Towards Deep Learning Prospects: Insights for Social Media Analytics

MALIK KHIZAR HAYAT[1], **ALI DAUD**[2,3], **(Fellow, IEEE), ABDULRAHMAN A. ALSHDADI**[4],
AMEEN BANJAR[4], **RABEEH AYAZ ABBASI**[5], **YUKUN BAO**[6], **(Member, IEEE),**
AND HUSSAIN DAWOOD[7]

[1]Department of Software Engineering, Foundation University Islamabad, Islamabad 44000, Pakistan
[2]Department of Computer Science and Software Engineering, International Islamic University at Islamabad, Islamabad 44000, Pakistan
[3]Department of Information Systems, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[4]Department of Information Systems and Technology, College of Computer Science and Engineering, University of Jeddah, Jeddah 23890, Saudi Arabia
[5]Department of Computer Science, Quaid-i-Azam University, Islamabad 44000, Pakistan
[6]School of Management, Huazhong University of Science and Technology, Wuhan 430064, China
[7]Department of Computer and Network Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah 23890, Saudi Arabia

Corresponding author: Malik Khizar Hayat (khizar.malik@fui.edu.pk)

ABSTRACT Deep learning (DL) has attracted increasing attention on account of its significant processing power in tasks, such as speech, image, or text processing. In order to the exponential development and widespread availability of digital social media (SM), analyzing these data using traditional tools and technologies is tough or even intractable. DL is found as an appropriate solution to this problem. In this paper, we keenly discuss the practiced DL architectures by presenting a taxonomy-oriented summary, following the major efforts made toward the SM analytics (SMA). Nevertheless, instead of the technical description, this paper emphasis on describing the SMA-oriented problems with the DL-based solutions. To this end, we also highlight the DL research challenges (such as scalability, heterogeneity, and multimodality) and future trends.

INDEX TERMS Social media data, dynamic network, deep learning, feature learning.

I. INTRODUCTION

With the dawn of SM data over the globe, the rate of growth of data-intensive problems has also been increased. The wide availability and exponential growth of digital data have made it almost challenging or even unendurable to be visualized, explored, managed, and analyzed by means of contemporary software tools and technologies. The abundant increase in data *volume*, the diversity in data *variety*, and the incoming/outgoing data *velocity* (the concept known as 3Vs) are the most prominent reasons, why and how the SM data escalated. For instance, more than 1K Petabytes of data per day is being processed over the Internet, according to a report by the National Security Agency. In 2006-2011, digitized data has grown nine times and by 2020, it will touch 35 trillion gigabytes [1]. This enormous digital data intensification unlocks significant research prospects for various sectors such as education, health, industrial sector, business, public administration, scientific research and so on. In addition, the emergence of SM also directs to a sensational paradigm shift in the current scientific research on the road to data-driven knowledge discovery.

Even though the SM has a key role in connecting people around the globe [2], it offers a vast variety of knowledge extracting tasks as mentioned earlier. Pulling out the information from data and reaping knowledge from this information is not even yet a trivial problem to solve. Machine learning techniques, accompanied by the advances in existing computing power, played an important role to leverage hidden information in this data. However, its enormousness and diversity invite such a solution to the problem, which would better be able to depict the obscured information and knowledge from the data. As an active sub-area of machine learning, DL is believed to be a powerful tool to deal with SMA problems. Apparently, together with other SM applications, web-based applications are increasing day-to-day as recent hotspots [3]. Significantly, it includes social computing such as online communities, reputation systems, question-answering systems, prediction systems, recommender systems, and Heterogeneous Information Network Analysis (HINA) [4]. Moreover, the graph theory illustrates better the semantic structure of SM data which represents the users as *nodes* and relationships among them as *links*.

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

SM data is enormously increasing every day which requires refined patterns and features extraction for healthier knowledge discovery. Most of the conventional learning methods use shallow-structured learning architectures. However, DL discusses supervised or unsupervised machine learning techniques which automatically learn the hierarchical representations for classification. Recently, DL has appealed significant consideration from the research community because of the inspiration from biological observations on human brain processing. Also, it has performed predominantly in numerous research areas such as digital image processing, speech to text, and collaborative filtering. Likewise, the DL has been applied in engineering and manufacturing products; successfully that get facilitated of the huge-volume digital data. Certainly, the well-known companies, such as Google, Apple, and Facebook have to deal with the heaps of data on a daily basis. Moreover, these companies are eagerly posing DL-oriented projects. For instance, using DL, Apple's Siri, an application in the iPhone works as a virtual assistant, proposes widespread services counting news about sports, answers to user's questions, reports for latest weather updates, and reminders. While Google applies DL to the multitudes of chaotic Google Translator's data.

In contrast, the SMA is one of the paramount, hot, and recent areas of study. Connecting DL with SMA can reveal evocative insights. Formerly, a number of reviews [5]–[7] show that DL is viable and efficient to solve substantial big data problems. However, most of the focus was on DL applications, for instance, image classification and speech recognition. None of them discussed the most paramount and developed SM platform, in particular. DL methods and applications [8] feast through the zeal of application domains including business, education, economics, health informatics etc. In this review article, we cover some noteworthy application domains of SM such as user behavior analysis, business analytics, sentiment analysis, and anomaly detection where DL has played a striking role to leverage rich knowledge. In terms of contribution, the following are the fundamental insights this study seeks to answer:

1) Provides a contemporary summary of existing DL methods that can exhibit a roadmap to extract useful insights for SMA.
2) Provides a classification scheme that identifies important features to study semantics of the particular problem which may be helpful for designing better future vision in diverse SMA application domains.
3) Investigates the pros and cons of existing techniques.
4) Enlightens the prominent application domains for applying DL.
5) Uncovers the noteworthy research challenges and future directions.

Following the introductory section, Section 2 describes relevant concepts and terminologies about SM and DL used in the study. Section 3 illustrates the taxonomy of methods; however, each taxonomic group is further sub-categorized based on different techniques in selected application domains. Section 4 describes the DL perspectives for benchmark datasets. Section 5 discusses the performance evaluation measures for DL models. Section 6 highlights how the SM is challenging the research community. Additionally, this section deliberates intuitive research challenges and future directions whereas Section 7 concludes the study.

II. BASICS OF SOCIAL MEDIA AND DEEP LEARNING
Before diving into the article details, we begin with an overview of the basic concepts, terminologies, data types, and architectures concerning SM and DL.

A. SOCIAL MEDIA
Here we explain some mostly practiced terminologies and concepts pertinent to the SM.

1) TERMINOLOGIE
a: SOCIAL MEDIA ANALYTICS
SMA is the web-based technologies used to transform the communications carried over virtual networks and communities into the interactive discussion. The interactive Web 2.0 Internet-based applications are composed of content generated by the user, such as comments or text posts, videos, and data generated through all online communications. Generally, users generate service-specific SM profiles that are governed by some organization. SM facilitates the growth of online social networks by linking profile of a user with other people or groups usually consuming similar interests. Accordingly, SMA is the approaches to assemble data from different SM platforms, for instance, Facebook, Twitter, etc., then evaluate and analyze the data to make business decisions. Significantly, this data is updating, expanding, and evolving, unceasingly [9] – perhaps, a good pathway to comprehend real-time customer experiences, intents, and sentiments.

b: SOCIAL NETWORK ANALYSIS
The social network lies in the subcategory of SM where two users are connected with a common interest. The individuals and groups are the nodes while the edges show connections between the nodes. Social Network Analysis (SNA) is the drawing and determining associations and drifts amongst individuals, groups, organizations, URLs, and other interlinked information entities using networks and graph theory [10]. It also helps to analyze human relationships both graphically and mathematically.

Comparatively, the SMA entitles to the Business Intelligence (BI) tools – reporting, searching, visualizing, text mining and so on – applied to the information sourced from SM platforms, for instance, Facebook and Twitter. It helps us to answer the questions such as, how much and from where you are driving the traffic, how much influential is your messaging etc. However, the SNA is explicitly focused on identifying the relationships, connections, interactions, influence amongst information engineering entities such as individuals, groups, organizations and so on. It helps us to

answer the questions – how closely an individual is linked to a network, how information drift occurs within a network etc. In particular, we investigate the former here in this study.

c: BIG DATA

The datasets which are so voluminous, diverse, rapid, and complex than the conventional data-processing and computational systems are unable to deal with, are termed as the Big Data. In order to store, process, and analyze such data entails plausible tools and techniques which can well-encompass the underlying data. Moreover, SM platforms, for instance, Facebook, Twitter, and YouTube are causing swift intensification in the data. The Big Data from SM can be used to extract trends, patterns, and associations particularly pertinent to human behavior, entities interactions, and complex integrations. Certainly, the Big Data is interwoven with the explicit 4Vs concepts, meaningfully, Volume, Variety, Velocity, and Veracity [5]–[7].

d: DYNAMIC NETWORK

Users linked together by friendship links, make a network of users at Facebook. This network varies over time by several online activities, for instance, adding or removing the friends, liking or disliking the products, joining or leaving the groups, and so on. In a dynamic network, nodes and edges vary over time. Statistical methods, computational investigation, or computerized simulations are often essential to discover how these networks grow, shrink, fine-tune, or deal with the peripheral interferences.

e: SIGNED SOCIAL NETWORK

People hold both kinds of sentiments against each other – positive and negative. With the advancement of SM platforms, individuals frequently prefer to express their sentiments using these platforms. The sentiments such as friends or foes, agreements or disagreements, likes or dislikes, trust or distrust, group joining or departing, and so on can be bi-categorized into positive connections such as friends, agreements, or likes and negative connections such as foes, disagreements, or dislikes. Such interactions yield to the development of Signed Social Networks (SSNs) [11], [12]. Certainly, SM signed networks are noisy and sparse usually with a massive number of users and multitudes of relationships. SM illustrations of SSNs comprise friends/foes in Slashdot[1] [13], and trust/distrust in Epinions[2] [14].

2) CONCEPTS
a: INFORMATION OBJECT

Real-world networks can be better indicated as graphs. The graph comprises vertices and relationships among vertices. For instance, two people (vertices) can be related to a network if they are connected via a friendship link in the

---

[1]http://www.slashdot.org/

[2]http://www.epinions.com/

Facebook graph. Real-world entities or nodes in these graphs are denoted as information objects [15].

b: DOMAIN ADAPTATION

In business, users buy and sell different products. Some buying and selling purely depend upon users' reviews, ratings etc. However, users' interested domains could vary among multiple types. Partaking a separate system for classifying sentiments for each domain would not likely be serviceable in terms of heavy cost, higher resources requirement, and more processing time as well. In order to learn a single system from the set of different domains, a substitute approach is used which solely depends upon the labeled or unlabeled data. Subsequently, the learned system is applied to any labeled or unlabeled target domain. Being a multi-source system, it is unviable to extract patterns that are not shared and meaningful across domains. The problem of learning systems on diverse domain distributions is termed as domain adaptation [16].

c: SENTIMENT ANALYSIS

The process of finding subjective information hidden in the users' content is known as sentiment analysis. Particularly, the content is in the form of text sourced from multiple SM platforms. Thus, it uses Natural Language Processing (NLP), computational linguistics, and other text analysis techniques to categorize the users' attitude. Meaningfully, it uses data mining techniques to extract and gather data for analysis. On purpose, this analysis is to determine the subjective knowledge from a piece of text or a corpus such as news articles, reviews, comments, blog posts, SM news feeds, tweets, and status updates [17].

d: LINK PREDICTION

Link prediction is the data-analysis method used to estimate the relationships between nodes in the SM networks. The nodes could be people, organization, or transactions. SM platforms are highly dynamic where nodes and edges are constantly evolving. In an SM network, the new interactions which are yet expected to befall, the problem is denoted as link prediction [18].

