An Overview on User Profiling in Online Social Networks

Vasanthakumar G. U.
University Visvesvaraya
College of Engineering,
Bangalore University,
Bangalore, India.

Sunithamma K.
University Visvesvaraya
College of Engineering,
Bangalore University,
Bangalore, India.

P. Deepa Shenoy
University Visvesvaraya
College of Engineering,
Bangalore University,
Bangalore, India.

Venugopal K. R.
University Visvesvaraya
College of Engineering,
Bangalore University,
Bangalore, India.

ABSTRACT

Advances in Online Social Networks is creating huge data day in and out providing lot of opportunities to its users to express their interest and opinion. Due to the popularity and exposure of social networks, many intruders are using this platform for illegal purposes. Identifying such users is challenging and requires digging huge knowledge out of the data being flown in the social media. This work gives an insight to profile users in online social networks. User Profiles are established based on the behavioral patterns, correlations and activities of the user analyzed from the aggregated data using techniques like clustering, behavioral analysis, content analysis and face detection. Depending on application and purpose, the mechanism used in profiling users varies. Further study on other mechanisms used in profiling users is under the scope of future endeavors.

Keywords

Behavior Analysis, Content Analysis, Face Detection, Online Social Networks, User Profiling.

1. INTRODUCTION

Huge amount of data is generated every minute from different social networking sites like Facebook, Twitter, MySpace, LinkedIn, Classmates, Instagram, Google+ etc., due to social connections and networking amongst their users.

Facebook allows users to create their profile for free, and then they can connect to different people who are organized in four categories: region, college, workplace and high school by using search engine or through web based email account. Users can post images, videos, media with others and friends can see and comment on each post. Privacy is maintained for the images posted like only friends can see or no one can download the image and so on. Google+ is another social networking site where user can post images with single user or with groups / communities. Here users can group different types of relationships into circles (other than friends) and form new circles. Activities like instant messaging, text and video chats through hangouts, geo-tagging, can be performed.

Twitter is simple yet powerful social networking service where its user can post or receive messages to a network of contacts in contrast to sending bulk emails [1] – [3]. Here, user can build their network by inviting others to receive one's tweets and can also follow others posts. By creating profile in MySpace you can invite others to join the network and can become friends with already profiled users so that they will be in your Initial Friend Space. Once the friendship is accepted, then your friend's friend space becomes part of your network. Hence your extended network involves your friend's friends also even though there is no direct connection between you both [4] – [7]. LinkedIn is a professional online social networking site which helps business professionals to find jobs, connect with potential partners. It needs creation of

profile where it requests for personal information. Unlike other social networks, it cannot be used for posting media or images. The site provides business professional an opportunity to find the right person for recruitment, business partners and to meet experts [8] – [14].

Since an user can use the social networking sites for various purposes like instant messaging, media posting etc., it helps in connecting with friends, family from a distance apart. Students staying away from parents, spouse working in different geographical location depend on these sites to be closer with their loved ones. Now a days, people generally depend on the data generated from the internet for their personal purpose as well. For example, if someone wants to find a good doctor or grocery shop or even to visit places nearby, they depend on the recommendations from their friends and others in their social network and also get opinion [15] – [23]. In this way the usage of social networking sites have become part and parcel of everyone's life especially for youth. Users also post their day-to-day activities and concerns, opening an avenue for emerging data mining research [24] – [34].

Since most youth are engaged in social networking, this is a revolution in the society to bring some political or technical changes through young minds. Most of the social networking sites have their own terms and conditions for users to join their sites. After possessing the criteria, users create their profile and start performing activities in the site. Since there is no connection or restriction for users to join different sites, an user can be involved in many or all social networking sites available. This may be useful in connecting to different kind of people, which may sometimes become danger for society [35] – [51].

Usually in social networking sites, an user can search their friends who may already have profiles in that site and get connected with each other through consent. Once the request is accepted then they both can share all types of media and even have private messaging. If an user finds some problem, then he posts it openly and others who have solution to it will reply with possible solutions to the problem. In this regard, if an user likes something then he can post regarding that and can also attract few others through his posts. Bloggers.com gives a platform for their users to create and publish their blogs / posts. The activities performed by the bloggers are of immense scope [52] – [54]. Most of the time, the posts by the users will be regarding their personal lives or some information in concern to society. Finding the topics under discussion in networking sites is still an open research topic.

Because of the wide popularity and fast information diffusion, social networking sites are being used for some malicious purposes as well. Literature reveals that the information in social networking sites spread faster than other media which motivated us to carry out this work. The key challenge lies in

finding such malicious users who intentionally create and spread false information in social networking sites [55] – [60].

Profiling user in online social network requires data pertaining to online activities of such user who performs different kinds of activities. Those activities may depend on one's interest or they can be the effect of some influence on them. In order to profile the users based on their activities on different social networks requires the whole aggregated data of all social networking sites which is practically difficult for the researchers to collect. But based on the activities performed in one particular site for some period of time, the user can be profiled using various approaches separately.

Rest of the paper is organized as under: An overall system architecture is presented in section 2, whereas the components like Clustering, Content Analysis, Face Detection, Geo-Tagging and Behavior Analysis are discussed in detail in sections 3 to 7 respectively. Section 8 concludes the paper with scope for future researchers whereas section 9 gives references of all those papers refered for this work.

2. SYSTEM ARCHITECTURE

Fig. 1. illustrates the architecture of user profiling. There are various ways by which an user can be profiled which depends on the application as well. The overall process followed in profiling the user is as follows: The raw data from online social networks is collected and analyzed with respect to various aspects like activities, images, text data and geographic location.

In this stage, the data related to number of activities performed, the images posed or profile photos, data posed by user and geographic location of the user are extracted distinctly. With respect to images, the details of the image and its features are extracted using appropriate algorithms. The respective data is analyzed and the related / unrelated data is classified using classification mechanisms. Then based on the classified data, the users are clustered to distinct groups with group name and functionality.

After obtaining the user details and their related groups, the number, type and pattern of activities of each user is extracted. Based on the highest number of activities and

influence made on others by an user, the user behavior is obtained. Thus by analyzing this pattern of behavior in online social networks, the user is profiled.

Next, from the extracted facial features, appropriate algorithms of face detection are applied and validation is performed. Challenges faced here is with images having different pose and expressions. Now by comparing the obtained results with database images, user is profiled. Since huge data is available in OSNs and the data is more or less short and sparse, the user generated content in the collected data is analyzed for similarity. Further, the data is analyzed semantically to get more insight to semantic relationship between the data of users, by way of which the user behavior is extracted.

Friends sitting distance apart perform activities which are interconnected in OSNs; all of which are extracted based on the geographical location of users which are analyzed using appropriate algorithms. In order to avoid getting highlighted / recognized for any activities performed from same location, users may use tricks like accessing social networks from various locations. Such type of users and their behaviors are tracked. Based on the behavior of users with respect their activities and data analyzed so far, users are profiled. Depending on the application, the methods to profile users in OSNs are selected.

Huge amount of data gets created each and every minute in Online Social Networks (OSNs). Wide variety of social networking sites are available where users can utilize the data present in the network or upload data which is available with them or can create their own information or can discuss their opinions on different issues, share their knowledge, train users online on some topics and can perform many more activities. Different social networking sites generate different types of data as the user activities vary in each site. Hence the method for profiling users also varies on type of social networking site considered. Profiling user requires a clear knowledge of their activities and their area of interest. In order to track either short time or long time interest or involvement of the user in different topics, clustering is the optimistic and efficient mechanism.
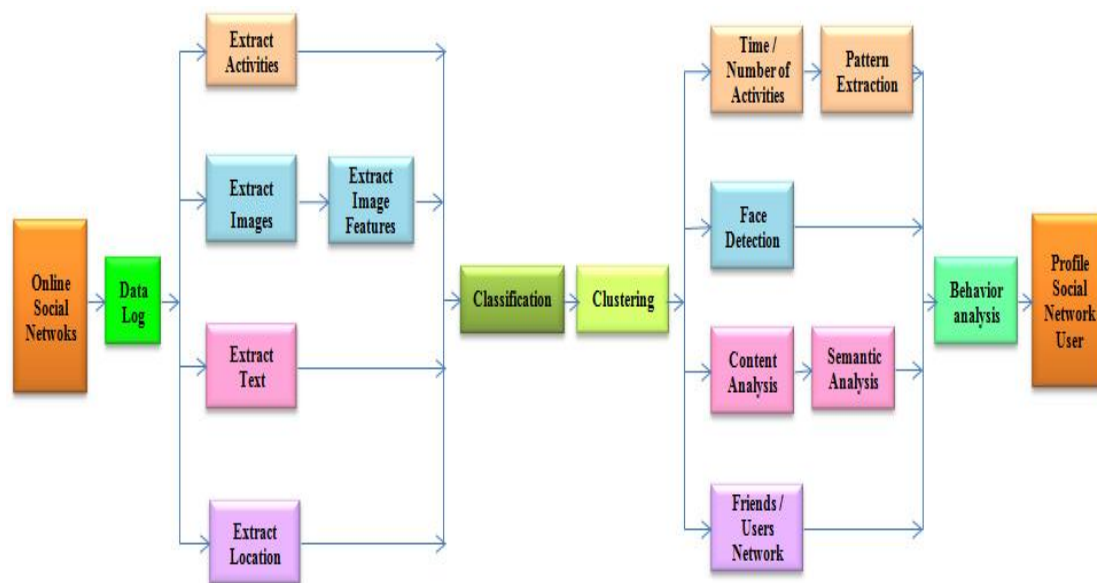

Fig. 1. Architecture of User Profiling

3. CLUSTERING

Based on the user involvement in various clusters and the theme under discussion in each cluster, the profiling of the user is easily done. The mechanism of grouping, also called as clustering has been achieved efficiently using various methods. Clustering the user generated text data and finding the topics underneath each cluster is still an active research topic. In order to form clusters from the data in the database, the relativity between the data is important. Based on the word or phrase or sentence similarity, the clusters may be formed. Unlike traditional clustering, for online social network analysis, objects' similarity cannot be applied.

