Dual Sequential Prediction Models Linking Sequential Recommendation and Information Dissemination

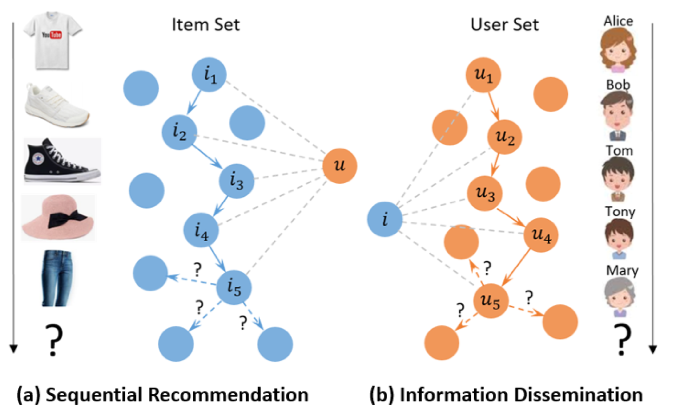
Qitian Wu, Yirui Gao, Xiaofeng Gao, Paul Weng, Guihai Chen

Shanghai Jiao Tong University

Introduction

Motivation

- An item's attributes (resp. user's interest) varies dynamically.
- A user's interests incline to change when exposed to different
- Concurrently modeling the temporal dependencies and targeting the distribution poses challenges



Dual structures in both user domain and item domain

- Sequential recommendation
- Information dissemination

Contribution

- **General Aspects:** We unify sequential recommendation and information dissemination to take advantage of dual information and conduct prediction.
- **Novel Methodologies:** We design a new training approach to integrate the dual models and use the output to define a hedge loss for training. The mechanism is helpful for distinguishing false negative samples.
- **Multifaceted Experiments:** We deploy DEEMS on four practical datasets. Experiment results verify the superiority of DEEMS over state-of-the-art techniques.

DEEMS model

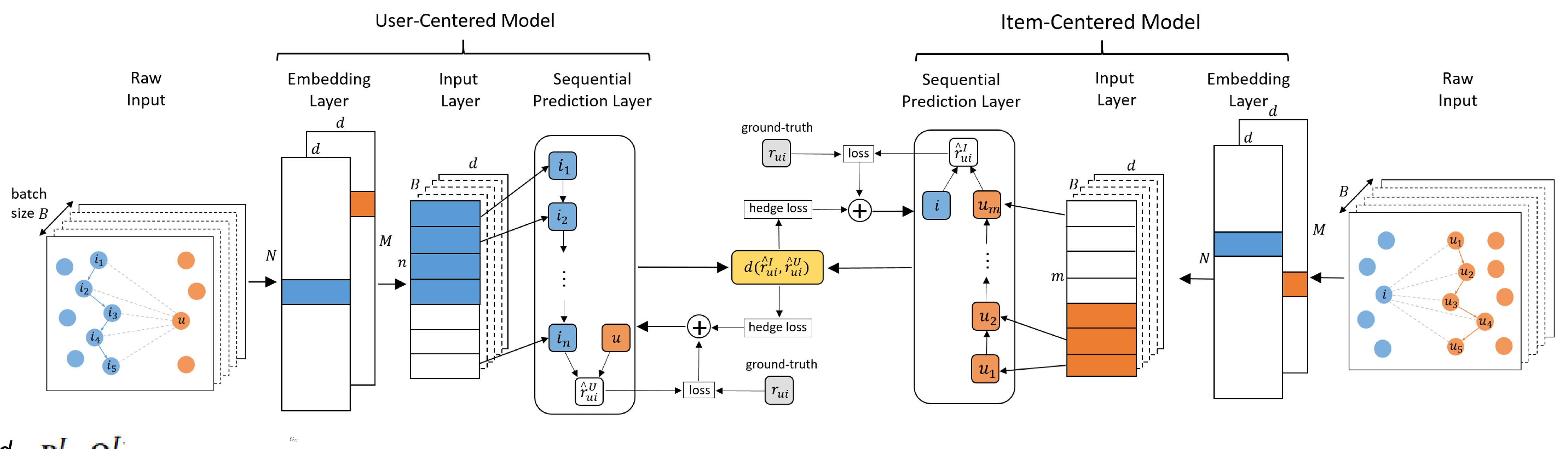
Raw Input and Item Implicit Network

- User-item interaction matrix R
- History of item sequence $\mathcal{I}_u^t$
- History of user sequence $\mathcal{U}_i^t$