3) TYPES OF SOCIAL MEDI
a: MICROBLOGGING

A blog is an informal-adapted discussion often depicted in the written, image, audio, or audiovisual forms on the World Wide Web. However, a microblog is a precise form of a blog which allows smaller content to be shared online as compared to a regular blog. It lets a user post slight content perhaps, small sentences, images, short length videos, or connections to the web pages. In order to keep the blog updated, bloggers usually use a number of services such as instant messaging, emailing, etc. These services are termed as *microposts* whereas bringing them in practice to keep the blog up-to-date is called microblogging [19]. SM websites such as Weibo, Twitter, Facebook also provides microblogging services like

status updates and tweets [20]. Some microblogging services allow users to regulate their privacy settings that who can read or comment on their microblogs. Microblogging applications spread in various domains such as e-commerce, online marketing, product searching, and advertising sales.

b: FRIENDSHIPS

Two people connected with each other using any online medium construct a social platform termed as friendship network. People are consuming tons of their energy being bridged with others on SM platforms. The friendship network members may or may not focus on a particular topic rather put more attention on remaining linked to their friends. Licamele and Getoor [21] considered these networks as friendship-event networks. For instance, in an academic collaboration network, the researchers are actors/friends, collaborations are friendships, and conferences are events. Moreover, the noteworthy online friendship platforms include but not limited to Facebook,[3] MySpace,[4] Badoo,[5] and Bebo.[6]

c: PROFESSIONAL

A professional network or particularly professional network service is a category of social network service that is concentrated merely on communications and associations of a corporate/professional nature [22]. Instead of counting private, non-corporate or unprofessional relationships, the focus is to either find a job or move a step ahead in professional career [20], [22]. Some substantial professional platforms include LinkedIn,[7] Xing,[8] and Data.com Connect.[9]

d: PHOTOS

Photo sharing platforms allow users to share photos online publicly or privately. Sharing refers to viewing the images but not necessarily download them [20]. Objectively, these sites allow users to backup images, make the images searchable, share images, and even give them the control, who can see their shared images. These platforms could also be used for multi purposes such as image resourcing and repository, visual literacy, and research purposes. Some substantial photo sharing platforms include Flickr,[10] Instagram,[11] and Pinterest.[12]

e: VIDEOS

Video sharing platforms help users to share their videos online publicly or privately. Some of the video sharing websites allow users to upload short length videos, however,

others allow lengthy video content as well. Objectively, these sites allow users to save, share, comment, and control viewership of their shared content. Online SM platforms are one of the extensively used media to share video content, these days. These platforms also offer live streaming features to exclusively get connected with other users. Moreover, they also allow users to link their existing SM accounts such as Facebook with the video-sharing websites to instantly share videos [23]. Some substantial video sharing platforms include YouTube,[13] Vimeo,[14] and Veoh.[15]

f: QUESTION/ANSWER FORUMS

Question/Answer (Q/A) forums tend to answer the questions asked by different users across multiple domains. Users are able to see the questions asked by other users if they help them, and similar is the case with posted answers usually by domain experts. Q/A forums are often fused by huge and professional corporations and inclined to be applied as a community which lets users in alike domains discuss questions provided with expert answers. In social Q/A services, any user can ask a question and also post an answer. Shah *et al.* [24] divided these services into three categories – *digital reference services*, where users can ask for help from librarians, *expert services*, where organizations offer Q/A service and experts are supposed to answer the questions, and *social or community Q/A*, where everybody can ask and answer a question. Sharing knowledge over online communities helps users to encourage self-presentation, peer recognition, and social learning [25]. Some substantial Q/A forums include StackExchange,[16] Quora,[17] Answers.[18]

B. DEEP LEARNING

Here we explain significantly practiced concepts and terminologies pertinent to the DL.

1) TERMINOLOGIE
a: NEURAL NETWORK

Neural Network (NN) or Artificial Neural Network (ANN) is a paradigm to process information specifically fortified by the biological neurons (like in human nervous systems). It attempts to find fundamental relationships in datasets by using the series of processor that imitates the way the human brain works. The significance of this paradigm – it comprises numerous extremely interrelated processing elements (artificial neurons) operating mutually to solve particular problems. A neural network usually encompasses numerous tier-wise arranged processors operating in parallel. Raw input is injected to the first tier, whereas the output from each preceding tier is injected as input to the succeeding tier. Eventually, the last tier generates the final output [26].

---

[3] http://www.facebook.com/

[4] http://www.myspace.com/

[5] http://www.badoo.com/

[6] http://www.bebo.com/

[7] http://www.linkedin.com/

[8] http://www.xing.com/

[9] http://connect.data.com/

[10] http://www.flickr.com/

[11] http://www.instagram.com/

[12] http://www.pinterest.com/

[13] http://www.youtube.com/

[14] http://www.vimeo.com/

[15] http://www.veoh.com/

[16] http://www.stackexchange.com/

[17] http://www.quora.com/

[18] http://www.answers.com/

In contrast to DL, training in neural networks always happens simultaneously for all layers.

b: CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN) is encompassed of numerous layers with few feature representation layers and other layers as in a typical neural network. Frequently, it starts off with two varying layer types namely; convolutional and sub-sampling. Nonetheless, the former layers accomplish convolution processes and the later layer reduces the size of earlier layers [1]. Owing to fewer parameters, training CNNs is rather easier than more fully linked networks. Moreover, with a fixed size input, CNNs generate absolute outputs which diminish the cost of preprocessing.

c: RECURRENT NEURAL NETWORK

Recurrent Neural Network (RNN), a class of ANN where links between components construct a directed graph. It lets an RNN parade a progressive dynamic behavior. Distinct from feedforward NNs, internal memory of RNNs can be used to process random input sequences. This establishes them appropriate for tasks such as joining handwriting and even speech to text conversion. Moreover, in RNNs, every node is connected to another node with a directed connection whereas nodes could be input, output, or hidden nodes that adjust the data between input and output nodes [7].

d: AUTO-ENCODER

An Auto-Encoder (AE), also known as auto-associator, or Diabolo network is an ANN used for unsupervised learning. The Auto-Encoder is aimed to learn an encoding (usually a representation) for a dataset. The learned representation can then be used for dimensionality reduction. Recently, the auto-encoder concept has become extensively used for generative models learning. Constructively, the modest form of and auto-encoder is a feedforward, non-RNN with a similar quantity of nodes in the input and output layer. Accordingly, an auto-encoder is an unsupervised learning model with binary parts as encoder and decoder [27].

e: RESTRICTED BOLTZMAN MACHINES

A Restricted Boltzman Machine (RBM) is a neural network with binary layers. It comprises a single non-connected hidden units' layer and connected visible units' layer. Additionally, the connections between hidden and visible units are symmetrical and undirected. There is a bias for hidden together with visible units in the network. Binary values are used for hidden and visible units (0 for hidden and 1 for visible). The applications of RBM lie around classification, dimensionality reduction, feature learning, collaborative filtering, and topic modeling [28].

f: DEEP BELIEF NETWORK

Deep Belief Network (DBN), a class of deep NN in the machine learning, or rather a generative graphical model, encompassed of various hidden layers with latent variables.

Within each layer, the connections are between the layers but not between units. When unsupervised training is performed, a DBN reconstructs its inputs probabilistically. The layers then act as feature extractors. Next to learning, to perform classification, a DBN can be further dealt with supervised training. DBNs can be observed as an alignment of merely unsupervised networks, for instance, RBM or Auto-Encoders, where each hidden layer of sub-network acts as the visible layer for the next [29].

2) CONCEPTS

a: DEEP LEARNING

DL deals with a collection of machine learning methods that train several levels of data representations in deep architectures. The learning could be supervised, semi-supervised, or unsupervised [30]. It uses a multiple-layered cascade of nonlinear processing units (neurons) for feature transformation and extraction. The output from each preceding layer is used as input for the succeeding layer as shown in FIGURE 1; x_1, x_2 and so on are the inputs, w_i up to w_n are weights where $i = 1, 2, \ldots, n$ and y is the output. Each gray circle is a neuron which processes input based on an activation function (described later in this sub-section). The neurons exchange messages between each other in a complexly interconnected schema. The links have adjustable numeric weights (w) based on training. Eventually, these layers form a hierarchy of concepts. DL is composed of more hidden layers while neural networks consist of up to 3 layers most of the time. The training is occurred in 2 parts for instance, after training layer 1, the output of layer 1 would then forward to the next layer. Finally, the output layer demonstrates the learnt representations [31].

b: ACTIVATION FUNCTION

The activation function of a node (neuron) describes its output given the input or a set of input values. These functions are an extremely important feature for Dynamic Neural Networks (DNNs) [32]. Fundamentally, these functions decide whether the neuron should be activated or not. Also, whether the incoming information for the neuron is relevant or ignorable. It is the non-linear transformation done over the input signals. The result of this transformation is then set to the next layer of neurons as input. Mathematically, the activation function can be depicted as:

$$Y = b + \sum_{i=1}^{n} x_i w_i \qquad (1)$$

where Y is output, b is bias factor which minimizes the loss function over the training set, x is the input, and w is the connection weight.

c: NETWORK EMBEDDING

Network embedding is a technique to learn the vertex representations of low-dimensions in networks [33]. It drives to demonstrate the data into a low-dimensional hidden space.
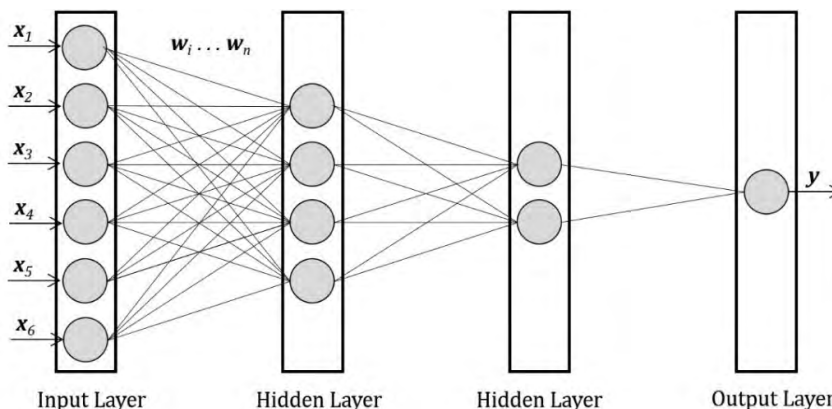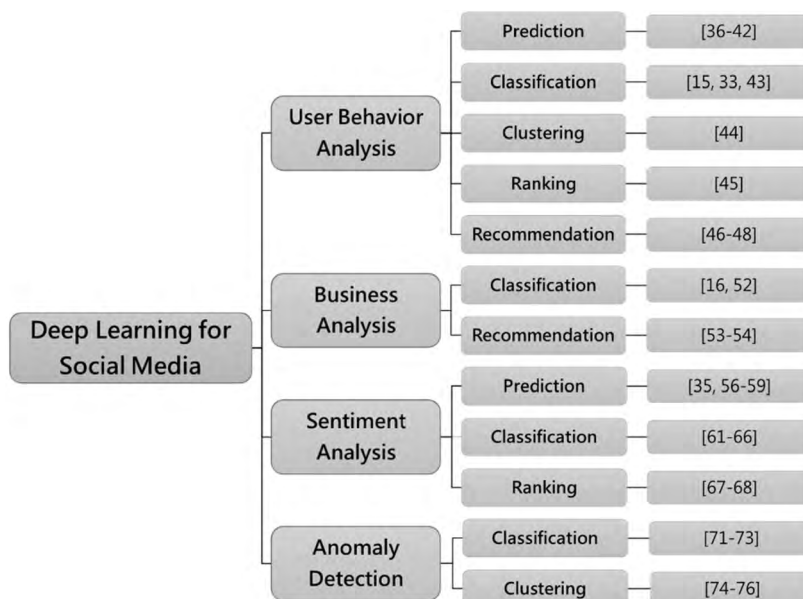
FIGURE 1. Deep learning.

FIGURE 2. Taxonomy of DL methods.

Each vertex in this space is denoted as a low-dimensional vector which helps in the direct computational processing of the network. For embedding a number of networks, it is essential to preserve the network structure both locally and globally [34].

d: WORD EMBEDDING
Applying DL to resolve NLP problems aims to obtain well-distributed words representations, particularly the words from vocabulary are mapped to real-valued vector representations. These vector representations are known as word embeddings [35]. Theoretically, it encompasses a mathematical embedding that revolves around a space opening at one dimension per word and ending at continuous and substantially low-dimensional vector space. Using these representations as input, these embeddings can boost the performance of NLP tasks, for instance, sentiment analysis, syntactic parsing, and NLP.