The social network data after being preprocessed, is clustered based on the need of its application. Fig. 2. shows various clustering techniques whereas Table. I. Shows various clustering algorithms and their features that are applied on the collected social network data for analysis.
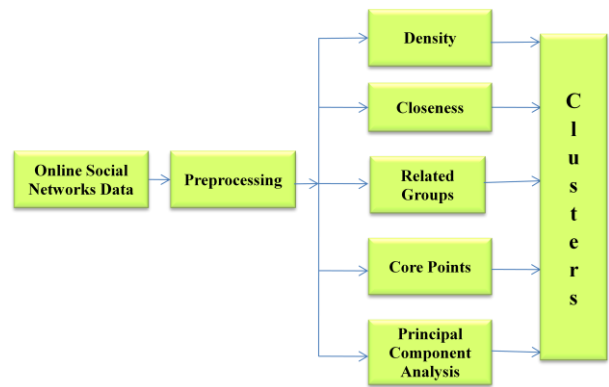

Fig. 2. Clustering Techniques

Table I. Clustering Algorithms

Algorithm	Concept	Data set	Advantages	Disadvantages
DBSCAN [61]	To discover clusters of arbitrary shapes with minimal knowledge of domain with density-based notion of clusters.	Both synthetic and real data of SEQUOIA 2000 benchmark.	1) It performs better than CLARANS. 2) It performs by a factor of more than 100 than CLARANS in efficiency.	1) Only point objects are considered. 2) Performance of Spatial databases may vary. 3) Not verified on high-dimensional feature space.
SDC [62]	This Web opinion clustering technique identifies themes within discussions in social networks.	MySpace web forum data	It performs better than DBSCAN by grouping less relevant clusters.	Adaptive techniques for SDC are not discussed.
Improved SDC [63]	Web opinion clustering with noisy content with sparse and short text messages	MySpace web forum data	It performs better than DBSCAN with filter noise.	It requires eps and MinPts; two parameters to identify the initial clusters.
GDBSCAN [64]	Clusters based on both point objects and spatial extended objects. 2D, 3D, 5D points and 2D polygons applications are presented.	MySpace web forum data, SEQUOIA 2000.	1) Clusters using point objects and spatially extended objects. 2) Its applicability to real-world problems are with of 4 applications.	1) Heuristics determining the parameters. 2) Simultaneous detection of hierarchical clustering.
Improved BSP [65]	Based on objects and links between the objects, social network analysis using PCA with BSP.	Zachary's Karate Club dataset	BSP with PCA outperforms in terms of classification time and accuracy.	It is challenging to divide the object classes based on object's links.
SNN [66]	Identifies clusters of varying densities, sizes and shapes even with outliers and noise.	2D point sets	1) Identifies dense clusters achieving run-time complexity of O (n^2). 2) Handles large data sets efficiently.	No comparison with any existing algorithms.
LSDBC [67]	Automates the process of clustering	Synthetic data set	1) It identifies the clusters as determined. 2) Compared with DBSCAN, k-means and spectral clustering. 3) It summarizes segments and images into meaningful regions.	Performance against compared methods is not statistically measured.
ACDE [68]	To automatically find	5 real-life grayscale images	1) It performs better than DCPSO or GCUK.	Method does not prove to be efficient on all

	optimal number of clusters	and 2 image data sets	2) It finds clusters of arbitrary size and shapes.	complexities of real-life data.
Improved DBSCAN [71]	To detect adjacent nested clusters	Database D	1) It detects embedded and adjacent nested clusters effectively. 2) It also performs better than DBSCAN, EnDBSCAN even on feature space with nested clusters.	Complexity of the algorithm is not defined.
Parallel K-means [72]	Parallel implementation of k-means on both COU and GPU.	MiniBoo, UCI HAR, USCensus1990 and RGB peppers images	1) CPU considered both Clik Plus and OpenMp and GPU considered CUDA. 2) It is implemented on various possibilities.	Performance and iteration time vary based on resolution of images.

Martin et al. [61] developed Density Based Spatial Clustering of Applications with Noise (DBSCAN) method to identify the clusters in large spatial databases and also to separately recognize noise in the database. Before this, the basic method for clustering was Clustering Algorithm based on Randomized Search (CLARANS) and DBSCAN outperforms the state-of-art methods. DBSCAN follows 'eps-neighborhood' mechanism to form clusters. It is also applicable for databases with high dimensional feature space.

This technique is further improved by many researchers to form more presized clusters. Yang et al. [62] improved DBSCAN and developed Scalable Density based Clustering (SDC) technique for web opinion clustering. Since there are many social networking sites available, the methods to form clusters may vary depending on the type of data in each site. SDC technique clusters the data initially based on the density between the data. The density levels *'minimum'* and *'maximum'* are predefined and hence the data should be of *'minimum density'* reachable from each other to form clusters. Once an initial small cluster is formed, then the scalable distance is used to expand the cluster size. After forming the clusters, the theme under each cluster is identified. This method is better than DBSCAN which is proved by the experiments conducted. Thus SDC achieved both microaccuracy and macroaccuracy since cluster formation was based on both density and scalable distance filtering all the noise present. But the disadvantage of SDC is that it is not suitable for adaptive techniques. Hence this may further be investigated to get adjusted to adaptive techniques.

Jeswin et al. [63] investigated DBSCAN further more and improved the SDC technique. Here the method to form initial cluster is changed from having *'minimum density'* and selected new parameter *'closeness between data'* to form clusters. Depending on the closeness, data is grouped to form clusters. This method is better compared to DBSCAN. Since the need of users is high which includes the adaptive techniques/methods for forming clusters in different sites and situations, thus investigated method has to be further refined to achieve this. Initial clustering is a major challenge since it requires similarity within the data to get clustered. In order to find the word or phrase or sentence similarity, there are many methods available out of which Pattern Extraction is one of the mechanisms used, which is discussed later in this paper.

The DBSCAN method is generalized in two directions and four applications. The Grid based DBSCAN (GDBSCAN) is developed by Jorg et al. [64] to determine the density-connected sets and noise in spatial databases by using 2D, 3D, 5D points and 2 polygons. The GDBSCAN algorithm clusters point and spatially extended objects which demonstrated that this algorithm may be applied to real-world problems. The other method to start initial clustering is in finding principal components from the database using Principal Component Analysis (PCA). Then the clustering algorithm is applied on the principal component, based on which all possible clusters are obtained.

Saranya et al. [65] developed improved BSP (Business System Planning) algorithm which groups objects in social networks into different classes through their links whereas the relations among the classes is also identified dynamically. The advantage of this improved BSP algorithm which is in integration with PCA technique is that it requires minimal memory to from classes or clusters and shows noticeable performance in classification accuracy and time.

To find the clusters naturally from a given set of parameters, Levent et al. [66] developed Shared Nearest Neighbor (SNN) algorithm, which finds clusters of different shape, size and even different densities in spite of noise. This method uses similarity measures to find core points and builds clusters around it. The data with high dimension and noise is used for experimentation and found clusters of varying shape, size and densities. Clusters found using SNN algorithm represent relatively consistent uniform regions with respect to their surroundings when compared to DBSCAN and k-means.

Ergun et al. [67] automated the process of clustering using local density information. Differential Evaluation (DE) is a classical method used in automatic determination of clusters in the database. Swagatam et al. [68], by tuning the parameters of DE i.e., modifying chromosome representation scheme to improve the convergence properties, developed Automatic Clustering using an improved Differential Evolution (ACDE) Algorithm. This algorithm automatically determines the optimal number of clusters.

Using visualization tools, C Yang et al. [69] have explored the complex network of Terrorists to determine the network structure. They have proposed a Spring Embedded algorithm which displays the Terrorist network by optimizing the usage of 2D space. This algorithm uses both fractal and fish eye view and the experiments conducted show that the combination of these two views yields more efficiency.

Search engines are effective in recommending suggestions for the users in an effective way.

To estimate temporal and geographical provenance of handwritten document and to discover correlations between historical documents Sheng et al. [70] have proposed Multi Label Self-Organizing Map (MLSOM) in a single framework. Experimental results on medieval paleographic scale data show promising results.

Based on neighbourhood difference and density based notion of clusters, Nagaraju et al. [71] proposed an improved version of DBSCAN for detecting embedded and nested adjacent clusters. Results of experiments showed that the proposed work outperforms when compared to DBSCAN. Baydoun.et al. [72] have presented parallel implementation of K-means clustering algorithm considering both Cilk Plus and OpenMP.

4. CONTENT ANALYSIS

Since it is easy with less conditions to create user profile in social networks, some malicious users find it easy to post their unrevealed topics and grab the attention of others. Hence some wrong information may also take publicity here.

The authors [73] have proposed solution for bloggers identification in community blog site that are both influential and productive. For the proposed solution, based on the number of incoming links, two matrices are considered characterizing the bloggers into recently influential / productive or both or none. By using proposed method, it is easy to identify blogger's activities, temporal patterns and behavioral patterns and the experiments conducted on Engadget dataset establish the facts.

The authors [74] have proposed MASS model for effectively mining and identifying the top-k influential bloggers based on parameters such as comments, domain of interest and page link network authority. The developed model can be used for multiple application scenarios. Eunyoung et al. [75] have demonstrated a framework to identify influential and popular bloggers and considered weightage of readers to determine influential bloggers. The Quantifying Influence Model (QIM) is also proposed to identify influential bloggers by an interdisciplinary way. To rank the bloggers, PageRank algorithm is adopted. SimRank algorithm is proposed by Chang Sun et al. [76] to recommend influential bloggers in the network. Experimental output proves that the algorithm can be used in recommendation applications of bloggers. The authors [77] have proposed a model that creates the similarity metric using Term Frequency and Inverse Document Frequency (TF-IDF) consisting of TF-IDF and cosine similarity which proved effective than existing models.

Emily Hill et al. [78] have conducted qualitative study of stemmers on software domain, source code, specifically on java. To check the impact of stemming and retrieval effectiveness, they used Mean Average Precision (MAP) and Rank measure and conducted quantitative study of query-by-query. The author [79] has developed a model to extract keywords in entire text posted by the users based on relativity weight measure. To find the strength of terms and their relationships, they also measured the semantic similarity. The parameters considered in the model were word pair similarity, average similarity, semantic relatedness, keywords list stability and similarity score normalization. The experimental results achieved enhanced recall and precision extraction values compared to other existing approaches.

Moral et al. [80] have presented few stemming algorithms used for retrieval of information. Advantages and disadvantages of algorithms were explained and current status of stemming process gives the detailed assessment with its historical evolution. TWIG stemming algorithm [81] which produces meaningful stems and reduces the error rate. With proportion of unmeaningful to meaningful words, it gives the error rate. Experiments were conducted on Alan Beale's Core Vocabulary Dictionary data and proved that TWIG algorithm gives best results compared to STANS and Porters stemming algorithm.

Boom et al. [82] have proposed hybrid method to reduce the impact and effect of terms which are less informative. The proposed method combines the strength of dense distributed representations and TF-IDF method's sparse term matching. A 3D visualization technique [83] is used to track and analyze changing activities and interests of bloggers in the network based on phrase dependency structures in sentences. A weighted link-based rank [84] is used to identify the influential bloggers. Bipartite graph is modeled based on relationship between posts and bloggers, thus relativity based each link has some weight. The bloggers are linked and their behaviors are analyzed through semantically analyzing documents. Closerknit groups [85] are formed in the social network for bloggers based on the topics and their behavioral analysis.

Parato model [86] is proposed for forecasting the behavior of bloggers based on the number of posts and activeness in the network. The impacts and effects of explosive communication by microblogging describe the evaluation model [87]. The public and personal fields have a substantial impact. It proved that the user with highest User Content Power (UCP) as the influential blogger. According to authors [88], UCP is the summation of Document Content Power (DCP) of all the documents of a blogger. The number of times activities performed on that document of a blogger by other bloggers in the network is the DCP of a document. A model to view real-time updates in social media and statistical data of their respective profiles collected on daily basis is proposed [89] which provides data from the user profiles of facebook, youtube, slideshare and twitter through its APIs.

