USTC at NTCIR-12 STC Task

Junbei Zhang Junfeng Hou Shiliang Zhang Lirong Dai

University of Science and Technology of China

Presented by Junbei Zhang

2016.06.09

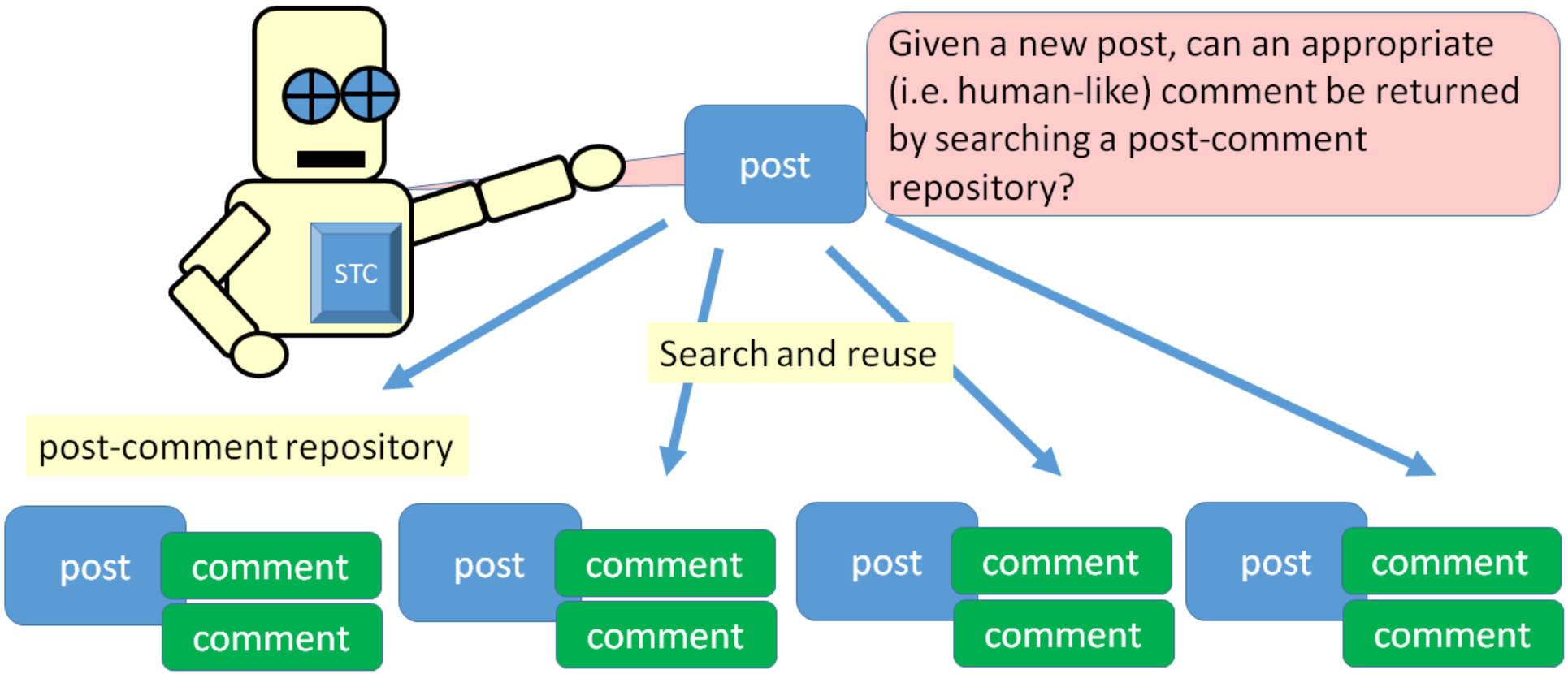
zjunbei@mail.ustc.edu.cn

Outline

- **Introduction**
- System Architecture
- Experiment and analysis
- Conclusion and future work

Task

- Short Text Conversation (STC)



From: http://ntcir12.noahlab.com.hk/stc.htm

Previous Model

- Neural Responding Machine for Short-Text Conversation [Shang et al.2015]
- A neural conversational model [Vinyals et al. 2015]

