International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Prediction of Group and Individual Behaviours on Social Web Sites

Siriplakota Balaji[1], Jhansi Lakshmi Vaddelle[2], B. Lakshmi[3]

[1]Department of CSE, MVR College of Engineering & Technology, Beside Hanuman Statue on NH - 9, Paritala,, Vijayawada, A. P. , India

[2]Department of CSE, MVR College of Engineering & Technology, Beside Hanuman Statue on NH - 9, Paritala,, Vijayawada, A. P. , India

[3]Department of Computer Applications, V. R. Siddhartha Engineering College, Kanuru, Vijayawada, A. P., India

Abstract: ***Today every user is using social media such as twitter, YouTube, Flicker, twitter for communication and what's going on around the world. In this paper we analyse the data regarding users of different social networks and predict their interesting areas and collective behaviour of each user. To predict the behaviour we collect the corpus from various social networking sites .Here we use the concept of edge centric clustering scheme to retrieve social dimensions. The behaviour is predicted based on social dimensions considered. This algorithm is used to address the scalability issue. The proposed approach can effectively maintain enormous amount of actors in demonstrating comparable prediction performance to other non-scalable methods.***

Keywords: Collective Behaviour, Edge centric clustering, Social Dimensions, Non-Scalable methods.

1. Introduction

Connections in networking media are not homogenous. 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.

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 ad-dressing this heterogeneity [1]. 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 [10]. With a huge number of actors, extracted dense social dimensions cannot even be held in memory, causing a serious computational problem. In this work, we propose an effective edge-centric approach to extract sparse social dimensions [2].

2. Collective behavior

Collective Behavior refers to the behaviors of individuals in a social networking environment. In a connected environment individuals behavior tend to be inter dependent, influenced by the behavior of friends. This naturally leads to behavioral correlation between the users [3]. Take marketing as an example: if our friends buy something, there is a better-than-average chance that we will buy it, too.

This behavior correlation can also be explained by homophily [4]. When people are exposed in a social network environment, their behaviors can be influenced by the behaviors of their friends. One special case is K = 1, indicating that the studied behavior can be described by a single label with 1 and 0. For example, if the event is the presidential election, 1 or 0 indicates whether or not a voter voted for Barack Obama. The problem we study can be described formally as follows:

Suppose there are K class labels Y = {c1, ··, cK}. Given network G = (V, E, Y) where V is the vertex set, E is the edge set and Y ⊆ Y are the class labels of a vertex vi ∈ V, and known values of Yi for some subsets of vertices V L, how can we infer the values of Yi (or an estimated probability over each label) for the remaining vertices V U = V − VL?

Table 1: Social Dimension Representation

Actors	Affiliation-1	Affiliation-2 ···	Affiliation-
1 2	0	1	··
		0.8 0.5	
.	0.3	···	0

3. Social Dimensions

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 (category) while others are not. This relation-type information, however, is often not readily available in social media. A direct application of collective inference [5] or label propagation [6] would treat connections in a social network as if they were homogeneous. To address the heterogeneity present in connections, a frame-work (SocioDim) [1] has been proposed for collective behavior learning.

The framework SocioDimis composed of two steps: 1) social dimension extraction, and 2) discriminative learning. In the first step, latent social dimensions are extracted based on network topology to capture the potential affiliations of actors.. One example of the social dimension representation is shown in Table 1. The entries in this table denote the degree of one user involving in an affiliation.

In the initial instantiation of the framework SocioDim, a spectral variant of modularity maximization [3] is adopted to extract social dimensions. It has been empirically shown that this framework outperforms other representative relational learning methods on social media data. However, there are several concerns about the scalability of SocioDim with modularity maximization:

Figure 1: Toy Example, Edge Clusters

a. Social dimensions extracted according to soft clustering, such as modularity maximization and probabilistic methods, are dense. Suppose there are 1 million actors in a network and 1, 000 dimensions are extracted. If standard double precision numbers are used, holding the full matrix alone requires 1M ×1K ×8 = 8G memory.

b. Networks in social media tend to evolve, with new members joining and new connections occurring between existing members each day. This dynamic nature of networks entails an efficient update of the model for collective behavior pre- diction.

Consequently, it is imperative to develop scalable methods that can handle large-scale networks efficiently without extensive memory requirements. Next, we elucidate on an edge-centric clustering scheme to extract sparse social dimensions. With such a scheme, we can also update the social dimensions efficiently when new nodes or new edges arrive.

Table 2: Social Dimension(s) of the Toy Example

Actors	*Modularity Maximization*	*Edge Partition*
1	-0.1185	1 1
2	-0.4043	1 0
3	-0.4473	1 0
4	-0.4473	1 0
5	0.3093	0 1
6	0.2628	0 1
7	0.1690	0 1
8	0.3241	0 1
9	0.3522	0 1

(i.e., the density of connectivity is very low), whereas the extracted social dimensions are not sparse. Let's look at the toy network with two communities in Figure 1. Its social dimensions following modularity maximization are shown in Table 2. Clearly, none of the entries is zero. The corresponding memory requirement hinders both the extraction of social dimensions and the subsequent discriminative learning. Hence, it is imperative to develop some other approach so that the extracted social dimensions are sparse.