Embedding Layer

- User embedding $\mathbf{P} \in \mathbb{R}^{d \times M}$
- Item embedding $\mathbf{Q} \in \mathbb{R}^{d \times N}$
- Embedding parameter

user-centered $\mathbf{P}^U$ $\mathbf{Q}^U$ *item-centered* $\mathbf{P}^I$ $\mathbf{Q}^I$



Sequential Prediction Layer

- Dynamic characterization

user-centered $\mathbf{s}_u^t = SU_U(\mathcal{I}_u^t, \mathbf{p}_u^U, \mathbf{q}_i^U; \mathbf{w}_U)$

item-centered $\mathbf{l}_i^t = SU_I(\mathcal{U}_i^t, \mathbf{p}_u^I, \mathbf{q}_i^I; \mathbf{w}_I)$

- Prediction score

user-centered $\hat{r}_{ui}^U = PU_U(\mathbf{p}_u^U, \mathbf{q}_i^U, \mathbf{s}_u^t; \mathbf{y}_U)$

item-centered $\hat{r}_{ui}^I = PU_I(\mathbf{p}_u^I, \mathbf{q}_i^I, \mathbf{l}_i^t; \mathbf{y}_I)$

Output Layer

Averaged scores to predict the probability that user *u* will interact with item *i*

e.g. $\hat{r}_{ui} = \frac{\hat{r}_{ui}^U + \hat{r}_{ui}^I}{2}$

Loss Function

Cross-entropy

$$\mathcal{L}_{sv}^U = -\sum_{(u,i,t)\in\mathcal{T}} r_{ui} \cdot \log \hat{r}_{ui}^U + (1 - r_{ui}) \cdot \log(1 - \hat{r}_{ui}^U).$$

$$\mathcal{L}_{sv}^I = -\sum_{(u,i,t)\in\mathcal{T}} r_{ui} \cdot \log \hat{r}_{ui}^I + (1 - r_{ui}) \cdot \log(1 - \hat{r}_{ui}^I).$$

Hedge loss

$$\mathcal{L}_{hd}^U = \frac{1}{K}\sum_{k=1}^{K} -\hat{r}_{u_k i_k}^I \cdot \log \hat{r}_{u_k i_k}^U - (1 - \hat{r}_{u_k i_k}^I) \cdot \log(1 - \hat{r}_{u_k i_k}^U).$$

$$\mathcal{L}_{hd}^I = \frac{1}{K}\sum_{k=1}^{K} -\hat{r}_{u_k i_k}^U \cdot \log \hat{r}_{u_k i_k}^I - (1 - \hat{r}_{u_k i_k}^U) \cdot \log(1 - \hat{r}_{u_k i_k}^I).$$

Final Regularized loss

$$\mathcal{L}^U = \mathcal{L}_{sv}^U + \alpha \mathcal{L}_{hd}^U + \frac{\lambda}{2} \|\theta_U\|^2$$

$$\mathcal{L}^I = \mathcal{L}_{sv}^I + \alpha \mathcal{L}_{hd}^I + \frac{\lambda}{2} \|\theta_I\|^2$$

Algorithm: Training for DEEMS

REQUIRE: interaction triple set $\mathcal{T}$.