Outline

- Introduction
- **System Architecture**
- Experiment and analysis
- Conclusion and future work

System Architecture

- Learning-to-rank
- **Lexical** Features:
  - Query-Response Similarity
  - Query-Post Similarity
  - Transition-p2c
- **Semantic** Features:
  - EncDec-Forward model
  - EncDec-Reverse model
  - Joint-Train model
- Ranking: linear RankingSVM

Lexical Features

- Query-Response Similarity
  - Map query and response to their own **TF-IDF score vector**
  - Calculate **cosine similarity** between query and response

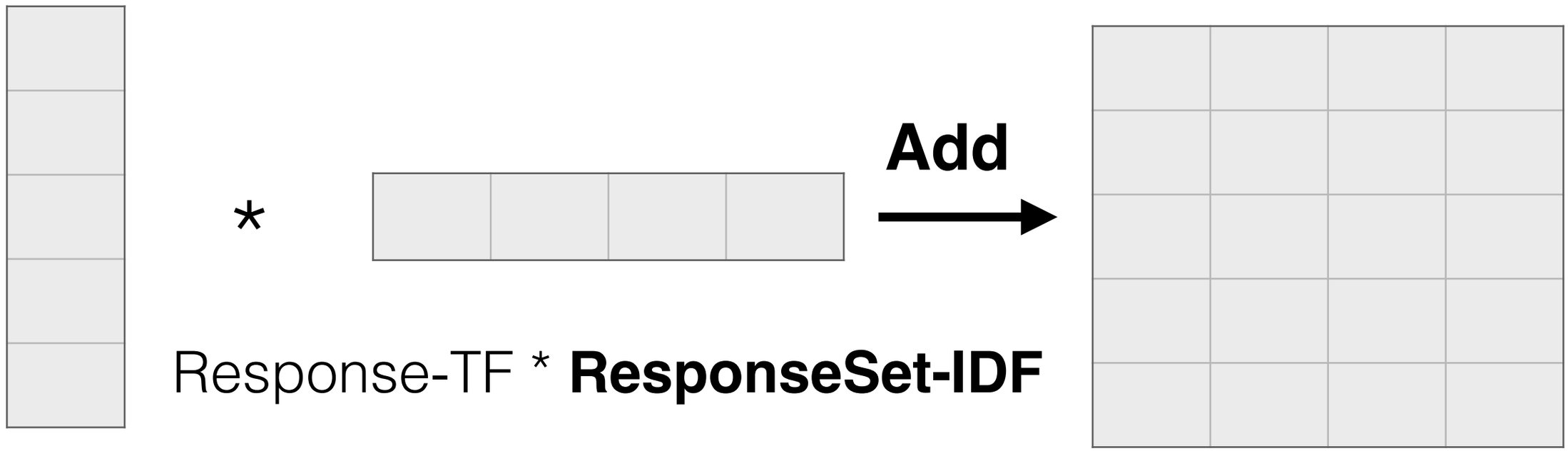
- Query-Post Similarity
  - TF-IDF score vector
  - The cosine similarity between query and post is regarded as response score.

Transition-p2c

- Model the transition probability between post words and response words
- Example:
  - Query: Where shall we eat?
  - Query-Response: **We eat** two apples.
  - Query-Post: **Where shall we** go? -> We go to the bank.
  - Transition-p2c: Restaurant.