The authors [90] have proposed a seller model for social network, where nine sub and three main category variables are considered for its content analysis. The network indicators are used to get access to the data, in-turn graph theory and extracted the interaction patterns through network analysis and adjacency matrix techniques. The experimental results proved that content analysis identifies the users and central actor who have not participated in the discussion. Boudiba et al. [91] demonstrated a new approach to profile users in online social networks using folksonomies. The keywords or tags of content indexing created a tripartite hypergraph.

A normalized measure is used to identify the interests of users and provides degree of preference based on vertices and weight of graph [92]. A visualization toolkit [93] is proposed to identify, analyze the propagation, to influence and prediction in multiple views of location distribution. The authors [94] have analyzed the features of information content instead of identifying individual's features. The proposed framework is used to analyze and visualize the social network gathered using query dependent exploration. Experiments were conducted with Twitter data on cancer topic and applied

social cognitive theory [95] to analyze the conversation pattern, information propagation and its nature.

Content Analysis Mechanism [96] presents features of available content analysis models in discussion forums for cognitive aspects of students learning. The proposed approach cannot execute due to scarcity of data, particularly in twitter due to the characters restriction. To overcome this drawback, Adham et al. [97] explored complete knowledge in tweets. Most sophisticated data gets bootstrapped from rich sources. The Dirichlet process presented the complex topic structure and describes the effectiveness of proposed model on autism-related tweets. Authors [98] have analyzed recent research topics from Taiwin in the field of Information Technology using content analysis mechanism. The papers were categorized into 5 catalogs by considering paper-tile and content. With this approach, they found increasing and decreasing researches in IT-related issues.

Nitin Agarwal et al. [99] demonstrated an approach to aggregate bloggers in innovative ways for collective wisdom and to employ contextual information. The existing challenge was in aggregating the similar bloggers. The parameters such as profile similarity and category similarity were considered which identifies the similar bloggers by using a collective wisdom based search approach. Faiza et al. [100] attempted to identify the gender of blogger in online social network based on text written in their documents and classified specific words into classes based on specific features. By using these parameters and assigning a score to each blogger, the gender of the blogger is found. Results proved its effectiveness using Corpus data set. Bi Chen et al. [101] proposed three models such as Extreme Learning Machine, based on two regressive techniques and Modified General Regression Neural Network to merge the social dimensions, content and temporal dimensions for modelling the bloggers behavior experimenting on DailyKos blog data set.

5. FACE DETECTION

In the given image, objects other than face(s) may also be available. Depending on the situation or other terms, the photos in the database may vary largely in illumination, occlusion, view, dimension etc. Based on the specific application, the mechanism used to detect face(s) in the given image varies. Fig. 3. shows Face Recognition Mechanism whereas Table. II. shows various face recognition algorithms and their features. From the basic knowledge, it is easy to detect faces if it gets differentiated from the other objects in the image.

For this purpose, by using gradient magnitude images, Kang-Seo et al. [102] have proposed a geographical face model. Geographically the existence of face is identified and by using 33 block rank patterns face detection algorithm, face in the given image is detected and even for the illumination changes, the method performs better. As the use of face detection increased, the need for it in different situations increased like real-time face detection [103], face detection in videos, live streams etc.
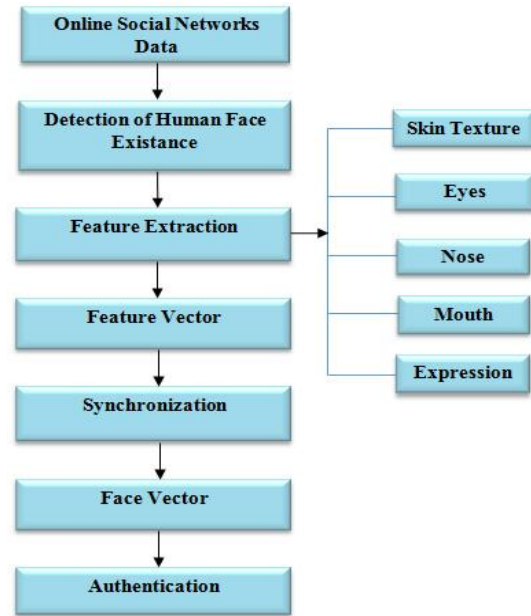
Fig. 3. Face Recognition Mechanism

Basically Viola-Jones algorithm and Haar-like features (Adaboost) are in existence since years for facial feature extraction. Researches have been using either of these two and improved upon it in a positive way for face detection. Arundati et al. [104] have followed the same way and improvised the algorithm to detect face in the live video. They first used Haar based algorithm to detect the face. Then from the detected face area using the proposed mechanism [105] they extracted eyes, eyebrows, nose, and lips. Experiments were conducted after rotating images horizontally and vertically proving that the method works better even with rotation and also during harsh light and unexpected occlusions.

With Viola-Jones as base algorithm and utilizing the Adaboost, Adrian et al. [106] have developed a framework on latest Graphics Processing Unit (GPU). This process also incorporates OpenCV and Compute Unified Device Architecture (CUDA) for frontal face detection. With this process the complete performance of the system improved for frontal face detection and results of experiment proved that this is 18 times better than other conventional GPU based algorithms. Ijaz et al. [107] presented an improved and hybrid mechanism for detecting face based on skin color. This process detects the face faster by differentiating other objects from face and this also reduces the time taken for the detection process. Face area detection utilizes more time since there can exists many other elements other than face. After the area detection, it becomes easy to extract the features in the face.

Considering these two as base, Liuliu et al. [108] for the purpose of decreasing the time taken during face detection, uses skin color detection process in UCS (Uniform Color Space). By this, the time taken by Adaboost in searching and detecting the multi-pose faces got reduced. Once the face area is detected, then Adaboost is used for candidate detection. Learning algorithm and learning code has been described by Yi-Qing [109] for face detection in colored images. Since skin color detection in colored images varies, this combined learning algorithm made face detection faster and accurate. In certain images, few objects resemble human face and hence

chances of detecting other objects as face are high. This kind of situation may vary image to image or it may also depend on application. To address this issue and to detect face accurately, Sureshkumar et al. [110] have presented a high performance image retrieval system. Here it uses indexing mechanism to identify the attributes of face.

Face detection methods have many challenges like non-frontal face, face with different expressions, image with illumination, varying light effects, age, occlusions, non-face objects, too many objects, multi-face etc. To overcome all the challenges by a face detection system and to build such an accurate system is still an open research topic. One of such mechanisms is localization [111], where the objects are localized and then objects are detected as face or non-face objects. This process improves speed of face detection by reducing time taken for unnecessary scanning of full image.

Active test framework [112] detects face faster and localizes it using mutation information gain heuristic approach proposed by novel search techniques and hierarchical model for reducing search space. This method increases the speed of face detection process by locating the faces efficiently which has been proved from the results of experiments conducted. In some scenarios, considering the whole image to localize the face and its attributes may be unnecessary since it takes more time. The solution for the said issue would be classification, where the given image is classified and then the localization is applied upon it. This reduces the time taken and increases the detection rate by accurately detecting faces with their attributes. Matting algorithm [113] segments human faces and then localizes the face attributes on the segmented parts individually. Its segmentation process expanded the application of face detection for computer vision, face recognition and even for security issues. As already mentioned, classification of objects in the given image will yield in high face detection rate.

Paul et al. [114] have proposed a framework which detects objects and increases image processing rate. It first creates an integral image representation and computes their features and transfers it to detector. Later, it is passed through an efficient classifier developed using learned algorithm based on Adaboost for identifying critical features. At last, all the classifiers are combined to detect the face. The proposed framework detects objects accurately by minimizing computational time. For the human body detection based on classification and synchronization, Hatice et al. [115] have proposed a model which determines temporal segments automatically for modality fusion synchronization. Here the individual models for detection purpose have been trained separately. Monomodel affect recognition is obtained as first level using neural and apex frames. Using fused apex phase, body modalities and affective face are synchronized. Obtaining different modalities and combining their probabilities, the decision is taken.

Automatic face segmentation techniques [116] acquired landmark detection in images and the method is robust in identifying faces with different facial expressions. Usually when there are more training sets available, face recognition becomes easier wherein the given image is compared with few possible matched training sets to accurately recognize and match the faces. But if there is only one single training sample available, then global face recognition will be challenging. To address this issue, Locality Preserving Projection (LPP) feature transfer algorithm [117] utilizes image segmentation mechanism and generalization ability promotion technique.

Here as a initial step, sources are screened using whitened cosine similarity to get selective sample source. With the help of feature transfer matrix, mapping between faces are calculated to find their relationship and using nearest neighbour classifier, faces are recognized. The proposed method is experimented on FERET, ORL and Yale data sets and results proved that the proposed single-sample face recognition algorithm LPP performs better than the available Block Fisher Linear Discriminant Analysis (FLDA) [118] and $(PC)^2A$ [119] algorithms.

Table. II. Face Recognition Algorithms

Algorithm	Concept	Data set	Advantages	Disadvantages
3 x 3 block rank patterns and a geometrical face model [102]	By dividing the illumination corrected image into 3 x 3 blocks, rank pattern identifies whether the divided region contains face or not.	Images of Caltech database and real images	1) 3 x 3 rank pattern in 2D obtained facial features. 2) It works better than Adaboost and NN algorithms during illumination changes.	May not perform as expected during occlusions.
Face detection and tracking [104]	Detects face(s) in live video stream with the help of improved Haar based algorithm	Video from webcam	1) Detects and tracks face(s) with different expressions. 2) It works pretty well during small occlusions.	When occlusions are high and image is rotated for high degree, then the method fails.
Hybrid Detector Design [107]	Combined Viola-Jones and skin color pixel detection technique for better results.	Own data of 10,000 images	1) Detects face and its features accurately. 2) It consumes less time compared to individual methods.	Not experimented on real time images.
Skin color detection and Adaboost [108]	First obtains candidate area using skin color detection and applies Adaboost algorithm to detect human face.	Own dataset	Detects single or multiple faces in the given image faster.	Lacks experiments on real time images and comparison methods.
Attribute enhanced sparse coding and	Automatically detects human attributes with high level	Large-scale database	1) Retrieves images from large scale database much	Real time implementation may

inverted indexing [110]	features.		faster. 2) May improve image retrieval up to 43.55%.	yield very low results.
Single-sample face recognition algorithm on LPP feature transfer [117]	Image segmentation and generalization ability	FERET face database	1) Recognizes face with single-sample. 2) Works accurately even during expression and illumination changes. 3) Proposed LPP is superior to PCA and LDA.	Feature transfer from source to target may consume more computational time.
Simultaneous face hallucination and identification [125]	Low resolutions considering hallucination	LFW, GT, AR and FERET databases	Recognizes face by retaining the high-frequency details and holistic structure of facial images.	Requires prior knowledge of HR-LR training pairs.
SLSP [126]	Overcomes noise problem in local structure pattern like modified census transform and local binary patterns.	MIT + CMU and face detection data set	1) SLSP is robust in detecting face(s) with illumination changes and noise in the image. 2) It has fusion of region and point features characteristics.	It yields low results when illumination is high.
2D-LDA and 2D-PCA [128]	Investigations on 3D face recognition methods which are robust to illumination changes and facial expressions.	Gavab database	1) 2D-LDA and 2D-PCA on 3D image database achieves less error rate compared to fisher-surface and eigen-surface approaches. 2) 8.96% optimized ERR obtained using 2D-LDA and proved to be best compared to other methods.	Based on light and illumination, sometimes instead of nose, chin is wrongly identified and it requires smoothing of images.
3D PFER-GEM [129]	For different angles and poses.	CMU-PIE, FERET, Bosphorus and LFW	1) Recognizes face with variation in pose and expressions rapidly. 2) Promising results on real-time images with non-control and control situations.	Illumination and occlusion changes may affect the results since it is not addressed.

