Improving Search Snippets in Context-aware Web Search Scenarios

Jia Chen, Jiaxin Mao, Yiqun Liu*, Min Zhang, Shaoping Ma

Department of Computer Science and Technology, Tsinghua University

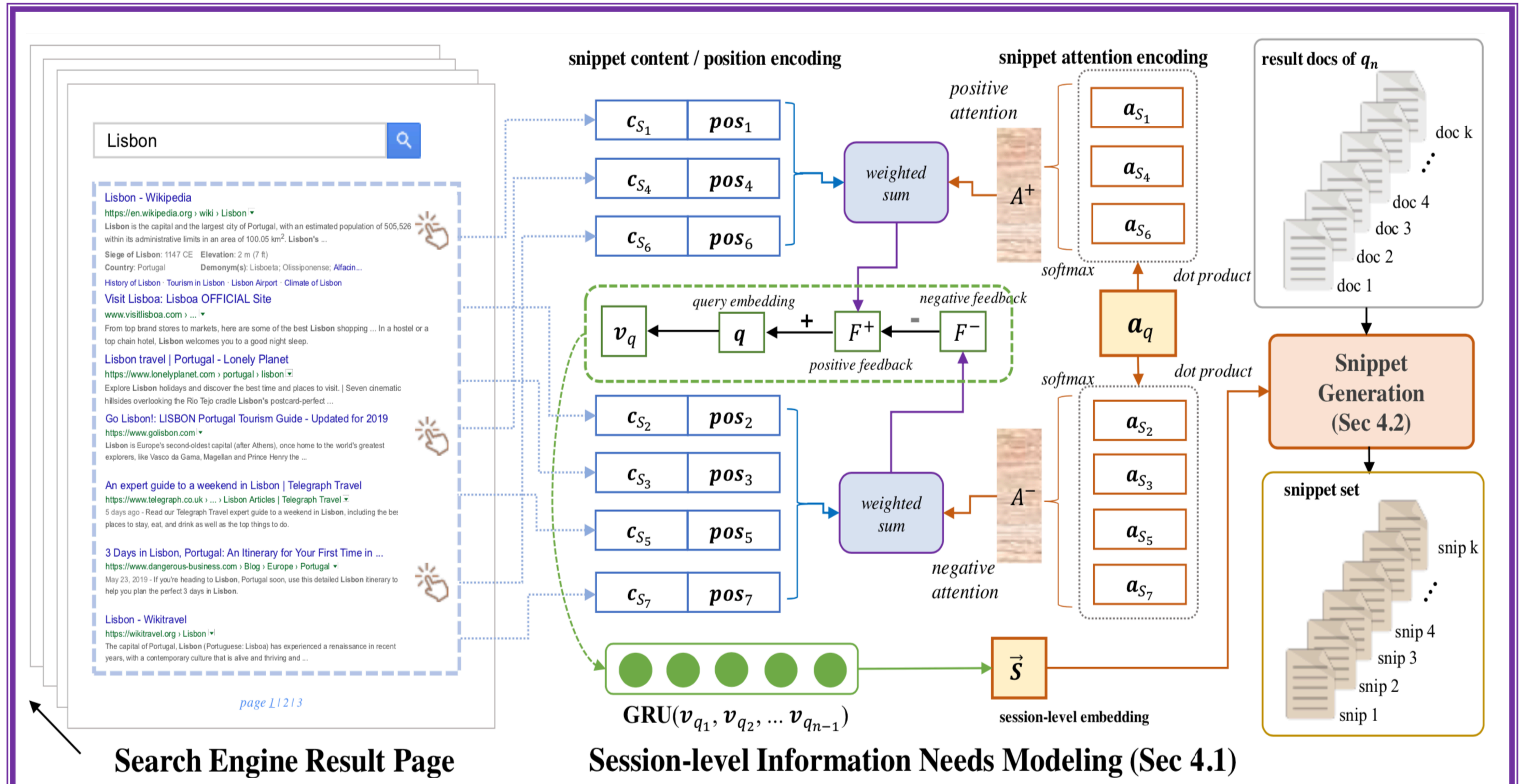
*yiqunliu@tsinghua.edu.cn.

