Collaborative Memory Network for Recommendation Systems

Travis Ebesu[1] Bin Shen[2] Yi Fang[1]

[1]Santa Clara University
California, USA

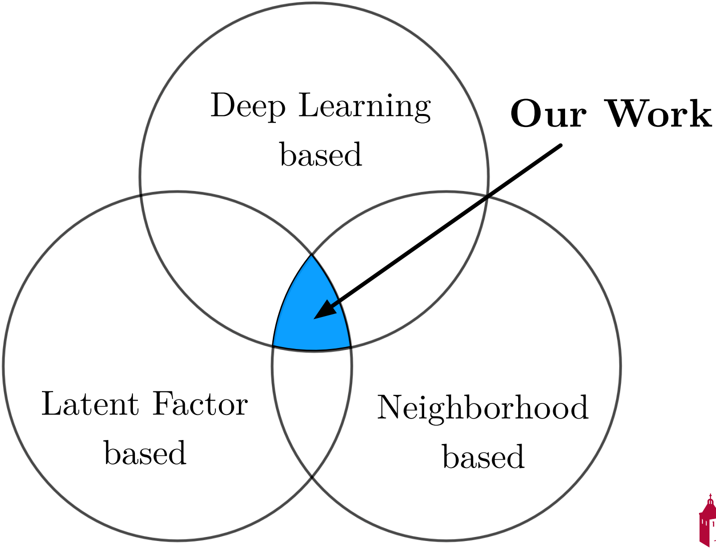
[2]Google
New York, USA

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



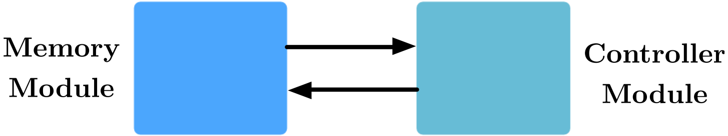
Contributions

- 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)

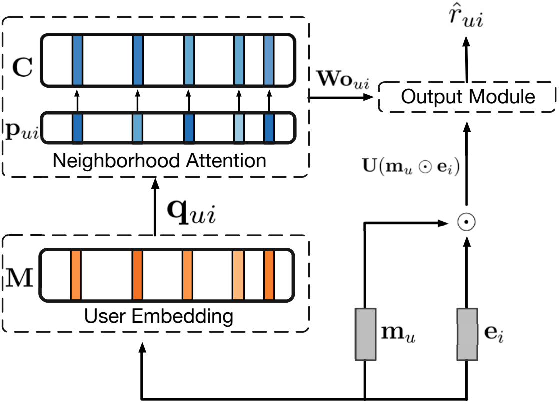


Collaborative Memory Network (CMN)

Model Overview

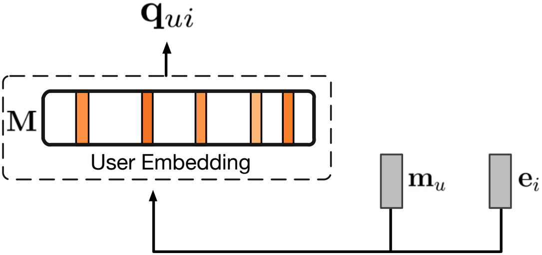
Three core components:

- User Embedding
- Neighborhood Attention
- Output Module



User Embedding

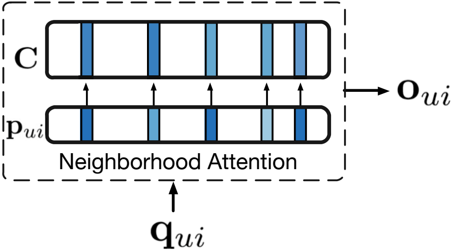
User u is embedded into a d dimensional memory slot $\mathbf{m}_u$ and similarly item i to $\mathbf{e}_i$



$$q_{uiv} = \mathbf{m}_u^\mathsf{T}\mathbf{m}_v + \mathbf{e}_i^\mathsf{T}\mathbf{m}_v \quad \forall\ v \in N(i)$$

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

Neighborhood Attention

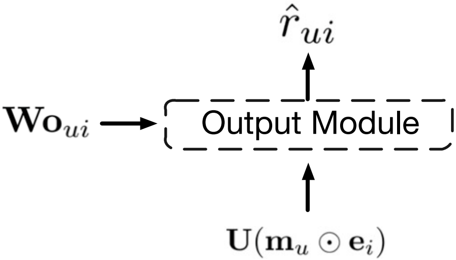


The neural attention mechanism weights each user's contribution to the neighborhood:

$$p_{uiv} = \frac{\exp(q_{uiv})}{\sum_{k \in N(i)} \exp(q_{uik})} \quad \forall\ v \in N(i) \qquad \mathbf{o}_{ui} = \sum_{v \in N(i)} p_{uiv} \mathbf{c}_v$$