III. DEEP LEARNING IN SOCIAL MEDIA
Connectionism is the philosophy of DL while an individual biological neuron or feature in a deep model is not intelligent. However, a large collection of connections related and acting together can subdue intelligent behavior. In this section, we discuss the problem domains of SM where DL has been used as a key technique to the problem solutions. Next, we discuss the pros and cons of existing methods. To end, we walk through the taxonomic details of these methods and provide a comprehensive summary of the DL methods for SMA. FIGURE 2 shows the taxonomy of the DL methods in SMA.

A. USER BEHAVIOR ANALYSIS
The current living society is a combination of various entities whereas human beings are one of them. Intuitively, the behavior of human beings can be majorly categorized into – individual behavior and group behavior. Both have their

own causes and consequences. However, human beings as users in society, behave differently in different social situations. Social behavior is the outcome of certain atmospheric changes, environmental events, or social influences. To obtain the knowledge of the well-being of a society, along with knowing the social changes, it is equally important to become familiar with the social behavior of individuals.

Moreover, it is worthy to determine the impact of social influences on users' behavior. As defined earlier, the SM is a prominent source of connecting people in society and primarily count on user-generated content. Accordingly, the DL offers captivating techniques to analyze users' behavior, learning correlations between their past and current characteristics based on SM. Here we go through some categorized tasks performed in SM to analyze users' behavior using DL.

1) PREDICTION USING DL

A number of studies exist which used DL to predict human behavior in social networks. Aramo-Immonen *et al.* [36] learned communities using data from Twitter which is considered as one of the prominent tools for information dissemination. DL could handle multi-dimensional data quite effectively. Aiming which, Zhang *et al.* [37] proposed a Tensor Auto-Encoder (TAE), deep computational model, to learn features from the heterogenetic YouTube data. Given a reference basis of vectors, a tensor is used to represent the linear relation between vectors. Arrays are one of the ways to represent tensors in computer memory. The dimensions of the array make the degree (rank) of a tensor. For instance, a 2D array can be used to represent a linear map between vectors, hence, a 2nd-order tensor. In order to represent the input data and the representations in all layers, the TAE extends the conventional DL model to high-order tensor space by means of tensors. In this model, tensors are used to fuse the learned heterogeneous data features inside the hidden layer. It benefits the tensor DL model to apprehend the multifaceted relationships of the input data. To train the TAE model, the authors designed a high-order back-propagation algorithm – an important tool to improve the prediction accuracy. However, in comparison with using homogeneous data on TAE, the heterogeneous data took more time as it requires more iterations to train the parameters.

SM data comprises a wealth of valued information usable to make noteworthy predictions. Apparently, heterogenous source learning is yet a non-trivial task in SM. Jia *et al.* [38] fusing social networks, proposed a novel deep model, Fuses sociAl netwoRks uSing dEep lEarnING (FARSEEING) which integrated the useful information from heterogeneous social networks. Particularly, this is an information fusion task which acquired DL to learn from the complexity of multiple data sources. In this model, the authors used different inner layers to learn complex representations from multiple social networks. First, the users are associated by relating their multiple social forum accounts. Second, the given users are characterized using extracted multi-faceted features such as linguistic, demographic, and behavioral features.

Apparently, the activities of users in SM are unbalanced which cause the data missing. This missing data is inferred before feeding the extracted features into FARSEEING using Non-negative Matrix Factorization (NMF). NMF is a set of algorithms in the multivariate analysis where a matrix M is factorized into two matrices X and H, without any negative element in all three matrices.

The low-level features are mapped into high-level features using deep layers then, the high-level features are fused together for learning the task. The users' confidence level and consistency among multiple social forums are measured to gain a comprehensive understanding of a user's interest, behavior and personality traits. The data sources used as a ground truth are Quora, About.me, and LinkedIn.

Social networks are the source of creating online relationships among users. Labeling these relationships as negative or positive drive these networks to SSNs. Liu *et al.* [39], [40] proposed deep belief network-based (DBN) techniques to predict links in SSNs. The prediction tasks such as co-authorship, friendship, trust, distrust, and further associations are considered. It is obvious that social networks have escalated usage rate for communication in a number of crises situations, these days. Lazreg *et al.* [41] presented a strong analysis of SM posts such as text, images to predict the crises situations communicated by users. The various SM posts are typically informal, brief, and heterogeneous (a mixture of languages, acronyms, and misspellings) in nature. Without loss of generality, identifying the context of the post is often necessary to infer its underlying meaning. Moreover, the posts on other ordinary events are also the data part, which grants supplementary training noise. DL better understands these complex representations to learn the crises situations. In addition, Liu and Zhu [42] used microblog data to predict users' behavior by proposing an unsupervised drawing of the Linguistic Representation Feature Vector (LRFV). This method could comprehensively and more objectively describe users' semantic information.

2) CLASSIFICATION USING DL

SM is one of the most prominent modes of interaction amongst people in which they produce, exchange and share ideas and information in networks and communities. In general, SM data is noisy, diverse, of low quality, in large quantity, and heterogeneous in nature. In order to record routine activities, users with diverse background practice SM platforms. This influences the SM data to be subjective. It also gives these data a wide collection of attributes such as the resources used, the appearance of entities in a specific context, information diffusion, link analysis, and so on. For instance, SM tasks such as image annotation and classification are non-trivial because of the diversity characteristic.

However, the SM is heterogeneous with multi-modal user-generated content which inspires to have joint representation for the data. For instance, a flower image could be associated with a number of textual tags which make latent feature learning for image classification quite complicated.

A joint representation could better deliver the information associated with the content. Yuan *et al.* [15] proposed a DL-based approach to classify SM data, particularly images, using latent feature learning. The authors used Flickr dataset and classified the images as linked and not linked with a tag which somehow articulated this as an image classification as well as link analysis problem. Apparently, dealing with such an enormous feature space is a non-trivial task, rather, DL could be a worthy tool to handle image data. The reasons include the unsupervised pre-training of diverse social data characteristics, fine-tuning of features, layer-wise learning structure, and the explanation of more abstract and robust semantics.

Yuan *et al.* [15] presented a Relational Generative Deep Belief Nets (RGDBN) model and investigated links between information objects which are generated by interactions of latent features. Initially, low-level representations are learned in RGDBN, then using a deep architecture with more layers, higher-level representations are used to learn better the links between images and associated textual tags. The authors believe that integrating the collective effect of latent features learning into deep model could better represent the diverse and heterogeneous data space. Keeping in view, learning useful network representations, Wang *et al.* [33] presented a Structural Deep Network Embedding (SDNE) model, to efficiently capture the enormously non-linear structure of the complex networks. In particular, it is a semi-supervised deep model with several layers of non-linear functions. The multiple deep layers in SDNE allowed this model to seizure the tremendously non-linear heterogeneous network structure. The objective of network embeddings is to learn complex representations for heterogeneous networks.

Using social networks, people share multi-typed diverse data. However, it is unlikely that users share their personal data, for instance, gender, birth year, demographics etc. User behavior prediction entails classifying the users on the basis of their age groups which can reveal valuable insights about user behaviors among different age groups. Guimarãs *et al.* [43] analyzed 7000 sentences from social networks. They used Deep Convolutional Neural Network (DCNN) to classify social network posts features such as hashtag, retweet, characters in a tweet, number of followers, number of tweets etc. After extensive experimentation using different machine learning algorithms, for instance, Random Forest, Decision Trees, Support Vector Machine (SVM) etc., they found that DCNN outperformed the counterparts in terms of large-scale data classification. Guimarãs *et al.* [43] also proposed enhanced Sentiment Metric (eSM) which could classify the users by age, who restrict their personal information.

3) CLUSTERING USING DL

In SM data, community detection is a realistic solution to determine the intrinsic grouping of information objects (defined under Social Media Concepts). In order to group information objects, diverse attributes may have varying contributions. The attribute values have pairs of interest with influence on the grouping task. For instance, the degree of qualification – a worthy attribute in an attempt to group users with corresponding institutions.

In social networks, Zin *et al.* [44] integrated clustering with ranking and proposed a novel deep model, Deep Learning Cluster Rank (DeepLCRank). This method better illustrates the ranked clusters in social networks. For each item of the cluster – there is a rank assigned on the basis of learned features of information objects in the network. The varied information objects in social networks form much complex representation where DeepCLRank could handle this quite effectively.

4) RANKING USING DL

Individuals tend to use social media as a marker to solve their problem by posting queries. These forums are known as Community Question Answer (CQA) forums. These help users to have satisfactory information. Nevertheless, it is not very likely that users would be able to get the desired content in fractions of time because there exists a lot of answers to the same questions at CQA. It necessitates ranking the answers provided by experts at CQA.

Chen *et al.* [45] proposed an approach to predict users' personalized satisfaction using multiple instance DL frameworks. The authors presented a novel model, Multiple Instance Deep Learning (MIDL) framework to predict personalized user satisfaction. In CQA, a single question could have a number of answers. The known aspect is one of the answers is assigned a *satisfied* tag, however, the known aspect is what exactly the answer is with *satisfied* tag. This situation intrigues to learn multiple instances for answers. Each answer to a question at CQA is considered as an instance in a bag where each question resolution acquires one satisfactory answer on the Stack Exchange dataset. In terms of the historical behavior of users, common user space is defined and initialized for the representation of each individual user. Subsequently, after extracting features, all are injected into deep recurrent neural networks to rank them as positive or negative.

5) RECOMMENDATION USING DL

SM data is a promising source of the incessant recommendation of pertinent content to multi-domain users. The influence of recommendation could be exaggerated if items from different domains are jointly learned and getting recommended. Elkahky *et al.* [46] proposed a Multi-View Deep Neural Network (MV-DNN), which mapped items and users into a semantic space which is shared and recommends items with capitalized similarity. For instance, people who visit espn-cricinfo.com would most likely to see news about cricket, like to play games relevant to cricket on PC or Xbox. The authors used several data sources such as Microsoft product logs such as Bing search log, Windows Store download history log, or movie view logs from Xbox to make an interest-oriented recommendation.

DNN is used to incorporate the high-dimensional to the lower-dimensional features space of users and items from different domains. Social network users often belong to different domains, hence always look for items of their interests. MV-DNN has the ability to recommend items based on categorical features such as movie genre, application category, country or region the item belongs to, and so on.

Collaborative Filtering (CF) is also a famous style of recommending the appropriate content to users. CF-based methods typically use users' ratings to recommend pertinent items to them. However, the sparsity of ratings causes noteworthy deprivation in recommendation performance. Wang *et al.* [47] proposed a Collaborative Deep Learning (CDL) model which assimilates the learning of deep representations for the content information (items) and collaborative filtering for users' ratings. The authors used diverse data domains such as Citeu-Like, Netflix, and IMDB to recommend items to users.

In addition, users' trust also plays an important role in finding trustworthy recommendations. Deng *et al.* [48] proposed Deep Learning based Matrix Factorization (DLMF) model to synthesize the interests of users and links of their trust. DLMF performed better in terms of recommendation accuracy with unusual data and cold-start users as well. By using Epinions data, the authors used an autoencoder to learn the users' initial feature vectors and items in the first phase, whereas final latent feature vectors are learned in the second phase. This method could also work for trusted-community detection. TABLE 1 shows the summary of DL methods for user behavior modeling.

B. BUSINESS ANALYSIS

With the dawn of SM such as social networks, blogs, review forums, ratings, and recommendations are swiftly thriving. To automatically filter them out is very critical for businesses tending to sell their products and recognize new market prospects. However, the large-scale [49], [50] social data makes it tough to classify users' sentiments automatically. For instance, reviews from two different domains would contain different vocabulary which surges different data distributions for diverse domains. Consequently, domain adaptation can play a transitional role to learn intermediate representations.

1) CLASSIFICATION USING DL

Using SM platforms, people used to manage customer relationship [51], and make hotel decision for outing/eating [52]. Actually, SM platform such as Facebook, Twitter are now an extensive source of input and highly valuable for marketing research corporations, public view associations, and other text mining units. It drags such entities to spend more on SM to gain more business [51], [53].

Glorot *et al.* [16] presented a DL approach for domain adaptation of sentiment classifiers. DL techniques learn the intermediate concepts between the source and target data. The domain adaptation drift enables DL to learn meaningful intermediate concepts such as product price or quality,

customer services, customer reviews about products and so on. Based on the features revealed at the preceding level, the features are learned, level-wise. Additionally, Amazon is a widely used platform for business where DL leveraged better learning representations across all domains.

