Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 8738613, 19 pages
https://doi.org/10.1155/2018/8738613

Review Article

A Survey on Machine Learning-Based Mobile Big Data Analysis: Challenges and Applications

Jiyang Xie,[1] Zeyu Song,[1] Yupeng Li,[2] Yanting Zhang,[3] Hong Yu,[1] Jinnan Zhan,[2] Zhanyu Ma,[1] Yuanyuan Qiao,[3] Jianhua Zhang,[2] and Jun Guo[1]

[1] *Pattern Recognition and Intelligent Systems Lab., Beijing University of Posts and Telecommunications, Beijing, China*
[2] *State Key Lab. of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China*
[3] *Center for Data Science, Beijing University of Posts and Telecommunications, Beijing, China*

Correspondence should be addressed to Zhanyu Ma; mazhanyu@bupt.edu.cn and Jianhua Zhang; jhzhang@bupt.edu.cn

Received 9 April 2018; Accepted 7 June 2018; Published 1 August 2018

Academic Editor: Liu Liu

This paper attempts to identify the requirement and the development of machine learning-based mobile big data (MBD) analysis through discussing the insights of challenges in the mobile big data. Furthermore, it reviews the state-of-the-art applications of data analysis in the area of MBD. Firstly, we introduce the development of MBD. Secondly, the frequently applied data analysis methods are reviewed. Three typical applications of MBD analysis, namely, wireless channel modeling, human online and offline behavior analysis, and speech recognition in the Internet of Vehicles, are introduced, respectively. Finally, we summarize the main challenges and future development directions of mobile big data analysis.

1. Introduction

With the success of wireless local access network (WLAN) technology (a.k.a. Wi-Fi) and the second/third/fourth generation (2G/3G/4G) mobile network, the number of mobile phones, which is 7.74 billion, 103.5 per 100 inhabitants all over the world in 2017, is rising dramatically [1]. Nowadays, mobile phone can not only send voice and text messages, but also easily and conveniently access the Internet which has been recognized as the most revolutionary development of mobile Internet (M-Internet). Meanwhile, worldwide active mobile-broadband subscriptions in 2017 have increased to 4.22 billion, which is 9.21% higher than that in 2016 [1]. Figure 1 shows the numbers of mobile-cellular telephone and active mobile-broadband subscriptions of the world and main districts from 2010 to 2017. The numbers which are up to the bars are the mobile-cellular telephone or active mobile-broadband subscriptions (million) in the world of the year which increase each year. Under the M-Internet, various kinds of content (image, voice, video, etc.) can be sent and received everywhere and the related applications emerge to satisfy people's requirements, including working, study, daily life, entertainment, education, and healthcare. In China, mobile applications giants, *i.e.*, Baidu, Alibaba, and Tencent, held 78% of M-Internet online time per day in apps which was about 2,412 minutes in 2017 [2]. This figure indicates that M-Internet has entered a rapid growth stage.

Nowadays, more than 1 billion smartphones are in use and producing a great quantity of data every day. This situation brings far-reaching impacts on society and social interaction and increases great opportunities for business. Meanwhile, with the rapid development of the Internet-of-Things (IoT), much more data is automatically generated by millions of machine nodes with growing mobility, for example, sensors carried by moving objects or vehicles. The volume, velocity, and variety of these data are increasing extremely fast, and soon they will become the new criterion for data analytics of enterprises and researchers. Therefore, mobile big data (MBD) has been already in our lives and is being enriched rapidly. The trend for explosively increased data volume with the increasing bandwidth and data rate in the M-Internet has followed the same exponential increase as Moore's Law for semiconductors [3]. The prediction [2] about the global data volume will grow up to 47 zettabytes

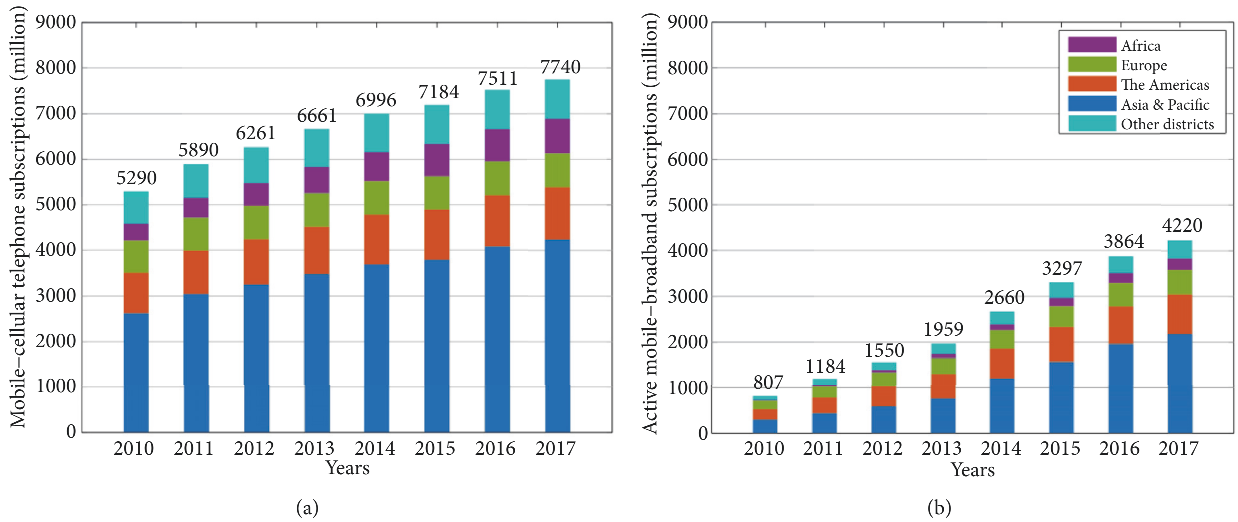


Figure 1: Mobile-cellular telephone subscriptions (million) in (a) and active mobile-broadband subscriptions (million) in (b) of the world and main districts [1].

(1 zettabyte = 1×10^{21} bytes) by 2020 and 163 zettabytes by 2025. For M-Internet, 3.7 exabytes (1 exabyte = 1×10^{18} bytes) data have been generated per month from the mobile data traffic in 2015 [4], 7.2 exabytes in 2016 [5], 24 exabytes by 2019 on forecasting [5], and 49 exabytes by 2021 on forecasting [5]. According to the statistical and prediction results, a concept called MBD has appeared.

The MBD can be considered as a huge quantity of mobile data which are generated from a massive number of mobile devices and cannot be processed and analyzed by a single machine [6, 7]. MBD is playing and will play a more important role than ever before by the popularization of mobile devices including smartphones and IoT gadgets especially in the era of 4G and the forthcoming the fifth generation (5G) [4, 8].

With the rapid development of information technologies, various data generated from different technical fields are showing explosive growth trends [9]. Big data has broad application prospects in many fields and has become important national strategic resources [10]. In the era of big data, many data analysis systems are facing big challenges as the volume of data increases. Therefore, analysis for MBD is currently a highly focused topic. The importance of MBD analysis is determined by its role in developing complex mobile systems which supports a variety of intelligently interactive services, for example, healthcare, intelligent energy networks, smart buildings, and online entertainments [4]. MBD analysis can be defined as mining terabyte-level or petabyte-level data collected from mobile users and wireless devices at the network-level or the app-level to discover unknown, latent, and meaningful patterns and knowledge with large-scale machine learning methods [11].

Present requirements of MBD are based on software-defined in order to be more scalable and flexible. M-Internet environment in the future will be even more complex and interconnected [12]. For this purpose, data centers of MBD need to collect user statistics information of millions of users and obtain meaningful results by proper MBD analysis methods. For the decreasing price of data storage and widely accessible high performance computers, an expansion of machine learning has come into not only theoretical researches, but also various application areas of big data. Even though, there is a long way to go for the machine learning-based MBD analysis.

Machine learning technology has been used by many Internet companies in their services: from web searches [13, 14] to content filtering [15] and recommendation [16, 17] on online social communities, shopping websites, or contend distribution platforms. Furthermore, it is also frequently appearing in products like smart cellphones, laptop computers, and smart furniture. Machine learning systems are used to detect and classify objects, return most relevant searching results, understand voice commands, and analyze using habits. In recent years, big data machine learning has become a hot spot [18]. Some conventional machine learning methods based on Bayesian framework [19–22], distributed optimization [23–26], and matrix factorization [27] can be applied into the aforementioned applications and have obtained good performances in small data sets. On this foundation, researchers have always been trying to fill their machine learning model with more and more data [28]. Furthermore, the data we got is not only big but also has features such as multisource, dynamic and sparse value; these features make it harder to analyze MBD with conventional machine learning methods. Therefore, the aforementioned applications implemented with conventional machine learning methods have fallen in a bottleneck period for low accuracy and generalization. Recently, a class of novel techniques, called deep learning, is applied in order to make the effort to solve the problems and has obtained good performances [29]. Machine learning, especially deep learning, has been an essential technique in order to use big data effectively.

Most conventional machine learning methods are shallow learning structures with one or none hidden layers.

These methods performed well in practical use and were precisely analyzed theoretically. But when dealing with high-dimensional or complicated data, shallow machine learning methods show their weakness. Deep learning methods are developed to learn better representations automatically with deep structure by using supervised or unsupervised strategies [30, 31]. The features extracted by deep hidden layers are used for regression, classification, or visualization. Deep learning uses more hidden layers and parameters to fit functions which could extract high level features from complex data; the parameters will be set automatically using large amount of unsupervised data [32, 33]. The hidden layers of deep learning algorithms help the model learn better representation of data; the higher layers learn specific and abstract features from global features learned by lower layers. Many surveys show that nonlinear feature extractors that are linked up as stacks such as deep learning methods always perform better in machine learning tasks, for example, a more accurate classification method [34], better learning of data probabilistic models [35], and the extraction of robust features [36]. Deep learning methods have proved useful in data mining, natural language processing, and computer vison applications. A more detailed introduction of deep learning is presented in Section 3.1.4.

Artificial Intelligence (AI) is a technology that develops theories, methods, techniques, and applications that simulate or extend human brain abilities. The research of observing, learning, and decision-making process in human brain motivates the development of deep learning, which was first designed aiming to emulate the human brain's neural structures. Further observation on neural signals processing and the effect on brain mechanisms [37–39] inspired the architecture design of deep learning network, using layers and neuron connections to generalize globally. Conventional methods such as support vector machines, decision trees, and case-based reasoning which are based on statistics or logic knowledge of human may fall short when facing complex structure or relationships of data. Deep learning methods can learn patterns and relationships from hidden layers and may benefit the signal processing study in human brain with visualization methods of neural network. Deep learning has attracted much attention from AI researchers recently because of its state-of-the-art performance in machine learning domains including no only the aforementioned natural language processing (NLP), but also speech recognition [40, 41], collaborative filtering [42], and computer vision [43, 44].

Deep learning has been successfully used in industry products which have access to big data from users. Companies in United States such as Google, Apple, Facebook, and Chinese companies like Baidu, Alibaba, and Tencent have been collecting and analyzing data from millions of users and pushing forward deep learning based applications. For example, Tencent YouTu Lab has developed identification (ID) card identification and bank card identification systems. These systems can read information from card images to check user information while registering and bank information while purchasing. The identification systems are based on deep learning model and large volume of user data provided by Tencent. Apple develops Siri, a virtual intelligent assistant in iPhones, to answer questions about weather, location, news according to voice commands and dial numbers or send text messages. Siri also utilizes deep learning methods and uses data from apple services [45]. Google uses deep learning on Google translation service with massive data collected by Google search engine.

