ELIXIR: Efficient and LIghtweight model for eXplaIning Recommendations

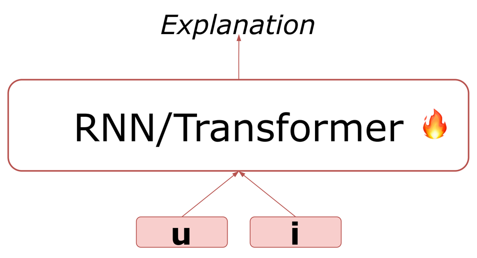
Ben Kabongo [1] Vincent Guigue [2] Pirmin Lemberger [3]

[1]Sorbonne University, CNRS, ISIR, MLIA [2]AgroParisTech, UMR MIA Paris-Saclay [3]Onepoint

Review-based Explainable Recommendation

- Users expect not only relevant recommendations but also transparent justifications → explanation required.
- **User reviews** = opinions on aspects of interest → making them a natural candidates for explanation.

Previous methods:



- Att2Seq [1] & NRT [4] (RNN-based) and PETER [2] (Transformer) fail to leverage the capabilities of pre-trained LLMs.
- PEPLER [3] (GPT-2, 124M) suffers from sub-optimal adaptation.
- Learning compact **2-token** user and item profiles to condition the entire generation.
- Neglect aspect modeling.

- → Fine-grained, faithful control of generation conditioned on user and item.

Problem Formulation

Notations

- $\mathcal{U}$ users, $\mathcal{I}$ items, $\mathcal{A}$ aspects of interest (e.g., quality of service, cleanliness, in the hotel domain).
- r_{ui} overall rating, r_{ui}^{a} aspect-level rating, t_{ui} review.
- $\mathcal{R} = \left\{ (u, i, r_{ui}, t_{ui}, \{r_{ui}^{a}\}_{a \in \mathcal{A}}) \right\}$ interaction data.

Objectives:

- **Recommendation**: Predict the overall rating r_{ui}.
- **Explainability**: Predict aspect ratings $\{r_{ui}^{a}\}_{a \in \mathcal{A}}$ and **generate review** t_{ui}.

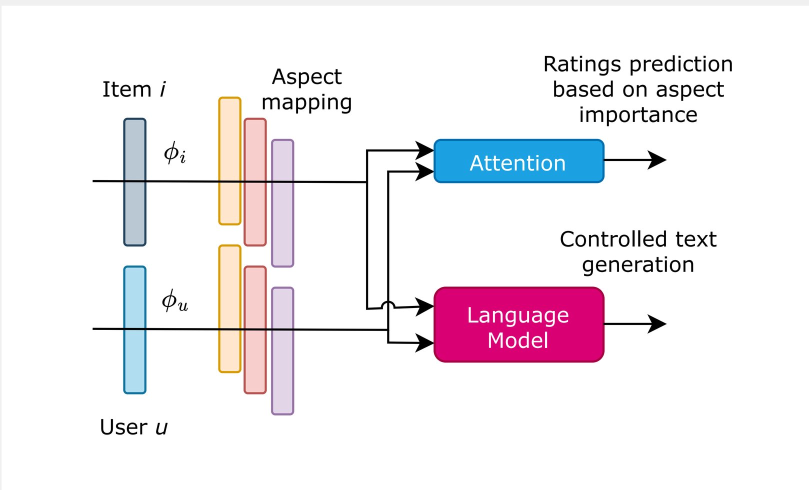


Figure 1. Schematic overview of our approach.

Our approach:

- Learning global-level and **aspect-level** user and item profiles.
- **Personalized attention** to estimate aspect importance.
- Two modules: **rating prediction** and **review generation**.

Hypothesis: Aspect information provides better guidance for review generation.

Review Generation Results

Table 1. Performance on review generation (%).

