Measuring Social Influence on Online Collaborative Communities

Zhe-Li Lin, Yu-Ming Lu, Ming-Fend Tsai
Department of Computer Science, National Chengchi University, Taiwan

Chuan-Ju Wang
Department of Computer Science, University of Taipei, Taiwan

Abstract

How to measure an individual influencing others within an online social network in a quantitative way is a critical problem in the field of computational social science. This paper attempts to observe collaborative events occurring at individuals in a social network to obtain such crucial knowledge. We propose a framework with Factorization Machines (FM) to model the social influence among the individuals based on their collaborations; meanwhile, due to the essence of FM, any auxiliary information can be integrated into the modeling process in a straightforward manner. We conduct the experiments on a dataset collected from GitHub, a web-based Git repository hosting service that provides programmers an effective way to collaborate on development projects. In the experiments, we utilize not only the collaborative information among programmers but incorporate various supplementary information, such as user profile (e.g., the number of owned repositories and followers), repository profile (e.g., the number of stars and forks), and textual information (e.g., the title of a repository). The experimental results verify that the effectiveness of the proposed framework on providing better predictive models than several baseline methods. Furthermore, through the experimental results, we observe some interesting social phenomena and provide further analyses and discussions.

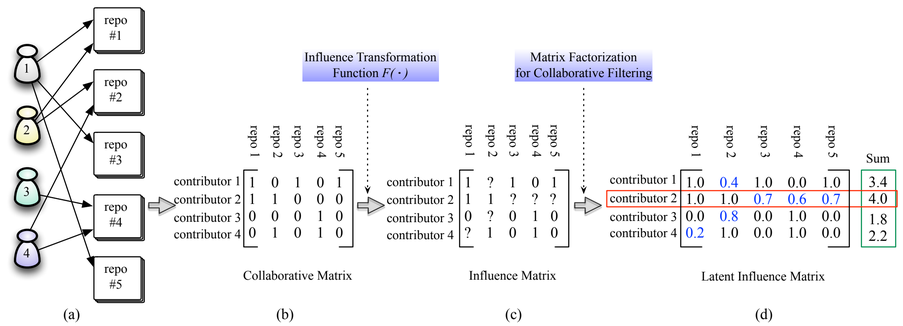
Methodology

Collaborative Latent Social Influence

Collaborative Filtering (CF) is a common technique adopted by recommendation systems. We attempt to model the latent social influence of people in a certain research community with this technique, which filters information or patterns involving collaboration among people. The figure as below gives an illustrative example to introduce the core idea of the proposed framework for modeling the latent social influence.

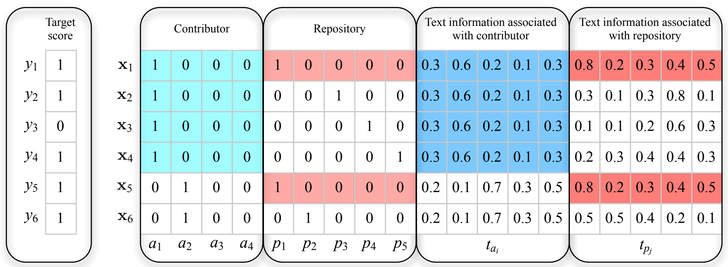
$$F(x_{a_i,p_j}) = \begin{cases} 1, & \text{if } a_i \text{ is the contributor of } p_j, \\ ?, & \text{if } \exists\, a_k \in C_{a_i} \text{ and } a_k \text{ is the contributor of } p_j, \\ 0, & \text{otherwise,} \end{cases}$$

where C_{ai} is the set of the contributors who have collaborated with contributor a_i.
These relationships can be transformed to the coauthor matrix in the following figure.



Modeling Social Influence with FM

Factorization Machines (FM) provides an advantage over other existing CF approaches, which makes it possible to incorporate with any auxiliary information that can be encoded as a real-valued feature vector. Thus, via using FM, we integrates with other supplementary information model latent social influence, and we use textual information as the supplementary information in our experiments.



Dataset and Experimental Setup

The experimental dataset is built from **Github**, which contains the information of repositories and the contributors of each repository. This dataset consists of **4568** programmers and **529** repositories. In the experiments , the gold standard adopted to evaluate the performance is the ranking list provided by Github Ranking (https://github-ranking.com).

Experimental Results

The experimental results are shown in the following table, in which we compare the results of two baselines and the proposed FM framework. The first two baselines are the ranking via the numbers of contributors and repositories.

Gold Standards	Evaluation Metrics	(∗) #Contributors	(†) #Repos	(§) FM
Github Ranking	ρ	0.6	0.539	$0.781^{*\dagger}$
	τ	0.467	0.378	$0.627^{*\dagger}$

Top 10 Learned Terms

We list the top 10 words learned from the FM model with textual information as below.

Repo	Contributor
generator	comment
hipchat	buildpack
notification	csv
jquery	db
element	comic
directory	green
stackoverflow	javascript
queue	legit
curator	Werkzeug
enhance	article