REQUIRE: θ_U, initial parameters for user-centered model. θ_I, initial parameters for item-centered model. $\eta = 0.1$, learning rate. $\lambda = 0.01$, regularization coefficient. $B = 64$, batch size. $K = 5$, sample size. $\alpha = 0.1$, weight for hedge loss.

while *not converged* **do**

  Sample B user-item interactions $\mathcal{B} = \{(u_b, i_b, t_b)\}_{b=1}^{B}$ from $\mathcal{T}$;

  $\mathcal{L}_{sv}^U \leftarrow 0$, $\mathcal{L}_{hd}^U \leftarrow 0$, $\mathcal{L}_{sv}^I \leftarrow 0$, $\mathcal{L}_{hd}^I \leftarrow 0$;

  for *each sampled interaction* $(u, i, t) \in \mathcal{B}$ **do**

    Generate K items i_k uniformly among the unobserved pairs $(u, i') \notin \mathcal{T}$;

    $\mathcal{L}_{sv}^U += -\log \hat{r}_{ui}^U - \frac{1}{K}\sum_{k=1}^{K} \log(1 - \hat{r}_{ui_k}^U)$;

    $\mathcal{L}_{hd}^U += -\frac{1}{K}\sum_{k=1}^{K} \hat{r}_{ui_k}^I \cdot \log \hat{r}_{ui_k}^U + (1 - \hat{r}_{ui_k}^I) \cdot \log(1 - \hat{r}_{ui_k}^U)$;

  $\theta_U \leftarrow \theta_U - \eta \left[\nabla_{\theta_U} \mathcal{L}_{sv}^U + \alpha \nabla_{\theta_U} \mathcal{L}_{hd}^U + \lambda \|\theta_U\|\right]$;

  for *each sampled interaction* $(u, i, t) \in \mathcal{B}$ **do**

    Generate K user u_k uniformly among the unobserved pairs $(u', i) \notin \mathcal{T}$;

    $\mathcal{L}_{sv}^I += -\log \hat{r}_{ui}^I - \frac{1}{K}\sum_{k=1}^{K} \log(1 - \hat{r}_{u_k i}^I)$;

    $\mathcal{L}_{hd}^I += -\frac{1}{K}\sum_{k=1}^{K} \hat{r}_{u_k i}^U \cdot \log \hat{r}_{u_k i}^I + (1 - \hat{r}_{u_k i}^U) \cdot \log(1 - \hat{r}_{u_k i}^I)$;

  $\theta_I \leftarrow \theta_I - \eta \left[\nabla_{\theta_I} \mathcal{L}_{sv}^I + \alpha \nabla_{\theta_I} \mathcal{L}_{hd}^I + \lambda \|\theta_I\|\right]$;

Available sources

Codes: https://github.com/echo740/DANSER

Qitian Wu: echo740@sjtu.edu.cn

Experiments

Table 1: Experiment results of DEEMS and competitors when holding out the last two clicked items of each user for test.

	Digital				Baby				Grocery				Clothing			
	P@3	AUC	GP@3	GAUC	P@3	AUC	GP@3	GAUC	P@3	AUC	GP@3	GAUC	P@3	AUC	GP@3	GAUC
PMF	0.511	0.752	0.691	0.830	0.523	0.765	0.631	0.765	0.518	0.754	0.615	0.753	0.481	0.738	0.639	0.775
FPMC	0.515	0.765	0.717	0.845	0.527	0.766	0.639	0.770	0.520	0.776	0.620	0.759	0.492	0.739	0.646	0.770
TransRec	0.518	0.776	0.691	0.842	0.531	0.787	0.617	0.771	0.521	0.765	0.605	0.754	0.495	0.746	0.624	0.771
DREAM	0.529	0.780	0.724	0.855	0.525	0.778	0.648	0.785	0.524	0.775	0.633	0.763	0.507	0.750	0.652	0.789
DIN	0.531	0.779	0.712	0.843	0.529	0.776	0.643	0.782	0.524	0.777	0.630	0.768	0.503	0.751	0.648	0.789
RMTP	0.528	0.771	0.722	0.861	0.521	0.781	0.646	0.784	0.523	0.776	0.634	0.767	0.493	0.745	0.656	0.791
TopoLSTM	0.523	0.777	0.716	0.845	0.524	0.784	0.648	0.782	0.523	0.778	0.634	0.768	0.497	0.747	0.652	0.782
RRN	0.524	0.775	0.716	0.859	0.531	0.789	0.653	0.781	0.521	0.775	0.638	0.775	0.495	0.751	0.660	0.790
DEEMS-RNN	**0.564**	**0.827**	**0.726**	**0.867**	0.533	0.791	**0.659**	**0.793**	**0.529**	**0.783**	0.640	**0.781**	**0.530**	**0.789**	**0.667**	**0.800**
DEEMS-SAN	0.524	0.793	0.716	0.867	**0.553**	**0.809**	0.655	0.792	0.521	0.762	**0.644**	0.780	0.516	0.780	0.663	0.799

