TheWebConf 2019
May 13-17
San Francisco

# Transfer Meets Hybrid: A Synthetic Approach for Cross-Domain Collaborative Filtering with Text

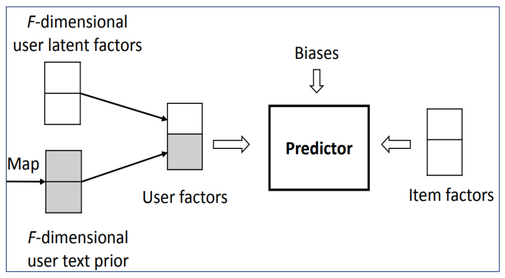
**Guangneng Hu**, **Yu Zhang**, and **Qiang Yang**
Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong
njuhgn@gmail.com, yu.zhang.ust@gmail.com, qyang@cse.ust.hk

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

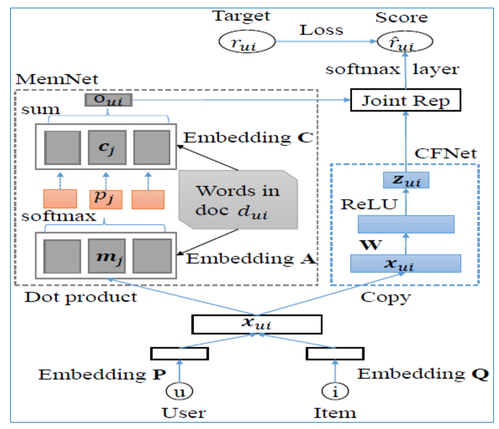
## 1 Motivation

Two threads to alleviate the data sparsity issue in collaborative filtering:

- Hybrid filtering methods integrate the content information, e.g. product reviews and news titles

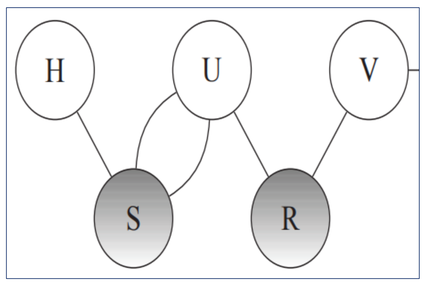



TBPR, Hu et al, PAKDD'17

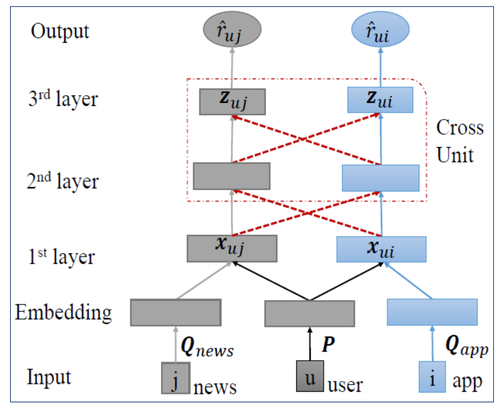



PNE, G. Hu, NAACL'19

- Cross-domain methods leverage the knowledge from a related domain, e.g. from App to News




eSMF, Hu et al, IJCAI'15




CoNet, Hu et al, CIKM'18

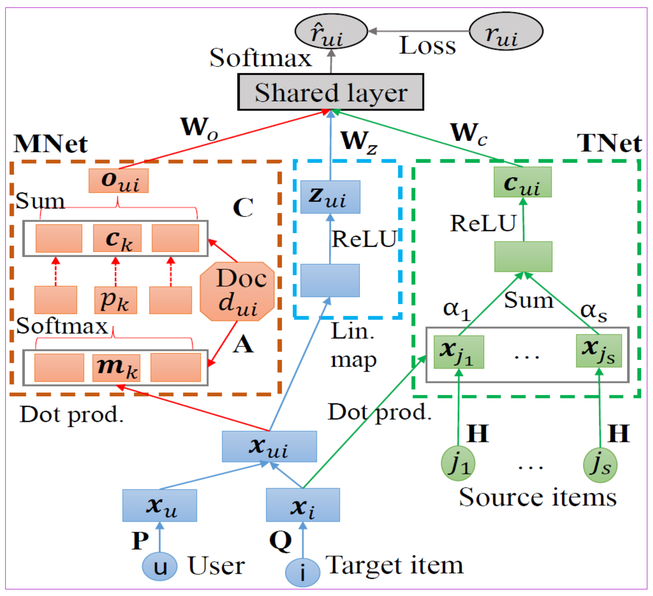
## 2 TMH: Transfer Meets Hybrid

TMH brings an effort on combining two threads to alleviate the data sparsity issues including cold-user and cold-item start.

