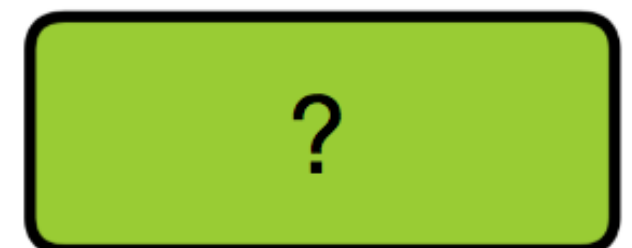


Next Item Recommendation w/ Self-Attention

Shuai Zhang, Lina Yao, Yi Tay, Aixin Sun

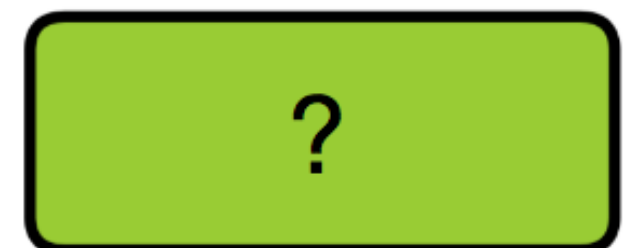
Next Item Recommendation (Sequence-aware)

- **Yehuda Koren**. 2009. Collaborative filtering with temporal dynamics.
- **Rendle et al**. 2010. Factorizing Personalized Markov Chains for Next-basket Recommendation.
- **Feng et al**. 2015. Personalized Ranking Metric Embedding for Next New POI Recommendation.
- **Wu et al**. 2017. Recurrent Recommender Networks.

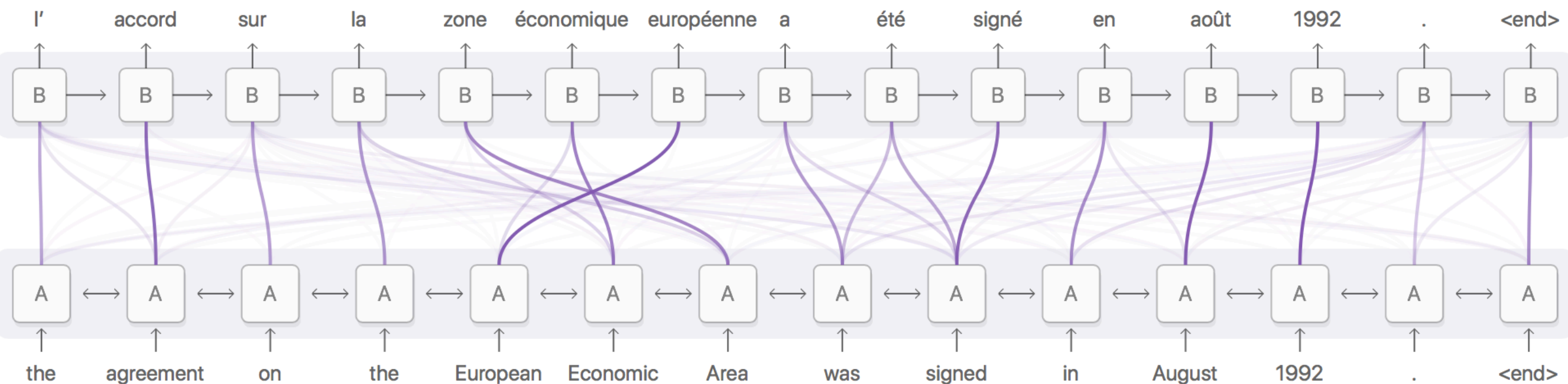


Next Item Recommendation (Deep Learning)

