Collaborative Memory Network for Recommendation Systems

Travis Ebesu\textsuperscript{1} \quad Bin Shen\textsuperscript{2} \quad Yi Fang\textsuperscript{1}

\textsuperscript{1}Santa Clara University
California, USA

\textsuperscript{2}Google
New York, USA

SIGIR 2018
Motivations

Collaborative Filtering (CF) Classes
- Latent Factor-based
- Memory/Neighborhood-based
- Hybrid Models

Deep Learning-based methods can model nonlinear user-item relations however, they ignore neighborhood information.

We propose a deep architecture to unify the 2 classes of CF.
Motivations

- Deep Learning based
- Latent Factor based
- Neighborhood based

Our Work
Contributions

- We propose Collaborative Memory Network (CMN), leveraging a memory network and neural attention mechanism.
- We reveal the connection between CMN with existing collaborative filtering methods and memory networks.
Memory Networks

Memory-based architectures generally consist of two components:

- Memory Module: stores long-term knowledge representations
- Controller Module: manipulates the memory (e.g. read/write operations)
Collaborative Memory Network (CMN)

Model Overview

Three core components:
- User Embedding
- Neighborhood Attention
- Output Module
User $u$ is embedded into a $d$ dimensional memory slot $m_u$ and similarly item $i$ to $e_i$

\[ q_{ui} = m_u^T m_v + e_i^T m_v \quad \forall \ v \in N(i) \]

where $N(i)$ represents the set of all users (neighborhood) who rated item $i$. 
Neighborhood Attention

The neural attention mechanism weights each user’s contribution to the neighborhood:

\[
p_{uiv} = \frac{\exp(q_{uiv})}{\sum_{k \in N(i)} \exp(q_{uik})} \quad \forall \ v \in N(i)
\]

\[
o_{ui} = \sum_{v \in N(i)} p_{uiv} c_v
\]
Output Module

For a given user $u$ and item $i$ the ranking score is:

$$ \hat{r}_{ui} = v^T \phi \left( U(m_u \odot e_i) + Wo_{ui} + b \right) $$

where $\odot$ is the elementwise product and $\phi(x) = \max(0, x)$.
Multiple Hops
User $u$ prefers the observed item $i^+$ over the unobserved or negative item $i^-$ forming triplet preferences $(u, i^+, i^-)$. 

$$
\mathcal{L} = - \sum_{(u, i^+, i^-)} \log \sigma(\hat{r}_{ui^+} - \hat{r}_{ui^-})
$$

where $\sigma(x) = 1/(1 + \exp(-x))$
Collaborative Memory Network (CMN)

Relation to Existing Models

Latent Factor Model

\[ \hat{r}_{ui} = v^T \phi(U(m_u \odot e_i) + Wo_{ui} + b) \]
Relation to Existing Models

Latent Factor Model

CMN
\[ \hat{r}_{ui} = v^T \phi(U(m_u \odot e_i) + Wo_{ui} + b) \]

GMF
\[ \hat{r}_{ui} = v^T \phi(m_u \odot e_i) \]
Relation to Existing Models

Latent Factor Model

\[
\hat{r}_{ui} = v^T \phi \left( U \left( m_u \odot e_i \right) + W_{ou} + b \right)
\]

CMN

\[
\hat{r}_{ui} = v^T \phi \left( m_u \odot e_i \right)
\]

GMF

\[
\hat{r}_{ui} = 1^T \left( m_u \odot e_i \right)
\]

MF
Relation to Existing Models

Neighborhood-based Similarity Models

Neighborhood-based similarity models estimate a user-user (item-item) similarity matrix $S$.

$$\hat{r}_{ui} = v^T \phi(U(m_u \odot e_i) + W \sum_{v \in N(i)} p_{uiv} c_v + b)$$

$$\hat{r}_{ui} = \alpha \sum_{v \in N(i)} S_{uv}$$
Relation to Existing Models

Hybrid Models

Hybrid models consist of two general terms: user-item latent factor interaction and a neighborhood component.

