Co-Attentive Multi-Task Learning for Explainable Recommendation

Zhongxia Chen

4/29/2019

Background

- Personalized recommendation has become a major technique for helping users handle huge amounts of online content
- Recommender systems remain mostly black boxes
- Except for accuracy, there is a growing interest in model explainability
- Providing explanations can increase user trust, improve satisfaction, and persuade the users to buy or try an item

Motivation

- A fundamental question of explainable recommendation: How we balance accuracy and explainability for explainable recommendation?

- Most existing methods consider the two goals in separate steps or only focus on one of the goals
  - Post-hoc
  - Embedded
  - Simple jointly learning method

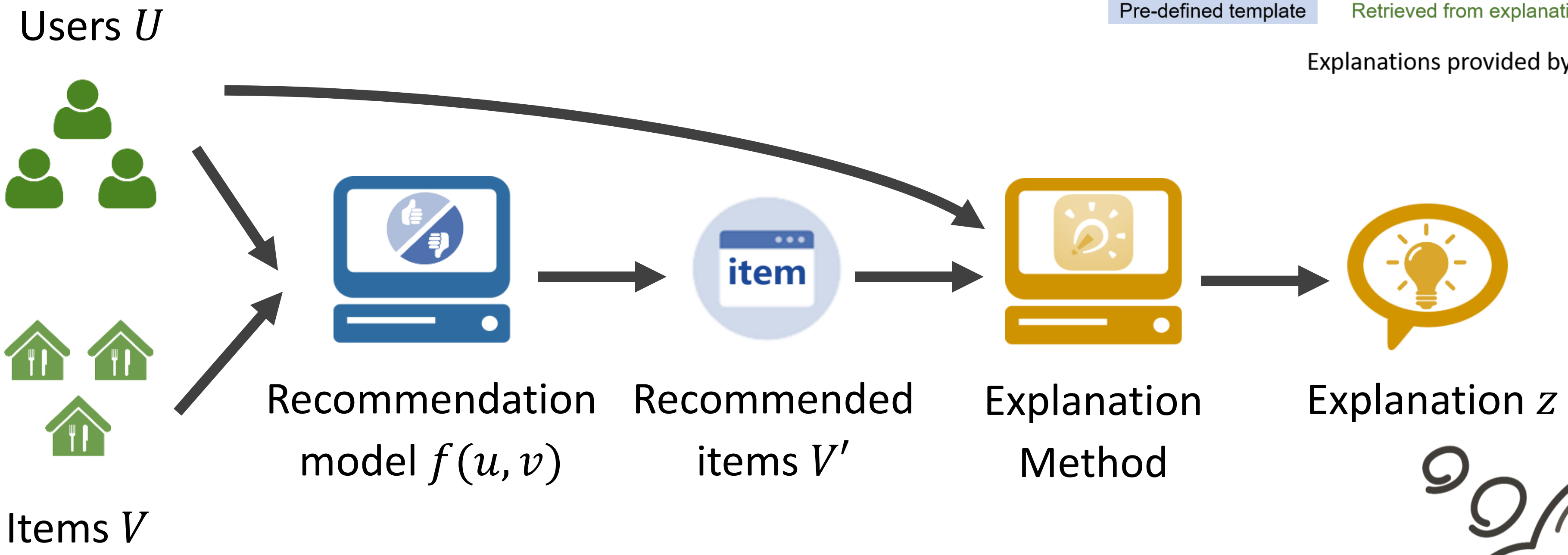
Related Works—Post-hoc

Item: Last Stand of the 300 User interest: war, history, documentary

(a) Post-hoc	Alice and 7 of your friends like this.
	Because you watched Spartacus, we recommend Last Stand of the 300.
(b) Embedded-F	You might be interested in documentary, on which this item performs well.
(c) Embedded-S	I agree with several others that this is a good companion to the movie.
(d) Joint	**This is a very good movie.**
(e) Ours	**This is a very good documentary about the battle of thermopylae.**

Pre-defined template Retrieved from explanations written by others **Generated by RNNs**

Explanations provided by different methods



Related Works—Post-hoc

Item: Last Stand of the 300 User interest: war, history, documentary

(a) Post-hoc	Alice and 7 of your friends like this. Because you watched Spartacus, we recommend Last Stand of the 300.
(b) Embedded-F	You might be interested in documentary, on which this item performs well.
(c) Embedded-S	I agree with several others that this is a good companion to the movie.
(d) Joint	**This is a very good movie.**
(e) Ours	**This is a very good documentary about the battle of thermopylae.**

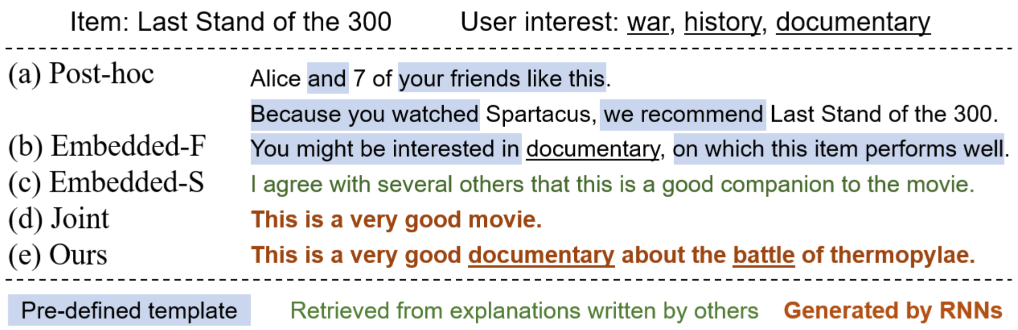
Pre-defined template Retrieved from explanations written by others **Generated by RNNs**

