Data mining package report

Kandati Vishnu Sai, Midhilesh E., Sanjay Seetharaman

November 2020

1 Introduction

Hi! This is the section for introduction. Create your own tex file and include the corresponding reference in main.tex.

2 Graph Mining

Hi! This is a new section.

3 DeepInf: Social influence prediction

3.1 Introduction

Deepinf focus on the prediction of user-level social influence. It aims to **predict the action status of a user given the action**. It is a deep learning based framework to represent both influence dynamics and network structures into a latent space and tries to minimize the negative likelihood that was defined in the section 1.

3.2 Model framework

3.2.1 Sampling Near Neighbour

Give a user v, a r-ego network $\mathcal{G}_v^r$ is extracted using breadth-first search (BFS) starting from user v. However, $\mathcal{G}_v^r$ may have different size due to the small world property in the social network. Since most deep learning models expects fixed size data, the graph $\mathcal{G}_v^r$ can be sampled to fixed size.

For sampling a fixed size graph **random walk the restart** (RWR) was used. RWR algorithm is defined as following steps.

- Start random walks from either the ego user v or one of her active neighbors randomly.

- The random walk iteratively travels to its neighborhood with the probability that is proportional to the weight of each edge.
- At each step, the walk is assigned a probability to return to the starting node, that is, either the ego user v or one of v's active neighbors.
- Run the algorithm until a fixed number of vertices denoted by $\hat{\Gamma_v^r}$ with $|\hat{\Gamma_v^r}|$=n.

After running this algorithm, a sub-graph $\hat{\mathcal{G}_v^r}$ and denote $\hat{S}_v^t$= { $s_u^t : u \in \hat{\Gamma_v^r}$ } be the action statuses of v's sampled neighbours. Therefore we re-define the optimization objective in section 1 as:

$$\mathcal{L}(\theta) = -\sum_{i=1}^{N} \log(P_\theta(s_{v_i}^{t_i+\Delta t} | \hat{\mathcal{G}_{v_i}^t}, s_{v_i}^{t_i})) \tag{1}$$

3.2.2 Embedded layer

3.2.3 Instane Normalization

3.2.4 Input Layer

3.2.5 GCN

3.3 Evaluation Metrics