The authors used Amazon dataset in [16] with reviews from domains together with books, kitchen, electronics, and DVDs. For feature extraction phase, the Stacked Denoising Autoencoder (SDA) is compared with multi-layer perceptron (MLP). The two variations of SDA-1 with one layer and SDA-3 with three layers are used in the comparison.

The MLP performance illustrated that being non-linear supports to extract information, however, is not adequate to accumulate all necessary information from data. It is more passable to use an unsupervised phase which can incorporate data from diverse domains. Obviously, on this wide-ranging problem, a single layer does not suffice to grasp optimal performance. The stacking of three layers together returns the best representation from data. It is worth noticing that the representation learned by $SDA_{sh}3$ is significant for diverse domains and is thus accurately tailored to domain adaptation. TABLE 2 shows the summary of DL methods for business analysis.

Ding *et al.* [54] presented a CNN-based model to classify users based on their products need that is expressed on an SM platform. A product consumed by a certain user is more likely to get endorsed, subsequently. From a range of products, the products' consumption makes it significant to classify products that are more likely to be consumed by the consumers. The proposed CNN-based product consumption intention model can better classify the words of intention from the text as compared to the SVM along with word embeddings or bag-of-words.

2) RECOMMENDATION USING DL

On account of the upsurge practice of SM, people tend to buy attires and dresses online, these days. Lin *et al.* [55] proposed a hierarchical deep CNN framework to recommend better and efficient clothing options for online customers. The authors used a large-scale image dataset sourced from Yahoo online shopping. The costumes' images depict high variability in poses and appearances with significantly noisy backgrounds. The reason, the clothing-specific tree is generated with the categories such as Men, Women, and so on, whereas sub-categories include top, dress, coat, outfit and so on. The solution intuition is to match the cloths images liked by customers to that of the images in the dataset. The deep CNN is used to learn discerning features representations automatically, talented of detecting heterogeneous types of clothing images. Additionally, the DL-based hierarchical search offered immediate retrieval response as compared to conventional CNN with manually constructed features.

The deep CNN models are becoming pervasive for learning feature representations. Kiapour *et al.* [56] proposed a DL-based model to match the precise shopping location with users' query. The large-scale cloth shop images are

TABLE 1. DL methods summary for user behavior modeling.

Model	Year	Main Idea	Findings	Limitations
Prediction using DL				
TAE [37]	2016	• Deep-computation model for learning features on big data	• Tensor DL model is effective to learn features from heterogeneous data. • Higher capability of multimodal features learning	• Tensor-based feature learning takes more time • Heterogeneity in data enhances the complexity of the model
FARSEEING [38]	2016	• Multiple social network learning	• DL is effective in dealing with multi-sourced SM data • FARSEEING performed better capturing the complex nature of SM by multi-sourced information fusion	• Measuring the confidence level for each information source • Consistency among different information sources • Multi-source multi-task learning
DBN-based link prediction model [39]	2015	• DBN-based link prediction and feature representation • Signed social networks	• Positive and negative links prediction with improved accuracy also in • Behavior learning for members in signed social networks	• Incorporate more significant features in the model • Update novel link prediction approaches
Unsupervised Linguistic Representation Feature Vector (LRFV) extraction [42]	2016	• Feature learning for personality prediction	• Relevance between the personality of users and their behaviors over social networks • Prediction of social linguistic behaviors of users through unsupervised DL • LRFV is more efficient, objective, and complete as compared to the supervised feature extraction	• Consider huge data to learn features • Novel unsupervised feature learning models
Clustering using DL				
DeepLCRank [44]	2017	• DL for enhanced clustering and wise ranking	• Integrating DL with the SM data for better ranking • Assimilating ranking with clustering or community member ranking is objectively informative	• Synthetic experimentation • Explore real-world datasets
Classification using DL				
RGDBN [15]	2013	• Deep architecture-based latent features learning	• RGDBN-based feature learning process is adequate to retain information about connections both for homogeneous and heterogeneous data • More effective prediction of links between connected media content • Non-feature-based learning methods are ineffective in user recommender systems and image annotation tasks	• Consider more sensible and multifaceted forms of latent features • Posterior inference techniques for model learning, and experimentation in different SM platforms
SDNE [33]	2016	• Deep networked embedding-based network structure-preserving	• Multi-layered semi-supervised deep models better preserve the network structure • Robustness to sparse networks	• Learn representations for a non-connected novel vertex
DCNN [43]	2017	• Classification of teenager and adults using deep CNN	• With the writing style, history, and profile-based users' history, the accuracy of age determination for Twitter users can be enhanced	• Consider other sentiment metrics such as time-based web-log to predict missing information in users' social profiles
Ranking using DL				
MIDL [45]	2017	• User-personalized satisfaction prediction	• Question-diversity increased the necessity of satisfaction prediction • DL-based methods predict with higher accuracy	• Applications to other information retrieval tasks and NLP
Recommendation using DL				
MV-DNN [46]	2015	• Content-based recommendation system • Enhance the quality and scalability of recommendation	• Augmented-recommendation with the MV-DNN framework • Incorporate multi-domain data for the recommendation	• Incorporate more users' features to improve recommendation quality, • Avoid dimensionality reduction with enhanced DNN-based models • Integrate collaborative filtering with DL for the recommendation
CDL [47]	2015	• Learning user ratings • Deep representation learning for content information	• CDL can better understand the tagging and collaborative filtering • Suitable recommendation increases the output precision	• Consider the context and order of words in text corpora
DLMF [48]	2017	• Trust-based recommendation in social networks	• DLMF predict improved latent features for the enhanced recommendation • Individual characteristics affect the users' decisions • Reliable recommendations by friends	• Temporal sensitivity to recommend trust-based content in social networks

used from Tamaraberg[19] and ModCloth.[20] The problem of matching users' clothing query with available and feasible

shopping location is intuited as computing the cosine similarity between the cloth features from query and features of online shop images. To recommend customers with better shopping locations, the shop retrievals are ranked based on the computed similarity. Accordingly, Chen *et al.* [57] also

[19]http://www.tamaraberg.com/street2shop
[20]http://www.ModCloth.com

TABLE 2. DL methods summary for business analysis.

Model	Year	Main Idea	Findings	Limitations
Classification using DL				
SDA [16]	2011	• Extract unsupervised meaningful representations for sentiments	• Less computation required for cross-domain transformations • Multi-domain data incorporation	• Design cost-effective and efficient DL-based model
CIMM [54]	2015	• Model user consumption intention for certain products • CNN model	• Semantic information with word embeddings in sentences, • CNN-based CIMM accomplished better performance than traditional SVM • Mid-level sentence representation by domain adaptation	• Personalized classification of users based on product usage
Recommendation using DL				
Deep CNN framework [55]	2015	• Better and efficient online clothing recommendation	• Learning domain-specific visual features • Supervised pre-training on the image dataset • CNN-based content recommendation frameworks	• Develop models with unlabeled images • Attribute annotations
Deep CNN-based similarity computation [56]	2015	• Enhance the similarity between users' queries and online clothing shop images	• Introduced a novel dataset 'Tamaraberg' • Similarity computation for users' queries and shop images.	• More precise configuration between user intended and shop images • Manual evaluation

presented a DL-based approach to describe people on the basis of granulated clothing features.

C. SENTIMENT ANALYSIS

Sentiment analysis, also stated as opinion mining involves the attitude prediction of users' that are generating massive textual content using multi-typed SM platforms, for instance, Facebook, Twitter, etc. Significantly, the intent of analyzing users' sentiments is prevailing, that is to classify their intellect towards a specific product or topic is positive, negative, neutral, or even to classify into some other category. It is commonly applied to the product survey answers, customer evaluations, user opinions, and in domains such as education, business, e-commerce, and healthcare. In this section, concerning sentiment analysis, we explore the techniques comprising prediction, classification, and ranking the sentiments.

1) PREDICTION USING DL

Predicting opinions from SM data is a prevalently active task. The English language has been commonly used for opinion mining task. However, Li *et al.* [58] predicted sentiment labels for Chinese sentiment corpus. The authors collected 2270 movies' reviews.[21] Subsequently, the movie reviews are filtered based on specific criteria such as with rude language, with special symbols, with more than one sentence, with typos, with short or long sentences, or with multi-languages.

[21]http://www.movie.douban.com/

Eventually, they constructed a dataset composed of Chinese sentiments and named it – Chinese Sentiment Treebank.

The sentiments are classified into 5 classes, specifically, very positive, positive, neutral, negative, and very negative. The authors proposed a recursive DL model namely Recursive Neural Deep Model (RNDM) to predict the labels for these classified sentiments. This model is compared with three baselines models, Naïve Bayes (NB), Maximum Entropy (ME), and Support Vector Machine (SVM). Equally, the RNDM outperformed all of the baselines as it predicted better sentiment labels for sentences with contrastive conjunction structure, like "X but Y" in English.

Recently, in order to resolve a wide range of NLP and text mining tasks, DL techniques are rapidly developed and have drawn persuasive consideration to training deep and complex models on abundant data. Since the text is created by humans, it already comprises morphological aspect such as grammatical rules, syntactic knowledge, for instance, Part-Of-Speech (POS) tagging, and also the semantic knowledge such as the relationship between words and entities, synonyms, and antonyms. However, DL leverages to generate fine word embeddings by incorporating knowledge [59].

Bian *et al.* [35] conducted an extensive study to show the injection of knowledge-based DL into word embeddings. Nevertheless, Stojanovski *et al.* [60] proposed a DL architecture for Twitter sentiment analysis using pre-trained word embeddings sourced from GloVe embeddings [61]. Bian *et al.* [35] used the root, affix, and syllable to infer the semantic meaning from text. In order to enhance the

word representations, the semantic and syntactic knowledge is used as additional inputs. The Continuous Bag-Of-Words (CBOW) model is used as the baseline method, however, the Morfessor, Longman, WordNet, and Freebase are the datasets used in the evaluation of word embeddings' quality learned with fused knowledge and without knowledge. In the same regard, this study also explores three tasks namely word similarity, analogical reasoning, and sentence completion task. After comparison, the DL framework directly generated the embeddings for each root/affix, and syllable by aggregating the morphological elements.

Significantly, because of the layered nature of DL-based models, they are well-equipped to incorporate both semantic and syntactic features learning. However, the authors concluded that syntactic knowledge delivered significant input information but it may be inappropriate for regularized objectives. Nevertheless, semantic knowledge can expand the performance of sentence completion and the word similarity task. In addition, though, for the analogical reasoning task, applying the semantic knowledge as additional input is quite influential.

Accordingly, owing to leverage the users' expressions and sentiments within 140 tweet characters' restriction by Twitter, it is rather a non-trivial task. In order to clean the unnecessary information from tweets dataset used for experimentation, Stojanovski *et al.* [60] removed all URLs and HTML entries from the tweets. The pre-trained word embeddings are used to construct lookup tables, where each is linked with the matching feature representation. In terms of DL, the authors fused two NN models; one is CNN used for feature extraction from tweets, and Gates Recurrent Neural Network GRNN that uses sequential data where input are dependent upon the previous outputs. The noteworthy properties and reasons using a Gated Recurrent Unit (GRU) comprise: it uses less number of parameters, needs fewer data to generalize, and also enables rapid learning. Architecturally, GRU contains gating units – to control the information flow inside the underlying units. Consequently, the fusion of CNN with GRNN dominated the performance of existing individual NN models.

2) CLASSIFICATION USING DL

In current ages, DL has arisen as an operative means for addressing the problems pertinent to sentiment classification. Without human efforts, except the labeling phase, a neural network innately learns a valuable representation, automatically. Nevertheless, the victory of DL extremely counts on the availability of extensive training data. With the thriving of e-commerce and Web 2.0, people start consuming SM increasingly and post comments about their buying experiences on review or merchant websites. This prejudiced stuff is a valuable resource to merchants for products improvement, service quality, and to prospective customers to make appropriate decisions.

Poria *et al.* [62] proposed a systematic approach to extract short text features. It is founded on the inner layer activation values of a deep CNN. The authors used CNN for textual data

feature extraction. However, utterances are translated from Spanish into English using Google translator. The CNN used is composed of 7 layers and is trained using a typical back-propagation procedure – usually convenient in improving the accuracy of the model.

