76 minutes in [9]).

2) DEVELOPMENTS IN SELF / PROACTIVE OPTIMISATION

In the 4G period, the 3GPP stated the significance of implementing automatic optimisation and introduced Self-Organising Network (SON) in Release 8 [16]. In the past decade, a significant number of SON implementations have been developed to enable cell sites to self-optimize their coverage and capacity [11], energy savings [32], and load balance [12], [15]. The commonly used optimisation methods included reinforcement learning [32], Fuzzy controllers [11], regression tree [15]. One challenge with machine learning approaches is that the integration of data is typically low dimensional (e.g. channel estimation or QoS reporting) and both the contextual information is missing to personalise services as well as the forecasting capability to enable proactive optimisation. As such, typically advanced SON engines reacted over 10 minutes after a severe event [15], [17].

In the 5G, to meet the fast-changing context in dense network deployment, the SON decision process has to converge to a satisfactory solution in a very short time. The optimisation time is influenced by algorithm time complexity, computing ability, the time to trigger the algorithm, as well as the uncertainty of the decision's benefit (regret function). This is important because many optimisation algorithms are time-sensitive that the time-scale highly influences the QoS delivery. For example, in a Power Load Sharing (PLS) research [17], the user dissatisfaction rate would increase nearly 20% if the time to trigger the algorithm was delayed from one minute to one hour. That is because network recovers after the degradation of performance. This outage period increases user dissatisfaction experience [17] and risks increased customer complaint and lower customer loyalty to

TABLE 1. Developments of cellular network optimisation from passive to proactive, data-driven, and self-optimisation.

Gen.	Background	Data	Source	Self / Proactive Optimisation Examples	Automation Methods	Time Scale	Summary
2G	Specialists monitor the network statistics to control KPIs and plan the route for drive test [7]	Network statistics and KPIs	BS and manual measurement	Alarm monitoring (intelligently set warning threshold) [7]	-Engineers use their filed knowledge to diagnose the origin of problem	Limited by the ability and knowledge of engineers from several minutes to hours	Passive optimisation by engineering expertise
3G	Manual troubleshooting and optimisation through analysing the statistics and drive surveys [8]	Antenna tilt [8] Transmitter power UE measurement [9] Interference [9]	UE [8], [9] Macro BS [8] Micro BS [8]	-3GPP Minimisation of Drive Testing (MDT) [10] -Use MDT based SON to maximise offload [8] Automatic control of coverage [9]	-Decision tree [8] -Continuous optimisation iterations [9]	-Converge after 65 iterations and take 76 min to execute [9] -Take 23-162 optimisation steps [8]	Reactive numerical optimisation
4G	Manual network parameters managements associated with high cost and long delay [12]	Load [11]–[14] Down tilt [11], [12] Handover parameters [15] Transmitter power [11], [14] QoS [19], [20]	eNB [11], [15] [12]–[14] UE measurement [15] [15]	-3GPP SON Release 8 [16] -3GPP Mobile Data Applications Impacts (Release 11) [18] -3GPP Context Aware Service Delivery (Release 14) [5] -Coverage and capacity [11] [19] [20] -Load balancing [12], [13], [15], [21] -Interference management [14]	-Fuzzy controllers [11], [13], [15], [17] -Decision tree [12] -Polynomial regression [14] [19] -Reinforcement learning [20] -Q learning [21]	-Converge in around 10 minutes [15] -Take around 500 episodes [11] -Take around 20 iterations [12] -Converge after 8 mins [17]	Self-optimisation
5G	Dense network deployment results in an immense amount of nodes [31] Manual and reactive optimisation face heavy burden [30]	Radio measurements [3], [24] Transmitter power [22] Mobility event [22] Load [23]	Femto cell [22] Small cells [23]	-3GPP service-based architecture and network slicing [6], [25], [26] -Femtocell coverage SON [22] [29] -User Centric CoMP clustering [23] -Interference management [23] -Mobile edge caching and computing [27] [28] -Sleeping cell detection [3]	Continuous optimisation iterations [22] [27] Data analysis [30] Fuzzy controller [29] Polynomial regression [24] K-nearest neighbour [3] Decision tree [23]	Take 5-12 iterations [23] Proactive [27] [28]	Big-data driven self-optimisation towards proactive

their network [33]. As we can see in the ‘Time Scale’ column of Table 1, the optimisation time is developed to be closer to real-time and even proactive to save cost and improve loyalty. Approaching this purpose requires heterogeneous data sources to sense the fast-changing network context. On the one hand, current developments in cloud-fog-edge computing have laid the foundation to enable large-scale big-data analysis on cloud and small-scale streaming-data analysis on edge [34]. On the other hand, proactive optimisation in 5G and beyond 5G offers low-latency and reliable communication services to transfer data in CPSS. The widely-used machine learning algorithms aim to configure model-free optimisation for reducing real-time complexity [3], [24], [35], and a current trend for improving SON decision time horizon is triggering the algorithm in advance (e.g. proactive). Table 1 summarised the network optimisation the general view of developments from 2G to 5G. It also provides the gaps of current optimisation.

3) GAPS IN DATA USAGE

Current research for RRM and network optimisation tend to be using mobile network data (subscriber-level and cell-level) as the only data source [14], [23], [24], [36]. The general user behaviour is needed but abstracted to homogeneous mathematical models or statistical assumptions which have limitations to represent the real-world changing complexity and diversity. The vast majority of these approaches self-simulate a single user-side measurement as a proxy for more complex behaviour aspects [3], [14] [24]. Such a fact indicates difficulties in collecting and integrating multi-dimensional data from different sources, especially if it involves both structured and unstructured metadata. The lack of consumer context causes *cold start* problem that the algorithms require some time to collect information and then start optimisation. Furthermore, in current passive load balancing, inter-cell user distribution

is unknown so that all neighbouring cells become active for offloading. It is inefficient when the over-load cell has a skewed user distribution.

B. ENABLERS FOR PROACTIVE OPTIMISATION

The 3GPP Mobile Data Applications Impacts (Release 11) [18] mentioned that network optimisation will be boosted if it can understand user behaviour and spatial-temporal traffic pattern through application data. In that case, the optimisation needs the status information of concrete entities in the social space (e.g., UEs) or virtual entities in the cyber space (e.g., software), such user-centric meta-information is called *context* in [37], [38]. The context represents all the user information indicating spatial-temporal network traffic characteristics for a user-centric network, including geolocation and user behaviour. The methods to gain the context are named *context-aware* or *context-awareness*. And the computing works rely on cloud-fog-edge computing. The optimisation algorithm requires a context-aware module which automatically collects and analyses data from different sources (e.g., online data and personal devices), then supplies context for adequately re-allocating communication resources [17].

1) ONLINE DATA

Many users on online social networks voluntarily generate personal data, and the data is interleaved with geographical, public, and other information [39]. The online data directly and tightly connects to users’ intents, and therefore appropriate for transforming network optimisation to be proactive and user-centric. As stated in [40], [41], CPSS aims to offer services to be not only high-quality but also proactive and personalised, which results in an irreplaceable role for the online big-data analysis. Online data can be divided into different types: social networks, video/photo

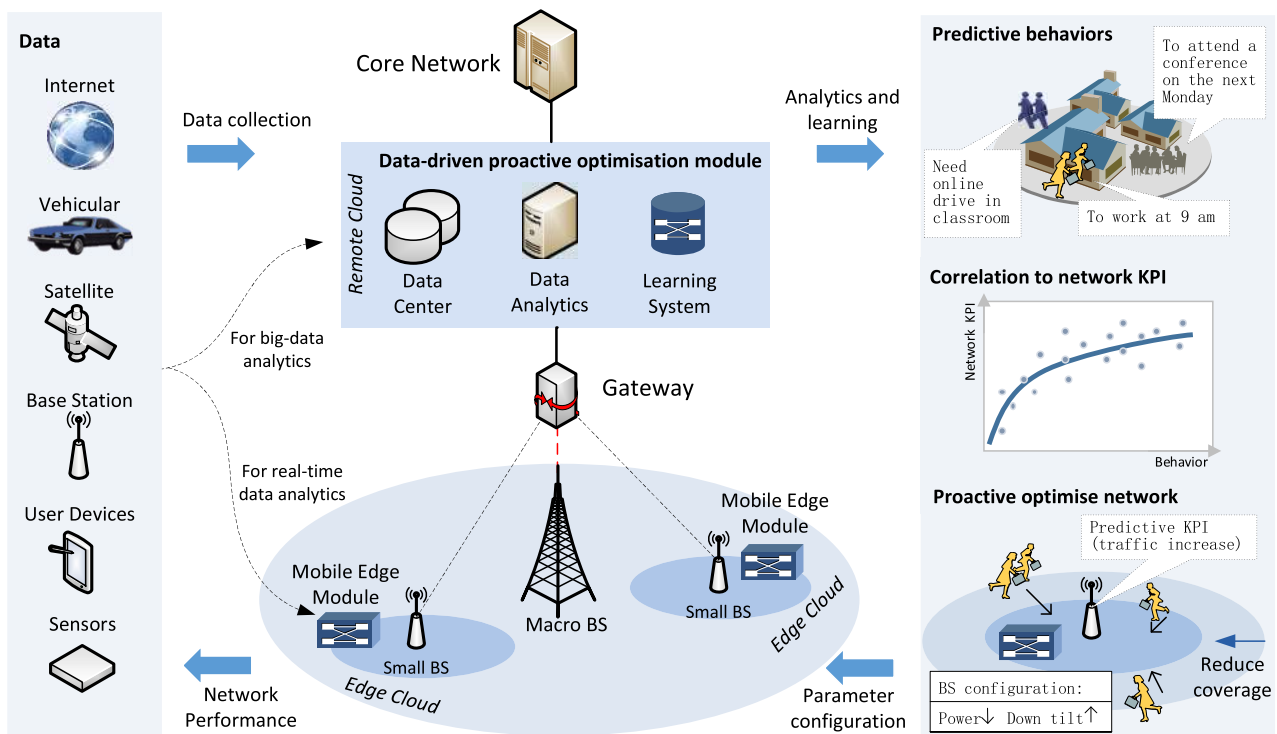


FIGURE 1. The framework of data-driven network for proactive optimisation.

sharing sites, online forums, product reviews/ratings, and wikis. This survey focuses on three of them, social network (e.g., Twitter, Facebook and Instagram), comments and reviews (e.g., Amazon customer review), and multi-media hosting sites (e.g., YouTube). These types of online data inherently contain substantial hidden information about users and hold different data merits and pitfalls.

2) SOCIAL NETWORK DATA

Online social platforms change users from content viewers to content creators and distributors. It not only owns plenty of shared information about public and individuals [17] but also supplies real-time details for forecasting spatial and temporal attributes of future events. The social network data consists of four data formats: geolocation, timeline, text, and photos/videos. Text and photos have plenty but noisy information to grasp users' desire, by contrast, timeline and geolocation have a clear and structural format but limited information. The social network also records the social relationship/tie which benefits the estimation of the weights of D2D links. For example, the forecasting of social ties in [42], [43] enable D2D in caching delivery with finding the most influential user [42] and sharing with friends [43].

C. THE FRAMEWORK OF DATA-DRIVEN PROACTIVE OPTIMISATION

To make the enablers drive proactive optimisation, we propose a framework in Fig. 1 that involves data acquisition,

integration, and using forecasting to drive RRM and network optimisation. The complex computation relies on cloud and edge computing. Example frameworks of combining the cloud-edge computing and the CPSS big data analysis is put forward in [34], [40], [44]. In these works, a cloud plane is responsible for global big data with large-scale and long-term (cloud computing) while local data is processed by an edge plane (edge computing) because of its small scale and short term. Our work follows a similar idea, the remote cloud and the edge cloud work as the cloud and edge planes respectively. In Fig. 1, the remote cloud (data-driven proactive optimisation module) is placed between the Gateway and the core network. An example of this implementation is shown in [45] that placed the caching module close to the small BS and between the core network and edge network. That will benefit the data storage, analysis and learning, but the precondition is that the latency should be acceptably low. The following parts illustrate the functions of the framework.

1) DATA

This step includes data collection, cleaning, and storage. For a more detailed process, the work [34] further divides it into organisation, representation, cleaning, reduction and integration. These aspects have been successfully implemented thanks to the support of the edge computing and well-accepted in CPSSs. The data collection is a process of gathering information on variables of interests through multiple online and offline available sources, such as the

TABLE 2. List of acronyms.

Acronyms	Description
3GPP	The 3rd Generation Partnership Project
5G	Fifth-Generation Mobile Networks
ABS	Almost Blank Subframe
API	Application Programming Interface
ARMA	Auto Regressive Moving Average model
BBU	Baseband Unit
BS	Base Station
CAPEX	Capital Expenditure
CBR	Call Block Ratio
CDF	Cumulative Distribution Function
CNN	Convolutional Neural Network
CoMP	Coordinated Multi-Point
D2D	Device to Device
DBSCAN	Density-based Spatial Clustering of Applications with Noise
eICIC	Enhanced Inter-cell Interference Coordination
EM	Expectation Maximisation
eMBB	enhanced Mobile Broadband
ESN	Echo States Networks
FLC	Fuzzy-Logic Controller
GP	Gaussian Process
GPS	Global Positioning System
GSM	Global System for Mobile Communications
HetNet	Heterogeneous Network
ICIC	Inter-cell Interference Coordination
KPI	Key Performance Indicators
LTE	Long Term Evolution
LTE-U	LTE-Unlicensed
MDT	Minimisation of Drive Testing
MIMO	Multiple-Input Multiple-Output
mMTC	Massive Machine Type Communications
NLP	Natural Language Processing
OPEX	Operational Expenditure
PLS	Power Load Sharing
PTS	Power Traffic Sharing
RAN	Radio Access Network
RNN	Recurrent Neural Networks
RRH	Remote Radio Heads
RRM	Radio Resource Management
RSRP	Reference Signal Received Power
SON	Self-Organising Network
SVM	Support Vector Machine
SVR	Support Vector Regression
QoE	Quality of Experience
UAV	Unmanned Aerial Vehicles
UE	User Equipment
UGC	User-Generated Content
uRLLC	ultra-Reliable and Low Latency Communications
WCDMA	Wideband Code Division Multiple Access

Internet, vehicular, satellite, base station, user devices, government and business databases, and sensors. The combined data may vary in sparsity and resolution across urban and industrial areas [46]. For the Internet data, the providers provide Application Programming Interfaces (API) for the third parties to access the open data. More details of data collection can be found in Section III.A.1. Then, the raw data can be noisy and redundant, so it needs a cleaning process (e.g., sort, filtering) to be established in a systematic fashion and stored in the edge data centre or cloud data centre for

further analytics. Tensor-based method is effective to analyse big data by focusing on typical features [47]. The extensible order tensors can represent unstructured and structured data. Typical applications of this method are in [48], [49]. These works illustrate how the tensors work in cloud-edge computing. Big streaming data is also a challenge for real-time big data processing. High-order singular value decomposition is proved efficient to avoid redundancy (see examples in [50]).

2) PREDICT CONSUMER BEHAVIOURS

The network traffic fluctuates according to consumer behaviour. This step is building models to predict demand changes. The input is the online data from the Internet, and the output is the probability of different demand levels across various behaviour contexts and slices. Understanding the posterior distribution of predictions will generate a spatial-temporal consumer demand distribution that helps predict the network KPI.

3) CORRELATE TO NETWORK KPI

The correlation models a path of mapping predictive behaviours to network KPIs, such as the network traffic. The polynomial regression (in the figure) and statistics analysis (e.g., Pearson correlation) are two commonly used approaches. Therefore, the model input is the behaviour probability, and the output is the probability of KPI, such as the probability of high-load occurrence.

4) PROACTIVE NETWORK OPTIMISATION

The predictive network KPIs are injected into this function. The optimisation algorithm needs to configure the parameters in advance for the upcoming condition changes to achieve targets. An example is proposed in [51] about optimising the network in a proactive and energy-efficient way. They presented a framework with implementing a big-data-aware intelligent platform between the core network and Baseband Unit (BBU) pool for analysing user behaviour and network patterns to output control strategies. Note that, these analytics and leanings are available to be carried out by both remote cloud plane and edge plane. The cloud computing can generate a general trend context of public behaviours, such as traffic of a city in rush hours, which suggests a macroscopic optimisation. At the same time, the edge computing processes personalised context with the help of edge data centre, edge tensors, edge data management and analysis. Finally, the network configuration in the physical space is decided according to both trend and individual context.

D. SURVEY OBJECTIVES

The core concept of this survey is to identify the user-oriented contexts that proactive network optimisation requires, then categorise the available online data and methods to acquire these contexts, finally to characterise the relationship between different proactive network optimisation and each online data. In this case, this survey probes further into these questions:

TABLE 3. A comparison of related reviews in the areas of data, context, and optimisation with this survey.

	Online Data (Web Content)	Operator Data	User Context (Geo & Social)	Network Traffic Prediction	5G Optimisation	SON
[39], [52]	✓	×	×	×	×	×
[53]	×	✓	✓	✓	×	✓
[54]	×	✓	✓	×	×	×
[55]	×	✓	×	×	✓	×
[56]	×	✓	×	×	×	✓
[31]	×	✓	×	×	×	✓
[57]	×	✓	✓	×	✓	✓
This Survey	✓	✓	✓	✓	✓	✓

- Which user and demand contexts can help personalise proactive network and RRM optimisation?
- How to retrieve or infer the CPSS contexts from online data?
- What are the inputs and outputs of the Cloud-Fog-Edge computing?
- What is the supply-demand relationship between online data and proactive network optimisation?
- What are the open research areas in proactive network optimisation for 5G?

The main academic contributions of this survey are as follows:

- To help readers understand the categories and characteristics of the available online data.
- To review and summarise the useful CPSS contexts provided by online data and the spatial-temporal traffic pattern.
- To clarify the role of Cloud-Fog-Edge computing in proactive network optimisation.
- To analyse each network optimisation to be proactive regarding the required contexts.
- To build a supply-demand business relationship between online data and network optimisation.
- To provide applications of proactive optimisation by the cloud and edge computing in CPSS.

The remainder of this survey is organised as follows: Section II discusses related survey papers about online data, contexts, and network optimisation. Then, Section III summarises the contexts, data sources, and analytic methods. In Section IV, the authors analyse each proactive network optimisation and build a link between contexts and optimisation. After that, an online data-proactive network optimisation supply-demand relationship map is put forward. Section V proposes open challenges and future research directions. Finally, Section VI concludes this survey.

II. RELATED WORK

This section discusses relevant surveys in three aspects (data-driven optimisation, consumer context, and online data), summarises and compares them with our survey (in Table 3).

There are several data-driven network optimisation works with focus areas of exploiting big-data [55], [57], constructing SON engine architecture [56], and developing in

machine learning [31], [97]. The authors of [55] exploited big data-driven 5G network optimisation. They first presented a general framework to integrate operator big data, then introduced optimisation cases (e.g., resource allocation, interference coordination, and cache deployment). Furthermore, self-optimisation is emerging in recent years. The survey [56] was one of the pioneering literature about reviewing cellular self-optimisation for a tutorial purpose. Readers were provided with projects, features, standards, taxonomy, solutions and design guidelines in detail to start their research in this area. Then, a hot topic of applying machine learning in SON has been reviewed in [31]. Moreover, the work [97] provided a comprehensive tutorial on applying artificial neural networks (ANN) in wireless communication scenarios (e.g., proactive caching). Several key types of the neural network were presented, such as feed-forward, recurrent, spiking, and deep neural networks. Besides, the authors in [57] proposed a review and two case studies on data-driven small cell RRM and deployment. Although the above works provide detailed reviews in data-driven optimisations, to our best knowledge, there is still no particular work to reveal the paths of taking the advantages of online data to proactive network optimisation. Compared with other works (as summarised in Table 3), ours mainly contributes to using online data to make user behaviour predictive to drive proactive optimisation.

The methods for forecasting consumer behaviours were reviewed [53], [54], including users geolocation, link quality, network traffic, and social information. The work [53] aimed at constructing a predictive frame to obtain intelligence for detecting user behaviour and network environmental changes. The authors presented a detailed context classification of prediction techniques but did not introduce data sources. Isolated examples of network optimisation were briefly discussed, such as mobility prediction for network offloading. The reviews of data sources from the online platform can be found in [39], [52]. Moreover, the researchers in [54] investigated existing big-data mobility works for geolocation prediction. The authors summarised the basic principles and common methods for mining users’ distribution and mobility patterns. Furthermore, GPS, Global System for Mobile Communications (GSM), and WiFi data records could all be used to track popular regions, so the authors compared them in the areas of data analysing and model building.