MBD contains a large variety of information of offline data and online real-time data stream generated from smart mobile terminals, sensors, and services and hastens various applications based on the advancement of data analysis technologies, such as collaborative filtering-based recommendation [46, 47], user social behavior characteristics analysis [48–51], vehicle communications in the Internet of Vehicles (IoV) [52], online smart healthcare [53], and city residents' activity analysis [6]. Although the machine learning-based methods are widely applied in the MBD fields and obtain good performances in real data test, the present methods still need to be further developed. Therefore, five main challenges facing MBD analysis regarding the machine learning-based methods include large-scale and high-speed M-Internet, overfitting and underfitting problems, generalization problem, cross-modal learning, and extended channel dimensions and should be considered.

This paper attempts to identify the requirement and the development of machine learning-based mobile big data analysis through discussing the insights of challenges in the MBD and reviewing state-of-the-art applications of data analysis in the area of MBD. The remainder of the paper is organized as follows. Section 2 introduces the development of data collection and properties of MBD. The frequently adopted methods of data analysis and typical applications are reviewed in Section 3. Section 4 summarizes the future challenges of MBD analysis and provides suggestions.

2. Development and Collection of the Mobile Big Data

2.1. Data Collection. Data collection is the foundation of a data processing and analysis system. Data are collected from mobile smart terminals and Internet services, or called mobile Internet devices (MIDs) generally, which are multimedia-capable mobile devices providing wireless Internet access and contain smartphones, wearable computers, laptop computers, wireless sensors, etc. [54].

MBD can be divided into two hierarchical data form: transmission and application data, from bottom to top. The transmission data focus on solving channel modeling [55, 56] and user access problems corresponding to the physical transmission system of M-Internet. On this foundation, application data focus on the applications based on the MBD including social networks analysis [57–59], user behavior analysis [48, 50, 60], speech analysis and decision in IoV [61–66], smart grid [67, 68], networked healthcare [53, 69, 70], finance services [46, 71], etc.

Due to the heterogeneity of the M-Internet and the variety of the access devices, the collected data are unstructured and usually in many categories and formats, which make data preprocessing become an essential part of a data processing and analysis system in order to ensure the input data complete

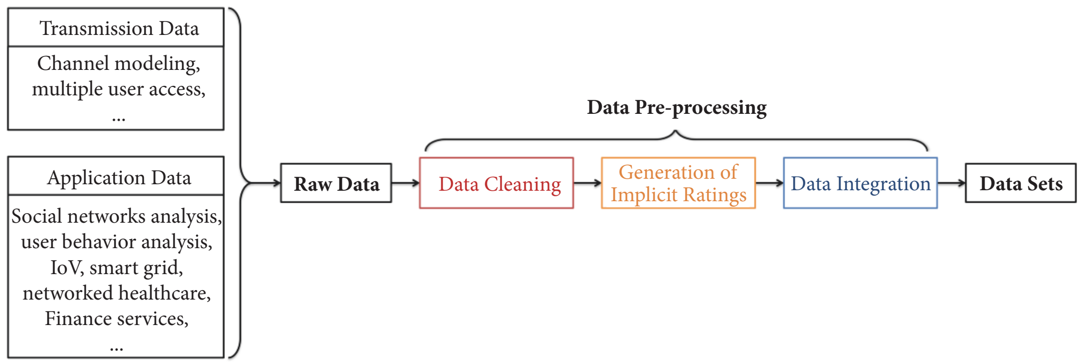


FIGURE 2: The procedures of data collection and preprocessing.

and reliable [72]. Data preprocessing can be divided into three steps which are data cleaning, generation of implicit ratings, and data integration [46].

(1) Data Cleaning. Due to possible equipment failures, transmission errors, or human factor, raw data are "dirty data" which cannot be directly used, generally [46]. Therefore, data cleaning methods including outlier detection and denoising are applied in the data preprocessing to obtain the data meet required quality. Manual removal of error data is difficult and impossible to accomplish in MBD due to the massive volume. Common data cleaning methods can alleviate the dirty data problem to some extent by training support vector regression (SVR) classifiers [73], multiple linear regression models [74], autoencoder [75], Bayesian methods [76–78], unsupervised methods [79], or information-theoretic models [79].

(2) Generation of Implicit Ratings. Generation of implicit ratings is mainly applied in recommend systems. The volume of rating data increases rapidly by analyzing specific user behaviors to solve data sparsity problem with machine learning algorithms, for example, neural networks and decision trees [46].

(3) Data Integration. Data integration is a step to integrate data from different resources with different formats and categories and to handle missing data fields [7].

Figure 2 represents the procedures of data collection and preprocessing.

2.2. Properties of Mobile Big Data. The MBD brings a massive amount of new challenges to conventional data analysis methods for its high dimensionality, heterogeneity, and other complex features from applications, such as planning, operation and maintenance, optimization, and marketing [57]. This section discusses the five Vs (short for volume, velocity, variety, value, and veracity) features [80] deriving from big data towards the MBD. The five Vs features have been improved in M-Internet, while it makes users access Internet anytime and anywhere [81].

(1) Volume: Large Number of MIDs, Exabyte-Level Data, and High-Dimensional Data Space. Volume is the most obvious feature of MBD. In the forthcoming 5G network and the era of MBD, conventional store and analysis methods are incapable of processing the 1000x or more wireless traffic volume [7, 82]. It is of great urgency to improve present MBD analysis methods and propose new ones. The methods should be simple and cost-effective to be implemented for MBD processing and analysis. Moreover, they should also be effective enough without requiring a massive amount of data for model training. Finally, they are precise to be applied in various fields [81].

(2) Velocity: Real-Time Data Streams and Efficiency Requirement. Velocity can be considered as the speed at which data are transmitted and analyzed [83]. The data is now continuously streaming into the servers in real-time and makes the original batch process break down [84]. Due to the high generating rate of MBD, velocity is the efficiency requirement of MBD analysis since real-time data processing and analysis are extremely important in order to maximize the value of MBD streams [7].

(3) Variety: Heterogeneous and Nonstructured Mobile Multimedia Contents. Due to the heterogeneity of MBD which means that mobile data traffic comes from spatially distributed data resources (*i.e.,* MIDs), the variety of MBD arises and makes the MBD more complex [4]. Meanwhile, the nonstructured MBD also causes the variety. The MBD can be divided into structured data, semistructured data, and unstructured data. Here, unstructured data are usually collected in new applications and have random data fields and contents [7]; therefore, they are difficult to analyze before data cleaning and integration.

(4) Value: Mining Hidden Knowledge and Patterns from Low Density Value Data. Value, or low density value of MBD, is caused by a large amount of useless or repeated information in the MBD. Therefore, we need to mine the big value by MBD analyzing which is hidden knowledge and patterns extraction. The purified data can provide comprehensive information to conduct more effectively analysis results about user demands, user behaviors, and user habits [85] and to achieve better system management and more accurate demand prediction and decision-making [86].

(5) Veracity: Consistency, Trustworthiness, and Security of MBD. The veracity of MBD includes two parts: data consistency and trustworthiness [80]. It can also be summarized as data quality. MBD quality is not guaranteed due to the noise of transmission channel, the equipment malfunctioning, and the uncalibrated sensors of MIDs or the human factor (for instance, malicious invasion) resulting in low-quality data points [4]. Veracity of MBD ensures that the data used in analysis process are authentic and protected from unauthorized access and modification [80].

3. Applications of Machine Learning Methods in the Mobile Big Data Analysis

3.1. Development of Data Analysis Methods. In this section, we present some recent achievements in data analysis from four different perspectives.

3.1.1. Divide-and-Conquer Strategy and Sampling of Big Data. The strategies dividing and conquering big data is a computing paradigm dealing with big data problems. The development of distributed and parallel computing makes divide-and-conquer strategy particularly important.

Generally speaking, whether the diversity of samples in learning data benefits the training results varies. Some redundant and noisy data can cause a large amount of storage cost as well as reducing the efficiency of the learning algorithm and affecting the learning accuracy. Therefore, it is more preferable to select representative samples to form a subset of original sample space according to a certain performance standard, such as maintaining the distribution of samples, topological structure, and keeping classification accuracy. Then learning method will be constructed on previous formed subset to finish the learning task. In this way, we can maintain or even improve the performance of big data analyzing algorithm with minimum computing and stock resources. The need to learn with big data demands on sample selection methods. But most of the sample selection method is only suitable for smaller data sets, such as the traditional condensed nearest neighbor [93], the reduced nearest neighbor [94], and the edited nearest neighbor [95]; the core concept of these methods is to find the minimum consistent subset. To find the minimum consistent subset, we need to test every sample and the result is very sensitive to the initialization of the subset and samples setting order. Li et al. [96] proposed a method to select the classification and edge boundary samples based on local geometry and probability distribution. They keep the space information of the original data but need to calculate k-means for each sample. Angiulli et al. [97, 98] proposed a fast condensation nearest neighbor (FCNN) algorithm based on condensed nearest neighbor, which tends to choose the classification boundary samples.

Jordan [99] proposed statistical inference method for big data. When dealing with statistical inference with divide-and-conquer algorithm, we need to get confidence intervals from huge data sets. By data resampling and then calculating confidence interval, the Bootstrap theory aims to obtain the fluctuation of the evaluation value. But it does not fit big data. The incomplete sampling of data can lead to erroneous range fluctuations. Data sampling should be correct in order to provide statistical inference calibration. An algorithm named Bag of Little Bootstraps was proposed, which can not only avoid this problem, but also has many advantages on computation. Another problem discussed in [99] is massive matrix calculation. The divide-and-conquer strategy is heuristic, which has a good effect in practical application. However, new theoretical problems arise when trying to describe the statistical properties of partition algorithm. To this end, the support concentration theorem based on the theory of random matrices has been proposed.

In conclusion, data partition and parallel processing strategy is the basic strategy to deal with big data. But the current partition and parallel processing strategy uses little data distribution knowledge, which has influence on the load balancing and the calculation efficiency of big data processing. Hence, there exists an urgent requirement to solve the problem about how to learn the distribution of big data for the optimization of load balancing.

3.1.2. Feature Selection of Big Data. In the field of data mining, such as document classification and indexing, the dataset is always large, which contains a large number of records and features. This leads to the low efficiency of algorithm. By feature selection, we can eliminate the irrelevant features and increase the speed of task analysis. Thus, we can get a better preformed model with less running time.

Big data processing faces a huge challenge on how to deal with high-dimensional and sparse data. Traffic network, smartphone communication records, and information shared on Internet provide a large number of high-dimensional data, using tensor (such as a multidimensional array) as natural representation. Tensor decomposition, in this condition, becomes an important tool for summary and analysis. Kolda [100] proposed an efficient use of the memory of the Tucker decomposition method named as memory-efficient Tucker (MET) decomposition decreasing time and space cost which traditional tensor decomposition algorithm cannot do. MET adaptively selects execution strategy based on available memory in the process of decomposition. The algorithm maximizes the speed of computation in the premise of using the available memory. MET avoid dealing with the large number of sporadic intermediate results proceeded during the calculation process. The adaptive selections of operation sequence not only eliminate the intermediate overflow problem, but also save memory without reducing the precision. On the other hand, Wahba [101] proposed two approaches to the statistical machine learning model which involve discrete, noisy, and incomplete data. These two methods are regularized kernel estimation (RKE) and robust manifold unfolding (RMU). These methods use dissimilarity between training information to get nonnegative low rank definite matrix. The matrix will then be embedded into a low dimensional Euclidean space, which coordinate can be used as features of various learning modes. Similarly, most online learning research needs to access all features of training instances. Such classic scenario is not always suitable for practical applications when facing high-dimensional data instances or expensive feature sets. In order to break through this limit, Hoi et al. [102] propose an efficient algorithm to predict online feature solving problem using some active features based on their study of sparse regularization and truncation technique. They also test the proposed algorithm in some public data sets for feature selection performance.

The traditional self-organizing map (SOM) can be used for feature extraction. But the low speed of SOM limits its usage on large data sets. Sagheer [103] proposed a fast self-organizing map (FSOM) to solve this problem. The goal of this method is to find a feature space where data is mainly distributed in. If there exits such area, data can be extracted in these areas instead of information extraction in overall feature spaces. In this way, we can greatly reduce extraction time.

