YAHOO! JAPAN

YJTI at the NTCIR-13 STC Japanese Subtask

Dec. 7, 2017
Toru Shimizu

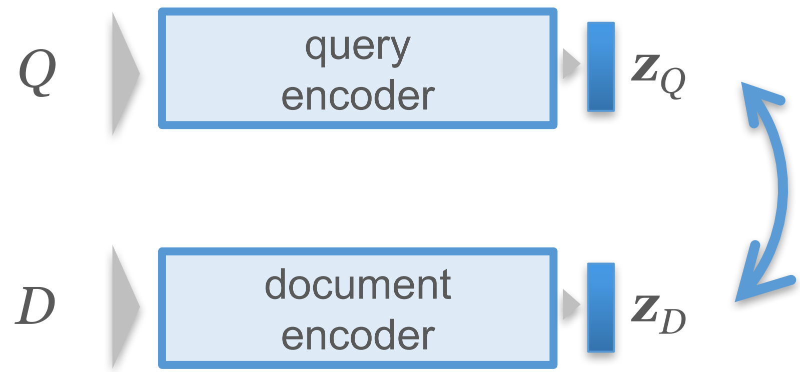
Overview

Retrieval or Generation

- Retrieval-based system
  - Effective if you have a good matching model and enough candidate responses
  - Pros
    - Human-written, fluent sentences for responses
    - The conversation can sometimes actually be interesting.
      - Hence more practical
  - Cons
    - Lack of flexibility
      - This can be mitigated with large amount of candidates and the variety in them.
      - 1.2M unique sentences in the training data

Architecture

- DSSM (Deep Structured Semantic Model)
  – Huang et al., 2013
  – A method for IR, query-document matching



$$R_{\Theta}(Q, D) = \frac{\mathbf{z}_Q^{\mathrm{T}} \mathbf{z}_D}{\|\mathbf{z}_Q\| \|\mathbf{z}_D\|}$$

- LSTM-DSSM
  – Palangi et al., 2014
  – LSTM-RNN for generating query and document representations

The Overall Process: Three Stages

Model Training

- Train two models:
  - a comment encoder
  - a reply encoder

Reply Text Preparation and Indexing

- Preprocess the training data to obtain candidate replies
- Generate vector representations of the replies
- Build the reply index

Runtime

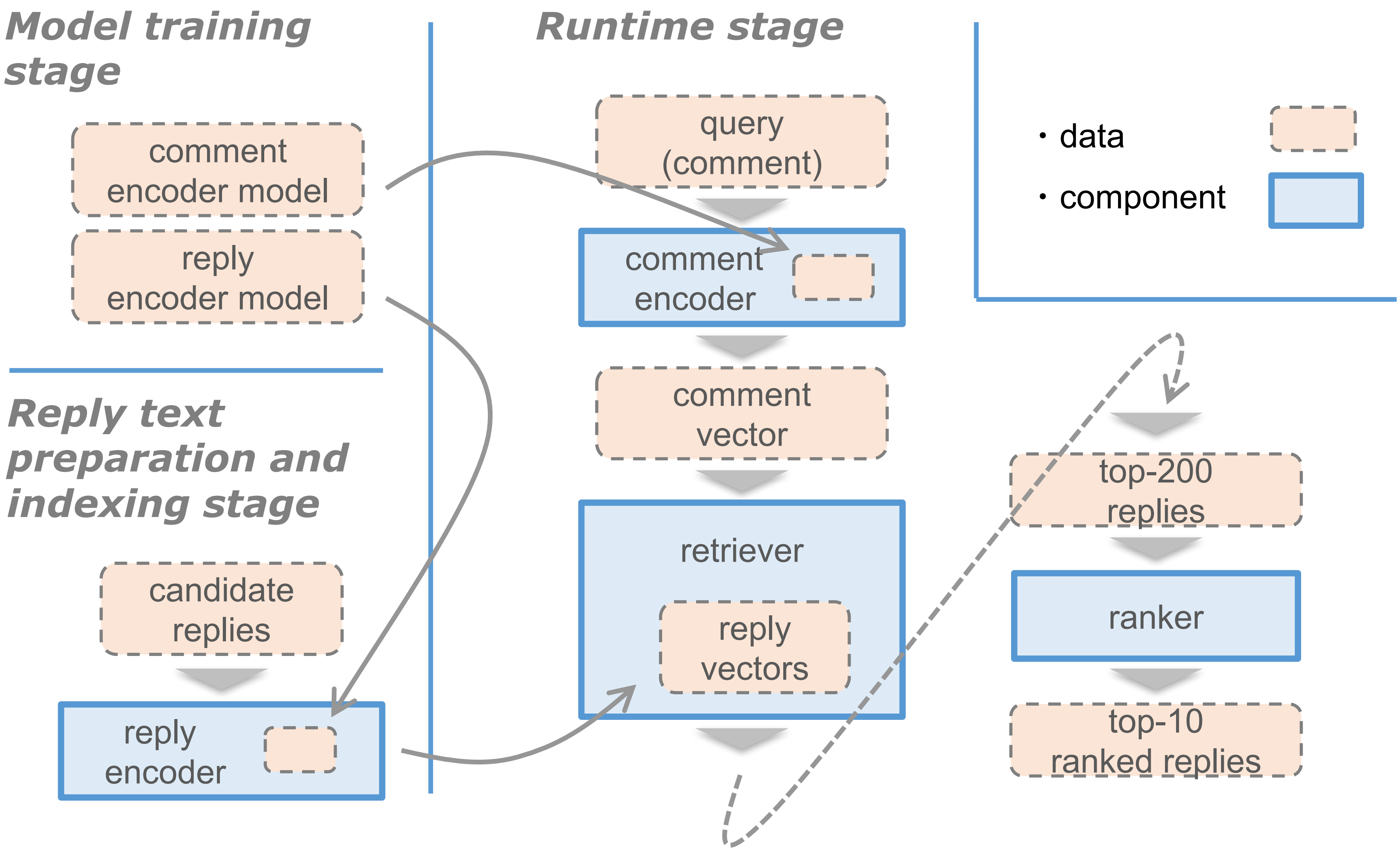
- Produce actual reply lists using the runtie system