TABLE 4. Common methods of data collection, cleaning, and analytics to retrieve useful network context from online information.

	Search API	Stream API	Data Set	Data Cleaning	NLP	Sentiment Analysis	Text Analysis	Statistics	Regression	Classification	Clustering	Neural Network
Popular Region	[58], [59]	[58] [62], [63]	[60], [61] [59]	[59], [60]	[59], [60]	[59], [60]	[58], [59] [63]	[59], [62], [63]	-	[62]	[60], [61] [58], [63]	-
Personal Trajectory	-	[64]–[66]	[61], [67]–[69] [66], [70]	[64], [65]	-	-	[64]–[65]	[69] [64]–[66]	-	[64], [65]	[61], [67] [68], [70]	-
Content Popularity	[71]	[72], [73] [71], [74], [76]	[74]	[71], [75]	[75]	[71], [75]	[71]	[72], [73] [71], [74]–[76]	[72], [76] [75]	[71], [73] [75]	[72], [73] [76]	[75]
User Preference	[77], [78] [79]	[79]	[33], [77], [78] [82], [83]	[33], [77]	-	[33]	[33]	[33], [77]–[79] [82], [84]	[80]	[33], [81] [80], [83]	[79]	[81], [82]
User Relationship	-	[85], [86] [91]	[87], [88] [80], [92]	-	-	-	-	[85], [86], [88]–[91] [92]	-	-	[92]	[93]
Network Traffic	-	[94]	[94]–[96]	-	-	-	-	[94]–[96]	-	-	-	-

The above surveys presented the feasibility of applying online data for forecasting consumer behaviours (demand), whereas it is still not clear how to couple these techniques with proactive network optimisation. This survey also contributes to this aspect through integrating the demand knowledge for a range of proactive optimisation techniques.

III. ONLINE DATA FOR BEHAVIOUR PREDICTION

The users' behaviour can be fully understood and even predicted based on meta-information such as: time, history, geolocation, identity, social status, schedule, and events. In this section, we will lead the way to forecast user behaviour and inferring data demand.

A. OVERVIEW OF ONLINE DATA ANALYTICS TECHNIQUES

Table 4 presents an overview of the common data-analytics tools in the following reviewed literature. We divide these methods into two main categories: data collection and data analytics.

1) DATA COLLECTION

The online data consists of three major categories, social network (e.g., Twitter and Facebook), media hosting site (e.g., YouTube and Instagram), comments and reviews (e.g., e-commerce reviews and topic talk). The geo-tagged social network data is the source for extracting the geo-location context. Forecasting social behaviours requires different kinds of online data for modelling user activities and corresponding meta information. The traditional method for collection is using an Application Programming Interface (API).

The API-collection method is widely used (see application examples in Table 4, 'Search API' and 'Stream API' columns). It allows automatic data collection from service providers in an economical way. The search API and the stream API are all in this category. However, it still has some challenges, such as poor efficiency and difficulty in gathering historical data. Also, service providers could limit the collection.

Building own datasets, such as collecting from volunteers or purchasing from service providers, can avoid the above limitations. This method provides a nearly complete dataset

and achieves a flexible setting of the environmental variables. A popular method is sending friend requests to other users with a statement that researchers are collecting the data for researches with privacy protection. The volunteers could feel free to accept or reject friend requests. For example, the researchers in [98] invited 19,484 users agreed to join the experiment. Sometimes, incentives can improve the performance of collection that contributors who would be rewarded with payment according to their contribution [99], [100]. This method can be costly and requires ethical approvals, or one can use open datasets to reduce this cost.

The open datasets can reduce the cost in data-collection. Many organisations make available their datasets for transparency or research purposes, such as Kaggle datasets [101] and European Union Open Data. One typical example is the Italia Telecom operator dataset (see data from [102]). It is used to forecast cellular traffic pattern [103]. Some public projects will also open their data to other researchers. For example, an EC H2020 RISE Project DAWN4IoE has opened the datasets, such as cellular traffic data [104] and Twitter density [95], [105]. Researchers are required to choose their methods for collection and pre-process the raw data for further data analytics.

2) MOST COMMON ONLINE-DATA ANALYTICS TECHNIQUES

Scalable machine learning emerged in recent years as it gives the computer system an ability to learn user behaviour from online data. We compare the most commonly used machine learning methods regarding proactive optimisation requirements in Table 5 and list the sections where these methods have been used. The complexity refers to the number of computation operations that should be performed to achieve the desired result. The training data and time indicate the required data amount and training efficiency. Then, the accuracy suggests the supposed performance that the algorithms generate. Finally, the evaluated levels (low/fair/high) is based on the previous literature cited after each name.

From Table 5, we can find some common characteristics of the machine learning usage. For example, the geo-location modelling usually requires unsupervised clustering methods, such as K-means and Density-based Spatial Clustering of Applications with Noise (DBSCAN). This has the

TABLE 5. A summary of the reviewed machine learning methods with the usage position in sections and a comparison of time complexity.

Machine Learning	Time Complexity	Description	Complexity	Training Data Amount	Training Time	Accuracy	Usage in Proactive Optimisation
K-means [106]–[108]	$O(KMNI)$	K: number of clusters M: dimension N: number observations I: number of iterations	Low	Low	High	Fair	III.B. Modelling Geo-location
DBSCAN [106], [107]	$O(N\log(N))$	N: number of observations	Fair	Low	Fair	High	III.B. Modelling Geo-location
Support Vector Machine [108]	$O(N^2)$	N: number of observations	High	High	Fair	High	III.C. Modelling Social Behaviour IV.C. Proactive Caching
Naive Bayes Classifier [107], [109]	$O(Nd)$	N: number of observations d: dimensionality of the features	Fair	High	Low	Low	III.C. Modelling Social Behaviour
K Nearest Neighbour [107], [109]	$O(Nkd)$	N: number of observations d: dimensionality of the features k: hyperparameter	High	High	Fair	Fair	III.C. Modelling Social Behaviour
Naive Bayes Neural Network [107], [109]	$O(nMPNe)$	n: input variables M: number hidden neurons P: number output values N: number of observations e: number of epochs	High	High	High	High	III.C. Modelling Social Behaviour
Linear Regression [108]	$O(NPI)$	N: number of observations P: number of variables I: number of epochs	Low	Fair	Fair	Fair	III.C. Modelling Social Behaviour IV.A. Network Traffic Prediction IV.C. Proactive Caching
Gaussian Process [110]	$O(N^3)$	N: number of observations	High	Low	Fair	High	III.D. Predictive User Behaviour IV.A. Network Traffic Prediction
Deep Gaussian Process [111]	$O(N^3)$	N: number of observations	High	Low	High	High	IV.A. Network Traffic Prediction
Long Short-Term Memory (LSTM) [112], [113]	$O(W)$ per time step	W: the total number of parameters in a standard LSTM network	High	Fair	-	High	IV.A. Network Traffic Prediction
Polynomial Regression [108]	$O(NPI)$	N: number of observations P: number of variables I: number of epochs	Low	Fair	Fair	Fair	IV.B. Proactive Load Balancing
Q-Learning [114], [115]	$O(S^2A)$	S: number of states A: number of actions	High	High	-	-	IV.B. Proactive Load Balancing
Reinforcement Learning [114], [115]	$O(S^2)$	S: number of states	High	High	-	-	IV.C. Proactive Caching
Neural Network [108], [114]	$O(nMPNe)$	n: input variables M: number hidden neurons P: number output values N: number of observations e: number of epochs	High	High	High	High	IV.C. Proactive Caching
Echo States Networks [116]	$O(2TU/H)$	A recurrent neural network T/H: time of executions U: number of predictions	Fair	High	Fair	-	IV.C. Proactive Caching
Convolutional Neural Network [117]	$O(\sum_{l=1}^d n_l m_l^2)$	d: number of layers n: number of filters m: size of the output feature s: size of the filter	High	High	High	-	IV.C. Proactive Caching
Transfer Learning [118]	$cO(n+m)$	c: time of calculating distance n,m: size of training set	Fair	-	Fair	Fair	IV.C. Proactive Caching
Meta Learning [119]	$O(l+t/\log t)$	l: length of program t: program halt time	High	High	Fair	High	Learn how to learn. Use conclusions of other learning.
Deep Reinforcement Learning [114], [115]	$O(S^2D)$	S: number of states D: deep learning complexity	High	High	High	High	Complex context perception by DL Decision making by RL
Neuro-evolution Deep Learning [120]	$O(n(1+\log n)D)$	n: the problem size D: deep learning complexity	High	High	High	High	Avoid local optimum in using gradient descent

advantage of not requiring labels in a sophisticated problem setting and rely on topological features as a compressed representation of high dimensional attributes. However, the ill-defined nature of clustering means initial parameterisation is highly related to researcher bias or intuition (see details in Section III. B). In contrast, social behaviours usually have limited categories (e.g., positive and negative in user sentiment), so supervised classification is commonly selected, such as Support Vector Machine (SVM) and K

Nearest Neighbour (KNN) (Section III. C). For time series forecasting, regression and Markov methods are good at predicting network traffic (Section IV.A). In network optimisation, more sophisticated methods are chosen, such as reinforcement learning (Section IV. B) and the neural network (Section IV. C). However, many of the non-Bayesian methods face challenges of catastrophic forgetting and dealing with high-dimensional inputs. That requires more-advanced learning methods, such as deep Gaussian process, meta learning,

deep reinforcement learning, and neuro-evolution deep learning. These approaches either provide quantitative uncertainty estimates, high-dimensional feature capture, and/or improved adaptation to the environment.

3) PREDICTION ERRORS AND OPTIMISATION OVERHEAD

The machine learning approaches thus far produce prediction errors because of the scarcity of training data or the mismatch of prediction functions. The probability of error can be described by uncertainty in the predictions. In further optimisation, the overhead of the system will be incurred because of such uncertainty. We draw lessons from other prediction and optimisation systems in other areas of science and engineering.

In prediction systems, such as climate science and structural mechanics, big data helps to inform the likelihood of outcomes of predictions that arise from dynamical systems. Probabilistic numerics translate input uncertainty into output uncertainty (e.g., probabilistic finite element). The uncertainty caused by data/estimation errors is required to monitor and control the computational overhead. In that way, the paper [121] provided an illustration of using the probabilistic numerics to describe the uncertainty with diagnosing error sources in computations. However, the Gaussian Process needs to be coupled with deep learning to face more complex tasks, so the deep Gaussian process emerges.

The deep Gaussian process acts as a deep neural network but with Gaussian Process governing the mappings between layers. It will give an empirical confidence interval to quantify the uncertainty. The higher uncertainty could mean a higher potential to cause overhead. In network optimisation, the forecasting associated with high potential of overhead could be discarded in the decision making. For example, the work [111] successfully learnt natural human motion by the Deep Gaussian Process even with scarce data. Besides the uncertainty (overhead) quantification, a parallel system offers a useful structure to improve the robustness.

The parallel system owns a reliability-wise structure. It allows the system to function with any mechanism working. For example, if the network unexpectedly operates in a bias condition, it still has time to alter to a reactive optimisation. Such a method works as the parallel system as introduced in [122] to improve reliability. Based on the above methods, this survey proposes a framework to create a regret function for the poor performance in the proactive optimisation (see IV.A.6. The Quantification of Uncertainty in Proactive Optimisation).

4) PRIVACY AND DATA UTILITY

The privacy problem is critical in data analysis. The challenge is to gain high utility in data while ensuring confidentiality, integrity, and availability [123]. Besides, the network operators and the data providers should achieve not only encryption of all data but also a strict access control to avoid the unpleasant data collection, storage, and usage. In that case,

the trade-off between data utility and privacy needs to be solved in three aspects.

Firstly, current users are usually unaware of the collection of personal data, which causes the anxiety about potential defraudation and hurt feelings. Appropriate notices and asking authorisation can relieve the anxiety during personal data collection. For example, in the Internet of Things (IoT), the users are notified about IoT privacy properties [124]. This 'right to know' alleviates anxiety and provides users with choice. An example is the current usage of Internet cookies (see cookies consent under the EU General Data Protection Regulation [125]), but it requires that consumers trust data storage and usage.

Secondly, the stored data must be provided with both privacy and authenticity. For this purpose, the encryption schemes will transform the data into a ciphertext with a symmetric-key mechanism to satisfy the two requirements (see Authenticated Encryption [126]).

Moreover, data protection not only needs to encrypt information but also protect them from attacks, which is a classification problem. For example, the classification of legal/illegal user can be achieved by utilising the radio channel information [127]. The work [128] used Recurrent Neural Networks (RNN) to detect various attack variations, and the work [129] provides a panoramic survey of security in cyber-physical systems.

Finally, during the data usage, researchers need to protect sensitive latent information while reserving utility. Such a trade-off is studied in [130] by measuring data utility loss and latent-data privacy matrices. Another method to secure outsourced data analytics is by applying the homomorphic encryption [131].

5) SUMMARY OF FINDINGS AND LESSONS LEARNED

In summary, the main findings and lessons learned from this sub-section include:

- Collecting online data from APIs is the most economical way, but the service providers could limit the process. In contrast, building their own datasets can suit the requirements well, but it will cost time or money to find or reward the contributors. The other way is to use the public datasets which are increasingly available due to transparency and reproducibility drives.
- Geo-location modelling is often an ill-defined unsupervised clustering challenge. In contrast, behaviour modelling is usually a supervised classification problem. Next, the traffic prediction can be addressed by the regression methods, including the polynomial regression, Gaussian Process, and RNN. In the network optimisation, the parameters and principles become dynamic and numerous, so high-complexity methods (e.g., reinforcement learning and the neural networks) are becoming increasingly suitable.
- The prediction errors cause undesirable overhead in the proactive optimisation modules. It is necessary to

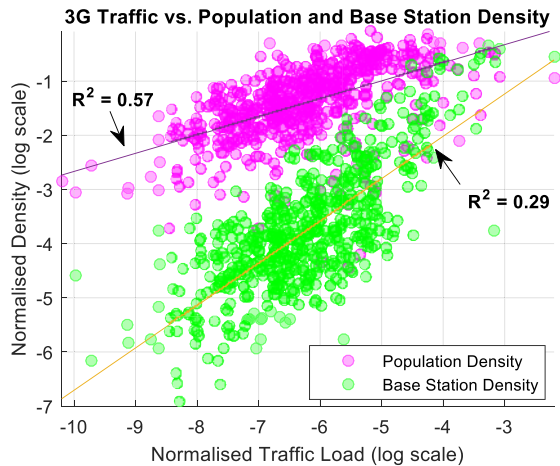


FIGURE 2. Population and BS density vs. 3G network traffic adapted from our work [95].

quantify such overhead and also take into account other utility functions such as privacy and security.

B. MODELLING GEOLOCATION

Geolocation represents the real-world measured location of users, which offers spatial traffic distribution. It contains three components, observation time, moving objects, and geolocation records [54], which can be mined from data of GPS, Base Station (BS), and landmarks. In that case, the online data is suitable for providing the context of the popular region [94] or forecasting the nodes of the personal trajectory [132].

1) POPULAR REGION

A popular region is a specific place with the potential to generate high communication traffic where a group of location records gather around a centre at a particular time. This region can be attractive all the time, such as commercial and tourists areas. In Fig. 2 from [95], we present the spatial correlations between 3G traffic and population density. This verifies the hypothesis that popular regions (with high population) have high probabilities of generating high demand. The network optimisation schemes should allocate resources in these areas to satisfy the imbalance traffic distribution, especially during events.

In the network optimisation, the prediction of the popular region represents the upcoming hotspots whose popularity is correlated with social network traffic (e.g., Twitter). It will benefit resource deployment [95], load balancing [94], and caching [133] by finding the place with high demand. The time requirement for hotspot prediction generally needs to be two hours in advance [94], [95] due to the achievable high accuracy (correlation > 0.85 in [95]). In contrast, the geographic resolution requirements for hotspot prediction depend on the requirements of different network optimisations. For example, in [133], the predicted resolution of hotspot decided the flying height (332 m) of the flying BSs for proactive caching. Moreover, the work [94] achieved

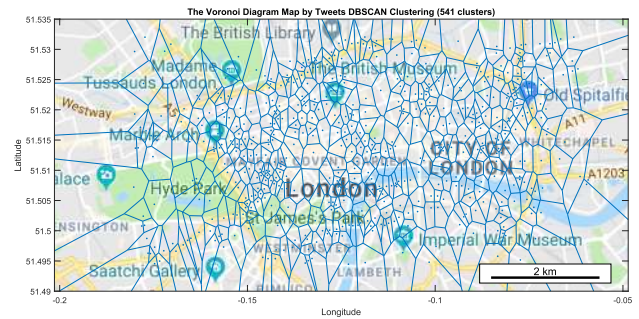


FIGURE 3. Popular region clustered by DBSCAN based on the geo-locations in Twitter of London adapted from our work [134]. The cells are visualised using Voronoi diagram. Smaller cells represent more-popular regions.

a load-resource matching with a 120-meter resolution. The popular regions (hotspots) are usually modelled by clustering, such as k-means in [133]. It is to maximise similarity in the same group and guarantee that the assigned objects in different clusters are as different as possible.

a: K-MEANS BASED SPATIAL MODEL

In using this model, researchers manually choose the number of clusters (k), then the algorithm groups GPS coordinates according to k centroids cooperated with map information. For example, the users can be clustered into different groups to guide the location and height of flying BS to cover them (see the research in [133]). However, the k has to be manually determined, and the cluster range is out of control. To determine the number of clusters and a cluster radius, Ashbrook and Starner [61] tried to use a variant of K-means which simulated radius regarding cluster numbers and picked the k at the convergence starting point.

In fact, popular regions' ranges can vary a lot in both size and shape, which requires automatic range optimisation and methods to reduce computation cost. Besides, this model cannot avoid the influence of noise data. In that case, the method named DBSCAN emerged.

b: DBSCAN BASED SPATIAL KERNEL

The DBSCAN is a density-based algorithm that groups the points with many nearby neighbours and ignores the points lying along in low-density area as outliers (noise). This model requires no prior knowledge of clusters and no radius and results in fitting cluster shapes. Researchers choose only a minimum range and the minimum number of points in this range. Then a cluster with a minimum density is generated with arbitrary shape. For instance, the work [135] used this method to search popular regions considering the diversity of users and adaptive density. We applied the DBSCAN-based method on a Twitter dataset in London. The result is proposed in Fig. 3. The density parameter is set as the average density, so the popular regions (high-density cluster) own higher density and smaller size. For convenience in visualisation, a Voronoi diagram is used, so only the centres

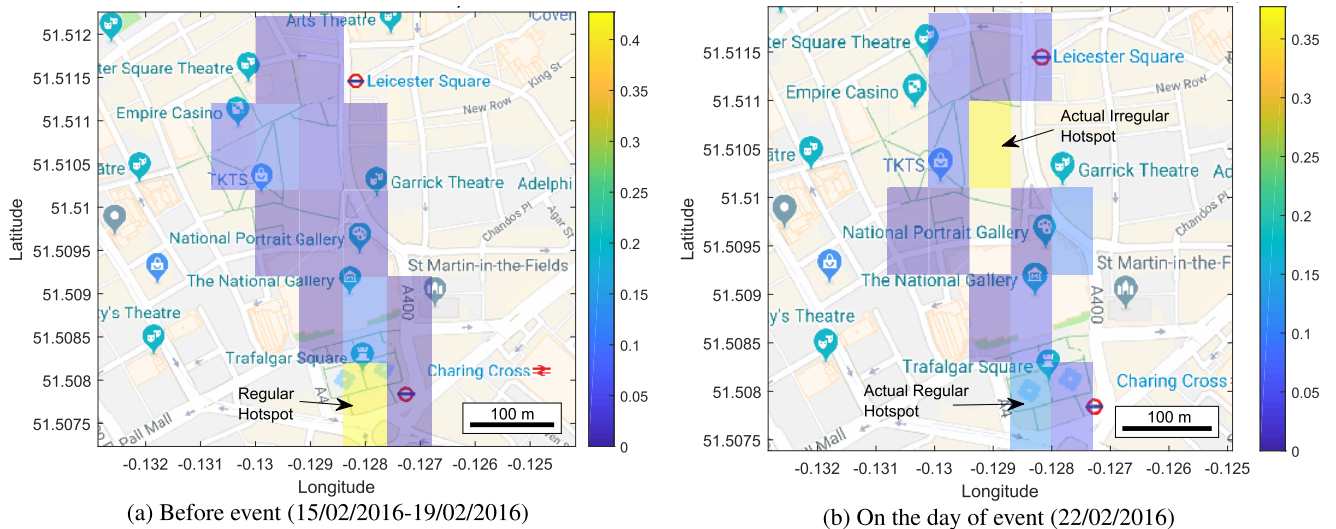


FIGURE 4. An example of Twitter traffic heatmap before and during event at a cinema, Leicester Square, London (adapted from our work [134]). Regular hotspots still exist, but the hottest spot changed to the event location. (a) Before event (15/02/2016-19/02/2016). (b) On the day of event (22/02/2016).

and borders are displayed. Popular regions, like the city of London, own denser small clusters indicating the high traffic demand. Even the spatial clustering algorithms are widely applied to find popular areas, only clustering in one dimension was not enough, such as recognising sub-areas for the evolution of events. Therefore, researchers considered using a spatial-temporal model.

c: SPATIAL-TEMPORAL CLUSTERING MODEL

Geolocation clustering has two main sub-categories, spatial clustering and temporal clustering, which make the objects gather regarding both dimensions (location and time). We need to consider the temporal dimension to find changing popular areas along with time. For example, the temporal dimension can be added to an extension of DBSCAN to take time changes into account to separate regions in both space and time. K. Tamura and T. Ichimura proposed an example work in [58] by analysing Twitter data.

d: EVENT-DETECTION BASED MODEL

A place with an attractive event becomes popular in a particular period. Detecting events means to retrieve necessary information of a planned public occasion, such as schedule, topics, and attendance. Thanks to the online information, the occurrence of events can be automatically detected [59], [62].