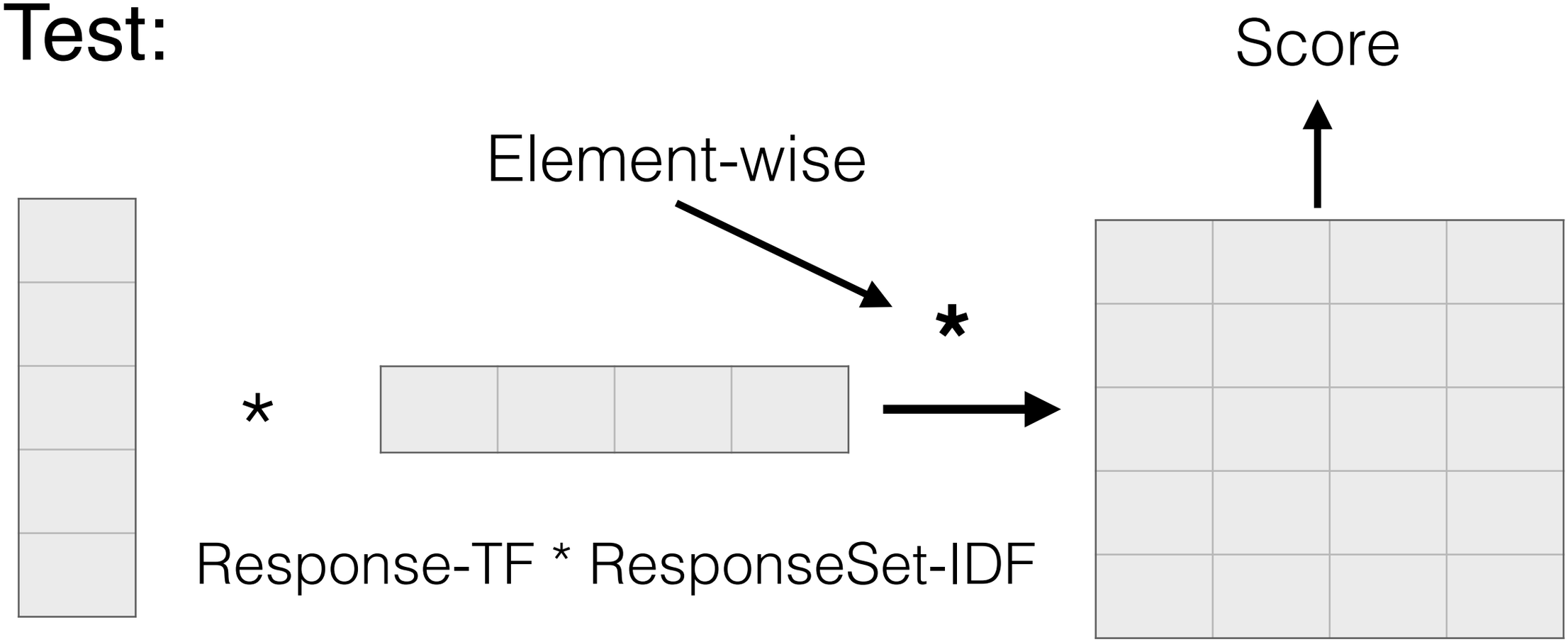
Transition-p2c

- Train:



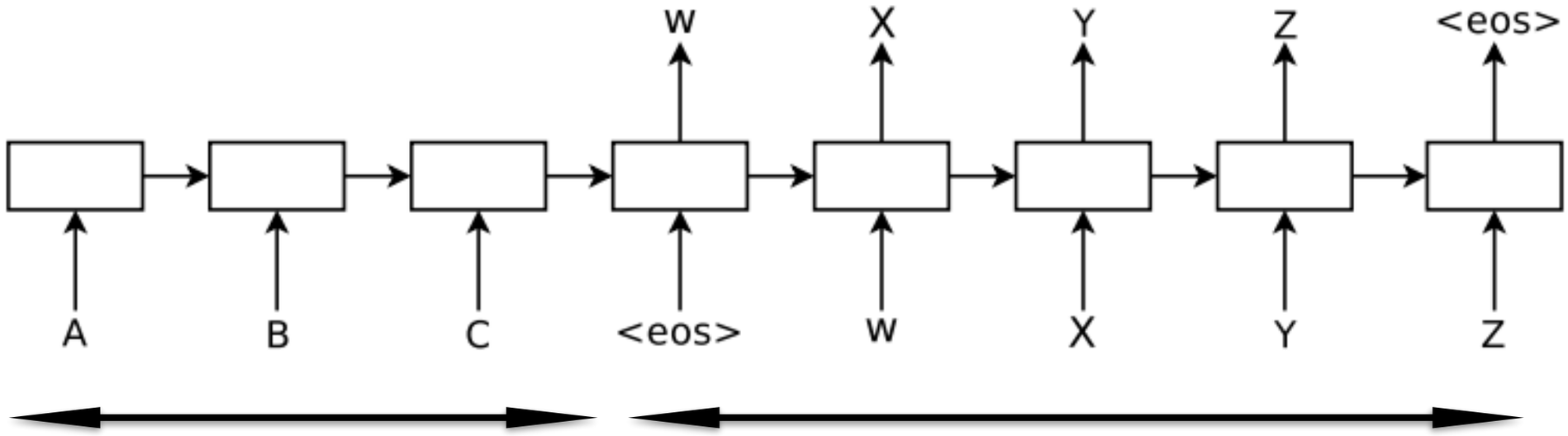
Transition-p2c

- Test:



