



INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING
2019, ICRTAC 2019

Towards User Profiling From Multiple Online Social Networks

B.GayathriDevi*, V.Pattabiraman

Vellore Institute of Technology, Chennai, Tamilnadu, India

Abstract

Social media networks are exponentially growing and it leads us to challenging issues in independent user assign assumption from the Big Data prospect. The main objective of the user profile is to generate a profile for the users by grouping user intelligence. For effective marketing and advertisement accurate user profiling is necessary in personalized endorsement system. For Instance, Twitter users can address their profile with `narrative tags. The tags deliver a user explanation and used to retrieve smoothen data and another implementation of Yelp. On food, users desire accurate profiling for example, it will notably upgrade presentation on suggesting restaurants to the user. This work, we endorse a hypothesis for providing a profile user perception in online social medias. The mechanism may vary in profiling users by depending on its purpose and applications. The study of dissimilar implementations used in user profiling is under the extent of fortune aspires. The proposed model is used to attain the advanced activity to endorse modules in groups of members.

© 2019 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the scientific committee of the INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019.

Keywords: Social network; user profiling; clustering; behavior analysis.

1. Introduction

Social media technologies and tools authorize the users to interface with other people. Social networks contain photo sharing, video, wikis and blogs. In human life communication extremely plays a key role for interaction and our society development. By using internet, it accompanies people together with same attraction and also assist them comfortably to link with other people. When comparing with traditional news sources, Online media have the potential to administer the information as fast as they can. The interaction is in the form of content sharing or messages.

* Corresponding author. Tel. 9894094644

E-mail address: gayathrimtechse@gmail.com

In today's world social media sites are being commercial as people can be in touch with each other although there is no face to face interaction or contact. They can debate about the topics, getting answers and asking queries.

1.1 Characteristics of Social Media Data

(a) Linked Data

Data objects are autonomous and uniformly distributed systems (i.i.d), and it is extensively used supposition in profiling and traditional mining techniques. The supposition does not contain social media data where data objects are potentially networked between user to user interconnections. The connection through data objects may carry textual functional data which is inaccessible in i,i,d. data. For instance, in certain cellphones the user's favor may be near to or guided by their companions. The logic fundamental hypothesis is described by two speculations. Homophily – people favor to befriend with who are alike to themselves, and they favor to be closer to their companion by time (Social Influence). The survey of social relationship in social network promotes user design conduct for various applications. Profiling techniques and existing mining concentrates on establishing a knowledgeable attribute area or successful knowledge design for information analysis. Applying the process directly in social network applications is challenging to manage the interconnected data. To traverse social connections some efforts to be made for non-identical tasks in social network. Jiliang Tang et.al [1] explores whether interconnected information can be used in a current attribute framework selection.

Based on the connections among indicators from social media they studied four types of connections, i.e., Following, CoPost, CoFollowed and CoFollowing. Interconnected data also involve exploration of location-based social media study, spammer detection, text mining, recommendation and active learning.

(b) Unstructured Data

Unstructured data forms are the key variation through social media information and information obtained from different manifest. This attribute can be explained in two methods. In First method, social media information can be portrayed as attribute of different data resources. For Instance, Twitter represents word aspects from the tweets users addressed, by Scale Invariant Feature Transform (SIFT) parametric attribute obtained from images transfer, and by friendship attribute attained from his social media. Heterogeneous space integrates attributes in a unified learning framework is an open issue for many applications. Second, the data generated by user in social media are no-formal. For Instance, tweets like “so cool” and “Happy Bday” are congenital. The un-organized congenital design represents imposing obstacles for an enlightening algorithm to exactly compute the textual sense of the posts.

(c) Dynamic Data

In most of the social media applications user-creates content which progress very rapidly as do social networks. For instance, users may coin new acronyms or abbreviations while writing a microblogging message which are hardly employed in other traditional documents. For example, “spoopy” is a deliberate misspelling of the text “spooky” to narrate a condition that inspire laughter and frighten at the identical time. In social networks Slang words are extremely widespread and retain altering. The membership, network groups and the individual influence of network may alter across time. For instances, machine learning analysis and deep learning communities may develop remarkably in a social network, along with deep learning performance methods.