Submissions

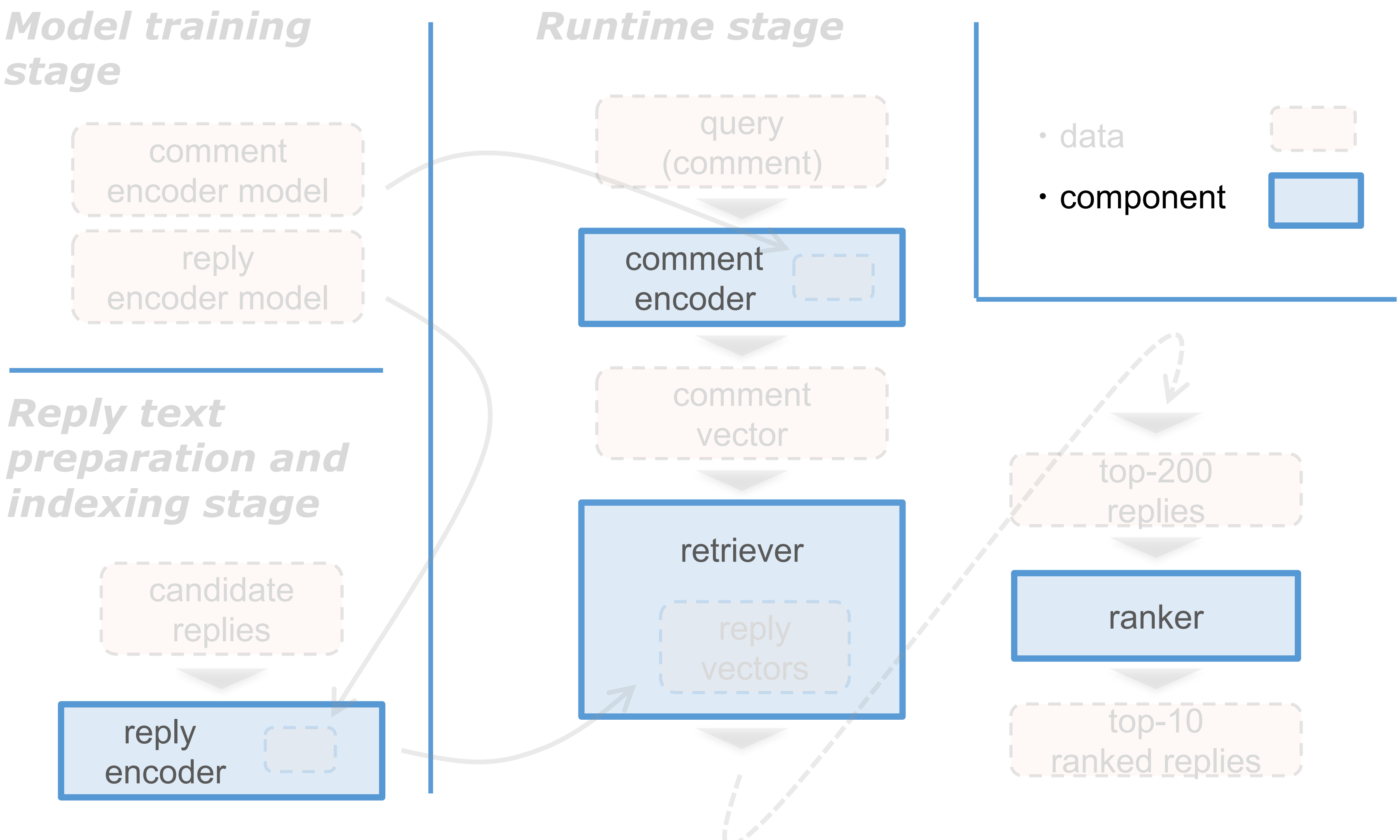
- Two runs:
  - YJTI-J-R1
    - Trained by Twitter conversation data
  - YJTI-J-R2
    - Trained mainly by Yahoo! Chiebukuro QA data
- The runtime system is the same.
- Only the models are different.

