Program to Principle: Neural Review mining for recommendation

Muthusamy Chelliah
Flipkart

Sudeshna Sarkar
IISc

Vishal Kakkar
Flipkart

Agenda

1. Introduction/topic models - cold start/explainability (10 min.s) (Chelliah)
2. Basic neural methods for recommendation (20 min.s) (Chelliah)
3. Attention - review ranking/aspect-based recommendation (20 min.s) (Sudeshna)
4. Programs - attention (20 min.s) (Vishal)
4. Generative approaches - review/tips (25 min.s) (Sudeshna)

5. Programs - generation (25 min.s) (Vishal)

Flipkart

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
RECRUITMENT 2018

NARRE: Neural Attentional Rating Regression with Review-level Explanations

- To predict a rating given a user and an item, as well as to select both useful and representative reviews.
- Useful reviews are obtained through a attention mechanism and provide explanations for users to make better and faster decisions.
- NARRE learns the usefulness of each review.

Chen C, Zhang M, Liu Y, Ma S. Neural Attentional Rating Regression with Review-level Explanations, WWW 2018.

Flipkart

TextCNN

- CNN text Processor: inputs a sequence of words and outputs a n-dimensional vector representation for the input.

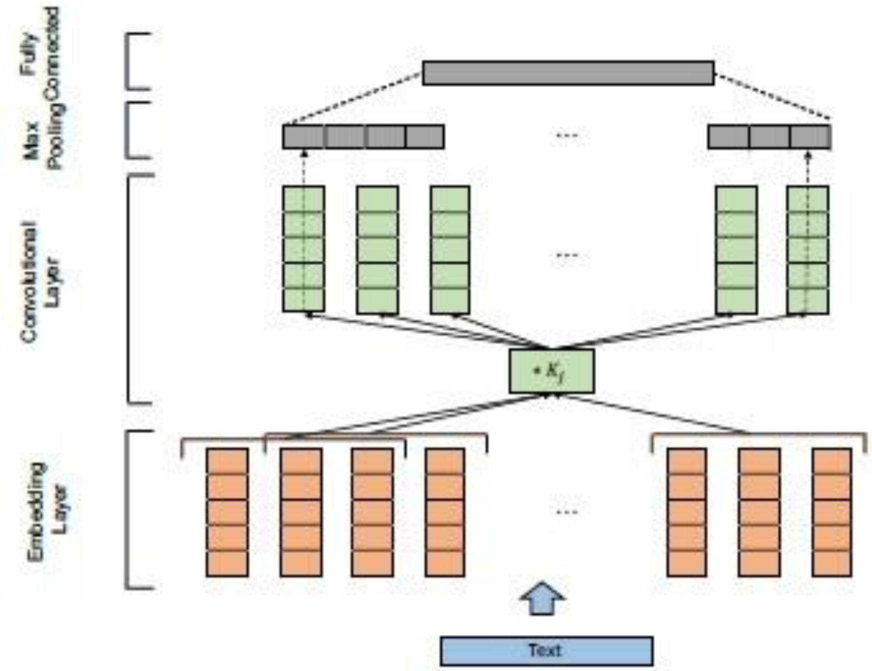


Figure 2: The CNN Text Processor architecture.

NARRE

- Utilize the attention mechanism to assign weights to reviews when modeling users and items.
- Two parallel neural networks for user modeling and item modeling.
- A prediction layer to let the hidden latent factors of user and item interact.
- The training data consists of users, items, and text reviews.
- Test stage: only users and items are available.

