

Query-aware Tip Generation for Vertical Search

Yang Yang*, Junmei Hao*, Canjia Li*, Zili Wang*(Equal Contribution), Jingang Wang, Fuzheng Zhang, Rao Fu, Peixu Hou, Gong Zhang, Zhongyuan Wang



- **Introduction**
- **Query-aware Tip Generation Framework**
- **Experiments**
- **Conclusion**

1

Introduction

Tip Generation: A **concise** form of user reviews which have unique advantages to **explain** search results, assist decision making, and further **improve user experience** in vertical search scenarios.



RELATED WORK

- Query-focused Summarization
QFS aims to summarize a document cluster in response to a specific user query or topic.
- Constrained Sentence Generation
Consider external information during generation.

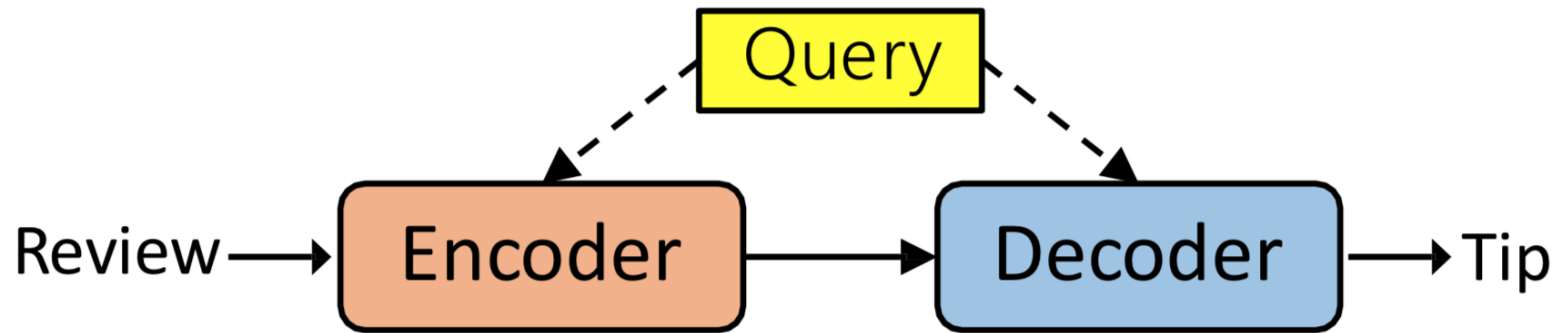
To take the **query impact** into account, this paper proposes query-aware tip generation for vertical search.

We consider query information in both encoder and decoder sides to generate query-aware tips, that are **intuitive but effective** and **of great business values** in vertical search scenarios.

2

Query-aware Tip Generation Framework

Query-aware tip generation framework



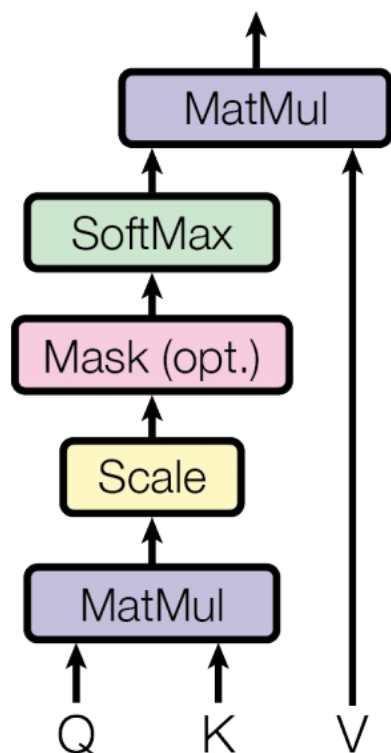
Query-aware Encoder ($Q_A_E_{NC}$).

- Transformer: incorporate the query representation into the self-attention computation
- RNN: introduce a selective gate network in the encoder to distill query-relevant information from the input sequence.

Query-aware Decoder ($Q_A_D_{EC}$).

