Predicting user behavior over social networks

Sk Ayesha Begum, Dr U Mohan Srinivas, M Ravi Teja, P Pratap Reddy, S Dharma Reddy, K Ashok

1 Asst. Professor, 2 Associate Professor, 3, 4, 5, 6 MCA Scholars, Department of MCA
QIS College of Engineering and Technology (Autonomous), Ongole
Corresponding Author: tejumoparthi@gmail.com

Abstract:

Online social networks have been more popular in recent years, especially Facebook, Twitter, and LinkedIn. There is a lot of interest right now in figuring out how people interact with one another on social media. As far as I know, existing link prediction algorithms forecast connections based on topological structure characteristics as well as on node attribute characteristics. This leaves out important aspects such as click-through rates and forwarding potential for a social network as a whole. A link prediction technique based on user behaviours such as clicks, shares, likes, forwarding and comments is proposed to fill this void. We provide a link prediction model that combines metrics for user activity with topological structure metrics in this article. New metrics for link prediction may be defined using this suggested measure, which can be used to bridge the gap between current techniques. While this study is still a work in progress, next directions include investigating how the suggested metrics might be implemented on common datasets by properly training the classifiers.

1. Introduction

The number of people using online social networks such as Google plus, LinkedIn, Facebook, and Twitter is increasing every year[1]. As a result of this large and varied user base, academics from both academia and business are increasingly interested in studying and understanding how these networks may be used for economic and social good[2]. In link prediction, the goal is to foresee future connections between the users. A few examples of link recommendations include "people you may know" on Facebook and LinkedIn, as well as local businesses. People who use these social networks communicate with one another by posting, commenting, liking, and forwarded. As a result, the owners of these social networks make a lot of money from showing advertisements on the candidate nodes (profiles), which are based on the preferences or choices of the users. As a result, it's reasonable to assume that link suggestion or link prediction helps both users and social networks alike. Other approaches to predicting connections take into account user characteristics and topological structures, with the goal of predicting the links that are most likely to be linked. The advantage of evaluating other current attribute data such as clicks, shares, likes and comments, etc., provided by these social networks is overlooked by them when predicting connections. If we want to suggest or forecast connections more correctly, we must look at current characteristics such as topological aspects of online social networks and define new metrics separate from them. As a result, social networks, such as follower networks,
followee networks, citation networks, wikipedia networks, twitter, and recommendation social networks, produce enormous amounts of data dynamically and quickly. These many kinds of networks need the classification of their users, consumers into various categories for analysis, the discovery of popular users for product sales data in order to improve network performance, design, and so on. This classification of users into distinct groups is based on numerous clustering criteria, decision rules, concealed flow and ultimately it utilises certain broad patterns of user behaviour, network access. These tendencies may be influenced by social network platforms, flexibility, and feature richness, all of which have an impact on how people use social networks in general. The publically accessible social network data may be used to compute some of these behaviours directly, such as persistent behaviour in our conception, while other behaviours may not be disclosed outright by the data itself. Non-persistent social network behaviour is what this is. Concealed information like web server access logs, hidden work flow, and so on may be used to infer nonpersistent behaviour.