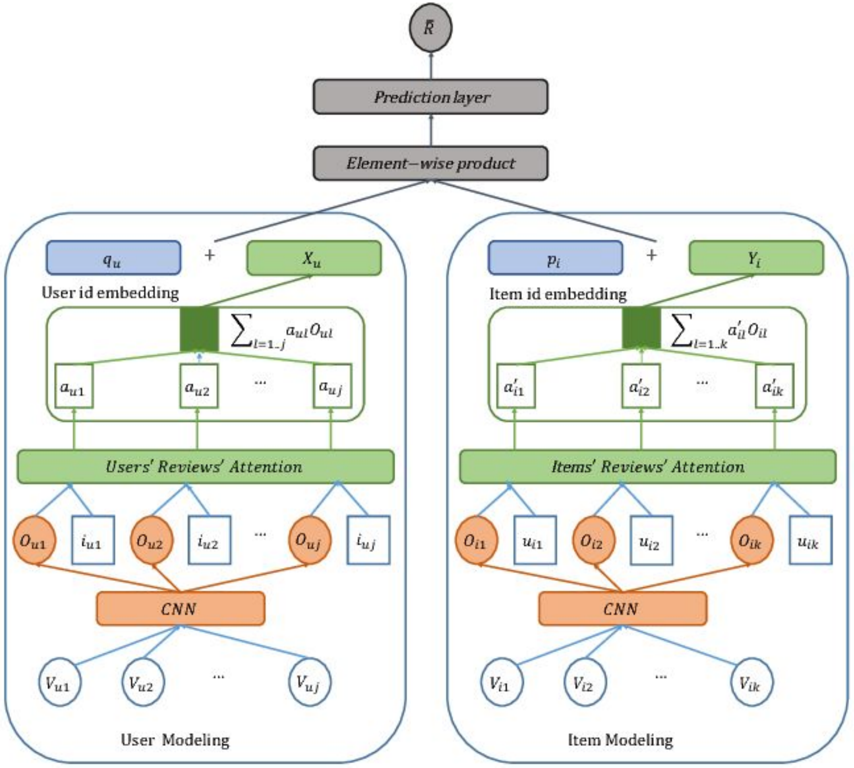
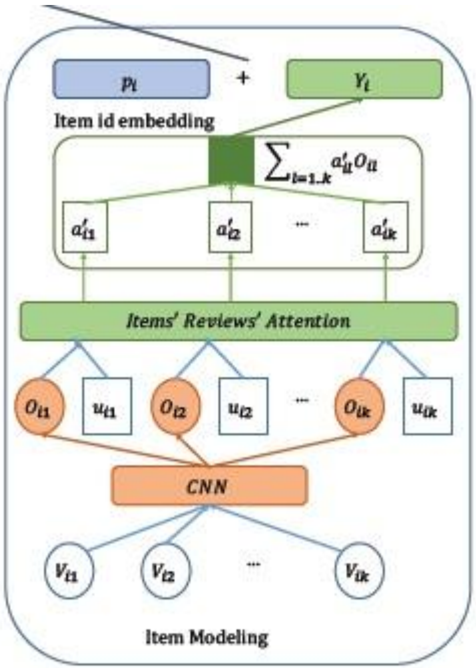


Figure 3: The neural network architecture of NARRE.

Flipkart

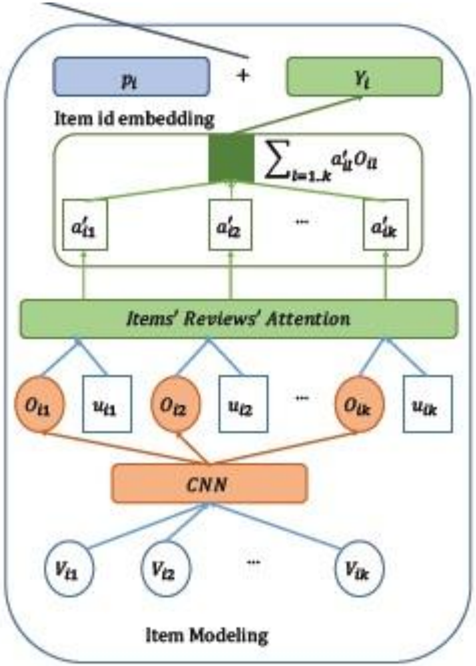
Item Modelling

- CNN Text Processor is applied to process the textual reviews of item i.
- Each review of i is transformed into a matrix of word vectors, (Vi1, Vi2,...Vik).
- These matrices are sent to the convolutional layer and the feature vectors of them can be obtained from the output as (Oi1, Oi2, ...Oi j) .
- Attention mechanism is used to learn the weight of each review.



Attention-based Review Pooling

- Select reviews that are representative to item i's features
- Aggregate the representation of informative reviews to characterize item i.
- A two-layer network is applied to compute the attention score a_{il} .
- The input contains the feature vector of the *l*th review of item *i* (O_{il}) and the user who wrote it
- (ID embedding, u_{il}) models the quality of users, helps identify users who always write less-useful reviews.