Explanations provided by different methods

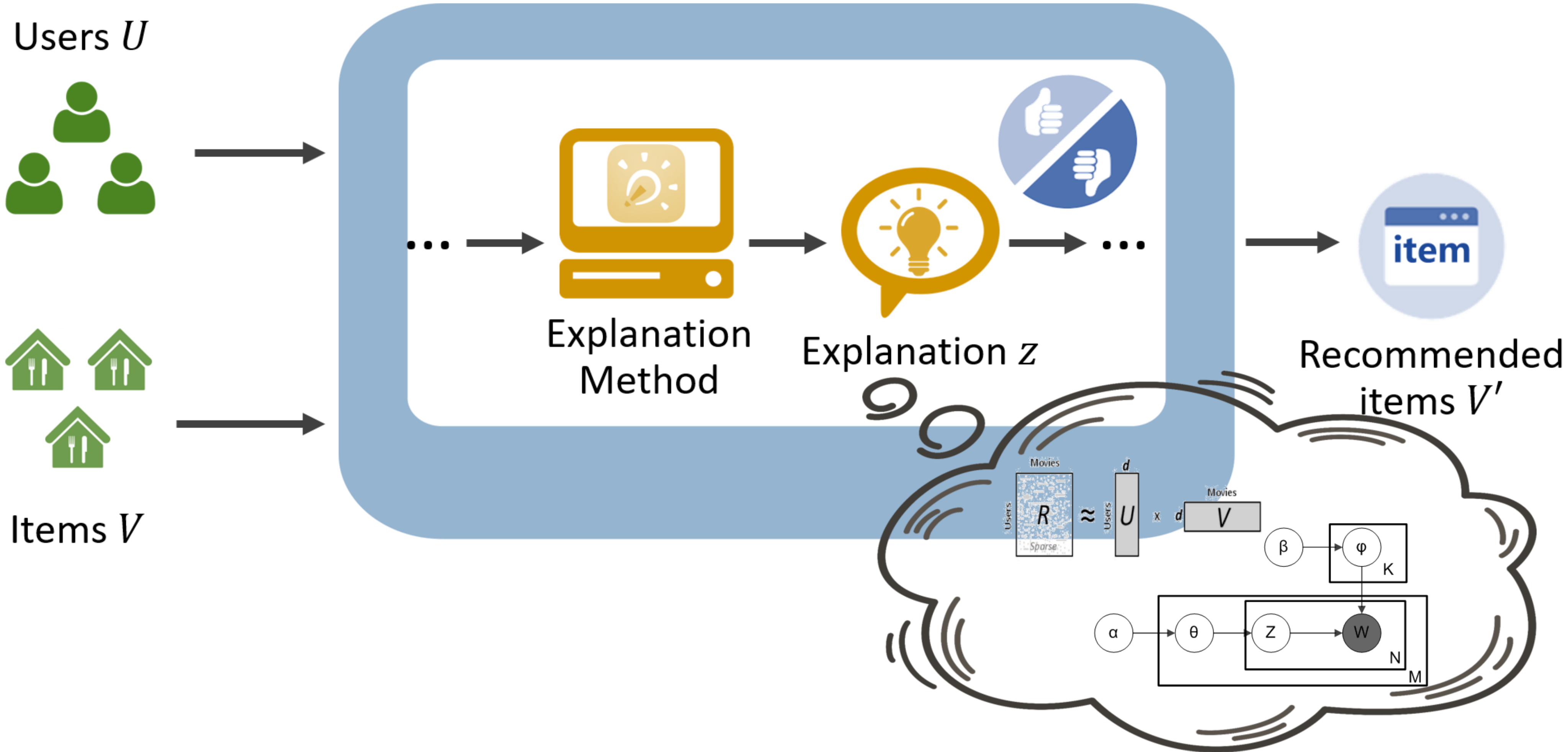
- Explain a black-box model after it is trained
  - Separately consider accuracy and explainability
  - Information embedded inside the recommendation models are ignored

- Pros and cons
  - Highly readable and persuasive
  - Not reflect model's actual reasoning
  - Difficult to generate in non-social scenarios
  - Limited diversity