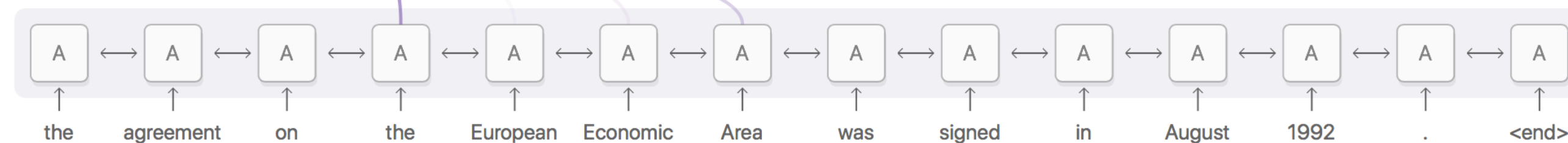
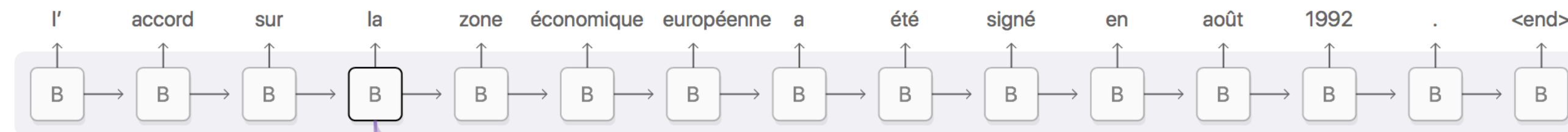
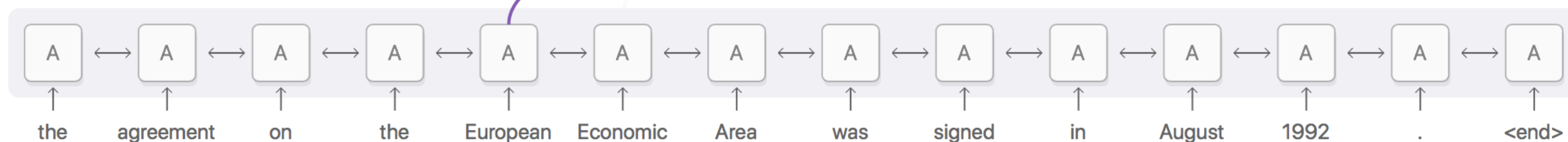
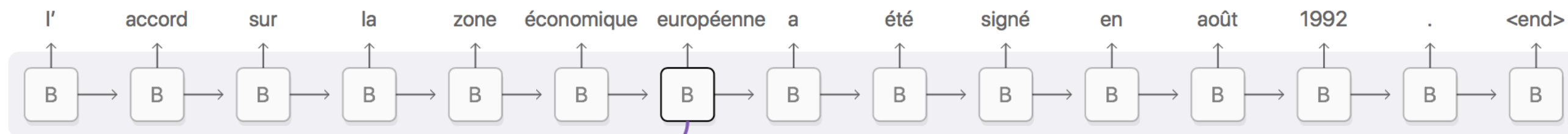
- MLPs, ConvNets, Auto Encoders, RNNs, Attention.
- **Wang et al.** 2015. Learning Hierarchical Representation Model for Next Basket Recommendation.
- **Tang et al.** 2018. Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding
- **Yuyun Gong and Qi Zhang.** 2016. Hashtag Recommendation Using Attention-based Convolutional Neural Network