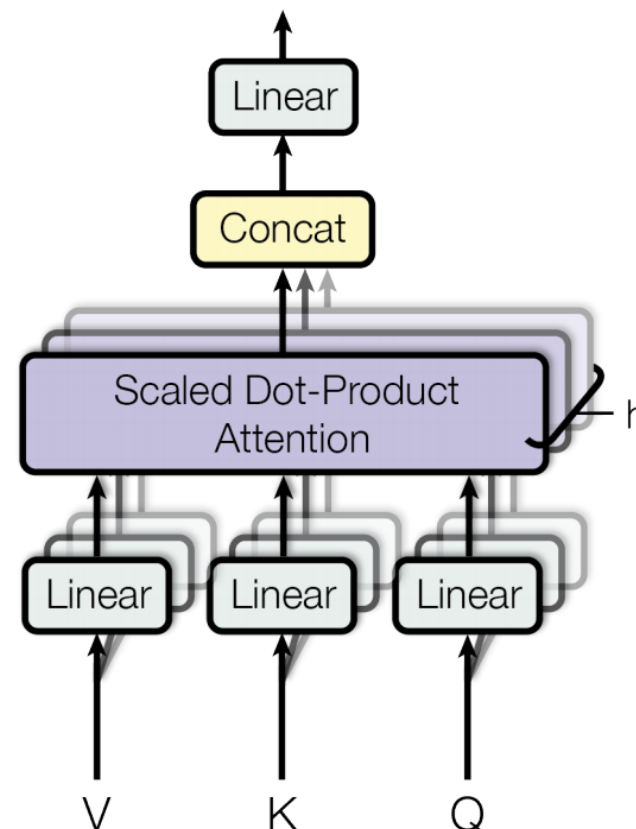
- Transformer: incorporate the query representation into the self-attention computation as the encoder.
- RNN: improve the attention mechanism by integrating query representation into the context vector to better direct the decoder.