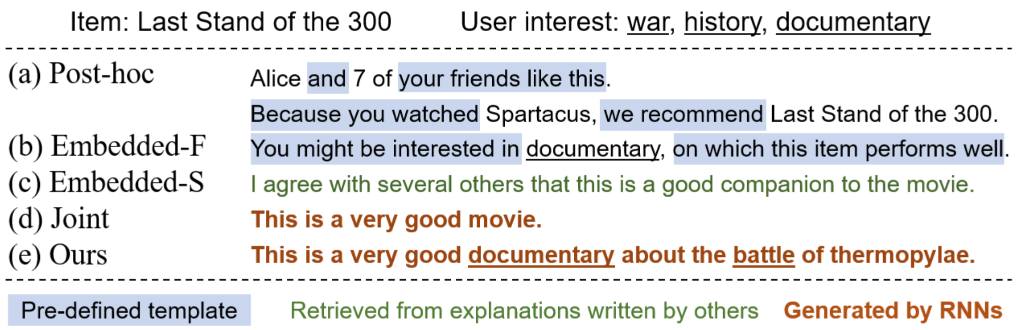
Related Works—Embedded



Explanations provided by different methods



Related Works—Embedded

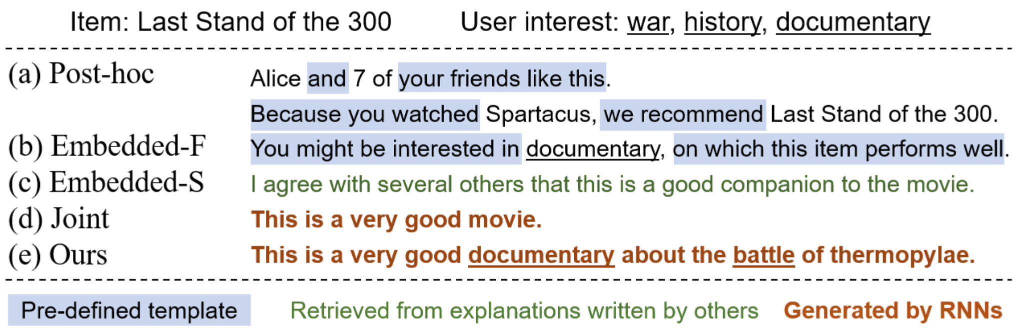


Explanations provided by different methods

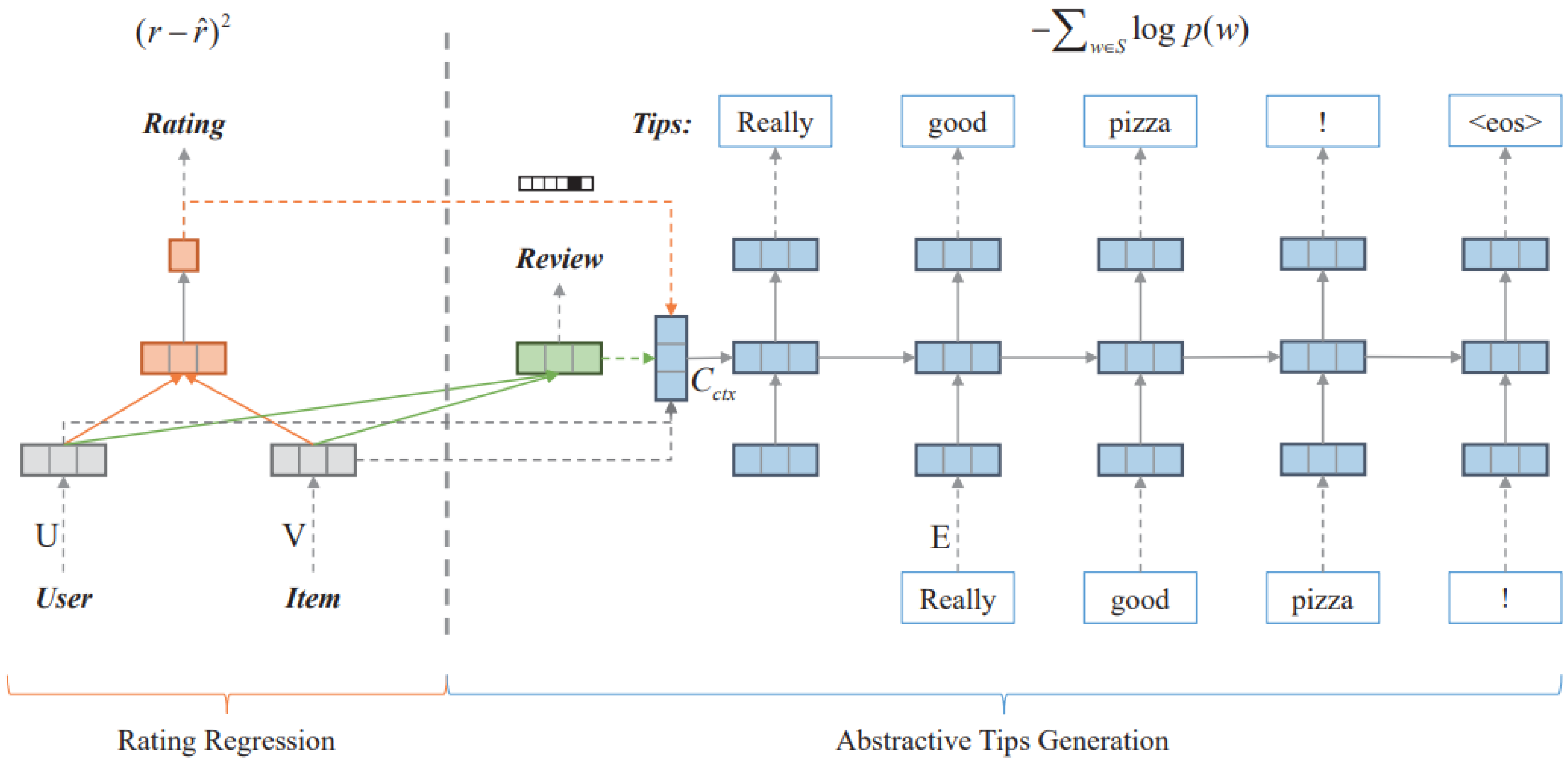
- Integrate the explanation process into the construction of the recommendation model
  - Retrieval-based
  - Consist of features or sentences
  - But only focus on recommendation accuracy
    - Explainability is not included in the optimization goal

- Issues
  - Difficult to guarantee the quality of the explanations
  - Fail to provide a highly personalized explanation when data is sparse
  - Legal issues (copyright)

Related Works—Joint



Explanations provided by different methods



Related Works—Joint

Item: Last Stand of the 300 User interest: war, history, documentary

(a) Post-hoc	Alice and 7 of your friends like this.
	Because you watched Spartacus, we recommend Last Stand of the 300.
(b) Embedded-F	You might be interested in documentary, on which this item performs well.
(c) Embedded-S	I agree with several others that this is a good companion to the movie.
(d) Joint	**This is a very good movie.**
(e) Ours	**This is a very good documentary about the battle of thermopylae.**

Pre-defined template Retrieved from explanations written by others **Generated by RNNs**

Explanations provided by different methods

- A simple jointly learning method
  - Only shares user/item latent representations
- Issues
  - The shared representations are not explainable and fail to provide explicit constraints on the explanations
  - User/item embeddings do not contain sufficient information about deep user-item interactions
  - Generated explanations are usually quite general

Contributions

- We propose a Co-Attentive Multi-task Learning (**CAML**) model that tightly couples the recommendation task and the explanation task
  - Design an encoder-selector-decoder architecture for multi-task learning based on **cognitive psychology**
  - Propose a hierarchical co-attentive selector to effectively control the cross knowledge transfer for both tasks by incorporating **multi-pointer networks**
  - Our method improves **both explainability and accuracy**