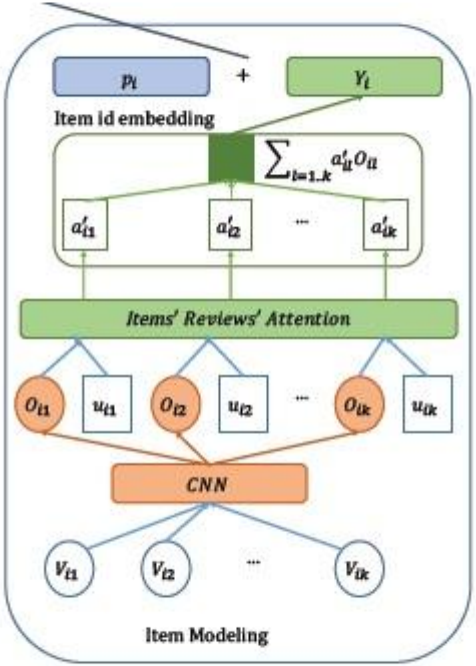
Flipkart

Attention-based Review Pooling

- The attention network:

$$a^*_{il} = h^T \, relu(W_O O_{il} + W_u u_{il} + b_1) + b_2$$

- Normalize: $a_{il} = \frac{\exp(a^*_{il})}{\sum_{i=0}^{k} \exp(a^*_{il})}$
- the feature vector of item i: $O_i = \sum_{l=1\ldots k} a_{il} O_{il}$
- The output of the attention-based pooling layer is a k1 dimensional vector
- A fully connected layer computes the final representation of item i: $Y_i = W_0 O_i + b_0$



NAPRE: Prediction Layer

- NAPRE extends user preferences and item features in LFM model to two components: one based on ratings while the other based on reviews.
- A neural form LFM used for predicting ratings.
- the latent factors of user and item are mapped to a shared hidden space.
- The interaction between u and i is modelled as:

$$h_0 = (q_u + X_u) \odot (p_i + Y_i)$$

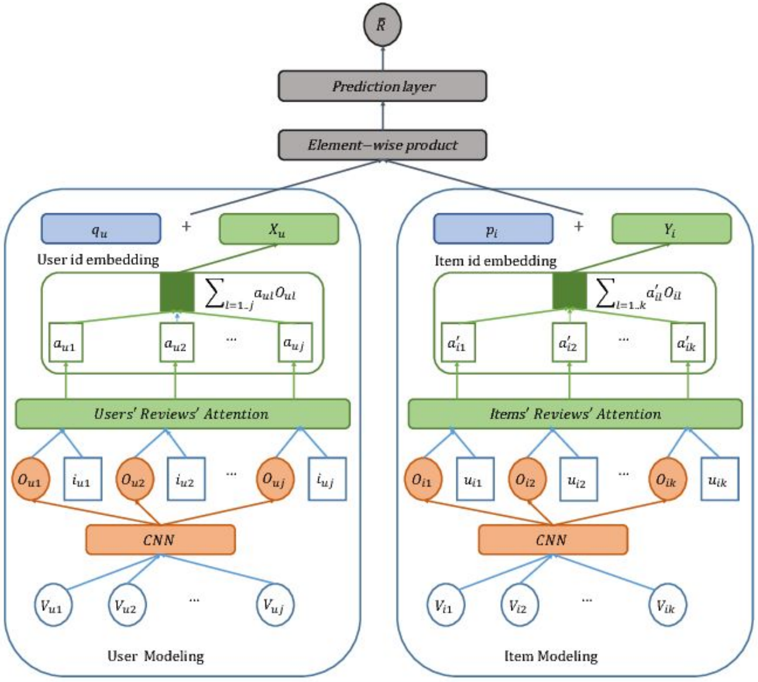


Figure 3: The neural network architecture of NARRE.

Flipkart

NAPRE: Prediction Layer

- The interaction between u and i is modelled as:

$$h_0 = (q_u + X_u) \odot (p_i + Y_i)$$

- q_u and p_i are user preferences and item features based on ratings from LFM
- X_u and Y_i are user preferences and item features obtained from the above method
- $\odot$ denotes the element-wise product of vectors.
- The output is a n-d vector, passed to prediction layer:

$$\hat{R}_{u,i} = {w_1}^T h_0 + b_u + b_i + \mu$$

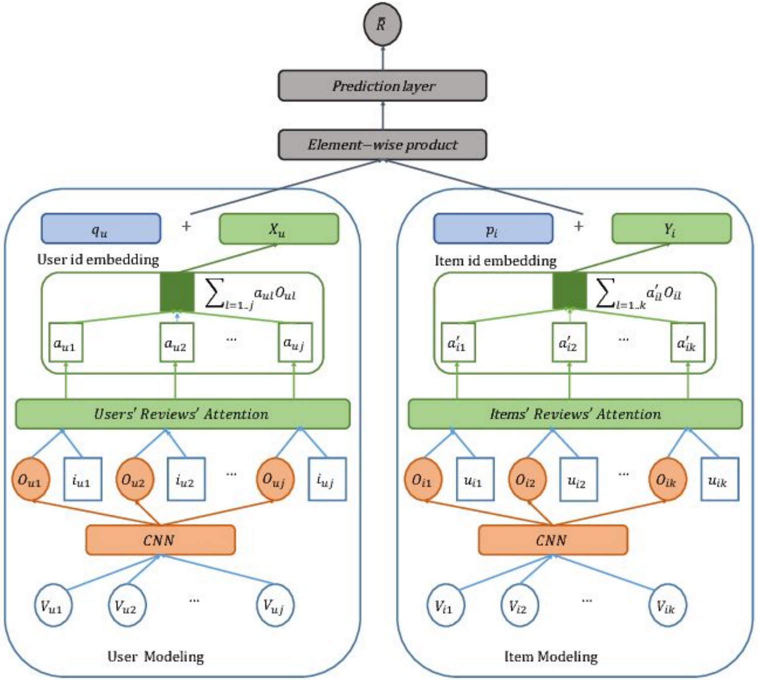


Figure 3: The neural network architecture of NARRE.

Usefulness of reviews

Table 4: Examples of the high-weight and low-weight reviews selected by our model (a_{ij} means attention weight).

Item 1	a (a_{ij}=0.1932)	These brushes are great quality for children's art work. They seem to last well and the bristles stay in place very well even with tough use.
	b (a_{ij}=0.0161)	I bought it for my daughter as a gift.
Item 2	a (a_{ij}=0.2143)	From beginning to end this book is a joy to read. Full of mystery, mayhem, and a bit of magic for good measure. Perfect flow with excellent writing and editing.
	b (a_{ij}=0.0319)	I like reading in my spare time, and I think this book is very suitable for me.

Table 5: Usefulness evaluation on Amazon datasets (taking rated usefulness of reviews as ground truth). **: p<0.01 in statistical significance test, compared to the best baseline.

	Toys_and_Games				Kindle_Store				Movies_and_TV			
	Latest	Random	Length	NARRE	Latest	Random	Length	NARRE	Latest	Random	Length	NARRE
Precision@1	0.1487	0.3255	0.2476	**0.3860****	0.2447	0.4574	0.4041	**0.5235****	0.3040	0.4908	0.3903	**0.6576****
Recall@1	0.0362	0.0952	0.0771	**0.1398****	0.0400	0.0992	0.0852	**0.1131****	0.0436	0.0976	0.0677	**0.1445****
Precision@10	0.1550	0.2000	0.2316	**0.2697****	0.2228	0.2707	0.2933	**0.3530****	0.2325	0.2925	0.3369	**0.3459****
Recall@10	0.4367	0.5763	0.6763	**0.8601****	0.4510	0.5551	0.6168	**0.8317****	0.3716	0.4673	0.5403	**0.7674****

A^{3}NCF: An Adaptive Aspect Attention Model for Rating Prediction

- Aspect aware recommendation
- For a user and an item Predict rating, as well as capture varying aspect attention user pays to item
- Topic model to extract user preference and item characteristics from review texts
- User preference and item characteristics used to guide latent user and item factors
- Attention network to capture user attention on each aspect of the targeted item

Cheng, Zhiyong, et al. A^3NCF: An Adaptive Aspect Attention Model for Rating Prediction IJCAI 2018.