Runtime System

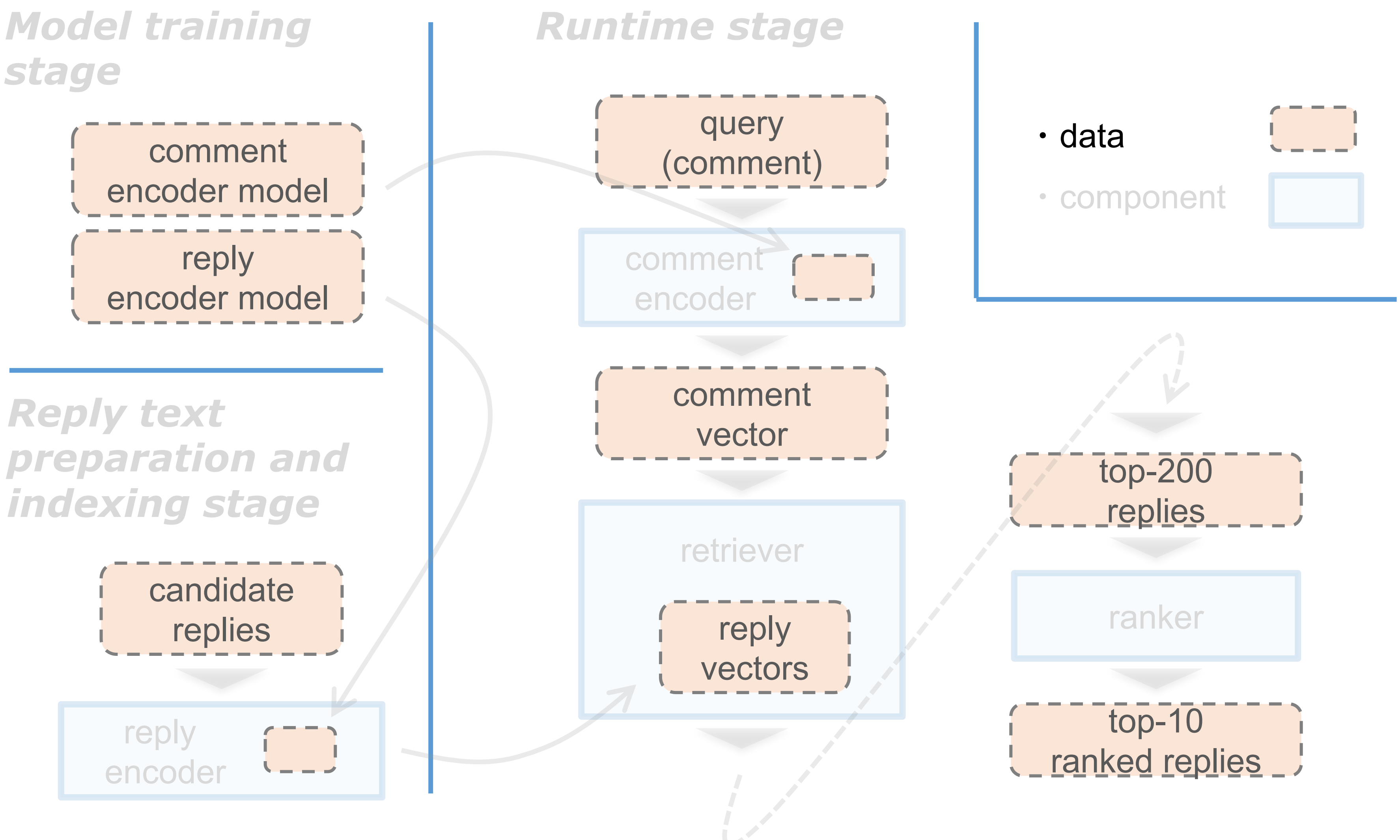
Runtime System Overview