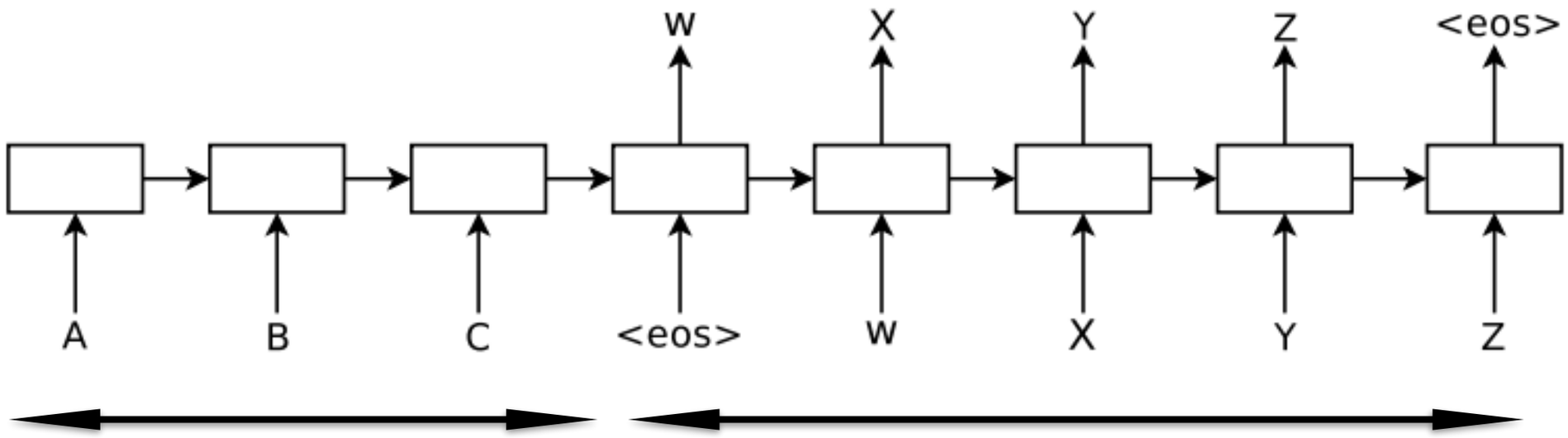
Semantic Features

- EncDec-Forward model:
  - seq2seq model to estimate **P(Response | Post)**



Semantic Features

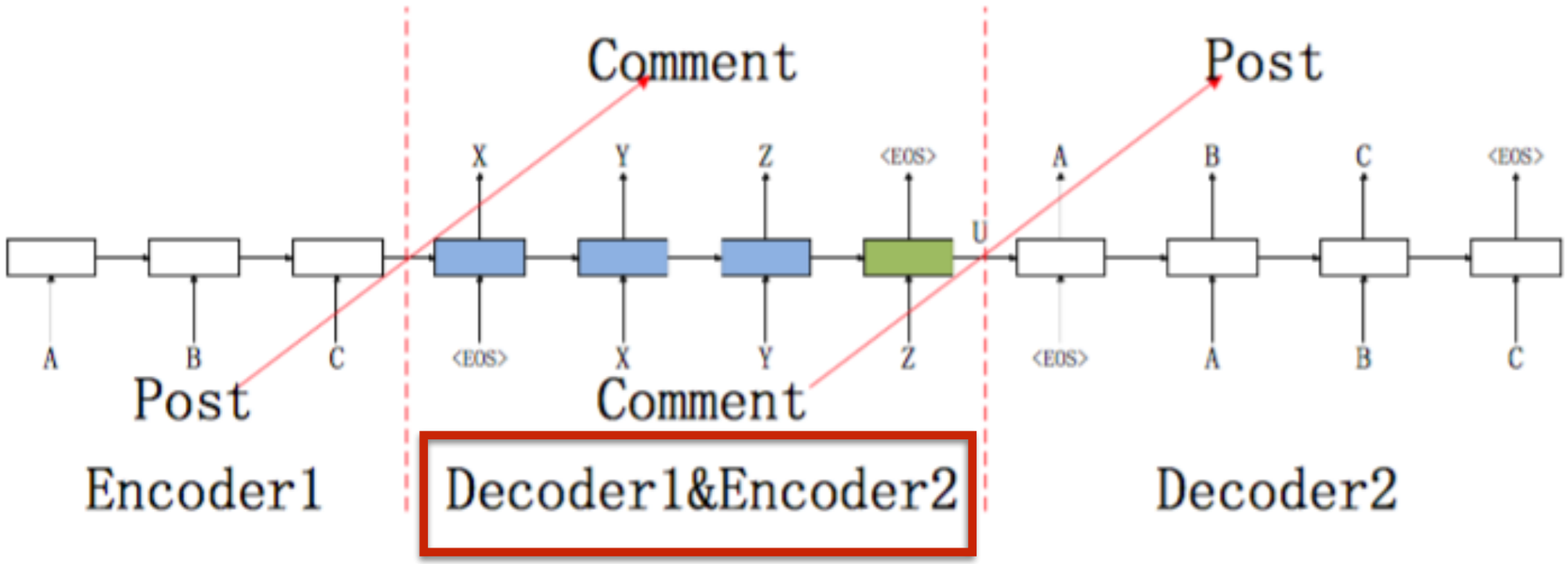
- EncDec-Reverse model:
  - **Many-to-many:** one post —> several responses, one response <— several posts
  - **Reconstruction:** max P(Post | Response)



~~hahaha~~

Semantic Features

- Joint-Train model:
  - **Generation & Reconstruction:** combine P(Response I Post) and P(Post I Response) in one model
  - Decoder1 is regarded as **both a decoder and an encoder** (Encoder2)



Ranking

- Linear RankingSVM
- Merge all the scores and outputs a final score for each query and response pair

Outline

- Introduction
- System Architecture
- **Experiment and analysis**
- Conclusion and future work

Implementation

- Subset:
  - Using 4 Models to get **top 10000 responses each** from the whole repository, and remove the same pairs
  - Query-Response Similarity, Query-Post Similarity, Transition-p2c and EncDec-Reverse

- loss function (Joint-Train):
