Report About BL-MNE Algorithm

Hoben Wong

January 30, 2018

1 Background

1.1 Problem Studied

BL-MNE means Broad Learning based eMerging Network Embedding. In the concurrent cmbedding process based on the on the broad learning setting, BL-MNE aims at distilling relevant information from both the emerging and other aligned mature networks to derive compliment knowledge and learn a good vector representation for user nodes in the emerging network.

1.2 Challenges

There are several great challenges to deal with in terms of the BL-MNE problem. They can be list as followed:

- *Problem Formulation:* To overcome the information sparsity problem, BL-MNE studies the concurrent embedding of multiple aligned social networks, which is still an open problem to this context so far. Formal definition and formulation of the BL-MNE problem is required before we introduce the solutions.
- *Heterogeneity of Network:* Besides the regular social connections among users, there also exist many other types of links as well as diverse attributes attached to the user nodes.
- *Multiple Aligned Network Embedding Framework:* Due to the significant differences between BL-MNE with the existing works, few existing network embedding models can be applied to address the BL-MNE directly. A new embedding learning framework is needed to learn the emerging network embedding vectors across multiple aligned networks synergistically.

2 Terminology Definition And Problem Formulation

Definition 1 (Attributed Heterogeneous Social Networks): It can be represented as a graph $G = (\nu, \varepsilon, \tau)$, where $\nu = \cup_i \nu_i$ denotes the set of nodes belonging to various categories and $\varepsilon = \cup_i \varepsilon_i$ represents the set of diverse links among the nodes. $\tau = \cup_i \tau_i$ denotes the set of attributes attached to the node in τ. For user u in the network, the i_{th} type of attribute of u is $T_i(u)$, and all attributes u has can be represented as $T(u) = \cup_i T_i(u)$.

Formally, Foursquare and Twitter can be both represented as the *Attributed Heterogeneous Social Networks* $G = (\nu, \varepsilon, \tau)$, where $\nu = \mathcal{U} \cup \mathcal{P}$ involves the user and post nodes, and $\varepsilon = \varepsilon_{u,u} \cup \varepsilon_{u,p}$ contains the links among users and those between users and posts. In addition, the nodes in ν are also attached with a set of attributes, i.e.,τ. For instance, for the posts written by users, we can obtain the contained textual contents, timestamps and checkins, which can all be represented as the attributes of the post nodes.

Between Foursquare and Twitter, there may exist a large number of shared common users, who can align the networks together. Meanwhile, the networks connected by the anchor links are called the multiple aligned attributed heterogeneous social networks (or aligned social networks for short).

Definition 2 (Multiple Aligned Social Networks): Formally, given n attributed heterogeneous social networks $\left\{G^{(1)}, \ldots, G^{(n)}\right\}$ with shared users, they can be defined as *Multiple Aligned Social Networks* $\mathcal{G} = ((G^{(1)}, \ldots, G^{(n)}), (\mathcal{A}^{(1,2)}, \ldots, A^{(n-1,n)}))$, where $\mathcal{A}^{(i,j)}$ represents the anchor links between $G^{(i)}$ and $G^{(j)}$. User pair $(u^{(i)}, v^{(j)} \in \mathcal{A}^{(i,j)}$ iff $u^{(i)}$ and $v^{(j)}$ are the accounts of the same user in networks $G^{(i)}$ and $G^{(j)}$ respectively.

For the Foursquare and Twitter social networks, they can be presented as two aligned social networks $\mathcal{G} = ((G^{(1)}, G^{(2)}), (\mathcal{A}^{(1,2)})$, which will be used as an example to illustrate the models. A simple extension of the proposed framework can be applied to *k aligned networks* very easily.

Problem Definition(BL-MNE Problem): Given two aligned networks $\mathcal{G} = ((G^{(1)}, G^{(2)}), (\mathcal{A}^{(1,2)})$, where $G^{(1)}$ is an emerging network and $G^{(2)}$ is a mature network, BL-MNE aims at learning a mapping function $f^{(i)} : \mathcal{U}^{(i)} \to \mathbb{R}^{d^{(i)}}$ to project the user node in $G^{(i)}$ to a feature space of dimension $d^{(i)}(d^{(i)} \ll |\mathcal{U}^{(i)}|)$. The objective of mapping functions $f^{(i)}$ is to ensure the embedding results can preserve the network structural information, where similar user nodes will be projected to **close** regions. Furthermore, in the embedding process, BL-MNE also wants to transfer information between $G^{(2)}$ and $G^{(1)}$ to overcome the information sparsity problem in $G^{(1)}$.