CMN

\[ \hat{r}_{ui} = v^T \phi \left( U ( m_u \odot e_i ) + W o_{ui} + b \right) \]

Hybrid CF

\[ \hat{r}_{ui} = v^T \phi \left( m_u \odot e_i + \sum_{v \in N(i)} p_{uv} c_v \right) \]

*Latent Factors*

*Neighborhood*
### Experimental Results

#### Experimental Settings

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Ratings</th>
<th>Users</th>
<th>Items</th>
<th>Sparsity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epinions</td>
<td>664,823</td>
<td>40,163</td>
<td>139,738</td>
<td>99.98%</td>
</tr>
<tr>
<td>citeulike-a</td>
<td>204,987</td>
<td>5,551</td>
<td>16,980</td>
<td>99.78%</td>
</tr>
<tr>
<td>Pinterest</td>
<td>1,500,809</td>
<td>55,187</td>
<td>9,916</td>
<td>99.73%</td>
</tr>
</tbody>
</table>

- **Top-$N$ Evaluation Metrics**
  - **Hit Rate (HR):** measures presence of the ground truth item in the ranked list.
  - **NDCG:** considers the relative ranking of ground truth item in the ranked list.
Baseline Comparison

**Neighborhood-based**
- Item-KNN (KNN)
- Factored Item Similarity Model (FISM)

**Latent Factor-based**
- Bayesian Personalized Ranking (BPR)

**Hybrid Collaborative Filtering**
- SVD++

**Deep Learning-based**
- Generalized Matrix Factorization (GMF)
- Collaborative Denoising Auto Encoder (CDAE)
- Neural Matrix Factorization (NeuMF)
### Table: Experimental results for different methods on the *Epinions* dataset. Best results highlighted in bold. † indicates the improvement over baselines is statistically significant on a paired t-test ($p < 0.01$).
Experimental Results

Baseline Comparison - citeulike-a dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>HR@5</th>
<th>HR@10</th>
<th>NDCG@5</th>
<th>NDCG@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN</td>
<td>0.6990</td>
<td>0.7348</td>
<td>0.5789</td>
<td>0.5909</td>
</tr>
<tr>
<td>FISM</td>
<td>0.6727</td>
<td>0.8072</td>
<td>0.5106</td>
<td>0.5545</td>
</tr>
<tr>
<td>BPR</td>
<td>0.6547</td>
<td>0.8083</td>
<td>0.4858</td>
<td>0.5357</td>
</tr>
<tr>
<td>SVD++</td>
<td>0.6952</td>
<td>0.8199</td>
<td>0.5244</td>
<td>0.5649</td>
</tr>
<tr>
<td>GMF</td>
<td>0.7271</td>
<td>0.8326</td>
<td>0.5689</td>
<td>0.6034</td>
</tr>
<tr>
<td>CDAE</td>
<td>0.6799</td>
<td>0.8103</td>
<td>0.5106</td>
<td>0.5532</td>
</tr>
<tr>
<td>NeuMF</td>
<td>0.7629</td>
<td>0.8647</td>
<td>0.5985</td>
<td>0.6316</td>
</tr>
<tr>
<td>CMN-1</td>
<td>0.6692</td>
<td>0.7809</td>
<td>0.5213</td>
<td>0.5575</td>
</tr>
<tr>
<td>CMN-2</td>
<td>0.7959†</td>
<td>0.8921†</td>
<td>0.6185†</td>
<td>0.6500†</td>
</tr>
<tr>
<td>CMN-3</td>
<td>0.7932†</td>
<td>0.8901†</td>
<td>0.6234†</td>
<td>0.6551†</td>
</tr>
</tbody>
</table>