Problem Formulation

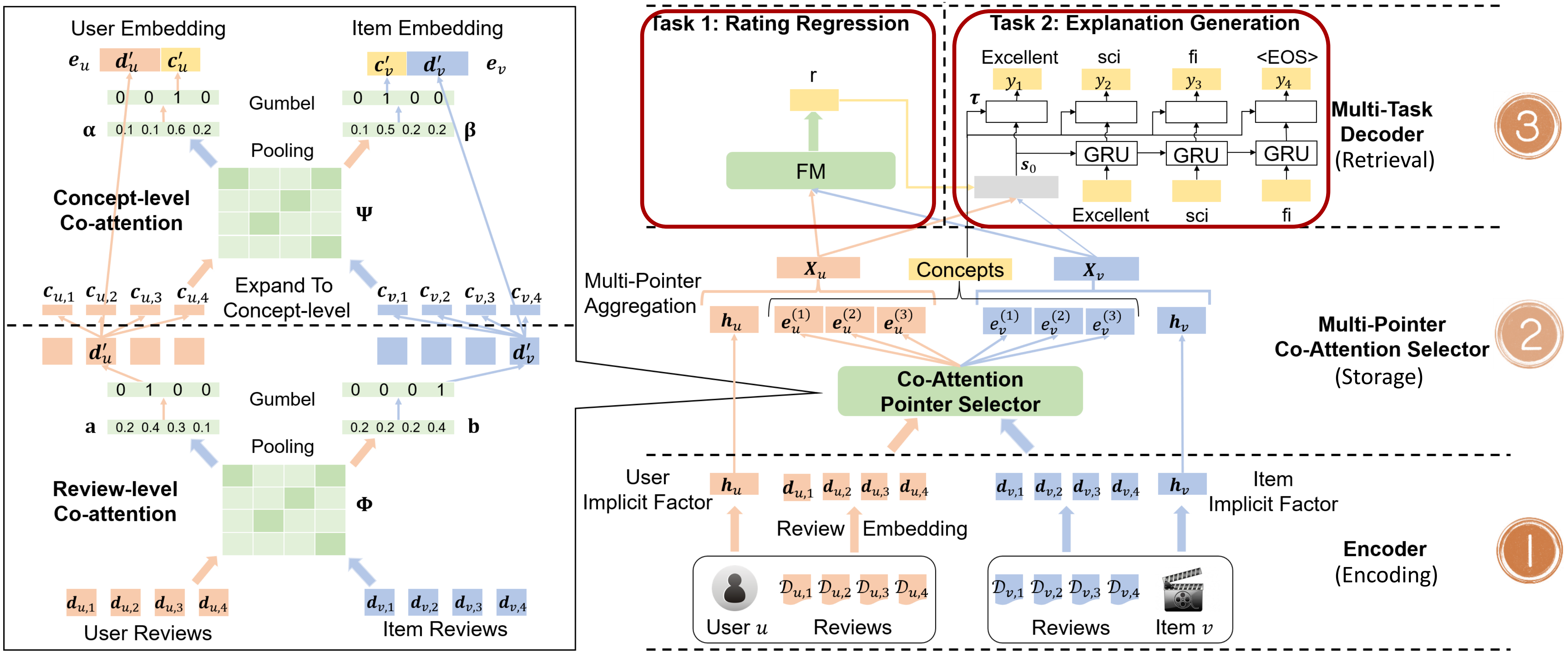
Microsoft Concept Graph

- Input
  - User set U, item set V
  - User reviews $D_{u,1}, \dots, D_{u,l_d}$, item reviews $D_{v,1}, \dots, D_{v,l_d}$
  - Concepts
    - a subset of words that correspond to important explicit features mentioned in the review

This is a great little comedy with a catchy song.

- Output
  - Rating r
  - Linguistic explanation $Y = (y_1, y_2 \dots, y_T)$
    - illustrates why user u likes or does not like item v

Method

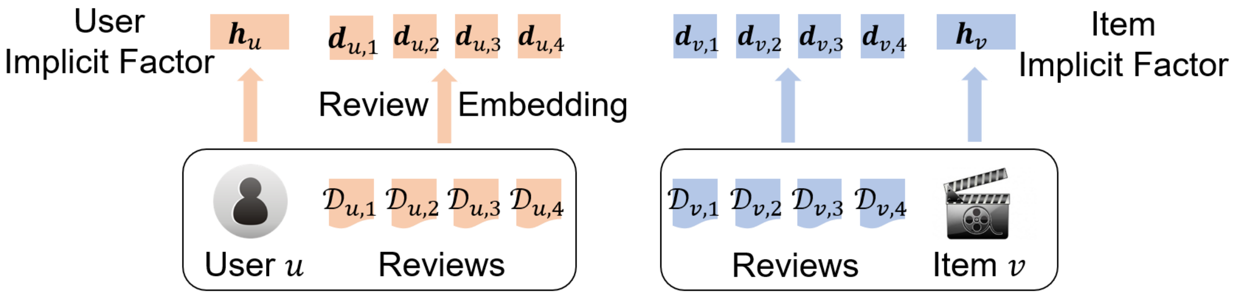


User Embedding
Item Embedding
Gumbel
Pooling
Concept-level Co-attention
Expand To Concept-level
Review-level Co-attention
User Reviews
Item Reviews
Task 1: Rating Regression
Task 2: Explanation Generation
FM
r
Excellent
sci
fi
<EOS>
GRU
Concepts
Multi-Pointer Aggregation
Co-Attention Pointer Selector
User Implicit Factor
Item Implicit Factor
Review Embedding
User u
Reviews
Item v
Multi-Task Decoder (Retrieval)
Multi-Pointer Co-Attention Selector (Storage)
Encoder (Encoding)

Encoder

- Word Encoder

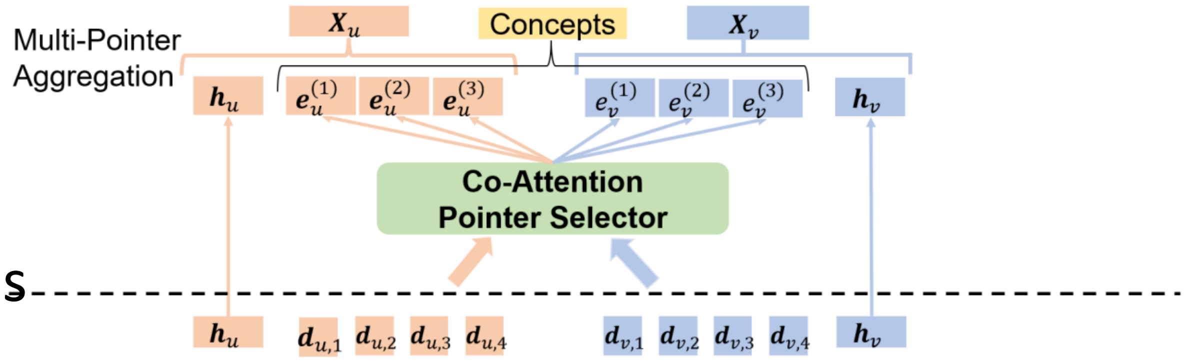
- Review Encoder
  - $\boldsymbol{d}_{u,i} = \sum_{w \in D_{u,i}} w$