Self-attention Scaled Dot-Product Attention



$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Multi-Head Attention



$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$

where $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$

Methodology - Transformer-based Adaptation

(1) Review-aware Query Encoder

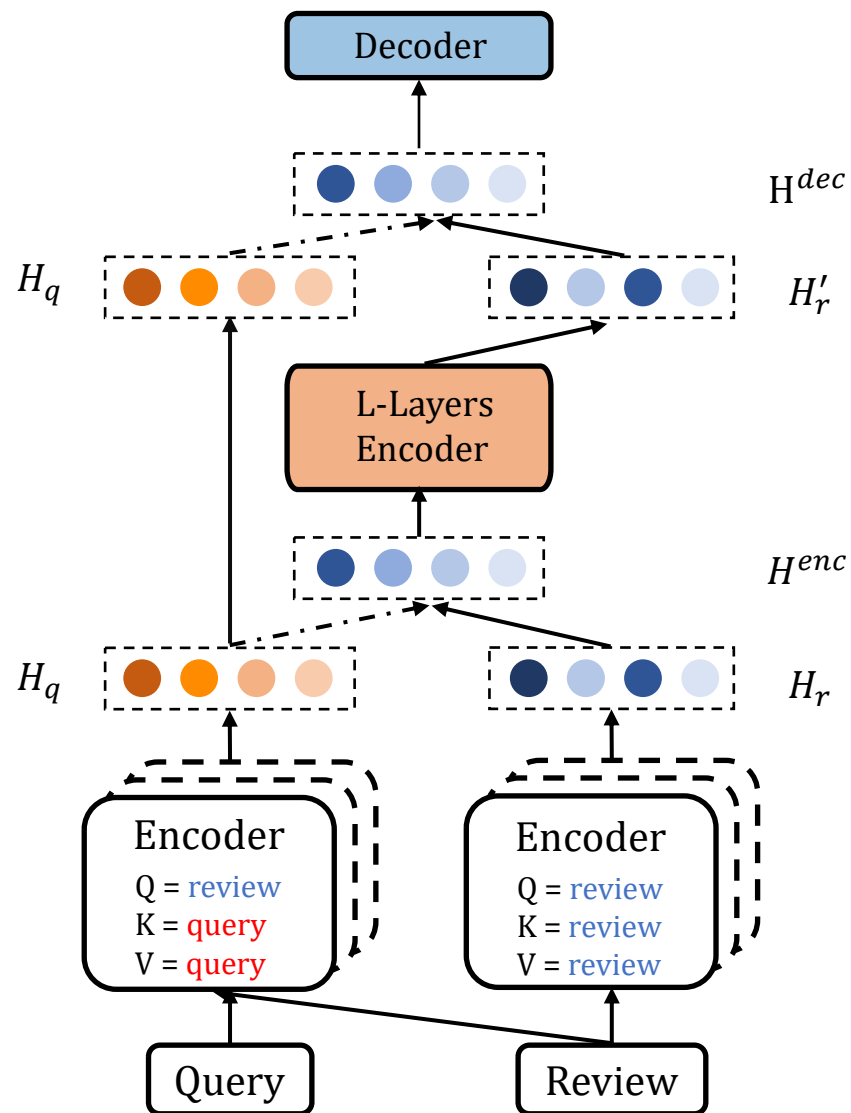
$$H = \text{MultiHead}(E_r, E_q, E_q)$$

(2) Query-aware Review Encoder

$$H^{\text{enc}} = [H_q; H_r]W$$

(3) Query-aware Tip Decoder

$$H^{\text{dec}} = [H_q; H'_r]W$$



Methodology - RNN-based Adaptation

(1) Query Encoder

(2) Query-Selective Encoder

$$g_t = \sigma(W_r [H_t; h_r] + W_q h_q + b_g)$$

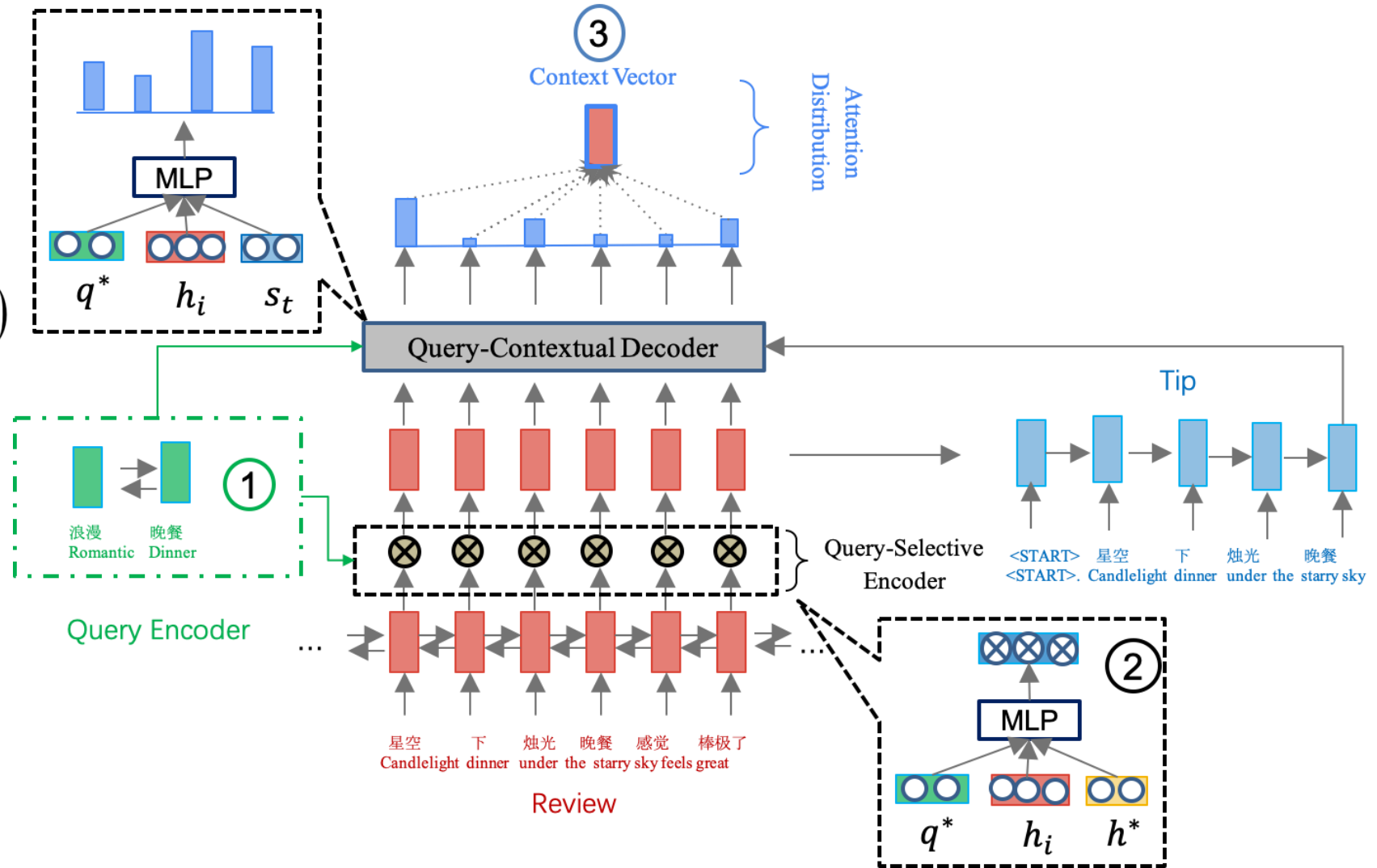
$$\tilde{H}_t = g_t \circ H_t$$

(3) Query-Contextual Decoder

$$c_i = W_c [\tilde{H}_i; s_t] + W_q h_q + b_c$$

$$a_i^t = \text{softmax}(v^T \tanh(c_i))$$

$$h_t^{\text{dec}} = \sum_i a_i^t \tilde{H}_i$$



3

Experiments

- Debate: An open-source query-based English summarization dataset which is created from Debatepedia.

Document	The “natural death” alternative to euthanasia is not keeping someone alive via life support until they die on life support. That would, indeed, be unnatural. The natural alternative is, instead, to allow them to die off of life support.
Query	"non-treatment" : Is euthanasia better than withdrawing life support?
Summary	The alternative to euthanasia is a natural death without life support.

- Dianping: crawled from the search log of the Dianping App.

POI	J Prime牛排海鲜餐厅	J Prime Steak & Seafood Restaurant