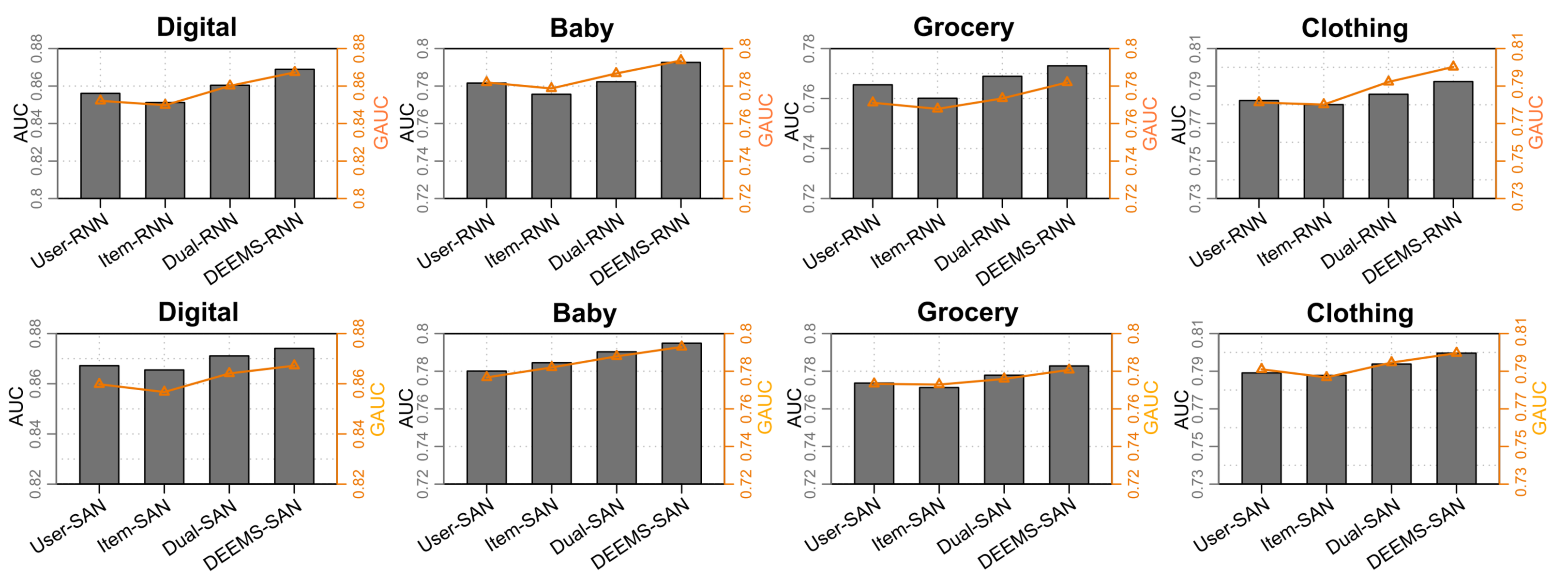


Figure 1: Ablation studies of dual models

	P@3	AUC	GP@3	GAUC
DEEMS - RNN - L1	0.5290	0.7841	0.6522	0.7807
DEEMS - RNN - L2	0.5312	0.7830	0.6529	0.7825
DEEMS - RNN - Log	0.5329	**0.7934**	0.6575	0.7902
DEEMS - RNN - Gauss	0.5301	0.7854	0.6510	0.7875
DEEMS - RNN	**0.5332**	0.7914	**0.6595**	**0.7966**
DEEMS - SAN - L1	0.5337	0.8014	0.6441	0.7850
DEEMS - SAN - L2	0.5356	0.8041	0.6414	0.7828
DEEMS - SAN - Log	0.5423	0.8097	0.6541	0.7892
DEEMS - SAN - Gauss	0.5375	0.8015	0.6508	0.7882
DEEMS - SAN	**0.5531**	**0.8091**	**0.6556**	**0.7928**