- User/Item implicit factor Encoder
  - Complement explicit factors of user/item

Multi-Pointer Co-Attention Selector

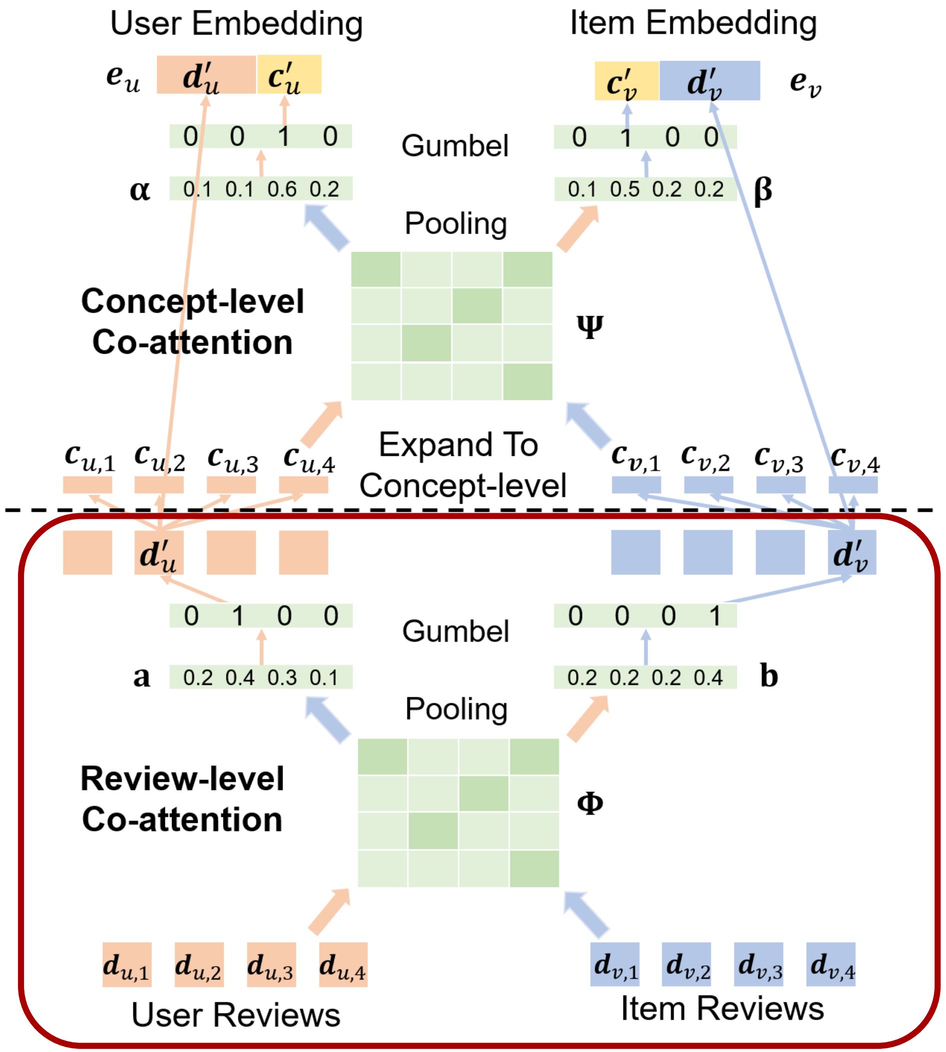
- Model the cross knowledge transferred for the two tasks
- Advantages of the multi-pointer co-attention networks (MPCN)
  - Faster convergence than REINFORCE
  - Model deep level user-item interactions



- Extend the MPCN to hierarchically select reviews and then concepts
  - Review-level co-attention pointer
  - Concept-level co-attention pointer
  - Multi-pointer aggregation

Review-level co-attention pointer

- Review-level Co-attention
  - $\phi_{i,j} = F(\boldsymbol{d}_{u,i})^T \boldsymbol{W}_d F(\boldsymbol{d}_{v,j})$
- Max-Pooling
  - $a_i = \max_{j=1,\ldots,l_d} \phi_{i,j}$
  - $b_j = \max_{i=1,\ldots,l_d} \phi_{i,j}$



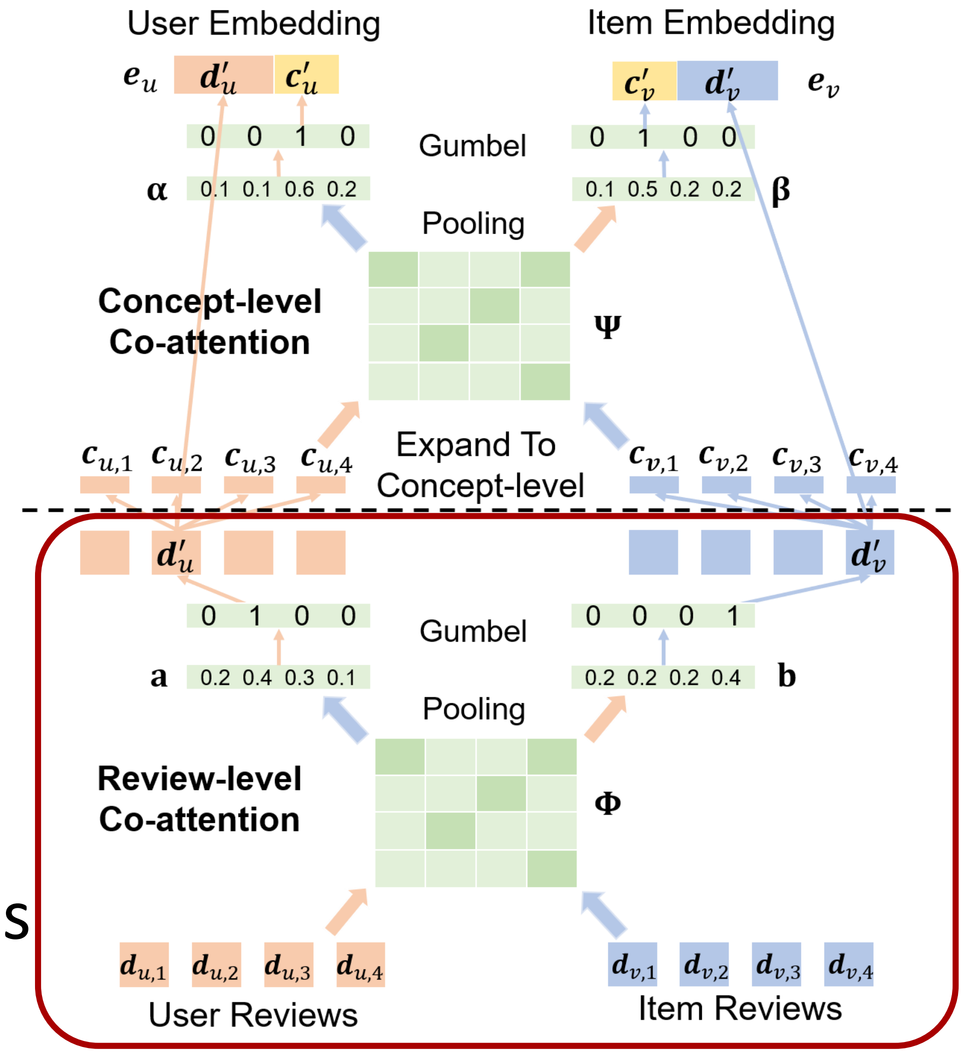
Review-level co-attention pointer

- Gumbel-Softmax
  - $q_i = \frac{\exp(\frac{a_i + g_i}{\tau})}{\sum_{j=1}^{n_m} \exp(\frac{a_j + g_i}{\tau})}$
  - g_i is the Gumbel noise