  - loss = **0.2** * post-comment loss (encoder1&decoder1) + **0.8** * comment-post loss(encoder2&decoder2)

Results

- 5 Runs:
  - **USTC-C-R1:** Query-Response Similarity + Query-Post Similarity + EncDec-Forward + **EncDec-Reverse + Transition-p2c**
  - **USTC-C-R2:** Query-Response Similarity + Query-Post Similarity + EncDec-Forward + **EncDec-Reverse + JointTrain**
  - **USTC-C-R3:** Query-Response Similarity + Query-Post Similarity + EncDec-Forward + **Transition-p2c**
  - **USTC-C-R4:** Query-Response Similarity + Query-Post Similarity + EncDec-Forward + **EncDec-Reverse**
  - **USTC-C-R5:** Query-Response Similarity + Query-Post Similarity + EncDec-Forward

Results

- Official Results:

Table 1: Official STC(Chinese) results

Run	nDCG@1	P+	nERR@10
R5	**0.2867**	**0.4509**	0.4160
R4	0.2767	0.4479	**0.4181**
R1	0.2733	0.4499	0.4169
R2	0.2567	0.4310	0.4001
R3	0.2267	0.4094	0.3848

Results

- Offline Results(using 6017 labeled data):

Table 2: STC(Chinese) training set results

Run	nDCG@1	P+	nERR@10
R5	0.4741	0.6529	0.6327
R4	0.4785	0.6582	0.6395
R3	0.4726	0.6570	0.6347
R2	**0.4889**	**0.6625**	0.6446
R1	0.4859	0.6618	**0.6449**