Anaraki [104] proposed a threshold method of fuzzy rough set feature selection based on fuzzy lower approximation. This method adds a threshold to limit the QuickReduct feature selection. The results of the experiment prove that this method can also help the accuracy of feature extraction with lower running time.

Gheyas et al. [105] proposed a hybrid algorithm of simulated annealing and genetic algorithm (SAGA), combining the advantages of simulated annealing algorithm, genetic algorithm, greedy algorithm, and neural network algorithm, to solve the NP-hard problem of selecting optimal feature subset. The experiment shows that this algorithm can find better optimal feature subset, reducing the time cost sharply. Gheyas pointed in as conclusion that there is seldom a single algorithm which can solve all the problems; the combination of algorithms can effectively raise the overall affect.

To sum up, because of the complexity, high dimensionality, and uncertain characteristics of big data, it is an urgent problem to solve how to reduce the difficulty of big data processing by using dimension reduction and feature selection technology.

3.1.3. Big Data Classification. Supervised learning (classification) faces a new challenge of how to deal with big data. Currently, classification problems involving large-scale data are ubiquitous, but the traditional classification algorithms do not fit big data processing properly.

(1) Support Vector Machine (SVM). Traditional statistical machine learning method has two main problems when facing big data. (1) Traditional statistical machine learning methods are always involving intensive computing which makes it hard to apply on big data sets. (2) The prediction of model that fits the robust and nonparameter confidence interval is unknown. Lau et al. [106] proposed an online support vector machine (SVM) learning algorithm to deal with the classification problem for sequentially provided input data. The classification algorithm is faster, with less support vectors, and has better generalization ability. Laskov et al. [107] proposed a rapid, stable, and robust numerical incremental support vector machine learning method. Chang et al. [108] developed an open source package called LIBSVM as a library for SVM code implementation.

In addition, Huang et al. [109] present a large margin classifier M4. Unlike other large margin classifiers which locally or globally constructed separation hyperplane, this model can learn both local and global decision boundary. SVM and minimax probability machine (MPM) has a close connection with the model. The model has important theoretical significance and furthermore, the optimization problem of maxi-min margin machine (M^4) can be solved in polynomial time.

(2) Decision Tree (DT). Traditional decision tree (DT), as a classic classification learning algorithm, has a large memory requirement problem when processing big data. Franco-Arcega et al. [110] put forward a method of constructing DT from big data, which overcomes some weakness of algorithms in use. Furthermore, it can use all training data without saving them in memory. Experimental results showed that this method is faster than current decision tree algorithm on large-scale problems. Yang et al. [111] proposed a fast incremental optimization decision tree algorithm for large data processing with noise. Compared with former decision tree data mining algorithm, this method has a major advantage on real-time speed for data mining, which is quite suitable when dealing with continuous data from mobile devices. The most valuable feature of this model is that it can prevent explosive growth of the decision tree size and the decrease of prediction accuracy when the data packet contains noise. The model can generate compact decision tree and predict accuracy even with highly noisy data. Ben-Haim et al. [112] proposed an algorithm of building parallel decision tree classifier. The algorithm runs in distributed environment and is suitable for large amount and streaming data. Compared with serial decision tree, the algorithm can improve efficiency under the premise of accuracy error approximation.

(3) Neural Network and Extreme Learning Machine (ELM). Traditional feedforward neural networks usually use gradient descent algorithm to tune weight parameters. Generally speaking, slow learning speed and poor generalization performance are the bottlenecks that restrict the application of feedforward neural network. Huang et al. [113] discarded the iterative adjustment strategy of the gradient descent algorithm and proposed extreme learning machine (ELM). This method randomly assigns the input weights and the deviations of the single hidden layer neural network. It can analyze the output weights of the network by one step calculation. Compared to the traditional feedforward neural network training algorithm, the network weights can be determined by multiple iterations, and the training speed of ELM is significantly improved.

However, due to the limitation of computing resource and computational complexity, it is a difficult problem to train a single ELM on big data. There are usually two ways to solve this problem: (1) training ELM [114] based with divide-and-conquer strategy; (2) introducing parallel mechanism [115] to train a single ELM. It is shown in [116, 117] that a single ELM has strong function approximation ability. Whether it is possible to extend this approximation capability to ELM based on divide-and-conquer strategy is a key index to evaluate the possibility that ELM can be applied to big data.

Some of the related studies also include effective learning to solve such problem [118].

In summary, the traditional classification method of machine learning is difficult to apply to the analysis of big data directly. The study of parallel or improved strategies of different classification algorithms has become the new direction.

3.1.4. Big Data Deep Learning. With the unprecedentedly large and rapidly growing volumes of data, it is hard for us to get hidden information from big data with ordinary machine learning methods. The shallow-structured learning architectures of most conventional learning methods are not fit for the complex structures and relationships in these input data. Big data deep learning algorithm, with its deep architectures and globally feature extracting ability, can learn complex patterns and hidden connections beyond big data [37, 119]. It has had state-of-the-art performances in many benchmarks and also been applied in industry products. In this section, we will introduce some deep learning methods in big data analytics.

Big data deep learning has some problems: (1) the hidden layers of deep network make it difficult to learn from a given data vector, (2) the gradient descent method for parameters learning makes the initialization time increasing sharply as the number of parameters arises, and (3) the approximations at the deepest hidden layer may be poor. Hinton et al. [32] proposed a deep architecture: deep belief network (DBN) which can learn from both labeled and unlabeled data by using unsupervised pretraining method to learn unlabeled data distributions and a supervised fine-tune method to construct the models, and solved part of the aforementioned problems. Meanwhile, subsequent researches, for example, [120], improved the DBN trying to solve the problems.

Convolutional neural network (CNN) [121] is another popular deep learning network structure for big data analyzing. A CNN has three common features including local receptive fields, shared weights, and spatial or temporal subsampling, and two typical types of layers [122, 123]. Convolutional layers are key parts of CNN structure aiming to extract features from image. Subsampling layers, which are also called pooling layers, adjust outputs from convolutional layer to get translation invariance. CNN is mainly applied in computer vision field for big data, for example, image classification [124, 125] and image segmentation [126].

Document (or textual) representation, also part of NLP, is the basic method for information retrieval and important to understand natural language. Document representation finds specific or important information from the documents by analyzing document structure and content. The unique information could be document topic or a set of labels highly related to the document. Shallow models for document representation only focus on small part of the text and get simple connection between words and sentences. Using deep learning can get global representation of the document because of its large receptive field and hidden layers which could extract more meaningful information. The deep learning methods for document representation make it possible to obtain features from high-dimensional textual data. Hinton et al. [127] proposed deep generative model to learn binary codes for documents which make documents easy to store up. Socher et al. [128] proposed a recursive neural network on analyzing natural language and contexts, achieving state-of-the-art results on segmentation and understanding of natural language processing. Kumer et al. [129] proposed recurrent neural networks (RNN) which construct search space from large amount of textual data.

With the rapid growth and complexity of academic and industry data sets, how to train deep learning models with large amount of parameters has been a major problem. The works in [40, 41, 43, 130–133] proposed effective and stable parameter updating methods for training deep models. Researchers focus on large-scale deep learning that can be implemented in parallel including improved optimizers [131] and new structures [121, 133–135].

In conclusion, big data deep learning methods are the key methods of data mining. They use complex structure to learn patterns from big data sets and multimodal data. The development of data storage and computing technology promotes the development of deep learning methods and makes it easier to use in practical situations.

3.2. Wireless Channel Modeling. As is well known, wireless communication transmits information through electromagnetic waves between a transmitting antenna and a receiving antenna, which is deemed as a wireless channel. In the past few decades, the channel dimension has been extended to space, time, and frequency, which means the channel property is comprehensively discovered. Another development is that channel characteristics can be accurately described by different methods, such as channel modeling [136].

Liang et al. [137] used machine learning to predict channel state information so as to decease the pilot overhead. Especially for 5G, wireless big data emerges and its related technologies are employed to traditional communication research to meet the demand of 5G. However, the wireless channel is essentially a physical electromagnetic wave, and the current 5G channel model research follows the traditional way. Zhang [138] proposed an interdisciplinary study of big data and wireless channels, which is a cluster-based channel model. In the cluster-nuclei based channel model, the multipath components (MPCs) are aggregated into a traditional stochastically channel model. At the same time, the scene is discerned by the computer and the environment is rebuilt by machine learning methods. Then, by matching the real propagation objects with the clusters, the cluster-nuclei, which are the key factors in contacting deterministic environment and stochastic clusters, can be easily found. There are two main steps employing the machine learning methods in the cluster-nuclei based channel model. The recent progress is shown as follows.

3.2.1. A Gaussian Mixture Model (GMM) Based Channel MPCs Clustering Method. The MPCs are clustered with the Gaussian mixture model (GMM) [87, 139]. Using sufficient statistic characteristics of channel multipath, the GMM can get clusters corresponding to the multipath propagation characteristics. The GMM assumes that all the MPCs consist

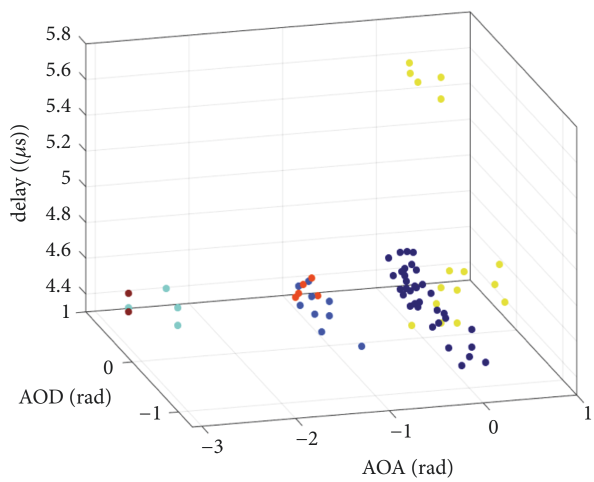


Figure 3: Clustering results of GMM [87].

of several Gaussian distributions in varying proportions. Given a set of N channel multipath X, the log-likelihood of the Gaussian mixture model is

$$L(X;\Theta)=\sum_{i=1}^{N}\log\sum_{k=1}^{K}\pi_k p\left(x_i \mid z_i;\mu_k,\Sigma_k\right), \tag{1}$$

where $\Theta=\{\pi_k,\mu_k,\Sigma_k,\ k=1,\cdots,K\}$ is the set of all the parameters and $\pi_k\in[0,1]$ is the prior probability satisfying the constraint $\sum_{k=1}^{K}\pi_k=1$. To estimate the GMM parameters, expectation maximization (EM) algorithm is employed to solve the log-likelihood function of GMM [87]. Figure 3 illustrates the simulation result of GMM clustering algorithm.

As seen in Figure 3, the GMM clustering obtains clearly compact clusters. As scattering property of the channel multipath obeys Gaussian distribution, the compact clusters can accord with the multipath scattering property. Moreover, corresponding to the clustering mechanism of GMM, paper [87] proposed a compact index (CI) to evaluate the clustering results shown as follows:

$$CI=\frac{\operatorname{tr}(B)/(K-1)}{\operatorname{tr}(W)/(L-K)}\bullet\left(\sum_{k=1}^{K}S_k^2\right), \tag{2}$$

where S_k^2 is the variance of the kth cluster and $\operatorname{tr}(B)$ and $\operatorname{tr}(W)$ are given as

$$\operatorname{tr}(B)=\sum_{k=1}^{K}L_k\bullet MCD\left(c_k,\bar{c}\right)^2, \tag{3}$$

$$\operatorname{tr}(W)=\sum_{k=1}^{K}\sum_{j\in C_n}MCD\left(x_j,c_k\right)^2, \tag{4}$$

where L_k is the number of multipaths corresponding to the kth cluster. Both the means and variances of the clusters are considered in CI. Considering sufficient statistics characteristics, CI can uncover the inherent information of multipath parameters and provide appropriate explanation to the clustering result. Besides, considering sufficient statistics characteristics, the CI can evaluate the clustering results more reasonably.