- Forward pass: Hard attention

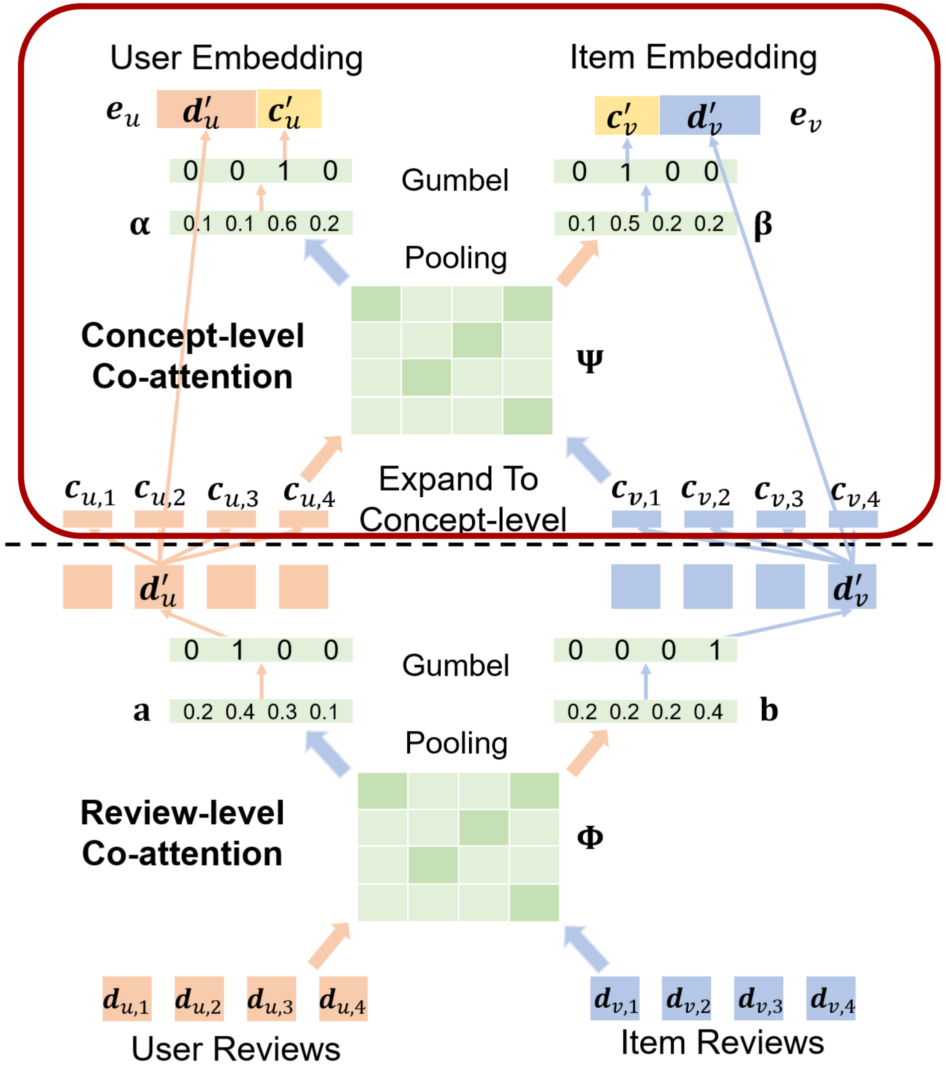
$$z_i = \begin{cases} 1, & i = \arg\max_j(q_j), \\ 0, & \text{otherwise} \end{cases}$$

- Backward pass: continuous gradients



Concept-level co-attention pointer

- Expand review to concept level
- Concept-level Co-attention
  - $\psi_{i,j} = F(\boldsymbol{c}_{u,i})^T \boldsymbol{W}_c F(\boldsymbol{c}_{v,j})$
- Mean-Pooling and Gumbel-Softmax

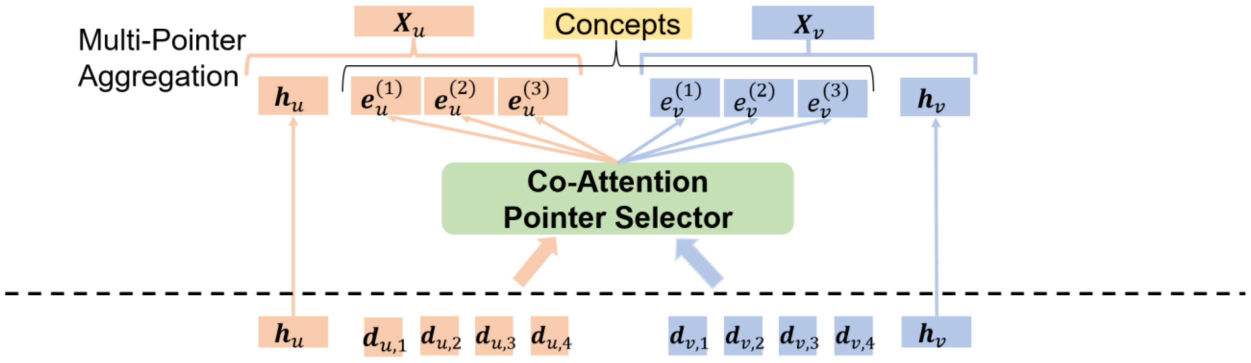


Multi-pointer aggregation

- Run selector multiple times

- Aggregate latent embeddings
  - Non-linear layer

$$\bar{\mathbf{e}}_u = \sigma(\mathbf{W}_p[\mathbf{e}_u^{(1)}, ..., \mathbf{e}_u^{(n_p)}] + \mathbf{b}_p)$$
$$\bar{\mathbf{e}}_v = \sigma(\mathbf{W}_p[\mathbf{e}_v^{(1)}, ..., \mathbf{e}_v^{(n_p)}] + \mathbf{b}_p)$$