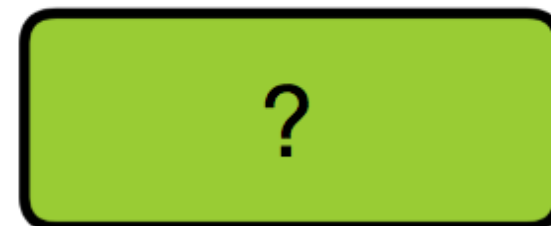
Attention



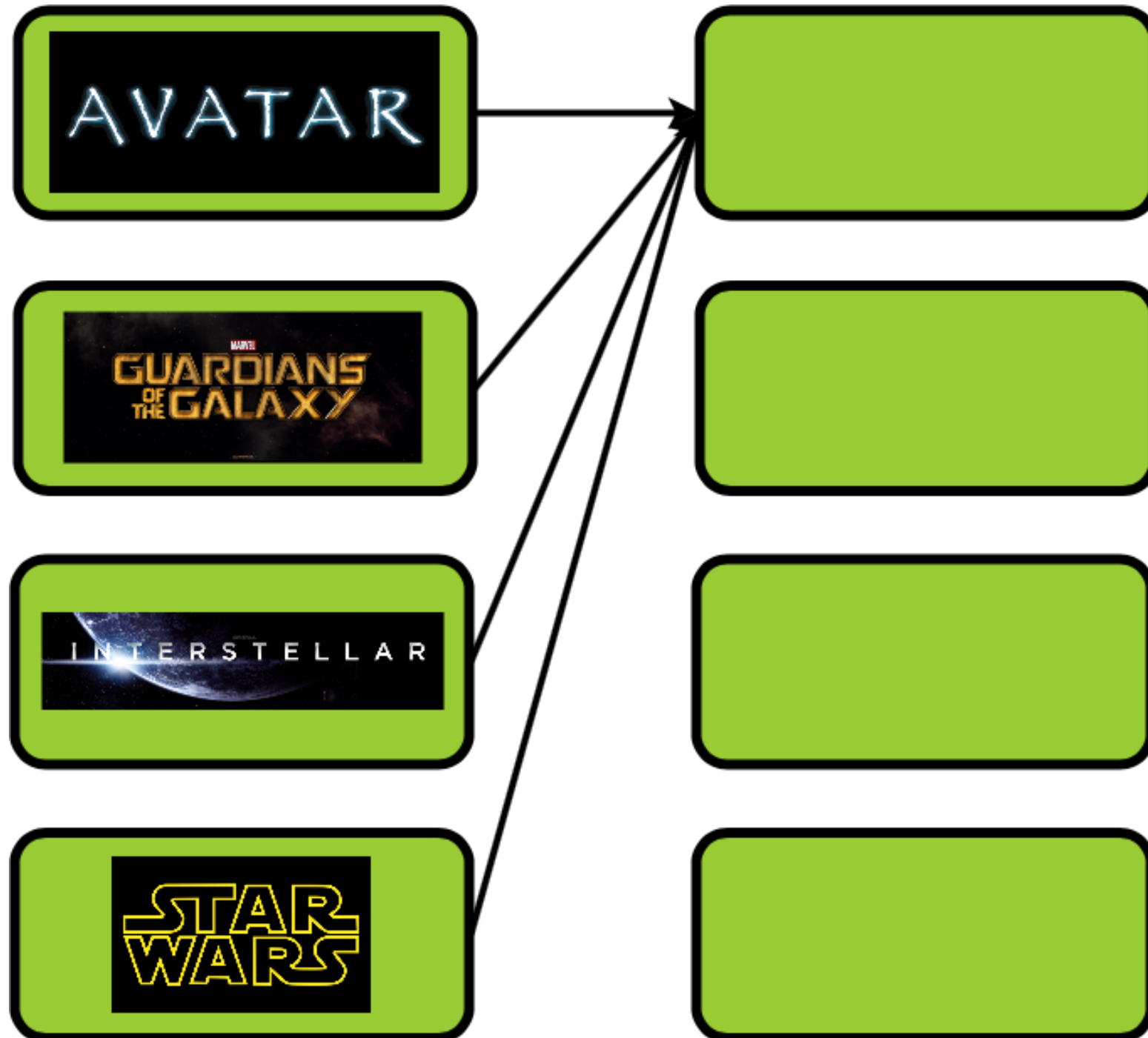
Attention



Self-Attention



Self-Attention



Self-Attention



Self-Attention

The FBI is chasing a criminal on the run .

The FBI is chasing a criminal on the run .

The FBI is chasing a criminal on the run .

The FBI is chasing a criminal on the run .

The FBI is chasing a criminal on the run .

The FBI is chasing a criminal on the run .

