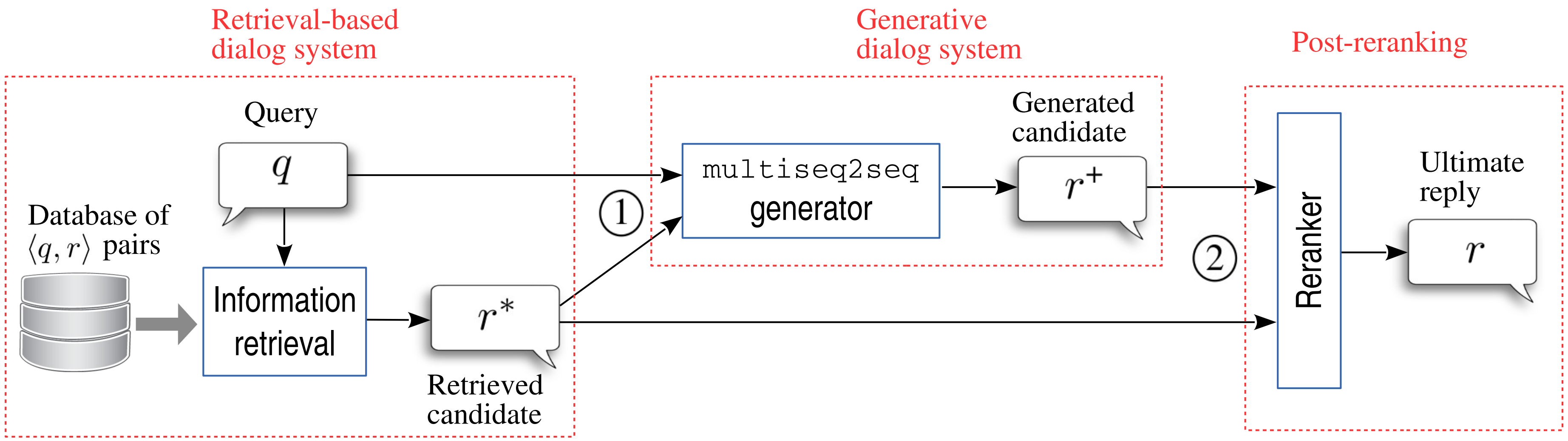
# Combine IR and Generative models

Jcykcai

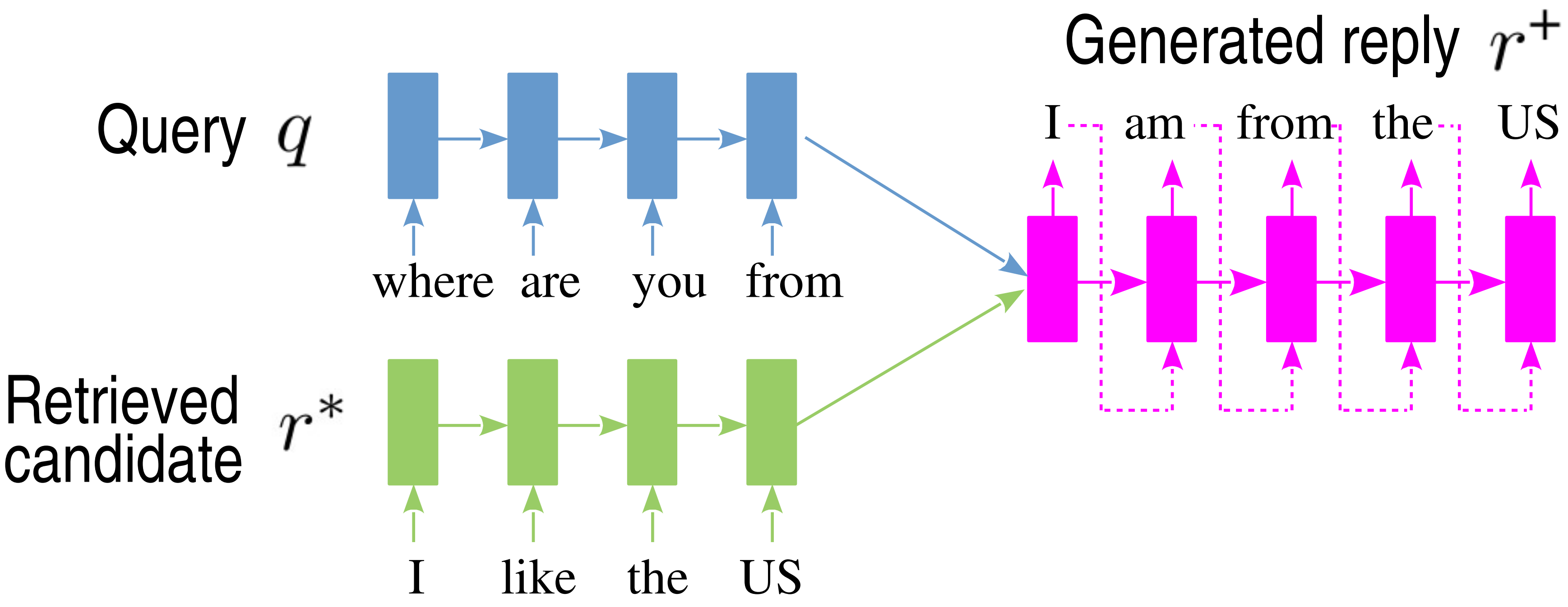
# Two are Better than One: An Ensemble of Retrieval-and Generation-Based Dialog Systems