2. Related Work

Dipon Kumar Ghosh et.al, [2] proposed a decision-making engine for online products based on sentiments of consumer reviews. Gong et al. [3] presented a collaborative distributed design filter deployed on peer-to peer network. By clustering method sizeable number of scalability difficulty has been resolved. Clusters are the group of data classes that are homogeneous to one other in the identical and are different to the classes in other networks. K-medoid methods are applied by Chakraborty [4] for clustering items as well as users. The item clustering prediction was proposed by Huang [5], which uses collaborated item based clustering, while Gong and Ye [6] presents user clustering technology.,

A cooperative filtering algorithm [9] was proposed by Goldberg et al. [7] relates Eigentaste which register PCA which dimensionality facilitates cutback for offline network of user and quick online cluster mission. Using angular similarity found on AS-INDEX, Yu [8] presents a collaborative filtering endorsement design.

In today's culture most of the youngsters are busy in social networking, this brings some technical or political revolution changes in the society through young minds. To join in the social networking sites, most of the network follows certain terms and conditions. After acquiring their standards, user can create their profile and begin accomplish task in the site. The user could be mixed up in various social networks, since there is no restriction or connection for users to add other sites. Connecting different sort of people, sometimes may leads to become insecurity for society [11] – [21].

Lee et al. [22] concentrated on the utilization of social site information and uses tranquil-based process for user profiling. This research, to create user profiles for news analysis recommendation twitter employs users' timelines. Sofia et al. [23] have progressed a technique which discovers the user members in online social networks. The performance characters are explained by using skeleton rule and ontology clustering and the users are identified with those performance measures.

To make automatic prediction Collaborative Filtering is an approach (filtering) about the absorption of a user by grouping heed from many users (collaborating). Gong et al. [24] provides a distributed collaborative filtering structure in the form of peer-to-peer network.

A cooperative filtering algorithm [9] was proposed by Goldberg et al. [7] relates Eigentaste which register PCA which dimensionality facilitates depletion for offline clustering of user and quick online cluster mission. Using angular similarity provided by AS-INDEX, Yu [8] presents a collaborative filtering endorsement design. Acilar et.al. [25] proposes a module with collaborative filtering, which is build stationed on simulated Immune Network design, to locate the scalability and sparsity issues.

Waseem Ahmad et al. [26] proposed the framework that “exploits the user-generated posts to match the user identity across social networks”.

3. System Architecture

Depending upon the applications there are numerous methods by which user can be profiled. Following is the overall function followed in profiling. The real information from online social sites is analyzed and grouped with regards to diverse features like images, movements, geographic location and text data. At this stage, numerous data activities can be executed, the profile photos or images posed, data attached by the user location are separated precisely. With regard to images, the image features and its details are separated using relevant algorithms. Specific information is examined and the unrelated / connected information is categorized by applying classification techniques [10]. Establishing the classification of data, users are grouped into specific class with functionality and class objects. After acquiring the features of the user and their connected class, type, the number and set of projects of each user is extricated. User behavior is acquired by the influence and numerous excessive actions made on others by a user. The user is profiled by examined the set off conduct patterns in social media. From expressing the facial attributes, particular algorithms are validated and applied for face detection. Having dissimilar expressions and pose with the images creates conflicts. By balancing the accessible outcome with database images, user is profiled. In OSNs large number of datasets are available and the information is added or below tiny and scant, for similarity user developed content data are grouped and analyzed. To attain more insight to semantic connection the user data are semantically analyzed, by which user behavior is derived. Figure 1 illustrates the architecture of user profiling.

