ISSN Online: 2319 – 9253
Print: 2319 – 9245

SCALABLE LEARNING MODELS FOR COLLECTIVE BEHAVIOUR IN SOCIAL MEDIA

Rachamalla Sridhar [1] **Abdul Rasool.MD** [2] **K.AnilKumar** [3]

1. Asst.Professor, Dept. of CSE, Vivekananda Institute of Technology and Science, karimnagar.
2. HOD, Dept. of CSE, Vivekananda Institute of Technology and Science, karimnagar.
3. Asst.Professor, Dept. of CSE, Vivekananda Institute of Technology and Science, karimnagar.

ABSTRACT

Now a days the social networks are playing vital role in world. The social networks has various dimensions the people were using that and gain the content as well as services. By given information about some individuals, how can we infer the behavior of unobserved individuals in the same network. A social-dimension-based approach has been shown effective in addressing the heterogeneity of connections presented in social media. We can predict online behaviors of users in a network, given the behavior information of some actors in the network. Many social media tasks can be connected to the problem of collective behavior prediction. Since connections in a social network represent various kinds of relations, a social-learning framework based on social dimensions is introduced.

The scale of these networks entails scalable learning of models for collective behavior prediction. To address the scalability issue, we propose an edge-centric clustering scheme to extract sparse social dimensions. With sparse social dimensions, the proposed approach can efficiently handle networks of millions of actors while demonstrating a comparable prediction performance to other non-scalable methods. This study of collective behavior is to understand how individuals behave in a social networking environment.

Index Terms: social media, Network Data, Collective Behavior, Community Detection, Social Dimensions

I. INTRODUCTION

Social media is not an end in and of itself. It is an enabler. Social media can enable new mass collaborative behaviors that unlock the power of the collective and deliver new paths to enterprise results. Enterprises can employ these collective behaviors as the link between business value and social media technologies. Enterprises can use them to examine a target community and formulate new ways that people can interact to achieve enterprise value.

Collective intelligence is the meaningful assembly of relatively small and incremental community contributions into a larger and coherent accumulation of knowledge. Collective intelligence is not new, but the mass collaboration enabled by social media provides collective intelligence at scales never before possible. Even the most modest individual contributions can be tremendously valuable when meaningfully combined at a large scale. Wikipedia, YouTube and Flickr are all Social Web examples of collective intelligence. Each Wikipedia article by itself is relatively insignificant, but a million articles collected and linked together are highly powerful.

Collective intelligence is the strongest of the collective behavior patterns, meaning that it is the most prevalent in the 200 social media cases analyzed. It most heavily impacts business results in the area of operational effectiveness. The business use case most benefiting from collective intelligence is product delivery. Blogs and wikis are the most prevalent social media technologies supporting collective intelligence.

ISSN Online: 2319 – 9253
Print: 2319 – 9245

Social networking sites (a recent phenomenon) empower people of different ages and backgrounds with new forms of collaboration, communication, and collective intelligence. Prodigious numbers of online volunteers collaboratively write encyclopedia articles of unprecedented scope and scale; online marketplaces recommend products by investigating user shopping behavior and interactions; and political movements also exploit new forms of engagement and collective action. In the same process, social media provides sample opportunities to study human interactions and collective behavior on an unprecedented scale.

In this work, we study how networks in social media can help predict some human behaviors and individual preferences. In particular, given the behavior of some individuals in a network, how can we infer the behavior of other individuals in the same social network. This study can help better understand behavioral patterns of users in social media for applications like social advertising and recommendation. Typically, the connections in social media networks are not homogeneous. Different connections are associated with distinctive relations. For example, one user might maintain connections simultaneously to his friends, family, college classmates, and colleagues. This relationship information, however, is not always fully available in reality. Mostly, we have access to the connectivity information between users, but we have no idea why they are connected to each other. Heterogeneity of connections limits the effectiveness of a commonly used technique — *collective inference for network classification*. A recent framework based on *social dimensions* is shown to be effective in addressing this heterogeneity.