2. Related Work

The realism of the anticipated actions would be significantly improved by taking into account the complicated user involvement behaviours places where people go to connect on social media. There has been a lot of development in this field of study. In this part, we'll dissect and explain the user involvement behaviour prediction based on a single message old-fashioned models like the ones that use neural networks in the last several years Prediction of user involvement is common in contemporary approaches behaviour considers the topology of the user's network and without considering the effect of messaging on the user's experience promoted under the guise of pressing issues. Sheikhahmadi and coworkers proposes a two-tiered approach for detecting and categorizing examining the interactions between users, we may learn about the psychological impact of users. This is similar to the work done by Colombo et al. Studying the transmission of information using a map social media website According to Salehi et al. [20], they discovered that the likelihood of using a multilayer network structure for forecasting sending a microblog on to someone else. Researchers [21, 22] disagreed, stating utilising a user's activity to forward to other people methods for learning from machines. Predictions made by Grabowicz et al. [23] included that filtering the variables that are relevant to the user's forwarding behavior greatly affected by user conduct. The vast majority of research to date predicts nonlinear connections subject data input and user interaction are separated by a topic gap results from conventional machine-learning techniques. User forwarding behaviour was predicted by Lee et al. [24], who found that the amount of time it takes for various machine-learning algorithms to transmit information. According to Sankaram et al. [25], an impact model was developed based on With a machine-learning system, they forecast clients as well as admirers. Huang et al. [26] looked at how much a user cared about a topic. Bayesian model applied to various types of tweets and predicted interest metrics-driven forwarding behaviour. Other In experiments, people have acted as though they were users in messages by utilizing Models of infectious diseases that don't depict the illness correctly the fact that things aren't always as they seem. An investigation of the event
was conducted by Huang and Su [27]. It anticipated the user's advancing based on the likelihood of their actions SIR (susceptible, infected, and recovered) model behavior infectious illness evolution. According to Xiong et al. [28], a new susceptibility has emerged. SCIR model (contact, infection, and resiliency) is a code that indicates "Contacted" and forecasts the user's behavior In-depth analysis of browsing and forwarding behaviours. Recent studies have predicted user involvement behaviour. the use of artificial neural networks. Despite the fact that the neural network fits the topic data input is linked in a nonlinear way a large portion of the neural output is determined by user involvement behaviour, Single messages are the focus of network research. After giving it some thought, the mechanism that makes multimessage exchanges possible, the The model's accuracy and complexity would rise as a result. Yang [29], for example, predicted internet download patterns. of people who are utilising neural networks that are created artificially. Li and colleagues [28] genetically enhanced the conventional BP neural network the algorithm and identified the frequent changes in the social environment network. Liu et al. [29] optimised a radial basis in a similar manner. Cloud-based neural network using a function-based (RBF) design It anticipated the user's involvement in fuzzy mathematics Message-specific behaviours.

3. Analyzing User Behavior

Computational social networks' study of user behaviour is crucial for a variety of reasons. Behavioral analysis in social networks includes things like spotting suspicious activity and spotting anomalies, as well as studying the sentiment of large social groups, estimating how popular a product will be, and rating it. These are user behaviour statistics that are gleaned from a social networking site. Figure 1 depicts our suggested behavioural analysis architecture. Locusts, of course The Survey's Standy As compared to other polls, we have a few unique features. Like [25], we examine the origins, sources, and features of social network user behaviour, as well as behaviour analysis in great depth. [26, 27]. Furthermore, rather than limiting ourselves to statistical techniques, we incorporate a wide collection of current social network user behaviour analysis methodologies under the categories of statistical, classification-based, knowledge-based, soft computing, and clustering-based (e.g.). In addition, we address a number of significant outstanding problems and difficulties. In addition to a categorization of social network platforms and their features, we, like [7] and [5,] try to offer a classification of different social network user behaviour analysis techniques, systems, and software tools released to date. We also compare and contrast the various approaches. We've also given a list of open research challenges, like [6]. For the first time, unlike [8,] our study goes beyond only looking at how people use social media to find and share content. It incorporates a wide range of cutting-edge techniques and methodologies for data processing. Our study includes a comprehensive discussion on social network dataset types, collection, and preparation. Unlike [19], however, we also provide a list of actual research problems and open tasks, as well as suggestions for analysing user behaviour. When compared to other studies, ours focuses on the behaviours of social network users, including how to characterise, recognise and identify them.

www.turkjphysiotherrehabil.org
4. The problem of User Behavior Analysis

The behaviour is a qualitative word, and expressing it mathematically is a difficult task. This, to our knowledge, may be described as a user's habitual participation in social network activities. To put it another way, this is the most common way that people utilise social networks. Actual communication between social network users includes interaction patterns that lead to the most impactful pattern for each member throughout the whole network in real life. It's critical to identify the network's influential patterns. These patterns show how the social network as a whole behaves. This is critical for social networks that include the interaction patterns of many people. As a result, we're on the lookout for the social network communities’ most important interaction structures. Given a graph $G$ with vertices $V$ and edges $E$, find the shortest path across the graph. We need to find the most common patterns of social network user interaction $P$ that are a part of $G$. Users' behaviour in a social network is categorised by the distinct patterns they see in the network. After identifying the common patterns of user interaction in the network, we must link these patterns to the social network platform in order to do a qualitative analysis on the user behaviour on that platform.

5. Our Contributions

An comprehensive review of studies on social network user behaviour analysis is provided in this article. The following are the survey's most significant contributions: The connection between user behaviour in online social networks has yet to be addressed properly. We made an effort to link our social network user behaviour research to real-world situations. (b) User behaviour analysis has been divided into two main categories: persistent and non-persistent views. This is a fresh approach to studying social network users' online behaviours. Many studies do not include all the methods to social network user behaviour analysis that have been developed up to this point, but we have examined a broad variety of methodologies that have been published up until now. It is important to characterise and reflect the behaviour of social network users in surveys, yet most of the current surveys do not do this. We describe and

www.turkjphysiotherrehabil.org
evaluate a number of methods for analysing accessible datasets to determine how they define behaviour. While discussing behavioural analysis methods, we present a number of different ways to characterise online social network user behaviour and compare them.