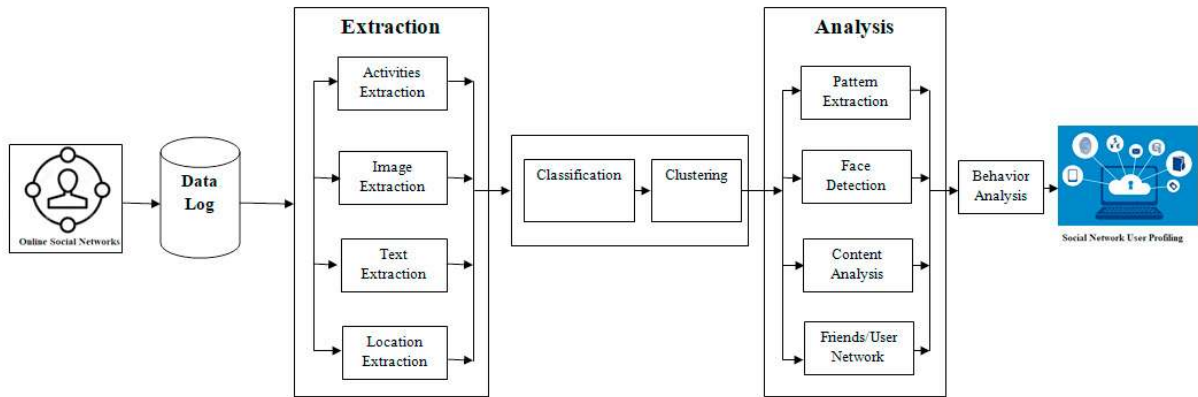


Fig.1. Architecture of User Profiling

4. Algorithms Used

4.1 Clustering Algorithm

The framework uses the clustering algorithm [10]

```

for each user
  Retrieve user loan and reservation data and store in list new items
  for each item in list new items
    Check if item is in existing cluster
    if yes
      Update item's rating in cluster
    if user not assigned to cluster
      Assign user to that cluster
      Number of users assigned to cluster + 1
    End if
    Indicate cluster has been modified
    Decrease all other item's rating in cluster
    Remove item from new items
  End if
End for
for each item in list new_items
  Check if item can be placed in existing clusters
  if yes
    Add item to cluster
    if user not assigned to cluster
      Assign user to that cluster
      Number of users assigned to cluster + 1
    End if
    Indicate cluster has been modified
    Remove item from new items
  End if
  If no
    Create new cluster base on subject description and author and keyword
    associated with item
    Remove item from new items
  End if
End for

```

4.2 Hybrid Method

A hybrid method, can also be called as hybrid filtering, which employs collaborative and tranquil-based techniques. Hybrid method assures the instant accessibility of a contour for each user. The network which uses the hybrid technique imparts the user preferences and interest error free explanation, as it constantly retrieves and monitors user associated information between the user-system interactivity. The hybrid technique generally allocates the latest standard contour user with the use of the method participation and magnifies the profile employing the tranquil-based technique. On the other hand, In dynamic collaborative profiling, information's grouping and gathering of users with homogeneous conduct has done according to the users positive responsiveness.

5. Experimental Result

Nearly twenty-five contenders were requested for this experiment to discover the representation of our work. The users are separated into five dissimilar user groups so that we can use integrated methods. Every group holds five individuals where each one is carrying a frequent task and each group has a common manifest. So that each member in a group performing certain roles. For example, one group is “to build a web portal to host user forums and chat groups”. Each user in the network holding dissimilar characters, for e.g. “Systems Engineer”, and was stated a explanation of their position range. Based on the profile allocated each user can employ the search facility throughout the experiment. And this demonstration was managed in two different categories. In each category the user was questioned to explore by employing the similar search question for each category. Category 1, without utilizing any profiling it searches by using content base method and in Category 2; it employs our advanced profiling method and searches by hybrid methods [10]. At every category, the users were questioned to document the find issues, the sum of information repaid by the search engine and the number of applicable information repaid. Figure 2 shows the relative precision (i.e. “ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. It is usually expressed as a percentage”) with number of users for our proposed method and existing method.

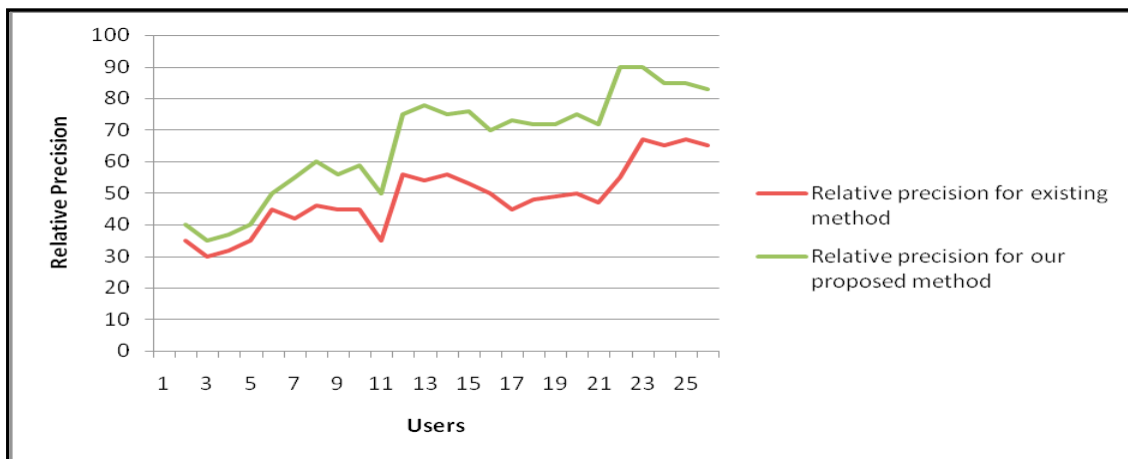


Fig. 2. Precision Graph for individual users

6. Conclusion and Future Enhancement

In this paper, by using user profiling and current concepts that we spotlighted will have the prospective to reduce and overcome the issues experienced by users. Based on the user profiling movements in line the future researchers, with ongoing incidents in society, personalization, advertising, e-marketing, friends' movements, search engine log information's are vital for duplicate discernment and establishing dominant user. For accurate profiling any social media user it is indeed for testing user activities and their data's simultaneously in online multi social networks.

