# Informative Neural Response Generation with Retrieval Memories

# Problems of IR and Generative Models

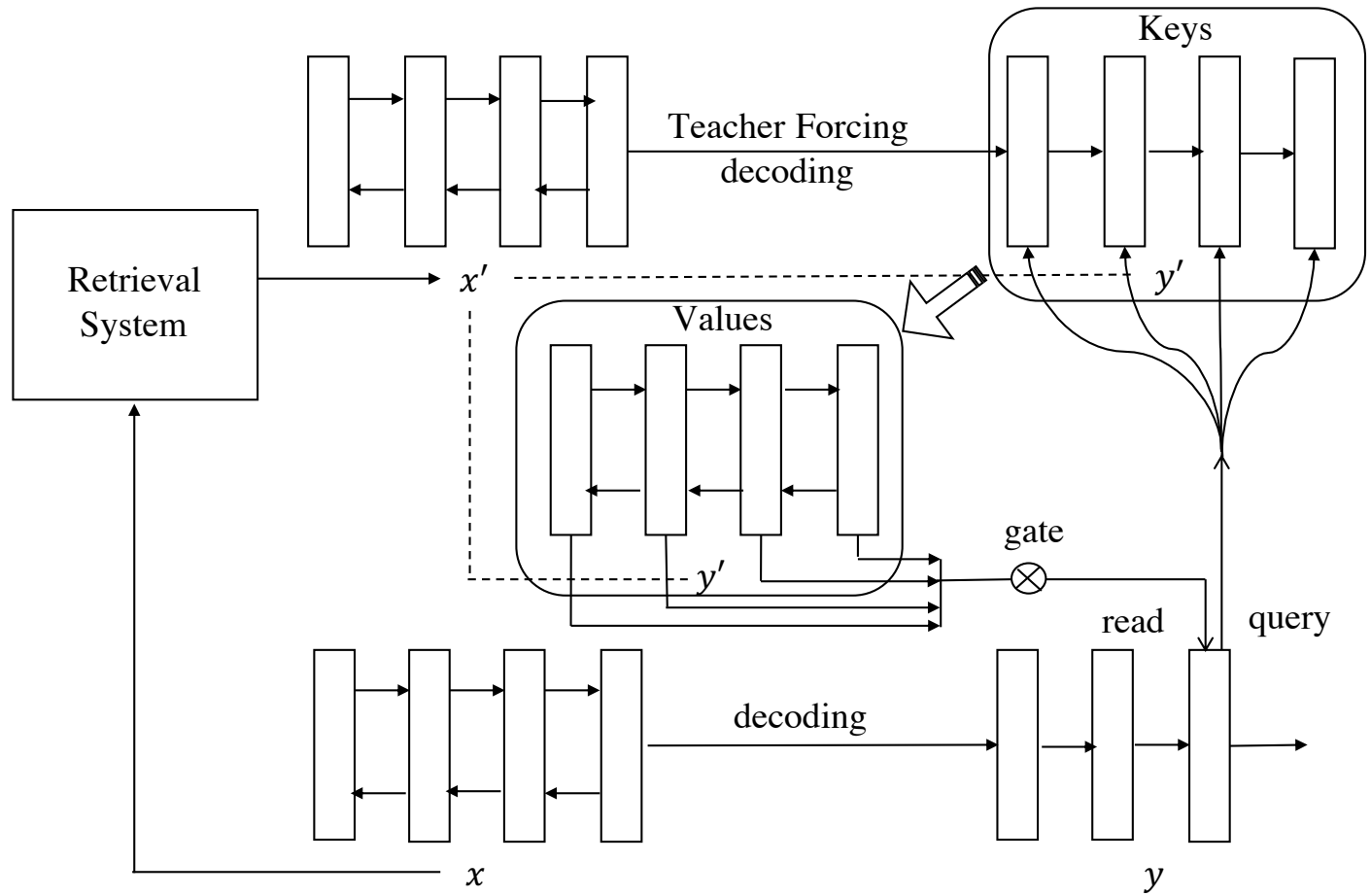|  | informative | appropriate |
|---|---|---|
| IR | √ | × |
| Generative | × | √ |

# Why Generative Models do not work

Generation – non-informative

- One-to-many problem, yet the generation model outputs responses with no specific preference.