天工研究院

论文摘要

As an essential part in web search, search snippets usually provide result previews for users to either gather useful information or make click-through decisions. In complex search scenarios, users may need to submit multiple queries to search systems until their information needs are satisfied. As user intents tend to be ambiguous, incorporating contextual information for user modeling has been proved effective in many session-level tasks. Therefore, the generation of search snippets may also benefit from the integration of context information. However, to our best knowledge, most existing snippet generation methods ignore user interaction and focus merely on the query content. Whether it is useful of exploiting session contexts to improve search snippets still remains inscrutable. To this end, we propose a snippet generation model which considers session contexts. The proposed method utilizes the query sequence as well as users' interaction behaviors within a session to model users' session-level information needs. We also adopt practical log-based search data to evaluate the performance of the proposed method. Experiment results based on both expert annotation and user preference test show the effectiveness of considering contextual information in search snippet generation.

系统模型



算法原理

snippet encoding

$$\mathcal{S}^+ = \{\mathcal{S}_p | p \in \mathcal{C}\}, \mathcal{S}^- = \{\mathcal{S}_p | p \le max(\mathcal{C}) + 1, p \notin \mathcal{C}\}$$

$$\mathbf{w}_j^p = \mathbf{E}_c^T \cdot \mathbf{x}_j^p,$$

$$\mathbf{c}_{\mathcal{S}_p} = \mathrm{GRU}_c(\mathbf{w}_1^p, ..., \mathbf{w}_j^p ..., \mathbf{w}_{|\mathcal{S}_p|}^p),$$

attention mechanism

$$\mathbf{w}_j^q = \mathbf{E}_q^T \cdot \mathbf{x}_j^q,$$

$$\mathbf{w}_j^p = \mathbf{E}_{a_{\mathcal{S}}}^T \cdot \mathbf{x}_j^p,$$

$$\mathbf{a}_q = \mathrm{GRU}_q(\mathbf{w}_1^q, ..., \mathbf{w}_j^q ..., \mathbf{w}_{|q|}^q),$$

$$\mathbf{a}_{\mathcal{S}_p} = \mathrm{GRU}_{a_{\mathcal{S}}}(\mathbf{w}_1^p, ..., \mathbf{w}_j^p ..., \mathbf{w}_{|\mathcal{S}_p|}^p).$$

$$\mathcal{A}_{\mathcal{S}}^{+/-} = softmax(\mathbf{a}_q^T \mathbf{a}_{\mathcal{S}_p} | \mathcal{S}_p \in \mathcal{S}^{+/-}).$$

feedback memory mechanism

$$\mathcal{F}^{+/-} = \sum_{\mathcal{S}_p \in \mathcal{S}^{+/-}} \mathcal{A}_{\mathcal{S}}^{+/-}(\mathbf{W}[\mathbf{c}_{\mathcal{S}_p} \oplus \mathbf{pos}_p] + \mathbf{b}_F),$$

$$\mathbf{v}_{q_k} = \mathbf{q}_k + \mathcal{F}^+ - \mathcal{F}^-,$$

$$\boldsymbol{S} = \mathrm{GRU}_s(\mathbf{v}_{q_1}, ...\mathbf{v}_{q_k}, ...\mathbf{v}_{q_K}).$$

snippet generation

$$\log \mathcal{L}(M) = \sum_{x_j \in q_s} \log P_M(x_j | x_1 : x_{j-1}, \boldsymbol{S}).$$

$$score(s) = \prod_{j, x_j \in s} P(x_j | x_1 : x_{j-1}, \boldsymbol{S}).$$

$$Snippet_{ss} = \arg\max_{\Omega'} \sum_{s \in \Omega'} S_{ss}(s), \;\; s.t. \sum_{s \in \Omega'} |s| \le L, s \in \Omega,$$

$$S_{ss}(s) = score(s) + \sum_{i=1, x_i \in q_s}^{n} N(x_i) \cdot TFIDF(x_i),$$

实验结果

- **Snippet quality evaluation**

Table 2: Intrinsic evaluation results of each model. All values are $F_{1.2}$-scores, where $*$ and $\dagger$ indicate a statistically significant improvement over the strongest baseline *SE* at $p < 0.05/0.01$ level, respectively.