Output Module

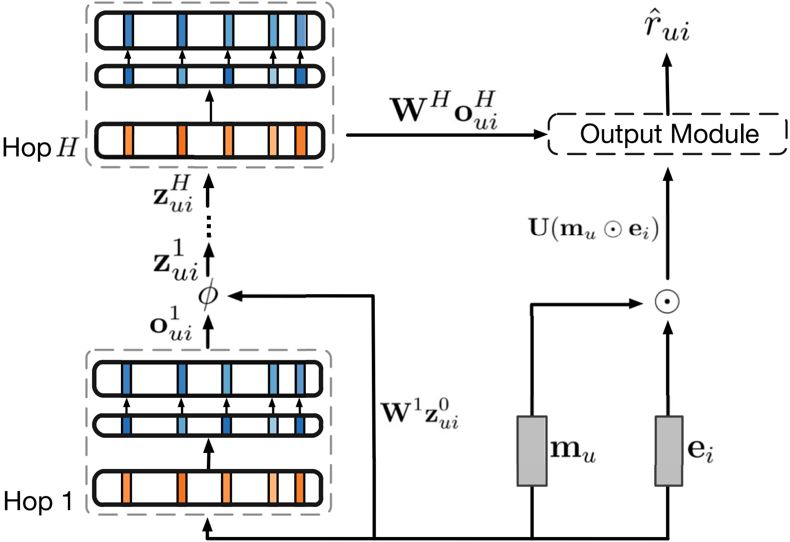
For a given user u and item i the ranking score is:



$$\hat{r}_{ui} = \mathbf{v}^{\mathsf{T}} \phi\big(\mathbf{U}(\mathbf{m}_u \odot \mathbf{e}_i) + \mathbf{Wo}_{ui} + \mathbf{b}\big)$$

where $\odot$ is the elementwise product and $\phi(x) = \max(0, x)$

Multiple Hops



Parameter Estimation

User u prefers the observed item i^+ over the unobserved or negative item i^- forming triplet preferences (u, i^+, i^-).

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

where $\sigma(x) = 1/\big(1 + \exp(-x)\big)$

Santa Clara University

Relation to Existing Models

Latent Factor Model

CMN $$\hat{r}_{ui} = \mathbf{v}^{\mathsf{T}} \phi\big(\mathbf{U}(\boxed{\mathbf{m}_u \odot \mathbf{e}_i}) + \mathbf{W}\mathbf{o}_{ui} + \mathbf{b}\big)$$

Santa Clara University

Relation to Existing Models

Latent Factor Model

CMN $\hat{r}_{ui} = \mathbf{v}^{\mathsf{T}} \phi\big(\mathbf{U}(\boxed{\mathbf{m}_u \odot \mathbf{e}_i}) + \mathbf{W}\mathbf{o}_{ui} + \mathbf{b}\big)$

GMF $\hat{r}_{ui} = \mathbf{v}^{\mathsf{T}} \phi\big(\mathbf{m}_u \odot \mathbf{e}_i\big)$

Relation to Existing Models

Latent Factor Model

CMN $\hat{r}_{ui} = \mathbf{v}^\top \phi(\mathbf{U}(\boxed{\mathbf{m}_u \odot \mathbf{e}_i}) + \mathbf{W}\mathbf{o}_{ui} + \mathbf{b})$

GMF $\hat{r}_{ui} = \mathbf{v}^\top \phi(\mathbf{m}_u \odot \mathbf{e}_i)$

MF $\hat{r}_{ui} = \mathbb{1}^\top (\mathbf{m}_u \odot \mathbf{e}_i)$

Relation to Existing Models

Neighborhood-based Similarity Models

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

CMN $$\hat{r}_{ui} = \mathbf{v}^{\mathsf{T}}\phi\big(\mathbf{U}(\mathbf{m}_u \odot \mathbf{e}_i) + \mathbf{W}\boxed{\sum_{v \in N(i)} p_{uiv}\mathbf{c}_v} + \mathbf{b}\big)$$

Neighborhood $$\hat{r}_{ui} = \alpha \sum_{v \in N(i)} \mathbf{S}_{uv}$$

Relation to Existing Models

Hybrid Models

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

CMN $$\hat{r}_{ui} = \mathbf{v}^{\mathsf{T}}\phi\left(\mathbf{U}(\boxed{\mathbf{m}_u \odot \mathbf{e}_i}) + \mathbf{W}\boxed{\mathbf{o}_{ui}} + \mathbf{b}\right)$$

Hybrid CF $$\hat{r}_{ui} = \mathbf{v}^{\mathsf{T}}\phi\left(\overbrace{\mathbf{m}_u \odot \mathbf{e}_i}^{\text{Latent Factors}} + \underbrace{\sum_{v \in N(i)} p_{uiv}\mathbf{c}_v}_{\text{Neighborhood}}\right)$$

Experimental Settings

Dataset	Ratings	Users	Items	Sparsity
Epinions	664,823	40,163	139,738	99.98%
citeulike-a	204,987	5,551	16,980	99.78%
Pinterest	1,500,809	55,187	9,916	99.73%