Using Local Vector Pattern, a new feature proposed by Huang et al. [120] is efficient to match *M* possible candidates. Using Bilateral Recognition feature point, they provide final result. By repetitive backward and forward recognition with feature point detection, the proposed method performs better compared to Elastic Bunch Graph Matching (EBGM) [121] and Patterns of Oriented Edge Magnitudes (POEM) [122] in real time. Using two field feature extraction scheme, Chih-Rung et al. [123] have developed a new cascade face detection architecture. This process reduced the memory space required for storing the features and hence resulted in high face detection rate. This method extracts features faster and reduced the overall average detection rate which is proved from the results of experiment conducted. For face detection with different facial expressions, Mandeep et al. [124] have proposed a Singular Value Decomposition (SVD) which classifies the emotions. By identifying the features of different emotions accurately, it achieves high detection rate. The results of experiments on real time and on JAFEE database proved the accuracy of the method. By using SVD, Jian et al. [125] have developed a framework which simultaneously recognizes hallucination and LR (Low Resolution) faces in the given image.

Based on regional local differences, a new Semi-Local Structure Pattern (SLSP) architecture is proposed by Kyungjoong et al. [126]. This feature extraction method transforms noise in the given image into uniform patterns. With the binary code, the method calculates the internal image faster when compared to conventional face detector methods. Existence of forensic elements in the image also reduces the accuracy of face detection. Juefie-Xu et al. [127] have conducted a study on forensic face processing techniques. This method is based on Commercial Off-The-Shelf (COTS) for face recognition, eliminating forensic elements like occlusion in the image. Results proved that COTS performs better in occlusion and also helps in cropping.

Researchers have developed many advanced methods for face detection in 2D images but presently few algorithms / mechanisms are available for face detection in 3D images. Challenges in 3D images are similar to that in 2D images. Dhara Marvadi et al. [128] have proposed a 3D face recognition approach with different facial expressions and variations in illumination. This approach employed 2-Dimensional Linear Discriminant Analysis (2D-LDA) and 2D-PCA approaches instead of local Linear Discriminant Analysis (LDA) and PCA. By using both the approaches, they recognized faces in 3D images. The results of experiment showed that 2D-LDA performs better than 2D-PCA.

As expression and pose changes, face recognition becomes difficult. For this purpose, Moeini et al. [129] developed a model which reproduces a 3D model using frontal human face. Then through iterative process, the features are extracted from rotated image and stored in Feature Library Matrix (FLM). Later, the values are computed to get the face value. Experiments on FERET, LFW, Bosphurus and CMU-PIE demonstrated that face with expression and pose variations are recognized efficiently and with better performance.

An android application Secure Me is developed by Christina et al. [130] for extracting similar faces accurately. This application provides security for the data storage and it uses authentication techniques for this purpose. Viola-Jones framework which is openly available in OpenCV can be integrated with other techniques to improve the efficiency of face detection. In order to improve the speed and accuracy of face detection, image scaling techniques like Bicubic [132], Nearest Neighbor [133], Bilinear [134], Extended Linear [135] and Piece-wise Extended Linear [136] are integrated with viola-jones framework and are examined by Himanshu et al. [131] and the results are compared. The two parameters used to test these algorithms are PSNR and SSIM. Results prove that Bicubic algorithm performs better compared to other algorithms considered here.

6. GEO-TAGGING

Geo-tagging mechanism adds access location (longitude and latitude) information of the place where the user is signed-in or the photo taken etc., which helps in understanding the user activities. The geo-tagging is achieved using Global Positioning System (GPS). The details like user login location, place of accessing various websites, photo location and GPS tagged photo sharing etc., helps in identifying the pattern of user activities. From the location of the user and the activities performed in various locations, the users are profiled. User's friends' activities and their location information may also be considered to profile the user. Fig. 4. shows various components of Geo-Tagging.

Zhang et al. [137] constructed a transmission graph by defining event interactions occurred among the population and strongly urge that the interaction duration is directly proportional to the number of individuals which differs completely from truncated power-law pattern. Recommendation plays a vital role in OSNs where users from different location provide huge information. Pavlos et al. [138] conducted a survey on algorithms used for Location Based Social Networks (LBSN). Three new perspectives: user's privacy issue, time awareness and explainability of recommendations are provided by authors and revealed entities like users, activities, groups, location, along with quality, strengths and weakness of LBSN.
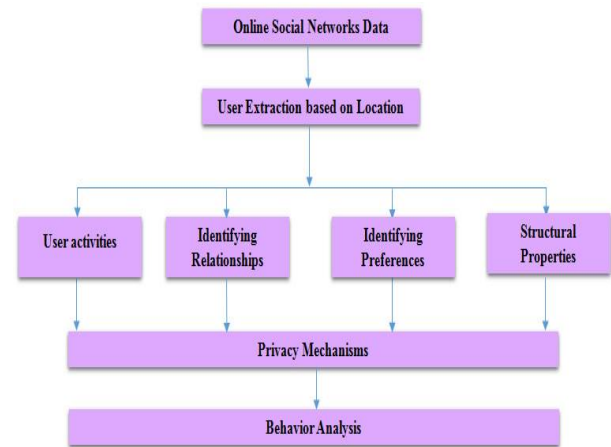

Fig. 4. Components of Geo-Tagging

As a solution for location prediction problems in friendship recommendation system in LBSN, Yang et al. [139] proposed probability model including both spatial and temporal perspectives. The model is developed based on Bayes' rule. To make users satisfied with the recommendations provided to them based on their preferences, Betin et al. [140] proposed a personalized recommender making use of location to understand users and their preferences. Experiments conducted on Gowalla data set proves that the incorporation of geo-location including different type of location into recommendation algorithm has improved the method in identifying preferences and behavior of users. To analyze the structural properties and relationship between users in social network along with their location traces, Justin et al. [141] evaluated a set of features depicting the social behavior of users applied to privacy control mechanisms.

Utilizing both location based social networks and cell phone location data, Eunjoon et al. [142] developed a model which explains human travel influenced by social network structure. They proved that the human mobility for short and periodic travels is not dependent on social network while long distance travel completely depends on ties of social network. Nitai et al. [143] developed a friend recommendation system based on analysis and characterization of network formed by friends and friends of friends. Narayanam et al. [144] proposed ShaPley value based Influential Nodes (SPIN) algorithm to compute network value for identifying the influential node in the network. It also provides insight of information diffusion in the network.

A combination of collaborative filtering content based recommendation system is introduced by Souvik et al. [145] to recommend users based on their interest. Sparse regression and isometric projection are dominant and based on the users' posts in social media and by applying collaborative filtering, the user's interest is recorded. By this collective information and ranking obtained, a friend recommendation system is proposed by Wan et al. [146]. Experiments conducted on Digg data set proves that the system outperforms.

Sequential user behavior [147] is analyzed which provides design mechanisms for social computing systems and to analyze the behavior of neighbors for reciprocating information in the network. Chung-Kai et al. [148] conducted a study on distributed information sharing network which helps self-interested agents to judge the neighbors, whereas a

similar decision making methodology is presented by Yunlong Wang et at. [149].

Jalal et al. [150] built a crowd-source 'qCrowd' which automatically selects strangers in Twitter to get response for the information requested, proving both recommendation algorithm and prediction model. Twitter is used to analyze the real time traffic-related events. In order to make the users of social networks secure from others who keep requesting for friendship in social ties, IntRank [151] recommends trustworthy friends to community members. On repeated analysis over the pattern of interactions, social ties and ranking, this method filters out to provide recommendations. For secure access and usage in social networks, symmetric key encryption mechanism [152] searches the data in decentralized mobile networks when given with preference search criteria.

7. BEHAVIOR ANALYSIS

Growing usage of social networks is both beneficial and harmful for the society. Availability of huge information, accessing family and friends from geographically far locations, fast information processing than other medias etc., are helpful to people. But this is wrongly being utilized by some people, like creating and spreading fake information and performing illegal activities through social networks. In order to safeguard the society and stop all harmful activities being performed by fake information creators, profiling of such users by analyzing the behavioral patterns of those users need to be determined. These intruders also play a vital role in influencing other users for their illegal activities. Fig. 5. shows various aspects of how behavior of users is analyzed.

To avoid being recognized as intruders in social networks, they play some tactics and behave differently in different social networks. In other situations, the behavioral analyses of those users who reach many other users and share information for a good cause are in need to be identified. Usually such situation comes in trading and marketing, which is one of the business promotional strategies. Based on the behavioral patterns of user in different social networks, their activities and involvement, the users are profiled. Table. III shows various behavior analyses algorithms used during user profiling.
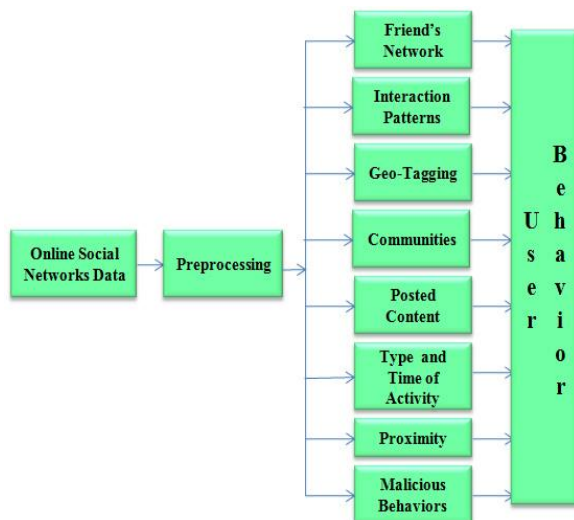

Fig. 5. Aspects of User Behavior

In order to promote some business and increase the market value for their products, business people take help of predictors in social networks. Predictors are those who identify the active users in social networks for the advertisement. Sharad et al. [153] have developed a method which contextually targets to assess the value of social predictors. Wrongly assessing the predictors will lead to loss in the business. Hence, predicting individual's behavior in social networks and identifying appropriate candidate to apply for marketing strategies became easy with this approach. Based on age, sex, demographic, geographic locality and actions performed in social networks, behavior of individuals is predicted and the user with high prediction value will be selected for marketing promotions. Comparatively very few people perform actions leading to disadvantage.

Various statistical methods can be applied to track the pattern of user activities in social networks. Identifying activities of user helps in characterizing the user, in turn gives higher contribution for profiling that user. Fabricio et al. [154] used a statistic named ClickStream which is based on the URL clicks of user and identified the pattern that the user has followed. They made use of social network aggregator to identify patterns and social interactions. Since an user may involve in different social networks and different activities, this algorithm manually classifies user activities and related statistics. Hence, the pattern of user in different networks will be traced appropriately. The disadvantage of this algorithm is that it limits the data set whereas the data log contains only the information that is part of HTTP or URL of the request which clearly tells that other activities will not be tracked and which mainly works on manual statistics only. In future, they also intend for spam analysis which includes multiple social networks and to automate the process.