*Yiping Song,*[1] *Rui Yan,*[2] *Xiang Li,*[1] *Dongyan Zhao,*[2] *Ming Zhang*[1]
[1] School of EECS, Peking University, China
[2] Institute of Computer Science and Technology, Peking University, China
{songyiping, ruiyan, lixiang.eecs, zhaody, mzhang_cs }@pku.edu.cn

# Two are Better than One: An Ensemble of Retrieval-and Generation-Based Dialog Systems



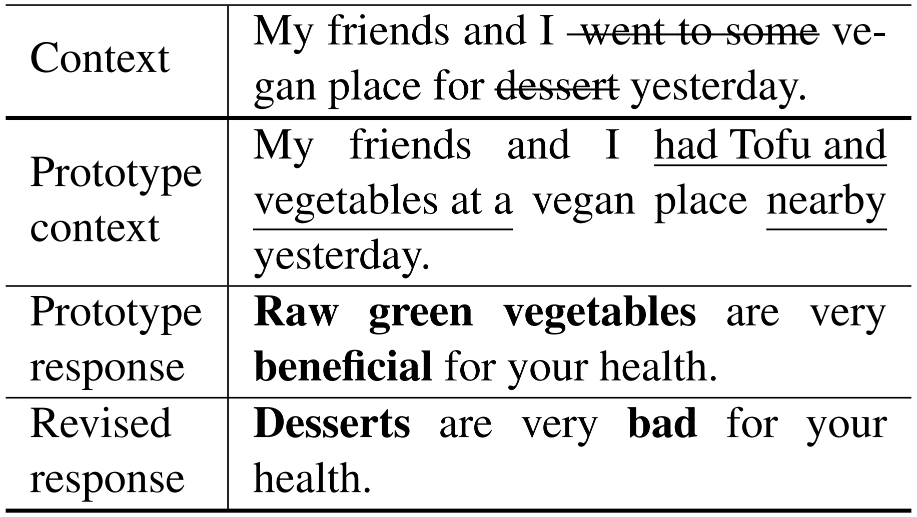
# Two are Better than One: An Ensemble of Retrieval-and Generation-Based Dialog Systems



