Social Influencer Analysis with Factorization Machines

Ming-Feng Tsai and Zhe-Li Lin
Department of Computer Science, National Chengchi University, Taiwan

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

Abstract
This work attempts to observe the collaborative events occurring at individuals involved in a social network to obtain such crucial knowledge. We propose a Factorization Machine approach to find out the latent social influence among the individuals based on their collaborations. Experiments conducted on a real-world DBLP dataset verify that the proposed approach can discover the latent social influence among individuals and provide a better predictive model than several baselines.

Methodology
Fig. 1 gives an illustrative example to introduce the core idea of the modeling process for the latent social influence. Fig. 1(a) depicts the relationships between the authors and their papers. These relationships can be transformed to the matrix representation in Fig. 1(b), in which each element \(x_{a,p} \) equals to 1 if \( a \) is the author paper \( p \) and otherwise that equals to 0. We then define an influence transformation function \( F(x_{a,p}) \) to build up the influence matrix (see Fig. 1(c)); this is a key step to transform the relationships in Fig. 1(a) to the input of a standard CF algorithm. The transformation function \( F(x_{a,p}) \) can be designed variously; in this paper, \( F(x_{a,p}) \) is defined as

\[
F(x_{a,p}) = \begin{cases} 
1, & \text{if } a \text{ is the author of } p, \\
0, & \text{otherwise},
\end{cases}
\]

where \( C_a \) is the set of the authors who have coauthored with \( a \). After the transformation, we can obtain the resulting matrix in Fig. 1(d) via any CF algorithms. In Fig. 1(d), each number in blue color can be explained as the estimated latent social influence; the numbers in the green box are the sum of the influence scores of each author on all papers. As shown in the figure, we can observe that although author 2 has only written 2 papers, his/her social influence score (i.e., 4) is larger than that of author 1 (i.e., 3.4), who has written the most papers among the 4 authors. Even though author 2 is not the author of papers 3, 4, and 5, we consider that author 2 should still have latent influence on these three papers and the influence can be modeled with the patterns of collaborations among the authors. FM provides an advantage over other existing CF approaches, which make it possible to incorporate with any auxiliary information that can be encoded as a real-valued feature vector. Thus, via using FM, this paper integrates with text information to model latent social influence.

Conclusions and Future Work
This study attempts to model the latent social influence among individuals based on their patterns of collaborations in a social network via the FM approach. Preliminary experimental results on the small DBLP dataset for the data mining community show that the proposed approach provides a better predictive model than several baselines. In future work, we will conduct experiments on larger data sets with various fields of research communities. In addition, other auxiliary information, such as the temporal information of the publications [5], will be included and analyzed in our further experiments.

Table 1: The Experimental Results. The notations *, †, and ‡ in Table 1 denote the result is significant better than the three baselines #coauthor, #paper, and #citation, respectively, with \( p < 0.05 \).

<table>
<thead>
<tr>
<th></th>
<th>Spearman’s Rho</th>
<th>Kendall’s Tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>#coauthor (baseline)</td>
<td>0.233</td>
<td>0.179</td>
</tr>
<tr>
<td>#paper (baseline)</td>
<td>0.388</td>
<td>0.284</td>
</tr>
<tr>
<td>#citation (baseline)</td>
<td>0.469</td>
<td>0.347</td>
</tr>
<tr>
<td>FM without texts</td>
<td>0.478**†</td>
<td>0.349†</td>
</tr>
<tr>
<td>FM with texts</td>
<td>0.556**†</td>
<td>0.409**†</td>
</tr>
</tbody>
</table>