Personalized Neural Embeddings for Collaborative Filtering with Text

Guangneng Hu

Tuesday, 4 June 2019
NAACL-19, Minneapolis
Outline

• Collaborative filtering
  • Matrix factorization & Neural approaches
• Collaborative filtering with text
  • Topic modelling & Word embeddings
• Personalized neural embeddings
• Conclusion
Recommendations: Products, Media, Entertainment, & Partners

- **Amazon**
  - 300 million customers
  - 564 million products

- **Netflix**
  - 480,189 users
  - 17,770 movies

- **Spotify**
  - 40 million songs

- **OkCupid**
  - 10 million members
A Typical CF Approach: Matrix Factorization (MF)  
(Koren KDD’08, KDD 2018 TEST OF TIME)

|       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|
| ✔     | ❓     | ✔     | ❓     | ✔     | ❓     |       |
| ✔     | ❓     | ✔     | ❓     | ❓     | ✔     |       |
| ❓     | ❓     | ✔     | ❓     | ✔     | ❓     |       |
| ❓     | ✔     | ❓     | ✔     | ❓     | ❓     |       |
| ❓     | ✔     | ❓     | ✔     | ❓     | ✔     |       |
| ❓     | ✔     | ❓     | ✔     | ❓     | ✔     | ✔     |

MF, SVD/PMF

\[ \hat{r}_{ui} = P^T_u Q_i \]
A Limitation of MF: As a Single-Layer Linear Neural Network

• **Input**: one-hot encodings of the user and item indices \((u, i)\)
• **Embedding**: embedding matrices \((P, Q)\)
• **Output**: Hadamard product between embeddings with a fixed all-one weight vector \(h\) and an identity activation function

\[
\hat{r}_{u,i} = \sigma(h^T(P_u \odot Q_i))
\]
CF Faces Challenges: Data Sparsity, Long Tail & Unbalanced

• Data sparsity issue
  • Netflix
    • 1.225%
  • Amazon
    • 0.017%

• Long tail & Unbalanced
  • Pareto principle (80/20 rule):
    • A small proportion (e.g., 20%) of products generate a large proportion (e.g., 80%) of sales
A Solution: Collaborative filtering with text

- Item reviews justify user ratings
- Item content reveals topic semantics
Topic Modelling: Hidden Factors & Topics (HFT)

• Using a transform that aligns latent item factors and item topics

\[
\theta_{i,k} = \frac{\exp(\kappa \gamma_{i,k})}{\sum_{k'} \exp(\kappa \gamma_{i,k'})}
\]

\[\gamma_i \in \mathbb{R}^K, \quad \theta_i \in \Delta^K \text{ (i.e., } \sum_k \theta_{i,k} = 1)\]

McAuley & Leskovec, Hidden factors and hidden topics, RecSys'13
Pre-extracted Word-embedding as Features (TBPR)

- Basic MF factorizes ratings into user/item *latent* factors
- Another MF factorizes reviews into user/item *text* factors

\[ f_i \equiv \frac{1}{|d_i|} \sum_{w \in d_i} e_w \]

\[ P_u^T Q_i + \theta_u^T (H f_i) \]

Hu & Dai, Integrating Reviews into Personalized Ranking for Cold Start Recommendation, PAKDD’17
Personalized Neural Embeddings (PNE)

• Inspired by neural CF and entity embeddings
  • PNE jointly learns embeddings of users, items, and words

• PNE estimates the probability that a user will like an item by two terms
  • behavior factors and semantic factors
Behavior Factors: Learning Neural Embeddings of Users & Items

- Recap: MF as a linear NN

\[ \hat{r}_{u,i} = \sigma(h^T(P_u \odot Q_i)) \]

- Learning weights \( h \) instead of fixing it
- Using non-linear activation instead of identity

\[ z_{ui}^{\text{behavior}} = \text{ReLU}(W x_{ui} + b) \]
Semantic Factors: Learning Personalized Word Embeddings

• Personalized word embedding encodes the importance of a word to the given user-item interaction

\[ a_{j}^{u,i} = x_{ui}^T m_{j}^{u,i} \]

\[ z_{ui}^{\text{semantic}} = \sum_{j: w_j \in d_{ui}} \text{Softmax}(a_{j}^{u,i}) c_{j} \]

\[ \sum_{j: w_j \in d_{ui}} \text{Softmax}(a_{j}^{u,i}) c_{j} \]
Jointly Learning Embeddings of Users, Items, & Words

- Sharing user and item embeddings
- Binary cross-entropy loss
Dataset and Baselines

• Datasets
  • Amazon: Product reviews by users
  • Cheetah Mobile: News reading by users

| Dataset     | #user | #item  | #rating | #word    | #density  | avg. words |
|-------------|-------|--------|---------|----------|-----------|------------|
| Amazon      | 8,514 | 28,262 | 56,050  | 1,845,387| 0.023%    | 65.3       |
| Cheetah     | 15,890| 84,802 | 477,685 | 612,839  | 0.035%    | 7.2        |

• Baselines

| Baselines       | Shallow method | Deep method   |
|-----------------|----------------|---------------|
| CF              | BPR            | MLP           |
| CF w/ text      | HFT, TBPR      | LCMR, PNE (ours) |
Evaluation Metrics