A^{3}NCF

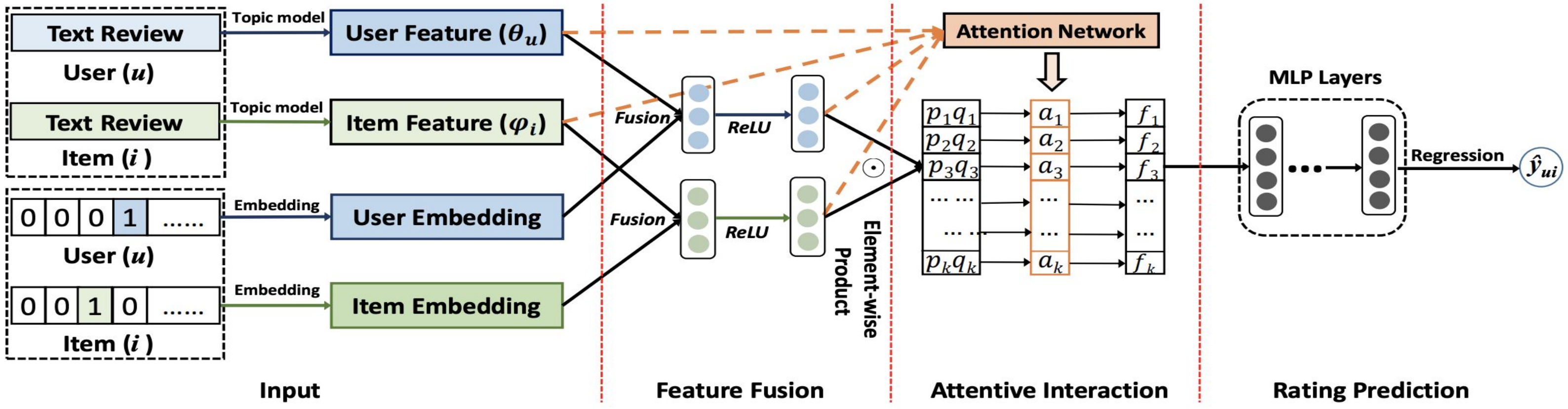


Figure 1: The structure of our A^{3}NCF model.

- Rating only reflects overall user satisfaction, but not the rationale
- Earlier techniques (MF,FM) can not achieve explainable fine-grained modeling
- Review contains rich information about user preference and item characteristics
- Topic model to extract user preference and item characteristics from reviews
- Aspect aware representation learning of users and items via Neural Network
- Adaptive aspect attention modelling via dedicated attention network

A^{3}NCF: Attention Network

- The attention network:

$$\alpha_{u,i} = v^t ReLU(W_a[\theta_u; \varphi_i; p_u; q_i] + b_a)$$

- Normalize: $$a_{u,i,k} = \frac{\exp(\alpha_{u,i,k})}{\sum_{j=1}^{K} \exp(\alpha_{u,i,j})}$$

- Attentive output representation:

$$F = a_{u,i} \odot (p_u \odot q_i)$$

- MLP:

$$z_L = \sigma_L(W_L(\sigma_{L-1}(W_{L-1}..\sigma_1(W_1 F + b_1)) + b_{L-1}) + b_L)$$

- Predictions: $\hat{r}_{u,i} = W z_L + b$

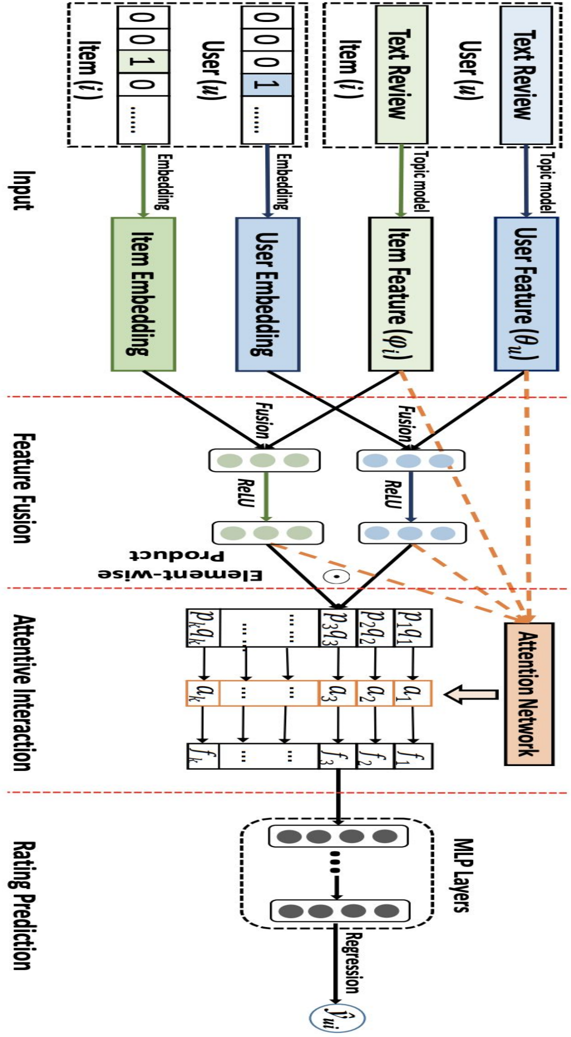


Figure 1: The structure of our A^{3}NCF model.