# Response Generation by Context-aware Prototype Editing

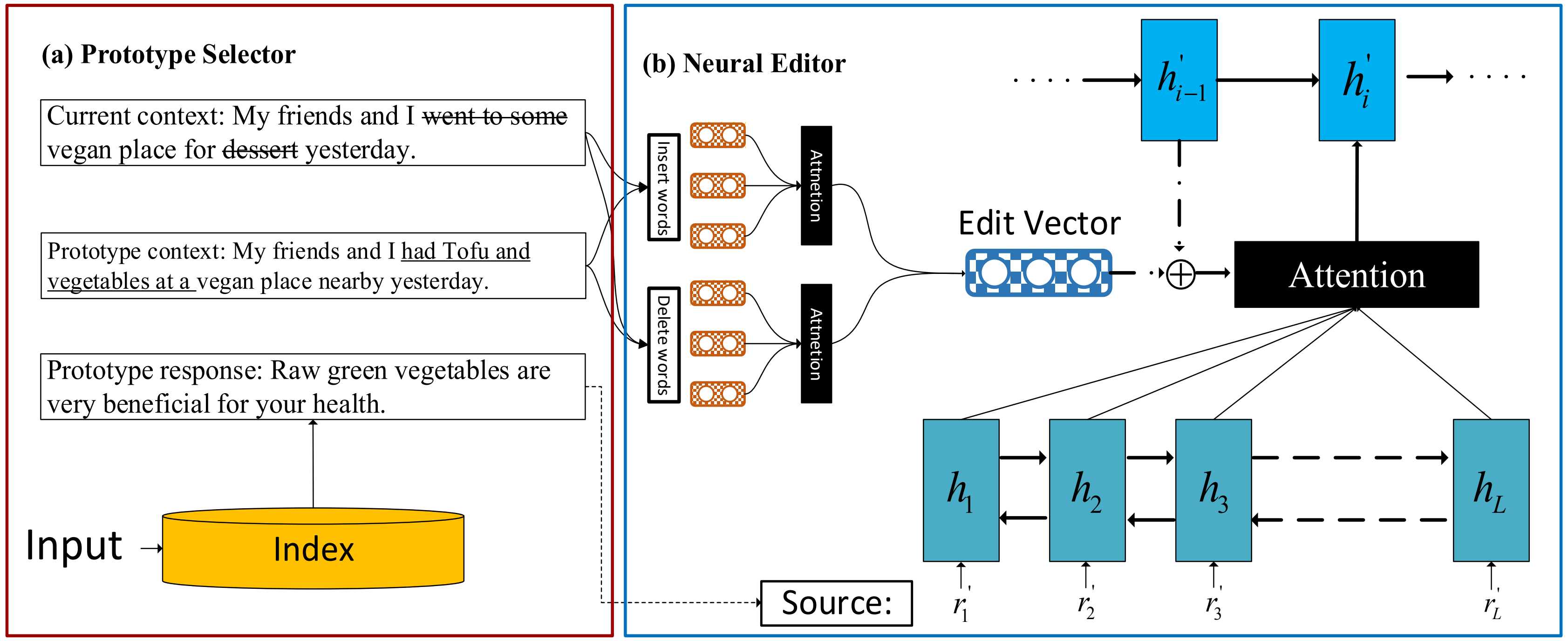
**Yu Wu[†], Furu Wei[‡], Shaohan Huang[‡] , Zhoujun Li[†], Ming Zhou[‡]**
[†]State Key Lab of Software Development Environment, Beihang University, Beijing, China
[‡] Microsoft Research, Beijing, China
{wuyu,lizj}@buaa.edu.cn {fuwei, shaohanh, mingzhou}@microsoft.com

# Response Generation by Context-aware Prototype Editing



| | |
|---|---|
| Context | My friends and I ~~went to some~~ vegan place for ~~dessert~~ yesterday. |
| Prototype context | My friends and I had Tofu and vegetables at a vegan place nearby yesterday. |
| Prototype response | **Raw green vegetables** are very **beneficial** for your health. |
| Revised response | **Desserts** are very **bad** for your health. |

# Response Generation by Context-aware Prototype Editing



(a) Prototype Selector

Current context: My friends and I ~~went to some~~ vegan place for ~~dessert~~ yesterday.

Prototype context: My friends and I had Tofu and vegetables at a vegan place nearby yesterday.

Prototype response: Raw green vegetables are very beneficial for your health.

(b) Neural Editor

# Data filtering

- The motivation behind filtering out instances with Jaccard similarity < 0.3 is that a neural editor model performs well only if a prototype is lexically similar to its ground-truth.