6. Characterization of Behavior

Characterizing social network user behaviour involves defining behaviour quantitatively and appropriately. This mathematical description of behaviour will assist in the identification of the most suitable pattern of behaviour from vast amounts of social network data. Because human behaviour differs significantly from that of social network users, this is a very difficult picture to create. The same user's activity patterns on multiple social network platforms vary from user to user, which affects the overall user's behaviour. Characterizing behaviour in a computable manner to model generic behaviour also presents challenges in terms of processing complexity, precise data representation for model input, and so on. In the past, people have attempted to categorise user behaviour by looking at the kind of activities they engage in, patterns they follow, and how they participate in social networks, among other things. The characterisation of user behaviour on different social network platforms used a variety of techniques. [21] Maria Kihl et al. 2010 examined traffic data from a Swedish municipal broadband access network and developed user behaviour models as a result of their research, "Traffic analysis and characterisation of Internet user Behaviour" [21]. Internet traffic patterns, volumes, and application use are all discussed in detail in this study. Session durations and traffic rate distributions, two user activity variables, are also examined and modelled. Modeling Users Activity on Twitter Networks: Validation of Dunbars Number[10] presented by Bruno Goncalves et al. in 2011 included a basic model of users' behaviour with limited priority queueing and time resources that reproduced the observed social behaviour. These researchers looked over 1.7 million Twitter interactions over six months to see how many solid social connections a person can have before hitting the theoretical cognitive limit. In their 2011 article, "Primary role identification in dynamic social networks," Radoslaw Brendel and Henryk Krawczyk presented a behavioural model based on user roles in social networks[4]. Social networks often take on very complicated structures, which evolve over time. Roles for players in these systems must be described by taking into consideration the behavioural features of these actors' dynamics. A role may be described as a set of interrelated tasks that a person does. Pattern sub-graphs represent many actions, whereas sequence diagrams describe the succession of these activities. It is suggested to give each player a main role performed in a dynamic social network for such clearly defined roles. To better understand how one user behaves across several social media sites, Erheng Zhong et al. conducted research in 2012. Many prior studies used just one historical user record to predict user behaviour [4]. According to their findings, many individuals are active users of several social networks at the same time. It's important to note that their actions and preferences in various networks have an impact on one another. In order to relieve the data sparsity issue and improve the prediction performance of user modelling, this creates an opportunity to use knowledge about user behaviour in various networks. They framed the issue
as a problem of composite network knowledge transfer, which models several networks. Using a hierarchical Bayesian model parameterized for individual users, they first identified the most appropriate networks within a composite social network. After that, they create topic models for predicting user behaviour based on the connections in the chosen networks and the corresponding data on behaviour. "Characterizing User Behaviour and Information Propagation on a Social Multimedia Network"[9] was published in 2013 by Francis T. Odonovan et. al. and investigated the online behaviours of users. The sharing of multimedia is one such activity, the popularity of which fluctuates greatly. They spoke about the preliminary results of scraping consenting Facebook users' demographic and psychological data and analysing it anonymously. They showed five groups of individuals with similar online habits and profile features that linked with one another. In 2013, Ryan A. Rossi released a paper titled "Modeling Dynamic Behaviour in Large Evolving Graphs"[17] on his study on dynamic behaviour in social network graphs. A huge time-evolving network was used to describe and characterise the temporal behaviours of individual nodes. Furthermore, they've looked at how nodes' behavioural transition patterns may be modelled. Nodes in the graph have responsibilities, and those roles change over time, as shown by their temporal behaviour model. Generalizable and computationally efficient behavioural roles are defined by the model. It uses temporal behaviour to discover patterns and trends in nodes and network states, forecast structural changes in the future, and detect transitions in temporal behaviour that aren't expected. In 2013, Vasanthan Raghavan et al. released a study titled "Modeling Temporal Activity Patterns in Dynamic Social Networks"[24] in which they created probabilistic models of user activity in social networks that included the perceived social network impact of the users. An study of microblog user behaviour based on social networks was published in 2013 by Qiang Yan et al. The impact of microblogs on the dissemination of information is becoming more apparent. When following and being followed are characterised as out-degree and in-degree behaviours, a microblog social network may be constructed as a result. Short linked network diameters, short average route lengths, and a high average clustering coefficient were discovered. There are power-law characteristics in the distribution of out-degree, in-degree, and total microblogs posted. There is a negative correlation between the exponent of total microblog numbers and the degree of each user. The exponent drops at a considerably slower rate as the degree rises. A human dynamics model built on social networks was presented in this article, based on empirical investigation, and it shown that induced drive and spontaneous drive led to posting microblogs. A user behaviour model was examined by Zhenhua Wang et al. in 2014 in their article "Analysis of user behaviours by mining massive network data sets"[7]. Understanding the intelligence of human behaviour by mining petabytes of network data reflects the trend in social behaviour research and demonstrates significant importance on the creation of Internet applications and service growth on the Internet. Meanwhile, mobile networks in use today, which generate enormous amounts of data, can serve as the best social sensors for future research. By processing large data intelligently, this article examines a real-world application of mobile network-assisted social sensing that reveals certain characteristics of mobile network users' behaviour. On the basis of huge user data sets, the study
examines users' communication, mobility, and consuming behaviours. Characterizing Group-Level User Behaviour in Major Online Social Networks by Reza Motamedi et. al. was published in 2014[15]. An in-depth measuring research was carried out to describe and compare the behaviour of Facebook, Twitter, and Google+ users on a group-level. User connection, activity, and responses were the primary indicators on which they honed down on user behaviour. Additionally, they carried out a two-year temporal study on various elements of user behaviour for all demographics. It turns out that Facebook and Google+ users express themselves rapidly, whereas Twitter users prefer to pass on a received post to other people, making it easier for the spread of a message. A popular Facebook user's post gets more re-shares than a popular Twitter user's post, despite Twitter's culture of re-sharing. An OSN's added functionality may significantly increase the pace of action and response among its users, as seen in (ii).