Review	果然是顶级牛排，口味超赞！前餐例汤也很有特色，无酒精鸡尾酒很解腻服务非常周到，每道菜细心讲解，吃的放心，环境也特别美，星空下烛光晚餐感觉棒极了朋友很满意！	It was indeed a top-grade steak, and its taste was superb! The soup before the meal was very special. The non-alcoholic cocktail really rid myself of the greasy feeling. The service was very satisfactory, and each dish was explained carefully. The environment is also very beautiful, the candlelight dinner under the stars is fantastic, my friends are very satisfied!
Query	浪漫 晚餐	Romance Dinner
Tip	星空下烛光晚餐感觉棒极了	The candlelight dinner under the stars is fantastic

- Statistics of both datasets

Dataset	Avg_Len			Query	Train	Valid	Test
	Review (Doc)	Query	Tip (Summary)				
DEBATE	72.61	11.54	9.93	w/o	10,846	1,356	1,356
				w/	10,975	1,372	1,372
DIANPING	101.50	3.39	12.20	w/o	137,208	17,151	17,151
				w/	179,784	22,473	22,473

- **Query_LEAD.** Taking the leading sentence(s) of a document is reported to be a strong baseline in summarization.

Extractive baseline

- **Extract_BM25.** Given the query, the sentences in the review are ranked by their BM25 scores and the top one is favored.
- **Extract_Embed.** The sentences in the review are ranked by their embedding-based cosine similarities with the query.

Abstractive baseline

- **RNN.** An abstractive baseline utilizing the pointer generator implementation, regardless of the query.
- **Trans.** An abstractive baseline utilizing the Transformer-based encoder-decoder implementation, regardless of the query.
- **RNN(Trans) + Qa_Enc/Qa_Dec/Both.** The RNN(Trans) with the proposed Qa_Enc and Qa_Dec separately or jointly.