The framework suggests a novel way of network classification: first, capture the latent affiliations of actors by extracting social dimensions based on network connectivity, and next, apply extant data mining techniques to classification based on the extracted dimensions. In the initial study, modularity maximization [3] was employed to extractsocial dimensions. The superiority of this framework over other representative relational learning methods has been verified with social media data in. The original framework, however, is not scalable to handle networks of colossal sizes because the extracted social dimensions are rather dense.

II. Related Work

As existing approaches to extract social dimensions suffer from scalability, it is imperative to address the scalability issue. Connections in social media are not homogeneous. People can connect to their family, colleagues, college classmates, or buddies met online. Some relations are helpful in determining a targeted behavior while others are not. This relation-type information, however, is often not readily available in social media. A direct application of collective inference or label propagation would treat connections in a social network as if they were homogeneous.

Problems with these systems are

- Social dimension suffer from scalable in heterogeneity.
- This heterogeneity of connections limits the effectiveness.

III. PROPOSED APPROACHES

A recent framework based on *social dimensions* is shown to be effective in addressing this heterogeneity. The framework suggests a novel way of network classification: first, capture the latent affiliations of actors by extracting social dimensions based on network connectivity, and next, apply extant data mining techniques to classification based on the extracted dimensions.

In the initial study, modularity maximization was employed to extract social dimensions. The superiority of this framework over other representative relational learning methods has been verified with social media data in. The original framework, however, is not scalable to handle networks of colossal sizes because the extracted social dimensions are rather dense. In social media, a network of millions of actors is very common. With a huge number of actors, extracted dense social dimensions cannot even be held in memory, causing a serious computational problem.

Sparsifying social dimensions can be effective in eliminating the scalability bottleneck. In this work, we propose an effective *edge-centric* approach to extract *sparse* social dimensions. We prove that with our proposed approach, sparsity of social dimensions is guaranteed. By this approach incomparable advantage of our model is that it easily scales to handle networks with millions of actors while the earlier models fail. This scalable approach offers a viable solution to effective learning of online collective behavior on a large scale.

SOCIAL DIMENSION EXTRACTION

The latent social dimensions are extracted based on network topology to capture the potential affiliations of actors. These extracted social dimensions represent how each actor is involved in diverse affiliations. These social dimensions can be treated as features of actors for subsequent discriminative learning. Since a network is converted into features, typical classifiers such as support vector machine and logistic regression can be employed. Social dimensions extracted according to soft clustering, such as modularity maximization and probabilistic methods, are dense.

DISCRIMINATIVE LEARNING

The discriminative learning procedure will determine which social dimension correlates with the targeted behavior and then assign proper weights. A key observation is that actors of the same affiliation tend to connect with each other. For instance, it is reasonable to expect people of the same department to interact with each other more frequently. A key observation is that actors of the same affiliation tend to connect with each other. For instance, it is reasonable to expect people of the same department to interact with each other more frequently. Hence, to infer actors' latent affiliations, we need to find out a group of people who interact with each other more frequently than at random.

CHART GENERATION FOR GROUP/ MONTH

Two data sets reported in are used to examine our proposed model for collective behavior learning. The first data set is acquired from user interest, the second from concerning behavior; we study whether or not a user visits a group of interest. Then generates chart the based on the user visit group in the month.

CHART GENERATION FOR USER/ GROUP

Two data sets reported in are used to examine our proposed model for collective behavior learning. The first data set is acquired from user interest, the second from concerning behavior; we study whether or not a user visits a group of interest. Then generates chart the based on the user visit group in the month.

ALGORITHM

Algorithm for Learning of Collective Behavior

Input: network data, labels of some nodes, number of social dimensions;

Output: labels of unlabeled nodes.

1. Convert network into edge-centric view.

2. Perform edge clustering

3. Construct social dimensions based on edge partition node belongs to one community as long as any of its neighboring edges is in that community.