Table 2: Performance comparison under different metrics for hedge loss on Baby dataset

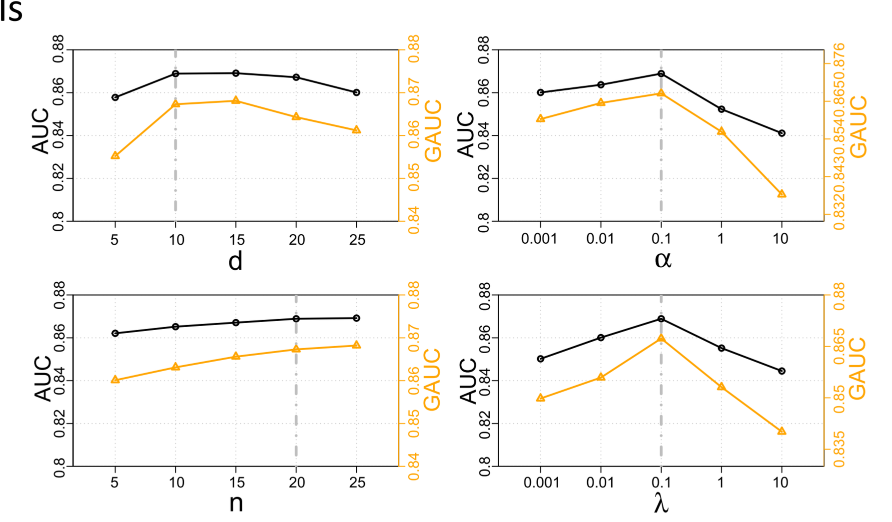


Figure 2: Performance of DEEMS on Baby dataset w.r.t different hyper-parameters.

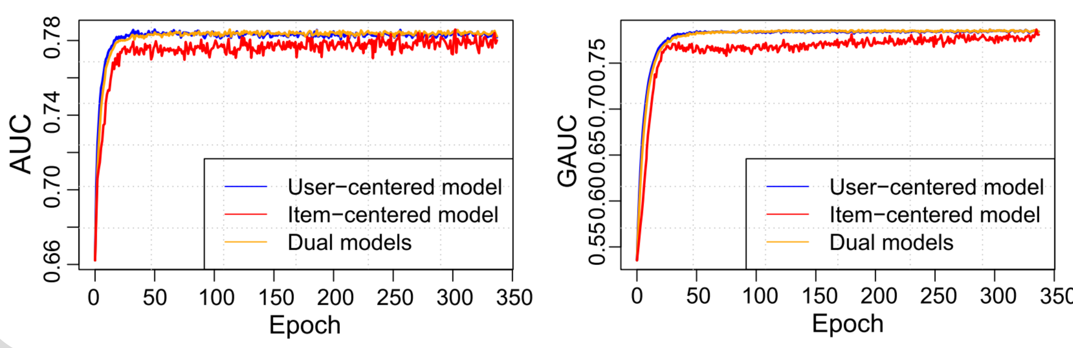


Figure 3: Learning curves of AUC and GAUC for user-centered, item-centered and dual models on Baby dataset