Automatic Evaluation

- **Semantic** : cosine similarities between the embeddings of the generated tips and the corresponding queries.
- **Lexicon** : the number of co-occurring tokens in the generated tip and query divided by the query length is used as a lexical proxy of the relevance.
- **BLEU** : measures the coherency

Manual Evaluation

- **Readability** measures whether a generated tip is fluent and grammatical.
- **Relevance** indicates whether a generated tip is relevant to the query.
- **Usefulness** demonstrates whether the generate tip is helpful for the user to make a decision.

- Automatic Evaluation Results

Group	Methods	DEBATE			DIANPING		
		Semantic	Lexicon	BLEU	Semantic	Lexicon	BLEU
RETRIEVAL	QUERY_LEAD	-	10.23	2.23	-	40.70	23.20
	EXTRACT_BM25	-	14.39	1.12	-	47.18	27.59
	EXTRACT_EMBED	-	14.43	1.13	-	37.04	28.29
RNN	RNN	83.87	8.91	11.02	60.08	40.94	40.74
	RNN + QA_ENC	84.37	9.23	15.72	62.65	41.37	48.29
	RNN + QA_DEC	84.17	9.07	15.37	62.77	43.92	46.88
	RNN + BOTH	84.43	9.34	16.58	64.86	44.11	48.38
TRANSFORMER	TRANS	87.17	10.52	30.41	65.64	47.49	48.71
	TRANS + QA_ENC	86.07	13.17	32.03	67.00	49.79	50.39
	TRANS + QA_DEC	84.70	13.46	32.52	62.70	42.61	52.66
	TRANS + BOTH	88.06	13.43	32.93	69.79	53.75	54.20

- Manual Evaluation Results

	Methods	Read.	Rel.	Useful.
RNN	RNN	2.12	32.00%	43.56%
	RNN + QA_ENC	2.67	40.33%	43.13%
	RNN + QA_DEC	2.71	52.31%	41.49%
	RNN + BOTH	2.63	53.50%	44.50%
Transformer	TRANS	2.55	39.25%	44.38%
	TRANS + QA_ENC	2.88	60.52%	47.36%
	TRANS + QA_DEC	2.80	60.53%	47.37%
	TRANS + BOTH	2.88	63.72%	54.35%

Examples of tip generation from Dianping dataset

POI	Nana蛋糕手工烘焙	Nana's cake handmade baking
Review	公司周年庆吃到的。从外观来说，不错，看得出是裱花的技术还是很好的。因为早上是想9点半拿到，但是因为堵车的原因，晚了一点，这里老板很用心，会提前跟我说，让我们不要着急。另外在味道很不错，我比较喜欢吃草莓那边的，巧克力好吃，哈哈	Eaten at the company's anniversary celebration. From the appearance, it is good, it can be seen that the technology of decorating is very good. I wanted to get it at 9:30 in the morning, but cause of the traffic jam, I was a bit late. The boss is very attentive and will tell me in advance do not worry about it. In addition, it tastes very good. I prefer to eat strawberry flavor. The chocolate flavor is delicious, haha.
Query	蛋糕	cake
Tip (Query_LEAD)	公司周年庆吃到的。	Eaten at the company's anniversary celebration.
Tip (RNN)	裱花的技术还是很好的	The technology of decorating is very good.
Tip (Transformer)	这里老板很不错	The boss is very nice.
Tip (RNN-Both)	蛋糕很好吃	The cake tastes very good.
Tip (Transformer-Both)	这里老板很用心蛋糕很松软	The boss is very nice and the cake tastes very fluffy.

The A/B testing system diverts 10% total query traffic and splits it equally into 4 separate buckets.

4 buckets

- a) No tip is displayed with POIs
- b) The tips are generated by Trans
- c) The tips are generated by Extract_BM25
- d) The tips are generated by Trans + Both.

$$CTR = \frac{\#Clicks_in_SRP}{\#Query}$$

a	b	c	d
65.72%	65.74%	65.77%	65.80%

4

Conclusion

Task: Query-aware tip generation for vertical search

Framework: an intuitive and effective query-aware tip generation framework

Methodology: Two specific adaptations for the Transformer and the RNN architectures are developed.

Experiments:

- Extensive experiments on both public and realistic datasets reveal the effectiveness of our proposed approach.
- The online deployment experiments on Dianping demonstrate the promising business value of the query-aware tip generation framework.

Thanks
Q&A