• Top-N item recommendation

• Metrics to measure the accuracy of rankings
  • Hit Ratio (HR)
  • Mean Reciprocal Rank (MRR)
  • Normalized Discounted Cumulative Gain (NDCG)

\[
HR = \frac{1}{|U|} \sum_{u \in U} \delta(p_u \leq \text{top} N),
\]

\[
MRR = \frac{1}{|U|} \sum_{u \in U} \frac{1}{p_u}
\]

\[
NDCG = \frac{1}{|U|} \sum_{u \in U} \frac{\log 2}{\log(p_u + 1)},
\]
Comparing Different Approaches: PNE vs Multilayer Perceptron

• Since CFNet of PNE is a neural CF (with one hidden layer), results show the benefit of exploiting unstructured text to alleviate the data sparsity issue faced by pure CF methods

| TopK | Metric | Method | BPR | HFT | TBPR | MLP | LCMR | PNE |
|------|--------|--------|-----|-----|------|-----|------|-----|
| 5    | HR     |        | 8.10| 10.77| 15.17| 21.00*| 20.24| 23.52|
|      | NDCG   | 5.83   | 8.15| 12.08| 14.86*| 14.51| 16.46|
|      | MRR    | 5.09   | 7.29| 11.04| 12.83*| 12.63| 14.13|
| 10   | HR     |        | 12.04| 13.60| 17.77| 28.36*| 28.36*| 31.86|
|      | NDCG   | 7.10   | 9.07| 12.91| 16.97*| 16.78| 19.15|
|      | MRR    | 5.61   | 7.67| 11.38| 13.71*| 13.56| 15.24|
| 20   | HR     |        | 18.21| 27.82| 22.68| 38.20| 39.51*| 42.21|
|      | NDCG   | 8.64   | 12.52| 14.14| 18.99| 19.18*| 21.75|
|      | MRR    | 6.02   | 8.54| 11.71| 14.26*| 14.20| 15.95|
Comparing Different Approaches: PNE vs HFT & TBPR

• Results show the benefit of integrating content text through MemNet (and also exploiting interactions through neural CF)

| TopK | Metric | Method  |
|------|--------|---------|
|      |        | BPR     | HFT    | TBPR   | MLP    | LCMR   | PNE    |
| 5    | HR     | 8.10    | 10.77  | 15.17  | 21.00* | 20.24  | 23.52  |
|      | NDCG   | 5.83    | 8.15   | 12.08  | 14.86* | 14.51  | 16.46  |
|      | MRR    | 5.09    | 7.29   | 11.04  | 12.83* | 12.63  | 14.13  |
| 10   | HR     | 12.04   | 13.60  | 17.77  | 28.36* | 28.36* | 31.86  |
|      | NDCG   | 7.10    | 9.07   | 12.91  | 16.97* | 16.78  | 19.15  |
|      | MRR    | 5.61    | 7.67   | 11.38  | 13.71* | 13.56  | 15.24  |
| 20   | HR     | 18.21   | 27.82  | 22.68  | 38.20  | 39.51* | 42.21  |
|      | NDCG   | 8.64    | 12.52  | 14.14  | 18.99  | 19.18* | 21.75  |
|      | MRR    | 6.02    | 8.54   | 11.71  | 14.26* | 14.20  | 15.95  |
Comparing Different Approaches: PNE vs LCMR

• Since MemNet of PNE is the same with Local MemNet of LCMR (with one-hop), results show the design of CFNet of PNE is more reasonable than that of Centralized MemNet of LCMR

• This also points out the challenge of effectively fusing ratings & text

| TopK | Metric | Method | BPR | HFT | TBPR | MLP | LCMR | PNE |
|------|--------|--------|-----|-----|------|-----|------|-----|
| 5    | HR     |        | 8.10| 10.77| 15.17| 21.00*| 20.24| 23.52|
|      | NDCG   |        | 5.83| 8.15 | 12.08| 14.86*| 14.51| 16.46|
|      | MRR    |        | 5.09| 7.29 | 11.04| 12.83*| 12.63| 14.13|
| 10   | HR     |        | 12.04| 13.60| 17.77| 28.36*| 28.36*| 31.86|
|      | NDCG   |        | 7.10 | 9.07 | 12.91| 16.97*| 16.78| 19.15|
|      | MRR    |        | 5.61 | 7.67 | 11.38| 13.71*| 13.56| 15.24|
| 20   | HR     |        | 18.21| 27.82| 22.68| 38.20| 39.51*| 42.21|
|      | NDCG   |        | 8.64 | 12.52| 14.14| 18.99| 19.18*| 21.75|
|      | MRR    |        | 6.02 | 8.54 | 11.71| 14.26*| 14.20 | 15.95|
PNE Learns Meaningful Word Embeddings

- Nearest neighbors of drug: shot, shoots, gang, murder, killing, rape, stabbed, truck, school, police, teenage
- Google word2vec: drugs, heroin, addiction, abuse, fda, alcoholism, cocaine, lsd, alcohol, schedule, substances

Pre-trained word embeddings [http://home.cse.ust.hk/~ghuac/](http://home.cse.ust.hk/~ghuac/)
Conclusion and Future Works

• Conclusion
  • Behavior interactions can be effectively integrated with unstructured text via jointly learning neural embeddings of users, items, and words

• Future works
  • User privacy
    • A user does not want to share the raw data with others
    • General data privacy regulatory (GDPR) and Federated learning
Thanks!

Q & A

Acknowledge: NAACL travel grant