Statistic method is chosen to forecast the regularity and the events. The city region can be partitioned into sub-areas by clustering. Then, in each sub-area, a geographical regularity estimation was executed, it was the usual condition of crowds moving pattern. Finally, the statistic method, such as boxplot, was chosen to find out the outliers. For example, in [62], R. Lee and K. Sumiya developed such an event detection algorithm to identify festival occurrence through analysing the Twitter data. Fig. 4 provides an example that we use this method to detect the event in London. It displays the

density of geo-tagged Tweets before and during the event. The popular region has changed from regular hotspot to the event hotspot. The network optimisation schemes require to be changed for fitting the new-emerging event hotspot.

However, the extracted features of events can be incomplete while using only one data source. To improve in this aspect, H. Becker et al. proposed an approach [59] for identifying scheduled events from not only the social networks (e.g., Twitter, Flickr) but also media hosting site (e.g., YouTube). Another challenge is that the majority of online data is not geo-tagged, which limits the upper boundary of the detection precision. K. Watanabe et al. proposed a real-time local event detection system in [63] using both geo-tagged and non-tagged Tweets.

2) PERSONAL TRAJECTORY

The personal trajectory describes the individual moving path that refers to an ordered time sequence of stops where a user pass or stay [54]. With this context, one can maximise network performance based on the adequately allocated resources. Social network data has the GPS records to find user's staying locations, such as home and workplaces.

In the wireless network, the personal trajectory directs the efficient resource distribution for continuous optimisations, such as cooperative caching. The mobile users may handover to other BSs before finishing the content transmission. In that case, it is better to predict user movement and distribute the caching segments along the trajectory. The required accuracy for a beneficial caching is $> 75\%$ or higher [116]. In proactive load balancing, the handover margin's re-configuration has to be finished before the high traffic (crowds) comes. For example, when a group of users moved to a destination cell in 5 minutes, the network can recover from over-load in less than 48 minutes with the prediction of the trajectory (more than one hour without trajectory) [17].

Moreover, the forecasting of high-mobility users would reduce the handover frequency in the HetNet [136].

a: CLUSTERING BASED MODEL

Clustering based model is to group personal location records into different stops or nodes and allocate the context tag to this location, such as home or work. The stops indicate the locations where users would require some services (e.g., using the network) while they were staying longer than a minimum stop duration. It firstly requires to collect the locations in the testing scenario. Then, the clustering is applied to model the sub-region with stops. The research [137] used affinity propagation clustering following the above procedures. Another work [138] also regarded the trajectory as stops and moves. However, for slow-moving events, such as city tour, previous methods miss some interesting places as objects are moving. Palma *et al.* [70] considered speed by a spatial-temporal clustering to achieve unknown stops discovery.

Another challenge is to avoid the influence of half-way data because the half-way location records distributed sparsely and cost less time for users to stay. According to this, we can filter the half-way coordinates out of raw data through time-based clustering. A time interval can be set to pick out the places with a longer time duration. J. H. Kang *et al.* studied this method in [68] according to the time characteristic using real WiFi-based location system. The WiFi-based trajectory modelling is also a kind of indoor fingerprinting localisation and tracking system whose accuracy can be improved by the Kalman filter [139].

b: MARKOV MODEL

In the Markov model, the future states only depend on their current state and have no relation with all previous states. In that way, it is an available choice to model personal trajectory with probability. Each node is a location, and a transition between two nodes means the probability of the user mobility between those two locations. D. Ashbrook and T. Starner proposed the paper [61] for forecasting multiple users' movement by the Markov model composed of nodes and transitions. However, the GPS data faces a problem that indoor positioning is not as accurate as outdoor. We still need other data source to compensate the model for loss. A work about using BS handover data was published recently by H. Farooq and A. Imran in [69] to predict the students' mobility using a Semi-Markov model.

c: SENTIMENT MODEL

The stops of trajectory have both physical and semantic meanings. As reviewed, the online data can be used to detect both of them [64], [65], such as eating at home or offering presentations at workplaces. Previously reviewed literature concentrated on physical trajectory but gave fewer efforts to know the semantic meanings of interesting places (e.g. home, work). The researchers in [66] utilised Bayesian networks to investigate GPS temporal patterns to find semantic meanings

of frequently visited places. Such personal trajectory model tracked visited places to predict future regular visits. However, we require to know not only the regular movements but also a judgement of the outliers of their daily trajectory. For example, cognitively-impaired elders or blind people will encounter problems if they have such irregular movements. Q. Lin *et al.* studied this problem in [67] based on mining historical GPS data. Moreover, the trajectory during events can also be different from the daily movements. The authors of [53] regarded modelling the user' movements as an event-based trajectory that records new geolocations when we detected new events.

3) SUMMARY OF FINDINGS AND LESSONS LEARNED

In summary, the main findings and lessons learned from the modelling of geolocation include:

- The popular regions indicate the hotspots distribution in a spatial traffic pattern. The proactive optimisation needs this context to decide the most profitable region for resource allocation and infrastructure deployment.
- The clustering methods, K-means and DBSCAN, are widely chosen in the popular region modelling. The K-means is easy to implement with low complexity but requires a manual selection of the cluster number k , leading to a degree of arbitrary parameterisation based on user bias/intuition. Some variants of K-means can mitigate this problem by re-simulating a series of k values, but still meet the negative influence caused by noisy samples. In that case, the DBSCAN based spatial kernel is selected to reduce the noise and highlight the high-density areas. The setting of minimum density determines the popular regions that can be found.
- The clustering method is good at modelling the stops in the trajectory, but it is difficult to track the slow-moving objects which have few stops. In that way, the speed of objects is taken into consideration. Besides, the half-way locations with less staying time require to be ignored. Therefore, it is better to also analyse the data in the temporal dimension through setting the minimum stay threshold. Another problem is that trajectory prediction also requires the transfer probability between the stops. The Markov model can provide such a transition probability. As this model assumes that the next stop is only correlated to the current one, it will ignore the potential influence of the previous stops.

C. MODELLING SOCIAL BEHAVIOUR

Humans have diverse behaviours and attitudes towards a specific object [140]. Understanding user behaviour can help choose the suitable communication service type (e.g., video, audio). Moreover, the attitude becomes the key to guide operators to improve service quality. Social behaviours include content popularity, preference, and relationship.

1) CONTENT POPULARITY

Frequent-accessed content generates the majority load of the network. Content popularity tells which content is liked, accessed and shared by a high number of users. There are two targets of modelling content popularity, the first one is to judge which content will be popular, and the other one is predicting the level of popularity [141]. The content popularity decides the content placement, transmission and storage in proactive caching. Traditional statistic popularity models (e.g., Zipf distribution) have no parameter optimisation, such as the gradient descent in machine learning. One of the promising methods for improvement is to use machine learning based on online data for modelling the popularity model instead of using Zipf distributions, such as applying regression methods to model published YouTube videos' popularity for new video's proactive caching [142]. The error tolerance is low in caching (error rate < 0.5% [116]) due to the high backhaul cost while meeting errors. Moreover, the models are required to be updated continuously to maintain a lower overhead (cost) [116].

a: CLASSIFICATION BASED MODEL

Classification is a supervised-learning method that identifies which category a new observation belongs, its training process is based on the data set whose categories are already known. The work [142] extracted video features as vectors and used SVM to classify videos and build the popularity model. The videos in the same classified category should have a similar popularity trend. Besides, we can also mark contents with tags of popular or not popular, then describe the content with popularity in the first hours and popularity in the later days, finally use classification to model the content popularity. The authors in [73] proposed an example of using the Naive Bayes classifier to recognise stably popular and highly popular YouTube video based on popularity patterns and content-requesting times.

b: REGRESSION BASED MODEL

The content popularity varies along with the changes of some independent variables, such as comments and visit counts. The regression-based model can make the network understand this variation. For example, after the content publication, the first-hour popularity can reflect its level after a month. Such correlation can be described by a linear regression [72]. The authors in [72] worked on an online newspaper article popularity prediction using readers' comments on Digg. However, online content in text-form occupies much lower cellular resources than online videos. The popularity of online videos was studied in [76].

2) USER SENTIMENT

User sentiment is a sense of a user's perception to certain contents or services (positive or negative), and it may affect its subsequent actions (e.g. stop usage or complaint). Understanding user sentiment can help best fit their tastes

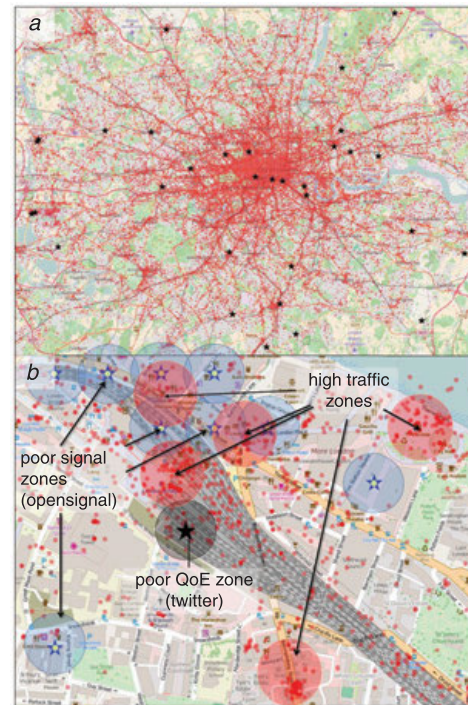


FIGURE 5. Case study of using user sentiment to detect cellular blackspots adapted from our work [33]. (a) The blackspots where users' complaints about the cellular network detected. (b) A comparison of the actual poor signal zones and the poor QoE zone from Twitter.

to increase QoS and even cache the preferred content in advance. In the network optimisation, user sentiment usually indicates the areas with poor network performance and the content a consumer likes or not. Users could complain about network experience through the social network, and such behaviour offers an opportunity to classify the complaints into different service categories. In a work about network coverage blackspots detection [33], the detection accuracy was achieved over 0.6 F1 score with 80 training size. In that work, we used Twitter data as the source to detect spatial user sentiment and provided a case study of London Bridge as shown in Fig. 5. Consumer requests reflect their sentiment on the contents. The works [116], [133] cached each user's interested content in the BS along the trajectory.

a: LEXICON DICTIONARY

Sentiment analysis is mining affective states subjective information. It is a combination of using natural language processing, text analysis and unsupervised learning. An example is given in [33] building an NLP based model using Twitter data to help telecom operators to find QoE blackspots where network optimisation or better deployment is required. They filtered the words related to complaints and located the blackspots with the geo-tags in the Tweets. Such blackspots can be translated to coverage holes caused by uneven BS deployment, power-exhaustion BSs, or emergence BSs under the attack by extremes weathers [143]. Sensors' coverage self-healing provides a path to solve the

coverage problem, such as coverage hole repairing of mobile edge nodes [144], [145].

b: SUPERVISED LEARNING BASED MODEL

Supervised learning produces an inferred model which maps the training data and its labels. Then, this model can allocate new observations with the existing labels according to the estimated probability. The authors in [116], [133] predicted the Youku (like YouTube) video's content-requesting probability by using a neural network. The input of this model was a vector of user context, and the output was the content request distribution. Other researchers estimated the user sentiment from feedback and ratings in [81] using such as Naive Bayes Neural Network, and Nearest Neighbour algorithm. Similar works using different data based on this model, such as the investigation on searching result preference [82], [83]. However, individual preference is not stable over time because of the influence of environments, experiences, and education. Generally, for machine intelligence at the individual level, calibration is needed on a daily basis [146].

3) USER RELATIONSHIP

The online platforms record the relationship between users like Facebook friendship or Twitter subscription. One can use friendship, interaction, latent, and following graphs for representation [147]. These relationships not only indicate the role of a user in society but also provide us with a model to reveal information spreading rules to benefit the dissemination of content [148]. According to [149], [150], the diffusion of information in Social-Physical Networks can be modelled by the strength of social ties. In the network optimisation, sharing the cached contents in friends circles dramatically alleviates the pressure of the core network. This can be done by finding the most influential user in social ties [42]. An example of using Facebook friends in a cooperate caching is [43].

a: INTERACTION GRAPH-BASED MODEL

Online relationship is less valuable than offline ones [89] because some of them might have little interactions after building connections. This skewed distribution of online friendship challenges in estimating close relationships. Therefore, the interaction graph takes visible interactions into account to build the user relationship model. The proactive caching work [43] considered both connections and interaction on Facebook to find users with similar requests. In [85], the interactions, such as wall posts and photo comments, were used to improve traditional friendship graphs on Facebook. However, not all the interactions are as visible as comments.

b: LATENT GRAPH-BASED MODEL

Besides visible interactions, there are also some 'latent' interactions, such as profile browsing or posts glancing, occur every time while using. In [86], J. Jiang et al. built latent interaction graphs based on social links and the tracked profile-visiting histories using statistic methods.

However, 'close relation' can be reflected from not only interactions but also the geolocation that effective close relationship should have overlapped trajectory in their daily life. Otherwise, its value will be discounted in network optimisation. The work [88] studied friendship and location by using statistics to estimate the probability of friendship that is roughly proportional to the inverse of the distance. Similar work is in [90] using Gaussian distribution and Expectation-Maximisation (EM) to fit the periodic and social mobility model which forecasts the relationship between friendship and mobility.

c: FOLLOWING GRAPH-BASED MODEL

In Twitter, users subscribe to others and see all of their public posts. It indicates a weaker offline relationship than Facebook but reflects more power in information spreading on the news media level. H. Kwak et al. modelled this relationship as the following graph in [91] and found that Twitter had a nature of broadcasting, which verified its role as the emerging news media. The information-spreading ability is influenced by numbers of the follower and influential user rankings. Understanding the spatial properties of the broadcasting graph can enable proactive content caching and D2D and P2P joint optimisation [151].

4) SUMMARY OF FINDINGS AND LESSONS LEARNED

In summary, the main findings and lessons learned from modelling the social behaviours include:

- In proactive caching, the popularity prediction is critical to guide the deployment, storage, and transmission of the segments for maximising the caching hit ratio. The forecasting of content popularity becomes a classification problem which matches the known popularity of published contents to the unknown popularity of the new similar contents. The Naive Bayes and SVM classifiers are all popular supervised choices. It should be noted that the selection of dimensions in the classification requires it to maximise the difference between the diverse kinds of content for avoiding the misclassification. It would be useful if the data sets are visualised with the dimensions. Besides, the popularity correlates to the number of visits and comments.
- The users' sentiments represent their tastes to different content or the satisfaction of their experience. In the proactive optimisation, it will enable the sliced virtual networks with functionality specific to the service or customer. Such a sentiment analysis model consists of natural language processing, text analysis and supervised learning, which has already been successfully used in detecting cellular blackspots. The problem is that the results highly rely on pre-defined text corpus, which can have ambiguity and understanding errors. As for modelling user preference, the supervised learning based models can use the user context as an input and produce a distribution of future user's requests of contents.

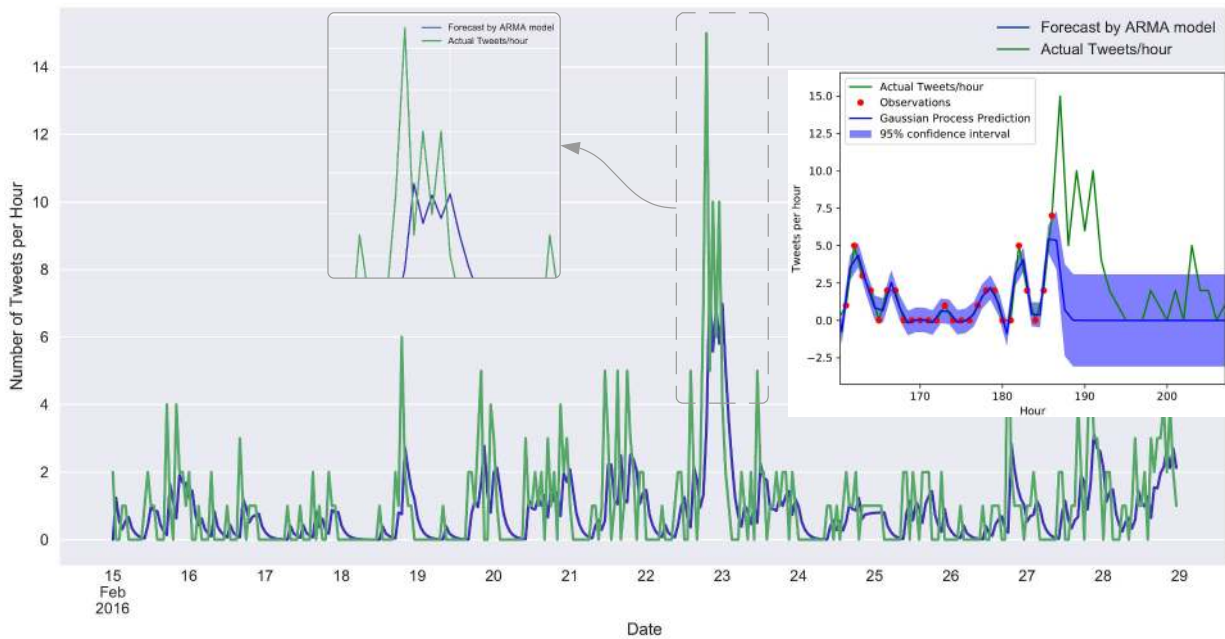


FIGURE 6. The number of geo-tagged Tweets per hour in two weeks (15/02/2016-28/02/2016) at a cinema in Leicester Square, London. A movie world premiere was held on 22/02/2016 where the highest peak of Tweets/hour appears. Such a burst of traffic is unique compared with all the other days in the two weeks, so it is irregular and not predictive according to only historical traffic data. We applied both ARMA and Gaussian Process methods for prediction, and results indicate that they are not capable of predicting the random burst.

However, the individual sentiment is unstable, so it requires a frequent calibration via user prompts [100] which is subject to inconsistent user participation.

- The modelling of user relationship can help to reveal the information spreading rules which are required in the dissemination of contents in the proactive caching, especially on the D2D level. It has three kinds of representations, interaction, latent, and following graphs. The intensity of interactions reflects their relationship and the potential of overlapped trajectories as well as similar requests. In that case, it represents a ratio to adjust the strength of users' relationship. However, not all the interactions are visible, such as the posts glancing or social mobility. The closer locations and similar behaviours increase the probability of a friendship. The Gaussian distribution can be used to fit the social mobility that the probability of friendship is proportional to the inverse of the geo-distance. The above models mainly focus on friendship on Facebook, but there is also another subscription relationship on Twitter. The following graph-based model can determine the most-influential user which may benefit the broadcasting of contents in proactive caching.

D. PREDICTIVE USER BEHAVIOUR

1) SEASONALITY IN USER BEHAVIOUR

The previous models represent the connection between online data and user context. Major of them aim to model the seasonality (regularities) in user behaviour, such as places with

daily visits and preferred content. The regular behaviour has a low possibility to change, which brings convenience for prediction. To track the habits in network usage, it is better to forecast the regular spatial-temporal pattern according to historical data [94], [132]. In general, these results indicate that regular behaviours are predictive because they repeat over time.