Results

- Inconsistent:

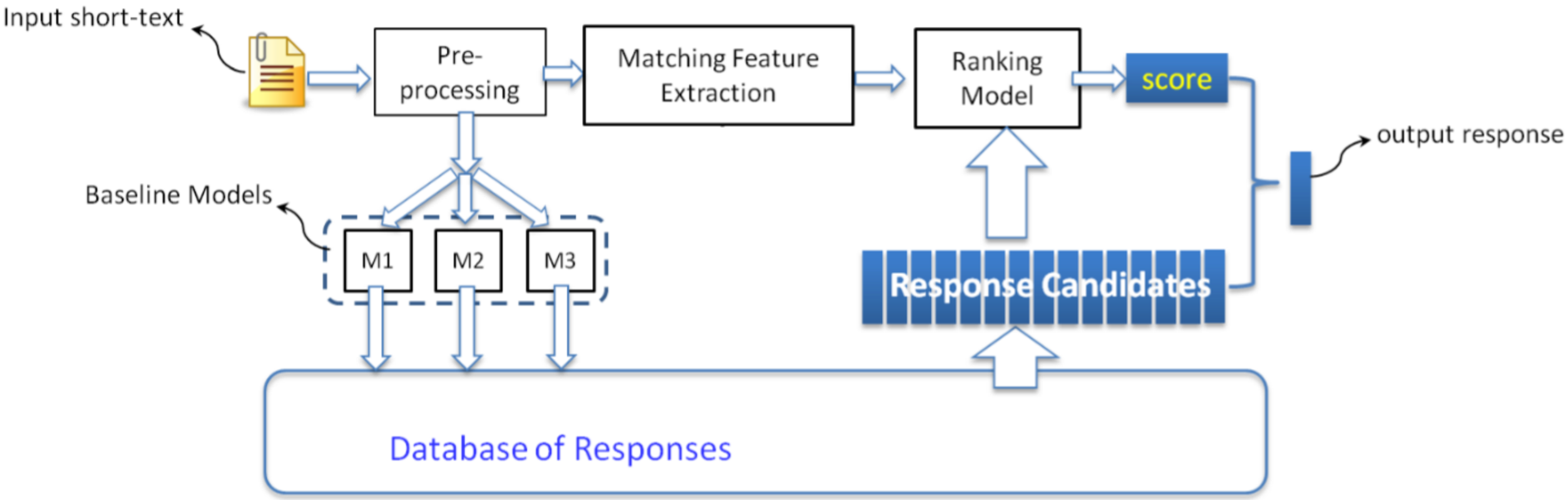


Figure 5: Diagram of the retrieval-based automatic response system.

A Dataset for Research on Short-Text Conversations. [Wang et al. EMNLP2013]

Results

- Transition-p2c Case Study:

Table 3: Transition score top10 of different word pairs

post words	comment words	transition score
运费 (freight)	代购 (purchasing agents)	0.3207
中型(medium)	谢谢(thanks)	0.1302
警报(alarm)	口水(saliva)	0.1273
元宵节 (Lantern Festival)	快乐 (happy)	0.1260
萌到(sprout)	可爱(lovely)	0.1180
拜年 (pay a New Year call)	新年快乐 (happy new year)	0.1177
王老吉 (Wong Lo Kat)	加多宝 (JDB Beverage)	0.1077
本地(native)	流量(traffic)	0.1066
小家伙(kiddy)	可爱(lovely)	0.1042
张国荣 (Leslie Cheung)	哥哥 (brother)	0.1007

Outline

- Introduction
- System Architecture
- Experiment and analysis
- **Conclusion and future work**

Conclusions

- Our model:
  - **Lexical** Features:
    - Query-Response Similarity
    - Query-Post Similarity
    - Transition-p2c
  - **Semantic** Features:
    - EncDec-Forward model
    - EncDec-Reverse model
    - Joint-Train model

Conclusions

- The results in training set and cases show the efficiency of our method.

- Inconsistent: different evaluation indexes & the subset selection problem.

Future Work

- Try deep encoder-decoder model
- Try word2vec instead of TF-IDF