3.2.2. Identifying the Scatters with the Simultaneous Localization and Mapping Algorithm (SLAM). In order to reconstruct three-dimensional (3D) propagation environment and to find the main deterministic objects, simultaneous localization and mapping (SLAM) algorithm is used to identify the texture from the measurement scenario picture [140, 141]. Figure 4 illustrates our indoor reconstruction result with SLAM algorithm.

The texture of propagation environment can be used to search for the main scatters in the propagation environment. Then, the three-dimensional propagation environment can be reconstructed with the deep learning method.

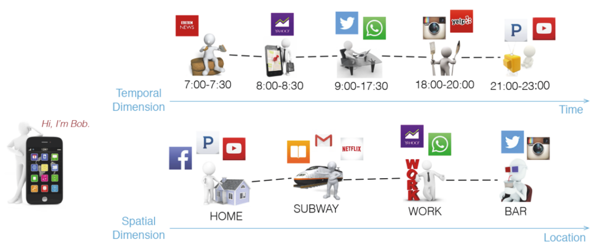
Then the mechanism to form the cluster-nuclei is clear. The channel impulse response can be produced by machine learning with a limited number of cluster-nuclei, *i.e.*, decision tree [142], neural network [143], and mixture model [144]. Based on the database from various scenarios, antenna configurations, and frequency, channel changing rules can be explored and then input into the cluster-nuclei based modeling. Finally, the predication of channel impulse response in various scenarios and configuration can be realized [138].

3.3. Analyses of Human Online and Offline Behavior Based on Mobile Big Data. The advances of wireless networks and increasing mobile applications bring about explosion of mobile traffic data. It is a good source of knowledge to obtain the individuals' movement regularity and acquire the mobility dynamics of populations of millions [145]. Previous researches have described how individuals visit geographical locations and employed mobile traffic data to analyze human offline mobility patterns. Representative works like [146, 147] explore the mobility of users in terms of the number of base stations they visited, which turned out to be a heavy tail distribution. Authors in [146, 148, 149] also reveal that a few important locations are frequently visited by users. In particular, these preferred locations are usually related to home and work places. Moreover, through defining a measure of entropy, Song et al. [150] believe that 93% of individual movements are potentially predictable. Thus, various models have been applied to describe the human offline mobility behavior [151]. Passively collecting human mobile traffic data while users are accessing the mobile Internet has many advantages like low energy consumption. In general, the mobile big data covers a wide range and a great number of populations with fine time granularity, which gives us an opportunity to study human mobility at a scale that other data sources are very hard to reach [152]. Novel offline user mobility models developed based on the mobile big data are expected to benefit many fields, including urban planning, road traffic engineering, telecommunication network construction, and human sociology [145].

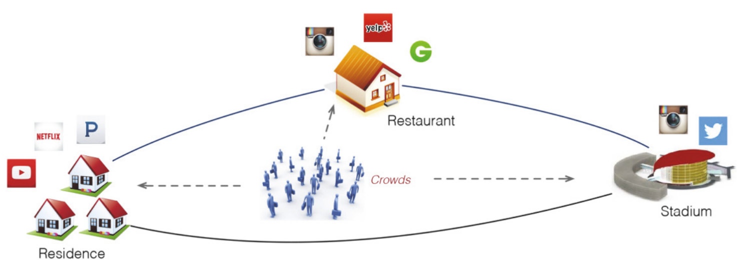
Online browsing behavior is another important facet regarding user behavior when it comes to network resource consumption. A variety of applications are now available on smart devices, covering all aspects of our daily life and

(a) (b)

Figure 4: Recognition of multiobjects with SLAM algorithm: (a) real indoor scene and (b) reconstruction result with SLAM algorithm.



(a) App usage behavior of Bob in temporal and spatial dimension



(b) App usage behavior of crowds at crowd gathering place

Figure 5: App usage behavior in daily life: (a) the app usage behavior of an individual and (b) app usage behavior of crowds at crowd gathering places [50].

providing convenience. For example, we can order taxies, shop, and book hotels using mobile phones. Yang et al. [49] provide a comprehensive study on user behaviors in exploiting the mobile Internet. It has been found that many factors, such as data usage and mobility pattern, may impact people's online behavior on mobile devices. It is discovered that the more the number of distinct cells a user visit, the more diverse applications user has visited. Zheng et al. [153] analyze the longitudinal impact of proximity density, personality, and location on smartphone traffic consumption. In particular, location has been proven to have strong influences on what kinds of apps users prefer to use [149, 153]. The aforementioned observations point out that there is a close relationship between online browsing behavior and offline mobility behavior.

Figure 5(a) is an example of how browsed applications and current location related to each other from the view of temporal and spatial regularity. It has been found that the mobility behaviors have strong influences on online browsing behavior [149, 153, 154]. Similar trends can also be observed for crowds at crowd gathering places, as is shown in Figure 5(b); *i.e.*, certain apps are favored at places that group people together and provide some specific functions. The authors in [50] tried to measure the relationship between human mobility and app usage behavior. In particular, the authors proposed a rating framework which can forecast the online app usage behavior for individuals and crowds. Building the bridge between human offline mobility and online mobile Internet behavior can tell us what people really need in daily life. Content providers can leverage this knowledge to appropriately recommend content for mobile users. At the same time, Internet service providers (ISPs) can use this knowledge to optimize networks for better end-user experiences.

In order to make full use of users' online and offline information, some researchers begin to quantize the interplay between online social network and offline social network and investigate network dynamics from the view of mobile traffic data [155–158]. Specifically, the online and offline social networks are, respectively, constructed based on online interest based and location based social network among mobile users. The two different networks are grouped into layers of a multilayer social network $M = \{G^{on}, G^{off}\}$, as shown in Figure 6. G^{off} and G^{on} depict offline and online

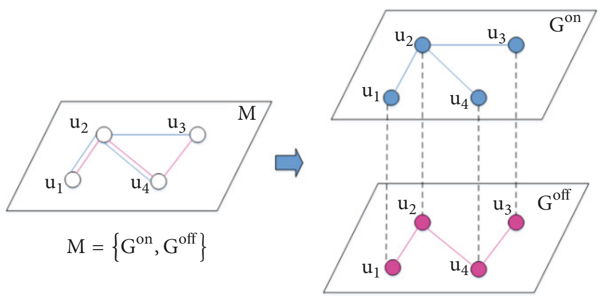


Figure 6: Multilayer model of a network [88].

social network separately. In each layer, the graph is described as $G = \langle V, E \rangle$, where V and E, respectively, represent node sets and edge sets. Nodes, such as $u_1, \ldots, u_4$, represent users. Edges exist among users when users share similar object-based interests [88]. Combining information from manifold networks in a multilayer structure provides a new insight into user interactions between virtual and physical worlds. It sheds light on the link generation process from multiple views, which will improve social bootstrapping and friend recommendations in various valuable applications by a large margin [158].

So far, we have summarized some representative works related to human online and offline behaviors. It is meaningful to note that owing to the highly spatial-temporal and nonhomogeneous nature of mobile traffic data, a pervasive framework is challenging yet indispensable to realize the collection, processing, and analyses of massive data, reducing resource consumption and improving Quality of Experience (QoE). The seminal work by Qiao et al. [60] proposes a framework for MBD (FMBD). It provides comprehensive functions on data collection, storage, processing, analyzing, and management to monitor and analyze the massive data. Figure 7(a) displays the architecture of FMBD, while Figure 7(b) shows the considered mobile networks framework. With the interaction between user equipment and 2G/3G/4G network, real massive mobile data can be collected by traffic monitoring equipment (TME). The implementation modules are employed based on Apache software [159]. FMBD builds a security environment and easy-to-use platform both for operators and data analysts, showing good performance on energy efficiency, portability, extensibility, usability, security, and stability. In order to meet the increasing demands on traffic monitoring and analyzing, the framework provides a solution to deal with large-scale mobile big data.

In conclusion, the prosperity of continuously emerging mobile applications and users' increasing demands on accessing Internet all bring about challenges for current and future mobile networks. This section surveys the literature on analyses of human online and offline behavior based on the mobile traffic data. Moreover, a framework has also been investigated, in order to meet the higher requirement of dealing with dramatically increased mobile traffic data. The analyses based on the big data will provide valuable information for the ISPs on network deployment, resource management, and the design of future mobile network architectures.

3.4. Speech Recognition and Verification for the Internet of Vehicles.

With the significant development of smart vehicle produces, intelligent vehicle based Internet of Vehicle (IoV) technologies have received widespread attention of many giant Internet businesses [160–162]. The IoV technologies include the communication between different vehicles and vehicles to sensors, roads, and humans. These communications can help the IoV system sharing and the gathering information on vehicles and their surrounds.

One of the challenges in the real-life applications of smart vehicles and IoV systems is how to design a robust interactive method between drivers and the IoV system [163]. The level of focusing on driving will directly affect the danger of driver and passengers; hence, the attention of drivers should be paid on the complex road situation in order to avoid accidents during an intense driving. So, using the voices transfer information to the IoV systems is an effective solution for assistant and cooperative driving. By building a speech recognition interactive system, the driver can check traffic jams near the destination or order a lunch in the restaurant near the rest stop through the IoV system by using voice-based interaction. The speech recognition interactive system for IoV system can reduce the risk of vehicle accident, and the drivers do not need to touch the control panels or any buttons. A useful speech recognition system in IoV can simplify the life of the drivers and passengers in vehicles [164]. In the IoV system, drivers want to use their own voice commands to control the driving vehicles, and the IoV system must recognize the difference between an authorized and unauthorized user. Therefore, an automatic speaker verification system is necessary in IoV, which can protect the vehicle from the imposters.

Recently, many deep learning methods have been applied in the speech recognition and speaker verification systems [41, 165–167], and published results show that speech processing methods driven by MBD and deep learning can obviously improve the performance of the existing speech recognition and speaker verification system [40, 168, 169]. In the IoV systems, millions of sensors collect abundant vehicles and environmental noises from engines and streets will significantly reduce the accuracy of speech processing system, while the traditional speech enhancement methods, for example, Wiener filtering [170] and minimum mean-square error estimation (MMSE) [171] which focus on advancing signal noise ratio (SNR), do not take full advantage of a priori distribution of noises around vehicles. With the help of machine learning and deep learning methods, we can use a priori knowledge of the noises to improve the robustness of speech processing systems.

For speech recognition task, deep-neural-network (DNN) can be applied to train an effective monophone classifier, instead of the traditional GMM based classifier. Moreover, the deep-neural-network hidden Markov model (DNN-HMM) speech recognition model can significantly improve the performance of Gaussian mixture model hidden Markov model (GMM-HMM) models [172–174]. As shown

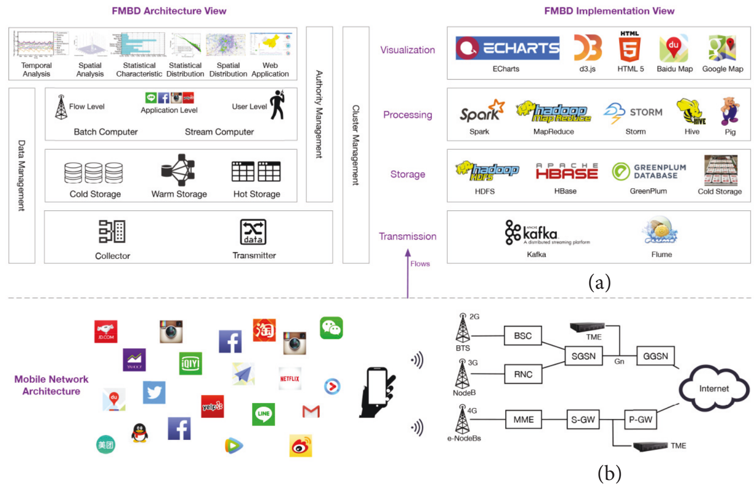


FIGURE 7: The overall architecture of framework for mobile big data (FMBD) and our considered mobile networks architecture [60].