- Besides, we hope the editor does not copy the prototype so we discard instances where the prototype and groundtruth are nearly identical (Jaccard similarity > 0.7)

- We do not use context similar- ity to construct parallel data for training

# Response Generation by Context-aware Prototype Editing

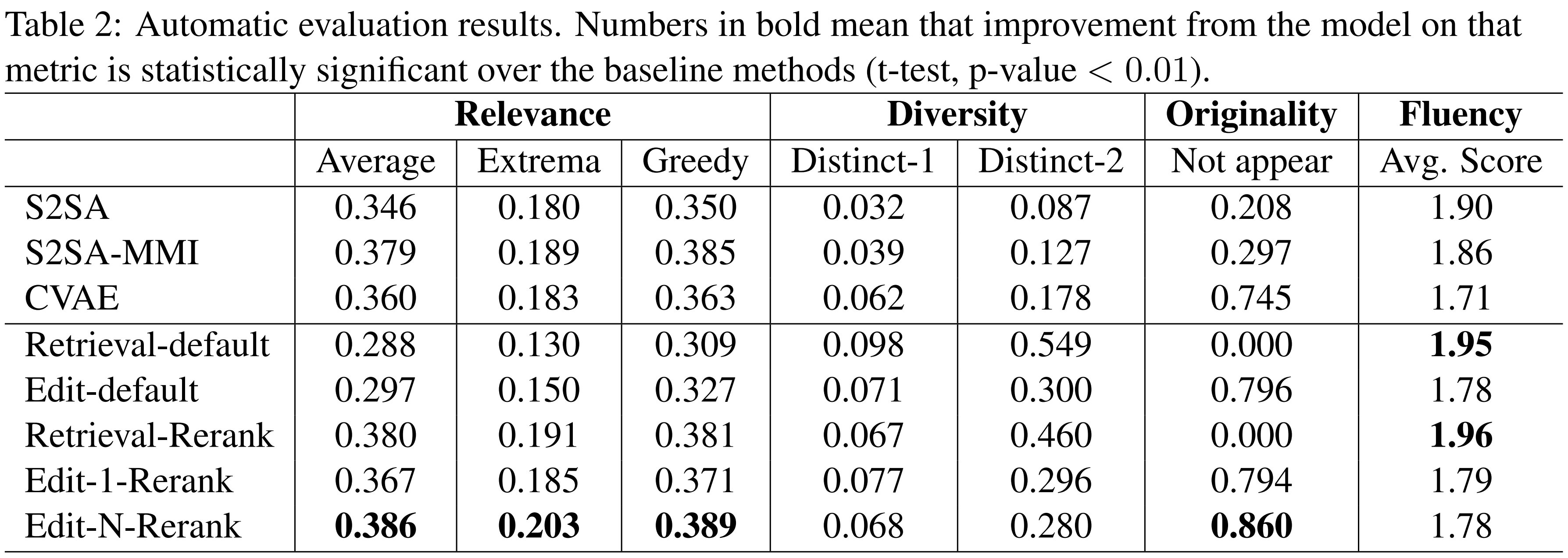


Table 2: Automatic evaluation results. Numbers in bold mean that improvement from the model on that metric is statistically significant over the baseline methods (t-test, p-value $< 0.01$).

| | **Relevance** | | | **Diversity** | | **Originality** | **Fluency** |
|---|---|---|---|---|---|---|---|
| | Average | Extrema | Greedy | Distinct-1 | Distinct-2 | Not appear | Avg. Score |
| S2SA | 0.346 | 0.180 | 0.350 | 0.032 | 0.087 | 0.208 | 1.90 |
| S2SA-MMI | 0.379 | 0.189 | 0.385 | 0.039 | 0.127 | 0.297 | 1.86 |
| CVAE | 0.360 | 0.183 | 0.363 | 0.062 | 0.178 | 0.745 | 1.71 |
| Retrieval-default | 0.288 | 0.130 | 0.309 | 0.098 | 0.549 | 0.000 | **1.95** |
| Edit-default | 0.297 | 0.150 | 0.327 | 0.071 | 0.300 | 0.796 | 1.78 |
| Retrieval-Rerank | 0.380 | 0.191 | 0.381 | 0.067 | 0.460 | 0.000 | **1.96** |
| Edit-1-Rerank | 0.367 | 0.185 | 0.371 | 0.077 | 0.296 | 0.794 | 1.79 |
| Edit-N-Rerank | **0.386** | **0.203** | **0.389** | 0.068 | 0.280 | **0.860** | 1.78 |