References

- [1] Tang, J., Lou, T., & Kleinberg, J. (2012). “Inferring social ties across heterogeneous networks”. Paper presented at the Proceedings of the fifth ACM international conference on Web search and data mining.
- [2] Dipon Kumar Ghosh , Prithwika Banik , Dr. S. Balakrishnan (2018), “Review-Guppy: A Decision-Making Engine for Ecommerce Products Based on Sentiments of Consumer Reviews”, International Journal of Pure and Applied Mathematics, Volume 119, No. 12, 2018, pp.1135-1141.
- [3] S. Gong, H. Ye and P. Su, A Peer-to-Peer Based Distributed Collaborative Filtering Architecture. In Proc of the International Joint Conference on Artificial Intelligence (IJCAI), 2009, pp. 305–307.
- [4] P.S. Chakraborty, A Scalable Collaborative Filtering Based Recommender System Using Incremental Clustering. In Proc of the IEEE International Advance Computing Conference (IACC), 2009, pp. 1526–1529.
- [5] Y. Huang, An item based collaborative filtering using item clustering prediction. In Proc of the International Colloquium on Computing Communication Control and Management, 2009, pp. 54–56.
- [6] S. Gong and H. Ye, Joining User Clustering and Item Based Collaborative Filtering in Personalized Recommendation Services. In Proc of the International Conference on Industrial and Information Systems (IIS), 2009, pp. 149–151.
- [7] K. Goldberg, T. Roeder, D. Gupta and C. Perkins, Eigentaste: A Constant Time Collaborative Filtering Algorithm, Inform Retrieval 4(2) (2001), 133–151.
- [8] X.P. Yu, As-index based collaborative filtering recommendation algorithm. In Proc of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 2009, vol. 3, pp. 1570–1576.
- [9] Balakrishnan. S and K L Shunmuganathan, An Agent Based Collaborative Spam Filtering Assistance Using JADE”, International Journal of Applied Engineering Research, ISSN 0973-4562, Volume 10, Number 21 (2015) pp 42476-42479.
- [10] Godoy, Daniela & Amandi, Analía. (2003). A User Profiling Architecture for Textual-Based Agents. Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial, ISSN 1137-3601, N° 21, 2003, pages. 27-36.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] Miro-llinares f and j. J. Rodríguez-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.
- [18] Dong, Lijun, Kui Wu and Guoming Tang, “A DataCentric Approach to Quality Estimation of Role Mining Results,” IEEE Transactions on Information Forensics and Security, vol. 11, no. 12, pp. 2678-2692, 2016.
- [18] Nguyen, T.H., Shirai, K. and Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. Expert Systems With Applications, 42 (2015), pp. 9603–9611.
- [19] 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.
- [20] 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 HumanMachine Systems, vol. 45, no. 1, pp. 39-50, 2015.
- [21] 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.
- [22] W. J. Lee, K. J. Oh, C. G. Lim and H. J. Choi (2014). User profile extraction from twitter for personalized news recommendation, International Conference on Advanced Communication Technology, pp. 779-783.
- [23] Sofia Angeletou, Matthew Rowe and Harith Alani, “Modelling and analysis of User behavior in online community,” International Semantic Web Conference (ISWC), pp. 35-50, Springer Berlin Heidelberg, 2011.
- [24] S. Gong, H. Ye and P. Su, A Peer-to-Peer Based Distributed Collaborative Filtering Architecture. In Proc of the international Joint Conference on Artificial Intelligence (IJCAI), 2009, pp. 305–307.
- [25] A.M. Acilar and A. Arslan, A collaborative filtering method based on artificial immune network, Expert Syst Appl 36(4) (2009), 8324–8332.
- [26] Waseem Ahmad, Rashid Ali, International Conference on Pervasive Computing Advances and Applications – PerCAA 2019, Procedia Computer Science 152 (2019) 222–229.