4. Apply regularization to social dimensions.

5. Construct classifier based on social dimensions of labeled nodes.

6. Use the classifier to predict labels of unlabeled ones based on their social dimensions.

V. CONCLUSION

It is well known that actors in a network demonstrate correlated behaviors. In this work, we aim to predict the outcome of collective behavior given a social network and the behavioral information of some actors.
In particular, we explore scalable learning of collectivebehavior when millions of actors are involved in the network. Our approach follows a social-dimensionbased learning framework. Social dimensions are extracted to represent the potential affiliations of actors before discriminative learning occurs. As existing approaches to extract social dimensions suffer from scalability, it is imperative to address the scalability issue. We propose an edge-centric clustering scheme to extract social dimensions and a scalable k-means variant to handle edge clustering. Essentially, each edge is treated as one data instance, and the connected nodes are the corresponding features. Then, the proposed k-means clustering algorithm can be applied to partition the edges into disjoint sets, with each set representing one possible affiliation. With this edge-centric view, we show that the extracted social dimensions are guaranteed to be sparse. This model, based on the sparse social dimensions, shows comparable prediction performance with earlier social dimension approaches.
An incomparable advantage of our model is that it easily scales to handle networks with millions of actors while the earlier models fail.
This scalable approach offers a viable solution to effective learning of online collective behavior on a large scale. In social media, multiple modes of actors can be involved in the same network, resulting in a multimode network . For instance, in YouTube, users, videos, tags, and comments are intertwined with each other in co-existence.
Extending the edge-centric clustering scheme to address this object heterogeneity can be a promising future direction. Since the proposed *EdgeCluster* model is sensitive to the number of social dimensions as shown in the experiment, further research is needed to determine a suitable dimensionality automatically. It is also interesting to mine other behavioral features (e.g., user activities and temporal spatial information) from social media, and integrate them with social networking information to improve prediction performance.

REFERENCES

[1] L. Tang and H. Liu, “Toward predicting collective behavior viasocial dimension extraction,” IEEE Intelligent Systems, vol. 25,pp. 19–25, 2010.
[2] Relational learning via latent social dimensions,” in KDD ’09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: ACM, 2009, pp. 817–826.
[3] M. Newman, “Finding community structure in networks usingthe eigenvectors of matrices,” Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), vol. 74, no. 3, 2006. [Online]. Available: http://dx.doi.org/10.1103/PhysRevE.74.036104
[4] L.Tang and H. Liu, “Scalable learning of collective behaviorbased on sparse social dimensions,” in CIKM ’09: Proceeding of the 18th ACM conference on Information and knowledge management. New York, NY, USA: ACM, 2009, pp. 1107–1116.
[5] P. Singla and M. Richardson, “Yes, there is a correlation: - from social networks to personal behavior on the web,” in WWW 08: Proceeding of the 17th international conference on World Wide Web. New York, NY, USA: ACM, 2008, pp. 655–664.
[6] M. McPherson, L. Smith-Lovin, and J. M. Cook, “Birds of a feather: Homophily in social networks,” Annual Review of Sociology, vol. 27, pp. 415–444, 2001.
[7] A. T. Fiore and J. S. Donath, “Homophily in online dating:when do you like someone like yourself?” in CHI ’05: CHI ’05 extended abstracts on Human factors in computing systems. New York, NY, USA: ACM, 2005, pp. 1371–1374.
[8] H. W. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas, “Homophily in the digital world: A LiveJournal case study,” IEEE

ISSN Online: 2319 – 9253
Print: 2319 – 9245

Internet Computing, vol. 14, pp. 15–23, 2010.
[9] S. A. Macskassy and F. Provost, Classification in networked data: A toolkit and a univariate case study," J. Mach. Learn. Res., vol. 8, pp. 935–983, 2007.
[10] X. Zhu, "Semi-supervised learning literature survey," 2006. [Online]. http://pag1es.cs.wisc.edu/jerryzhu/pub/ssl survey 12 9 2006.pdf

AUTHORS PROFILE

1. **Rachamalla Sridhar**
   MCA,(M.Tech(cs))
   Vivekananda Institute of Technology and Sciences(N9)
2. **K..AnilKumar**
   Vivekananda Institute of Technology and Sciences(N9)
3. **Abdul Rasool.MD**
   Head of the Department Of CSE
   Vivekananda Institute of Technology and Sciences(N9)