2) ANOMALY DETECTION IN USER BEHAVIOUR

However, the anomaly (irregularity) also exists in user behaviour, where traffic-burst randomly occurs on the time-line. It is difficult to model such behaviour along with time, such as parades. For example, Fig. 6 plots the number of geo-tagged Tweets per hour around a cinema in Leicester Square, London in two weeks. A movie world premiere caused an unusually high peak on 22/02/2016. It is difficult to predict such an event according to the regular traffic in previous days (15/02/2016-21/02/2016). To verify this view, we use an Auto-Regressive Moving Average model (ARMA) and a Gaussian Process (GP) model to forecast future confidence interval based on the data from 15/02/2016-21/02/2016. In the Gaussian Process model, the prediction of Tweets per hour y_* is based on the observations y before 18:00, 22/02/2016. Therefore, the probability follows a Gaussian distribution $y_*|y \sim \mathcal{N}(\bar{y}_*, \text{var}(y_*))$ [152], in which \bar{y}_* is the mean indicating the best estimate of y_* and $\text{var}(y_*)$ is the variance representing the uncertainty. The 95% confidence interval is $\bar{y}_* \pm 1.96\sqrt{\text{var}(y_*)}$. This model helps understand seasonal characteristic in historical data. However, the burst on the event day does not obey the seasonal trend.

TABLE 6. Top 5 term frequency of Tweets before and during event 'Grimby World Premiere and Red Carpet' at cinema on Leicester Square, London.

Dates	Top 5 Term Frequency
18/02/2016 Thursday	(Square, 8), (Leicester, 8), (I, 6), (London, 5), (#tgifridays, 3)
19/02/2016 Friday	(Leicester, 11), (Square, 11), (London, 11), (amp, 11), (Greater, 4)
22/02/2016 Monday (Event Day)	(#GrimbyWorldPremiere, 27), (#London, 25), (#redcarpet, 25), (Leicester, 19), (Square, 17)

The key to solve is using online contents to detect the future popular event. The event is the dominating disturbance that enters early in the process of Tweeting. Therefore, the ARMAX (ARMA with exogenous terms) model can be used to track the irregular burst caused by the event. The events are regarded as unusual outliers in the regular traffic pattern. For example, as shown in Fig. 4, the hottest traffic spots will change when the event is approaching, and the term frequency will also alter to event-related words as shown in Table 6. Therefore, we need to find irregular conditions in regularity. We can use machine learning, such as SVM, to classify the Tweets according to temporal-spatial-textual dimensions. Then the algorithm detects upcoming popular events as well as predicts the irregular behaviour of UEs. The paper [17] proposed a context-aware load balancing based on predicting an event in simulation, and [153] also studied that unexpected real-time road traffic prediction and control based the Tweets by waiting drivers.

3) SUMMARY OF FINDINGS AND LESSONS LEARNED

In summary, the main findings and lessons learned from the predictive user behaviour include:

- Majority of the prediction models focus on the seasonality (regularity) in the behaviours because the demand components have periodic variations. Machine learning approaches attempt to balance between best fitting of parameters and avoid overfitting via Bayesian methods. The application examples include: the small cells can be turned on or off according to the periodic variations of network traffic. However, the challenge appears when the network traffic does not follow the seasonality. In that condition, random components break the rules of the seasonality and cause prediction errors.
- The non-periodic random components in the social behaviours cannot be predicted based on the training data in the regularity. We applied the Gaussian Process and the ARMA model to examine the degree to which uncertainty exists. The conclusion is that extra information is required to highlight the time with an anomaly. For example, in event detection, the event information has been posted several weeks or months ahead. We can use this information to determine the anomaly period and apply the detection methods to track the irregularity. In that case, the proactive load balancing strategy has to

balance the newly emerging event's hotspots which can be very different from the daily hotspots (see Fig. 4). This proactive action will increase the convergence ability of optimisation, but it also needs to deal with the prediction errors and the associated overhead.

- We summarise the reviewed papers of the whole Section III in Table 7. This table classified the papers with the fields of used models, data types and amount. In this section, we also introduce how each context is required in the proactive optimisation, which is the link between Section III and Section IV. For example, the popular region prediction needs to satisfy the requirements of proactive load balancing with a minimum spatial granularity of 120 m and 2 hours ahead. Such a prediction can be executed by using the clustering methods on the geo-tagged social network data. The proactive caching also needs the user trajectory prediction with accuracy > 75%. Such requirements can be achieved by analysing the GPS data using the neural networks.

IV. DATA-DRIVEN PROACTIVE NETWORK OPTIMISATION USING CELLULAR AND ONLINE INFORMATION

The context from online data-analysis is the enabler to infer predictive user behaviours, which also enable cellular traffic predictions and further shift current reactive network optimisation to proactive. This section will analyse how to implement them to achieve proactive optimisation.

A. CELLULAR TRAFFIC PREDICTION

Network traffic prediction becomes significant in the proactive optimisation, especially the proactive load balancing. Traditional algorithms construct regression models for the one-step prediction based on records, such as the core network [154] or cell-level prediction [155]. However, these researches face bottlenecks to step further as the resolution is limited to cell-level. One of the solutions is analysing the high-resolution GPS data from heterogeneous datasets, then correlating it to cellular traffic. The work [95] has verified that the network traffic and the size of online data are both positively correlated to the number of involved network users. In that way, online social networks could not only predict flash crowds' needs but also offer operators suggestions about traffic forecasting for resource allocation [87]. This part reviews the development and provides the findings about online-data driven traffic prediction.

1) NETWORK-LEVEL TRAFFIC PREDICTION

The network-level traffic indicates the amount of exchanging information through the backbone network. Such data record the past traffic as a vector in the temporal dimension, which is the training data for neural networks. Then, the trained network forecasts the quantification of traffic at the next time stamp. The researches in [154], [156] followed this way by using a feedforward deep neural network or a Long Short Term Memory (LSTM) recurrent neural network for

TABLE 7. References summary of online data type, amount, and analysing models.

		Section III		Section IV
Context	Models	Data Type	Data Amount	Data-Driven Proactive Optimisation
Popular Region	-K-means [133] [60] [61] -DBSCAN [58] [135] -Spatial-temporal Clustering [58] -Event-detection based model [59] [62] [63]	Social Network [133] [95] [58] [60] [135] [59]	-2000 geo-tagged Flickr images [60] -86 million geo-tagged photos [135] -0.48 million Tweets [58] -21.6 million geo-tagged Tweets [62] -7135 geo-tagged Tweets and 2.45 million non-geo-tagged Tweets [63]	-Resource deployment: [95] hotspot prediction accuracy (correlation > 0.85), 2 hours ahead -Load balancing: [94] hotspot prediction accuracy (120 m), 2 hours ahead -Caching: [133] real-time hotspots clustering covered by flying BS at 332 m height
Personal Trajectory	-Clustering based model [138] [70] -Sentiment Model [66] [67] [53] -Markov Model [61] [69]	GPS [116] [70] [61] [66] [67] WiFi [68] BS data [69] Social Network [53]	125 thousand location points [70] 0.8 million GPS records [61] BS handover data [69]	-Load balancing: [17] group movement to the destination in 5 min -Caching: [116] user trajectory prediction accuracy > 75%
Content Popularity	Classification based model [142] [73] Regression based model [72] [76]	Social network [33] Online news [72] Online videos [76] [73]	YouTube [76] [73] 29 million Diggs [72]	-Caching: [116] content request distribution error < 0.5% converged in 50 iterations and updated in each period
User Sentiment	Supervised learning-based model [81] [33]	Online feedback and ratings [81] Online searching history [82] [83]	1.4 million geo-tagged Tweets [33]	-Coverage blackspots detection: [33] accuracy over 0.6 F1 score with 80 training size
User Relationship	Interaction graph based model [43] [85] Following graph based model [91] Latent graph based model [86] [88] [90]	Social network [85] [86] [91]	-408 million links from Facebook [85] -1.47 billion Twitter relations [91] -1.66 billion Renren links [86]	-Caching: [42] Content sharing with finding most influential user in social ties -Caching: [43] Cooperate caching

this purpose. The results are satisfied in predictions, but they only provide the one-step prediction, which means that the network needs to be re-trained for multiple-step predictions. It may negatively impact the time for further optimisation. If the traffic seasonality and random spike can be decomposed in the training process, the multi-step traffic prediction can be transformed into a combination of seasonal prediction and adding external random information. The Non-linear Auto-Regressive with exogenous model (NARX) makes predictions in this way and solve the one-step problem [157].

2) CELL-LEVEL TRAFFIC PREDICTION

This traffic includes both spatial (BS location) and temporal dimensions. The granularity of prediction is usually in hour-cell level to have a stable seasonality. In other words, based on the hourly data collected from the BSs, the cell-traffic will be modelled by statistic models or machine learning methods.

The temporal traffic consists of the trend, seasonality, and random components. In detail, the trend indicates the overall direction in which the traffic is developing or changing. The seasonality is that the traffic experiences regular and predictable changes which recur every calendar day or other periods. In that case, the cell-level traffic becomes predictable if the trend and seasonality are modelled, which can be easily implemented by ARMA model or exponential smoothing [158], [159]. However, these models only consider a constant time range, which is described by a ‘window’, so the long-term memory of all the training data is neglected. The LSTM is designed for solving this problem by feedback connections in RNN to process the entire sequence of data with selectively remembering patterns. It has a forget gate to disable the meaningless information in recurrent states, such as the random fluctuations in the traffic pattern. One successful example is in [155], but this technique also has some limitations. One of them is that the knowledge learnt

from one cell can not be shared with other cells, which is not intelligent with repeating training effort. It is a promising way to use meta learning to use the conclusions of other learning’s results. The records of other learning methods will be stored and help current training in different cells for learning both temporal and spatial traffic.

The spatial traffic can be described by a probabilistic distribution whose parameters are adjusted for fitting the training data with minimum errors. The traditional method is to formulate using mathematical statistics, such as Zipf distribution [160]. This method finds relations between traffic and locations, and the significance of this relationship. For example, the work [161] designed an α -stable traffic model with parameter tuning for a city-wide scale. The common shortage of this method is that it approximates the parameters without the optimisation like gradient descent so some important details will be ignored. With the development of machine learning, this shortage has been overcome by the neural networks (e.g., LSTM) which owns intelligent weights’ fine-tuning methods like back-propagation [155]. Although the neural networks performed better in prediction accuracy, it lacks the ability for quantifying uncertainty as the mappings between layers governed by weights but not a stochastic process. The usage of Gaussian Process addresses this problem [155]. This non-parametric method trains its hyper-parameters to produce a posterior distribution of prediction with uncertainty quantified. Although the Gaussian Process may not surpass the performance of neural networks, it can quantify the risks via the posterior distribution. In that case, it becomes promising to use deep Gaussian process to couple the advantages of both deep learning and Gaussian Process [111].

3) TRAFFIC PREDICTION USING ONLINE DATA

The problem of previous traffic prediction is the lack of predictive user behaviours, so it becomes difficult to explain and

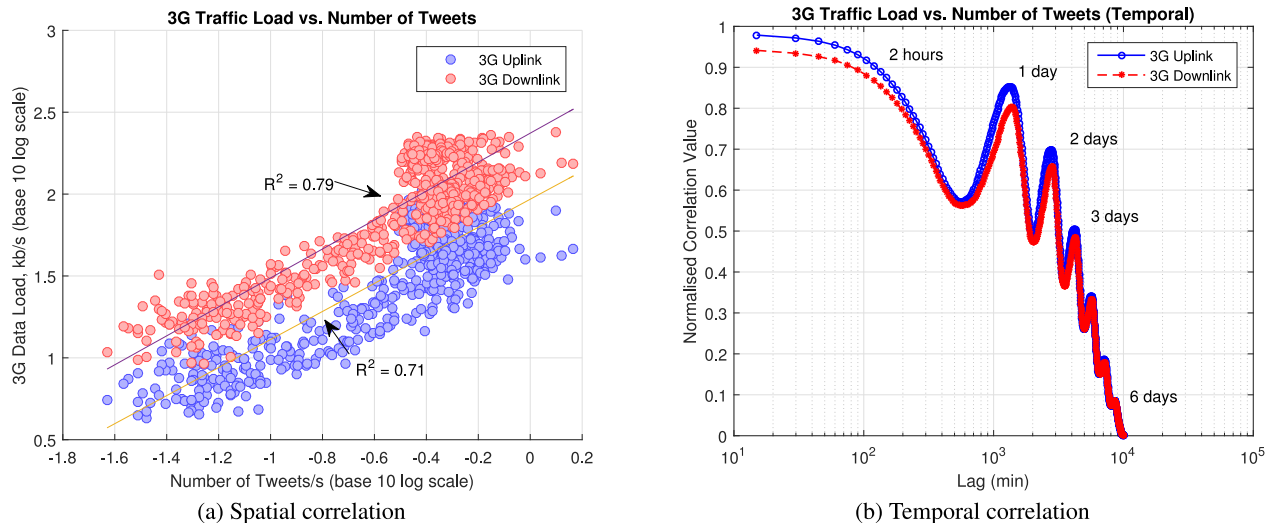


FIGURE 7. The spatial and temporal correlation between Tweets and network KPI adapted from our work [95]. (a) The spatial correlation of a ward-average (a ward is nearly the coverage of a BS) 3G traffic and number of Tweets per second in this ward. It shows a log-linear relationship. In other words, if the Tweets/s (user behaviour) is predictive, we can approximately know the mobile traffic in advance. (b) The temporal correlation indicated the traffic prediction up to 2 hours on the same day and for the same time over the following 1-2 days for both the uplink and downlink channels.

predict the random traffic spike caused by changed services. In that case, heterogeneous data become more important because they contain not only cellular information but also the individual-level geolocation and behaviours (see models in Section III). In the general methods, the traffic is decomposed into the trend, seasonality, and random components. The random components may be the holiday traffic or the traffic spike during popular events. If such traffic is estimated, one will combine the quantification of the three components for the final prediction [162], [163].

This gap is filled by estimating cellular traffic based on geo-tagged social network using machine learning, such as using linear regression [95], [96]. Before training, it is required to select the interested dimensions first, such as cellular traffic and amount of Tweets. Then, the data will be fitted by regression models with minimising residuals like using least squares [96]. The work [96] found a strong correlation between Tweets and mobile traffic in a stadium even though the Twitter traffic takes only a small partition in the whole traffic. However, the correlation is not fixed when the temporal or spatial scale changes. The strength of correlation increases along with the decrease of spatial resolution. It is a trade-off between better spatial resolution or higher correlation. The current method for this problem is re-calculating the correlation with different resolutions to pick an acceptable one [94].

Based on the positive correlation, network traffic becomes predictive using heterogeneous data. The regression methods shall provide the optimised parameters for the fitting correlations. One can formulate the model according to the parameters. For example, the work [95] predicted spatial-temporal traffic based on the estimation of correlation (Fig. 7) between 3G network load and Tweets using log-linear regression.

This estimated Down-Link (DL) traffic load \hat{r}_{DL} in cluster c in time interval t can be described as

$$\hat{r}_{DL}^{ct} = 10^{b_{DL}} \left(\frac{n_{ct}}{\tau}\right)^{a_{DL}} \tau \quad (1)$$

where $[a_{DL} = 0.88 \text{ kb/Tweet} \quad b_{DL} = 2.37 \text{ kbps}]$ and τ is ratio between time interval and second (e.g, in this work $\tau = 3600 \text{ s/hour}$). This formulation couples Tweets and cellular traffic but considers only general conditions. Sometimes, an anomalous traffic emerges without a holiday-like obvious signal. Therefore, an anomaly detection in traffic prediction emerges as another research direction.

4) ANOMALY DETECTION

The anomaly traffic does not follow the model of the trend or the seasonality because user behaviours become different during irregular conditions. The general method is to model the regular traffic first, then detect the outliers based on the modelled regularities. Finally, the outliers will be treated as a particular group with another model to fit the traffic. In that case, we need two machine learning models for both regular and anomalous conditions, which are usually combined with a clustering and a regression. The clustering methods automatically distinguish the regular and anomalous conditions in the selected dimension. For example, using K-means on grouping the BSs with similar traffic will present the BSs with extremely high or low load [164], [165]. The extreme values are useful for proactive optimisations, such as load balancing for extremely high-load cells and BSs turn-off for extremely low-load cells. Another model for modelling anomalous traffic is usually undertaken by regression methods, such as Gaussian Process, neural networks, or NARX model [157]. These methods performed well but faced a

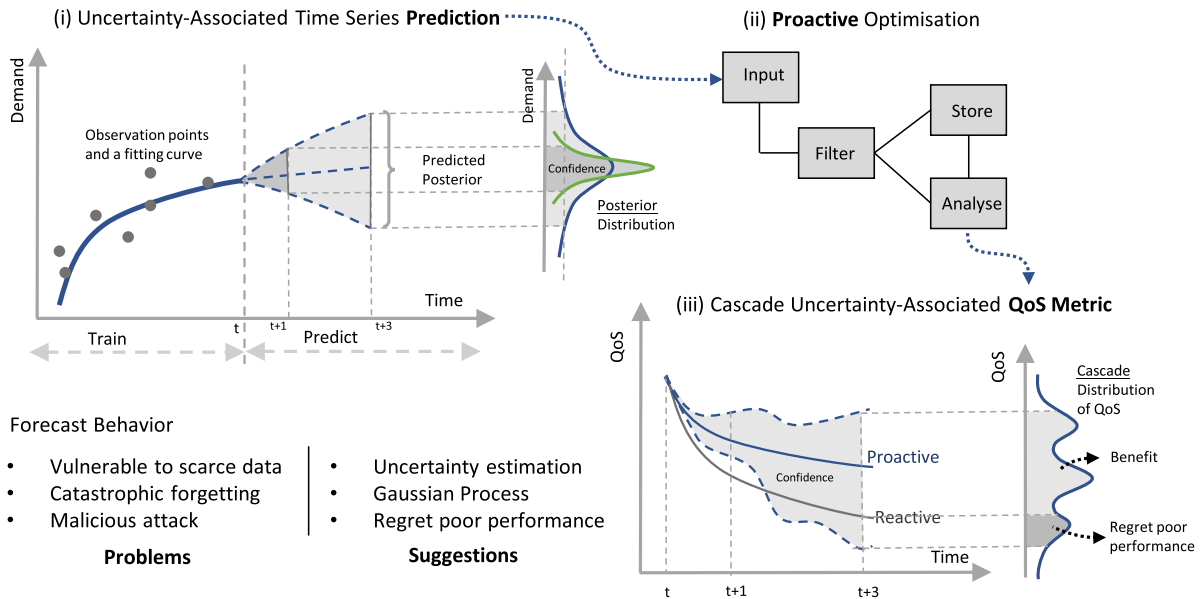


FIGURE 8. The framework and sketch diagram of forecasting behaviours with uncertainty estimation. The proactive optimisation is fed with prediction and its confidence range. It then provides a cascade distribution of QoS. The system can estimate the potential cost and profit to quantify the overhead and make decisions.

challenge in optimising the weights that they can not avoid the local optimum with using gradient descent. This is because the start of the gradient is randomly allocated in the global space. It is not controlled to walk iteratively to a close local optimum then converge. One of the solutions is applying evolution algorithm in parameters' optimisation. The scheme is inspired by biological evolution in which the generations will finish the procedures of reproduction, mutation, recombination, and selection. The mutation provides chances to jump out of the local optimum. In that case, it is promising to use neuro-evolution deep learning to better tune network weights.

5) THE QUANTIFICATION OF UNCERTAINTY IN PROACTIVE OPTIMISATION

To estimate the overhead caused by data scarcity and malicious attack, the certainty of the prediction needs to be quantified to regret the poor performance. We describe the above process in Fig. 8. The detail steps are shown as follow:

- (i) Gaussian Process and Bayesian learning can generate the posterior distribution based on the observations [111], [166]. Such distribution describes the certainty (confidence region) of demand prediction.
- (ii) Then, in the proactive optimisation, the input is the demand (e.g., traffic) samples generated by the posterior distribution, and the output is the corresponding network quality which can be statistically counted by the histograms or the Kernel Density Estimation after several simulations. In general, the simulator outputs the QoS metric according to the traffic posterior distribution and provides a cascade QoS distribution.

- (iii) Such a cascade distribution will quantify the confidence region of the proactive optimisation to compare with the reactive optimisation. In other words, its confidence area describes how the network can operate in future and its probability.

If the QoS of proactive optimisation is estimated to have poor performance worse than reactive optimisation, the regret of poor performance occurs. In contrast, the better performance area becomes benefit or profit. Based on the framework, the final decision is made according to the difference between profit and cost. In general, the uncertainty offers a probabilistic numerical estimation of the profit of proactive-optimisation decisions while facing data scarcity or malicious attacks.

6) SUMMARY OF FINDINGS AND LESSONS LEARNED

In summary, the main findings and lessons learned from the traffic prediction are highlighted as the following items. We also provide a summary Table 8 to compare the methods and suggest the solutions of current pitfalls.