7. Behavior Recognition

It is possible to deduce underlying user behaviour from a variety of observations of social network users' data, activities, and web usage content. Many people have attempted to learn about the online behaviour of others by using various social media platforms. Human behaviour was described by Marcelo Maia et al. in their 2009 research paper, "Identifying User Behaviour in Online Social Networks"[20]. They clarified how to classify various user behaviours. User behaviour characterization methods based on individual user characteristics have not been used for online networking sites in the past because of this. Users can upload and view content, choose friends, rank favourite content, subscribe to users, and engage in a variety of other interactions through a series of multiple interfaces in these environments. Different user groups have different patterns of interaction, which can be observed. To characterise and identify user behaviours in online social networks, the authors proposed a methodology in this paper (PDF). In order to group users who have similar patterns of behaviour, they used a clustering algorithm. They've also demonstrated that characteristics derived from social interaction are useful for making distinctions and spotting relevant user behaviours. K. R. Suneetha R. Krishnamoorthi used data mining techniques on web server log data to discover web access patterns in their 2009 research paper, "Identifying User Behaviour by Analyzing Web Server Access Log File"[1]. User personalization is possible by remembering previously visited pages, they said. These pages can be used to determine a user's typical browsing habits and then predict what pages they want to see. Important links can be found to improve the overall performance of future accesses by analysing user access behaviour. Traditional data mining parameters like clustering and classification, association, and examination are used to evaluate the information obtained through Web mining in order to discover a sequential pattern of user activity. In 2012, Sofia Angeletou, Matthew Rowe, and Harith Alani published an article titled "Modeling and Analysis of User Behaviour in Online Communities"[23]. They studied online user behaviour in online communities. A semantic model and rules for representing and computing behaviour in online communities were combined with statistical analysis by the researchers in this study. A number of forum communities from Boards were used to test out this model. In order to classify
community members' behaviour over time and report on how different behaviour compositions correlate with positive and negative community growth in these forums. "Social Media Data Analysis for Revealing Collective Behaviours" was published in 2012 by Aoying Zhou and Weinig Qian Haixin Ma. They found that, given enough social media data, the collective behaviour of users could be sensed, studied, and even predicted under certain conditions. They experimented with data from Twitter and Sina Weibo, two popular social media platforms. In collective behaviour, people act in ways that are neither predictable nor predictable. Social media is used to study a wide range of collective behaviours. Researchers have found a variety of information flow patterns in social media, some of which are similar to traditional media like newspapers, while others are deeply ingrained in the social network structure themselves. External stimuli, social network structure, and individual users' activities all have a significant impact on hotspot evolution. Social media, on the other hand, is resistant to repeated exposure to the same external stimuli.