Flipkart

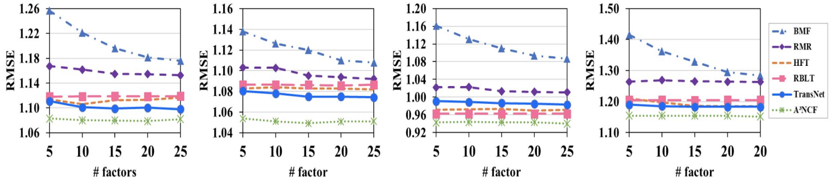
A^{3}NCF: Advantages of Aspects aware & Attention

Dataset	BMF	RMR	HFT	RBLT	TransNet	A^{3}NCF
Baby	1.176	1.152	1.117	1.119	1.098	**1.082***
Grocery	1.126	1.063	1.009	1.011	0.993	**0.985**
H & K	1.108	1.092	1.082	1.086	1.074	**1.051***
Garden	1.099	1.074	1.037	1.034	1.040	**1.021***
Sports	1.087	1.011	0.972	0.963	0.983	**0.940***
Yelp2017	1.284	1.263	1.185	1.204	1.183	**1.152***

Table 2: Comparisons of adopted methods in terms of RMSE with $K = 25$. The best performance is highlighted with bold face. The symbol * denotes a significance with $p - value < 0.05$ over the second best model based on a two-tailed paired t-test. "H & K" denotes "Home & Kitchen".

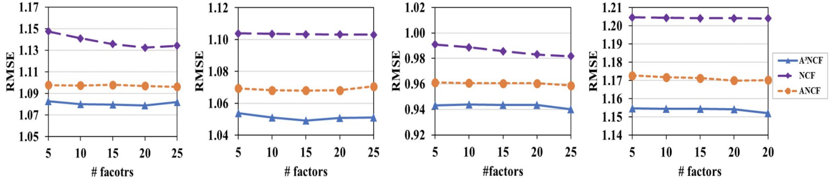
- A^{3}NCF outperforms state of the art methods: RBLT, TransNet
- Vanilla NCF is worst.
- Using user preference and item characteristics from Review text boost the results sharply (ACNF).
- Attention over user preferences and item characteristics outperforms ACNF .
- Improvements are more with the increase of # of latent factors.

Flipkart



(a) Baby (b) Home & Kitchen (c) Sports (d) Yelp 2017

Figure 3: Performance of all competitors *w.r.t.* the number of latent factors on four relatively larger datasets.



(a) Baby (b) Home & Kitchen (c) Sports (d) Yelp 2017

Figure 4: Performance of variants *w.r.t.* the number of latent factors on four larger datasets.

MRG: Multimodal Review Generation for Recommender System

- To predict a rating given a user and an item, as well as to generate review text
- Sensitize review generation to sentiment features based on a user and and item
- Improve the generation with review photos
- Useful reviews generation by using fusion gate as soft attention to weigh the sentiment and visual features

Truong QT,Lauw H. Multimodal Review Generation for Recommender System WWW 2019.