The dataset is composed of 498 short video fragments where a person utters one sentence. For the sake of polarity of sentiments, the items are manually tagged as positive, negative, or neutral. However, by discarding neutral items, in total 447 items are processed. The combined feature vectors of visual, audio, and textual modalities are used to train a classifier grounded on multiple kernel learning (MKL) algorithm, a well-known heterogeneous data processor.

Actually, the authors combined the results of feature and decision-level fusion. Moreover, in feature-level fusion, the features are fed into a supervised classifier namely SVM after extraction, while in decision-level fusion, the extracted features are fed into separate classifiers, and then decisions are combined. The significance of CNN feature extractor is being automatic and it does not rely on handcrafted features. In particular, it well-adapts to the distinctiveness of the specific dataset, in a supervised manner.

Being a subtask of analyzing sentiments, aspect extraction involves identifying the targets of opinions in prejudiced text, particularly in detecting that the judgment holder is either endorsing or complaining about the particular features of a product or service. Poria *et al.* [63] proposed a novel method for aspect-oriented opinion mining using a 7-layered deep CNN. The traditional methods for feature extraction from text such as conditional random fields (CRF) have limitations. These limitations include several features required to perform even better. The linguistic pattern (LP) needs to be crafted manually which depends on the sentences' grammatical accuracy. Owing to the automatic feature extraction nature, a CNN framework could effectively overcome such limitations for feature extraction.

The aspect-term features are based on its neighboring words. While, for aspect extraction, a 5-word window is used about all the words in a sentence, particularly ± 2 words. The window features which lie locally are considered as middle word features. Next, the feature vector is served to a CNN. Similar to [35], CBOW model is used for creating word embeddings. Google and Amazon embeddings are used as dataset particularly in electronics (e.g., cell phones, laptops), or food-chains (e.g., fast-food, restaurant) domains on which five LP rules defined in [63] are applied. Though, the performance of CBOW over food-chains found to be better than that of electronics' domain. Since electronics' domain encompassed less aspect-oriented terms.

Guan *et al.* [64] also presented a novel framework, Weakly-supervised Deep Embedding (WDE) to classify customer reviews. The focus is on the semantic orientation for each sentence. In this framework, first, a high-level representation is learned which captured the overall distribution of sentences in terms of sentiments using rating information. Next, a classification layer is supplemented on the upper

end of the embedding layer for supervised fine-tuning using labeled sentences. Practicing review ratings for sentiment classification [64] is an initial effort in the sentiment analysis community.

A large amount of unlabeled data is trained by using RBM/Auto-encoders. As ratings are noisy labels, so which would mislead classifier. Hence, the following simple 5-star rating scale rule is adopted:

$$ l(s) = \begin{cases} pos, & \text{if } s \text{ is in a } 4 \text{ or } 5 - \text{stars review} \\ neg, & \text{if } s \text{ is in a } 1 \text{ or } 2 - \text{stars review} \end{cases} $$

where $l(s)$ represents the weak label of sentence s. Amazon customer reviews of three domains namely, digital cameras, laptops, and cell phones are used as a dataset. However, 3-star reviews are ignored. The proposed WDE method is compared with several baseline methods, however, in terms of accuracy, WDE outperformed the existing methods. In a nutshell, WDE actually well trained the DNN by manipulating reviews' rating information that is commonly accessible on social websites.

Araque *et al.* [65] proposed a Classifier Ensemble Model (CEM) to depict the significance of multi-sourced information. Significantly, this composition has further information as compared to its base components. The model aimed to enhance the overall performance of sentiment classification by integrating surface (traditional ML classifier) and deep (DL-based) features which could not be attained by using the classifiers separately. Overall, seven public datasets are used from movie reviews and microblogging domain.

Using SM platform such as Twitter, occasionally, users post nonsensical content. It could be categorized as hate or abusive speech, targeting people such as politicians, celebrities, or products. Detecting such hateful intents of users belongs to a certain group or a set of users fit in another group is thus important. Moreover, it is equally significant to recommend appropriate content to users. Badjatiya *et al.* [66] used several DL architectures namely CNNs, FastText, and Long Short-Term Memory Networks (LSTMs) to classify tweets as hate speech or not using labels – sexist, racist, or neither. Functionally, CNNs are used for hate intent detection, FastText can quickly represent a document in the form of word vectors to tune the word representations using pre-trained word embeddings sourced from GloVe embeddings as in [60], and LSTMs are used to track long-standing dependencies in the tweets. Detecting applicability of DL to detect hate intent in the tweets is the major contribution in [66].

Instead, Pitsilis *et al.* [67] also presented a deep architecture to classify short text from tweets as hatred speech. Nonetheless, this approach does not solely count on pre-trained word embeddings. The intuition of classifying short text is established on the historical tendency of users to post hatred content in the form of offensive messages. Usually, users prefer to use short terms or words to express their slang intents. Accordingly, tweets of length 30 are used in

training deep proposed model, the reason – frequency of word vectors is preferred over pre-trained word embeddings. Alali *et al.* [68] proposed a Multi-Layered CNN (MLCNN) to classify tweets into five scales – highly positive, positive, neutral, negative, and highly negative. Nevertheless, after empirical evaluation of the proposed model, the authors found 3-layered CNN, the best among all other combinations of layers.

3) RANKING USING DL
CQA forums practice several questions from users and answers from experts every day. For instance, if a user posts her question on a CQA forum, the purpose is to seek for the best answer against her question. However, a number of experts inclined towards posting a quality answer to the same question. This establishes a link between question-answers text pairs. Likewise, in several information retrieval tasks, the links between query-documents exist as short text pairs. These impose the requirement to rank the answer-question pairs or query-document pairs of text. In addition, feature engineering is a protruding aspect of learning these ranks.

Above all, it has been lately depicted that CNNs are quite effective in resourceful learning and embedding input sentences into low-dimensional vector space. It preserves the important semantic and syntactic characteristics of the input sentences. Besides, it also heads towards the contemporary results in many text processing tasks. Severyn and Moschitti [69] proposed a convolution neural network-based (ConvNets) approach to learning the ranks for short text pairs. This model is aimed to learn decent intermediate representations of the documents and queries, later on, used for calculating their semantic matching. The network comprised a wide convolutional layer tracked by a non-linear and simple max pooling – which is used to reduce the dimensionality. The raw words are used as input to the network. This input is needed to be interpreted into real-valued feature vectors. The successive network layers are then used to process these real-valued feature vectors. Furthermore, extracting significant patterns is the objective of the convolutional layer, specifically, discriminative word sequences established within the common training input sentence instances. Purposefully, the authors used two well-known retrievals TREC benchmarks: TREC microblog retrieval and answer sentence selection.

CNN supported in learning more intermediate representations which therefore improved learning of high-quality sentence models. This architecture comprises intermediate representations of the questions/answers, which together establish a much richer representation. ConvNets also do not require manual feature engineering, virtual preprocessing, and peripheral resources, which might be costly or rather not available. The same architecture model has applications in other domains as well.

Tan *et al.* [70] proposed biLSTM-based method to select a suitable answer from a pool of answers to a question asked at the SM platform. TREC QA question answer selection

dataset is used in this study. The authors used distributed deep representations to match the questions in open domain question answering systems with appropriate answers by bearing their semantical structure. The problem is solved with the intuition of higher cosine similarity with the question and the answer is chosen with the highest cosine similarity from a descending similarity ranked list. The biLSTM, a DL-based model played the role where a question has multiple relations to the words or terms used in the answers along with several relations and ideas. The noticeable feature of the proposed model is that it does not depend upon linguistic features engineering or tools whereby it can be applied to any domain. TABLE 3 shows the summary of DL methods for sentiment analysis.

D. ANOMALY DETECTION

Anomaly detection is the field of detecting abnormalities from data. While investigating the real-life datasets, a collective necessity is to detect such instances in the data which stand out from the rest of the data instances [71]. Such data instances are known as abnormalities, outliers, or anomalies.

Usually, anomalies are instigated by bugs in the underlying data but occasionally, anomalies are produced due to formerly unknown fundamental process – rather, Hawkins [72] defined an outlier as: ''an outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism''. In order to detect anomalies, several efforts have been made using DL. We discuss here some significant DL-based anomaly detection techniques in SM.

1) CLASSIFICATION USING DL

Supervised anomaly detection can play a pivotal role in improving security systems. It can also help security organizations and law enforcement agencies to act proactively and detect destructive and predatory actions and conversations in the cyberspace. Ebrahimi *et al.* [73] proposed a CNN-based classifier to efficiently detect such actions in large-volume chat logs. The authors used PAN-2012[22] public dataset for experimentation. CNNs, as a binary classifier, is used to accomplish the conversational text classification task. In CNN, the convolutional layer usually functions on several input regions. However, the pooling layer is used to sub-sample the higher levels of abstraction in each convolutional layer. Being a text classification task, max-pooling is used as it beat the other method that is average-pooling [74]. It is found that using two convolutional layers is less effective as compared to the one convolutional layer in text classification tasks. However, for image classification tasks, several numbers of convolutional layers might help, indeed. Since the model tended to overfit the data with the greater number of convolutional layers.

Ribeiro *et al.* [75] proposed a deep Convolutional Auto-Encoder (CAE) for anomaly detection in video data.

22 http://pan.webis.de/clef12/pan12-web/

The authors used publicly available videos including SM platforms. The proposed CAE model does not necessitate the labeled data since all the training instances fit in the non-anomalous group. The sliding windows are used to sub-sample the video frames from video clips, wherein both motion and appearance features are extracted. The frame reconstruction error is presented as the anomaly score. The Area Under Curve (AUC) is used to calculate the performance of the proposed model. It is found that the accumulation of high-level information to unprocessed data can augment the performance of the proposed classifier, CAE. Particularly, supplementing high-level information can be valued given that the types of anomalies known beforehand that are envisioned to be detected.

2) CLUSTERING USING DL

Detecting anomalous events particularly in videos is of yet utmost significance. The scenes are greater in number, since detecting abnormal scenes from video data can be considered as a clustering problem. Feature learning is non-trivial in video surveillance. However, Xu *et al.* [76] proposed a DNN-based Appearance and Motion DeepNet (ADMN) to automatically and efficiently learn feature representations. A double fusion framework is presented coalescing the goods from both early and late fusion strategies. In early fusion strategy, stacked denoising auto-encoder is proposed to discretely learn both motion and appearance features. Then, one-class SVM is used in order to predict anomaly scores. Towards the end, the anomaly scores are grouped together to detect abnormal events. It is found that the proposed ADMN model does not rely on former knowledge for feature representation learning. Also, it is more powerful than existing hand-crafted learning of video features representations. The downside of this approach includes higher computational cost and also co-occurrence of several patterns in the videos.

Anomalies are nonetheless, a result of multiple events co-occurring or multiple factors causing instances to be declared as anomalies. One of the drawbacks of [76] includes a lack of consideration of co-occurrence of patterns in videos. Hayat and Daud [77] detected co-evolutionary anomalies in heterogeneous bibliographic information networks. Co-evolutionary anomaly (target object) is associated with a number of linked attributes (attribute object). Correspondingly, the influence of each of the attribute objects is calculated on the target objects which helped to identify the cause of the occurrence of anomalies in the underlying data.

Likewise, Feng *et al.* [78] detected anomalies in the crowded scenes. The authors proposed a deep Gaussian Mixture Model (GMM) with a different combination of layers on top of each other. The motion and appearance features are extracted using Principle Component Analysis (PCA) and applied with clustering. The clusters with less number of members and/or far away from the regular groups are picked as anomalies. It is found that using deep-GMM is way valuable as compared to hand-crafted feature learning. However, short and long-term temporal motion features are

TABLE 3. DL methods summary for sentiment analysis.