- Collect selected concepts

Decoder

- Rating prediction
  - Factorization machine

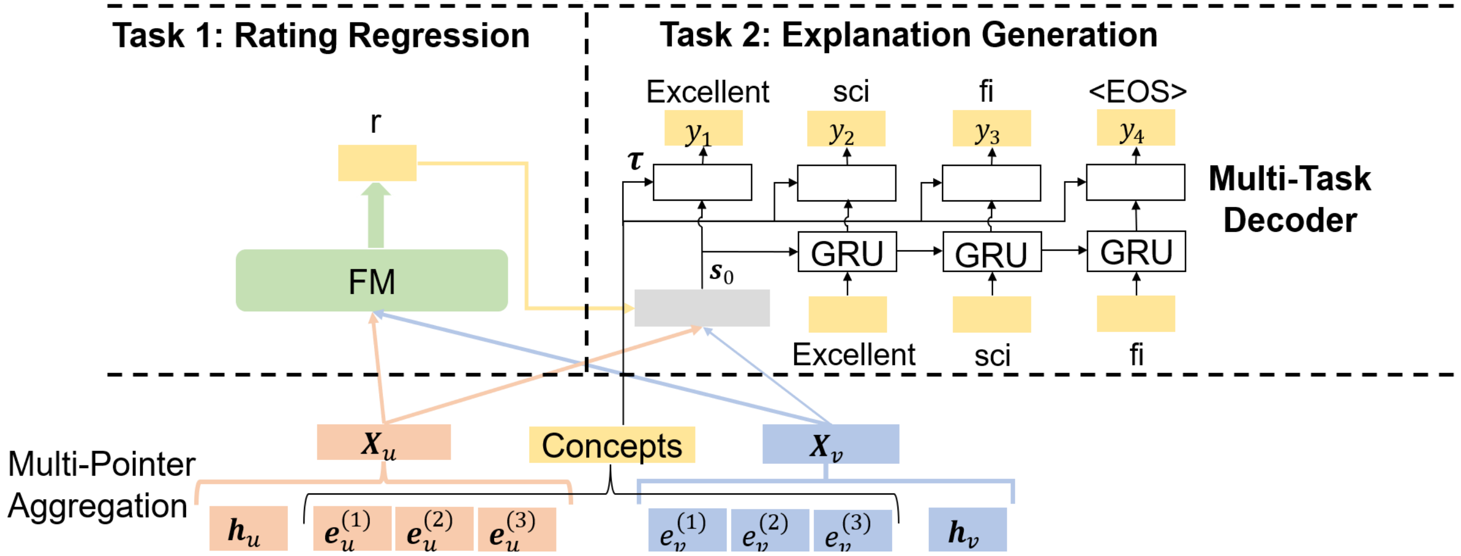
$$\mathcal{L}_r = \frac{1}{2|\Omega|}\sum_{(u,v)\in\Omega}(r - r_*)^2$$

- Explanation Generation
  - RNN decoder
  - Concept relevance loss

$$\mathcal{L}_c = \frac{1}{|\Omega|}\sum_{(u,v)\in\Omega}\sum_{t=1}^{T}(\max_k \ (-\tau_k \log o_{t,k}))$$

  - Negative log-likelihood loss

$$\mathcal{L}_n = \frac{1}{|\Omega|}\sum_{(u,v)\in\Omega}\sum_{t=1}^{T}(-\log o_{t,y_{t*}})$$



- Joint learning

$$\mathcal{L} = \mathcal{L}_r + \lambda_c\mathcal{L}_c + \lambda_n\mathcal{L}_n + \lambda_l\|\Theta\|_2^2$$

Experiments

- Datasets
  - Amazon Electronics, Movies&TV
  - Yelp

- Baselines
  - Explainability
    - Retrieval-based: Lexrank, NARRE, RLRec
    - Generative: NRT
  - Rating prediction
    - CF: PMF, NMF, SVD++
    - Neural: MPCN, NARRE, RLRec, NRT

	Electronics	Movies&TV	Yelp-2016
Users	192,403	123,960	677,379
Items	63,001	50,052	84,693
Reviews	1,688,117	1,697,533	2,530,843
Concepts	652	791	1,004

- Metrics
  - Explainability: Bleu and ROUGE, Human evaluation
  - Rating prediction: RMSE

Explainability Results

	Electronics	Movies&TV	Yelp-2016
Users	192,403	123,960	677,379
Items	63,001	50,052	84,693
Reviews	1,688,117	1,697,533	2,530,843
Concepts	652	791	1,004