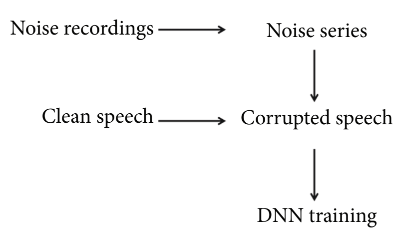


FIGURE 8: Multitraining DNN [89].

in Figure 8, making full use of the self-adaption power of DNN, we can use the multitraining methods to improve the robustness of DNN monophone classifier by adding noise into the training data [89]. The experimental results in [89, 175] show that the multitraining method can build a matched training and testing condition which can improve the accuracy of noisy speech recognition, especially for the prior knowledge of noise types that we can easily obtain in vehicles.

As shown in Figure 9, a DNN can also be used to train a feature mapping network (FMN) which uses noisy features as input and corresponding clean features as training target. Enhanced features extracted by the FMN can improve the performance of speech recognition systems. Han et al. [176] used FMN to extract one enhanced Mel-frequency cepstral coefficient (MFCC) frame from 15 noisy MFCCs frames. Xu et al. [90] built a FMN which learned the mapping from a log spectrogram to a log Mel filter bank. The enhanced feature can remarkably reduce the word error rate in speech recognition.

Besides getting the mapping feature directly, the DNN can also be used to train an ideal binary mask (IBM) which can be used to separate the clean speech from background noise as shown in Figure 10 [91, 177, 178]. With a priori knowledge of noise types and SNR, we can generate IBMs as training targets and use noisy power spectral as training data. In the test phase, we can use the learned IBMs to get enhanced features which can improve the robustness of speech recognition.

In speaker verification tasks, the classical GMM based methods, for example, Gaussian mixture model universal background model (GMM-UBM) [179] and i-vector systems [180], need to build a background GMM, firstly, using a large quantity of speaker independent speeches. Then, by computing the statistics information on each GMM component of enrollment speakers, we can get speaker models or speaker i-vectors. However, a trained monophone classification DNN can replace the function of GMM by computing the statistics information on each monophone instead of on GMM components. Many published papers [181–184] show that the DNN-i-vector based speaker verification systems work better than the GMM-i-vector method on detection accuracy and robustness.

Unlike in the speech recognition tasks where the DNNs are used to get enhanced features from noisy features, researchers more prefer to use a DNN or convolutional neural network (CNN) to generate noise robustness bottleneck feature directly in speaker verification tasks [185–187]. As shown in Figure 11, acoustic features or feature maps are used to train a DNN/CNN with a bottleneck layer which has less nodes and closes to the output layer. Speaker ID, noise types, monophone labels, or combination of these labels are used as training targets. Outputs of bottleneck layers include abundant differentiated information and can be used as speaker verification features which improve the performance of classical speaker verification methods such as the aforementioned GMM-UBM and i-vector. Similar to the multitraining method, adding noisy speeches into the training data can also improve the robustness of extracted bottleneck features [65, 92].

Recently, some adversarial training methods are introduced to extract noise invariant bottleneck features [64, 188]. As shown in Figure 12, the adversarial network includes two parts, *i.e.*, an encoding network (EN) which can extract noise invariant features and a discriminative network (DN) which can judge noise types of the noise invariant feature generated from EN. Therefore, we can get robustness noise invariant features from EN which can improve the performance of

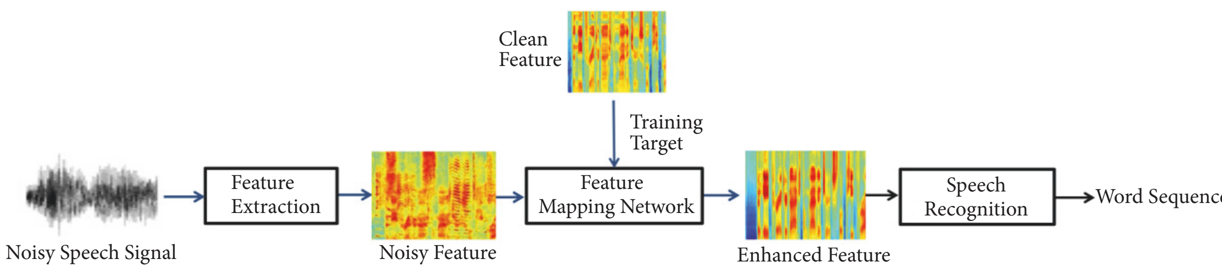


Figure 9: DNN used for feature mapping [90].

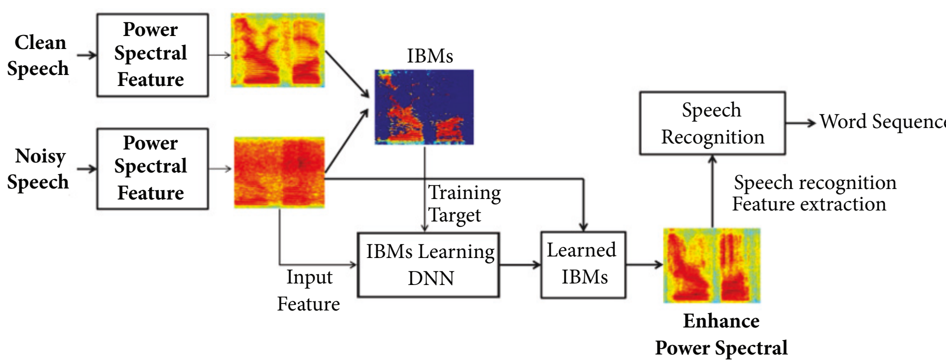


Figure 10: DNN used for IBMs learning [91].

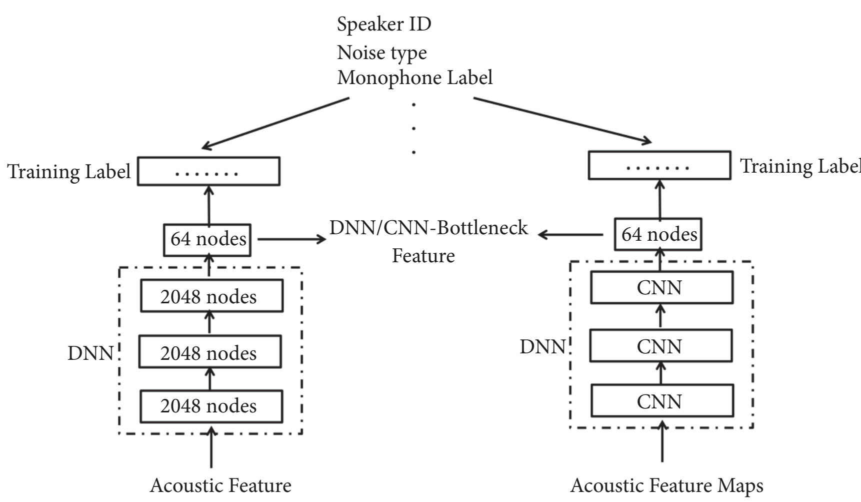


Figure 11: DNN/CNN used for extracting bottleneck feature [92].

speaker verification system by adversarial training these two parts in turn [64, 188].

In conclusion, using DNN and machine learning methods can make full use of the MBD collected from the IoV systems. Moreover, it improves the performance of speech recognition and speaker verification methods applied in the voice interactive systems.

4. Conclusions and Future Challenges

Although the machine learning-based methods introduced in Section 3 are widely applied in the MBD fields and obtain good performances in real data test, the present methods still need to be further developed. Therefore, five main challenges facing MBD analysis regarding the machine learning-based methods should be considered as follows.

(1) Large-Scale and High-Speed M-Internet. Due to the growth of MIDs and high speed of M-Internet, increasingly various mobile data traffic is introduced and results in a heavy load to the wireless transmission system, which leads us to improve wireless communication technologies including WLAN and cellular mobile communication. In addition, the requirement of real-time services and applications depends

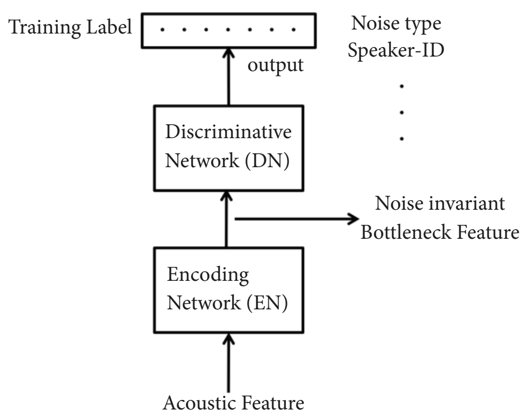


Figure 12: Adversarial training network for noise invariant bottleneck feature extraction [64].

on the development of machine learning-based MBD analysis methods towards high efficiency and precision.

(2) Overfitting and Underfitting Problems. A benefit of MBD to machine learning and deep learning lies in the fact that the risk of overfitting becomes smaller with more and more data available for training [28]. However, underfitting is another problem for the oversize data volume. In this condition, a larger model might be a better selection, while the model can express more hidden information of the data. Nevertheless, larger model which generally implies a deeper structure increases runtime of the model which affects the real-time performance. Therefore, the model size in machine learning and deep learning, which represents number of parameters, should be balanced to model performance and runtime.

(3) Generalization Problem. As the massive scale of MBD, it is impossible to gain entire data even if they are only in a specific field. Therefore, the generalization ability which can be defined as suitable of different data subspace, or called scalability, of a trained machine learning or deep learning model is of great importance for evaluating the performance.

(4) Cross-Modal Learning. The variety of MBD causes multiple modalities of data (for example, images, audios, personal location, web documents, and temperature) generated from multiple sensors (correspondingly, cameras, sound recorders, position sensor, and temperature sensor). Multimodal learning should learn from multimodal and heterogeneous input data with machine learning and deep learning [4, 189] and obtain hidden knowledge and meaningful patterns; however, it is quite difficult to discover.

(5) Extended Channel Dimensions. The channel dimensions have been extended to three domains, *i.e.,* space, time, and frequency, which means that the channel property is comprehensively discovered. Meanwhile, the increasing antenna number, high bandwidth, and various application scenarios bring the big data of channel measurements and estimations, especially for 5G. The finding channel characteristics need to be precisely described by more advanced channel modeling methodologies.

In this paper, the applications and challenges of machine learning-based MBD analysis in the M-Internet have been reviewed and discussed. The development of MBD in various application scenarios requires more advanced data analysis technologies especially machine learning-based methods. Three typical applications of MBD analysis focus on wireless channel modeling, human online and offline behavior analysis, and speech recognition and verification in the Internet of Vehicles, respectively, and the machine learning-based methods used are widely applied in many other fields. In order to meet the aforementioned future challenges, three main study aims, *i.e.,* accuracy, feasibility, and scalability [28], are highlighted for present and future MBD analysis research. In future work, accuracy improving will be also the primary task on the basis of a feasible architecture for MBD analysis. In addition, as the aforementioned discussion of the generalization problem, scalability has obtained more and more attentions especially in a classification or recognition problem where scalability also includes the increase in the number of inferred classes. It is of great importance to improve the scalability of the methods with the high accuracy and feasibility in order to face the analysis requirements of MBD.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was supported in part by the National Natural Science Foundation of China (NSFC) [Grant no. 61773071]; in part by the Beijing Nova Program Interdisciplinary Cooperation Project [Grant no. Z181100006218137]; in part by the Beijing Nova Program [Grant no. Z171100001117049]; in part by the Beijing Natural Science Foundation (BNSF) [Grant no. 4162044]; in part by the Funds of Beijing Laboratory of Advanced Information Networks of BUPT; in part by the Funds of Beijing Key Laboratory of Network System Architecture and Convergence of BUPT; and in part by BUPT Excellent Ph.D. Students Foundation [Grant no. XTCX201804].