- The network-level traffic prediction can be addressed by the regression methods. These methods can provide high-accuracy results in the one-step forecast but accumulating errors for multi-step prediction. The problem is that the network optimisation requires multi-step prediction to reduce redundant re-training efforts. The solution is using the NARX model regarding random spike as exogenous inputs and combine the multi-step seasonality prediction with the exogenous information. In that way, it alleviates the influence of errors caused by random components.

TABLE 8. Summary of the literature about network traffic prediction.

Reference	Year	Spatial Scale Country: Δ Province: \square City: \circ	Temporal Scale L: Multi-hour M: Hourly S: Minutes	Traffic Type Online Traffic: O Cellular Traffic: C	Prediction T: Temporal S: Spatial TS: Both	Decomposition	Model	Section	Suggested Solutions
[154]	2018	Δ	M	C	T	\times	Deep Learning	IV.A.1)	- Multi-step prediction => NARX model
[156]	2017	Δ	S	C	T	\times	LSTM		
[158]	2007	Δ	M	C	TS	\checkmark	Exponential Smoothing	IV.A.2)	- Long-term selective memory => LSTM
[160]	2011	Δ	M	C	TS	\times	Statistics		
[159]	2016	\circ	M	C	TS	\checkmark	ARMA		- Knowledge share between BSs => Meta learning
[161]	2017	\circ	M	O	TS	\times	α -stable		
[155]	2018	\circ	M	C	TS	\times	LSTM		- Uncertainty Quantification => GP
[167]	2018	Δ	M	C	TS	\times	Neural Network, GP		
[168]	2011	\square	M	C	TS	\times	Markov Chain		
[162]	2017	\circ	L	C	TS	\checkmark	ARMA, Decision Tree	IV.A.3)	- Social behaviours => Heterogeneous data analytics
[95]	2016	\circ	S	O	TS	\times	Statistic		
[94]	2017	\circ	M	O	TS	\times	Statistic		- Estimation of random spike => Anomaly detection
[96]	2017	\circ	S	O	TS	\times	Regression		
[164]	2017	\circ	S	O	TS	\times	K-Means, Neural Network		
[165]	2018	\square	M	C	TS	\times	K-Means, GP	IV.A.4)	- Avoid local optimum of gradient descent
[157]	2018	\square	M	O	TS	\checkmark	K-Means, NARX		=> Neuro-evolution deep learning

- For the temporal dimension, the traffic is decomposed into the trend, the seasonality, and the random components. The first two items are predictable through using the ARMA or the exponential smoothing, but only part of the training data is used to deduce the prediction. Such a requirement about flexible long-term memory makes the LSTM a feasible choice. Its forget gate is trained to remember the meaningful items. Moreover, one of the future researches is to avoid the knowledge catastrophic forgetting between BSs by meta learning.
- In the spatial traffic prediction, traditional methods modelled it by mathematical statistics (e.g., Zipf distribution and α -stable). Compared with machine learning (e.g., neural networks), the traditional ones have no parameter optimisation (e.g., gradient descent). Instead, the parameters are determined using general approaches, such as maximum likelihood estimation which approximates the parameters without finding a path to the minimum gradient. The problem of current machine learning is that the predictions are generated without a quantification of the uncertainty. It causes difficulties for future decision makings to quantify the cost and profit considering potential errors. Such a problem is estimated to be solved by Gaussian Process or deep Gaussian process to produce predictions as posterior distributions (uncertainty).
- The random components in traffic are difficult to explain and predict due to the lack of user behaviours meta-information. This difficulty drives data analytics from cellular only to heterogeneous data (e.g., Twitter data). Current methods concentrate on using linear regression to formulate the positive correlation between cellular traffic and social network. Some statistic methods can quantify the strength of correlation but not formulate the model. Current gap of predicting the random traffic spike is that many anomalous conditions are unknown in advance (e.g., non-periodic events such as protests).

It needs the anomaly detection to distinguish regular and anomalous conditions automatically. The general method is combined with a clustering method (for distinguishing) and a regression model (for traffic modelling), such as a combination of K-means and neural networks. However, the weights selection in the neural network may meet the local optimum by the gradient descent. In that case, the neuro-evolution deep learning can jump out of the local optimum, which is a promising method to model traffic with fine-tuned weights in the long-term.

B. PROACTIVE LOAD BALANCING

Load balancing is required to cope with the imbalance distribution of users’ demand [31]. Specifically, the goal is to handover the UEs at the edge of overlapping or adjacent cells from congested cell to idle cell through optimising handover offset values [169], thresholds [12], and the number of handovers [170]. Afterwards, the SON algorithm was developed to take advantages of Fuzzy logic controllers for auto-tuning handover margins [13], [171]. However, these methods face some challenges:

- The controller-based methods are reactive to random traffic spike, which delays the convergence and results in the limited ability in adapting the fast-changing load [17].
- The controller-based methods have a potential occurrence of oscillations, which may cause re-overload occurrence for target cells [171].

Several proactive load-balancing works use machine learning to deal with the above challenges by deriving, predicting, and adjusting key parameters (e.g., call blocking ratio (CBR)).

1) BALANCE LOAD WITH PREDICTIVE TRAFFIC

The first requirement to be proactive is forecasting the cell load (similar to the number of UEs and the BS traffic). The machine learning methods are selected to forecast the cell

load and decide the offsets according to their correlation. The offset was adjusted automatically based on the cell load with being subject to the minimisation of packet loss ratio. The knowledge learnt from the training process highlighted profitable choices of parameter adjustment. For example, in [19], they used the polynomial regression to formulate the relationship and adjust the small cell offset value. However, the relationship between parameters can be complicated and related to a lot of parameters. Reinforcement learning performs better than polynomial regression in complex scenarios.

Q-learning is a reinforcement learning to solve the problem by learning the state-action table from training data. The state is indicated by the cell load, and the action represents the optimisation decisions (e.g., offset values or antenna down-tilt). With the employment of this model, the Reference Signal Received Power (RSRP) margin can be continuously adjusted according to the state-action table to maximise the user QoS. The works [21], [172], [173] followed this way to self-tune the cell margin or the antenna down-tilt.

2) BALANCE LOAD WITH CALL BLOCK RATIO OR UE LOCATIONS

The CBR is an indicator of cell load because such a ratio will increase along with the rising cell load. Another indicator is the average distance between the BS and the neighbouring UEs. If the UEs are far from the BS, the cell load is estimated to be low. According to the indicators, similar state-action tables can be trained. The works [20] used such table to update the CBR-offset fuzzy rules. The work [174] implemented the distance-based target BS-selection algorithm in the UE to handover to the BS with maximum QoS. The problem of these methods is that the state-action table is generated in regular traffic conditions. The networks also have event-like random conditions with very different traffic pattern. One of the solutions is considering extra user-context information, such as the event traffic pattern.

3) MACRO CELL OFF-LOADING

Another way to balance load is to offload traffic from macro-cells to small-cells, WiFi and D2D connections to maximise energy efficiency. The offloading process was coupled with proactive content caching and finally resulted that most energy savings were from prediction and delay tolerance. The first step is to forecast context information, such as network traffic, user mobility and preference. Then, the offloading target function is formulated with maximising energy efficiency or the QoS. For example, the work [175] proposed this study in the 3G scenario with collecting data from YouTube and Apple iTunes to forecast consumer mobility and preference for forming the hot zones. Moreover, the papers [176], [177] focused on offloading to small cells and D2D connections toward high energy efficiency.

The Q-learning method builds a table of states (indicated by the cell load) by actions (the offloading decisions) in the training process and suggests the best action next time [178]. However, the social relationships and content popularity are

not considered, which can forecast the repetitive content downloads. Rerouting this mobile traffic to other access networks is a good choice to offload the macro cell and the core network. For example, in [179], they denoted the offload ratio of content j of user i as w_{ij} , the size of content j as L_j , the requested times of content j from user i as m_j^i , and the backhaul capacity as C_{BN} . In that case, the backhaul utilisation U_{BN} is

$$U_{BN} = \frac{\sum_i \sum_j w_{ij} L_j m_j^i}{C_{BN}} \quad (2)$$

The backhaul utilisation can be forecast if user preference m_j^i is predicted (such prediction models can be found in Section III), and the target is to minimise the backhaul utilisation in the future actions.

4) SMALL CELL SWITCH-OFF

The switch-off algorithm can disable idle devices for a sleeping interval [32], [180], which reserves the energy consumption in low-load scenarios. The resource allocation for this purpose is similar to the marketing models with bidding for maximum profit. The objects toward high profit are mapped to the trade-off among conflicting financial interests, such as cost (e.g., power) and profit (e.g., capacity). In detail, the network needs to 'bid' for the resource of third parties to carry their low loads. Therefore, we need to forecast the offloading traffic to 'buy' the requested capacity. The work [181] utilised this model considering time-varying traffic to switch off HetNet nodes and offload the remain UEs to third parties' cells.

Traffic prediction decides the involved cells and the time to wake up. In the temporal traffic prediction, the low-load intervals provide the basis for a system to determine which cells need to be turned off for how long. The work [182] did a similar work with the target of maximising energy efficiency. Beside load prediction, the sleeping intervals should be set in advance according to overlapped areas, battery condition, and previous settings. The traffic prediction is not the only requirement. The work [183] studied this problem considering extra context information.

The appropriate number of nodes to be switched off depends on the traffic pattern. The traffic changes to a low level at night, so its forecasting will suggest the maximum number of nodes to be switched off [184]. In a cell, the average distance between BS and all the UEs can indicate the power consumption because the BS has to increase its power to maintain stable links for further users. Therefore, researchers choose to switch-off the cells with the highest average distance with avoiding the QoS degradation [185].

5) DATA-DRIVEN USER ASSOCIATION

The users in the overlapping area become the load to be balanced. While the cell's capacity is limited, these users will still occupy the rare resource in this crowded cell even through they can be transferred to the adjacent idle cell without much increase of pathloss. Such traditional user-cell

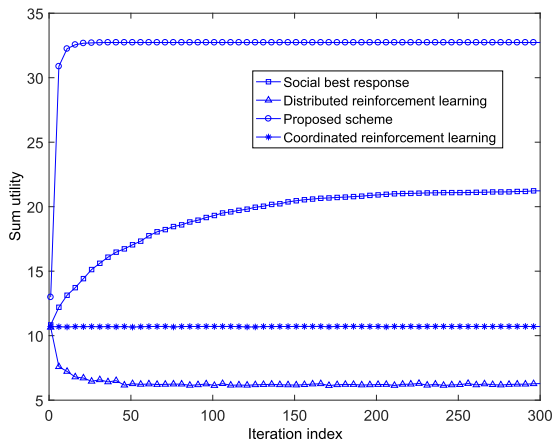


FIGURE 9. A comparison of convergence speeds adapted from our work [57]. The data-driven user association scheme owns the best performance to balance load in the cell overlapping areas.

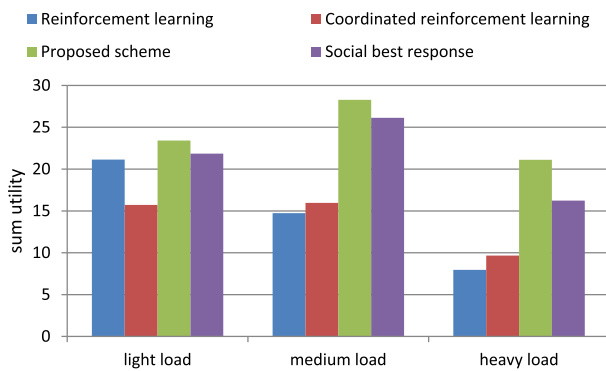


FIGURE 10. A comparison of total network utility with different loads adapted from our work [57]. The proposed method achieved the highest utility, especially in the condition of heavy load.

association schemes lose the opportunities to connect users on the far side. To solve this problem, better user-cell association will benefit the total utility of resources by making rooms for more potential candidates.

To maximise the total utility of users, the users in overlapping areas should be carefully associated even the pathloss is a little bit worse, but the overall maximum target is achieved. In detail, each pair of overlapping cells cooperatively re-associate users based on analysing the QoE data. Our work [57] proposed an iterative optimization method to implement this idea. The analysed QoE information was shared through X2 interface or cloud message exchange, then the involved small cells were triggered to optimise current association profile. Fig. 9 presents the convergence speed and compares with other three methods, two reinforcement learning and a social best response. It shows a fast convergence and high sum utility (above 30). These advantages make the data-driven methods compatible with proactive optimization. Under the circumstances with different loads, as shown in Fig. 10, data-driven method also results in the best-profiting association.

6) PROACTIVE LOAD BALANCING BASED ON ONLINE INFORMATION

The above proactive algorithms can relieve the pressure of oscillations and re-overload. However, the learning process is only cellular data-driven, which has some problems:

- There exists unnecessary cell expansion for the cells far from a hotspot.
- It lacks user behaviour, such as events.

These problems limit the roof of existing proactive algorithms. To address them, we need a proactive load balancing based on the analysis of online information.

The context-aware module in this algorithm should provide predictive user distribution of potential high-loaded areas [8]. The research [17] proposed a heterogeneous data-driven distribution-aware proactive load balancing study aiming to solve the unnecessary cell-expansion problem. In the design, the context-aware module collected and consolidated data from Twitter and GPS, then output user distribution which was also the forecasting of load difference between the loaded cell i and its nearest idle cell j , $LR_{diff}(i, j)$ (it is the input of the integration module). In the integration module, the forecasting of $LR_{diff}(i, j)$ suggested the loaded cell i to increase its speed to share the load to its nearest idle cell j . For example, the loaded cell i reduced its transmission power by $k \times \Delta P_{TX}(i)$. Such power reduction $\Delta P_{TX}(i)$ was adjusted according to the controller's output. Besides, the integration module could adjust the strength k to enhance or weaken the effect of load balancing algorithm according to the prediction of environmental changes.

Compared with the performance of Fuzzy controller-based optimisation, the context-aware optimisation could reduce at least 1.3% more user dissatisfaction rate, and nearly 50% convergence time. This research indicates a path to couple proactive load balancing with the online information.

7) SUMMARY OF FINDINGS AND LESSONS LEARNED

In summary, the main findings and lessons learned from the proactive load balancing include:

- The cell offset needs to be determined according to the learnt correlation between the offset and the cell traffic subjecting to achieving the minimum packet loss. The polynomial regression and the Q-learning are the commonly chosen tools, where the cell load is an input, and it outputs the adjustment of the offset. One problem of this method is that the regular conditions and the random components are mixed in the modelling, which causes a slow convergence in an anomalous condition (because the anomaly is not fully learnt). In that case, the anomaly detection needs to be implemented here to learn the anomalous traffic and alert the system to fit it.
- The behaviour information will enable the network to fit customers' demands and forecast the upcoming change of traffic. The localisation systems provide a high-resolution intra-cell traffic pattern, which will

TABLE 9. A summary of proactive load balancing according to the focused parameters.

Proactive Load Balancing Works	- Determine RSRP margin based on number of active UEs to minimise the packet loss ratio [19] - Determine BS transmit power and RSRP margin based on CBR and OR to maximise reward of LB [20] [186] - Determine RSRP margin based on cell load to maximise reward of LB [21] [187] [172] - Determine RSRP margin based on UEs' and BSs' distribution to maximise reward of LB [174] - Determine antenna down-tilt based on cell traffic [173]
Learning Methods	1) Polynomial regression [19] 2) Q-learning [20] [21] [174] [172] 3) Unsupervised learning [23] 4) Neural network [187]
Required Parameters	1) Number of active UEs, cell load [19], [21], [187] [172], [173] 2) CBR, OR [20] [186] 3) UE and BS distribution [174]
Online Data Analysis	1) Prediction of spatial-temporal network traffic [95] 2) Prediction of UEs' distribution [94] 3) Prediction of high-load [17]
Suggested Solutions	1) Cold-start controller => Start with future network traffic prediction 2) Active all neighbour BSs for LB => Active the hotspot-nearest BS for LB 3) Cellular data learning => Heterogeneous data learning 4) React to a burst of traffic demand => Predict events and reduce the range of cell in advance

suggest to enhance or weaken the effect of the controller-based load balancing according to the forecasting load difference of cells. Currently, the context-aware module is not well-designed yet due to the uncertain predictions and the difficulties to quantify the cost of taking risks to optimise network following the predictions. One method to avoid the potentially heavy cost is to use a parallel model to operate proactive optimisation while reserving the chance to be shifted back to traditional optimisation for minimising the risks. This method is reliable but does not quantify risk in a probabilistic framework. Another way is quantifying the cost based on the posterior distribution of predictions. It is a promising direction to address the concerns of overhead by the Gaussian Process or deep Gaussian process.

- In the Table 9, we summarise the proactive optimisation researches' topic, learning methods, optimised parameters, and the findings to use online information to fill the gaps. For example, current proactive methods need to active all neighbour BSs to participate in load balancing. This problem can be solved with the high-resolution hotspots detection and enabling only the nearest cell for this work. Such hotspots modelling methods can be found in Section III. B. 1. Popular Region and Section IV. A. 5. Social Network-based Traffic Prediction.

C. PROACTIVE CACHING

Mobile Edge Caching algorithm stores relevant data to the nearby BSs based on predicting user demands. This technique increases network performance regarding throughput

and latency. To maximise the hit-ratio of the cached content, the issues about content selection, content placement, content delivery, and storage utilisation should be addressed in a proactive way [31].

1) CONTENT PLACEMENT

To maximise the profit, the systems prefer selecting to cache the most popular content to the closest routers (BSs) to user-side and distributing content packets along with the path. This is popularity-based content dissemination which aims to improve the system performance regarding the server-hit rate and expected round-trip time. The content popularity represents the probability that the BS-associated UEs will request the content in future. The Zipf distribution is a traditional model to forecast popularity as the accumulated request probability in simple conditions (e.g., slow changing popularity). The contents are cached according to their estimated popularity [188]. To improve the performance in complex scenarios, reinforcement learning is another way to execute actions (cache contents) according to a trained state table (service delay). The reward is given with decreasing the delay (see example in [189]). Generally, the machine learning methods are used to model user-files correlations to guide the cache deployment [190].

However, it is inefficient to make the BSs with similar popularity repeat the accumulating works. Instead, it is better to update the estimation according to the knowledge of other BSs. The BSs can be grouped by clustering methods according to traffic, content preference, and storage. Then, the modelled popularity is shared through the control plan interface (e.g., X2 interface) and updated according to own preference [191]. This scheme also benefits the cooperative caching for sharing content between adjacent BSs and distributed caching for fetching contents from multiple BSs (see example in [45]). The general idea is following the trajectory prediction of users and deploy the segments. Nevertheless, an optimisation is required to maximise profit according to the above learnt knowledge. It is a matching game by considering both BS preference (most popular contents) and server preference (low transmission time). The best result is to cache the most-requested content with the lowest transmission time to minimise the backhaul usage. An example is in [192], they regarded this problem as a many-to-many matching game.

2) CONTENT DELIVERY

The content delivery represents an efficient strategy to prepare the contents before being requested. This process is usually scheduled during an off-peak time [45]. The traditional way is broadcasting the most popular contents at the off-peak time [192]. It is efficient at the beginning but becomes costly while facing fast changing popular contents. The developing direction is user-oriented. If the cache placement is in UE-level, the D2D link becomes more efficient for content delivery. The UEs with close social ties are potentially interested in the same contents, so the content delivery based on social ties becomes a new research direction. It is useful

to highlight the influential users by using a centrality metric to estimate the social influence of users (higher centrality means more influential). Moreover, the content dissemination models can be modelled by a stochastic Dirichlet process, such as the Chinese restaurant process [42].

3) STORAGE UTILISATION

The issue of storage utilisation rises as the storage of caching is limited. Some contents have to be removed for refreshing up-to-date contents. A traditionally used method is the Least Recently Used (LRU). The LRU algorithm lists a ranking from most recently used content to least recently used content and deletes the bottom item for the extra storage [27]. However, continuously tracking the accessing information of content is expensive. The best way to solve is forecasting the size of cache and deleting the least-popular items. This scheme preserves local memory for the upcoming popular contents and combines the prediction of storage with traffic for flexible transmission. The reinforcement learning can build a storage-cache table considering maximising the storage usage [193].

4) DYNAMIC CACHING WITH MOBILITY PREDICTION

The probability of content requesting differs when time and area change. Testing the caching strategy with dynamic popularity and behaviours becomes more important these days even though current simulations are executed with many ideal assumptions [194]. User mobility is one of the influential factors. The predictions of cell transition and cell sojourn time suggest the best segment to be stored and the time to delete it [195].

