Journal Review on “Collaborative Boosting for Activity Classification in Microblogs”

Yangqiu Song§, Zhengdong Lu‡, Cane Wing-ki Leung‡, Qiang Yang‡

yqsong@cse.ust.hk {lu.zhengdong,cane.leung,qiang.yang}@huawei.com

§Department of Computer Science and Engineering,

Hong Kong University of Science and Technology, Hong Kong

Noah’s Ark Lab, Huawei, Hong Kong

Objective/Abstract:

This paper describes how to recognize activity of individual at a particular time based on comments posted in blogs. It has used data sources from a well-known site called Sina Weibo (a hybrid version of Twitter and Facebook in China). The main goal of this paper is to focus on “activity recognition” and the need for personalized classifiers to prove that their method out-performs similar algorithms that do not take into account of “users’ individuality or social connections”. The interesting point that this paper has mentioned is the importance of social connections such as the connected friends activities like response or comments to the individual plays an important part in strengthening the possibility of the mentioned activity in blogs is true.

The paper claims the focus in activity recognition has not been greatly explore rather that majority of the research is done on event recognition. I believe this study will provide an new prediction method on individual activities at a particular time thus it will benefit groups such as target advertisers or in providing live data to public transportation authorities for crowd control or scheduling of train departures.

Style:

I believe this paper is intended for other data scientist as it highlights their approach in algorithm for activity recognition will out-performs others in the market. It also has a detail section to introduce related research and identifies how it differs from them. On the technical side, it emphasized the need of personal classifier to improve the results of activity recognition whereas the typical method is based on pure global classifiers.

Methods:

The methodology related to their work focus namely on collective classification and multi-task/transfer learning approach in achieving its objective. The paper describes how they fine tune each method to enhance the performance of their tailor algorithm known as SocialBoost. As mentioned earlier, the data set came from Weibo and their experiment was able to access 10 million users. Even though the data set seems more than sufficient they had issues with their training set as each individual message is quite short to conclude the activity at that point of time. To overcome this they rely on the inter-dependent instances and in this case the individual’s connection such as friends or chat buddies. With the message exchange between the connections supported by their algorithm it was able to increase the accuracy to determine the activity at that particular time. The finding in this paper clearly shows their SocialBoost algorithm out-performs all other six baseline methods in every comparison.

.

Conclusions:

I must give great credit to the paper even just based on the formatted structure with the detailing of each method and how they differentiate from current research. The numbers from the finding strengthens or confirms the explanation on how the personal classifier improves the accuracy on activity recognition. In addition, they were able to use real-life data from Weibo giving them almost unlimited amount of training, test and new data from random individuals that produce endless combination of message exchange. My only question if I had to ask is how good would this algorithm be when it is tested against blogs in other language which this experiment is done on a Chinese based blog.