References

[1] International Telecommunication Union (ITU), "ICT Facts and Figures 2017," https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx, 2017.

[2] Meeker, "Internet Trend 2017," http://www.kpcb.com/internet-trends, 2017.

[3] G. Fettweis and S. Alamouti, "5G: personal mobile internet beyond what cellular did to telephony," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 140–145, 2014.

[4] M. A. Alsheikh, D. Niyato, S. Lin, H.-P. Tan, and Z. Han, "Mobile big data analytics using deep learning and apache spark," *IEEE Network*, vol. 30, no. 3, pp. 22–29, 2016.

[5] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021 White Paper," https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html, 2017.

[6] Y. Guo, J. Zhang, and Y. Zhang, "An algorithm for analyzing the city residents' activity information through mobile big data mining," in *Proceedings of the Joint 15th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 10th IEEE International Conference on Big Data Science and Engineering and 14th IEEE International Symposium on Parallel and Distributed Processing with Applications, IEEE TrustCom/BigDataSE/ISPA 2016*, pp. 2133–2138, China, August 2016.

[7] Z. Liao, Q. Yin, Y. Huang, and L. Sheng, "Management and application of mobile big data," *International Journal of Embedded Systems*, vol. 7, no. 1, pp. 63–70, 2015.

[8] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

[9] W. Li and Z. Zhou, "Learning to hash for big data: current status and future trends," *Chinese Science Bulletin (Chinese Version)*, vol. 60, no. 5-6, p. 485, 2015.

[10] V. Mayerschönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Eamon Dolan/Houghton Mifflin Harcourt, Boston, 2013.

[11] D. Z. Yazti and S. Krishnaswamy, "Mobile big data analytics: research, practice, and opportunities," in *Proceedings of the 15th IEEE International Conference on Mobile Data Management, IEEE MDM 2014*, pp. 1-2, Australia, July 2014.

[12] E. Zeydan, E. Bastug, M. Bennis et al., "Big data caching for networking: moving from cloud to edge," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 36–42, 2016.

[13] Z. Liu, Y. Qi, Z. Ma et al., "Sentiment analysis by exploring large scale web-based Chinese short text," in *Proceedings of the International Conference on Computer Science and Application Engineering (CSAE)*, pp. 21–23, 2017.

[14] Z. Wang, Y. Qi, J. Liu, and Z. Ma, "User intention understanding from scratch," in *Proceedings of the 1st International Workshop on Sensing, Processing and Learning for Intelligent Machines, SPLINE 2016*, Denmark, July 2016.

[15] C. Zhang, Z. Si, Z. Ma, X. Xi, and Y. Yin, "Mining sequential update summarization with hierarchical text analysis," *Mobile Information Systems*, vol. 2016, Article ID 1340973, 10 pages, 2016.

[16] C. Zhang, Y. Zhang, W. Xu, Z. Ma, Y. Leng, and J. Guo, "Mining activation force defined dependency patterns for relation extraction," *Knowledge-Based Systems*, vol. 86, pp. 278–287, 2015.

[17] C. Zhang, W. Xu, Z. Ma, S. Gao, Q. Li, and J. Guo, "Construction of semantic bootstrapping models for relation extraction," *Knowledge-Based Systems*, vol. 83, pp. 128–137, 2015.

[18] M. Jordan, "Message from the president: the era of big data," *ISBA Bull*, vol. 18, pp. 1–3, 2011.

[19] W. Chen, D. Wipf, Y. Wang, Y. Liu, and I. J. Wassell, "Simultaneous Bayesian sparse approximation with structured sparse models," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6145–6159, 2016.

[20] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, "Projection design for statistical compressive sensing: a tight frame based approach," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2016–2029, 2013.

[21] H. Yong, D. Meng, W. Zuo, and L. Zhang, "Robust online matrix factorization for dynamic background subtraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[22] Q. Xie, D. Zeng, Q. Zhao et al., "Robust low-dose CT sinogram prepocessing via exploiting noise-generating mechanism," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2487–2498, 2017.

[23] M. O'Connor, G. Zhang, W. B. Kleijn, and T. D. Abhayapala, "Function splitting and quadratic approximation of the primal-dual method of multipliers for distributed optimization over graphs," *IEEE Transactions on Signal and Information Processing over Networks*, pp. 1-1, 2018.

[24] G. Zhang and R. Heusdens, "Distributed optimization using the primal-dual method of multipliers," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 1, pp. 173–187, 2018.

[25] G. Zhang and R. Heusdens, "Linear coordinate-descent message passing for quadratic optimization," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp. 20055-2008, 2012.

[26] G. Zhang, R. Heusdens, and W. B. Kleijn, "Large scale LP decoding with low complexity," *IEEE Communications Letters*, vol. 17, no. 11, pp. 2152–2155, 2013.

[27] Z. Ma, A. E. Teschendorff, A. Leijon, Y. Qiao, H. Zhang, and J. Guo, "Variational bayesian matrix factorization for bounded support data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 4, pp. 876–889, 2015.

[28] Z.-H. Zhou, N. V. Chawla, Y. Jin, and G. J. Williams, "Big data opportunities and challenges: discussions from data analytics perspectives," *IEEE Computational Intelligence Magazine*, vol. 9, no. 4, pp. 62–74, 2014.

[29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[30] Y. Bengio and S. Bengio, "Modeling high-dimensional discrete data with multi-layer neural networks," in *Proceedings of the 13th Annual Neural Information Processing Systems Conference, NIPS 1999*, pp. 400–406, USA, December 1999.

[31] M. Ranzato, Y.-L. Boureau, and Y. Le Cun, "Sparse feature learning for deep belief networks," in *Advances in Neural Information Processing Systems*, pp. 1185–1192, 2008.

[32] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[33] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proceedings of the 20th Annual Conference on Neural Information Processing Systems (NIPS '06)*, pp. 153–160, Cambridge, Mass, USA, December 2006.

[34] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks," *Journal of Machine Learning Research*, vol. 10, pp. 1–40, 2009.

[35] R. Salakhutdinov and G. Hinton, "Deep boltzmann machines," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, vol. 24, pp. 448–455, 2009.

[36] I. Goodfellow, H. Lee, and Q. V. Le, "Measuring invariances in deep networks," *Neural Information Processing Systems*, pp. 646–654, 2009.

[37] Y. Bengio and Y. LeCun, "Scaling learning algorithms towards, AI," *Large Scale Kernel Machines*, vol. 34, pp. 321–360, 2007.

[38] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.

[39] I. Arel, D. C. Rose, and T. P. Karnowski, "Deep machine learning—a new frontier in artificial intelligence research," *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13–18, 2010.

[40] G. E. Dahl, D. Yu, L. Deng et al., "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 30–42, 2012.

[41] G. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[42] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted Boltzmann machines for collaborative filtering," in *Proceedings of the 24th International Conference on Machine learning (ICML '07)*, vol. 227, pp. 791–798, Corvallis, Oregon, June 2007.

[43] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep, big, simple neural nets for handwritten digit recognition," *Neural Computation*, vol. 22, no. 12, pp. 3207–3220, 2010.

[44] M. D. Zeiler, G. W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *Proceedings of the 2011 IEEE International Conference on Computer Vision, ICCV 2011*, pp. 2018–2025, Spain, November 2011.

[45] A. Efrati, "How deep learning works at Apple, beyond," https://www.theinformation.com/How-Deep-Learning-Works-at-Apple-Beyond, 2013.

[46] Z. Yang, B. Wu, K. Zheng, X. Wang, and L. Lei, "A survey of collaborative filtering-based recommender systems for mobile internet applications," *IEEE Access*, vol. 4, pp. 3273–3287, 2016.

[47] K. Zhu, L. Zhang, and A. Pattavina, "Learning geographical and mobility factors for mobile application recommendation," *IEEE Intelligent Systems*, vol. 32, no. 3, pp. 36–44, 2017.

[48] S. Jiang, B. Wei, T. Wang, Z. Zhao, and X. Zhang, "Big data enabled user behavior characteristics in mobile internet," in *Proceedings of the 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–5, Nanjing, October 2017.

[49] J. Yang, Y. Qiao, X. Zhang, H. He, F. Liu, and G. Cheng, "Characterizing user behavior in mobile internet," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 1, pp. 95–106, 2015.

[50] Y. Qiao, X. Zhao, J. Yang, and J. Liu, "Mobile big-data-driven rating framework: measuring the relationship between human mobility and app usage behavior," *IEEE Network*, vol. 30, no. 3, pp. 14–21, 2016.

[51] Y. Qiao, J. Yang, H. He, Y. Cheng, and Z. Ma, "User location prediction with energy efficiency model in the Long Term-Evolution network," *International Journal of Communication Systems*, vol. 29, no. 14, pp. 2169–2187, 2016.

[52] M. Gerla and L. Kleinrock, "Vehicular networks and the future of the mobile internet," *Computer Networks*, vol. 55, no. 2, pp. 457–469, 2011.

[53] M. M. Islam, M. A. Razzaque, M. M. Hassan, W. N. Ismail, and B. Song, "Mobile cloud-based big healthcare data processing in smart cities," *IEEE Access*, vol. 5, pp. 11887–11899, 2017.

[54] Texas Instruments, "Wireless Handset Solutions: Mobile Internet Device," http://www.ti.com/solution/handset_smartphone, 2008.

[55] X. Ma, J. Zhang, Y. Zhang, and Z. Ma, "Data scheme-based wireless channel modeling method: motivation, principle and performance," *Journal of Communications and Information Networks*, vol. 2, no. 3, pp. 41–51, 2017.

[56] X. Ma, J. Zhang, Y. Zhang, Z. Ma, and Y. Zhang, "A PCA-based modeling method for wireless MIMO channel," in *Proceedings of the 2017 IEEE Conference on Computer Communications: Workshops (INFOCOM WKSHPS)*, pp. 874–879, Atlanta, GA, May 2017.

[57] X. Zhang, Z. Yi, Z. Yan et al., "Social computing for mobile big data," *The Computer Journal*, vol. 49, no. 9, pp. 86–90, 2016.

[58] K. Zhu, Z. Chen, L. Zhang, Y. Zhang, and S. Kim, "Geo-cascading and community-cascading in social networks: comparative analysis and its implications to edge caching," *Information Sciences*, vol. 436-437, pp. 1–12, 2018.

[59] S. Gao, H. Luo, D. Chen et al., "A cross-domain recommendation model for cyber-physical systems," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 2, pp. 384–393, 2013.

[60] Y. Qiao, Z. Xing, Z. M. Fadlullah, J. Yang, and N. Kato, "Characterizing flow, application, and user behavior in mobile networks: a framework for mobile big data," *IEEE Wireless Communications Magazine*, vol. 25, no. 1, pp. 40–49, 2018.

[61] H. Yu, Z. Tan, Z. Ma, R. Martin, and J. Guo, "Spoofing detection in automatic speaker verification systems using DNN classifiers and dynamic acoustic features," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12.

[62] H. Yu, Z.-H. Tan, Y. Zhang, Z. Ma, and J. Guo, "DNN filter bank cepstral coefficients for spoofing detection," *IEEE Access*, vol. 5, pp. 4779–4787, 2017.

[63] Z. Ma, H. Yu, Z.-H. Tan, and J. Guo, "Text-independent speaker identification using the histogram transform model," *IEEE Access*, vol. 4, pp. 9733–9739, 2016.