Model	Year	Main Idea	Findings	Limitations
Prediction using DL				
RNDM [58]	2014	• Predict sentiment label • Recursive DL	• Focus on sentiment analysis for the Chinese language • Chinese treebank introduction • Sentence-level sentiment label prediction with higher accuracy • Contrastive conjunction structure in sentences	• Consider more forms/languages of text with respect to natural language • Incorporate more textual contexts
CNN-GRNN [60]	2016	• Predict tweets' labels • Fused DL models	• Fusion of DL models performed better than using the models individually	• Pre-train the word embeddings on equally large-scale datasets, • Consume bi-directional GRNN for enhanced accuracy
Classification using DL				
CNN-based feature extraction [62]	2015	• Feature extraction for short texts	• Effective extraction of sentiment polarity from speaking video clips using CNN-based method • Facial expressions, and words' utterances	• Extract more tempting features from the visual modality • Feature selection method for extracting key features
7-layered deep CNN [63]	2016	• Aspect extraction from textual corpora • Opinion mining	• Better data fitting with non-linear Deep CNN, • Linear models such as Conditional Random Field (CRF) are weak in data fitting • Vanishing feature engineering requirement diminishes the time and development cost	• Diverse linguistic patterns (LPs) for aspect extraction
WDE [64]	2016	• Review sentiment classification • Ratings as weak supervision signals	• Learned an embedding space to capture the sentiment distribution of sentences • Inference of weak labels from ratings • Penalizing relative distances among sentences for improved learning	• Apply weak labels extraction on other types of deep networks
CEM [65]	2017	• DL-based sentiment classification • Word embeddings	• Multi-sourced information such as surface features with word vectors • Improved sentiment analysis	• Apply deep models to the aspect-based sentiment analysis • Emotion analysis
Fusing FastText, CNNs, LSTMs [66]	2017	• Hate speech detection • Twitter with multiple DL-based techniques	• Enhanced performance of DNN methods • Effective embeddings' learning with LSTMs	• User's network features to detect hate speech
RNN-based hate speech detection [67]	2018	• Hate speech detection • RNN-based ensemble learning	• Better short text • Count of word frequency vectors	• Evaluate and analyze texts in other foreign languages
MLCNN [68]	2018	• Classify tweets into a five-point scale • Multi-layered CNN	• Three convolutional layers better classify the tweets into a five-point scale • Other combinations of layers are not effective	• Explore further methods to classify text • Devise new pooling strategies other than max and average pooling
Ranking using DL				
ConvNets [69]	2015	• Re-ranking pairs of short texts	• No requirement of any external resources • No need for manual feature engineering • Cost-effective feature engineering	• Apply the model to different social networks other than texts
LSTM-based DL framework [70]	2015	• Answer selection form a pool of answers • TREC QA	• Enhanced appropriate answer selection from the pool of answers • Traditional IR models are not effective	• Answer quality prediction and ranking in community QA

Model	Year	Main Idea	Findings	Limitations
Classification using DL				
CNN-based predatory text classifier [73]	2016	• Classification of predatory discussions in chat logs	• Single convolutional layer for text classification is more efficient • Consuming several convolutional layers are advantageous in image classification tasks	• RNNs and LSTMs for large-scale documents training
CAE [75]	2017	• Anomaly detection in videos • Convolution auto-encoders	• Supplement information to raw data enhances the performance of CAE	• Devise methods to imitate the semantic conception of visuals by humans
Clustering using DL				
ADMN [76]	2016	• Anomalous event detection in video • DL-based architectures	• Prior knowledge is not reliable in DL • Especially, for feature representation	• High computational cost • Fuse more multi-modal data using stacked denoising auto-encoders • Co-occurrence of several patterns • Inclusion of contextual information
Deep-GMM [78]	2017	• Anomaly detection from crowded scenes • Deep-GMM	• Enhanced learning of motion and appearance features • Hand-crafted feature learning is not effective	• Build deep models for short-term motion feature learning • Long-term motion learning • Appearance feature learning

still having room to explore. TABLE 4 shows the summary of DL methods for anomaly detection.

IV. DATASETS AND BENCHMARKS

In this section, we go through significant SM datasets usually used as benchmarks in the SM tasks.

A. FACEBOOK

In the era of social connectionism, people prefer to connect with their friends online using SM platforms. Facebook[23] is one of such platforms which people use privately to express and share their content. Though, the SM usage of users may vary due to the wide-ranging services offered by these platforms. Certainly, a number of platforms are available to download and use publicly accessible SM datasets. Particularly, the features of Facebook dataset comprise – demographic features such as gender, education, relationship status; user topics such as dispositions, professions; user posting behavior such as slang posts, photos posts, video posts, shares, likes and so on.

An abundant amount of multimodal data is being posted at Facebook every day. However, most of the data is in unstructured form. In order to extract meaningful insights from unstructured data, DL can service in representing it in a form which can answer questions such as – how often does a certain company's product appear in pictures to a particular user? Correspondingly, the profusion of multimodal data can be sheltered by applying powerful DL models in several

noteworthy SMA tasks – textual analysis, image analysis, targeted advertising, and so on.

B. TWITTER

People tend to express their opinions about a certain topic. Twitter, an SM platform which people use to publicly express and share their personal viewpoints and interpretations about a certain event, personality, organization, or happenings. The features of tweets include positive emotions, negative emotions, relationship expressions, or health status.

DL has striking implications in NLP. In terms of SM insights, the text is still a dominating factor among multi-typed data. Users practice Twitter to express their personal viewpoint about a certain social, political, educational, personal, or professional event. In addition, most of the Twitter data is posted in the form of text. DL can help in a number of NLP tasks such as crises situation identification, users' behavior analysis, quantifying users' enthusiasm about a certain event, and so on.

C. LINKEDIN

Companies seek individuals who are professional in their field of interest and tend to excel in future career. LinkedIn is a type of professional SM platform where people build their profiles to get hired or to hire the appropriate candidates. Fundamentally, it is an employment-oriented and business-focused online social service to create a network of professionals. Usually, the access to this dataset is not truly public unless requested or subscribe by a registered entity.

DL can play an important role in analyzing the textual, image, or video data provided by the LinkedIn platform.

[23]http://multiplesocialnetworklearning.azurewebsites.net/

Usually, people use this platform for job search, however, applying DL to scan through the candidates resumes and filter out the eligible ones can be of significant support for human resource managers.

D. FLICKR

Photo sharing is one of the most used online service, these days. Flickr is an image and short video sharing platform where people share their experiences and emotions in the form of photos and short videos. These platforms are available in a number of languages including English, German, French, Korean, Spanish, and so on. Account creation is not mandatory to access the content on such SM platforms but uploading imposes the account creation.

DL can be applied for image and video data analytics. On account of the recurrent nature - the RNN-based models can be used to analyze users' likes, dislikes, recommending the most anticipated content to users, and surrounding the users with advertisements of their personal and pertinent interests.

E. YOUTUBE

Videos and moving objects are the most substantial type of content to fascinate users. YouTube is one of the most prominent SM platforms to share the content in the form of videos. It allows the consumers to upload (with an account), share, rate, or view the video posts and manage their account privacy settings. The user-generated content uploaded at YouTube ranges from educational, artistic, trailers, documentaries, towards official live video streaming.

Analyzing dynamic and live streams is a non-trivial problem, however, deep auto-encoders can leverage dynamic video analysis significantly. The underlying nature of the data anticipates more powerful frame representations which simple ML approaches cannot provide. However, DL can be used to extract meaningful insights from dynamic video streaming platforms such as YouTube, Metacafe, and Vimeo.

F. STACKEXCHANGE

Individuals may face a number of problems, every day, for which they sought the trivial and reliable solutions. Stack Exchange is a type of SM platform where people can ask questions from varied fields and topics and learn solutions from community question-answer systems. However, a family of experts, sometimes the normal users tend to answer the questions asked by various users. The reputation can be achieved by posting highly upvoted and quality answers to the questions. Particularly, the answer with a maximum number of upvotes is chosen as the best answer. Stack Overflow, Super User, and English Language and Usage are some prominent sites used for learning programming, information technology, and English linguistics respectively.

CQA is a platform where users seek for the most relevant answers to their questions, the best answers, to-the-point answers, and most importantly to search such answers rapidly. This constitutes a multimodal search altogether.

The deep hidden layers can be used to construct a model which can satisfy users multitude of necessities.

V. PERFORMANCE EVALUATION

There a number of ways to evaluate the performance of DL methods. The recall, precision, and F1-score are a noteworthy performance measure for prediction, ranking and classification-based tasks [38], [45]. The F1-score stabilizes the recall and precision as:

$$F_1 = \frac{precision \cdot recall}{precision + recall} \tag{2}$$

where *precision* – the portion of pertinent instances among instances retrieved while *recall* – the ratio of retrieved pertinent instances with the total number of pertinent instances in the data [79]. The classification tasks also evaluated by ROC curve. With ROC curve as a the ground, the Area Under Curve (AUC) is used as an evaluation metric for each class [39].

To measure the difference between actual observed values to that of predicted values, Root Mean Squared Error (RMSE) is often used to evaluate the model performance [42], [48]. The larger RMSE error values signify greater performance loss. Nevertheless, RMSE is a performance evaluation technique that is highly sensitive to anomalies. It can be represented as:

$$RMSe_{fo} = \sqrt{\sum_{i=1}^{N} \frac{(z_{f_i} - z_{o_i})^2}{N}} \tag{3}$$

where z_{f_i} and z_{o_i} are differences between predicted and real values, and N is the sample size [80].

Usually, the proposed method performance is also evaluated by an empirical comparison of the proposed and existing methods' results [76]. However, three core performance metrics plays a significant role in order to measure the performance of a DL model; accuracy, size of the model, and the rate of learning.

A. ACCURACY

Predominantly, the core objective of DL-based models is to make precise predictions. The diverse nature of SM data demands either to communicate the complete information with long delays owing to the excessive and complex processing or to have data samples in order to make the predictions. However, the later could lead towards imprecise diagnostics and predictions.

For instance, a DL-based model which predicts voter fraud. A binary classifier algorithm process individual votes and classify them as legit or fraudulent. Obviously, the error cost for a fraudulent vote as legit is more severe than a parallel error that marks a legit vote as potentially fraudulent. In case, after processing a million votes, if the model fails to capture 100 fraudulent votes, even then, there will be a high accuracy. However, the end result could affect the context in a greater sense.

B. MODEL SIZE

In order to progress the prediction accuracy of the model, the size of the NN matters equally. Also, it is mounting exponentially. Therefore, to process abundant data using deep models, they need to be really efficient, scalable, and robust, so that the computation accuracy may not be transformed if the model needs enhancement.

C. LEARNING RATE

Learning rate matters in a great deal most of the times for any DL-based method. The reason – DL is used to process plentiful data with improved accuracy and rapid processing. Also, the dynamic nature of DL models, they involve training, deployment, and sometimes re-training – which is a requirement when new data come to the lodge. Therefore, it is quite important – how fast a deep model can learn representations aligned with the speed of new data arrival.

VI. CHALLENGES AND FUTURE DIRECTIONS

The preceding sections explored different domains of SM where DL methods are applied, described the datasets used as benchmarks for specific research purposes, and deliberated some performance evaluation measures used in the literature. However, the features allied with SM pose challenges in adjusting DL methods in such a way that they can solve those problems. In this section, we present some topics where DL needs supplementary investigation for SMA. Undoubtedly, these are dealing with high-dimensional data, learning with streaming data, scalability of models, and distributed computing.

A. TRUST-AWARE SOCIAL RECOMMENDATION

SM platforms encompass a number of social relations which needs trusted recommendations. People are always ought to get connected in a type of relationship that can be trusted or receive such information that is reliable. Trust metrics have to play a significant role in recommender systems [81]. Nevertheless, disseminating trusted information can be helpful for to conquer unswerving recommendations. Deng *et al.* [48] made a DL-based effort to recommend trust-aware relations to SM users. However, that study is limited to Epinions and Flixster data only.

Deliberately, the trust in a pool of relations may vary with respect to time. Also, the content rated by users also significantly sensitive to time. More importantly, the ratings which have become obsolete may cause the data noisy and unreliable to be used for social recommendations. Efforts are still needed to proposed reliable and scalable DL-based techniques to recommend trusted information and trust-aware relations in SM platforms.

B. REFINING DL AND AVOIDING DIMENSIONALITY REDUCTION

Machine learning models use a wide range of parameters to predict, learn, recommend, classify, cluster, or group different data items. Sometimes, these random variables or parameters are often denoted as dimensions. Most of the machine learning algorithms perform dimensionality reduction before processing a dataset. However, DL provides the room to avoid dimensionality reduction by improving and undertaking more robust and scalable learning.

Elkahky *et al.* [46] though used multiple data sources to train and learn DL methods. However, they suggested making DL learning more scalable that one could avoid reducing dimensions. While using an unabridged set of features could depict more durable results through learning a model.

However, it is also indispensable to reduce dimensions [82] which can enrich the network structure. Apparently, an enhanced network structure can help in processing the nodes of DL-based network with greater ease and less complexity.