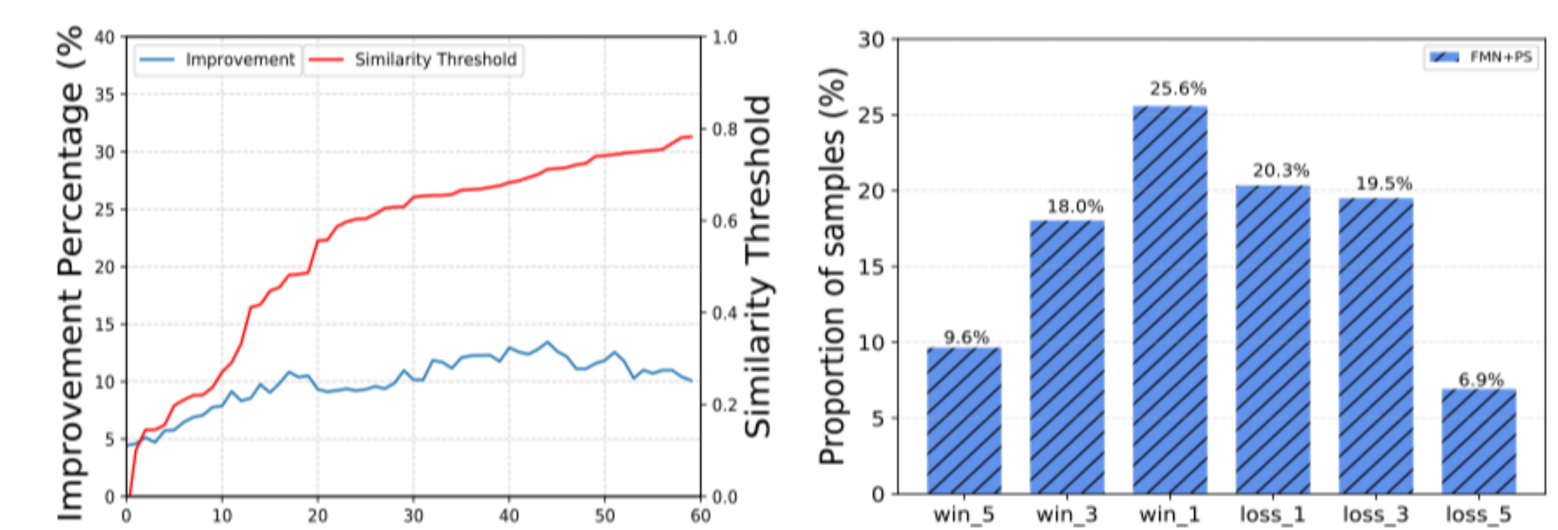
Sys.\Met.	RG-1	RG-2	RG-3	RG-4	RG-L	RG-W	RG-S	RG-SU
Luhn	0.236	0.167	0.148	0.136	0.215	0.101	0.119	0.122
KL-Sum	0.239	0.1663	0.147	0.136	0.215	0.101	0.118	0.122
Sumbasic	0.240	0.169	0.149	0.137	0.218	0.103	0.120	0.124
TextRank	0.242	0.171	0.151	0.139	0.220	0.103	0.122	0.125
PRF	0.361	0.209	0.151	0.111	0.309	0.120	0.155	0.161
SE(Sogou)	0.361	0.204	0.161	0.138	0.298	0.130	0.149	0.154
FMN_{NS}	0.352	0.216	0.185*	0.167†	0.283	0.131	0.157	0.161
FMN_{DS}	0.353	0.218	0.189*	0.172†	0.291	0.135	0.160	0.165
FMN_{SS}	**0.385** +6.48%	**0.259**† +26.54%	**0.226**† +40.15%	**0.205**† +48.34%	**0.314** +5.36%	**0.148*** +13.25%	**0.183**† +22.68%	**0.188**† +21.73%

⋆⋆ Note that RG is short for ROUGE. Here we only present the results of four systems with highest performances in SUMY: *Luhn*, *KL-Sum*, *SumBasic* and *TextRank*.

- **User preference test**

Table 4: The average proportion and boolean scores of *SE* and FMN_{SS} in all and top 60% sessions conditions (* indicates a statistical significant improvement over *SE* with an independent t-test at $p < 0.01$, and σ^2 denotes the variance).

Score\Model	*SE*	FMN_{SS}	σ^2
Boolean(All)	0.4779	**0.5221**	0.2495
Percentage(All)	0.4804	**0.5196***	0.0825
Boolean(Top 60%)	0.4412	**0.5588***	0.2465
Percentage(Top 60%)	0.4683	**0.5317***	0.0800



(a) System improvement across various filtering conditions. (b) User preference distribution on FMN_{SS}.

论文结论

- **Firstly, compared to existing snippet generation methods which mainly rely on exact matching with query terms, our methods are better at capturing semantic features.**
- **Secondly, the model utilizes previous queries and clicks to model users' session-level information needs, which can boost its performance in multi-query sessions.**
- **Last but not least, the proposed model has a low inference latency and needs not any human labels to train thus can be easily adopted in commercial search engines.**
- **since our model is feedback-based, it may not adapt to sessions with dramatic intent shifts. Intent detection and search task identification methods should be explored for further improvement for context-aware snippets**