[64] H. Yu, Z.-H. Tan, Z. Ma, and J. Guo, "Adversarial network bottleneck features for noise robust speaker verification," in *Proceedings of the 18th Annual Conference of the International Speech Communication Association, INTERSPEECH 2017*, pp. 1492–1496, Sweden, August 2017.

[65] H. Yu, A. Sarkar, D. A. L. Thomsen, Z.-H. Tan, Z. Ma, and J. Guo, "Effect of multi-condition training and speech enhancement methods on spoofing detection," in *Proceedings of the 1st International Workshop on Sensing, Processing and Learning for Intelligent Machines, SPLINE 2016*, Denmark, July 2016.

[66] H. Yu, Z. Ma, and M. Li, "Histogram transform model Using MFCC features for text-independent speaker identification," in *Proceedings of the IEEE Asilomar Conference on Signals, Systems*, pp. 500–504, 2014.

[67] Z. Ma, J. Xie, H. Li et al., "The role of data analysis in the development of intelligent energy networks," *IEEE Network*, vol. 31, no. 5, pp. 88–95, 2017.

[68] Z. Ma, H. Li, Q. Sun, C. Wang, A. Yan, and F. Starfelt, "Statistical analysis of energy consumption patterns on the heat demand of buildings in district heating systems," *Energy and Buildings*, vol. 85, pp. 464–472, 2014.

[69] D. West, "How mobile devices are transforming healthcare," *Issues in Technology Innovation*, vol. 18, no. 1, pp. 1–11, 2012.

[70] L. A. Tawalbeh, R. Mehmood, E. Benkhlifa, and H. Song, "Mobile cloud computing model and big data analysis for healthcare applications," *IEEE Access*, vol. 4, pp. 6171–6180, 2016.

[71] S. Sagiroglu and D. Sinanc, "Big data: a review," in *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS '13)*, pp. 42–47, IEEE, San Diego, Calif, USA, May 2013.

[72] K. Zheng, L. Hou, H. Meng, Q. Zheng, N. Lu, and L. Lei, “Soft-defined heterogeneous vehicular network: architecture and challenges,” *IEEE Network*, vol. 30, no. 4, pp. 72–80, 2016.

[73] H. Hsieh, V. Klyuev, Q. Zhao, and S. Wu, “SVR-based outlier detection and its application to hotel ranking,” in *Proceedings of the 2014 IEEE 6th International Conference on Awareness Science and Technology (iCAST)*, pp. 1–6, Paris, France, October 2014.

[74] S. Rahman, M. Sathik, and K. Kannan, “Multiple linear regression models in outlier detection,” *International Journal of Research in Computer Science*, vol. 2, no. 2, pp. 23–28, 2012.

[75] H. A. Dau, V. Ciesielski, and A. Song, “Anomaly Detection using replicator neural networks trained on examples of one class,” in *Simulated Evolution and Learning*, vol. 8886 of *Lecture Notes in Computer Science*, pp. 311–322, Springer International Publishing, Cham, 2014.

[76] Z. Ma, J.-H. Xue, A. Leijon, Z.-H. Tan, Z. Yang, and J. Guo, “Decorrelation of neutral vector variables: theory and applications,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 1, pp. 129–143, 2018.

[77] Z. Ma, S. Chatterjee, W. B. Kleijn, and J. Guo, “Dirichlet mixture modeling to estimate an empirical lower bound for LSF quantization,” *Signal Processing*, vol. 104, pp. 291–295, 2014.

[78] Z. Ma and A. Leijon, “Bayesian estimation of beta mixture models with variational inference,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2160–2173, 2011.

[79] C. C. Aggarwal, “Outlier analysis,” in *Data Mining*, Springer, 2015.

[80] Y. Demchenko, P. Grosso, C. de Laat, and P. Membrey, “Addressing big data issues in scientific data infrastructure,” in *Proceedings of the IEEE International Conference on Collaboration Technologies and Systems (CTS ’13)*, pp. 48–55, May 2013.

[81] C. Zhou, H. Jiang, Y. Chen, L. Wu, and S. Yi, “User interest acquisition by adding home and work related contexts on mobile big data analysis,” in *Proceedings of the IEEE INFOCOM 2016 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 201–206, San Francisco, CA, USA, April 2016.

[82] X. Ge, H. Cheng, M. Guizani, and T. Han, “5G wireless backhaul networks: challenges and research advances,” *IEEE Network*, vol. 28, no. 6, pp. 6–11, 2014.

[83] S. Landset, T. M. Khoshgoftaar, A. N. Richter, and T. Hasanin, “A survey of open source tools for machine learning with big data in the Hadoop ecosystem,” *Journal of Big Data*, vol. 2, no. 1, pp. 24–59, 2015.

[84] D. Soubra, “The 3Vs that define Big Data,” http://www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data, 2012.

[85] L. Ma, F. Nie, and Q. Lu, “An analysis of supply chain restructuring based on big data and mobile internet—a case study of warehouse-type supermarkets,” in *Proceedings of the IEEE International Conference on Grey Systems and Intelligent Services, GSIS 2015*, pp. 446–451, UK, August 2015.

[86] A. McAfee and E. Brynjolfsson, “Big data: the management revolution,” *Harvard Business Review*, vol. 90, no. 10, pp. 60–128, 2012.

[87] Y. Li, J. Zhang, and Z. Ma, “Clustering in wireless propagation channel with a statistics-based framework,” in *Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, Barcelona, April 2018.

[88] P. Kazienko, K. Musiał, and T. Kajdanowicz, “Multidimensional social network in the social recommender system,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 41, no. 4, pp. 746–759, 2011.

[89] A. Abe, K. Yamamoto, and S. Nakagawa, “Robust speech recognition using DNN-HMM acoustic model combining noise-aware training with spectral subtraction,” in *Proceedings of the 16th Annual Conference of the International Speech Communication Association, INTERSPEECH 2015*, pp. 2849–2853, Germany, September 2015.

[90] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, “An experimental study on speech enhancement based on deep neural networks,” *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 65–68, 2014.

[91] A. Narayanan and D. Wang, “Ideal ratio mask estimation using deep neural networks for robust speech recognition,” in *Proceedings of the 2013 38th IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2013*, pp. 7092–7096, Canada, May 2013.

[92] D. Serdyuk, K. Audhkhasi, and P. Brakel, “Invariant representations for noisy speech recognition,” *Computation and Language*, 2016, arXiv:1612.01928.

[93] P. E. Hart, “The condensed nearest neighbor rule,” *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 515-516, 1968.

[94] G. Gates, “The reduced nearest neighbor rule,” *IEEE Transactions on Information Theory*, vol. 18, no. 3, pp. 431–433, 1972.

[95] H. Brighton and C. Mellish, “Advances in instance selection for instance-based learning algorithms,” *Data Mining and Knowledge Discovery*, vol. 6, no. 2, pp. 153–172, 2002.

[96] Y. Li and L. Maguire, “Selecting critical patterns based on local geometrical and statistical information,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 6, pp. 1189–1201, 2011.

[97] F. Angiulli, “Fast nearest neighbor condensation for large data sets classification,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 11, pp. 1450–1464, 2007.

[98] F. Angiulli and G. Folino, “Distributed nearest neighbor-based condensation of very large data sets,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 12, pp. 1593–1606, 2007.

[99] M. I. Jordan, “Divide-and-conquer and statistical inference for big data,” in *Proceedings of the the 18th ACM SIGKDD international conference*, p. 4, Beijing, China, August 2012.

[100] T. G. Kolda and J. Sun, “Scalable tensor decompositions for multi-aspect data mining,” in *Proceedings of the 8th IEEE International Conference on Data Mining, ICDM 2008*, pp. 363–372, Italy, December 2008.

[101] G. Wahba, “Dissimilarity data in statistical model building and machine learning,” in *Proceedings of the 5th International Congress of Chinese Mathematicians*, pp. 785–809, 2012.

[102] S. C. Hoi, J. Wang, P. Zhao, and R. Jin, “Online feature selection for mining big data,” in *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, pp. 93–100, Beijing, China, August 2012.

[103] A. Sagheer, N. Tsuruta, R.-I. Taniguchi, D. Arita, and S. Maeda, “Fast feature extraction approach for multi-dimension feature space problems,” in *Proceedings of the 18th International Conference on Pattern Recognition, ICPR 2006*, pp. 417–420, China, August 2006.

[104] J. R. Anaraki and M. Eftekhari, “Improving fuzzy-rough quick reduct for feature selection,” in *Proceedings of the 2011 19th*

Iranian Conference on Electrical Engineering, ICEE 2011, Iran, May 2011.

[105] I. A. Gheyas and L. S. Smith, "Feature subset selection in large dimensionality domains," *Pattern Recognition*, vol. 43, no. 1, pp. 5–13, 2010.

[106] K. W. Lau and Q. H. Wu, "Online training of support vector classifier," *Pattern Recognition*, vol. 36, no. 8, pp. 1913–1920, 2003.

[107] P. Laskov, C. Gehl, S. Krüger, and K.-R. Müller, "Incremental support vector learning: analysis, implementation and applications," *Journal of Machine Learning Research*, vol. 7, pp. 1909–1936, 2006.

[108] C. Chang and C. Lin, "LIBSVM: a Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.

[109] K. Huang, H. Yang, I. King, and M. R. Lyu, "Maxi-min margin machine: learning large margin classifiers locally and globally," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 19, no. 2, pp. 260–272, 2008.

[110] A. Franco-Arcega, J. A. Carrasco-Ochoa, G. Snchez-Daz et al., "Building fast decision trees from large training sets," *Intelligent Data Analysis*, vol. 16, no. 4, pp. 649–664, 2012.

[111] H. Yang and S. Fong, "Incrementally optimized decision tree for noisy big data," in *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications (BigMine '12)*, pp. 36–44, Beijing, China, August 2012.

[112] Y. Ben-Haim and E. Tom-Tov, "A streaming parallel decision tree algorithm," *Journal of Machine Learning Research (JMLR)*, vol. 11, pp. 849–872, 2010.

[113] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.

[114] N. Liu and H. Wang, "Ensemble based extreme learning machine," *IEEE Signal Processing Letters*, vol. 17, no. 8, pp. 754–757, 2010.

[115] Q. He, T. Shang, F. Zhuang, and Z. Shi, "Parallel extreme learning machine for regression based on mapReduce," *Neurocomputing*, vol. 102, pp. 52–58, 2013.

[116] R. Zhang, Y. Lan, G.-B. Huang, and Z.-B. Xu, "Universal approximation of extreme learning machine with adaptive growth of hidden nodes," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 2, pp. 365–371, 2012.

[117] H.-J. Rong, G.-B. Huang, N. Sundararajan, P. Saratchandran, and H.-J. Rong, "Online sequential fuzzy extreme learning machine for function approximation and classification problems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 4, pp. 1067–1072, 2009.

[118] Y. Yang, Y. Wang, and X. Yuan, "Bidirectional extreme learning machine for regression problem and its learning effectiveness," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 9, pp. 1498–1505, 2012.

[119] W. X. Chen and X. Lin, "Big data deep learning: challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525, 2014.

[120] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *The American Association for the Advancement of Science: Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[121] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[122] D. C. Ciresan, U. Meier, and J. Masci, "Flexible, high performance convolutional neural networks for image classification," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 1237–1242, 2011.

[123] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Proceedings of the International Conference on Artificial Neural Networks*, pp. 92–101, 2010.

[124] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS '12)*, pp. 1097–1105, Lake Tahoe, Nev, USA, December 2012.

[125] J. Dean, G. Corrado, and R. Monga, "Large scale distributed deep networks," in *Neural Information Processing Systems*, pp. 1223–1231, 2012.

[126] G. Papandreou, L.-C. Chen, K. P. Murphy, and A. L. Yuille, "Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation," in *Proceedings of the 15th IEEE International Conference on Computer Vision, ICCV 2015*, pp. 1742–1750, Chile, December 2015.

