

Incremental learning with latent factor models for attribute prediction in social-attribute networks

Duc Kinh Le Tran^{*,**} Cécile Bothorel^{*}
Pascal Cheung Mon Chan^{**}

^{*}UMR CNRS 3192 Lab-STICC
Département LUSI – Télécom Bretagne
{duc.letran, cecile.bothorel}@telecom-bretagne.eu
^{**}Orange Labs
{duckinh.letran, pascal.cheungmonchan}@orange.com

Abstract. In this paper, we are interested in the problem of predicting attributes on the nodes in a social network. Most of the existing techniques addressing this problem are offline learning techniques and are not suitable in situations where massive data come in stream like social media. In this work, we use *latent factor models* to predict unknown attributes of the nodes in a social network and propose a method to incrementally update the prediction model on the arrivals of new data. Experiments on a real social media dataset show that our method is more rapid and can guarantee acceptable performances in comparison with state-of-the-art non-incremental techniques.

1 Introduction and problem statement

With the explosion of social media on the Internet in recent years, mining social media content has become more and more critical for many domains. One of the challenges of mining social media is how to leverage *relational information* (e.g. friendships, interactions between social media users) and simultaneously *attributes* (e.g. users' interests, textual or any other additional information). Another challenge lies in the fact that these media provide vast and continuous streams of data. Using offline learning techniques, we have to aggregate all the data available from the past until the present. This approach is not suitable in this situation because (1) as new data come, the size of the dataset grows, it gets more and more expensive to learn and to apply the model (2) this approach cannot capture the dynamic of the data stream: old data and recent data are treated uniformly.

In this paper, we address both challenges by introducing an incremental learning method for the task of predicting attributes of social actors in a social network. This problem has many real world applications, for example to predict users' interests or hobbies using social media. We build a graph of interactions among the social media users and enrich the graph with a set of attributes on nodes. As the data (nodes, links) arrive as a permanent stream, we want to build models to periodically predict unknown attributes on the nodes.

To formulate our problem, we adopt the *social-attribute network* (Yin et al. (2010)). A *social-attribute network (SAN)* contains a social network $G_s=(V_s, E_s)$ where V_s is the set of