C. CUT COST AND PUT PRODUCTIVITY

With the exponential growth in SM, the availability of online reviews and recommendations have significantly increased, which thereby made the sentiment classification, an interesting topic in academic and industrial research. However, the existence of an even bigger pool of multi-domain online reviews caused it difficult to gather annotated training data. Glorot *et al.* [16] proposed a deep model, SDA, which is based on stacked denoising autoencoders which use the output of one layer as the input of the next stacked layer, hence, improving representation learning.

SDA can use data from different domains which cuts the computation required to transfer to numerous domains. However, the number of domains adapted is not promising, while there exist several social networks acquiring to adapt data as a single source. In addition, training a DL model using large dataset requires more powerful resources which could increase the cost. It shows us the scope of building more cost-effective and efficient DL methods to incorporate yet more social network domains [76].

D. BEYOND ENGLISH AS TEXT

Sentiment analysis is the research area that investigates the opinions of people, individual evaluations, attitudes, sentiments, appraisals, and emotions towards objects such as products, services, individuals, organizations, events, issues, topics, and their attributes. With the rise of Web 2.0 and enormous growth in SM content, sentiment analysis has now developed as a prevalent and challenging research problem. Users are now able to post their opinions and views on a wide range of social websites such as Twitter and Facebook. However, the information in users' reviews and opinions is of great significance and needs to be explored.

Today, SM users are able to post their reviews and opinions in a wide variety of natural languages such as English, Chinese, Turkish, Spanish, and so on. Li *et al.* [58] made an effort to explore Chinese language sentiments using recursive DL. The challenges are two-fold here: one is to create data banks for languages other than English as [58] build

the Chinese Sentiment Treebank, and the other is to use and build apposite DL methods to explore SM natural languages beyond English [67]. In addition, there is still a decent pitch to play with DL methods on a wide ground of SM websites.

E. ASPECT EXTRACTION

Being a subtask of sentiment analysis, the aspect extraction comprises categorizing target of opinions in prejudiced text. Specifically, in distinguishing the particular aspects of a product, the opinion originator is either endorsing or complaining about. This problem needs a set of linguistic patterns to classify words as an aspect or non-aspect words in sentences.

Accordingly, Poria *et al.* [63] presented a 7-layered deep CNN model for efficient aspect extraction from Google and Amazon embeddings. Significantly, LPs play a prominent role in aspect extraction in SM data. Hence, it enquires DL for more robust LPs crafting for efficient and precise aspect extraction.

Generally, the recommendation systems recommend items on the basis of the customer's purchase history, for instance. However, designing deep-models which can learn the representations for the recommendation on the basis of multiple latent aspects [83] or personalized aspects [84] could be of great worth.

F. HETEROGENEOUS ANOMALY DETECTION

Social networks grasp the capital of useful information. While the same user could have her accounts on a number of social networking platforms [69]. Integrating users' information from heterogeneous SM data sources is always intriguing. It can illustrate a comprehensive understanding of the interests and behaviors of users – which are dynamic, heterogeneous, and multi-contextual. If some data instance is an anomaly in a specific context, however, it may be not in the other context. Also, the same data instance could be anomalous from the multi-points of view – which is significant to detect. Implementing DL approaches for joint representation learning can better seizure the complexity of networks for heterogeneous anomaly detection [85]. However, maintain consistency among heterogeneous social networks is challenging [38].

G. FUSING ANOMALY DETECTION AND SOCIAL INFLUENCE

Escalated usage of social media platforms caused an unprecedented intensification in the social data and provides an extraordinary prospect to study social behaviors of users. However, in the pool of social media users, few are influential while the majority are normal. Similar is the case with the anomalies in social networks. A number of earlier studies propose to solve conventional anomaly detection problems using the activity-based, graph-based, community-based, distance-based, and statistical-based methodologies. Influence analysis and anomaly detection problems mostly investigated independently with the domain specific applications. However, rather studying both problems separately,

combinatorial consideration may capture rich semantics from social networks with diverse and more efficient applications.

Likewise, fusing two or more problem solutions can arise to a new valuable solution, diminish the preceding or individual solution cost, and as well as enhance the learning rate for the DL model. Importantly, DL has the ability to learn more complex and heterogeneous representation for networks. The network diffusion based embeddings [86] methods can solve a number of limitations of the DL-based methods including heterogeneity, scalability, and multimodality.

H. TOPIC-SENTIMENT MIXTURE ANALYSIS

With the rise in SM data such as product reviews, business forums, and so on – analyzing sentiments along with the topics of the text is equally significant to swiftly and efficiently recapitulate the textual data [87]. Social and public availability of Web 2.0 generates a path to the unprecedented intensification of SM data whereas a significant part of that data is of textual nature. The SM platforms such as described in section II, the users are unrestricted to express their opinions. As compared to the conventional documents – the SM data (documents) entails the sentiment along with topic-oriented analysis.

Instead, exploring the power of DL-based models for NLP tasks, the distributed representation learning for words can be used to effectively identify topics and predict sentiments for document summarization and sentiments analysis tasks respectively [88]. Existing works focused on topic identification and sentiment analysis individually, however, studying them together and developing a unified model is effective, indeed. It could also support in summarizing textual data and also in CQA platforms.

VII. CONCLUSIONS

SMA has grown widespread attention, recently. While going through the literature, we found several articles studying diverse aspects of SM problems. However, there are no articles revealing DL prospects using insights from SM analytics. In this article, we particularly enlighten this gap, also the pertinent models and algorithms, comprehensively. In terms of DL, we present the state-of-the-art research accomplishments in SM analytics. We also present the current research challenges and future directions in this domain.

In conclusion, the SM platforms present a number of noteworthy challenges to DL. We provide a detailed depiction of varied SM domain. The DL-based methods have significant power to learn valuable data representations from multi-domain SM platforms such as user behavior analysis, business analysis, sentiment analysis, anomaly detection and many more. However, the aspects including powerful resource requirement to deal with the data heaps, improving productivity and putting down the computational costs, learning efficient data representations from heterogeneous social data sources and so on, still need efficient and reliable DL-based techniques. Using DL, these challenges need to be addressed in a canonical way that proves to be an edge

for the scientific community. We have faith that these posed challenges will bring plentiful research prospects to the DL community. Also, they will deliver key developments in various real-life fields such as education, business, e-commerce, medicine etc.

REFERENCES

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

[2] R. Hanna, A. Rohm, and V. L. Crittenden, "We're all connected: The power of the social media ecosystem," *Bus. Horizons*, vol. 54, no. 3, pp. 265–273, 2011.

[3] C. L. P. Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Inf. Sci.*, vol. 275, pp. 314–347, Aug. 2014.

[4] C. Shi, Y. Li, J. Zhang, Y. Sun, and P. S. Yu, "A survey of heterogeneous information network analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 1, pp. 17–37, Jan. 2017.

[5] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, no. 1, p. 1, Feb. 2015.

[6] S. Pal, Y. Dong, B. Thapa, N. V. Chawla, A. Swami, and R. Ramanathan, "Deep learning for network analysis: Problems, approaches and challenges," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2016, pp. 588–593.

[7] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Trans. Signal Inf. Process.*, vol. 3, pp. 1–29, Apr. 2014.

[8] L. Deng and D. Yu, "Deep learning: Methods and applications," *Found. Trends Signal Process.*, vol. 7, nos. 3–4, pp. 197–387, Jun. 2014.

[9] J. A. Obar and S. Wildman, "Social media definition and the governance challenge: An introduction to the special issue," *Telecommun. Policy*, vol. 39, no. 9, pp. 745–750, 2015.

[10] E. Otte and R. Rousseau, "Social network analysis: A powerful strategy, also for the information sciences," *J. Inf. Sci.*, vol. 28, no. 6, pp. 441–453, 2002.

[11] M. Shahriari and R. Klamma, "Signed social networks: Link prediction and overlapping community detection," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2015, pp. 1608–1609.

[12] N. Girdhar and K. Bharadwaj, "Signed social networks: A survey," in *Proc. Int. Conf. Adv. Comput. Data Sci.*, 2016, pp. 326–335.

[13] J. Kunegis, A. Lommatzsch, and C. Bauckhage, "The slashdot zoo: Mining a social network with negative edges," in *Proc. 18th Int. Conf. World Wide Web*, 2009, pp. 741–750.

[14] P. Massa and P. Avesani, "Trust metrics on controversial users: Balancing between tyranny of the majority," *Int. J. Semantic Web Inf. Syst.*, vol. 3, no. 1, pp. 39–64, 2007.

[15] Z. Yuan, J. Sang, Y. Liu, and C. Xu, "Latent feature learning in social media network," in *Proc. 21st ACM Int. Conf. Multimedia*, Barcelona, Spain, 2013, pp. 253–262.

[16] X. Glorot, A. Bordes, and Y. Bengio, " Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proc. 28th Int. Conf. Int. Conf. Mach. Learn. (ICML)*, Bellevue, WA, USA, 2011, pp. 513–520.

[17] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.

[18] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, pp. 1019–1031, May 2007.

[19] A. M. Kaplan and M. Haenlein, "The early bird catches the news: Nine things you should know about micro-blogging," *Bus. Horizons*, vol. 54, no. 2, pp. 105–113, 2011.

[20] T. Aichner and F. Jacob, "Measuring the degree of corporate social media use," *Int. J. Market Res.*, vol. 57, no. 2, pp. 257–276, 2015.

[21] L. Licamele and L. Getoor, "Social capital in friendship-event networks," in *Proc. 6th Int. Conf. Data Mining (ICDM)*, 2006, pp. 959–964.

[22] J. E. Vascellaro, "Social networking goes professional," *Wall Street J.*, pp. D1–D2, Aug. 2007.

[23] H. Li, H. Wang, J. Liu, and K. Xu, "Video sharing in online social networks: Measurement and analysis," in *Proc. 22nd Int. Workshop Netw. Oper. Syst. Support Digit. Audio Video*, 2012, pp. 83–88.

[24] C. Shah, S. Oh, and J. S. Oh, "Research agenda for social Q&A," *Library Inf. Sci. Res.*, vol. 31, pp. 205–209, Dec. 2009.

[25] J. Jin, Y. Li, X. Zhong, and L. Zhai, "Why users contribute knowledge to online communities: An empirical study of an online social Q&A community," *Inf. Manage.*, vol. 52, pp. 840–849, Nov. 2015.

[26] M. van Gerven and S. Bohte, *Artificial Neural Networks as Models of Neural Information Processing*. Lausanne, Switzerland: Frontiers Media, 2018.

[27] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.

[28] H. Larochelle and Y. Bengio, "Classification using discriminative restricted Boltzmann machines," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 536–543.

[29] G. E. Hinton, "Deep belief networks," *Scholarpedia*, vol. 4, no. 5, p. 5947, 2009.

[30] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[32] F. Agostinelli, M. Hoffman, P. Sadowski, and P. Baldi. (2014). "Learning activation functions to improve deep neural networks." [Online]. Available: https://arxiv.org/abs/1412.6830

[33] D. Wang, P. Cui, and W. Zhu, "Structural deep network embedding," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 1225–1234.

[34] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding," in *Proc. 24th Int. Conf. World Wide Web*, 2015, pp. 1067–1077.

[35] J. Bian, B. Gao, and T.-Y. Liu, "Knowledge-powered deep learning for word embedding," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2014, pp. 132–148.

[36] H. Aramo-Immonen, J. Jussila, and J. Huhtamäki, "Exploring co-learning behavior of conference participants with visual network analysis of Twitter data," *Comput. Hum. Behav.*, vol. 51, pp. 1154–1162, Oct. 2015.

[37] Q. Zhang, L. T. Yang, and Z. Chen, "Deep computation model for unsupervised feature learning on big data," *IEEE Trans. Services Comput.*, vol. 9, no. 1, pp. 161–171, Jan./Feb. 2016.

[38] Y. Jia, X. Song, J. Zhou, L. Liu, L. Nie, and D. S. Rosenblum, "Fusing social networks with deep learning for volunteerism tendency prediction," in *Proc. AAAI*, 2016, pp. 165–171.

[39] F. Liu, B. Liu, C. Sun, M. Liu, and X. Wang, "Deep belief network-based approaches for link prediction in signed social networks," *Entropy*, vol. 17, no. 4, pp. 2140–2169, 2015.