Illustration of DEEMS

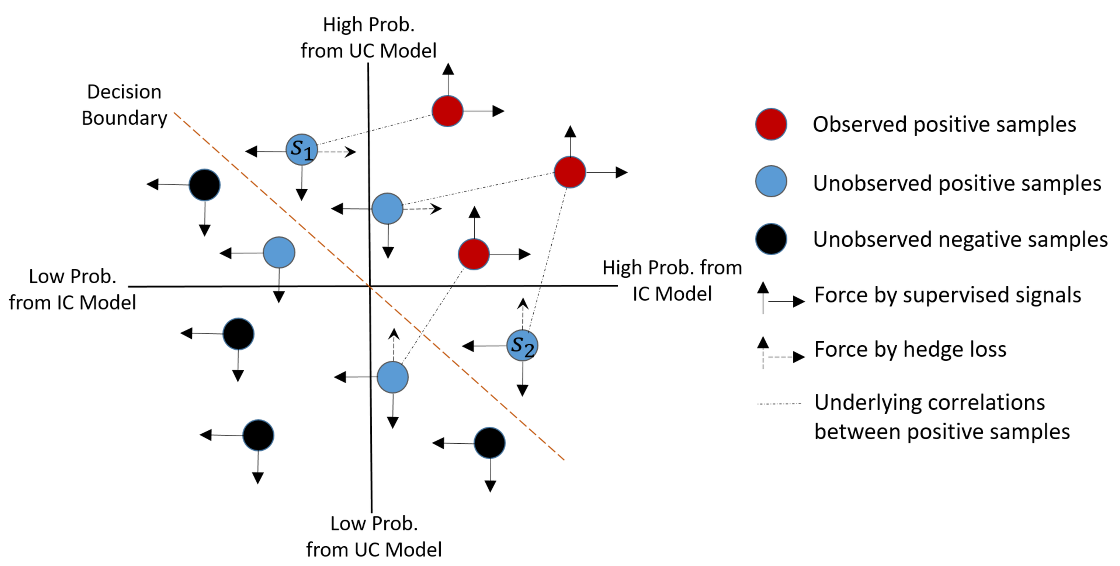


Figure: Illustration of training of DEEMS with user-centered model and item-centered model.

In the DEEMS model, the supervised signals from dual models would push the observed positive samples upward and rightward and unobserved samples downward and leftward. Strong interactions between items would give a relatively high probability in the user-centered model. Then the influence to the item-centered model increases the probability. Reversely, the hedge loss from item-centered model would guide the user-centered model and push the sample upward.

Implementation Details

Sequential Models

- Recurrent Neural Network Model (DEEMS-RNN)

$$\mathbf{h}_k = \tanh(\mathbf{W}'_I \mathbf{p}_k^I + \mathbf{W}'_H \mathbf{h}_{k-1} + \mathbf{b}'_I)$$

$$\hat{r}_{ui}^I = \sigma(\mathbf{W}'_U \cdot [\mathbf{p}_u^I, \mathbf{q}_i^I, \mathbf{l}_i^t] + \mathbf{b}'_U)$$

$$\mathbf{l}_i^t = \mathbf{W}'_O \mathbf{h}_m + \mathbf{b}'_O$$

- Self-Attention Model (DEEMS-SAN)

$$e_{uv} = \frac{\exp(\mathbf{W}'_A[\mathbf{p}_v^I, \mathbf{p}_u^I] + \mathbf{b}'_A)}{\sum_{k=1}^{m} \exp(\mathbf{W}'_A[\mathbf{p}_w^I, \mathbf{p}_u^I] + \mathbf{b}'_A)}, \quad \mathbf{l}_i^t = \sum_{v=1}^{n} e_{uv} \mathbf{p}_v^I$$

$$\hat{r}_{ui}^I = \sigma(\mathbf{W}'_U \cdot [\mathbf{p}_u^I, \mathbf{q}_i^I, \mathbf{l}_i^t] + \mathbf{b}'_U)$$

Different discrepancy metrics for hedge loss

Method	$d(r_{ui}^I, d_{ui}^U)$	$d(r_{ui}^U, d_{ui}^I)$
DEEMS-L1	$\lvert r_{ui}^I - r_{ui}^U \rvert$	$\lvert r_{ui}^U - r_{ui}^I \rvert$
DEEMS-L2	$\lvert r_{ui}^I - r_{ui}^U \rvert^2$	$\lvert r_{ui}^U - r_{ui}^I \rvert^2$
DEEMS-Log	$\frac{1}{1+\exp(-\lvert r_{ui}^I - r_{ui}^U \rvert)}$	$\frac{1}{1+\exp(-\lvert r_{ui}^U - r_{ui}^I \rvert)}$
DEEMS-Gauss	$\exp(\lvert r_{ui}^I - r_{ui}^U \rvert^2)$	$\exp(\lvert r_{ui}^U - r_{ui}^I \rvert^2)$