In social networking sites, influence plays a vital role and here, the users can influence others in many ways and on many different things. For marketing, statistics is a boon whereas the intruders take this as an opportunity and influence others in the network and involve in illegal activities. Writing blogs, sharing videos, posting photos etc., in different ways, users influence others. By analyzing the photos that user has posted, the geographical locations that the photos got posted from, geotagging is achieved.

Manos et al. [155] have developed this Geotagging system which uses GPS manual to detect the locations of photos getting uploaded in the social media. This method helps in mapping and avoids duplication. By this method, user activities, type of photos and the place from where they access social site is tracked which helps in finding the user behavior. The kind of photos that get uploaded depicts the user character and the kind of influence they try to create on others. Since the use of mobile phones have increased drastically especially by youth, the communication through text, photos, videos between friends has also increased. Here, the interest of user will depend on time. Example, the post or photo from a friend can influence the other and can lead to some actions, hence the correlation between users and the time of action are the parameters needed to find the influence factor and the list of activities by users.

ACTPred [156] predicts user's activity at '$t+1$' based on either his or his friends' activities until 't'. Correlation analysis depends on attribute, temporal and social activities of users. This algorithm works by constructing a graph model and then the learned parameter is applied on the graph to

make an inference. The experiments conducted shows clear improvement over the comparison methods and accurately predict users' activities. Apart from individuals influencing others, the trend has changed to form a community and then trying to influence others through communities by posting photos, videos etc.,. Here, the actions from an user in community is considered as the action of community, and very importantly such communities may last long or vanish within a short period of time.

Sofia et al. [157] have developed a method which identifies the online social communities and its user members. Using ontology and skeleton rule clustering, the behavioral roles are defined and the users are labeled with those behavioral roles. Behavioral pattern of communities are noted which predicts whether the communities formed will flourish or diminish. This application includes 4 steps: Constructing features, Deriving bins, Compiling Rule Base and Applying Rules. The only disadvantage is, about 29% of the users remain unclassified. Ahmed et al. [158] have proposed a model called Intrusion detection system which detects the unusual activities

and depicts the behavior of users with less human intervention. Bianconi et al. [159] have innovated the integration of GIS in S.G.RTUP tool for monitoring user activity.

This approach helps in dynamic data visualization and statistical analysis. Since it is applied on health care domain, it also helps in co-relating local clusters with socio-economic data. Paek et al. [160] have presented a mechanism to understand the behavior of users in accessing the information present in the internet. Yasemin et al. [161] have explored machine learning based classification approach to identify malicious behavior using network flows. C4.5 and Naïve Bayes are the two techniques used. David et al. [162] have proposed measures and proved that this approach can be used effectively in extracting future interactions using network topology. To analyze the student's behavior in online learning selection, Educational Data Mining (EDM) is used by Ratnapala et al. [163] and then students were grouped using K-means clustering.

Table. III. Behavior Analysis Algorithms

Algorithm	Concept	Data set	Advantages	Disadvantages
Improved targeting and prediction [137]	Social influence on individuals behavior	Yahoo! Data	1) Reveals social data and influences others. 2) It complements both demographic and behavioral Predictors.	1) Study is based on only two communication networks. 2) It may yield low results when individuals have less contact.
Clickstream Model [138]	Identifying patterns in social interactions and social network workloads	Clickstream (Orkut, MySpace, Hi5 and LinkedIn) data and Orkut crawled data	Characterized social network workloads.	Specific nature of respective social networks is not considered.
Methodology on Geotagging innovation [139]	Examining how individual's behavior affects decisions of friends.	Flickr Data	1) Method proposed is generic and may be used in real time marketing systems. 2) It provides both macroscopic and microscopic views.	It cannot be applied on large scale data sets.
ACTPred model [140]	Activity prediction in mobile social networks for predicting and modeling user's activities	University personal mobility data	1) A unified factor graph model is developed which formalizes discovered intuitions. 2) 10-20% improvement than compared methods.	Result may greatly vary if applied on real time data.
Behaviour Ontology and Rules [141]	To model semantically the behavioral characteristics of users	Boards forum data	Provides the pattern of communities and their activities.	Patterns may vary based on different communities from different forums.
Web based management system [143]	To integrate oncology network and IT	Umbria Cancer Registry	1) Geographical analysis in detail. 2) Accommodates data from different sources.	Other than statistical analysis, other factors are not considered.
Educational Data Mining techniques [147]	To analyze student's behavior with online e-learning system	Moodle modules	It has been proved that students are not self-motivated for e-learning.	Other characteristics are not been analyzed.

Analyzing the behavioral patterns of users in one social network is insufficient to profile them accurately. Hence, different social networks have been analyzed here to identify users' behavior and it is achieved using different methods which are covered and discussed in this paper.

8. CONCLUSIONS

This paper provides an insight into profiling users in online social networks. Extensive literature review has been carried out on various aspects to profile user(s) and depending on the

application, users are profiled. Year on year, the web content is increasing and the structure of web content varies broadly. Hence each application requires different techniques to analyze and conclude. Each application has its own terms and procedures in profiling and since the data available in the web is from semi-structured to unstructured, various approaches namely Clustering, Face Detection, User Activities, Content Analysis and Behavioral Analysis to profile user(s) in online social networks are presented in this paper. The concept and its applications under each topic are summarized in respective

tables. Information flow path plays a vital role in content analysis while robust algorithms acquire high position for face detection. Pattern recognition in behavioral analysis, type and number of activities in analyzing the user activities are essential key terms in profiling the user(s) in any online social network.

For future researchers, profiling users based on their activities in line with current affairs in society, friend's activities, search engine log data are essential for fraud detection, determining influential user, personalization, e-marketing and advertising etc.,. The need for research on user data and their activities in online multi social networks simultaneously is essential for accurate profiling of any social network user.

9. REFERENCES

[1] Zhang, Lejun, Tongxin Zhou, Qi Zhixin, Lin Guo and Li Xu. "The research on e-mail Users' behavior of participating in Subjects based on social network analysis," *China Communications*, vol. 13, no. 4, pp. 70-80, 2016.

[2] Veena H Bhat, V R Malkani, P Deepa Shenoy, K R Venugopal and L M Patnaik, "Classification of Email using BeaKS: Behavior and Keyword Stemming," *IEEE TENCON*, 2011.

[3] Leena Giri G, Praveen Gowda I V, Manjula S H, Venugopal K R and L M Patnaik, "In Page Semantic Ranking of Snippets for WebPages," *Sixth International Conference on advances in Computing, Control and Telecommunication Technologies – ACT 2015*, Trivandrum, India, October 2015.

[4] Itoh, Masahiko, Daisaku Yokoyama, Masashi Toyoda, Yoshimitsu Tomita, Satoshi Kawamura and Masaru Kitsuregawa. "Visual Exploration of Changes in Passenger Flows and Tweets on Mega-City Metro Network," *IEEE Transactions on Big Data*, vol. 2, no. 1, pp. 85-99, 2016.

[5] Kim, Youngsoo, Felicia Natali, Feida Zhu and Eepeng Lim, "Investigating the Influence of Offline Friendship on Twitter Networking Behaviors," *49th IEEE Hawaii International Conference on System Sciences (HICSS)*, pp. 736-745, 2016.

[6] Roy, Suman Deb, Gilad Lotan and Wenjun Kevin Zeng, "The Attention Automaton: Sensing Collective User Interests in Social Network Communities," *IEEE Transactions on Network Science and Engineering*, vol. 2, no. 1, pp. 40-52, 2015.

[7] D'Andrea, Eleonora, Pietro Ducange, Beatrice Lazzerini and Francesco Marcelloni, "Real-time detection of traffic from twitter stream analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2269-2283, 2015.

[8] Fabijan, Aleksander, Helena Holmström Olsson and Jan Bosch. "Customer feedback and data collection techniques in software R&D: a literature review," *In International Conference of Software Business*, pp. 139-153, Springer International Publishing, 2015.

[9] Li, Bo, Duoyong Sun, Julei Fu and Zihan Lin, "Social Network Construction and Analysis Based on Community Photo Collections with Face Recognition," *IEEE Sixth International Conference on Business*

Intelligence and Financial Engineering (BIFE), pp. 463-467, 2013.

[10] Riek, Markus, Rainer Bohme and Tyler Moore, "Measuring the influence of perceived cybercrime risk on online service avoidance," *IEEE Transactions on Dependable and Secure Computing*, vol. 13, no. 2, pp. 261-273, 2016.

[11] Ramachandra A C, Pavithra K, Yashasvini K, Raja K B, Venugopal K R and Lalit M Patnaik, "Cross-validation for graph matching based offline signature verification," *IEEE Annual India Conference, INDICON 2008*, vol. 1, pp. 17–22, 2008.

[12] Chetana Hegde, P Deepa Shenoy, K R Venugopal and L M Patnaik, "Authentication Using Finger Knuckle Prints", *Signal, Image and Video Processing, Springer,P-ISSN:1863-1703, E-ISSN: 1863-1711*, vol. 7, no. 4, pp. 633–645, July 2013.

[13] Ramesha K, K B Raja, Venugopal K R and L M Patnaik, "Template based Mole Detection for Face Recognition," *International Journal of Computer Theory and Engineering, (IJCTE), ISSN : 1793-8201*, vol. 2, no. 5, pp. 797-804, October 2010.

[14] Vibha L, Chetana Hegde, Deepa Shenoy P, Venugopal K R and L M Patnaik, "Dynamic Object Detection, Tracking and Counting in Video Streams for Multimedia Mining," *In IAENG International Journal of Computer Science, e-ISSN : 1819-9224, p-ISSN : 1819-656x*, vol. 35, no. 3, pp. 382-391, 2008.

[15] Akay, Altug, Andrei Dragomir and Björn-Erik Erlandsson, "A novel data-mining approach leveraging social media to monitor consumer opinion of sitagliptin," *IEEE Journal of Biomedical and Health Informatics,* vol. 19, no. 1, pp. 389-396, 2015.

[16] Naini, Farid M., Jayakrishnan Unnikrishnan, Patrick Thiran and Martin Vetterli, "Where you are is who you are: User identification by matching statistics," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 358-372, 2016.

[17] Qiao, Xiuquan, Wei Yu, Jinsong Zhang, Wei Tan, Jianchong Su, Wangli Xu and Junliang Chen, "Recommending Nearby Strangers Instantly Based on Similar Check-In Behaviors," *IEEE Transactions on Automation Science and Engineering,* vol. 12, no. 3, pp. 1114-1124, 2015.

[18] Valenza, Gaetano, Vladimir Carli, Antonio Lanata, Wei Chen, Roozbeh Jafari and Enzo Pasquale Scilingo, "Guest Editorial Sensor Informatics for Managing Mental Health," *IEEE Journal of Biomedical and Health Informatics,* vol. 20, no. 4, pp. 975-976, 2016.