TripAdvisor	METEOR ↑	BLEU ↑	ROUGE-1 ↑	ROUGE-2 ↑	BERT-F1 ↑
Att2Seq	18.611	04.690	28.783	06.473	84.490
NRT	17.219	03.405	25.833	05.194	82.161
PETER	17.955	03.943	27.974	05.906	84.406
PEPLER	24.340	11.400	33.831	11.679	83.726
ELIXIR	**42.752**	**33.544**	**53.285**	**37.878**	**89.554**
w/o Aspects	27.642	10.029	39.076	21.970	86.540
RateBeer	**METEOR ↑**	**BLEU ↑**	**ROUGE-1 ↑**	**ROUGE-2 ↑**	**BERT-F1 ↑**
Att2Seq	18.611	04.690	28.783	06.473	84.490
NRT	24.963	08.737	32.589	11.472	83.985
PETER	28.818	11.518	35.504	13.620	86.448
PEPLER	28.266	10.143	32.444	11.182	84.990
ELIXIR	**40.763**	**24.160**	**46.371**	**25.818**	**89.792**
w/o Aspects	32.675	13.652	39.068	17.106	88.310

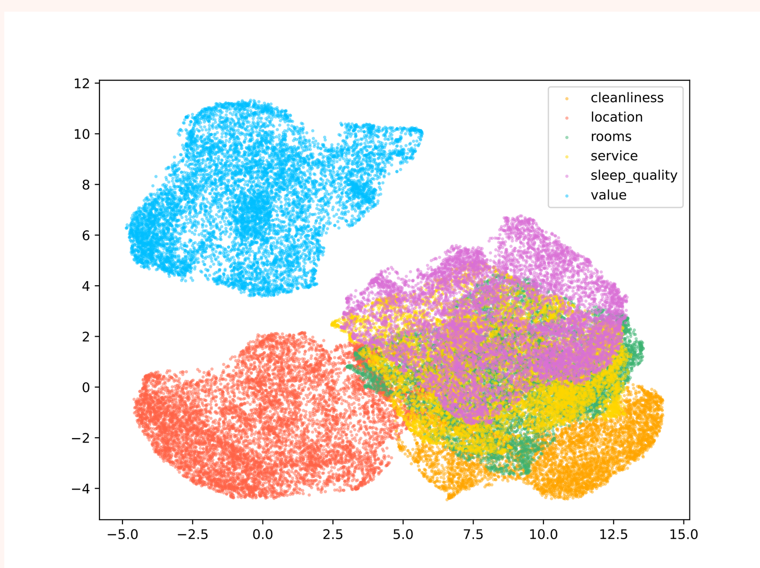
- ELIXIR achieves better scores than all baseline models across all considered metrics.
- **Efficiency of our lightweight architecture**: ELIXIR w/o Aspects outperforms all baselines.
- **Importance of aspect integration**: ELIXIR performs better than its aspect-ablated version.

Aspect Analysis and Impact of the Number of Tokens

Table 2. Impact of the number of tokens in the personalized prompt (η) on review generation (TripAdvisor dataset).

η	METEOR ↑	BLEU ↑	ROUGE-2 ↑
PEPLER	24.340	11.400	11.679
2	12.159	01.282	04.436
5	16.730	03.546	06.046
10	21.160	07.692	10.330
20	29.379	17.018	20.200
50	**42.752**	**33.544**	**37.878**

Figure 2. Projection and clustering of user aspect representations for the TripAdvisor dataset.



- **Impact of the number of tokens**: Starting from 20 tokens, ELIXIR outperforms PEPLER.
- **Aspect analysis**: We observe a consistent semantic separation of aspects.
- Aspect integration leads to performance improvements across all tasks.