4. Sparse Social Dimensions

In this section, we first show one toy example to illustrate the intuition of communities in an "edge" view and then present potential solutions to extract sparse social dimensions.

Though SocioDim with soft clustering for social dimension extraction demonstrated promising results, the dashed edges represent one affiliation, and the remaining edges denote the second affiliation. The disjoint edge clusters in Figure 2 can be converted into the representation of social dimensions as shown in the last two columns in Table 2, where an entry is 1 (0) if an actor is (not) involved in that corresponding social dimension. Node 1 is affiliated with both com-munities because it has edges in both sets.

In order to partition edges into disjoint sets, one way is to look at the "dual" view of a network, i.e., the line graph [8]. We will show that this is not a practical solution. In a line graph L(G), each node corresponds to an edge in the original network G, and edges in the line graph represent the adjacency between two edges in the original graph. The line graph of the toy example is shown in Figure 4. For instance, e(1, 3) and e(2, 3) are connected in the line graph as they share one terminal node 3.

5. Experimental Setup

In this section, we present the data collected from social media for evaluation and the baseline methods for comparison. Another related approach to finding edge partitions is bi-connected components [9]. Each bi-connected component is considered a community, and converted into one social dimension for learning.

5.1 Social Media Data

Two data sets reported in [1] are used to examine our proposed model for collective behavior learning. The first data set is acquired from BlogCatalog3, the second from a popular photo sharing site Flicker.

5.2 Social Media Data

From the above discussion for large scale networks constructing a line graph is prohibitive.
3.http://www.blogcatalog.com/
4. http://www.flickr.com/
5.http://socialnetworks.mpi-sws.org/data-imc2007.html
6.http://www.youtube.com/learning methods based on collective inference. We study how the sparsity in social dimensions affects the prediction performance as well as the scalability.

6. Results and Observations

In this section, we first examine how prediction performances vary with social dimensions extracted following different approaches. Then we verify the sparsely of social dimensions and its implication for scalability and is

shown in Figure 2. We also study how the performance varies with dimensionality. Finally, concrete examples of extracted social dimensions are given.

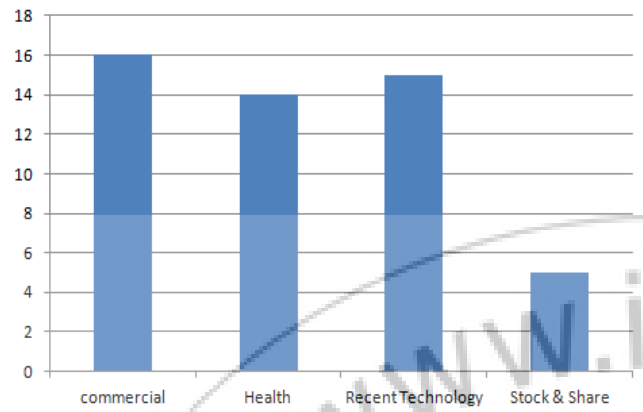

Figure 2: Social dimension and its implication Scalability

7. Conclusion and Future Work

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 net-work and the behavioral information of some actors. In particular, we explore scalable learning of collective behavior when millions of actors are involved in the network. Our approach follows a social-dimension-based learning framework. 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. This is based on the sparse social dimensions, shows comparable prediction performance with earlier social dimension approaches.

References

[1] "Relational learning via latent social dimensions," in KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. New Y ork, NY, USA: ACM, 2009, pp. 817–826.L.

[2] L. T ang and H. Liu, "Scalable learning of collective behavior based on sparse social dimensions," in CIKM '09: Proceeding of the 18th ACM conference on Information and knowledge management. New Y ork, NY, USA: ACM, 2009, pp. 1107–1116.

[3] 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.

[4] 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.

[5] 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.

[6] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in ICML, 2003.

[7] M. Newman, "Finding community structure in networks using the eigenvectors of matrices," Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), vol. 74, no. 3, 2006. [Online]. A vailable: http://dx.doi.org/10.1103/PhysRevE.74.036104.

[8] F . Harary and R. Norman, "Some properties of line digraphs," Rendiconti del Circolo Matematico di Palermo, vol. 9, no. 2, pp. 161–168, 1960.

[9] J. Hopcroft and R. T arjan, "Algorithm 447: efficient algorithms for graph manipulation," Commun. ACM, vol. 16, no. 6, pp. 372–378, 1973.

[10] Tang and H. Liu, "Toward predicting collective behavior via social dimension extraction," IEEE Intelligent Systems, vol. 25, pp. 19–25, 2010.

Author Profile

S. Balaji, persuing M.Tech (CSE) at M.V.R. College of Engineering and Technology, Paritala. He is interested in various topics like Data Mining RDBMS, and Computer Networks.

Ms. V. Jhansi Rani is Previously worked as an Asst. professor in SRK Institute of Technology and Sciences. She is expert in Advanced Computer Architecture.

Ms. B. Lakshmi is an Asst. Professor, Department of Computer Applications, VRSEC (Autonomous), Vijayawada, Andhra Pradesh. Her areas of interest are Database Management Systems, Operating Systems.