Generally, the mobility prediction continuously provides a ranking of content request percentage. Then, the contents are cached into the edge BSs (e.g., Remote Radio Heads (RRH)) according to this ranking. Additionally, with mobility prediction, some UEs who are difficult to be supported by BS caching can be served by the caching on Unmanned Aerial Vehicles (UAV). RNN is a useful tool to model temporal dynamic behaviour, such as mobility. The works [116], [133] applied this algorithm with the Echo States Networks (ESN) due to its easier training process and less computational effort. Future improvements could combine the ESN with reinforcement learning to build a deep reinforcement learning network for better cache actions considering the complex scenarios.

5) COLD START PROBLEM

Cold start problem means that the system starts with no prior information and needs some time to collect sufficient information and converge. The traditional methods chose to cache random content or even all files at the start to collect the popularity of these content [27], [42], [190]. This cold-start popularity generation required time for enough accuracy and had difficulties in supporting frequent user movement.

To overcome the cold start problem, transfer learning is one of the methods. The transferred knowledge can be learnt

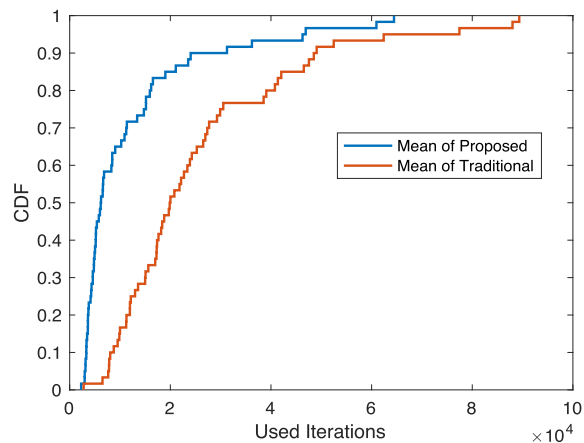


FIGURE 11. A comparison of computation iterations used for traditional resource deployment method and data-driven method. The reduction of computation time helps to avoid the cold start problem.

from other communications (e.g., D2D interactions) to support caching at initial stages with content popularity and social ties. An example of using transfer learning is in [196]. However, this method is hard to quantify the principle of the required knowledge to transfer in all cases. For example, the knowledge of rural areas can be very different from urban areas, but the definition of rural/urban can not be understood if they are not manually defined. Another method can avoid this problem by directly gaining prior knowledge from online data, such as video popularity prediction. During an off-peak time, the content selection and placement were already done, and there would be no cold start problem during peak time [116]. Here, we present an example to show the performance of using data-driven method to avoid the cold start problem. A spatial popularity map is analysed and provided to the caching resource deployment. Traditionally, meta-heuristic algorithms, like the Simulated Annealing algorithm, are chosen for deployment. They will waste efforts in searching low-demand spaces. The spatial popularity map is effective to reduce the searching space and used iterations. The computation iterations are compared in Fig. 11. A Cumulative Distribution Function (CDF) is given, an average 53.85% reduction of computation iterations is achieved.

6) PROACTIVE CACHING BASED ON ONLINE INFORMATION

Online information provides the popularity of Internet videos, which can be represented as $p_i(t)$ of the video i at time t with the total size s_i and the cached portion $\alpha_i(t)$. In that way, the load of backhaul can be expressed as $\sum_i (1 - \alpha_i(t)) s_i p_i(t)$, $0 \leq \alpha_i(t) \leq 1$. The caching algorithms aim to increase the $\alpha_i(t)$ to minimise the backhaul load according to content popularity $p_i(t)$.

The general method of popularity forecasting of a new video is to be classified into the categories of the published videos. Firstly, the unpublished-video features are extracted by a CNN. Then, the features clustering is applied by treating multiple published videos with similar features as a

TABLE 10. Summary of proactive caching literature.

Cache Placement	- Algorithm: BBU [116], [133], core network [45], BS [42] - Content: Small BS [27], [42], [45], [189], [190], [192], RRH [116], [133], cloud [116], UE [42], [45], UAV [133] - Popularity&mobility based placement: [116], [133], [195]
Content Delivery	- MmWave BBU-UAV-Users links [133] - Off-peak transmission [45], [116], [113] - Cooperative caching [45] - Distributed caching [45] - Share with social ties [42], [195]
Storage Utilisation	- LRU [194] - Myopic algorithm [27]
Data	- Youku China video index and University student mobility (map the data sets arbitrarily) [116], [133] - Social network [42], [192] - University user trajectory [195]
Targets	- Maximum long-term effective capacity [116] - Enhance the QoE while minimizing the transmit power [133] - Maximum content hit ratio [27], [45], [195] - Minimum service delay [189]
Complexity	- $O(UT_\tau/\tau)$, where U for number of ESN algorithms and T_τ/τ for number of content replace-time [116] - $O(2UT/H)$, where U for number of users, and the prediction triggers at every H in a period T [133] - $O(MF \log F)$, where M is the number of small cells and F is the number of files [190]
Learning Methods	- ESN [116], [133] - Conceptor based ESN [133] - K-means [133] - Collaborative filtering [42], [45], [190] - Supervised learning [42], [190] - Reinforcement learning [27], [189] - Spectral clustering [189] - Transfer learning [192] - Zipf distribution [45], [189], [190], [195]
Cold Start	- Yes [27], [42], [190] - No [116], [133], [192]

single video category with a representation vector. Next, high dimensional features extracted from unpublished videos are transformed to a representation vector to be classified. Finally, the popularity is predicted by a regression model.

The researchers in [142] worked on this and got a good result which presented the offloading ratio as a function of time sequence. The popularity of published video is updated along with time in the condition of ‘with prediction’, so its performance is approaching the genie case. Current method uses the perception ability of CNN to learn the video features, but it is still short at deciding the optimal cache placement scheme. In contrast, reinforcement learning is good at decision making. In that case, deep reinforcement learning is a promising way to improve the cache decision ability due to the coupling of deep learning and reinforcement learning.

7) SUMMARY OF FINDINGS AND LESSONS LEARNED

The Table 10 summaries the details of the reviewed literature in this subsection. The main findings and lessons learned from the proactive caching include:

- The content placement is a delay-minimising problem solved by the reinforcement learning or a matching game to best link contents to BSs. However, the traditional methods trained the model without knowledge sharing between BSs. This problem is estimated to be addressed

by cooperative caching based on popularity and mobility. Then, the content delivery needs the modelling of traffic seasonality to forecast the upcoming off-peak time. If the content should be delivered to the UE-level, the social ties and high influential users will enable the cache delivery through D2D links or the broadcast of the most popular content. Finally, the storage management has a similar target to save the transferred content and empty the spaces for another high-popularity cache. We find that the proactive caching not only needs to predict content popularity but also the mobility and relationship. Fortunately, all of them are available in online data.

- The mobility enables the dynamic caching updated with the user trajectory predicted by RNN. General RNN methods require high computational efforts and more training time, so the ESN is designed for improvement in this aspect. The gap of using only ESN is that its ability of decision making is not as good as reinforcement learning. In that case, deep reinforcement learning becomes a promising way to combine the advantages of both learning methods. Besides, the cold start problem exists in current proactive caching schemes. One can solve it by transferring knowledge from other networks, but the ‘appropriate’ principles need to be pre-defined to avoid mis-learning of useless knowledge. Online data-analytics can avoid this problem by directly learning from the requested services.
- The online information is the basis for forecasting new content’s popularity. Based on the popularity, the cache size will be optimised by the caching schemes to minimise the backhaul load. However, there still exists some challenges in further developments. For example, the prediction errors or fast-changing popularity can generate more burden in the fronthaul. The decision should not only be made according to the probability of occurrence but also the certainty of this prediction. Gaussian Process is estimated to be a feasible solution for this problem.

V. FUTURE RESEARCH DIRECTION: ONLINE DATA ANALYTICS-BASED 5G PROACTIVE NETWORK OPTIMISATION

In a 5G network, the explosive communication demands urge current network optimisation to transfer from passive to proactive. In detail, the proactive algorithm dynamically follows the changes of circumstances and executes optimising strategies to maximise the efficiency of resource usage. In that case, how to monitor rapidly changing environments (contexts) is the first problem to be solved. Online data is no doubt a valuable source with all the users as environment-monitoring sensors to provide the essential context for future network optimisation. In our survey, an optimisation-context-data map is proposed to clarify which kind of proactive optimisation requires what contexts through analysing which online data. However, despite the

work about data analysing (extracting contexts from online data) and simulated proactive optimisation (assuming that the contexts are already known), there are still open issues and challenges to complete the circle of making online data as a reliable data source. In this section, we discuss these challenges and issues along with future research directions.

A. REALISATION IN 5G

In the past years, lots of researchers have discussed the 5G network about the key technologies, such as millimetre wave, network densification, and massive Multiple-Input Multiple-Output (MIMO). Among all of them, densification is regarded to bring significant changes to the current cellular network because of the increasing inter-cell and intra-cell handover [29]. Therefore, the proactive trend should fit into the change of paradigm in 5G with a platform with security, collection, analysis, measurements, and efficiency. This section presents some research directions about achieving 5G proactive network optimisation.

1) CONTEXT GRANULARITY

Granularity represents the measure of the distinguishable scale of detail in the context. Specifically, in the popular area prediction, various levels of granularity of locations are provided according to different data sizes, geo-tags accuracy, and data-mining methods. For example, in [62] the festivals' areas were estimated in the scale of ward or town, while in [33] the Quality of Experience (QoE) blackspots and high traffic zones were detected in a much smaller granularity (in the London Bridge Station). There is no best granularity for every optimisation, but the real-world data sets indeed have granularity limitations. Therefore, the first challenge is to judge if the online data source has an achievable granularity in expectation.

Furthermore, the granularity of context should be consistent with the requirements of proactive network optimisation. For load balancing examples in [23] and [17], in [23] the granularity of cell load (traffic or area popularity) was at a small-cell level for the cell-offloading research, while the traffic distribution granularity in [17] was finer as at the intra-small-cell level for more efficient offloading. In other words, each network optimisation has required granularity, so it is necessary to determine it before data analysing.

For the 5G, the cell size varies, and there exist dense deployments of small cells. In that case, the context granularity becomes the premise factor, and the context extracted from data sets should achieve at least a small-cell level. Besides, the authors expect that real-time network data to be collected and used, so new challenges emerge about the prediction errors.

2) PREDICTION ERROR IMPACT

Prediction is the process of future context deduction according to experience, so the prediction error refers to the context information that is irrelevant for the future. Such unwanted variation always exists during sampling, training and testing.

Moreover, both noisy data quality and rapid changing circumstance can cause it. For example, in geolocation prediction based on Twitter data, there are a lot of GPS coordinates shifts caused by weak signals, especially for indoor Tweets. In that case, these indoor Tweets provide a location range (approximate a rectangular area) if Twitter cannot estimate an accurate coordinate. Such indoor data takes a large number of all geo-tagged Tweets. On the one hand, the geolocation prediction is inaccurate if we consider all location records. On the other hand, the prediction cannot represent the indoor user' context after filtering.

Accordingly, reducing the prediction errors and limiting the impact of unavoidable ones deserve researching efforts. Online data is unavoidably noisy and related to uncertain predictions. Therefore, future online-data based proactive optimisation algorithms have to take the impact of prediction errors into consideration.

B. AGGREGATION OF ONLINE DATA AND CELLULAR KPIS

Cellular network KPIs refer to radio network performance monitoring, performance degradation detection, and network resources optimisation. Such data is usually regarded as the first choice of cellular big-data mining due to its accurate measurements, significant amount and tidy format. However, network KPIs have two shortages. Firstly, the data source is not open to the public because of privacy and security. Secondly, the KPIs cannot reflect user's intents so they cannot detect demand burst in a long-term period. Therefore, the following two research directions are proposed to alleviate the negative impact.

1) COLLABORATION WITH DATA PROVIDERS

If the network is expected to learn from past user behaviour, the choice of data collection becomes critical. Most researchers collect non-real-time and limited data by crawling. For example, researchers mainly collected Twitter data by two APIs, Twitter search API (search Tweets with keywords in the previous seven days) and Twitter streaming API (up to 1% of all Tweets) [197]. This method is neither sufficient nor efficient. Besides, the crawled data is not representative because of the characteristics of different regions, cultures, and languages diverse a lot [198]. In that case, we need to build collaboration with service providers.

Another essential data source is the network KPI owned by mobile network operators. It directly reflects network status in a short period and becomes the only indicator for network-optimising specialists. In other words, these parameters are the foundation for current passive network optimisation and so as the future proactive optimisation. However, network operators forbid external access to such data, so current researchers rely on either historical cellular data [33], [95] [94] or simulation. Therefore, exploring collaboration with network operators is also necessary.

Traditional mobile network operators only provide communication services, so they face difficulties collecting meta-information needed to forecast context-aware demand

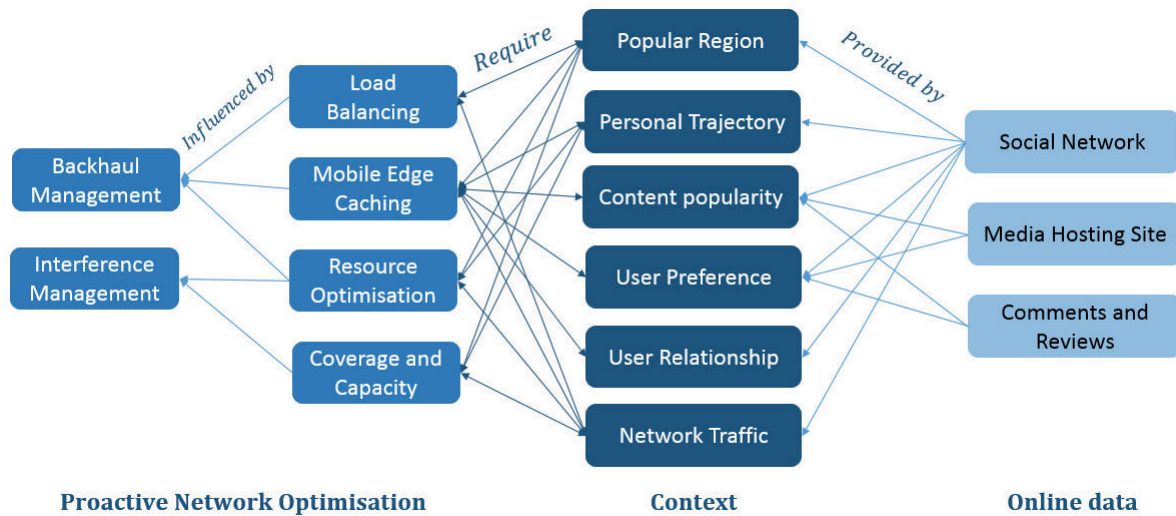


FIGURE 12. An optimisation-contexts-data relationship map. Proactive optimisation algorithms connect to online data through different contexts. Each optimisation own unique characteristic but have relations with others.

for proactive optimisation accurately. However, as Internet traffic becomes the majority of daily usage, there appears more cooperation and combination of mobile operators and Internet service providers. For example, in China, the Mobile Virtual Network Operators (e.g., Tencent, Weibo, and Taobao) are also Internet service companies in the realm of the social network, video, and e-commerce [199]. Such a combination of the service provider and the mobile provider brings convenience for the heterogeneous data analysis. Besides, the mobile network operators, such as China Unicom, also cooperate with service providers to serve users with particular application goody bags, such as unlimited mobile data for Youku videos. In that case, the operators can have traffic data for social behaviour analysis.

Future proactive optimisation should take advantages of both network KPIs and online data. Accordingly, finding a reliable and sufficient data collection method becomes another challenge.

2) OTHER OPTIMISATION SCHEMES

There are still other aspects in network optimisation, such as interference management, coverage and capacity, and resource optimisation. Most of the current attempts are real-time self-optimisation, but we can find some chances to benefit them using online data analysis. For example, H. Claussen et al. proposed a femtocell self-optimisation on coverage and capacity in [22]. They simulated user mobility in an indoor scenario as the context for coverage adaption, which can be modelled by personal trajectory with context (including way-points and spent time). In interference management, the researchers use polynomial regression [14], [200] or a neural network-based cognitive engine [201] to model the relationship between traffic distribution and transition power. With traffic distribution prediction, the system

can select the potential high-interference areas and trigger interference management. For reference, we summarise the context and potential applications as a map (Fig. 12) with possible connections for all proactive optimisation.

VI. CONCLUSION

In summary, this survey demonstrated the essential state-of-the-art technologies in online-data analytics that can offer promising drivers to shift network optimisation from passive to proactive in 5G.

Increasingly amounts of online data contain rich meta information of individual demand context (e.g., personal trajectory, user preference, and user relationship) and the wider social context (e.g., popular region, content popularity, and network traffic). The information when appropriately processed through machine learning mechanisms and environmental data can be used to forecast traffic patterns across multiple population and spatial scales. Then, in turn, it provides the capability for proactive optimisation that could allocate resources to suit diverse service requirements and the complex dynamics in advance by using the forecast information.

In this survey, to reveal the potential data to optimisation mapping, the authors virtualised the context as paths for connection to help readers find the most valuable context and its available data sources. Different models are proposed to retrieve predictive user behaviour and further correlate to network KPI.

The authors strongly believe that in future 5G networks, the optimisation will be proactive, service-oriented as well as user-oriented. In that case, online data becomes an indispensable source to increase QoE and reduce OPEX. However, open challenges still exist, such as context granularity in the 5G scenario, prediction errors, real-time data analytics, and taking full advantage of both online and cellular data.

