

Abstract

1. Reddit project
2. Display Advertising project

Description

1. **Reddit project**

We implemented the algorithm again to make it faster and support more configuration like train / validation splitting or different initialization states or balanced input classes. Unfortunately it came up much slower. Our results were 85% AUC and 95% AUPR. Our model at its best threshold performed great having accuracy = 87.5%, recall = 92.3% and precision = 91.7%.

We reviewed many papers related to gender classification and compared their results by ours but there wasn't any approach having such performance with so little data we used (our approach uses only network data).

Our approach can be classified into semi-supervised methods and can work in absence of large amounts of data. All this algorithm needs is a fraction of correctly profiled users and the network structure of a social network (e.g. reddit).

2. **Display Advertising project**

I worked on major questions of my thesis project. I reviewed some papers and discussed some ideas with Ali Osia about how to manage the ad historical data. I think its a good time to get involved with real data and see what its look like.

Next Week

- Reddit project (Find more data using network alignment - add more attributes to data - find a way to treat non-binary attributes)
- Display Advertising project (Meeting with Tapsell and looking at data - editing thesis papers table)

References