ELIXIR

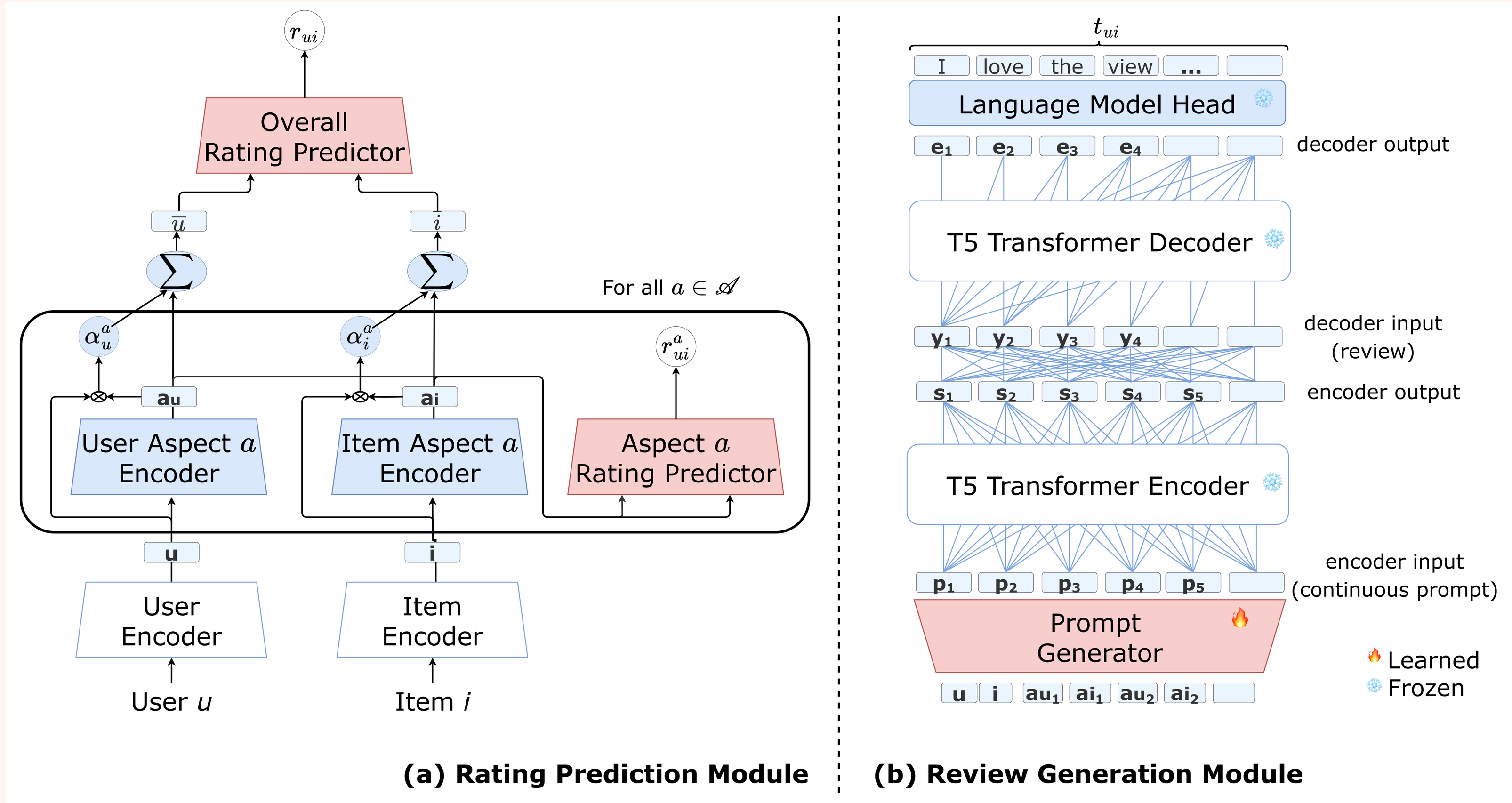


Figure 3. ELIXIR consists of two modules: (a) the rating prediction module on the left and (b) the personalized review generation module on the right.

- **Global-Level Representations**: $\mathbf{u}, \mathbf{i} \in \mathbb{R}^d$
- **Aspect-Level Representations**: $\mathbf{a}_u = \phi_{\mathcal{U}}^{a}(\mathbf{u})$, $\mathbf{a}_i = \phi_{\mathcal{I}}^{a}(\mathbf{i})$, $\phi_{*}^{a} : \mathbb{R}^d \to \mathbb{R}^d$.
- **Personalized Attention**:

$$\alpha_u^a = \frac{\exp(q_{\mathcal{U}}(\mathbf{u}) \cdot k_{\mathcal{U}}(\mathbf{a}_u))}{Z_u}, \quad \tilde{\mathbf{u}} = \sum_{a \in \mathcal{A}} \alpha_u^a v_{\mathcal{U}}(\mathbf{a}_u).$$

Similar for $\tilde{\mathbf{i}} \in \mathbb{R}^d$.
$q_*, k_*, v_* : \mathbb{R}^d \to \mathbb{R}^d$ (linear functions); Z_* (normalization terms).