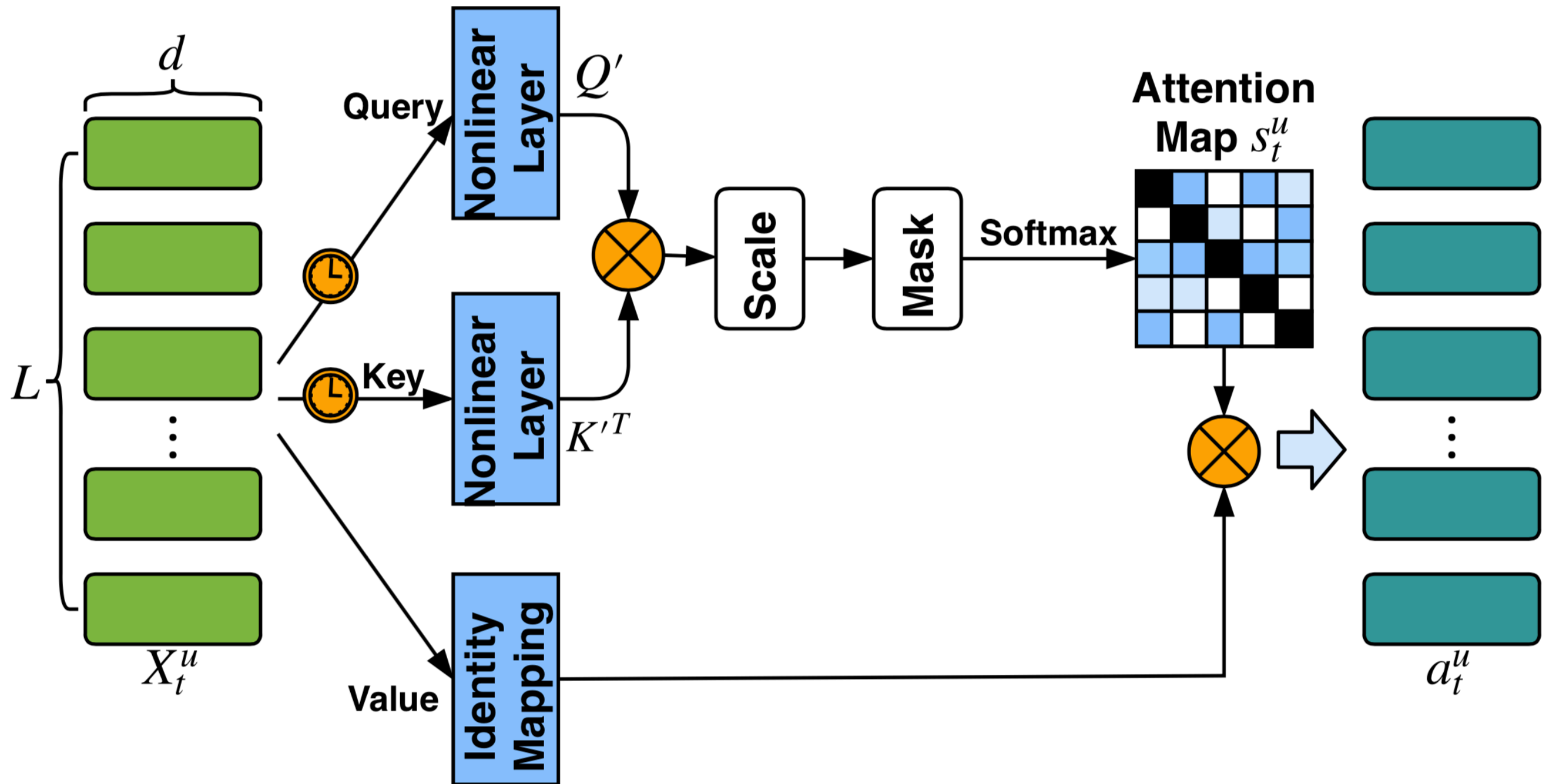
The FBI is chasing a criminal on the run .

The FBI is chasing a criminal on the run .

