Modeling Heterogeneous Trust in Recommendation Systems

Pulkit Agrawal, Agrim Gupta and Jayanth Ramesh

Department of Electrical Engineering, Stanford University

Idea

People trust different sets of social connections for different types of recommendations. Can we model this heterogeneous trust to improve collaborative filtering algorithms?

Data

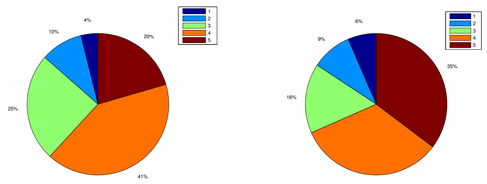


Figure 1: Yelp and Epinions Rating Distribution

Yelp

5614 **users**, 27688 (7598-S,20090-R) **businesses**, 247551 **reviews** (**sparsity**-0.16%),45634 **edges** (1712 zero degree nodes, maximum degree - 568), 88 **attributes**

Epinions

6874 **users**, 14412 (4307-E,10105-M) **products**, 74246 (57059-M, 17187-E) **reviews** (**sparsity**-0.07%),43733 **edges** (917 zero in-degree nodes, 641 zero out-degree nodes maximum degree - 314)

Modeling Heterogeneous Trust

Cosine Trust

Similar connections are more trusted

$$\mathbf{S}_{ij}^{l} = \begin{cases} \frac{\sum_b \mathbf{R}_{i,b}\mathbf{R}_{j,b}\mathbf{C}_{b,l}}{\sum_b \mathbf{R}_{i,b}\mathbf{C}_{b,l}\sum_b \mathbf{R}_{j,b}\mathbf{C}_{b,l}} \text{ if } \mathbf{T}_{ij} \neq 0 \\ 0 \text{ if } \mathbf{T}_{ij} = 0 \end{cases}$$

$\mathbf{S}_{ij}^{l}$ - trust strength for category l, $\mathbf{T}_{ij}$ - whether i trusts j, $\mathbf{C}_{b,l}$ - whether product b belongs to catrogry l

Recommendation Power

Two step random walk on weighted bipartite graph of users and products

$$\mathbf{S}_{ij}^{l} = \begin{cases} \frac{1}{\mathbf{R}_i}\sum_b \frac{\mathbf{R}_{i,b}\mathbf{R}_{j,b}\mathbf{C}_{b,l}}{\mathbf{R}_b\mathbf{C}_{b,l}} \text{ if } \mathbf{T}_{ij} \neq 0 \\ 0 \text{ if } \mathbf{T}_{ij} = 0 \end{cases}$$

$\mathbf{R}_b$ - Sum of all ratings received by b, $\mathbf{R}_i$ - Sum of all ratings given by i

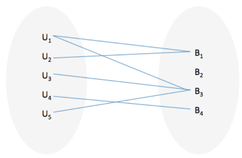


Figure 2: Recommendation Power using Bipartite Graph

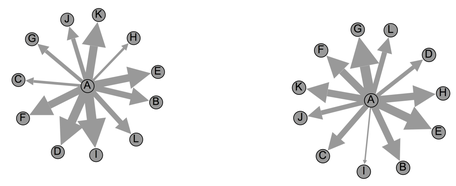


Figure 3: Heterogeneous trust learnt for movies and electronics

Global Trust

Social network analysis to estimate trust using centrality measures of nodes - PageRank, Closeness Centrality

$$\hat{\mathbf{R}}_{u,b} = \frac{\sum_v \mathbf{R}_{v,b}\mathbf{T}_v}{\sum_v \mathbf{T}_v}$$

Latent+Trust

Use SGD to learn trust strengths and latent factors for prediction

$$\min_{\mathbf{S},\mathbf{P},\mathbf{Q}} \sum_{\mathbf{R}_{(u,b)}>0} (\mathbf{R}_{u,b} - \alpha \mathbf{p}_u \mathbf{q}_b - (1-\alpha)\frac{\sum_{v|T_{u,v}=1} \mathbf{S}_{u,v}\mathbf{R}_{v,b}}{\sum_{v|T_{u,v}=1} \mathbf{S}_{u,v}})^2 + \frac{\beta}{2}\left(||\mathbf{P}||_F^2 + ||\mathbf{Q}||_F^2\right)$$

$\mathbf{P}_{U\times K}$ and $\mathbf{Q}_{K\times B}$ - latent factors of users and products, $\mathbf{p}_u$ - u^{th} row of $\mathbf{P}$ and $\mathbf{q}_b$ - b^{th} column of $\mathbf{Q}$.

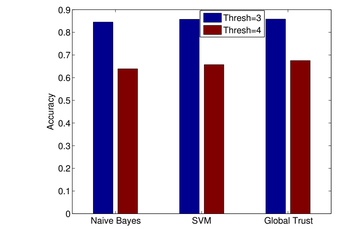
Latent+Heterogeneous Trust

Use SGD to learn heteregeneous trust strengths and latent factors for prediction

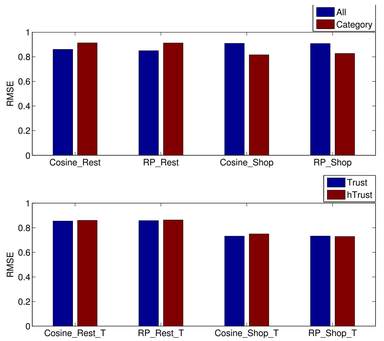
$$\min_{\mathbf{S},\mathbf{P},\mathbf{Q}} \sum_{\mathbf{R}_{(u,b)}>0} (\mathbf{R}_{u,b} - \alpha \mathbf{p}_u \mathbf{q}_b - (1-\alpha)\frac{\sum_l \sum_{v|T_{u,v}=1} \mathbf{S}_{u,v}^{l}\mathbf{R}_{v,b}\mathbf{C}_{b,l}}{\sum_l \sum_{v|T_{u,v}=1} \mathbf{S}_{u,v}^{l}\mathbf{C}_{b,l}})^2 + \frac{\beta}{2}\left(||\mathbf{P}||_F^2 + ||\mathbf{Q}||_F^2\right)$$

$\mathbf{S}_{u,v}^{l}$ - trust strength in category l, $\mathbf{C}_{b,l}$ - whether product b belongs to category l.

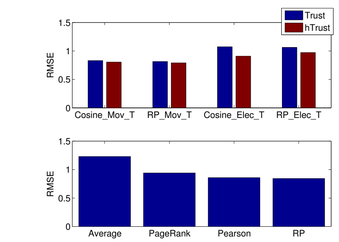
Results



Content based binary classifiers on Yelp treating treating (Thresh,5) as positive labels



Trust Models on Yelp Dataset



Trust models on Epinions Dataset

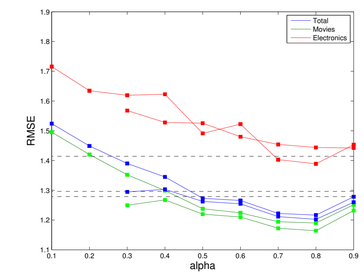


Figure 4: Hybrid Trust Models on Epinions