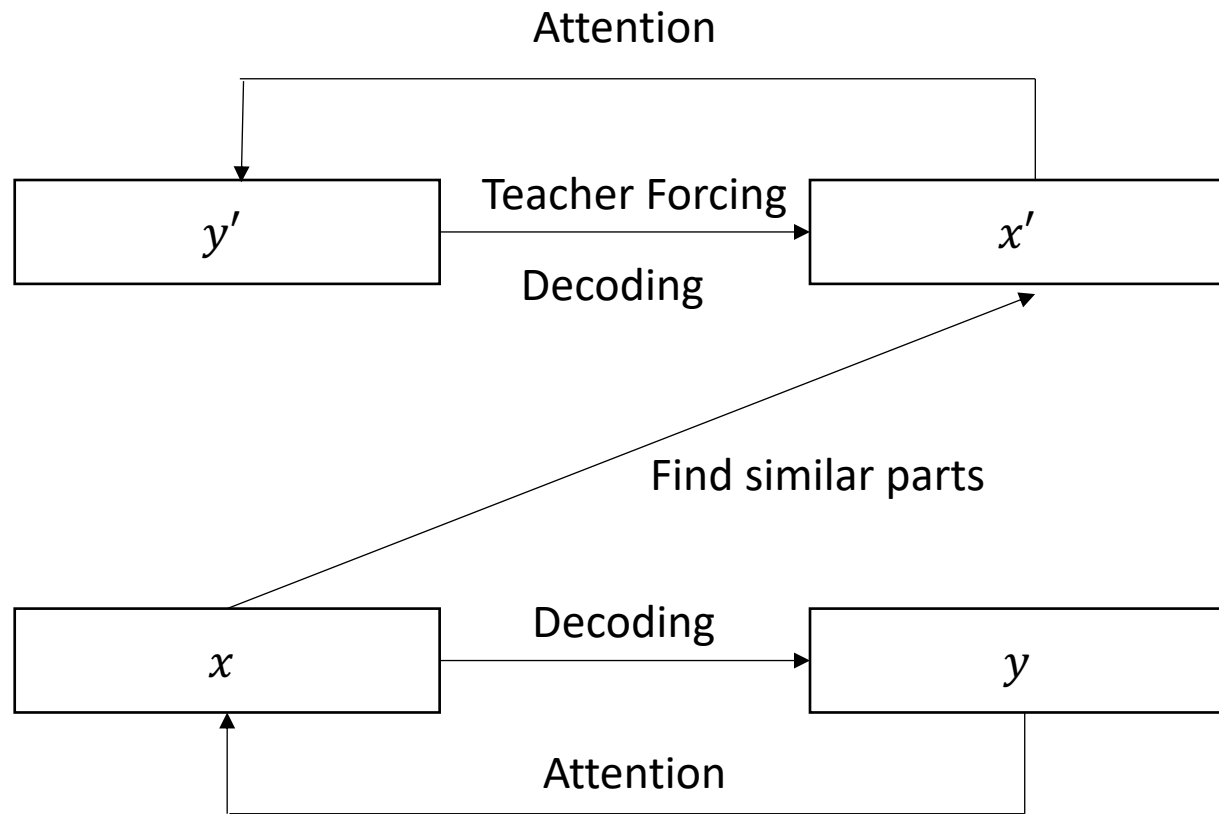- The most frequent responses are the safe and dull responses such as "I don't know".

