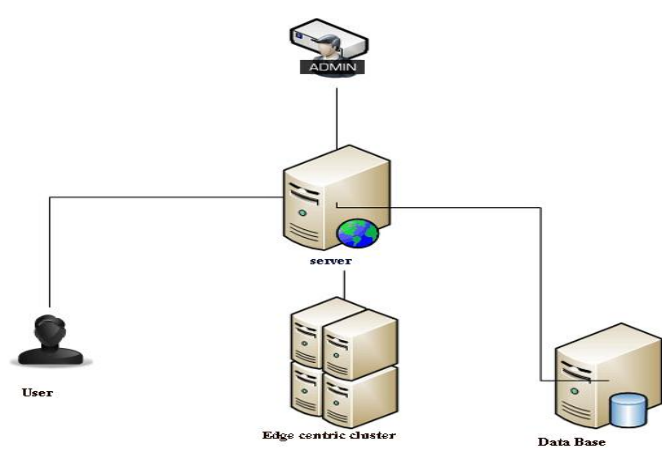
SCALABLE LEARNING OF COLLECTIVE

BEHAVIOUR

Abstract:

This study of collective behavior is to understand how individuals behave in a social networking environment. Oceans of data generated by social media like Face book, Twitter, Flicker, and YouTube present opportunities and challenges to study collective behavior on a large scale. In this work, we aim to learn to predict collective behavior in social media. In particular, 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. However, the networks in social media are normally of colossal size, involving hundreds of thousands of actors. 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.

Architecture:



Algorithm:

1. 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 as in Figure 5.

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.

Existing System:

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.

Disadvantages:

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

Proposed System:

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.

Advantages:

- 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.

Modules:

1. **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.

2. **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.

3. **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.

4. 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.

System Requirements:

Hardware Requirements:

Processor	:	Intel Duel Core.
Hard Disk	:	60 GB.
Floppy Drive	:	1.44 Mb.
Monitor	:	LCD Colour.
Mouse	:	Optical Mouse.
RAM	:	512 Mb.

Software Requirements:

Operating system	:	Windows XP.
Coding Language	:	ASP.Net with C#
Data Base	:	SQL Server 2005