[127] G. Hinton and R. Salakhutdinov, "Discovering binary codes for documents by learning deep generative models," *Topics in Cognitive Science*, vol. 3, no. 1, pp. 74–91, 2011.

[128] R. Socher, C. C.-Y. Lin, C. D. Manning, and A. Y. Ng, "Parsing natural scenes and natural language with recursive neural networks," in *Proceedings of the 28th International Conference on Machine Learning (ICML '11)*, pp. 129–136, Bellevue, Wash, USA, June 2011.

[129] R. Kumar, J. O. Talton, and S. Ahmad, "Data-driven web design," in *Proceedings of the International Conference on Machine Learning*, pp. 3-4, 2012.

[130] R. Raina, A. Madhavan, and A. Y. Ng, "Large-scale deep unsupervised learning using graphics processors," in *Proceedings of the 26th International Conference On Machine Learning, ICML 2009*, pp. 873–880, Canada, June 2009.

[131] J. Martens, "Deep learning via Hessian-free optimization," in *Proceedings of the 27th International Conference on Machine Learning (ICML '10)*, pp. 735–742, June 2010.

[132] K. Zhang and X.-W. Chen, "Large-scale deep belief nets with mapreduce," *IEEE Access*, vol. 2, pp. 395–403, 2014.

[133] L. Deng, D. Yu, and J. Platt, "Scalable stacking and learning for building deep architectures," in *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '12)*, pp. 2133–2136, Kyoto, Japan, March 2012.

[134] K. Kavukcuoglu, M. Ranzato, R. Fergus, and Y. LeCun, "Learning invariant features through topographic filter maps," in *Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, pp. 1605–1612, USA, June 2009.

[135] B. Hutchinson, L. Deng, and D. Yu, "Tensor deep stacking networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1944–1957, 2013.

[136] J. Zhang, "Review of wideband MIMO channel measurement and modeling for IMT-Advanced systems," *Chinese Science Bulletin*, vol. 57, no. 19, pp. 2387–2400, 2012.

[137] C. Liang, H. Li, Y. Li, S. Zhou, and J. Wang, "A learning-based channel model for synergetic transmission technology," *China Communications*, vol. 12, no. 9, pp. 83–92, 2015.

[138] J. Zhang, "The interdisciplinary research of big data and wireless channel: a cluster-nuclei based channel model," *China Communications*, vol. 13, no. supplement 2, Article ID 7833457, pp. 14–26, 2016.

[139] Y. Li, J. Zhang, Z. Ma, and Y. Zhang, "Clustering analysis in the wireless propagation channel with a variational gaussian mixture model," *IEEE Transactions on Big Data*, pp. 1-1, 2018.

[140] F. Bai, T. Vidal-Calleja, and S. Huang, "Robust incremental SLAM under constrained optimization formulation," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1–8, 2018.

[141] I. Z. Ibragimov and I. M. Afanasyev, "Comparison of ROS-based visual SLAM methods in homogeneous indoor environment," in *Proceedings of the 2017 14th Workshop on Positioning, Navigation and Communications (WPNC)*, pp. 1–6, Bremen, October 2017.

[142] U. M. Fayyad, *On the Induction of Decision Trees for Multiple Concept Learning*, University of Michigan, 1992.

[143] G. Cybenko, "Approximation by Superpositions of a sigmoidal function," *Mathematics of Control Signals & Systems*, vol. 2, no. 4, pp. 303–314, 1989.

[144] Z. Ma, P. K. Rana, J. Taghia, M. Flierl, and A. Leijon, "Bayesian estimation of dirichlet mixture model with variational inference," *Pattern Recognition*, vol. 47, no. 9, pp. 3143–3157, 2014.

[145] D. Naboulsi, M. Fiore, S. Ribot, and R. Stanica, "Large-scale mobile traffic analysis: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 124–161, 2016.

[146] E. Halepovic and C. Williamson, "Characterizing and modeling user mobility in a cellular data network," in *Proceedings of the PE-WASUN'05 - Second ACM International Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks*, pp. 71–78, Canada, October 2005.

[147] S. Scepanovic, P. Hui, and A. Yla-Jaaski, "Revealing the pulse of human dynamics in a country from mobile phone data," *NetMob D4D Challenge*, pp. 1–15, 2013.

[148] S. Isaacman, R. A. Becker, and R. Caceres, "Identifying important places in people's lives from cellular network data," in *Proceedings of the International Conference on Pervasive Computing*, pp. 133–151, 2011.

[149] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci, "Measuring serendipity: connecting people, locations and interests in a mobile 3G network," in *Proceedings of the 2009 9th ACM SIGCOMM Internet Measurement Conference, IMC 2009*, pp. 267–279, USA, November 2009.

[150] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[151] Q. Lv, Y. Qiao, N. Ansari, J. Liu, and J. Yang, "Big data driven hidden markov model based individual mobility prediction at points of interest," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5204–5216, 2017.

[152] Y. Qiao, Y. Cheng, J. Yang, J. Liu, and N. Kato, "A mobility analytical framework for big mobile data in densely populated area," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1443–1455, 2017.

[153] L. Meng, S. Liu, and A. Striegel, "Analyzing the longitudinal impact of proximity, location, and personality on smartphone usage," *Computational Social Networks*, vol. 1, no. 1, 2014.

[154] M. Böhmer, B. Hecht, J. Schöning, A. Krüger, and G. Bauer, "Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage," in *Proceedings of the 13th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '11)*, pp. 47–56, September 2011.

[155] D. Hristova, M. Musolesi, and C. Mascolo, "Keep your friends close and your facebook friends closer: a multiplex network approach to the analysis of offline and online social Ties," in *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, pp. 206–215, USA, June 2014.

[156] R. I. M. Dunbar, V. Arnaboldi, M. Conti, and A. Passarella, "The structure of online social networks mirrors those in the offline world," *Social Networks*, vol. 43, pp. 39–47, 2015.

[157] D. Hristova, M. J. Williams, M. Musolesi, P. Panzarasa, and C. Mascolo, "Measuring urban social diversity using interconnected geo-social networks," in *Proceedings of the the 25th International Conference*, pp. 21–30, Canada, April 2016.

[158] D. Hristova, A. Noulas, C. Brown, M. Musolesi, and C. Mascolo, "A multilayer approach to multiplexity and link prediction in online geo-social networks," *EPJ Data Science*, vol. 5, no. 1, 2016.

[159] Apache, "Apache software foundation," http://apache.org, 2017.

[160] M. Gerla, E.-K. Lee, G. Pau, and U. Lee, "Internet of vehicles: from intelligent grid to autonomous cars and vehicular clouds," in *Proceedings of the IEEE World Forum on Internet of Things (WF-IoT '14)*, pp. 241–246, March 2014.

[161] F. Yang, S. Wang, J. Li, Z. Liu, and Q. Sun, "An overview of internet of vehicles," *China Communications*, vol. 11, no. 10, pp. 1–15, 2014.

[162] K. M. Alam, M. Saini, and A. El Saddik, "Toward social internet of vehicles: concept, architecture, and applications," *IEEE Access*, vol. 3, pp. 343–357, 2015.

[163] J. D. Lee, B. Caven, S. Haake, and T. L. Brown, "Speech-based interaction with in-vehicle computers: the effect of speech-based e-mail on drivers' attention to the roadway," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 43, no. 4, pp. 631–640, 2001.

[164] C. Y. Loh, K. L. Boey, and K. S. Hong, "Speech recognition interactive system for vehicle," in *Proceedings of the 13th IEEE International Colloquium on Signal Processing and its Applications, CSPA 2017*, pp. 85–88, Malaysia, March 2017.

[165] D. Amodei, S. Ananthanarayanan, and R. Anubhai, "Deep speech 2: end-to-end speech recognition in English and mandarin," in *Proceedings of the International Conference on Machine Learning*, pp. 173–182, 2016.

[166] E. Variani, X. Lei, E. McDermott, I. L. Moreno, and J. Gonzalez-Dominguez, "Deep neural networks for small footprint text-dependent speaker verification," in *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2014*, pp. 4052–4056, Italy, May 2014.

[167] K. Chen and A. Salman, "Learning speaker-specific characteristics with a deep neural architecture," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 22, no. 11, pp. 1744–1756, 2011.

[168] L. Deng, G. E. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: an overview," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '13)*, pp. 8599–8603, IEEE, Vancouver, Canada, May 2013.

[169] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '13)*, pp. 6645–6649, May 2013.

[170] J. Meyer and K. U. Simmer, "Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction," in *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP. Part 1 (of 5)*, pp. 1167–1170, April 1997.

[171] R. C. Hendriks, R. Heusdens, and J. Jensen, "MMSE based noise PSD tracking with low complexity," in *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2010*, pp. 4266–4269, USA, March 2010.

[172] F. Seide, G. Li, and D. Yu, "Conversational speech transcription using context-dependent deep neural networks," in *Proceedings of the 12th Annual Conference of the International Speech Communication Association (INTERSPEECH '11)*, vol. 33, pp. 437–440, August 2011.

[173] H. Ze, A. Senior, and M. Schuster, "Statistical parametric speech synthesis using deep neural networks," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '13)*, pp. 7962–7966, IEEE, Vancouver, Canada, May 2013.

[174] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '13)*, pp. 8609–8613, May 2013.

[175] Y. Qian, M. Bi, T. Tan, and K. Yu, "Very deep convolutional neural networks for noise robust speech recognition," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 24, no. 12, pp. 2263–2276, 2016.

[176] K. Han, Y. He, D. Bagchi et al., "Deep neural network based spectral feature mapping for robust speech recognition," *INTERSPEECH*, pp. 2484–2488, 2015.

[177] B. Li and K. C. Sim, "Improving robustness of deep neural networks via spectral masking for automatic speech recognition," in *Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2013*, pp. 279–284, Czech Republic, December 2013.

[178] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "A regression approach to speech enhancement based on deep neural networks," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 23, no. 1, pp. 7–19, 2015.

[179] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1, pp. 19–41, 2000.

[180] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.

[181] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, and G. Penn, "Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '12)*, pp. 4277–4280, IEEE, March 2012.

[182] M. McLaren, Y. Lei, and L. Ferrer, "Advances in deep neural network approaches to speaker recognition," in *Proceedings of the 40th IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2015*, pp. 4814–4818, Australia, April 2014.

[183] C. Yu, A. Ogawa, M. Delcroix, T. Yoshioka, T. Nakatani, and J. H. L. Hansen, "Robust i-vector extraction for neural network adaptation in noisy environment," in *Proceedings of the 16th Annual Conference of the International Speech Communication Association, INTERSPEECH 2015*, pp. 2854–2857, Germany, September 2015.

[184] N. Li, M.-W. Mak, and J.-T. Chien, "Deep neural network driven mixture of PLDA for robust i-vector speaker verification," in *Proceedings of the 2016 IEEE Workshop on Spoken Language Technology, SLT 2016*, pp. 186–191, USA, December 2016.

[185] Z. Zhang, L. Wang, A. Kai, T. Yamada, W. Li, and M. Iwahashi, "Deep neural network-based bottleneck feature and denoising autoencoder-based dereverberation for distant-talking speaker identification," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2015, no. 1, p. 12, 2015.

[186] M. McLaren, Y. Lei, and N. Scheffer, "Application of convolutional neural networks to speaker recognition in noisy conditions," in *Proceedings of the Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[187] T. N. Sainath, B. Kingsbury, and B. Ramabhadran, "Auto-encoder bottleneck features using deep belief networks," in *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2012*, pp. 4153–4156, Japan, March 2012.

[188] Y. Shinohara, "Adversarial multi-task learning of deep neural networks for robust speech recognition," *INTERSPEECH*, pp. 2369–2372, 2016.

[189] N. D. Lane and P. Georgiev, "Can deep learning revolutionize mobile sensing?" in *Proceedings of the the 16th International Workshop*, pp. 117–122, Santa Fe, NM, USA, Feburary 2015.