- 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)

Baseline Comparison - *Epinions* dataset

	HR@5	HR@10	NDCG@5	NDCG@10
KNN	0.1549	0.1555	0.1433	0.1435
FISM	0.5542	0.6717	0.4192	0.4573
BPR	0.5584	0.6659	0.4334	0.4683
SVD++	0.5628	0.6754	0.4112	0.4477
GMF	0.5365	0.6562	0.4015	0.4404
CDAE	0.5666	0.6844	0.4333	0.4715
NeuMF	0.5500	0.6660	0.4214	0.4590
CMN-1	0.5962	0.6943	0.4684	0.5003
CMN-2	0.6017†	**0.7007**†	0.4724†	0.5045†
CMN-3	**0.6020**†	0.6985†	**0.4748**†	**0.5062**†

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 ($p < 0.01$).

Santa Clara University

Baseline Comparison - *citeulike-a* dataset

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

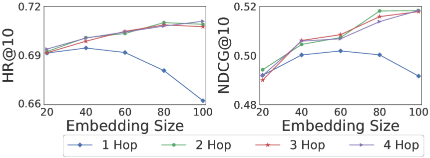
Table: Experimental results for different methods on the *citeulike-a* dataset. Best results highlighted in bold. † indicates the improvement over baselines is statistically significant on a paired *t*-test ($p < 0.01$).

Baseline Comparison - *Pinterest* dataset

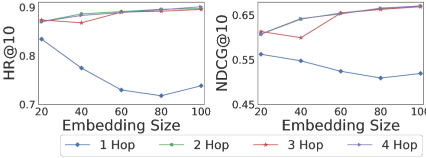
	HR@5	HR@10	NDCG@5	NDCG@10
KNN	0.5738	0.8376	0.3450	0.4310
FISM	0.6783	0.8654	0.4658	0.5268
BPR	0.6936	0.8674	0.4912	0.5479
SVD++	0.6951	0.8684	0.4796	0.5362
GMF	0.6726	0.8505	0.4737	0.5316
CDAE	0.7008	0.8722	0.4966	0.5525
NeuMF	0.7041	0.8732	0.4978	0.5530
CMN-1	0.6984	0.8662	0.4960	0.5507
CMN-2	0.7267†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

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 ($p < 0.01$).

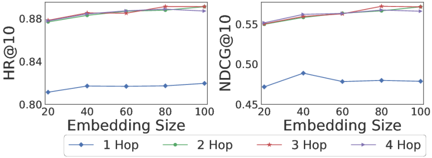
Embedding Size



(a) *Epinions* dataset



(b) *citeulike-a* dataset



(c) *Pinterest* dataset

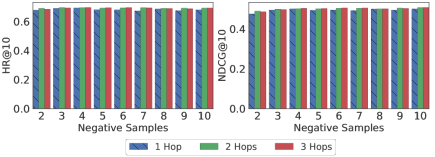
Figure: Experimental results for CMN varying the embedding size from 20-100 and hops from 1-4.

Effects of Attention & Nonlinearity

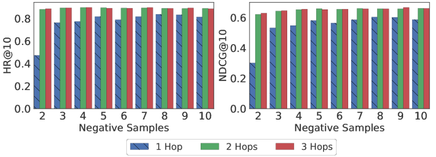
	Epinions		*citeulike-a*		*Pinterest*	
	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
CMN	0.7007	0.5045	0.8921	0.6500	0.8904	0.5714
CMN-Attn	0.6948	0.4809	0.8589	0.5887	0.8773	0.5530
CMN-Linear	0.6977	0.4992	0.8665	0.6282	0.8777	0.5534
CMN-Linear-Attn	0.6937	0.4816	0.8676	0.6256	0.8775	0.5527

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

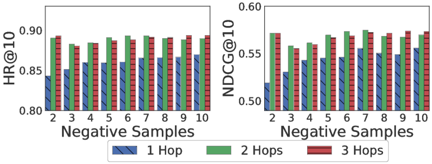
Negative Sampling



(a) *Epinions* dataset



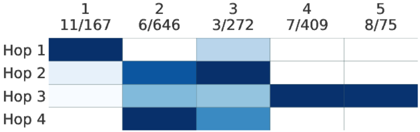
(b) *citeulike-a* dataset



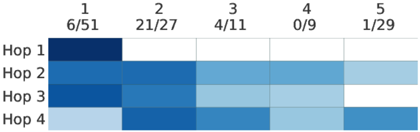
(c) *Pinterest* dataset

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

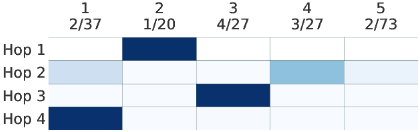
Attention Visualization



(a) *Epinions* dataset



(b) *citeulike-a* dataset



(c) *Pinterest* dataset

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.

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

Thank you ACM SIGIR for the generous travel grant.

Santa Clara University

Questions

Questions?

tebesu@scu.edu

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