3 Proposed Method

3.1 Heterogeneous Network Meta Proximity

3.1.1 Friendship based Meta Proximity

In online social networks,the friendship links are the most obvious indicator of the social closeness among users. Online friends tend to be closer with each other compared with the user pairs who are not friends. Users friendship links also carry important information about the local network structure information,

which should be preserved in the embedding results. Based on such an intuition, the *friendship based meta proximity* concept can be proposed as follows.
Definition 3(Friendship based Meta Proximity): For any two user node $u_i^{(1)}, u_j^{(1)}$ in an online social network $G^{(1)}$, there exists a definition

$$p^{(1)}(u_i^{(1)}, u_j^{(1)}) = \begin{cases} 1, & \text{if } (u_i^{(1)}, u_j^{(1)}) \in \varepsilon_{u,u}^{(1)} \\ 0, & \text{otherwise.} \end{cases}$$

Here $p^{(1)}(u_i^{(1)}, u_j^{(1)})$, or in another expression, $\mathcal{P}_{\phi}^{(1)}(i,j)$, represents the *friendship based meta proximity score* between users $u_i^{(1)}, u_j^{(1)}$. Thus, based on the above definition, the *friendship based meta proximity* scores among all the users in the network $G^{(1)}$ can be represented as matrix $\mathbf{P}_{\phi_0}^{(1)} \in \mathbb{R}^{|\mathcal{U}^{(1)} \times \mathcal{U}^{(1)}|}$. ϕ_0 denotes the simplest meta path of length 1 in the form U $\xrightarrow{follow}$ U. Here, if $G^{(1)}$ is an emerging online social network, $\mathbf{P}_{\phi_0}^{(1)}$ will have few value 1 and most 0, in other words, it is a sparse matrix and as a result, most existing embedding models will fail to work. The reason is that the sparse friendship information available in the network can hardly categorize the relative closeness relationships among the users (especially for those who are even not connected by friendship links), which renders these existing embedding models may project all the nodes to random regions.

3.1.2 Attribute Augmented Meta Path

To handle the diverse links and attributes simultaneously in a unified analytic, we propose to treat the attributes as nodes as well and introduce the attribute augmented network.
Definition 4 (Attribute Augmented Network Schema): Formally, the network schema of a given online social network $G^{(1)} = (\nu, \varepsilon)$ can be represented as $s_{G^{(1)}} = (\mathcal{N}_{\nu} \cup \mathcal{N}_{\tau}, \mathcal{R}_{\varepsilon} \cup \{have\})$, where $\mathcal{N}_{\nu}$ and $\mathcal{N}_{\tau}$ denote the set of node and attribute categories in the network, while $\mathcal{R}_{\varepsilon}$ represents the set of link types in the network, and {have} means the relationship between node and attribute node types. The node type set $\mathcal{N}_{\nu}$ involves node type User,Post while the attribute type set $\mathcal{N}_{\tau}$ includes Word, Time, Location. The link type set $\mathcal{R}_{\varepsilon}$ contains follow, write, which represents the friendship link type and the write link type respectively.
Definition 5 (Attribute Augmented Meta Path): Given a network schema $s_{G^{(1)}}$, the *Attribute Augmented Meta Path* denotes a sequence of node/attribute types connected by the link types or the have relation type. Formally, the *Attribute Augmented Meta Path* can be represented as $\phi : N_1 \xrightarrow{R1} N_2 \xrightarrow{R2} \dots \xrightarrow{R_{k-1}} N_k$, where $N_1, \dots, N_k \in \mathcal{N}_{\nu} \cup \mathcal{N}_{\tau}$ and $R_1, \dots, R_{k-1} \in \mathcal{R}_{\varepsilon} \cup \mathcal{R}_{\varepsilon}^{-1} \cup \{have, have^{-1}\}$(-1 denotes the reverse of relation type direction). If $N_1 = N_K = U$, for example, meta paths starts and ends with the user node type, the meat path will be called the *social meta paths.*

Based on the above definition, a set of different *social meta paths* $\{\phi_0, \phi_1, \phi_2, \dots, \phi_7\}$ can be extracted from the network, it can be shown on the following table.

ID	Notation	Heterogeneous Network Meta Path	Semantics
ϕ_0	$U \rightarrow U$	$User \xrightarrow{follow} User$	Follow
ϕ_1	$U \rightarrow U \rightarrow U$	$U \xrightarrow{follow} U \xrightarrow{follow} U$	Follower of Follower
ϕ_2	$U \rightarrow U \leftarrow U$	$U \xrightarrow{follow} U \xrightarrow{follow^{-1}} U$	Common Out Neighbor
ϕ_3	$U \leftarrow U \rightarrow U$	$U \xrightarrow{follow^{-1}} U \xrightarrow{follow} U$	Common In Neighbor
ϕ_4	$U \leftarrow U \leftarrow U$	$U \xrightarrow{follow^{-1}} U \xrightarrow{follow^{-1}} U$	Followee of Followee
ϕ_5	$U \rightarrow P \rightarrow W \leftarrow P \leftarrow U$	$User \xrightarrow{write} Post \xrightarrow{have} Word \xrightarrow{have^{-1}} Post \xrightarrow{write^{-1}} User$	Posts Containing Common Words
ϕ_6	$U \rightarrow P \rightarrow T \leftarrow P \leftarrow U$	$User \xrightarrow{write} Post \xrightarrow{have} Time \xrightarrow{have^{-1}} Post \xrightarrow{write^{-1}} User$	Posts Containing Common Timestamps
ϕ_7	$U \rightarrow P \rightarrow L \leftarrow P \leftarrow U$	$User \xrightarrow{write} Post \xrightarrow{have} Location \xrightarrow{have^{-1}} Post \xrightarrow{write^{-1}} User$	Posts Attaching Common Location Check-i