[19] Mertz Leslie, "What Can Big Data Tell Us About Health?: Finding Gold Through Data Mining," *IEEE Pulse,* vol. 7, no. 5, pp. 40-44, 2016.

[20] Rubiano, Sandra Milena Merchan and Jorge Alberto Duarte Garcia, "Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic Performance," *IEEE Latin America Transactions*, vol. 14, no. 6, pp. 2783-2788, 2016.

[21] Da Silva, Luis Alexandre Estevao, "A Data Mining Approach for Standardization of Collectors Names in

Herbarium Database," *IEEE Latin America Transactions*, vol. 14, no. 2, pp. 805-810, 2016.

[22] Dong, Boxiang, Ruilin Liu and Hui Wendy Wang, "Trust-but-Verify: Verifying Result Correctness of Outsourced Frequent Itemset Mining in Data-Mining-As-a-Service Paradigm," *IEEE Transactions on Services Computing*, vol. 9, no. 1, pp. 18-32, 2016.

[23] Liu Bing, "Opinion mining and sentiment analysis," *Web Data Mining*, pp. 459-526, Springer Berlin Heidelberg, 2011.

[24] P Deepa Shenoy, Srinivasa K G, Venugopal K R and Lalit M Patnaik, "Evolutionary approach for mining association rules on dynamic databases," *Advances in Knowledge Discovery and Data Mining*, pp. 325–336, April 2003.

[25] P Deepa Shenoy, Srinivasa K G, Venugopal K R and Lalit M Patnaik, "Dynamic association rule mining using genetic algorithms," *Intelligent Data Analysis*, vol. 9, no. 5, pp. 439–453, September 2005.

[26] Heredia, Diana, Yegny Amaya and Edwin Barrientos, "Student Dropout Predictive Model Using Data Mining Techniques," *IEEE Latin America Transactions*, vol. 13, no. 9, pp. 3127-3134, 2015.

[27] Saâdaoui, Foued, Pierre R. Bertrand, Gil Boudet, Karine Rouffiac, Frédéric Dutheil and Alain Chamoux, "A Dimensionally Reduced Clustering Methodology for Heterogeneous Occupational Medicine Data Mining," *IEEE Transactions on Nanobioscience*, vol. 14, no. 7, pp. 707-715, 2015.

[28] Chen, Hongmei, Tianrui Li, Chuan Luo, Shi-Jinn Horng and Guoyin Wang, "A decision-theoretic rough set approach for dynamic data mining," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 6, pp. 1958-1970, 2015.

[29] Veena H Bhat, Prashanth G Rao, Abhilash, P Deepa Shenoy, Venugopal K R and L M Patnaik, "A Novel Data Generation Approach for Digital Forensic Application In Data Mining," *IEEE Second International Conference on Machine Learning and Computing*, February-2010.

[30] Kalegele, Khamisi, Kazuto Sasai, Hideyuki Takahashi, Gen Kitagata and Tetsuo Kinoshita, "Four Decades of Data Mining in Network and Systems Management," *IEEE Transactions on Knowledge and Data Engineering,* vol. 27, no. 10, pp. 2700-2716, 2015.

[31] Cereda, Paulo Roberto Massa and J. José Neto, "Adaptive data mining: Preliminary studies," *IEEE Latin America Transactions*, vol. 12, no. 7, pp. 1258-1270, 2014.

[32] Baker and Ryan S, "Educational Data Mining: An Advance for Intelligent Systems in Education," *IEEE Intelligent Systems*, vol. 29, no. 3, pp. 78-82, 2014.

[33] Fisch, Dominik, Edgar Kalkowski and Bernhard Sick, "Knowledge fusion for probabilistic generative classifiers with data mining applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 3, pp. 652-666, 2014.

[34] Shmiel, Oren, Tomer Shmiel, Yaron Dagan and Mina Teicher, "Processing of multichannel recordings for data-mining algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 3, pp. 444-453, 2007.

[35] Zhou, Xiaoping, Xun Liang, Haiyan Zhang and Yuefeng Ma, "Cross-Platform Identification of Anonymous Identical Users in Multiple Social Media Networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 2, pp. 411-424, 2016.

[36] Alameda-Pineda, Xavier, Yan Yan, Elisa Ricci, Oswald Lanz and Nicu Sebe, "Analyzing free-standing conversational groups: a multimodal approach," *23rd ACM International Conference on Multimedia*, pp. 5-14, 2015.

[37] Ruan, Xin, Zhenyu Wu, Haining Wang and Sushil Jajodia, "Profiling Online Social Behaviors for Compromised Account Detection," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 176-187, 2016.

[38] Cao, Nan, Conglei Shi, Sabrina Lin, Jie Lu, Yu-Ru Lin and Ching-Yung Lin, "Targetvue: Visual analysis of anomalous user behaviors in online communication systems," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 280-289, 2016.

[39] Shen, Haiying, Ze Li, Jinwei Liu and Joseph Edward Grant, "Knowledge sharing in the online social network of yahoo! answers and its implications," *IEEE Transactions on Computers*, vol. 64, no. 6, pp. 1715-1728, 2015.

[40] Liu, Siyuan, Shuhui Wang and Feida Zhu, "Structured Learning from Heterogeneous Behavior for Social Identity Linkage," *IEEE Transactions on Knowledge and Data Engineering,* vol. 27, no. 7, pp. 2005-2019, 2015.

[41] Miro-llinares f and j. J. Rodriguez-sala, "Cyber Hate Speech On Twitter: Analyzing Disruptive Events From Social Media To Build A Violent Communication And Hate Speech Taxonomy," *International Journal of Design & Nature and Ecodynamics*, vol. 11, no. 3, pp. 406-415, 2016.

[42] Dong, Lijun, Kui Wu and Guoming Tang, "A Data-Centric Approach to Quality Estimation of Role Mining Results," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2678-2692, 2016.

[43] Univaso, Pedro, Juan Maria Ale and Jorge A. Gurlekian, "Data Mining applied to Forensic Speaker Identification," *IEEE Latin America Transactions*, vol. 13, no. 4, pp. 1098-1111, 2015.

[44] Lin, Cheng-Jhe, Changxu Wu and Wanpracha A. Chaovalitwongse, "Integrating human behavior modeling and data mining techniques to predict human errors in numerical typing," *IEEE Transactions on Human-Machine Systems,* vol. 45, no. 1, pp. 39-50, 2015.

[45] Angiulli, Fabrizio, Stefano Basta, Stefano Lodi and Claudio Sartori, "Distributed strategies for mining outliers in large data sets," *IEEE Transactions on Knowledge and Data Engineering,* vol. 25, no. 7, pp. 1520-1532, 2013.

[46] Marquez-Vera, Carlos, Cristóbal Romero Morales and Sebastián Ventura Soto, "Predicting school failure and dropout by using data mining techniques," *IEEE Revista*

Iberoamericana de Tecnologias del Aprendizaje, vol. 8, no. 1, pp. 7-14, 2013.

[47] Cao Longbing, "Social security and social welfare data mining: An Overview," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 837-853, 2012.

[48] Huang, Xiaoge, Liping Chen, Qianbin Chen and Bin Shen, "Joint malicious user detection and resource allocation in cognitive radio networks," *IEEE 10th International Conference on Communications and Networking in China (ChinaCom)*, pp. 278-282, 2015.

[49] Terzi, Ramazan, Uraz Yavanoglu, Duygu Sinanc, Dogac Oguz and Semra Cakir, "An Intelligent Technique for Detecting Malicious Users on Mobile Stores," *IEEE 13th International Conference on Machine Learning and Applications (ICMLA)*, pp. 470-477, 2014.

[50] Jnanamurthy H. K and Sanjay Singh, "Detection and filtering of collaborative malicious users in reputation system using quality repository approach," *IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 466-471, 2013.

[51] Rabinovitch Eddie, "Protect your users against the latest web-based threat: malicious code on caching servers [Your Internet Connection]," *IEEE Communications Magazine*, vol. 45, no. 3, pp. 20-22, 2007.

[52] Vasanthakumar G U, Bagul Prajakta, P Deepa Shenoy, Venugopal K R and Lalit M Patnaik, "PIB: Profiling Influential Blogger in Online Social Networks, A Knowledge Driven Data Mining Approach," *Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015), Procedia Computer Science, Elsevier B.V.*, vol. 54, pp. 362–370, August 2015.

[53] Vasanthakumar G U, P Deepa Shenoy and Venugopal K R, "PTIB: Profiling Top Influential Blogger in Online Social Networks," *International Journal of Information Processing (IJIP-2016), IK International Publishing*, vol. 10, no. 1, pp. 77–91, June 2016.

[54] Vasanthakumar G U, Priyanka R, Vanitha Raj K C, Bhavani S, Asha Rani B R, P Deepa Shenoy and Venugopal K R, "PTMIB: Profiling Top Most Influential Blogger using Content Based Data Mining Approach," *IEEE International Conference on Data Science and Engineering (ICDSE-2016)*, Cochin, India, August 2016.

[55] Zhu, Jiaqi, Kaijun Wang, Yunkun Wu, Zhongyi Hu and Hongan Wang, "Mining User-Aware Rare Sequential Topic Patterns in Document Streams," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 7, pp. 1790-1804, 2016.

[56] Wang, Yufeng, Athanasios V. Vasilakos, Jianhua Ma and Naixue Xiong, "On studying the impact of uncertainty on behavior diffusion in social networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 2, pp. 185-197, 2015.

[57] Rawassizadeh, Reza, Elaheh Momeni, Chelsea Dobbins, Joobin Gharibshah and Michael Pazzani, "Scalable Daily Human Behavioral Pattern Mining from Multivariate Temporal Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 11, 2016.

[58] Agreste, Santa, Pasquale De Meo, Emilio Ferrara, Sebastiano Piccolo and Alessandro Provetti, "Analysis of a heterogeneous social network of humans and cultural objects," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 4, pp. 559-570, 2015.

[59] Chen, Yi-Cheng, Wen-Chih Peng and Suh-Yin Lee, "Mining Temporal Patterns in Time Interval-Based Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 12, pp. 3318-3331, 2015.

[60] Soft Computing for Data Mining Applications, Venugopal K R, Srinivasa K G and L M Patnaik, *ISBN 978-3-642-00192-5, e-ISBN 978-3-642-00193-2, Do1 10.1007/978-3-642, 00193-2*, pp. 342, Springer Verlag, 2009.

[61] Martin Ester, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *2nd International Conference on Knowledge Discovery and Data Mining*, 1996.

[62] Christopher C. Yang and Tobun Dorbin Ng, "Analysing and Visualizing Web Opinion Development and Social Interactions with Density-based clustering," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 6, pp. 1144-1155, 2011.

[63] J.R. Dcouth and T. Mohanraj, "Analyzing and Extracting Social Mining Trends Through Web Opinion Developments Via Density Based Clustering," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, Vol. 2, Issue 1, Jan 2013.

[64] Sander F, M Ester and P. Kriegelh, "The algorithm GDBSCAN and its application," *Data Mining and Knowledge Discovery*, no.2, pp. 178-192, 1998.