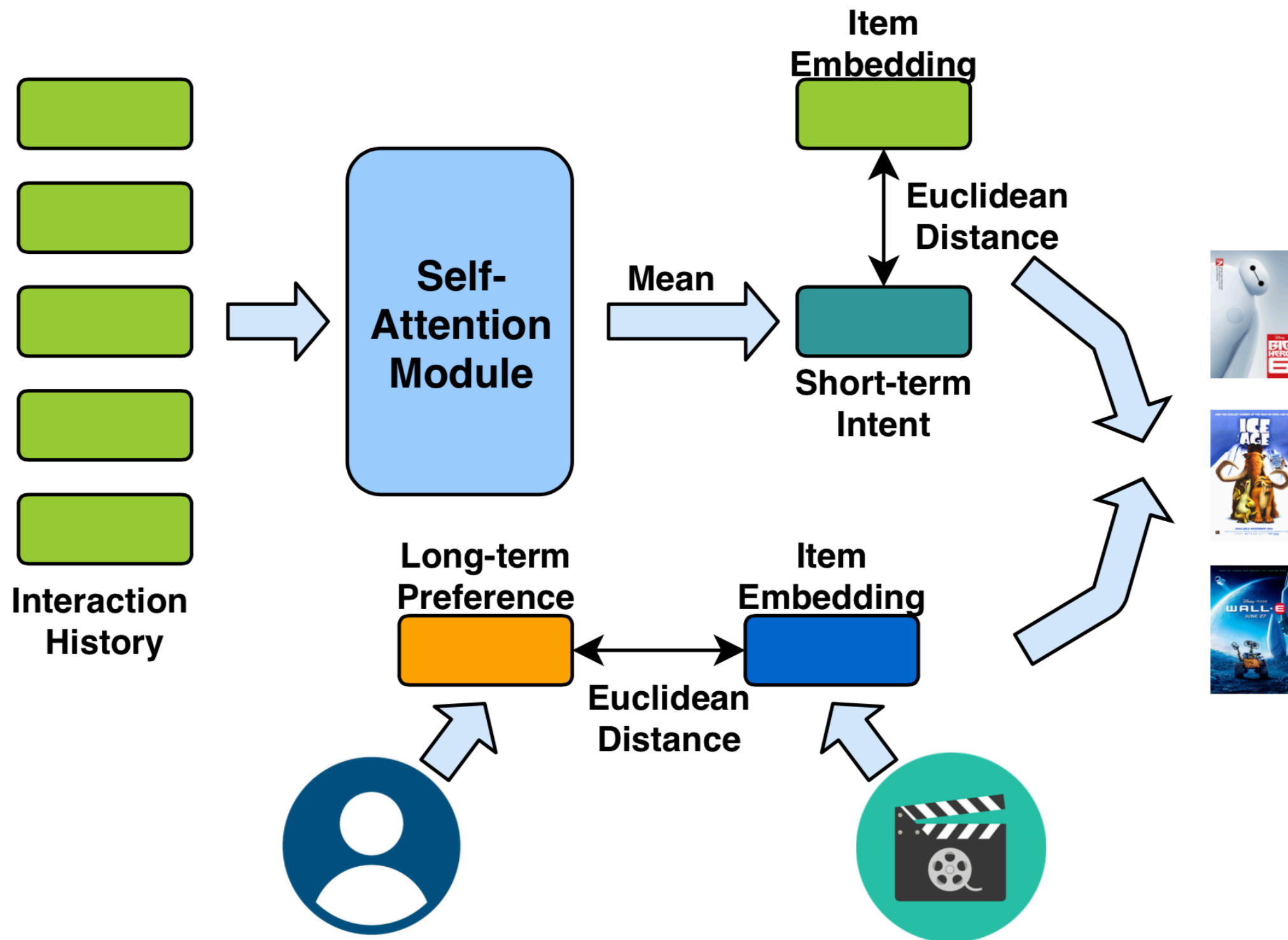
The FBI is chasing a criminal on the run .

The FBI is chasing a criminal on the run .

Self-Attention



Full Model



Evaluation Metrics

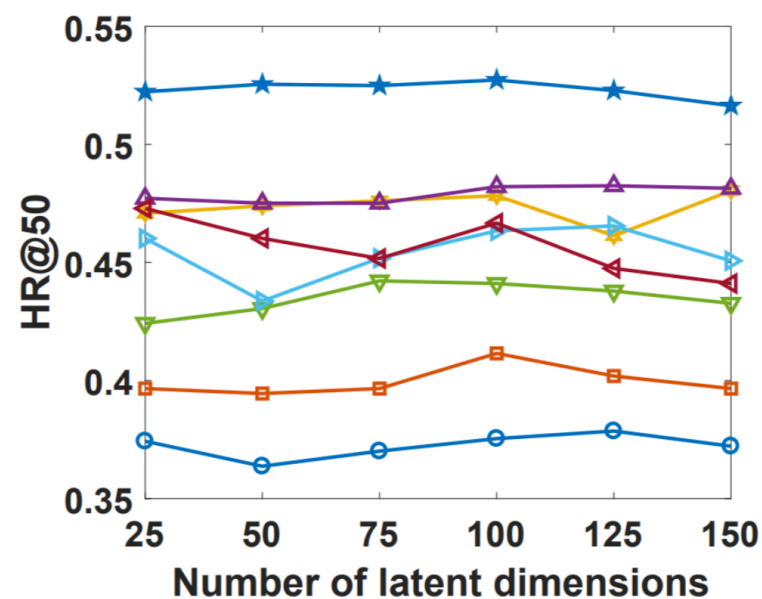
Results

- Hit ratio (HR@N)
- mean reciprocal rank (MRR)