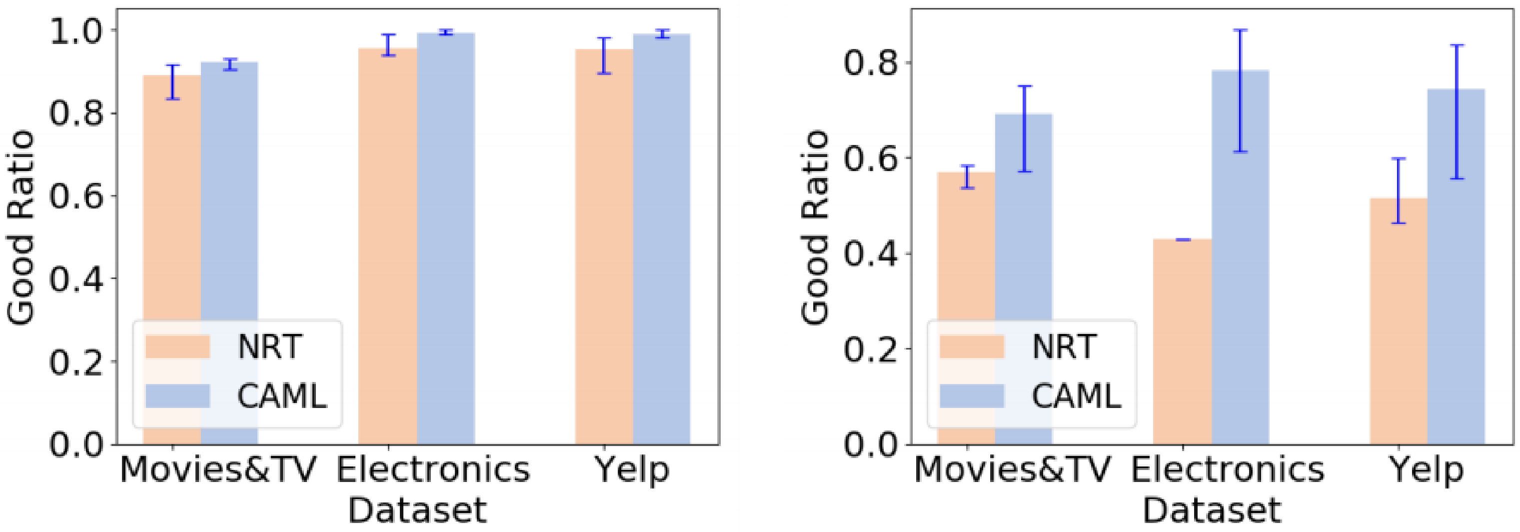
- Overall Performance

Datasets	Criteria	**Retrieval**			**Generative**	**Ours**			Improvement (%)	
		LexRank	NARRE	RLRec	NRT	CAML-G	CAML-C	CAML	$\Delta_{Retrieval}$	$\Delta_{Generative}$
Electronics	BLEU	1.44	1.45	1.45	1.33	1.79	1.92	**1.97**	+35.9	+48.1
	ROUGE-1	14.22	15.19	11.12	17.39	18.64	19.00	**19.26**	+26.8	+10.8
	ROUGE-2	3.60	3.29	1.60	3.50	3.63	3.78	**3.81**	+5.8	+8.9
	ROUGE-L	13.70	13.28	9.70	15.71	16.37	16.63	**16.75**	+22.3	+6.6
	ROUGE-SU4	4.38	5.25	3.13	5.97	6.26	6.43	**6.47**	+23.2	+8.4
Movies&TV	BLEU	1.78	1.75	1.73	1.60	1.94	**2.04**	**2.04**	+14.6	+27.5
	ROUGE-1	15.68	15.31	11.61	18.09	18.86	19.14	**19.32**	+23.2	+6.8
	ROUGE-2	2.45	3.62	3.84	4.30	4.48	4.43	**4.58**	+19.3	+6.5
	ROUGE-L	12.46	12.99	10.06	16.00	16.41	16.48	**16.69**	+28.5	+4.3
	ROUGE-SU4	5.24	5.79	4.98	6.29	6.52	6.57	**6.71**	+15.9	+6.7
Yelp	BLEU	0.97	1.13	1.13	1.31	1.50	1.57	**1.58**	+39.8	+20.6
	ROUGE-1	11.06	10.55	9.28	13.31	13.93	14.15	**14.24**	+28.8	+7.0
	ROUGE-2	2.42	2.66	1.93	3.05	3.42	3.47	**3.50**	+31.6	+14.8
	ROUGE-L	10.15	9.30	8.18	12.13	12.70	12.84	**12.90**	+27.1	+6.3
	ROUGE-SU4	3.58	3.85	3.10	4.60	4.92	5.00	**5.03**	+30.6	+9.3

Explainability Results

	Electronics	Movies&TV	Yelp-2016
Users	192,403	123,960	677,379
Items	63,001	50,052	84,693
Reviews	1,688,117	1,697,533	2,530,843
Concepts	652	791	1,004

- Human Evaluation
  - 3 assessors, 100 test cases



(a) Fluency (b) Usefulness

Case Study

Case 1. User interest: **horror**, **night**, **fun**	
NRT	I'll admit it.
CAML	I am a huge fan of **horror** movies, and this is one of my favorite movies.
Truth	Remember when **horror** movies were **fun**?
Case 2. User interest: **humor**, **scenery**, **main character**	
NRT	I love this series.
CAML	If you like british **humor**, you will love this series.
Truth	This is very british **humor**.
Case 3. User interest: **story line**, **cartoon**, **worth**	
NRT	I enjoyed this series as much as the first one.
CAML	I enjoyed this movie, the **animation** was great and the **story line** was very good.
Truth	Great **price** and the **animation** was cool.

Case 1	1	If you are a **fan** of the 80's, you'll love this.
	2	Not the best of the old **horror** movies, but it's still a good one.
	3	Nice to see the old classic **horror** movies.
	GT	I remember watching this film when I was a 8 years **kid**, I was so terrified, i didn't want to go to the bath alone!
Case 2	1	This movie is a total waste of **time**.
	2	Save your **money**.
	3	You know the movie is a **joke**.
	GT	Embarassingly painful is what this crap.
Case 3	1	As a **fan** of the phantom of the opera, I was very excited to see this movie.
	2	If you are a **fan** of the phantom, you will love this movie.
	3	The **story** was a bit rushed to the end.
		What a great **cast**.
	GT	As a **person** who really enjoyed phantom of the opera, I was expecting quite a bit from this **piece**.

Accuracy Results

	Electronics	Movies&TV	Yelp-2016
Users	192,403	123,960	677,379
Items	63,001	50,052	84,693
Reviews	1,688,117	1,697,533	2,530,843
Concepts	652	791	1,004

Datasets	**CF**			**Neural Models**				**Ours**			Improvement (%)	
	PMF	NMF	SVD++	MPCN	NARRE	RLRec	NRT	CAML-R	CAML-C	CAML	Δ_{CF}	Δ_{Neural}
Electronics	2.065	1.170	1.105	1.105	1.103	1.102	1.091	**1.085**	1.086	**1.085**	+1.8	+0.6
Movies&TV	1.250	1.089	1.013	1.001	0.999	1.012	0.990	0.990	**0.987**	**0.987**	+2.6	+0.3
Yelp	1.829	1.290	1.193	1.193	1.190	1.220	1.186	1.180	**1.173**	**1.173**	+1.7	+1.1

Conclusion

- We propose a co-attentive multi-task learning which fully exploits the correlations between the recommendation task and the explanation task

- We propose an encoder-selector-decoder architecture and a hierarchical co-attentive selector to effectively control the cross knowledge transfer for both tasks

- Experiments show that our approach outperforms state-of-the-art baselines on both the accuracy of rating prediction and the quality of generated explanations

Thank You!