Table: Experimental results for different methods on the citeulike-a dataset. Best results highlighted in bold. † indicates the improvement over baselines is statistically significant on a paired t-test ($p < 0.01$).
## Baseline Comparison - *Pinterest* dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>HR@5</th>
<th>HR@10</th>
<th>NDCG@5</th>
<th>NDCG@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN</td>
<td>0.5738</td>
<td>0.8376</td>
<td>0.3450</td>
<td>0.4310</td>
</tr>
<tr>
<td>FISM</td>
<td>0.6783</td>
<td>0.8654</td>
<td>0.4658</td>
<td>0.5268</td>
</tr>
<tr>
<td>BPR</td>
<td>0.6936</td>
<td>0.8674</td>
<td>0.4912</td>
<td>0.5479</td>
</tr>
<tr>
<td>SVD++</td>
<td>0.6951</td>
<td>0.8684</td>
<td>0.4796</td>
<td>0.5362</td>
</tr>
<tr>
<td>GMF</td>
<td>0.6726</td>
<td>0.8505</td>
<td>0.4737</td>
<td>0.5316</td>
</tr>
<tr>
<td>CDAE</td>
<td>0.7008</td>
<td>0.8722</td>
<td>0.4966</td>
<td>0.5525</td>
</tr>
<tr>
<td>NeuMF</td>
<td>0.7041</td>
<td>0.8732</td>
<td>0.4978</td>
<td>0.5530</td>
</tr>
<tr>
<td>CMN-1</td>
<td>0.6984</td>
<td>0.8662</td>
<td>0.4960</td>
<td>0.5507</td>
</tr>
<tr>
<td>CMN-2</td>
<td>0.7267†</td>
<td>0.8904†</td>
<td>0.5180†</td>
<td>0.5714†</td>
</tr>
<tr>
<td>CMN-3</td>
<td>0.7277†</td>
<td>0.8931†</td>
<td>0.5175†</td>
<td>0.5715†</td>
</tr>
</tbody>
</table>

**Table:** Experimental results for different methods on the *Pinterest* dataset. Best results highlighted in bold. † indicates the improvement over baselines is statistically significant on a paired $t$-test ($p < 0.01$).
Embedding Size

(a) *Epinions* dataset  
(b) *citeulike-a* dataset  
(c) *Pinterest* dataset

Figure: Experimental results for CMN varying the embedding size from 20-100 and hops from 1-4.
## Experimental Results

### Effects of Attention & Nonlinearity

<table>
<thead>
<tr>
<th></th>
<th>Epinions</th>
<th>citeulike-a</th>
<th>Pinterest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR@10</td>
<td>NDCG@10</td>
<td>HR@10</td>
</tr>
<tr>
<td>CMN</td>
<td>0.7007</td>
<td>0.5045</td>
<td>0.8921</td>
</tr>
<tr>
<td>CMN-Attn</td>
<td>0.6948</td>
<td>0.4809</td>
<td>0.8589</td>
</tr>
<tr>
<td>CMN-Linear</td>
<td>0.6977</td>
<td>0.4992</td>
<td>0.8665</td>
</tr>
<tr>
<td>CMN-Linear-Attn</td>
<td>0.6937</td>
<td>0.4816</td>
<td>0.8676</td>
</tr>
</tbody>
</table>

**Table:** CMN variants without attention (CMN-Attn); linear activation with attention (CMN-Linear); and linear without attention (CMN-Linear-Attn).
Negative Sampling

(a) *Epinions* dataset
(b) *citeulike-a* dataset
(c) *Pinterest* dataset

**Figure**: Experimental results for CMN varying the number of negative samples from 2-10 and hops from 1-3.
Experimental Results

Attention Visualization

Figure: Heatmap of the attention weights over four hops. The color scale indicates the intensities of the weights, darker representing a higher weight and lighter a lower weight. Each column represents a user in the neighborhood.

(a) *Epinions* dataset

(b) *citeulike-a* dataset

(c) *Pinterest* dataset
Conclusion and Future Work

- We introduced a novel hybrid architecture augmented with a memory module and neural attention mechanism.
- We showed the strong connections between CMN and the classes of collaborative filtering methods.

In future work, we plan to

- address dialogue systems
- integrate content and contextual information
- perform adversarial training
Thank you ACM SIGIR for the generous travel grant.
Questions?
tebesu@scu.edu

Code Online:
http://github.com/tebesu/CollaborativeMemoryNetwork