REFERENCES

- [1] J. Yao, T. Han, and N. Ansari, "On mobile edge caching," *IEEE Commun. Surveys Tuts.*, vol. 21, pp. 2525–2553, 3rd Quart., 2019.
- [2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [3] A. Imran and A. Zoha, "Challenges in 5G: How to empower SON with big data for enabling 5G," *IEEE Netw.*, vol. 28, no. 6, pp. 27–33, Nov. 2014.
- [4] Juniper Research Finds. (Nov. 2014). *Mobile Operator Costs Pass \$800bn Annually as Margins Become Squeezed*. [Online]. Available: [https://www.juniperresearch.com/press/press-releases/mobile-operator-costs-pass-\\$800bn-annually-as-marg](https://www.juniperresearch.com/press/press-releases/mobile-operator-costs-pass-$800bn-annually-as-marg)
- [5] *Technical Specification Group Radio Access Network; Study on Context Aware Service Delivery in RAN for LTE; (Release 14)*, document 3GPP TR 36.933 V14.0.0, 3rd Generation Partnership Project, 2017.
- [6] *Technical Specification Group; Services and System Aspects; System Architecture for the 5G System; Stage 2 (Release 15)*, document 3GPP TS 23.501 V0.0.0, 3rd Generation Partnership Project, 2017.
- [7] A. R. Mishra, *Fundamentals of Cellular Network Planning and Optimization: 2G/2.5G/3G... Evolution to 4G*. Hoboken, NJ, USA: Wiley, 2007.
- [8] R. Joyce and L. Zhang, "Self organising network techniques to maximise traffic offload onto a 3G/WCDMA small cell network using MDT UE measurement reports," in *Proc. IEEE Global Commun. Conf.*, Dec. 2014, pp. 2212–2217.
- [9] D. Fagen, P. A. Vicharelli, and J. Weitzen, "Automated wireless coverage optimization with controlled overlap," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2395–2403, Jul. 2008.
- [10] *Technical Specification Group TSG RAN Minimization of Drive Tests for E-UTRAN and UTRAN; Overall Description; Stage 2 (Release 10)*, document 3GPP TS 37.cde V0.3.0, 3rd Generation Partnership Project, 2010.
- [11] R. Razavi, S. Klein, and H. Claussen, "Self-optimization of capacity and coverage in LTE networks using a fuzzy reinforcement learning approach," in *Proc. 21st Annu. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2010, pp. 1865–1870.
- [12] I. Viering, M. Dötting, and A. Lobinger, "A mathematical perspective of self-optimizing wireless networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2009, pp. 1–6.
- [13] J. Rodriguez, I. De la Bandera, P. Munoz, and R. Barco, "Load balancing in a realistic urban scenario for LTE networks," in *Proc. IEEE 73rd Veh. Technol. Conf. (VTC Spring)*, May 2011, pp. 1–5.
- [14] L.-C. Wang, S.-H. Cheng, and A.-H. Tsai, "Bi-SON: Big-data self organizing network for energy efficient ultra-dense small cells," in *Proc. IEEE 84th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–5.
- [15] A. Lobinger, S. Stefanski, T. Jansen, and I. Balan, "Coordinating handover parameter optimization and load balancing in LTE self-optimizing networks," in *Proc. IEEE 73rd Veh. Technol. Conf. (VTC Spring)*, May 2011, pp. 1–5.
- [16] *3GPP Work Items on Self-Organizing Networks*, document v0.1.3, 2014.
- [17] A. Aguilar-Garcia, S. Fortes, A. Fernandez Duran, and R. Barco, "Context-aware self-optimization: Evolution based on the use case of load balancing in small-cell networks," *IEEE Veh. Technol. Mag.*, vol. 11, no. 1, pp. 86–95, Mar. 2016.
- [18] *Technical Specification Group Services and System Aspects; Study on Non-MTC Mobile Data Applications Impacts (Release 11)*, document 3GPP TR 22.801 V2.0.0, 3rd Generation Partnership Project, 2011.
- [19] C. A. S. Franco and J. R. B. de Marca, "Load balancing in self-organized heterogeneous LTE networks: A statistical learning approach," in *Proc. 7th IEEE Latin-Amer. Conf. Commun. (LATINCOM)*, Nov. 2015, pp. 1–5.
- [20] P. Munoz, R. Barco, J. M. Ruiz-Aviles, I. de la Bandera, and A. Aguilar, "Fuzzy rule-based reinforcement learning for load balancing techniques in enterprise LTE femtocells," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1962–1973, Jun. 2013.
- [21] S. S. Mwanje and A. Mitschele-Thiel, "A Q-Learning strategy for LTE mobility load balancing," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 2154–2158.
- [22] H. Claussen, L. T. W. Ho, and L. G. Samuel, "Self-optimization of coverage for femtocell deployments," in *Proc. Wireless Telecommun. Symp.*, Apr. 2008, pp. 278–285.
- [23] S. Bassoy, M. Jaber, M. A. Imran, and P. Xiao, "Load aware self-organising user-centric dynamic CoMP clustering for 5G networks," *IEEE Access*, vol. 4, pp. 2895–2906, 2016.
- [24] B. Keshavamurthy and M. Ashraf, "Conceptual design of proactive SONs based on the big data framework for 5G cellular networks: A novel machine learning perspective facilitating a shift in the SON paradigm," in *Proc. Int. Conf. Syst. Modeling Adv. Res. Trends (SMART)*, 2016, pp. 298–304.
- [25] *Technical Specification Group Services and System Aspects; Procedures for the 5G System; Stage 2 (Release 15)*, document 3GPP TS 23.502 V1.1.0, 3rd Generation Partnership Project, 2017.
- [26] *Technical Specification Group Services and System Aspects; Policy and Charging Control Framework for the 5G System; Stage 2 (Release 15)*, 3GPP TS 23.503 V15.1.0, 3rd Generation Partnership Project, 2018.
- [27] P. Blasco and D. Gündüz, "Learning-based optimization of cache content in a small cell base station," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 1897–1903.
- [28] V. Pejovic and M. Musolesi, "Anticipatory mobile computing: A survey of the state of the art and research challenges," *ACM Comput. Surv.*, vol. 47, no. 3, pp. 1–29, Apr. 2015.
- [29] T. Alsedairy, Y. Qi, A. Imran, M. A. Imran, and B. Evans, "Self organising cloud cells: A resource efficient network densification strategy," *Trans. Emerg. Telecommun. Technol.*, vol. 26, no. 8, pp. 1096–1107, Apr. 2014.
- [30] J. Moysen and L. Giupponi, "From 4G to 5G: Self-organized network management meets machine learning," *Comput. Commun.*, vol. 129, pp. 248–268, Sep. 2018.
- [31] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of machine learning techniques applied to self-organizing cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2392–2431, 4th Quart., 2017.
- [32] W. Guo and T. O'Farrell, "Dynamic cell expansion with self-organizing cooperation," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 851–860, May 2013.
- [33] W. Guo and J. Zhang, "Uncovering wireless blackspots using Twitter data," *Electron. Lett.*, vol. 53, no. 12, pp. 814–816, Jun. 2017.
- [34] X. Wang, L. T. Yang, H. Liu, and M. J. Deen, "A big data-as-a-service framework: State-of-the-art and perspectives," *IEEE Trans. Big Data*, vol. 4, no. 3, pp. 325–340, Sep. 2018.
- [35] R.-F. Liao, H. Wen, J. Wu, H. Song, F. Pan, and L. Dong, "The Rayleigh fading channel prediction via deep learning," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–11, Jul. 2018.
- [36] W. Zuo, H. Xie, Y. Lin, H. Hu, and Z. Cai, "A load balancing optimization algorithm for context-aware wireless sensor networks based on fuzzy neural networks," *Amer. J. Neural Netw. Appl.*, vol. 2, no. 2, pp. 6–16, 2016.
- [37] J. Wu, I. Bisio, C. Gniady, E. Hossain, M. Valla, and H. Li, "Context-aware networking and communications: Part 1 [Guest Editorial]," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 14–15, Jun. 2014.
- [38] J. Wu, I. Bisio, C. Gniady, E. Hossain, M. Valla, and H. Li, "Context-aware networking and communications: Part 2 [Guest Editorial]," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 64–65, Aug. 2014.
- [39] J. Krumm, N. Davies, and C. Narayanaswami, "User-generated content," *IEEE Pervasive Comput.*, vol. 7, no. 4, pp. 10–11, Oct. 2008.
- [40] X. Wang, L. T. Yang, X. Xie, J. Jin, and M. J. Deen, "A cloud-edge computing framework for cyber-physical-social services," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 80–85, Nov. 2017.
- [41] X. Wang, L. T. Yang, H. Li, M. Lin, J. Han, and B. O. Apduhan, "NQA: A nested anti-collision algorithm for RFID systems," *ACM Trans. Embedded Comput. Syst.*, vol. 18, no. 4, pp. 1–21, Jul. 2019.
- [42] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.
- [43] K. Zhu, W. Zhi, L. Zhang, X. Chen, and X. Fu, "Social-aware incentivized caching for D2D communications," *IEEE Access*, vol. 4, pp. 7585–7593, 2016.
- [44] X. Wang, L. T. Yang, J. Feng, X. Chen, and M. J. Deen, "A tensor-based big service framework for enhanced living environments," *IEEE Cloud Comput.*, vol. 3, no. 6, pp. 36–43, Nov. 2016.
- [45] D. Liu, B. Chen, C. Yang, and A. F. Molisch, "Caching at the wireless edge: Design aspects, challenges, and future directions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 22–28, Sep. 2016.
- [46] N. Gupta, H. Crosby, D. Purser, S. Jarvis, and W. Guo, "Twitter usage across industry: A spatiotemporal analysis," in *Proc. IEEE 5th Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Mar. 2018, pp. 64–71.
- [47] X. Wang, L. T. Yang, X. Chen, J.-J. Han, and J. Feng, "A tensor computation and optimization model for cyber-physical-social big data," *IEEE Trans. Sustain. Comput.*, vol. 4, no. 4, pp. 326–339, Oct. 2019.

- [48] X. Wang, L. T. Yang, L. Kuang, X. Liu, Q. Zhang, and M. J. Deen, "A tensor-based big-data-driven routing recommendation approach for heterogeneous networks," *IEEE Netw.*, vol. 33, no. 1, pp. 64–69, Jan. 2019.
- [49] L. Kuang, L. T. Yang, X. Wang, P. Wang, and Y. Zhao, "A tensor-based big data model for QoS improvement in software defined networks," *IEEE Netw.*, vol. 30, no. 1, pp. 30–35, Jan. 2016.
- [50] X. Wang, W. Wang, L. T. Yang, S. Liao, D. Yin, and M. J. Deen, "A distributed HOSVD method with its incremental computation for big data in cyber-physical-social systems," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 2, pp. 481–492, Jun. 2018.
- [51] Y. Li, Y. Zhang, K. Luo, T. Jiang, Z. Li, and W. Peng, "Ultra-dense HetNets meet big data: Green frameworks, techniques, and approaches," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 56–63, Jun. 2018.
- [52] G. W. Kim, Y. M. Kim, J. S. Han, Y. L. Rhie, and M. H. Yun, "Understanding characteristics of user-generated content as a source of extracting user value," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage. (IEEM)*, Dec. 2016, pp. 1751–1755.
- [53] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1790–1821, 3rd Quart., 2017.
- [54] G. Xu, S. Gao, M. Daneshmand, C. Wang, and Y. Liu, "A survey for mobility big data analytics for geolocation prediction," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 111–119, Feb. 2017.
- [55] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5G," *IEEE Netw.*, vol. 30, no. 1, pp. 44–51, Jan. 2016.
- [56] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans, "A survey of self organisation in future cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 336–361, 1st Quart., 2013.
- [57] Z. Du, Y. Sun, W. Guo, Y. Xu, Q. Wu, and J. Zhang, "Data-driven deployment and cooperative self-organization in ultra-dense small cell networks," *IEEE Access*, vol. 6, pp. 22839–22848, 2018.
- [58] K. Tamura and T. Ichimura, "Density-based spatiotemporal clustering algorithm for extracting bursty areas from georeferenced documents," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2013, pp. 2079–2084.
- [59] H. Becker, D. Iyer, M. Naaman, and L. Gravano, "Identifying content for planned events across social media sites," in *Proc. 5th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2012, pp. 533–542.
- [60] K. Yanai, K. Yaegashi, and B. Qiu, "Detecting cultural differences using consumer-generated geotagged photos," in *Proc. 2nd Int. Workshop Location Web (LOCWEB)*, New York, NY, USA, 2009, pp. 1–4.
- [61] D. Ashbrook and T. Starner, "Using GPS to learn significant locations and predict movement across multiple users," *Pers. Ubiquitous Comput.*, vol. 7, no. 5, pp. 275–286, Oct. 2003.
- [62] R. Lee and K. Sumiya, "Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection," in *Proc. the 2nd ACM SIGSPATIAL Int. Workshop Location Based Social Netw. (LBSN)*, 2010, p. 1.
- [63] K. Watanabe, M. Ochi, M. Okabe, and R. Onai, "Jasmine: A real-time local-event detection system based on geolocation information propagated to microblogs," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, New York, NY, USA, 2011, p. 2541.
- [64] J. Mahmud, J. Nichols, and C. Drews, "Home location identification of Twitter users," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 1–21, Jul. 2014.
- [65] J. Mahmud, J. Nichols, and C. Drews, "Where is this Tweet from? inferring home locations of Twitter users," in *Proc. ICWSM*, vol. 12, 2012, pp. 511–514.
- [66] M. Lv, L. Chen, and G. Chen, "Discovering personally semantic places from GPS trajectories," in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, New York, NY, USA, 2012, p. 1552.
- [67] Q. Lin, D. Zhang, K. Connelly, H. Ni, Z. Yu, and X. Zhou, "Disorientation detection by mining GPS trajectories for cognitively-impaired elders," *Pervasive Mobile Comput.*, vol. 19, pp. 71–85, May 2015.
- [68] J. H. Kang, W. Welbourne, B. Stewart, and G. Borriello, "Extracting places from traces of locations," in *Proc. 2nd ACM Int. Workshop Wireless Mobile Appl. Services WLAN Hotspots (WMASH)*, New York, NY, USA, 2004, p. 110.
- [69] H. Farooq and A. Imran, "Spatiotemporal mobility prediction in proactive self-organizing cellular networks," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 370–373, Feb. 2017.
- [70] A. T. Palma, V. Bogorny, B. Kuijpers, and L. O. Alvares, "A clustering-based approach for discovering interesting places in trajectories," in *Proc. ACM Symp. Appl. Comput. (SAC)*, New York, NY, USA, 2008, p. 863.
- [71] E. Tunggawan and Y. E. Soelistio, "And the winner is ...: Bayesian Twitter-based prediction on 2016 U.S. presidential election," in *Proc. Int. Conf. Comput., Control, Informat. Appl. (IC3INA)*, Oct. 2016, pp. 33–37.
- [72] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Commun. ACM*, vol. 53, no. 8, p. 80, Aug. 2010.
- [73] T. Tanaka, S. Ata, and M. Murata, "Analysis of popularity pattern of user generated contents and its application to content-aware networking," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2016, pp. 1–6.
- [74] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas. (IMC)*, New York, NY, USA, 2007, p. 1.
- [75] J. Du, H. Xu, and X. Huang, "Box office prediction based on microblog," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1680–1689, Mar. 2014.
- [76] M. Ahmed, S. Spagna, F. Huici, and S. Niccolini, "A peek into the future: Predicting the evolution of popularity in user generated content," in *Proc. 6th ACM Int. Conf. Web Search Data Mining (WSDM)*, New York, NY, USA, 2013, p. 607.
- [77] Z. Xu, L. Ru, L. Xiang, and Q. Yang, "Discovering user interest on Twitter with a modified author-topic model," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol.*, Aug. 2011, pp. 422–429.
- [78] M. Yan, J. Sang, and C. Xu, "Unified YouTube video recommendation via cross-network collaboration," in *Proc. 5th ACM Int. Conf. Multimedia Retr. (ICMR)*, New York, NY, USA, 2015, pp. 19–26.
- [79] D. Lu and Q. Li, "Personalized search on flickr based on searcher's preference prediction," in *Proc. 20th Int. Conf. Companion World Wide Web (WWW)*, New York, NY, USA, 2011, p. 81.
- [80] T. Tominaga and Y. Hijikata, "Exploring the relationship between user activities and profile images on Twitter through machine learning techniques," *J. Web Sci.*, vol. 5, no. 1, pp. 1–13, Mar. 2018.
- [81] M. De Gemmis, L. Iaquinta, P. Lops, C. Musto, F. Narducci, and G. Semeraro, "Preference learning in recommender systems," *Preference Learn.*, vol. 41, pp. 41–55, Sep. 2009.
- [82] E. Agichtein, E. Brill, S. Dumais, and R. Ragno, "Learning user interaction models for predicting Web search result preferences," in *Proc. 29th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, 2006, pp. 3–10.
- [83] Z. Yu, D. Zhang, X. Zhou, and C. Li, "User preference learning for multimedia personalization in pervasive computing environment," in *Proc. Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst.*, 2005, pp. 236–242.
- [84] A. M. Rashid, I. Albert, D. Cosley, S. K. Lam, S. M. McNee, J. A. Konstan, and J. Riedl, "Getting to know you: Learning new user preferences in recommender systems," in *Proc. 7th Int. Conf. Intell. User Inter. (IUI)*, 2002, pp. 127–134.
- [85] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao, "User interactions in social networks and their implications," in *Proc. 4th ACM Eur. Conf. Comput. Syst. (EuroSys)*, New York, NY, USA, 2009, p. 205.
- [86] J. Jiang, C. Wilson, X. Wang, W. Sha, P. Huang, Y. Dai, and B. Y. Zhao, "Understanding latent interactions in online social networks," in *Proc. 10th Annu. Conf. Internet Meas. (IMC)*, New York, NY, USA, 2010, p. 369.
- [87] A. Sathiseelan, M. S. Seddiki, S. Stoyanov, and D. Trossen, "Social SDN: Online social networks integration in wireless network provisioning," in *Proc. ACM Conf. (SIGCOMM)*, New York, NY, USA, 2014, vol. 44, no. 4, pp. 375–376.
- [88] L. Backstrom, E. Sun, and C. Marlow, "Find me if you can: Improving geographical prediction with social and spatial proximity," in *Proc. 19th Int. Conf. World Wide Web (WWW)*, New York, NY, USA, 2010, p. 61.
- [89] J. N. Cummings, B. Butler, and R. Kraut, "The quality of online social relationships," *Commun. ACM*, vol. 45, no. 7, pp. 103–108, Jul. 2002.
- [90] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2011, p. 1082.
- [91] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *Proc. 19th Int. Conf. World Wide Web (WWW)*, New York, NY, USA, 2010, p. 591.
- [92] C. Wu, X. Chen, Y. Zhou, N. Li, X. Fu, and Y. Zhang, "Spice: Socially-driven learning-based mobile media prefetching," in *Proc. IEEE INFOCOM 35th Annu. IEEE Int. Conf. Comput. Commun.*, Apr. 2016, pp. 1–9.

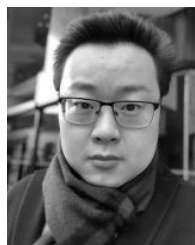
- [93] J. Leskovec, "Large-scale graph representation learning," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, p. 4.
- [94] H. Klessig, H. Kuntzschmann, L. Scheuven, B. Almeroth, P. Schulz, and G. Fettweis, "Twitter as a source for spatial traffic information in big data-enabled self-organizing networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–5.
- [95] B. Yang, W. Guo, B. Chen, G. Yang, and J. Zhang, "Estimating mobile traffic demand using Twitter," *IEEE Wireless Commun. Lett.*, vol. 5, no. 4, pp. 380–383, Aug. 2016.
- [96] F. Botta, H. S. Moat, and T. Preis, "Quantifying crowd size with mobile phone and Twitter data," *Roy. Soc. Open Sci.*, vol. 2, no. 5, May 2015, Art. no. 150162.
- [97] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," 2017, *arXiv:1710.02913*. [Online]. Available: <http://arxiv.org/abs/1710.02913>
- [98] R. Xie, Y. Chen, Q. Xie, Y. Xiao, and X. Wang, "We know your preferences in new cities: Mining and modeling the behavior of travelers," *IEEE Commun. Mag.*, vol. 56, no. 11, pp. 28–35, Nov. 2018.
- [99] F. Zeng, R. Wang, and J. Wu, "How mobile contributors will interact with each other in mobile crowdsourcing with word of mouth mode," *IEEE Access*, vol. 7, pp. 14523–14536, 2019.
- [100] A. Tsakalidis, M. Liakata, T. Damoulas, B. Jellinek, W. Guo, and A. I. Cristea, *Combining Heterogeneous User Generated Data to Sense Well-Being*. Osaka, Japan: The COLING 2016 Organizing Committee, 2016.
- [101] *Datasets on Kaggle*. Accessed: Feb. 19, 2020. [Online]. Available: <https://www.kaggle.com/datasets>
- [102] *Open Big Data Italia Cellular Traffic and Tweets*. Accessed: Feb. 19, 2020. [Online]. Available: <https://dandelion.eu/datamine/open-big-data/>
- [103] C. Zhang, X. Ouyang, and P. Patras, "ZipNet-GAN: Inferring fine-grained mobile traffic patterns via a generative adversarial neural network," in *Proc. 13th Int. Conf. Emerg. Netw. Exp. Technol. (CoNEXT)*, 2017, pp. 363–375.
- [104] G. Khodabandelou, V. Gauthier, M. Fiore, and M. A. El-Yacoubi, "Estimation of static and dynamic urban populations with mobile network metadata," *IEEE Trans. Mobile Comput.*, vol. 18, no. 9, pp. 2034–2047, Sep. 2019.
- [105] B. Yang, W. Guo, B. Chen, G. Yang, and J. Zhang, *Data From: Estimating Mobile Traffic Demand Using Twitter*. 2016. [Online]. Available: <https://doi.org/10.5061/dryad.35m1f4q>
- [106] *Comparing Different Clustering Algorithms on Toy Datasets—Scikit-Learn 0.20.3 Documentation*. Accessed: Feb. 19, 2020. [Online]. Available: https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html
- [107] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
- [108] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1153–1176, 2nd Quart., 2016.
- [109] *Comparing Different Clustering Algorithms on Toy Datasets—Scikit-Learn 0.20.3 Documentation*. [Online]. Available: https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html
- [110] J. Hensman, N. Fusi, and N. D. Lawrence, "Gaussian processes for big data," 2013, *arXiv:1309.6835*. [Online]. Available: <http://arxiv.org/abs/1309.6835>
- [111] A. Damianou and N. Lawrence, "Deep Gaussian processes," in *Proc. 16th Int. Conf. Artif. Intell. Statist.* Proceedings of Machine Learning Research, vol. 31. Scottsdale, AZ, USA: PMLR, 2013, pp. 207–215.
- [112] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," 2014, *arXiv:1402.1128*. [Online]. Available: <http://arxiv.org/abs/1402.1128>
- [113] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in *Proc. 15th Annu. Conf. Int. Speech Commun. Assoc.*, 2014, pp. 338–342.
- [114] L. Nechepurenko, V. Voss, and V. Gritsenko, "Comparing knowledge-based reinforcement learning to neural networks in a strategy game," 2019, *arXiv:1901.04626*. [Online]. Available: <http://arxiv.org/abs/1901.04626>
- [115] S. Koenig and R. G. Simmons, "Complexity analysis of real-time reinforcement learning," in *Proc. AAAI*, 1993, pp. 99–107.
- [116] M. Chen, W. Saad, C. Yin, and M. Debbah, "Echo state networks for proactive caching in cloud-based radio access networks with mobile users," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3520–3535, Jun. 2017.
- [117] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5353–5360.
- [118] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [119] N. Jankowski, "Complexity measures for meta-learning and their optimality," in *Algorithmic Probability and Friends. Bayesian Prediction and Artificial Intelligence*. Berlin, Germany: Springer, 2013, pp. 198–210.
- [120] J. He and X. Yao, "Drift analysis and average time complexity of evolutionary algorithms," *Artif. Intell.*, vol. 127, no. 1, pp. 57–85, Mar. 2001.
- [121] P. Hennig, M. A. Osborne, and M. Girolami, "Probabilistic numerics and uncertainty in computations," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 471, no. 2179, Jul. 2015, Art. no. 20150142.
- [122] K. B. Misra, "Reliability optimization of a series-parallel system," *IEEE Trans. Rel.*, vol. R-21, no. 4, pp. 230–238, Nov. 1972.
- [123] L. M. Kaufman, "Data security in the world of cloud computing," *IEEE Secur. Privacy Mag.*, vol. 7, no. 4, pp. 61–64, Jul. 2009.
- [124] R. Chow, "The last mile for IoT privacy," *IEEE Secur. Privacy*, vol. 15, no. 6, pp. 73–76, Nov. 2017.
- [125] *Cookies Consent Under the GDPR—EU GDPR Compliant*. Accessed: Feb. 19, 2020. [Online]. Available: <https://eugdprcompliant.com/cookies-consent-gdpr/>
- [126] D. Maimut and R. Reyhanitabar, "Authenticated encryption: Toward next-generation algorithms," *IEEE Secur. Privacy*, vol. 12, no. 2, pp. 70–72, Mar. 2014.
- [127] S. Chen, H. Wen, J. Wu, J. Chen, W. Liu, L. Hu, and Y. Chen, "Physical-layer channel authentication for 5G via machine learning algorithm," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–10, Oct. 2018.
- [128] K. Hamedani, L. Liu, R. Atat, J. Wu, and Y. Yi, "Reservoir computing meets smart grids: Attack detection using delayed feedback networks," *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 734–743, Feb. 2018.
- [129] R. Atat, L. Liu, J. Wu, G. Li, C. Ye, and Y. Yang, "Big data meet cyber-physical systems: A panoramic survey," *IEEE Access*, vol. 6, pp. 73603–73636, 2018.
- [130] Z. He, Z. Cai, and J. Yu, "Latent-data privacy preserving with customized data utility for social network data," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 665–673, Jan. 2018.
- [131] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proc. 41st Annu. ACM Symp. Theory Comput. (STOC)*, vol. 9, 2009, pp. 169–178.
- [132] G. Chen, S. Hoteit, A. C. Viana, M. Fiore, and C. Sarraute, "Spatio-temporal predictability of cellular data traffic," INRIA Saclay-Ile-de-France, Res. Rep. RT-0483, Jan. 2017, p. 17. [Online]. Available: <https://hal.inria.fr/hal-01393361> and https://hal.inria.fr/hal-01393361/file/mobihoc_INRIA_RT_hal.pdf
- [133] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, "Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1046–1061, May 2017.
- [134] B. Ma, B. Yang, Y. Zhu, and J. Zhang, "Context-aware proactive 5G load balancing and optimization for urban areas," *IEEE Access*, vol. 8, pp. 8405–8417, 2020.
- [135] S. Kisilevich, F. Mansmann, and D. Keim, "P-DBSCAN: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos," in *Proc. 1st Int. Conf. Exhib. Comput. Geospatial Res. Appl. (COM.Geo)*, 2010, p. 274.
- [136] J. An, K. Yang, J. Wu, N. Ye, S. Guo, and Z. Liao, "Achieving sustainable ultra-dense heterogeneous networks for 5G," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 84–90, Dec. 2017.
- [137] G. Ding, Z. Tan, J. Wu, and J. Zhang, "Efficient indoor fingerprinting localization technique using regional propagation model," *IEICE Trans. Commun.*, vol. E97-B, no. 8, pp. 1728–1741, 2014.
- [138] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vaganot, "A conceptual view on trajectories," *Data Knowl. Eng.*, vol. 65, no. 1, pp. 126–146, Apr. 2008.
- [139] G. Ding, Z. Tan, J. Wu, J. Zeng, and L. Zhang, "Indoor fingerprinting localization and tracking system using particle swarm optimization and Kalman filter," *IEICE Trans. Commun.*, vol. E98-B, no. 3, pp. 502–514, 2015.