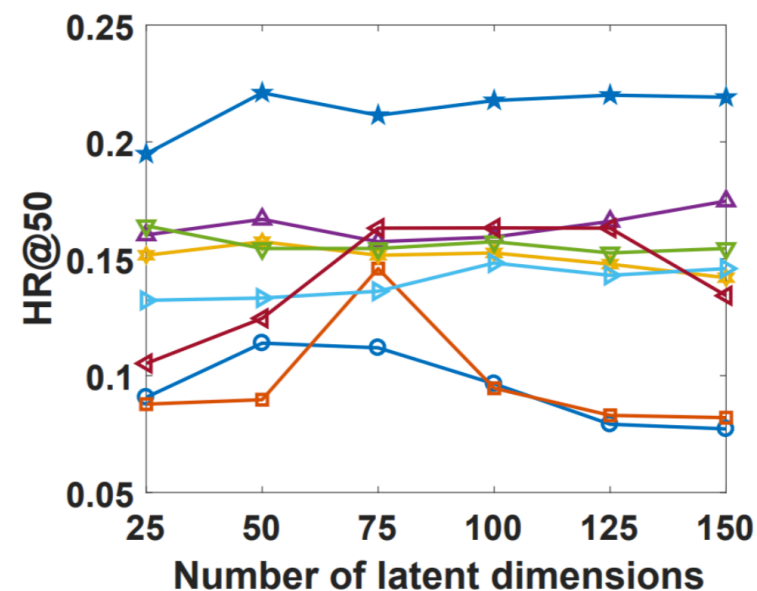
Results

Dataset	Metric	POP	BPRMF	MC	FPMC	HRM	PRME	TransRec	Caser	AttRec	Improv.
ML-100K	HR@50	0.2142	0.3754	0.4115	0.4783	0.4821	0.4411	0.4634	0.4667	0.5273	9.38%
	MRR	0.0388	0.0616	0.0662	0.0925	0.0889	0.0837	0.0827	0.0799	0.0981	10.35%
ML-HetRec	HR@50	0.1065	0.1462	0.1903	0.2321	0.2380	0.2357	0.1912	0.2144	0.2964	24.54%
	MRR	0.0177	0.0215	0.0359	0.0489	0.0486	0.0500	0.0337	0.0387	0.0592	18.40%
ML-1M	HR@50	0.1440	0.2378	0.3419	0.4209	0.4311	0.4449	0.3358	0.4811	0.5223	8.56%
	MRR	0.0231	0.0368	0.0654	0.1022	0.0873	0.1044	0.0561	0.0925	0.1172	12.26%
Android App	HR@50	0.1194	0.1738	0.1802	0.1990	0.2001	0.1686	0.2016	0.1426	0.2187	8.48%
	MRR	0.0228	0.0287	0.0355	0.0355	0.0295	0.0237	0.0306	0.0231	0.0415	16.90%
Health / Care	HR@50	0.0337	0.0900	0.0786	0.1128	0.0965	0.0843	0.0962	0.0768	0.1272	12.77%
	MRR	0.0171	0.0188	0.0245	0.0258	0.0183	0.0119	0.0232	0.0146	0.0277	7.36%
Video Game	HR@50	0.0609	0.1630	0.1708	0.2226	0.2150	0.1855	0.2035	0.1438	0.2414	8.45%
	MRR	0.0126	0.0277	0.0381	0.0451	0.0337	0.0263	0.0349	0.0248	0.0496	9.98%
Tools / Home	HR@50	0.0319	0.0559	0.0384	0.0535	0.0488	0.0465	0.0658	0.0424	0.0775	17.78%
	MRR	0.0061	0.0099	0.0093	0.0129	0.0086	0.0076	0.0112	0.0071	0.0148	14.73%
Digital Music	HR@50	0.0436	0.1621	0.1307	0.1580	0.1998	0.1559	0.1894	0.1327	0.2205	10.36%
	MRR	0.0073	0.0277	0.0320	0.0322	0.0310	0.0243	0.0300	0.0228	0.0467	45.03%
Garden	HR@50	0.0319	0.0965	0.0946	0.1525	0.1593	0.1573	0.1486	0.1632	0.2177	33.39%
	MRR	0.0049	0.0105	0.0333	0.0408	0.0255	0.0266	0.0257	0.0277	0.0459	12.50%
Instant Video	HR@50	0.1240	0.2350	0.1650	0.2120	0.2430	0.1910	0.2570	0.1620	0.2790	8.56%
	MRR	0.0173	0.0376	0.0426	0.0541	0.0414	0.0367	0.0441	0.0275	0.0634	17.19%
LastFM	HR@50	0.1314	0.3659	0.1682	0.2808	0.3733	0.2503	0.3785	0.1756	0.3901	3.06%
	MRR	0.0224	0.1062	0.0645	0.0869	0.1209	0.1276	0.1147	0.0343	0.1312	2.82%
MovieTweetings	HR@50	0.1687	0.1749	0.3314	0.3417	0.3105	0.3286	0.2755	0.3332	0.3602	5.41%
	MRR	0.0204	0.0231	0.0700	0.0674	0.0534	0.0476	0.0421	0.0585	0.0811	15.86%