MRG

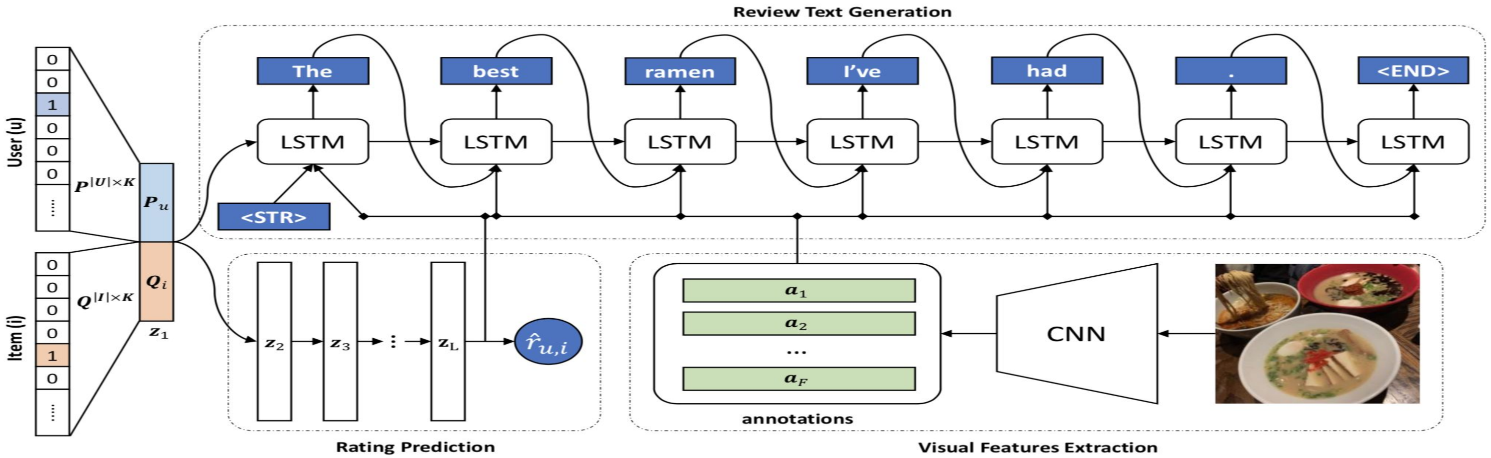
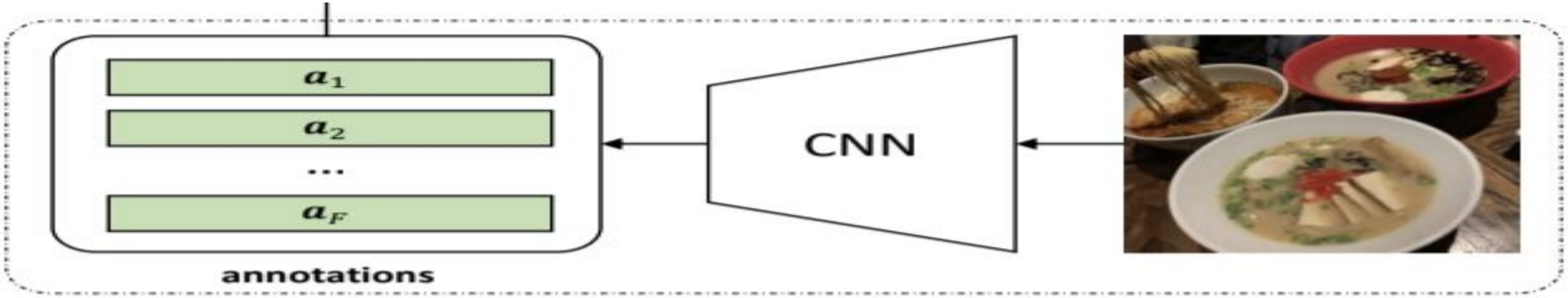


Figure 2: Overall Architecture of Multimodal Review Generation (MRG) model

- Rating Prediction using latent factor of user and item.
- Visual Feature Extraction using pretrained CNN
- Fusing visual features and latent factors using attention mechanism.
- Review Generation using fusion of visual and latent features.
- Training data consists of users, items and review photos.

MRG: Visual Features



Visual Features Extraction

- Get annotations from pretrained CNN.
- Each annotation a_j represents small region of image and contribute unequally.
- Apply soft attention to weight on more relevant annotations.
- Thus, for each time step t, get context vector v_t.

$$\mathbf{e}_{tj} = \phi(\mathbf{W}_{ea}\mathbf{a}_j + \mathbf{W}_{eh}\mathbf{h}_{t-1} + b_e) \tag{17}$$

$$\alpha_{tj} = \frac{\exp(\mathbf{W}_\alpha \mathbf{e}_{tj})}{\sum_{k=1}^{F} \exp(\mathbf{W}_\alpha \mathbf{e}_{tk})} \tag{18}$$

$$\mathbf{v}_t = \sum_{j=1}^{F} \alpha_{tj}\mathbf{a}_j \tag{19}$$

Flipkart

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
RECRUITMENT 2018

Fusion Gate as Soft Attention b/w Sentiment & Visual features

- To predict a rating given a user and an item, as well as to generate review text:

$$\gamma_t = \sigma(W_{\gamma y} E y_{t-1} + W_{\gamma h} h_{t-1} + b_\gamma)$$
$$s_t = \gamma_t z_L + (1 - \gamma_t) v_t$$

- z_L and v_t are the sentiment and visual features.

$$x_t = \begin{pmatrix} E y_{t-1} \\ s_t \end{pmatrix}$$

- x_t is the input to LSTM at time step t.