- [140] I. Ajzen and M. Fishbein, *Understanding Attitudes and Predicting Social Behavior*. 1980, vol. 278.
- [141] A. Tatar, M. D. de Amorim, S. Fdida, and P. Antoniadis, "A survey on predicting the popularity of Web content," *J. Internet Services Appl.*, vol. 5, no. 1, p. 8, Aug. 2014.
- [142] K. N. Doan, T. Van Nguyen, T. Q. S. Quek, and H. Shin, "Content-aware proactive caching for backhaul offloading in cellular network," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3128–3140, May 2018.
- [143] X. Deng, Y. Jiang, L. T. Yang, M. Lin, L. Yi, and M. Wang, "Data fusion based coverage optimization in heterogeneous sensor networks: A survey," *Inf. Fusion*, vol. 52, pp. 90–105, Dec. 2019.
- [144] X. Deng, M. Xu, L. T. Yang, M. Lin, L. Yi, and M. Wang, "Energy balanced dispatch of mobile edge nodes for confident information coverage hole repairing in IoT," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4782–4790, Jun. 2019.
- [145] X. Deng, L. T. Yang, L. Yi, M. Wang, and Z. Zhu, "Detecting confident information coverage holes in industrial Internet of Things: An energy-efficient perspective," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 68–73, Sep. 2018.
- [146] A. Tsakalidis, M. Liakata, T. Damoulas, B. Jellinek, W. Guo, and A. Cristea, "Combining heterogeneous user generated data to sense well-being," in *Proc. Int. Conf. Comput. Linguistics (COLING)*, 2016, pp. 3007–3018.
- [147] L. Jin, Y. Chen, T. Wang, P. Hui, and A. V. Vasilakos, "Understanding user behavior in online social networks: A survey," *IEEE Commun. Mag.*, vol. 51, no. 9, pp. 144–150, Sep. 2013.
- [148] S. Ioannidis, A. Chaintreau, and L. Massoulié, "Distributing content updates over a mobile social network," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 13, no. 1, p. 44, Jun. 2009.
- [149] Y. Yi, Z. Zhang, and C. Gan, "The effect of social tie on information diffusion in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 509, pp. 783–794, Nov. 2018.
- [150] Y. Yi, Z. Zhang, and C. Gan, "The outbreak threshold of information diffusion over social-physical networks," *Phys. A, Stat. Mech. Appl.*, vol. 526, Jul. 2019, Art. no. 121128.
- [151] Y. Wu, S. Wang, W. Liu, W. Guo, and X. Chu, "Iunius: A cross-layer peer-to-peer system with device-to-device communications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 7005–7017, Oct. 2016.
- [152] C. E. Rasmussen, "Gaussian processes in machine learning," in *Advanced Lectures on Machine Learning*. Berlin, Germany: Springer, 2004, pp. 63–71.
- [153] K. R. Pandhare and M. A. Shah, "Real time road traffic event detection using Twitter and spark," in *Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, Mar. 2017, pp. 445–449.
- [154] S. Narejo and E. Pasero, "An application of internet traffic prediction with deep neural network," in *Multidisciplinary Approaches to Neural Computing* (Smart Innovation, Systems and Technologies), vol. 69, A. Esposito, M. Faudez-Zanuy, F. Morabito, and E. Pasero, Eds. Cham, Switzerland: Springer, 2018.
- [155] C. Qiu, Y. Zhang, Z. Feng, P. Zhang, and S. Cui, "Spatio-temporal wireless traffic prediction with recurrent neural network," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 554–557, Aug. 2018.
- [156] Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu, and H. Zhang, "Traffic prediction based on random connectivity in deep learning with long short-term memory," 2017, *arXiv:1711.02833*. [Online]. Available: <http://arxiv.org/abs/1711.02833>
- [157] M. Xu, Q. Wang, and Q. Lin, "Hybrid holiday traffic predictions in cellular networks," in *Proc. NOMS IEEE/IFIP Netw. Oper. Manage. Symp.*, Apr. 2018, pp. 1–6.
- [158] D. Tikunov and T. Nishimura, "Traffic prediction for mobile network using Holt-Winter's exponential smoothing," in *Proc. 15th Int. Conf. Softw., Telecommun. Comput. Netw.*, 2007, pp. 1–5.
- [159] F. Xu, Y. Lin, J. Huang, D. Wu, H. Shi, J. Song, and Y. Li, "Big data driven mobile traffic understanding and forecasting: A time series approach," *IEEE Trans. Services Comput.*, vol. 9, no. 5, pp. 796–805, Sep. 2016.
- [160] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 882–890.
- [161] R. Li, Z. Zhao, J. Zheng, C. Mei, Y. Cai, and H. Zhang, "The learning and prediction of application-level traffic data in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3899–3912, Jun. 2017.
- [162] S. Zhang, S. Zhao, M. Yuan, J. Zeng, J. Yao, M. R. Lyu, and I. King, "Traffic prediction based power saving in cellular networks: A machine learning method," in *Proc. 25th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2017, p. 29.
- [163] J. Riihijarvi and P. Mahonen, "Machine learning for performance prediction in mobile cellular networks," *IEEE Comput. Intell. Mag.*, vol. 13, no. 1, pp. 51–60, Feb. 2018.
- [164] M. S. Parwez, D. B. Rawat, and M. Garuba, "Big data analytics for user-activity analysis and user-anomaly detection in mobile wireless network," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 2058–2065, Aug. 2017.
- [165] L.-V. Le, D. Sinh, B.-S.-P. Lin, and L.-P. Tung, "Applying big data, machine learning, and SDN/NFV to 5G traffic clustering, forecasting, and management," in *Proc. 4th IEEE Conf. Netw. Softwarization Workshops (NetSoft)*, Jun. 2018, pp. 168–176.
- [166] H. Owhadi, C. Scovel, and T. Sullivan, "On the brittleness of Bayesian inference," *SIAM Rev.*, vol. 57, no. 4, pp. 566–582, Jan. 2015.
- [167] L.-V. Le, D. Sinh, L.-P. Tung, and B.-S.-P. Lin, "A practical model for traffic forecasting based on big data, machine-learning, and network KPIs," in *Proc. 15th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2018, pp. 1–4.
- [168] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, "Characterizing and modeling Internet traffic dynamics of cellular devices," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 39, no. 1, pp. 265–276, 2011.
- [169] A. Lobinger, S. Stefanski, T. Jansen, and I. Balan, "Load balancing in downlink LTE self-optimizing networks," in *Proc. IEEE 71st Veh. Technol. Conf.*, May 2010, pp. 1–5.
- [170] H. Hu, J. Zhang, X. Zheng, Y. Yang, and P. Wu, "Self-configuration and self-optimization for LTE networks," *IEEE Commun. Mag.*, vol. 48, no. 2, pp. 94–100, Feb. 2010.
- [171] J. M. Ruiz-Avilés, S. Luna-Ramírez, M. Toril, and F. Ruiz, "Traffic steering by self-tuning controllers in enterprise LTE femtocells," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, p. 337, Nov. 2012.
- [172] J. Xu, L. Tang, Q. Chen, and L. Yi, "Study on based reinforcement Q-Learning for mobile load balancing techniques in LTE—A HetNets," in *Proc. IEEE 17th Int. Conf. Comput. Sci. Eng.*, Dec. 2014, pp. 1766–1771.
- [173] S. Tomforde, A. Ostrovsky, and J. Hähner, "Load-aware reconfiguration of LTE-antennas—dynamic cell-phone network adaptation using organic network control," in *Proc. 11th Int. Conf. Informat. Control, Autom. Robot.*, vol. 1, 2014, pp. 236–243.
- [174] T. Kudo and T. Ohtsuki, "Q-learning based cell selection for UE outage reduction in heterogeneous networks," in *Proc. IEEE 80th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2014, pp. 1–5.
- [175] N. Ristanovic, J.-Y. Le Boudec, A. Chaintreau, and V. Erramilli, "Energy efficient offloading of 3G networks," in *Proc. IEEE 8th Int. Conf. Mobile Ad-Hoc Sensor Syst.*, Oct. 2011, pp. 202–211.
- [176] R. Atat, L. Liu, H. Chen, J. Wu, H. Li, and Y. Yi, "Enabling cyber-physical communication in 5G cellular networks: Challenges, spatial spectrum sensing, and cyber-security," *IET Cyber-Phys. Syst., Theory Appl.*, vol. 2, no. 1, pp. 49–54, Apr. 2017.
- [177] R. Atat, L. Liu, J. Wu, J. Ashdown, and Y. Yi, "Green massive traffic offloading for cyber-physical systems over heterogeneous cellular networks," *Mobile Netw. Appl.*, vol. 24, no. 4, pp. 1364–1372, Feb. 2018.
- [178] X. Chen, J. Wu, Y. Cai, H. Zhang, and T. Chen, "Energy-efficiency oriented traffic offloading in wireless networks: A brief survey and a learning approach for heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 4, pp. 627–640, Apr. 2015.
- [179] H.-R. Cheon and J.-H. Kim, "Social context-aware mobile data offloading algorithm via small cell backhaul networks," *IEEE Access*, vol. 7, pp. 39030–39040, 2019.
- [180] S. Wang and W. Guo, "Energy and cost implications of a traffic aware and quality-of-service constrained sleep mode mechanism," *IET Commun.*, vol. 7, no. 18, pp. 2092–2101, Dec. 2013.
- [181] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Multiobjective auction-based switching-off scheme in heterogeneous networks: To bid or not to bid?" *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9168–9180, Nov. 2016.
- [182] F. Han, S. Zhao, L. Zhang, and J. Wu, "Survey of strategies for switching off base stations in heterogeneous networks for greener 5G systems," *IEEE Access*, vol. 4, pp. 4959–4973, 2016.
- [183] K. Wang, Y. Wang, Y. Sun, S. Guo, and J. Wu, "Green industrial Internet of Things architecture: An energy-efficient perspective," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 48–54, Dec. 2016.
- [184] A. Bousia, E. Kartsakli, L. Alonso, and C. Verikoukis, "Dynamic energy efficient distance-aware base station switch on/off scheme for LTE-Advanced," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2012, pp. 1532–1537.
- [185] A. Bousia, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Green distance-aware base station sleeping algorithm in LTE-Advanced," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 1347–1351.

- [186] P. Munoz, R. Barco, I. de la Bandera, M. Toril, and S. Luna-Ramirez, "Optimization of a fuzzy logic controller for handover-based load balancing," in *Proc. IEEE 73rd Veh. Technol. Conf. (VTC Spring)*, May 2011, pp. 1–5.
- [187] A. A. Atayero, M. K. Luka, and A. A. Alatishe, "Neural-encoded fuzzy models for load balancing in 3GPP LTE," *Int. J. Appl. Inf. Syst.*, vol. 4, no. 1, pp. 34–40, Sep. 2012.
- [188] K. Suksomboon, S. Tarnoi, Y. Ji, M. Koibuchi, K. Fukuda, S. Abe, N. Motonori, M. Aoki, S. Urushidani, and S. Yamada, "PopCache: Cache more or less based on content popularity for information-centric networking," in *Proc. 38th Annu. IEEE Conf. Local Comput. Netw.*, Oct. 2013, pp. 236–243.
- [189] M. S. ElBamby, M. Bennis, W. Saad, and M. Latva-Aho, "Content-aware user clustering and caching in wireless small cell networks," in *Proc. 11th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2014, pp. 945–949.
- [190] E. Bastug, J.-L. Guenego, and M. Debbah, "Proactive small cell networks," in *Proc. ICT*, May 2013, pp. 1–5.
- [191] S. Guo, H. Xie, and G. Shi, "Collaborative forwarding and caching in content centric networks," in *Proc. Int. Conf. Res. Netw.* Berlin, Germany: Springer, 2012, pp. 41–55.
- [192] K. Hamidouche, W. Saad, and M. Debbah, "Many-to-many matching games for proactive social-caching in wireless small cell networks," in *Proc. 12th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, May 2014, pp. 569–574.
- [193] Z. Zhang, Y. Yang, M. Hua, C. Li, Y. Huang, and L. Yang, "Proactive caching for vehicular multi-view 3D video streaming via deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2693–2706, May 2019.
- [194] M. Garetto, E. Leonardi, and S. Traverso, "Efficient analysis of caching strategies under dynamic content popularity," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2015, pp. 2263–2271.
- [195] R. Wang, X. Peng, J. Zhang, and K. B. Letaief, "Mobility-aware caching for content-centric wireless networks: Modeling and methodology," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 77–83, Aug. 2016.
- [196] E. Bastug, M. Bennis, and M. Debbah, "A transfer learning approach for cache-enabled wireless networks," in *Proc. 13th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, May 2015.
- [197] L. Madlberger and A. Almansour, "Predictions based on Twitter—A critical view on the research process," in *Proc. Int. Conf. Data Softw. Eng. (ICODSE)*, Nov. 2014, pp. 1–6.
- [198] Q. Gao, F. Abel, G.-J. Houben, and Y. Yu, "A comparative study of users' microblogging behavior on Sina Weibo and Twitter," in *Proc. Int. Conf. User Modeling, Adaptation, Personalization*, 2012, pp. 88–101.
- [199] TechNode. *Five Years on, China Finally Issues Licenses to First Batch of 15 Virtual Telecom Operators*. Accessed: Feb. 19, 2020. [Online]. Available: <https://technode.com/2018/07/23/china-licenses-virtual-telecom-operators/>
- [200] A. Zoha, A. Saeed, H. Farooq, A. Rizwan, A. Imran, and M. A. Imran, "Leveraging intelligence from network CDR data for interference aware energy consumption minimization," *IEEE Trans. Mobile Comput.*, vol. 17, no. 7, pp. 1569–1582, Jul. 2018.
- [201] A. Adeel, H. Larijani, A. Javed, and A. Ahmadinia, "Critical analysis of learning algorithms in random neural network based cognitive engine for LTE systems," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.



BO MA received the bachelor's degree in automatic control engineering from Southeast University, China, in 2014, and the M.Sc. degree (Hons.) in communication and signal processing from Newcastle University, U.K., in 2016. He is currently pursuing the Ph.D. degree in wireless systems with the Department of Electronic and Electrical Engineering, The University of Sheffield, U.K. His main research interests include proactive network optimization, the application of data analysis, and machine learning algorithms in 5G networks.



WEISI GUO (Senior Member, IEEE) received the M.Eng., M.A., and Ph.D. degrees from the University of Cambridge. He is currently a Professor with The University of Warwick and a Turing Fellow with The Alan Turing Institute, U.K. In Fall 2019, he joined the Centre for Autonomous and Cyber-Physical Systems, Cranfield University. He is also the Head of the Data-Embedded-Networks Laboratory and has been PI on over \$2m of funding and an Investigator on over \$6m. He has published over 110 peer-reviewed articles. He specializes in network science, wireless networks, machine learning, data science, and molecular communications. He was a recipient of the IET Innovation Award. He is also an Editor of two IEEE journals.



JIE ZHANG has been the Chair of wireless systems with the Department of Electronic and Electrical Engineering, The University of Sheffield, since January 2011. He is currently the Founder, the Board Chairman, and the Chief Scientific Officer (CSO) of Ranplan Wireless, a public company listed on Nasdaq OMX. Ranplan Wireless produces a suite of world leading in-building DAS, indoor-outdoor small cell/HetNet network design, and optimisation tools Ranplan Professional, Tablet and Collaboration-Hub, which are being used by the world's largest mobile operators and network vendors across the globe. Along with his students and colleagues, he has pioneered research in small cell and heterogeneous network (HetNet) and has published some of the landmark articles and book on these topics, widely used by academia and industry. His Google scholar citations are in excess of 7000, with an H-index of 35. In his current roles as CSO at Ranplan and the Chair Professor of The University of Sheffield, he leads the team of over 30 researchers at Ph.D. candidate level or above. The team has 12 on-going research projects funded by the European Commission and Innovate U.K., on data-driven proactive network optimization, millimeter wave small cell communications in the built environments, and modeling, and design of smart environments. Prior to his current appointments, he studied and worked at the Imperial College London, Oxford University, the University of Bedfordshire, and the East China University of Science and Technology, reaching the rank of a Lecturer, the Reader, and a Professor, in 2002, 2005, and 2006, respectively.

...