[65] Sharma Sanjeev and R. K. Gupta. "Improved BSP clustering algorithm for social network analysis," *International journal of grid and Distributed Computing*, vol. 3, no. 3, pp. 67-76, 2010.

[66] Ertöz, Levent, Michael Steinbach and Vipin Kumar, "Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data," *SDM*, pp. 47-58. 2003.

[67] Biçici, Ergun, and Deniz Yuret, "Locally scaled density based clustering," *International Conference on Adaptive and Natural Computing Algorithms*, pp. 739-748. Springer Berlin Heidelberg, 2007.

[68] Das, Swagatam, Ajith Abraham and Amit Konar, "Automatic clustering using an improved differential evolution algorithm," *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, vol. 38, no. 1, pp. 218-237, 2008.

[69] Christopher C. Yang, Nan Liu and Marc Sageman, "Analyzing the Terrorist Social Networks with Visualization Tools," *IEEE International Conference on Intelligence and Security Informatics*, pp. 331-342, 2006.

[70] Sheng He, Petros Samara, Jan Burgers and Lambert Schomaker, "A Multiple-Label Guided Clustering

Algorithm for Historical Document Dating and Localization," *IEEE Transactions on Image Processing*, vol. 25, no. 11, 2016.

[71] Nagaraju S, Manish Kashyap and Mahua Bhattacharya, "A variant of DBSCAN algorithm to find embedded and nested adjacent clusters," *IEEE 3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 486-491, 2016.

[72] Baydoun, Mohammed, Mohammad Dawi and Hassan Ghaziri, "Enhanced parallel implementation of the K-Means clustering algorithm," *IEEE 3rd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*, pp. 7-11, 2016.

[73] Leonidas Akritidis, Dimitrios Katsaros and Panayiotis Bozanis, "Identifying the productive and influential bloggers in a community," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 41, no. 5, pp. 759–764, 2011.

[74] Yichuan Cai and Yi Chen, "Mass: a multi-facet domain-specific influential blogger mining system," *2010 IEEE 26th International Conference on Data Engineering (ICDE)*, pp. 1109–1112, 2010.

[75] Eunyoung Moon and Sangki Han, "A qualitative method to find influencers using similarity-based approach in the blogosphere," *International Journal of Social Computing and Cyber-Physical Systems*, vol. 1, no. 1, pp. 56–78, 2011.

[76] Chang Sun, Bing-quan Liu, Cheng-jie Sun, De-Yuan Zhang and Xiaolong Wang, "Simrank: A link analysis based blogger recommendation algorithm using text similarity," *2010 International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 6, pp. 3368–3373, 2010.

[77] Mohammad Alodadi and Vandana P Janeja, "Similarity in patient support forums using tf-idf and cosine similarity metrics," *IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 521– 522, 2015.

[78] Emily Hill, Shivani Rao and Avinash Kak, "On the use of stemming for concern location and bug localization in java," *IEEE 12th International Working Conference on Source Code Analysis and Manipulation (SCAM)*, pp. 184–193, 2012.

[79] Mohamed H Haggag, "Keyword Extraction using Semantic Analysis," *International Journal of Computer Applications*, vol. 61, no. 1, 2013.

[80] Cristian Moral, Ange´lica de Antonio, Ricardo Imbert and Jaime Ram´ırez, "A survey of stemming algorithms in information retrieval," *Information Research: An International Electronic Journal*, vol. 19, no. 1, p. 1, 2014.

[81] S Megala, A Kavitha and A Marimuthu, "Improvised Stemming Algorithm–TWIG," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 7, pp. 168–171, 2013.

[82] De Boom, Cedric, Steven Van Canneyt, Steven Bohez, Thomas Demeester and Bart Dhoedt, "Learning Semantic Similarity for Very Short Texts," *IEEE International Conference on Data Mining Workshop (ICDMW)*, pp. 1229-1234, November 2015.

[83] Itoh, Masahiko, Naoki Yoshinaga, Masashi Toyoda and Masaru Kitsuregawa, "Analysis and visualization of temporal changes in bloggers' activities and interests," *Visualization Symposium (PacificVis)*, pp. 57-64, February 2012.

[84] Lu Fuxi Zhu, "Discovering the important bloggers in blogspace," *IEEE International Conference on Artificial Intelligence and Education (ICAIE)*, pp. 151-154, October 2010.

[85] Macskassy and Sofus A, "Leveraging Contextual Information to Explore Posting and Linking Behaviors of Bloggers," *IEEE International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 64-71, August 2010.

[86] Rui, Cai, Qi Jia-yin and Wang Mian, "Forecasting bloggers' online behavior based on improved Pareto/NBD model," *IEEE International Conference on Management Science and Engineering (ICMSE)*, pp. 8490, July 2013.

[87] Zhang, Yuan and Yuqian Bai, "Research on the Influence of Microbloggers, Take Sina Celebrity Micro-blog as an Example," *IEEE Eighth International Conference on Semantics, Knowledge and Grids (SKG)*, pp. 189-192, October 2012.

[88] Seung-Hwan Lim, Sang-Wook Kim, Sunju Park and Joon Ho Lee, "Determining content power users in a blog network: an approach and its applications," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 41, no. 5, pp. 853–862, 2011.

[89] Riccardo Cognini, Damiano Falcioni and Alberto Polzonetti, "Social networks: Analysis for integrated social profiles," *IEEE Internet Technologies and Applications (ITA)*, pp. 68–72, 2015.

[90] B. Erlin, Norazah Yusof and Azizah Abdul Rahman, "Analyzing online asynchronous discussion using content and social network analysis," *IEEE Ninth International Conference on Intelligent Systems Design and Applications*, pp. 872–877, 2009.

[91] Boudiba Tahar-Rafik and Ahmed-Ouamer Rachid, "Towards a new approach for generating user profile from folksonomies," *IEEE 4th International Symposium on ISKO-Maghreb: Concepts and Tools for knowledge Management (ISKO-Maghreb)*, pp. 1–6, 2014.

[92] Yi Cai and Qing Li, "Personalized search by tag-based user profile and resource profile in collaborative tagging systems," *19th ACM International Conference on Information and Knowledge Management*, pp. 969–978, 2010.

[93] Bo Wang, Yingjun Sun, Cheng Tang and Yang Liu, "A visualization toolkit for online social network propagation and influence analysis with content features," *IEEE International Conference on Orange Technologies (ICOT)*, pp. 129–132, 2014.

[94] Christopher C Yang and Tobun D. Ng, "Terrorism and crime related weblog social network: Link, content analysis and information visualization," *IEEE Intelligence and Security Informatics*, pp. 55–58, 2007.

[95] Hong-Jun Yoon and Georgia Tourassi, "Analysis of online social networks to understand information sharing

behaviors through social cognitive theory," *IEEE Annual Oak Ridge National Laboratory Biomedical Science and Engineering Center Conference (BSEC)*, pp. 1–4, 2014.

[96] Noor Izzati Ariff and Zaidatun Tasir, "Meta-analysis of content analysis models for analysing online problem solving discussion," *IEEE Conference on e-Learning, e-Management and e-Services (IC3e)*, pp. 148–152, 2015.

[97] Adham Beykikhoshk, Ognjen Arandjelovic, Dinh Phung and Svetha Venkatesh, "Overcoming data scarcity of Twitter: using tweets as bootstrap with application to autism-related topic content analysis," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 1354–1361, 2015.

[98] Yung-Chung Tsao, Kevin Chihcheng Hsu and Yin-Te Tsai, "Using content analysis to analyze the trend of information technology toward the academic researchers at the design departments of universities in Taiwan," *IEEE 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, pp. 3691–3694, 2012.

[99] Nitin Agarwal, Huan Liu, Shankara Subramanya, John J. Salerno and S. Yu Philip, "Connecting sparsely distributed similar bloggers," *Ninth IEEE International Conference on Data Mining*, pp. 11–20, 2009.

[100] Faiza Belbachir, Khadidja Henni and Lynda Zaoui, "Automatic detection of gender on the blogs," *IEEE/ACS International Conference on Computer Systems and Applications (AICCSA)*, pp. 1–4, 2013.

[101] Bi Chen, Qiankun Zhao, Bingjun Sun and Prasenjit Mitra, "Predicting blogging behavior using temporal and social networks," *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pp. 439–444, 2007.

[102] Kang-Seo Park,Young-Gon Kim and Rae-Hong Park, "Face Detection Using The 33 Block Rank Patterns Of Gradient Magnitude Images," *Signal and Image Processing : An International Journal (SIPIJ)*, vol. 4, no. 5, October-2013.

[103] Kailash Devrari and K.Vinay Kumar, "Fast Face Detection Using Graphics Processor," *International Journal of Computer Science and Information Technologies (IJCSIT)*, vol. 2, no. 3, pp. 1082–1086, 2011.

[104] Arundhati Das, Mameeta Pukhrambam and Ashim Saha, "Real-Time Robust Face Detection and Tracking using extended Haar functions and improved Boosting Algorithm," *IEEE International Conference on Green Computing and Internet of Things (ICGCIoT),* pp. 981-985, 2015.

[105] Shi Jianbo and Carlo Tomasi, "Good features to track." *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pp. 593-600, 1994.

[106] Adrian Wong Yoong Wai, Shahirina Mohd Tahir and Yoong Choon Chang, "GPU Acceleration of Real Time Viola-Jones Face Detection," *IEEE International Conference on Control System, Computing and Engineering*, pp. 27-29, 2015.

[107] Ijaz Khan, Hadi Abdullah and Mohd Shamian Bin Zainal, "Efficient Eyes and Mouth Detection Algorithm using Combination of Viola Jones and Skin Color Pixel Detection," *International Journal of Engineering and Applied Sciences*, vol. 3, no. 4, June-2013.

[108] Wei Liuliu and Liu Mingyang, "Multi-pose Face Detection Research based on Adaboost," *IEEE International Conference on Measuring Technology and Mechatronics Automation*, pp. 409-412, 2015.

[109] Wang Yi-Qing. "An Analysis of the Viola-Jones face detection algorithm," *Image Processing On Line,* vol. 4, pp. 128-148, 2014.

[110] R.Sureshkumar and N.Arthi, "Generate Attribute-Enhanced Sparse Codewords To Retrieve Image From Large Image Database," *International Journal of Engineering Science Invention*, vol. 2, no. 1, pp. 2319–6726, October 2013.

[111] Seyed Mohammad Hassan Anvar, Wei-Yun Yau and Eam Khwang Teoh, "Multiview Face Detection and Registration Requiring Minimal Manual Intervention," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 10, October-2013.

[112] Raphael Sznitman and Bruno Jedynak, "Active Testing for Face Detection and Localization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, October-2010.

[113] Hongliang Li, King N. Ngan and Qiang Liu, "FaceSeg: Automatic Face Segmentation for Real-Time Video," *IEEE Transactions on Multimedia*, vol. 11, no. 1, pp. 77 – 88, 2009.

[114] Paul Viola and Michael Jones, "Robust Real-time Object Detection," *Second International Workshop on Statistical and Computational Theories of Vision Modeling, Learning, Computing and Sampling*, July-2001.