- **Rating Prediction Module**: overall + aspect-level ratings. $\hat{r}_{ui} = f(\tilde{\mathbf{u}}, \tilde{\mathbf{i}})$, $\hat{r}_{ui}^a = g_a(\mathbf{a}_u, \mathbf{a}_i)$
- **Review Generation Module**: personalized prompt + review.

$$\mathbf{p}_{ui} = \psi(\mathbf{u}, \mathbf{i}, \{\mathbf{a}_u, \mathbf{a}_i\}_{a \in \mathcal{A}}) \in \mathbb{R}^{\eta \times d_w}, \qquad P_{\theta_P, \theta_{LM}}(t_{ui} | \mathbf{p}_{ui}) = \prod_{k=1}^{|t_{ui}|} P_{\theta_P, \theta_{LM}}(y_k | \mathbf{p}_{ui}, y_{<k}).$$

η: number of prompt tokens; θ_P: prompt parameters; θ_{LM}: language model parameters.
Parameter-efficient fine-tuning: We freeze the language model parameters during training and only optimize the prompt generation parameters.
Lightweight model: We use T5-Small (60M) as language model.

Overall Rating Results

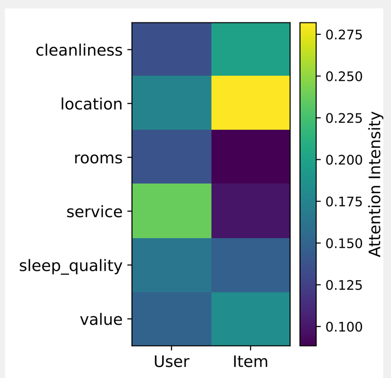
Table 3. Performance on overall rating prediction.

	Average	MF	MLP	NeuMF	NRT	PETER	PEPLER	**ELIXIR**	*w/o Attn*	*Global-only*
TripAdvisor										
RMSE ↓	0.932	0.840	0.833	0.840	0.859	0.807	0.779	**0.748**	0.771	0.865
MAE ↓	0.645	0.646	0.565	0.570	0.548	0.532	0.478	**0.447**	0.513	0.632
RateBeer										
RMSE ↓	0.571	**0.411**	0.464	0.473	0.420	0.415	0.430	0.416	0.421	0.443
MAE ↓	0.424	**0.300**	0.324	0.329	0.306	**0.300**	0.305	0.305	0.311	0.332

- ELIXIR w/o Attention > ELIXIR Global-only: Learning aspect-level representations in addition to global representations to better capture fine-grained preferences.
- ELIXIR > ELIXIR w/o Attention: Personalized attention allows for better aggregation of aspect information.

Personalized Attention Analysis

Table 4. Attention visualization on aspects and alignment with the review (TripAdvisor)



Aspect	Rating	Ground truth review
Cleanliness	4.9 (5.0)	if we go back to paris, we are staying here again. the place is so charming and
Location	5.0 (5.0)	overlooks the beautiful luxembourg gardens. the staff were sooo hospitiable.
Rooms	5.0 (5.0)	always asking what they could do to help us. they arranged two tours for us,
Service	5.0 (5.0)	recommended places to eat and then made the reservations for us, arranged
Sleep	5.0 (5.0)	transportation from and to the airport, etc. royce and xavier, i can't thank you
Value	5.0 (5.0)	enough! also, so many places are in walking distance, like notre dame and the
Overall	4.9 (5.0)	louvre. you can't help but fall in love with this place!

- **Personalized attention** infers the relative importance of aspects for each user–item pair, also enabling explainable recommendations.

References

[1] Dong, L., Huang, S., Wei, F., Lapata, M., Zhou, M., and Xu, K.
Learning to generate product reviews from attributes.
In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (2017), pp. 623–632.

[2] Li, L., Zhang, Y., and Chen, L.
Personalized transformer for explainable recommendation.
arXiv preprint arXiv:2105.11601 (2021).

[3] Li, L., Zhang, Y., and Chen, L.
Personalized prompt learning for explainable recommendation.
ACM Transactions on Information Systems 41, 4 (2023), 1–26.

[4] Li, P., Wang, Z., Ren, Z., Bing, L., and Lam, W.
Neural rating regression with abstractive tips generation for recommendation.
In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval* (2017), pp. 345–354.

Our paper

https://github.com/BenKabongo25/aspect_explainable_recommender

ben.kabongo@sorbonne-universite.fr