[40] F. Liu, B. Liu, C. Sun, M. Liu, and X. Wang, "Deep learning approaches for link prediction in social network services," in *Proc. Int. Conf. Neural Inf. Process.*, 2013, pp. 425–432.

[41] M. B. Lazreg, M. Goodwin, and O.-C. Granmo, "Deep learning for social media analysis in crises situations," in *Proc. 29th Annu. Workshop Swedish Artif. Intell. Soc. (SAIS)*, Malmö, Sweden, Jun. 2016, pp. 1–46.

[42] X. Liu and T. Zhu, "Deep learning for constructing microblog behavior representation to identify social media user's personality," *PeerJ Comput. Sci.*, vol. 2, p. e81, Sep. 2016.

[43] R. G. Guimarãs, R. L. Rosa, D. De Gaetano, D. Z. Rodríguez, and G. Bressan, "Age groups classification in social network using deep learning," *IEEE Access*, vol. 5, pp. 10805–10816, 2017.

[44] T. T. Zin, P. Tin, and H. Hama, "Deep learning model for integration of clustering with ranking in social networks," in *Proc. Int. Conf. Genetic Evol. Comput.*, 2016, pp. 247–254.

[45] Z. Chen, B. Gao, H. Zhang, Z. Zhao, H. Liu, and D. Cai, "User personalized satisfaction prediction via multiple instance deep learning," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 907–915.

[46] A. M. Elkahky, Y. Song, and X. He, "A multi-view deep learning approach for cross domain user modeling in recommendation systems," in *Proc. 24th Int. Conf. World Wide Web*, 2015, pp. 278–288.

[47] H. Wang, N. Wang, and D.-Y. Yeung, "Collaborative deep learning for recommender systems," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 1235–1244.

[48] S. Deng, L. Huang, G. Xu, X. Wu, and Z. Wu, "On deep learning for trust-aware recommendations in social networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1164–1177, May 2017.

[49] J. Dean *et al.*, "Large scale distributed deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1223–1231.

[50] A. Coates, B. Huval, T. Wang, D. J. Wu, B. Catanzaro, and A. Y. Ng, "Deep learning with COTS HPC systems," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1337–1345.

[51] F. Piller, A. Vossen, and C. Ihl, "From social media to social product development: The impact of social media on co-creation of innovation," *Die Unternehmung*, vol. 65, no. 1, pp. 1–22, 2011.

[52] L. McCarthy, D. Stock, and R. Verma, "How travelers use online and social media channels to make hotel-choice decisions [electronic article]," *Cornell Hospitality Rep.*, vol. 10, no. 18, pp. 6–18, 2010.

[53] B. D. Weinberg and E. Pehlivan, "Social spending: Managing the social media mix," *Bus. Horizons*, vol. 54, no. 3, pp. 275–282, 2011.

[54] X. Ding, T. Liu, J. Duan, and J.-Y. Nie, "Mining user consumption intention from social media using domain adaptive convolutional neural network," in *Proc. AAAI*, 2015, pp. 2389–2395.

[55] K. Lin, H.-F. Yang, K.-H. Liu, J.-H. Hsiao, and C.-S. Chen, "Rapid clothing retrieval via deep learning of binary codes and hierarchical search," in *Proc. 5th ACM Int. Conf. Multimedia Retr.*, 2015, pp. 499–502.

[56] M. H. Kiapour, X. Han, S. Lazebnik, A. C. Berg, and T. L. Berg, "Where to buy it: Matching street clothing photos in online shops," in *Proc. ICCV*, Dec. 2015, pp. 3343–3351.

[57] Q. Chen, J. Huang, R. Feris, L. M. Brown, J. Dong, and S. Yan, "Deep domain adaptation for describing people based on fine-grained clothing attributes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5315–5324.

[58] C. Li, B. Xu, G. Wu, S. He, G. Tian, and H. Hao, "Recursive deep learning for sentiment analysis over social data," in *Proc. IEEE/WIC/ACM Int. Joint Conferences Web Intell. (WI), Intell. Agent Technol. (IAT)*, vol. 2, Aug. 2014, pp. 180–185.

[59] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for twitter sentiment classification," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics (Volume 1: Long Papers)*, 2014, pp. 1555–1565.

[60] D. Stojanovski, G. Strezoski, G. Madjarov, and I. Dimitrovski, "Finki at SemEval-2016 Task 4: Deep learning architecture for Twitter sentiment analysis," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 149–154.

[61] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.

[62] S. Poria, E. Cambria, and A. Gelbukh, "Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 2539–2544.

[63] S. Poria, E. Cambria, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowl.-Based Syst.*, vol. 108, pp. 42–49, Sep. 2016.

[64] Z. Guan, L. Chen, W. Zhao, Y. Zheng, S. Tan, and D. Cai, "Weakly-supervised deep learning for customer review sentiment classification," in *Proc. IJCAI*, 2016, pp. 3719–3725.

[65] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," *Expert Syst. Appl.*, vol. 77, pp. 236–246, Jul. 2017.

[66] P. Badjatiya, S. Gupta, M. Gupta, and V. Varma, "Deep learning for hate speech detection in tweets," in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 759–760.

[67] G. K. Pitsilis, H. Ramampiaro, and H. Langseth. (2018). "Detecting offensive language in tweets using deep learning." [Online]. Available: https://arxiv.org/abs/1801.04433

[68] M. Alali, N. M. Sharef, H. Hamdan, M. A. A. Murad, and N. A. Husin, "Multi-layers convolutional neural network for twitter sentiment ordinal scale classification," in *Proc. Int. Conf. Soft Comput. Data Mining*, 2018, pp. 446–454.

[69] A. Severyn and A. Moschitti, "Learning to rank short text pairs with convolutional deep neural networks," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf.*, 2015, pp. 373–382.

[70] M. Tan, C. D. Santos, B. Xiang, and B. Zhou. (2015). "LSTM-based deep learning models for non-factoid answer selection." [Online]. Available: https://arxiv.org/abs/1511.04108

[71] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, p. 15, 2009.

[72] D. M. Hawkins, *Identification of Outliers*, vol. 11. Dordrecht, The Netherlands: Springer Science+Business Media, 1980. doi: 10.1007/978-94-015-3994-4.

[73] M. Ebrahimi, C. Y. Suen, and O. Ormandjieva, "Detecting predatory conversations in social media by deep convolutional neural networks," *Digit. Invest.*, vol. 18, pp. 33–49, Sep. 2016.

[74] Y. Zhang and B. Wallace. (2015). "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification." [Online]. Available: https://arxiv.org/abs/1510.03820

[75] M. Ribeiro, A. E. Lazzaretti, and H. S. Lopes, "A study of deep convolutional auto-encoders for anomaly detection in videos," *Pattern Recognit. Lett.*, vol. 105, pp. 13–22, Apr. 2017.

[76] D. Xu, Y. Yan, E. Ricci, and N. Sebe, "Detecting anomalous events in videos by learning deep representations of appearance and motion," *Comput. Vis. Image Understand.*, vol. 156, pp. 117–127, Mar. 2016.

[77] M. K. Hayat and A. Daud, "Anomaly detection in heterogeneous bibliographic information networks using co-evolution pattern mining," *Scientometrics*, vol. 113, pp. 149–175, Oct. 2017.

[78] Y. Feng, Y. Yuan, and X. Lu, "Learning deep event models for crowd anomaly detection," *Neurocomputing*, vol. 219, pp. 548–556, Jan. 2017.

[79] D. M. W. Powers, "Evaluation: From Precision, recall and f-measure to roc, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.

[80] A. G. Barnston, "Correspondence among the correlation, RMSE, and Heidke forecast verification measures; refinement of the Heidke score," *Weather Forecasting*, vol. 7, pp. 699–709, Dec. 1992.

[81] P. Massa and P. Avesani, "Trust metrics in recommender systems," in *Computing With Social Trust*. London, U.K.: Springer, 2009, pp. 259–285.

[82] L. Qiao, H. Zhao, X. Huang, K. Li, and E. Chen, "A structure-enriched neural network for network embedding," *Expert Syst. Appl.*, vol. 117, pp. 300–311, Mar. 2019.

[83] X. Han, C. Shi, S. Wang, S. Y. Philip, and L. Song, "Aspect-level deep collaborative filtering via heterogeneous information networks," in *Proc. IJCAI*, 2018, pp. 3393–3399.

[84] J. Tang and K. Wang, "Personalized top-n sequential recommendation via convolutional sequence embedding," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, 2018, pp. 565–573.

[85] S. Liu, Q. Qu, and S. Wang, "Heterogeneous anomaly detection in social diffusion with discriminative feature discovery," *Inf. Sci.*, vol. 439, pp. 1–18, May 2018.

[86] Y. Shi, M. Lei, H. Yang, and L. Niu, "Diffusion network embedding," *Pattern Recognit.*, vol. 88, pp. 518–531, Apr. 2019.

[87] D. Tang, Z. Zhang, Y. He, C. Lin, and D. Zhou, "Hidden topic–emotion transition model for multi-level social emotion detection," *Knowl.-Based Syst.*, vol. 164, pp. 426–435, Jan. 2019.

[88] R. K. Amplayo, S. Lee, and M. Song, "Incorporating product description to sentiment topic models for improved aspect-based sentiment analysis," *Inf. Sci.*, vol. 454, pp. 200–215, Jul. 2018.

MALIK KHIZAR HAYAT received the master's degree from the International Islamic University at Islamabad, Islamabad, Pakistan, in 2016. He is currently an Active Member with the Data Mining and Information Retrieval Group, IIU, Islamabad. He has published his M.S. thesis in *Scientometrics*. His research interests include heterogeneous social networks, anomaly detection, scientometrics, and deep learning.

ALI DAUD (M'76–SM'81–F'87) received the Ph.D. degree from Tsinghua University, in 2010. He is currently an Associate Professor, and also the Head of the Data Mining and Information Retrieval Group, Department of Computer Science and Software Engineering, International Islamic University, Islamabad, Pakistan. He has published about 70 papers in reputed international Impact Factor journals and conferences. He has taken part in many research projects, and is the Principal Investigator (PI) of two projects. His research interests include data mining, social network analysis and mining, probabilistic models, scientometrics, and natural language processing.

ABDULRAHMAN A. ALSHDADI received the Ph.D. in cloud computing from the University of Southampton, Southampton, U.K., in 2018. He is currently the Head of the Computer Science and Artificial Intelligent Department (CSAI) as well as the Vice Dean of the College of Computer Science and Engineering (CCSE), University of Jeddah, Jeddah, Saudi Arabia. He has published numerous conference papers, journal papers, and one book chapter. His research interests include Industry 4.0 prestaining issues of cloud computing and fog computing security, the Internet of Things (IoT), smart cities, intelligent systems, deep learning, data science analytics, and modeling.

AMEEN BANJAR received the Ph.D. degree from the Faculty of Engineering and Information Technology, University of Technology Sydney. He is currently an Assistant Professor with the Department of Information Systems and Technology, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia. His research interests include network management, intelligent systems, machine learning, and data science and analytics.

RABEEH AYAZ ABBASI received the Ph.D. degree in computer science from the University of Koblenz-Landau, Germany, in 2010. He is currently an Assistant Professor with the Department of Computer Science, Quaid-i-Azam University, Islamabad, Pakistan. During his Ph.D., he worked on extracting and utilizing semantics in social media. His current research interests include social media analytics and social network analysis. He received the HEC-DAAD Scholarship for his Ph.D.

YUKUN BAO (M'04) received the B.Sc., M.Sc., and Ph.D. degrees in management science and engineering from the Huazhong University of Science and Technology, China, in 1996, 1999, and 2002, respectively, where he is currently an Associate Professor with the Department of Business and Management, School of Management. His current research interests include time series modeling and forecasting, business intelligence, and data mining. He has served as a Referee of paper review for several IEEE journals and international journals and a PC Member for several international academic conferences. He has been a Principal Investigator for two research projects funded by the Natural Science Foundation of China.

HUSSAIN DAWOOD received the M.S. and Ph.D. degree in computer application technology from Beijing Normal University, Beijing, China, in 2012 and 2015, respectively. He is currently an Assistant Professor with the Department of Computer and Network Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia. His current research interests include image processing, pattern recognition, machine learning, and feature extraction.

● ● ●