### Contributions

- MNet (a memory network) can learn high-level representations of unstructured text w.r.t. the given user-item interaction.
- TNet (a transfer network) can selectively transfer knowledge from source items in a coarse-to-fine way.

### Architecture




$$\hat{r}_{ui} = \frac{1}{1 + \exp(-\boldsymbol{h}^T \boldsymbol{y}_{ui})}$$

$$\boldsymbol{y}_{ui} = [\boldsymbol{W}_o \boldsymbol{o}_{ui}, \boldsymbol{W}_z \boldsymbol{z}_{ui}, \boldsymbol{W}_c \boldsymbol{c}_{ui}].$$

$$a_j^{(i)} = \boldsymbol{x}_i^T \boldsymbol{x}_j$$

Quadrant II | Quadrant I | Quadrant III | Quadrant IV

## 3 Datasets and Baselines

| Dataset | Domain | Statistics | Amount |
|---|---|---|---|
| Mobile News | Shared | #Users | 15,890 |
| | Target | #News | 84,802 |
| | | #Reads | 477,685 |
| | | Density | 0.035% |
| | | #Words | 612,839 |
| | | Avg. Words Per News | 7.2 |
| | Source | #Apps | 14,340 |
| | | #Installations | 817,120 |
| | | Density | 0.359% |
| Amazon Men | Shared | #Users | 8,514 |
| | Target | #Clothes (Men) | 28,262 |
| | | #Ratings/#Reviews | 56,050 |
| | | Density | 0.023% |
| | | #Words | 1,845,387 |
| | | Avg. Words Per Review | 32.9 |
| | Source | #Products (Sports) | 41,317 |
| | | #Ratings/#Reviews | 81,924 |
| | | Density | 0.023% |

- Preprocessing:
  - Shared users between two domains
  - Filter users who have too few interactions
  - Filter items which have too few interactions
- Check Python script on my homepage http://home.cse.ust.hk/~ghuac/

| Baselines | Shallow method | Deep method |
|---|---|---|
| Single-domain | BPRMF | MLP |
| Cross-domain | CDCF, CMF | MLP++, CSN |
| Hybrid | HFT, TextBPR | LCMR |
| Cross + Hybrid | CDCF++ | TMH (ours) |

## 4 Results

**Table 2: Results on Amazon data.**

| Method | $topK = 5$ | | | $topK = 10$ | | | $topK = 20$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | HR | NDCG | MRR | HR | NDCG | MRR | HR | NDCG | MRR |
| BPRMF | .0810 | .0583 | .0509 | .1204 | .0710 | .0561 | .1821 | .0864 | .0602 |
| CDCF | .1295 | .0920 | .0797 | .2070 | .1167 | .0897 | .3841 | .1609 | .1015 |
| CMF | .1498 | .0950 | .0771 | .2224 | .1182 | .0863 | .3573 | .1521 | .0957 |
| HFT | .1077 | .0815 | .0729 | .1360 | .0907 | .0767 | .2782 | .1252 | .0854 |
| TextBPR | .1517 | .1208 | .1104 | .1777 | .1291 | .1138 | .2268 | .1414 | .1171 |
| CDCF++ | .1314 | .0926 | .0800 | .2102 | .1177 | .0901 | .3822 | .1605 | .1016 |
| MLP | .2100 | .1486 | .1283 | .2836 | .1697 | .1371 | .3820 | .1899 | .1426 |
| MLP++ | .2263 | .1626 | .1417 | .2992 | .1862 | .1514 | .3810 | .2069 | .1570 |
| CSN | .2340* | .1680* | .1462* | .3018* | .1898* | .1552* | .3944* | .2091* | .1605* |
| LCMR | .2024 | .1451 | .1263 | .2836 | .1678 | .1356 | .3951 | .1918 | .1420 |
| TMH | **.2575** | **.1796** | **.1550** | **.3490** | **.2077** | **.1666** | **.4443** | **.2311** | **.1727** |
| Our improve | 10.04% | 6.90% | 6.01% | 15.63% | 9.43% | 7.34% | 12.65% | 10.52% | 7.60% |