IR - informative

- narrow the wide output space
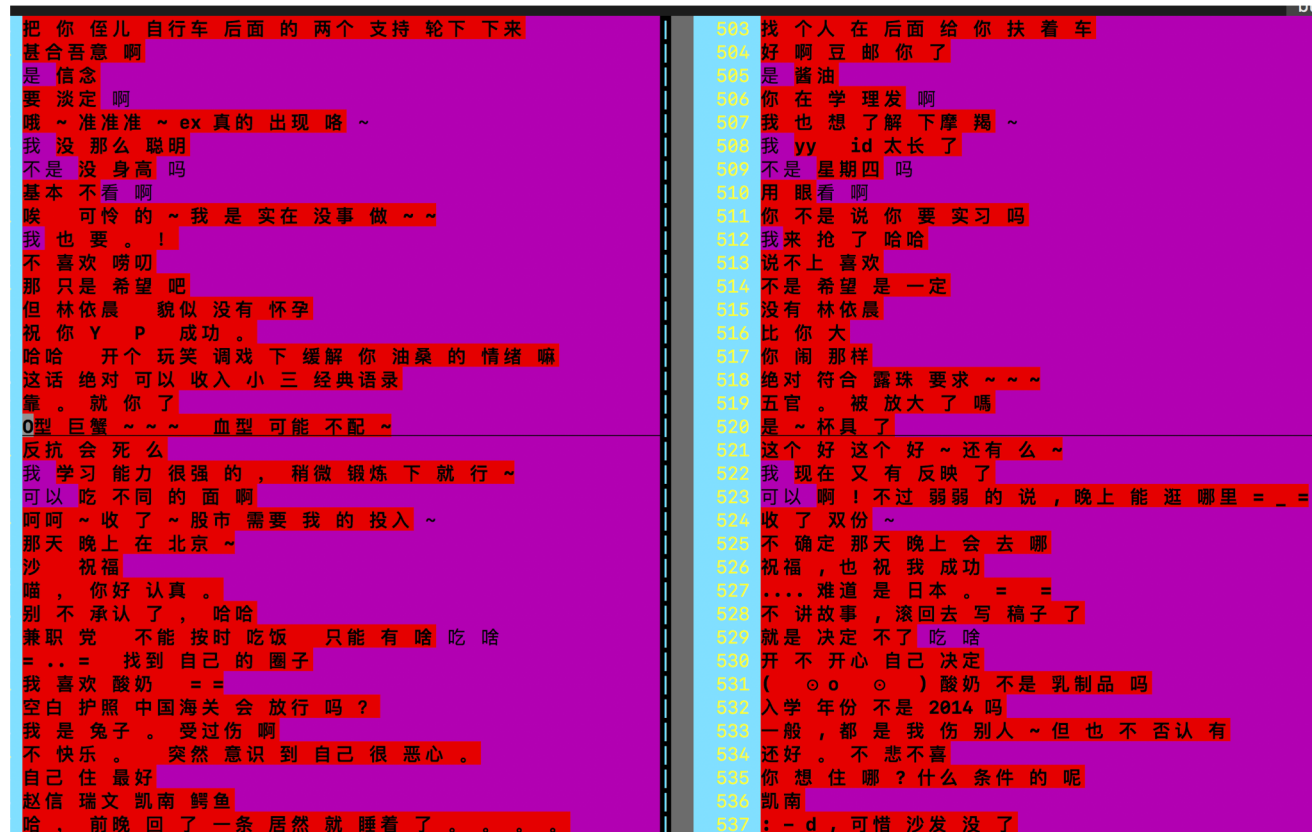
- provide missing facts

# Model-I

# Model-II

Attention

$y'$ → Teacher Forcing → $x'$

Decoding

Find similar parts

$x$ → Decoding → $y$

Attention

# Pitfalls in training

- the retrieved response does not necessarily ensembles the target response.