8. Prediction of Behavior

So far, we've talked about how to characterise and recognise user behaviour on social media platforms. There have been numerous methods and algorithms reported for the tasks listed above. After that, we'll talk about recent developments in the field of behaviour prediction research. Future behaviour can be predicted as outlined below by examining various social network data sets and looking back at previous activities. According to M. Vasudevan and M. Tamilarasi's research paper "Collective behaviour prediction in social media: A survey"[19], they conducted a survey on social network users' collective behaviour in 2012. Individuals' actions in a social network environment are referred to as their collective behaviour. This collective behaviour allows for the prediction of users' online behaviour if the behaviour of some actors in the network is known. It investigates how social media networks can be used to make predictions about human behaviour and personal preference, for example. For tasks like social networking advertising and recommendation, this can help decipher the patterns of behaviour displayed in social media. Adam Sadilek studied "Human behaviour at large scale in various social networks" for his doctoral thesis [1]. The unification and data mining of diverse, noisy and incomplete sensory data over a large number of individuals is the central theme of this thesis. Aside from sensory data, they discovered that user online communication and interpersonal relationships are rich information sources on which strong machine learning models can be built, as are explicit and implicit online social interactions. As an example, understanding human activities and predicting people's location and social ties from their online behaviour are examples where such models apply. Another example is predicting global epidemics from day-to-day interpersonal interactions will emerge. In 2012, Jihang Ye et al. used the check-in category information to model the underlying user movement pattern in their research work "Whats Your Next Move: User Activity Prediction in Location-based Social Networks"[12]. According to their framework, they can predict user activity categories using a mixed hidden Markov model, and then use the model's estimated distribution to guess where users are most likely to be located in the following
step. Modeling at the category level has several advantages, including a significantly smaller prediction space and a more precise expression of the semantic meaning of user activities. In their study, "Manifestations of user personality in website choice and behaviour on online social networks," Michal Kosinski et al. looked at how people behave in social networks. Using data from one million users, this study examines the relationship between online user behaviour and personality as seen in their website choices and Facebook profile features [24]. Results show that users' personalities, website preferences, and Facebook profile features have psychologically significant connections. Predicting an individual's personality profile can be used to improve online advertising, personalise content, and improve search engine results. This idea can be utilised to make predictions about a user's social network behaviour. Prediction of individual behaviour was studied in 2013 by Sharad Goel and Daniel G. Goldstein (co-authors). When it came to online department store shopping, they used a communications network with over 100 million users to forecast a wide range of behaviours.

9. Prediction and Analysis of Results

The results of the BP neural network prediction model optimised by the various simulations of an annealing algorithm baseline methods, namely the traditional neural network other traditional classifiers, such as the model The display used the three evaluations to conduct a more thorough evaluation of metrics (precision, recall, and the F1-measure, all of which were first used in the horizontal as well as vertical coordinates. The number 6 represents the percentage of the Message combinations with multiple messages and evaluation indicators are two examples of this. We can predict user behaviour because it uses the C-RBF model during single messaging, it failed to resolve the overfitting problem during multiple-message interactions, so the prediction Multimessaging lacked precision. The convolution kernel of CNN emphasises the window in space. Due to the applicability of data density, the accuracy of using CNN to predict that the message's reach will be limited. Due to the advantages of LSTM for time-series data processing, it can learn the user's historical behaviour nicely. However, the multimessage attribute is not fully captured. As opposed to that, the SA+BP model works quite well predicted how users would participate in a multimessaging session. To comprehensively evaluate the predictive performance, this article adds a comparative experiment of the receiver operating characteristic (ROC) curves. Fig. 2 is limited to tracking the key performance indicators for only the model unilaterally. In Fig. 3, the false positive rate is represented by the abscissa and ordinate. (FPR) and true positive rate (TPR), respectively.
10. Conclusion

We researched usage patterns in OSNs from three perspectives—behaviour categorization, behaviour recognising & behaviour estimation. Users' different tasks on social network platforms were taken into account, such as interconnections, traffic activities, and the arrangement of users in the network as a whole. We looked at the current representative schemes and made suggestions for where they might go in the future. The two broad research lines that we focused on were persistent and non-persistent social network user behaviours. As many of the published works up to 2016 are included in this report, this survey will aid in understanding the chronological progression of user behaviour analysis research.

References


[21] Maria Kihl1, Christina Lagerstedt, Andreas Aurelius and Per dling 2010, Traffic analysis and characterization of Internet user behavior, In Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Moscow, 18-20 Oct., 224- 231, IEEE.


[29] Nhat Hai Phan, Dejing Dou, Hao Wang , David Kil and Brigitte Piniewski 2015, Ontology-based Deep Learning for Human Behavior Prediction in Health Social Networks