Project Report
Graphical Models LAB

Maurice Wenig
Friedrich Schiller University Jena
Germany
maurice.wenig@uni-jena.de

1 INTRODUCTION

Some introduction.

2 METHODS

2.1 Choice of optimized score

For the score function $p(G|D) \propto p(D|G) \cdot p(G)$, i used the approximation

$$p(D|G) \approx p(D|G, \hat{\theta})$$
$$\hat{\theta} := \arg\max_{\theta} p(D|G, \theta)$$

and the prior

$$p(G) \propto \frac{1}{|E|^{\lambda}}.$$

With this, the objective function can be decomposed into the sum of independent node scores and a regularization term.

$$\begin{aligned}\arg\max_G p(G|D) &= \arg\max_G \log p(D|G) + \log p(G) \\ &\approx \arg\max_G p(D|G, \hat{\theta}) - \lambda\, |E_G| \\ &= \arg\max_G \sum_{i \in [n]} S_i(G) - \lambda\, |E_G|\end{aligned}$$

where

$$S_i(G) := -|D| \cdot \log \hat{\sigma}_i - \frac{1}{2} \sum_{x \in D} \left(\frac{x_i - (\hat{\beta}_i^{\top} x_{\mathrm{pa}(i)} + \hat{\beta}_i^{*})}{\hat{\sigma}_i} \right)^2$$

and $\hat{\beta}_i, \hat{\beta}_i^{*}, \hat{\sigma}_i$ are the respective ML estimates for

$$p(x_i | x_{\mathrm{pa}(i)}) \sim \mathcal{N}(\hat{\beta}_i^{\top} x_{\mathrm{pa}(i)} + \hat{\beta}_i^{*}, \sigma_i^2).$$

A derivation of this can be found in subsection A.1.

2.2 Implications for Hill Climbing

Elementary changes (addition, substraction, flip of an edge) only influence local distributions. That means if we construct a graph G', where G' was made by applying an elementary change to an edge (u, v) in G, the comparison $p(G'|D) > p(G|D)$ can be evaluated locally:

$$\begin{aligned} & \sum_{i \in [n]} S_i(G') - \lambda\, |E_{G'}| > \sum_{i \in [n]} S_i(G) - \lambda\, |E_G| \\ \iff & \sum_{i \in [n]} S_i(G') - S_i(G) > \lambda(|E_{G'}| - |E_G|) \\ \iff & \underbrace{\sum_{i \in \{u,v\}} S_i(G') - S_i(G)}_{:=\Delta_S(G',G)} > \lambda \Delta_E(G', G) \end{aligned}$$

where

$$\Delta_E(G', G) := \left|E'_G\right| - |E_G| = \begin{cases} 1 & \text{addition} \\ 0 & \text{flip} \\ -1 & \text{deletion} \end{cases}$$

Note that $\Delta_S(G', G)$ measures the improvement of node scores when the change is applied to G.

Similarly, two alterations G_1 and G_2 of G can be compared:

$$\begin{aligned} & \sum_{i \in [n]} S_i(G_1) - \lambda \left|E_{G_1}\right| > \sum_{i \in [n]} S_i(G_2) - \lambda \left|E_{G_2}\right| \\ \iff & \Delta_S(G_1, G) - \Delta_S(G_2, G) > \lambda\left(\Delta_E(G_1, G) - \Delta_E(G_2, G)\right) \end{aligned}$$

A derivation of this can be found in subsection A.2.

The implication of this is that hill climbing stops whenever the best improvement of the node scores is smaller than difference in edges that the change causes, scaled by the regularization constant.

3 RESULTS

3.1 Performance

Some performance analysis

3.2 Likelihood

Cross validation time.

Table 1: Error Score Comparison

Recommender	RMSE	MAE
user based	0.670	0.301
item based	0.569	0.222
factorization	0.512	0.207
hybrid	0.496	0.193

4 CONCLUSION

Some conclusion.

REFERENCES

A CALCULATIONS

A.1 Score Function

Here we derive

$$\max_G p(D|G,\hat{\theta}) = \max_G \sum_{i\in[n]} S_i(G)$$

$$\begin{aligned}
&\max_G p(D|G,\hat{\theta}) \\
&= \max_G \prod_{x\in D} p(x|G,\hat{\theta}) \\
&= \max_G \prod_{x\in D} \prod_{i\in[n]} p(x_i|x_{\mathrm{pa}(i)}, \hat{\beta}_i, \hat{\beta}_i{}^*, \hat{\sigma}_i) \\
&= \max_G \sum_{x\in D} \sum_{i\in[n]} \log p(x_i|x_{\mathrm{pa}(i)}, \hat{\beta}_i, \hat{\beta}_i{}^*, \hat{\sigma}_i) \\
&= \max_G \sum_{x\in D} \sum_{i\in[n]} \log\left[\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{1}{2}\left(\frac{x_i - (\hat{\beta}_i{}^\top x_{\mathrm{pa}(i)} + \hat{\beta}_i{}^*)}{\sigma_i}\right)^2\right)\right] \\
&= \max_G \sum_{x\in D} \sum_{i\in[n]} \left[-\frac{1}{2}\left(\frac{x_i - (\hat{\beta}_i{}^\top x_{\mathrm{pa}(i)} + \hat{\beta}_i{}^*)}{\sigma_i}\right)^2 - \log\hat{\sigma}_i - \frac{1}{2}\log 2\pi\right] \\
&= \max_G \sum_{i\in[n]} \underbrace{\left[-|D|\cdot\log\hat{\sigma}_i - \frac{1}{2}\sum_{x\in D}\left(\frac{x_i - (\hat{\beta}_i{}^\top x_{\mathrm{pa}(i)} + \hat{\beta}_i{}^*)}{\hat{\sigma}_i}\right)^2\right]}_{=S_i(G)} \qquad \square
\end{aligned}$$

Note that $\hat{\theta}$ $(\hat{\beta}_i, \hat{\sigma}_i)$ depends on G $(\mathrm{pa}(i))$.

A.2 Graph Comparison

For a graph G' that was made by applying one elementary change to a graph G, it holds that

$$\sum_{i\in[n]} S_i(G') = \sum_{i\in[n]} S_i(G) + \Delta_S(G',G) \tag{1}$$

and

$$|E_{G'}| = |E_G| + \Delta_E(G',G) \tag{2}$$

Therefore

$$\begin{aligned}
&\sum_{i\in[n]} S_i(G_1) - \lambda\left|E_{G_1}\right| > \sum_{i\in[n]} S_i(G_2) - \lambda\left|E_{G_2}\right| \\
\overset{(1),(2)}{\iff}\ & \Delta_S(G_1,G) - \lambda\Delta_E(G_1,G) > \Delta_S(G_2,G) - \lambda\Delta_E(G_2,G) \\
\iff\ & \Delta_S(G_1,G) - \Delta_S(G_2,G) > \lambda\left(\Delta_E(G_1,G) - \Delta_E(G_2,G)\right)
\end{aligned}$$