[115] Hatice Gunes and Massimo Piccardi, "Automatic Temporal Segment Detection and Affect Recognition From Face and Body Display," *IEEE Transactions on Systems, Man and Cybernetics Part B: Cybernetics*, vol. 39, no. 1, February 2009.

[116] Mauricio Pamplona Segundo, Luciano Silva, Olga Regina Pereira Bellon and Chaua C. Queirolo, "Automatic Face Segmentation and Facial Landmark Detection in Range Images," *IEEE Transactions on Systems, Man and Cybernetics Part B: Cybernetics*, vol. 40, no. 5, October-2010.

[117] Jie Pan, Xue-Song Wang and Yu-Hu Cheng, "Single-sample Face Recognition Based on LPP Feature Transfer," *IEEE Access*, vol. 4, pp. 2873 – 2884, 2016.

[118] M. Ko and A. Barkana, "A new solution to one sample problem in face recognition using FLDA," *Applied Mathematical Computations*, vol. 217, no. 24, pp. 10368-10376, Aug. 2011.

[119] J. Wu and Z. H. Zhou, "Face recognition with one training image per person," *Pattern Recognition Letters*, vol. 23, no. 14, pp. 1711-1719. Dec. 2002.

[120] Huang, Yea-Shuan and Suen-Yu Chen, "A geometrical-model-based face recognition," *IEEE*

International Conference on Image Processing (ICIP), pp. 3106-3110, 2015.

[121]   Wiskott L, Fellous, J.M, Kuiger N, and von der Malsburg C, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779, July 1997.

[122]   Ngoc-Son Vu and Alice Caplier, "Face Recognition with Patterns of Oriented Edge Magnitudes," *European Conference on Computer Vision*, pp. 313–326, Springer Berlin Heidelberg, 2010.

[123]   Chih-Rung Chen, Wei-Su Wong and Ching-Te Chiu, "A 0.64mm RealTime Cascade Face Detection Design Based on Reduced Two-Field Extraction," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 19, no. 11, November 2011.

[124]   Mandeep Kaur, Rajeev Vashisht and Nirvair Neeru, "Recognition of Facial Expressions with Principal Component Analysis and Singular Value Decomposition," *International Journal of Computer Applications*, vol. 9, no. 12, pp. 36–40, November 2010.

[125]   Muwei Jian and Kin-Man Lam, "Simultaneous Hallucination and Recognition of Low-Resolution Faces Based on Singular Value Decomposition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 11, pp. 1761-1772, 2015.

[126]   Kyungjoong Jeong, Jaesik Choi and Gil-Jin Jang, "Semi-Local Structure Patterns for Robust Face Detection," *IEEE Signal Processing Letters*, vol. 22, no. 9, September 2015.

[127]   Felix Juefei-Xu, Dipan K. Pal, Karanhaar Singh and Marios Savvides, "A Preliminary Investigation on the Sensitivity of COTS Face Recognition Systems to Forensic Analyst-style Face Processing for Occlusions," *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 25-33, 2015.

[128]   Dhara Marvadi, Maulin Joshi, Chirag Paunwala and Aarohi Vora, "Comparative Analysis of 3D Face Recognition Using 2D-PCA and 2D-LDA Approaches," *IEEE 5th Nirma University International Conference on Engineering (NUiCONE)*, pp. 1-5, 2015.

[129]   Ali Moeini and Hossein Moeini, "Real-World and Rapid Face Recognition towards Pose and Expression Variations via Feature Library Matrix," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 5, pp. 969-984, May 2015.

[130]   Christina Joy, Roshlin Anie Abraham and Raji, "A Survey on Face Matching and Retrieval of Images," *International Journal of Computer Science and Mobile Computing*, vol. 4, no. 2, pp. 33–37, February 2015.

[131]   Himanshu Sharma, Sumeet Saurav, Sanjay Singh, Anil K Saini and Ravi Saini, "Analyzing Impact of Image Scaling Algorithms on Viola-Jones Face Detection Framework," *IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1715-1718, 2015.

[132]   H. S. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Transaction on Acoustic, Speech, Signal Processing*, vol. 26, no. 6, pp. 508–517, 1978.

[133]   J. A. Parker, R. V. Kenyon and D.E. Troxel, "Comparison of interpolation methods for image resampling," *IEEE Transaction on Medical Imaging*, vol. 2, no. 1, pp. 31 – 39, 1983.

[134]   T. M. Lehamann, C. Gonner and K. Spitzer, "Survey: Interpolation Methods in Medical Image Processing," *IEEE Transaction on Medical Imaging*, vol. 18, no. 11, pp. 1049- 1075, 1999.

[135]   P. N. Gour, S. Narumanchi, S. Saurav and S. Singh, "Hardware accelerator for real- time image resizing," *IEEE 18th International Symposium on VLSI Design and Test*, pp. 1-6, 2014.

[136]   C. Lin, M. Sheu, H. Chiang, W. Tsai and Z. Wu, "Real-Time FPGA architecture of extended linear convolution for digital image scaling," *IEEE International Conference on Field- Programmable Technology*, pp. 381–384, 2008.

[137]   Zhang, Yi-Qing and Xiang Li, "Temporal dynamics and impact of event interactions in cyber-social populations," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 23, no. 1, 2013.

[138]   Kefalas, Pavlos, Panagiotis Symeonidis and Yannis Manolopoulos, "New perspectives for recommendations in location-based social networks: Time, privacy and explainability," *Fifth ACM International Conference on Management of Emergent Digital EcoSystems*, pp. 1-8, 2013.

[139]   Song, Yang, Zheng Hu, Xiaoming Leng, Hui Tian, Kun Yang and Xin Ke, "Friendship influence on mobile behavior of location based social network users," *Journal of Communications and Networks*, vol. 17, no. 2, pp. 126-132, 2015.

[140]   Berjani, Betim and Thorsten Strufe, "A recommendation system for spots in location-based online social networks," *ACM 4th Workshop on Social Network Systems*, p. 4, 2011.

[141]   Cranshaw, Justin, Eran Toch, Jason Hong, Aniket Kittur and Norman Sadeh, "Bridging the gap between physical location and online social networks," *ACM 12th International Conference on Ubiquitous Computing*, pp. 119-128, 2010.

[142]   Cho, Eunjoon, Seth A. Myers and Jure Leskovec, "Friendship and mobility: user movement in location-based social networks," *ACM 17th SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1082-1090, 2011.

[143]   Silva, Nitai B., Ren Tsang, George DC Cavalcanti and Jyh Tsang, "A graph-based friend recommendation system using genetic algorithm," *IEEE Congress on Evolutionary Computation*, pp. 1-7, 2010.

[144]   Narayanam, Ramasuri and Yadati Narahari. "A shapley value-based approach to discover influential nodes in social networks," *IEEE Transactions on Automation Science and Engineering*, vol. 8, no. 1, pp. 130-147, 2011.

[145]   Debnath, Souvik, Niloy Ganguly and Pabitra Mitra, "Feature weighting in content based recommendation system using social network analysis," *17th ACM International Conference on World Wide Web*, pp. 1041-1042, 2008.

[146]   Wan, Shengxian, Yanyan Lan, Jiafeng Guo, Chaosheng Fan and Xueqi Cheng, "Informational friend recommendation in social media," *36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1045-1048, 2013.

[147]   Gao Yang, Yan Chen and KJ Ray Liu, "Understanding sequential user behavior in social computing: To answer or to vote?," *IEEE Transactions on Network Science and Engineering*, vol. 2, no. 3, pp. 112-126, 2015.

[148]   Yu Chung-Kai, Mihaela van der Schaar and Ali H. Sayed, "Information-Sharing Over Adaptive Networks With Self-Interested Agents," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 1, pp. 2-19, 2015.

[149]   Yunlong Wang and Petar M. Djuric, "Social Learning With Bayesian Agents and Random Decision Making," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, 2015.

[150]   Mahmud, Jalal, Michelle X. Zhou, Nimrod Megiddo, Jeffrey Nichols, and Clemens Drews, "Recommending targeted strangers from whom to solicit information on social media," *ACM International Conference on Intelligent User Interfaces*, pp. 37-48, 2013.

[151]   Zhang, Lizi, Hui Fang, Wee Keong Ng and Jie Zhang, "IntRank: Interaction ranking-based trustworthy friend recommendation," *IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*, pp. 266-273, 2011.

[152]   Lan Zhang, Xiang-Yang Li, Kebin Liu, Taeho Jung and Yunhao Liu, "Message in a Sealed Bottle: Privacy PreservingFriending in Mobile Social Networks," *IEEE Transactions on Mobile Computing*, vol. 14, issue 9, 2013.

[153]   Sharad Goel and Daniel G. Goldstein, "Predicting Individual Behavior with Social Networks," *Marketing Science*, vol. 33, no. 1, pp. 82–93, 2014.

[154]   Benevenuto, Fabrício, Tiago Rodrigues, Meeyoung Cha and Virgílio Almeida, "Characterizing user behavior in online social networks," *9th ACM SIGCOMM Conference on Internet Measurement*, pp. 49-62, 2009.

[155]   Papagelis, Manos, Vanessa Murdock and Roelof van Zwol, "Individual behavior and social influence in online social systems," *22nd ACM Conference on Hypertext and Hypermedia*, pp. 241-250, 2011.

[156]   Jibing Gong, Jie Tang and A. C. M. Fong, "ACTPred: Activity Prediction in Mobile Social Networks," *Tsinghua Science and Technology*, vol. 19, no. 3, pp. 265-274, 2014.

[157]   Sofia Angeletou, Matthew Rowe and Harith Alani, "Modelling and analysis of User behaviourin online community," *International Semantic Web Conference (ISWC)*, pp. 35-50, Springer Berlin Heidelberg, 2011.

[158]   Ahmed Youssef and Ahmed Emam, "Network Intrusion Detection Using Data Mining and Network Behavior Analysis," *International Journal of Computer Science and Information Technology (IJCSIT)*, vol. 3, no. 6, 2011.

[159]   Fortunato Bianconi, Valerio Brunori, Paolo Valigi, Francesco La Rosa and FabrizioStracci, "Information Technology as Tools for Cancer Registry and Regional Cancer Network Integration," *IEEE Transactions on Systems, Man and Cybernetics—Part A: Systems and Humans*, vol. 42, no. 6, 2012.

[160]   Paek, Hye-Jin and Thomas Hove, "Determinants of vertical and horizontal online health information behavior," *IEEE 47th Hawaii International Conference on System Sciences*, pp. 2597-2606, 2014.

[161]   Gokcen, Yasemin, Vahid Aghaei Foroushani and A. Nur Zincir Heywood, "Can we identify NAT behavior by analyzing Traffic Flows?," *IEEE Security and Privacy Workshops (SPW)*, pp. 132-139, 2014.

[162]   Liben-Nowell, David and Jon Kleinberg, "The link-prediction problem for social networks," *Journal of American Society for Information Science and Technology*, vol. 58, no. 7, pp. 1019-1031, 2007.

[163]   Ratnapala I P, R G Ragel and S Deegalla, "Students behavioural analysis in an online learning environment using data mining," *IEEE 7th International Conference on Information and Automation for Sustainability*, pp. 1-7, 2014.