TrNews: Heterogeneous User-Interest Transfer Learning for News Recommendation

Guangneng Hu & Qiang Yang

April 21 (Wed), 2021
News feeds online platforms

Microsoft news

Google news

Toutiao

personalized modules on platforms

retrieved on March 27
Recommender systems: engine & sources

- **User**
- **Items of Interest**

**Recommender engine**
- Learn from:
  - rating data
  - item info
  - auxiliary data
  - E.g., user gender, age, occupation
  - E.g., item tag, category, texts
  - E.g., cross domains

**Data sources**
- Item corpus
  - Millions
  - Tens

**Recommender engine**
- Items of Interest
  - User
Deep learning meets recommender systems

- Learning user and item representations (recsys) by representation learning (deep learning)
Deep learning with small data

Extensive rating data in **English domain**

Limited rating data in **Chinese domain**

knowledge transfer
Deep knowledge transfer

• **Base network**
  • Answering how to model the individual domain

• **Transfer unit**
  • Answering “what to transfer” across domains

• **Joint learning**
  • Training base networks and transfer unit in end-to-end way
Feature representation transfer: strategies

Identity transfer: No adaptation (Pan et al, AAAI’10) pre-trained features: No fine-tuning

Linear transfer (Liu et al, AAAI’18; Li & Tuzhilin, WSDM’20)

Nonlinear transfer: Dilated MLP (Man et al, IJCAI’17)
Challenges

• Heterogeneity
  • Different word distributions and feature space across domains

• Effective neural architecture
  • Avoiding negative transfer
  • Reducing risks of overfitting
Proposed TrNews

- The key: Translator
- Translator: mapping user reps from source to target domain
  - attack heterogeneous feature space
- Translator: bottleneck architecture
  - deep layers
  - reduce overfitting risks
  - compress feature representations
Methodology

• Training phase: learn translator

Step 1: mapping  \( \tilde{\phi}_S(u) = \text{De} \left( \text{En} \left( \phi_S(u) \right) \right) \)

Step 2: supervised loss  \( \| \phi_T(u) - \tilde{\phi}_S(u) \|_2^2 \)

• Inference phase: use translator

\( \phi_T(u^*) \triangleq \text{De} \left( \text{En} \left( \phi_S(u^*) \right) \right) \)

where \( u^* \) is cold start user in target domain
Datasets

• Cheetah Mobile (geolocations: New York, Florida, Texas, California)
  • News reading logs of users collected in January of 2017 with
  • Top two categories (political & daily) used as the cross corpora
  • mean length of news articles = 12 words
  • mean length of user history = 45 articles

• Microsoft MIND (small version)
  • The clicked historical news articles are positive examples
  • Top two categories (news & sports) of news used as the cross corpora
  • mean length of news articles = 40 words
  • mean length of user history = 13 articles
Experiments with SOTA baselines

- POP is competitive
- Neural methods are better than shallow LR (logistic regression)
- TrNews is the best with large improvement

| MIND    | HR@5 | HR@10 | NDCG@5 | NDCG@10 | MRR  | AUC  |
|---------|------|-------|--------|---------|------|------|
| POP     | 84.18| 92.80 | 69.61  | 72.43   | 66.33| 95.13|
| LR      | 92.69*| 96.66*| 85.81  | 87.11*  | 84.33*| 97.92*|
| TANR    | 89.94| 95.34 | 89.94* | 83.38   | 79.84| 97.86|
| DeepFM  | 89.16| 94.78 | 79.36  | 81.19   | 77.12| 97.63|
| DIN     | 89.28| 94.88 | 80.16  | 82.03   | 78.22| 97.63|
| TrNews  | 97.36| 99.02 | 94.16  | 94.74   | 93.45| 99.47|
Experiments with transfer baselines

- Adaptation is better than No adaptation
- Nonlinear transfer may worse than linear transfer
- Bottleneck architecture is superior than dilated one
Visualization & explanation

- **Short-term** user interests: latest three matter the most
- **Long-term** user interests: oldest two still impact
- Filter out *irrelevant* articles from history

| Historical news                                                                 | Attn. weight |
|--------------------------------------------------------------------------------|--------------|
| hillary clinton makes a low-key return to washington                            | 0.04         |
| the hidden message in obama’s ‘farewell’ speech                                 | 0.12*        |
| here’s why sasha obama skipped the farewell address                             | 0.00         |
| donald trump’s ‘prostitute scandal’ was filmed by cameras and recorded with microphones hidden behind the walls | 0.00         |
| white house official explains sasha obama’s absence at father’s farewell speech | 0.00         |
| irish bookie puts odds on trump’s administration, inauguration and impeachment | 0.00         |
| heads are finally beginning to roll at the clinton foundation                  | 0.00         |
| donald trump’s incoming administration considering white house without press corps | 0.76         |
| donald trump says merkel made ‘big mistake’ on migrants                         | 0.05         |
| controversial clinton global initiative closing its doors for good             | 0.00         |
| army chief gen. bipin rawat talks about equal responsibility for women in the frontlines. we couldn’t agree more | N/A          |
Thanks!

Q & A