# Pitfalls in training

- Gated retrieval memories.
- Build a more effective training set

# Dataset-I

- 200W Jaccard(y, y') < 0.6 + 200W random samples
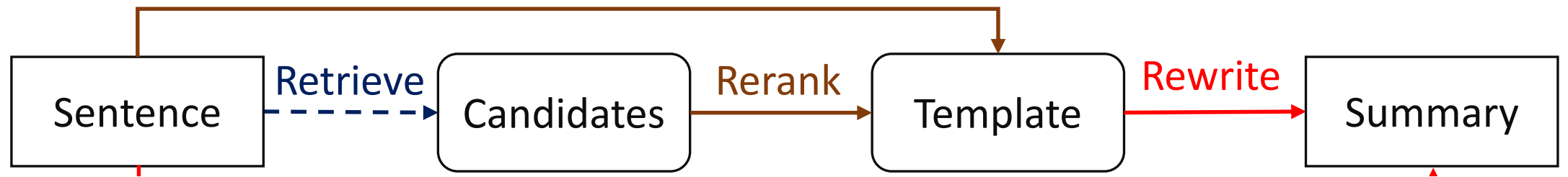
- Copy retrieval results

# Dataset-II

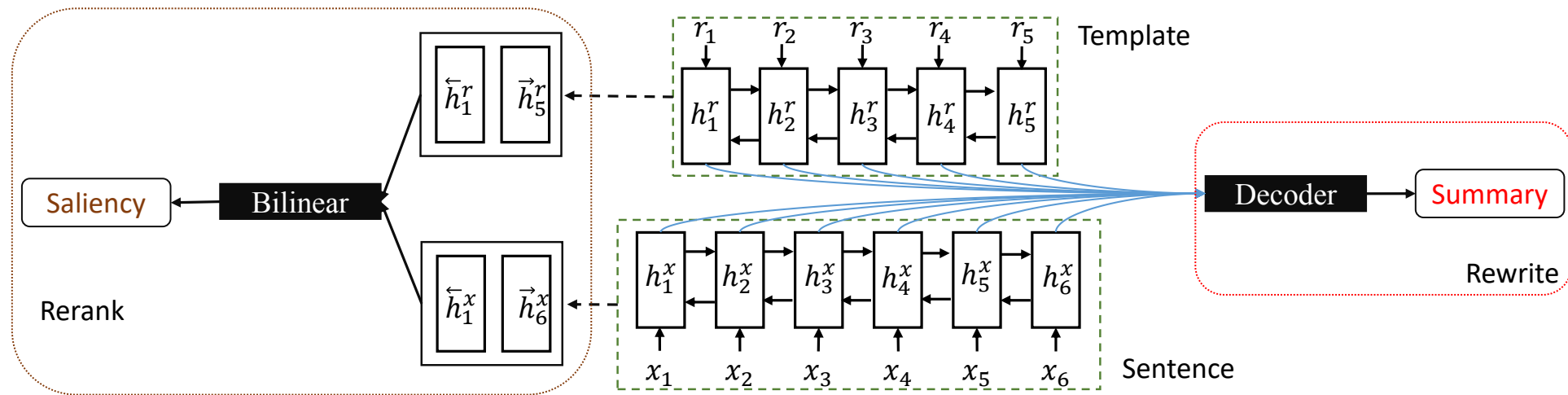- 400W Jaccard(y, y') < 0.65


- Copy retrieval results

# Dataset-III

- 100W Jaccard(y, y') < 0.2 and  0.2< Jaccard(x, x') < 0.5

- Safe responses

# Retrieve, Rerank and Rewrite: Soft Template Based Neural Summarization

# Overview

# Model

# Rerank

- Use x and r to predict the similarity (saliency) of y and r.

- The golden saliency is measured by ROUGE.

# Results

| Type | RG-1 | RG-2 | RG-L |
|---|---|---|---|
| Random | 2.81 | 0.00 | 2.72 |
| First | 24.44 | 9.63 | 22.05 |
| Max | 38.90 | 19.22 | 35.54 |
| Optimal | 52.91 | 31.92 | 48.63 |
| Rerank | 28.77 | 12.49 | 26.40 |

IR

| Type | RG-1 | RG-2 | RG-L |
|---|---|---|---|
| +Random | 32.60 | 14.31 | 30.19 |
| +First | 36.01 | 17.06 | 33.21 |
| +Max | 41.50 | 21.97 | 38.80 |
| +Optimal | 46.21 | 26.71 | 43.19 |
| +Rerank($Re^3Sum$) | 37.04 | 19.03 | 34.46 |

generation