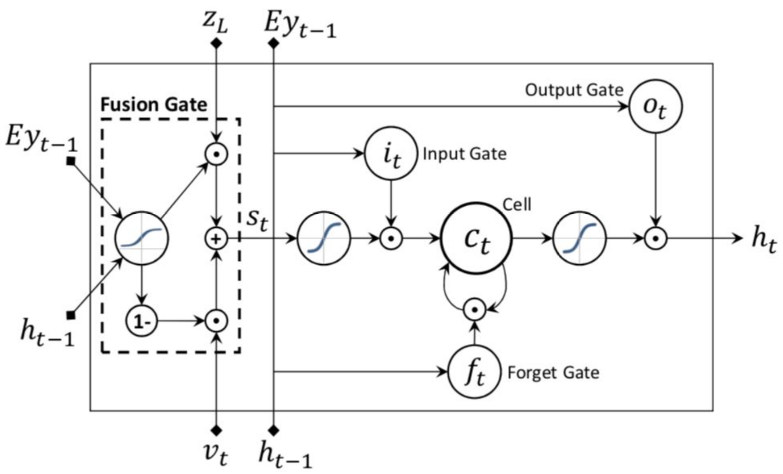


Figure 3: LSTM Cell with Fusion Gate

MRG: Advantages of Multimodal

Table 6: Ablation analysis: rating prediction performance (*lower is better*)
(The results are statistically significant with $p < 0.01$ based on the paired sample t-test)

	Dataset	Chicago			Los Angeles			New York			San Francisco		
	User + Item	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Text		✓	✓		✓	✓		✓	✓		✓	✓
	Photo			✓			✓			✓			✓
MAE	Training 80%	0.751	0.741	**0.729**	0.749	0.713	**0.710**	0.695	0.677	**0.673**	0.685	0.667	**0.664**
MAE	Training 60%	0.779	0.751	**0.734**	0.766	0.719	**0.716**	0.717	0.693	**0.683**	0.703	0.678	**0.671**
MAE	Training 40%	0.832	0.788	**0.767**	0.801	0.732	**0.727**	0.754	0.704	**0.696**	0.743	0.686	**0.678**
MAE	Training 20%	0.920	0.817	**0.791**	0.881	0.746	**0.738**	0.820	0.740	**0.721**	0.811	0.728	**0.708**
RMSE	Training 80%	1.006	0.952	**0.942**	0.985	0.936	**0.934**	0.912	0.881	**0.875**	0.889	0.857	**0.853**
RMSE	Training 60%	1.032	0.961	**0.945**	1.004	0.943	**0.938**	0.937	0.900	**0.887**	0.941	0.872	**0.864**
RMSE	Training 40%	1.083	1.010	**0.975**	1.043	0.957	**0.950**	0.981	0.910	**0.900**	0.989	0.879	**0.874**
RMSE	Training 20%	1.165	1.039	**1.003**	1.126	0.988	**0.968**	1.082	0.953	**0.926**	1.058	0.932	**0.905**

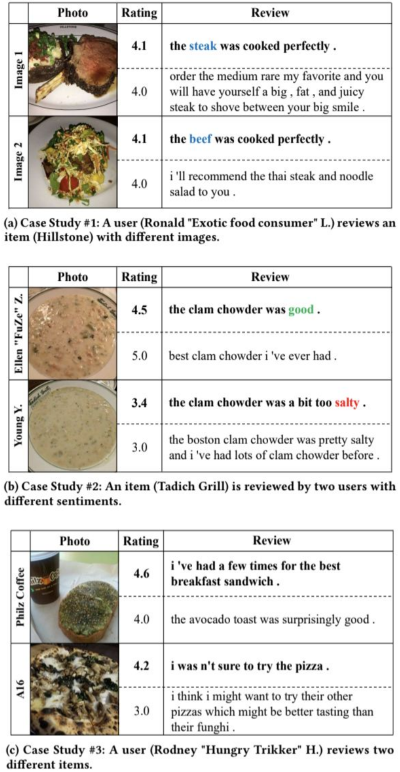


Figure 6: Multimodal review generation. The first line next to each photo (bold) is generated rating & text, and the second line is the ground truth. Photos are best seen in color.

- User + Item is worst.
- Review text generation boost the results sharply.
- Adding visual features improves the scores, as it provides alignment between words and sentiment features.
- Improvements are more clearer when data is sparse.

Flipkart

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
RECRUITMENT 2018