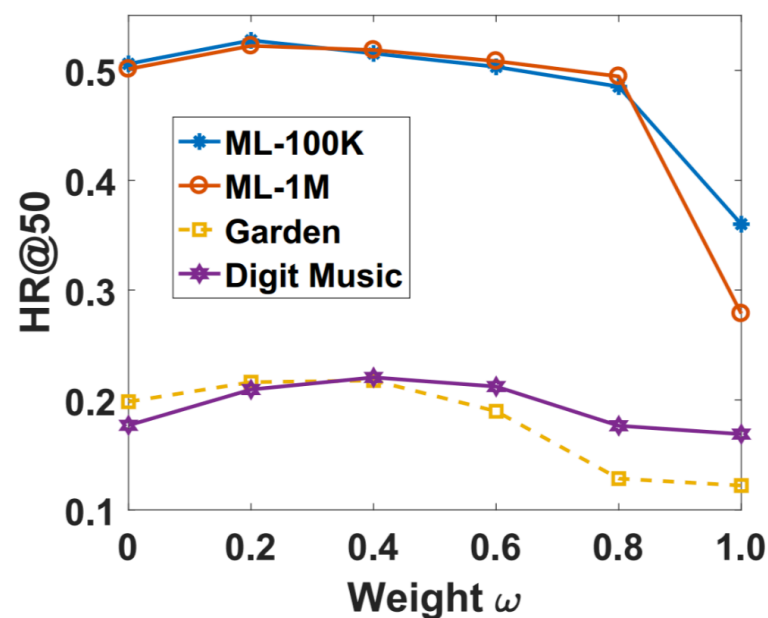
Parameter Sensibility



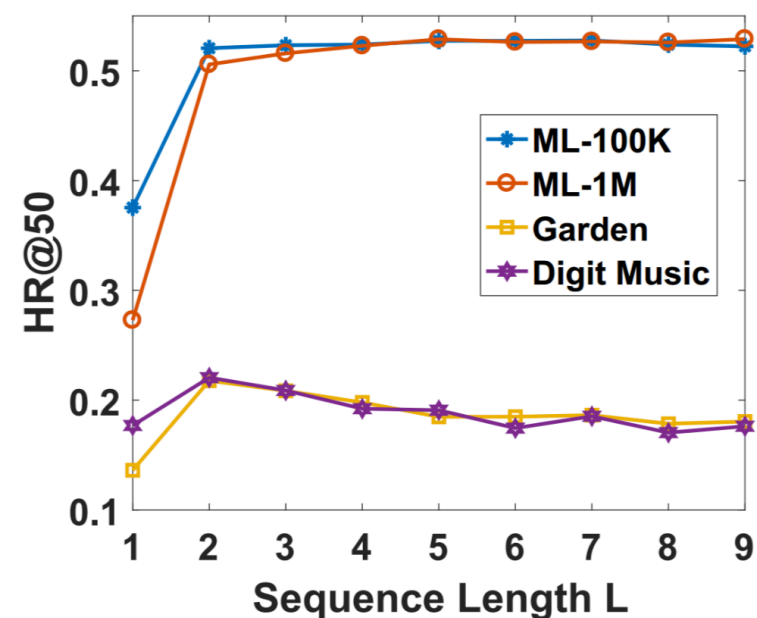
(a) Movielens 100K



(b) Amazon Garden



(a) Effects of the weight ω



(b) Effects of the sequence length L

Possible Extensions

- Attention instead of pondered mean for combination of long and short term modules.
- Attention instead of mean (transformer, etc)
- Longer ?histories?
- Extension to implicit feedback datasets.