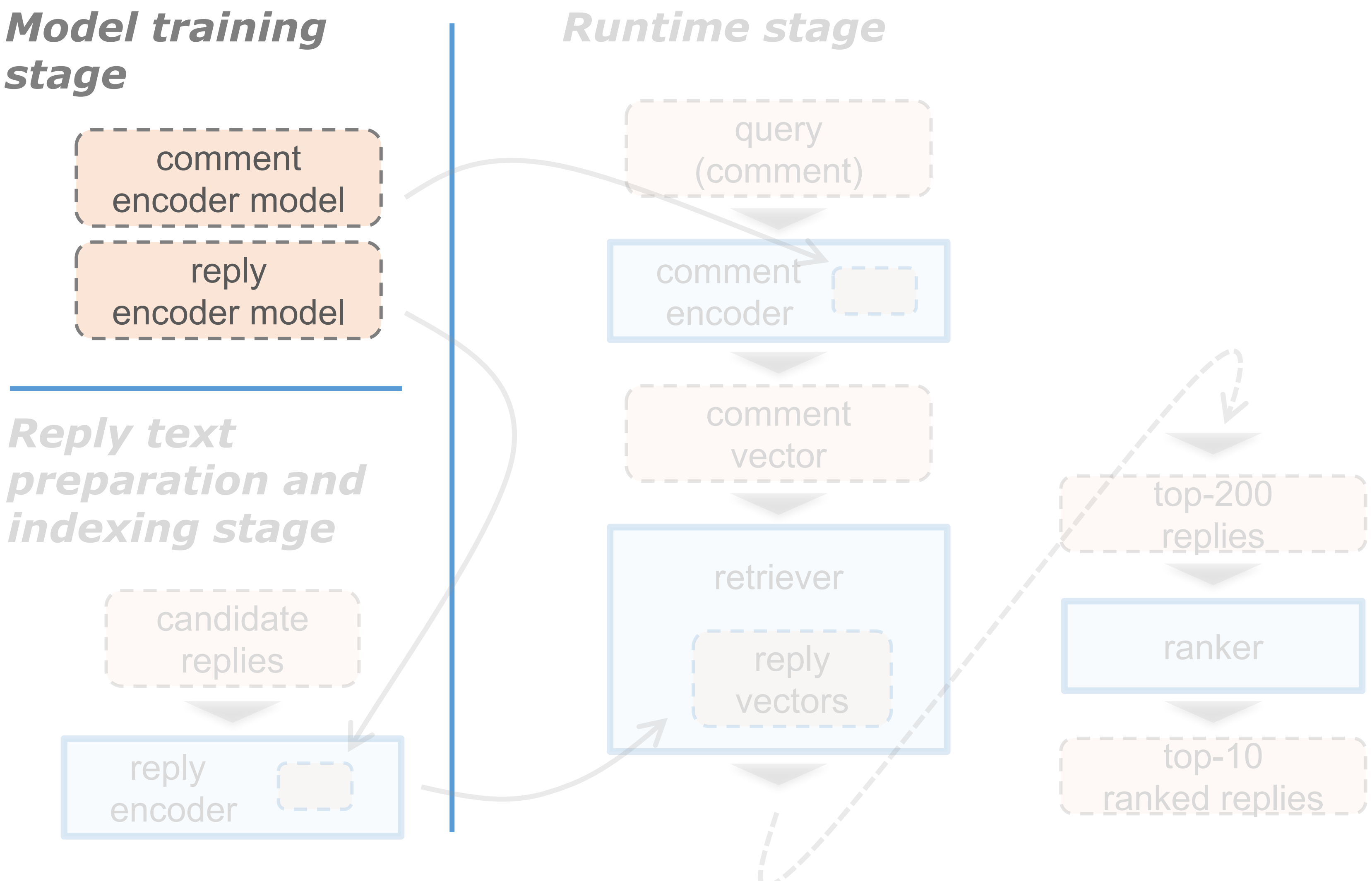
Runtime System Overview: Software Components



Runtime System Overview: Data

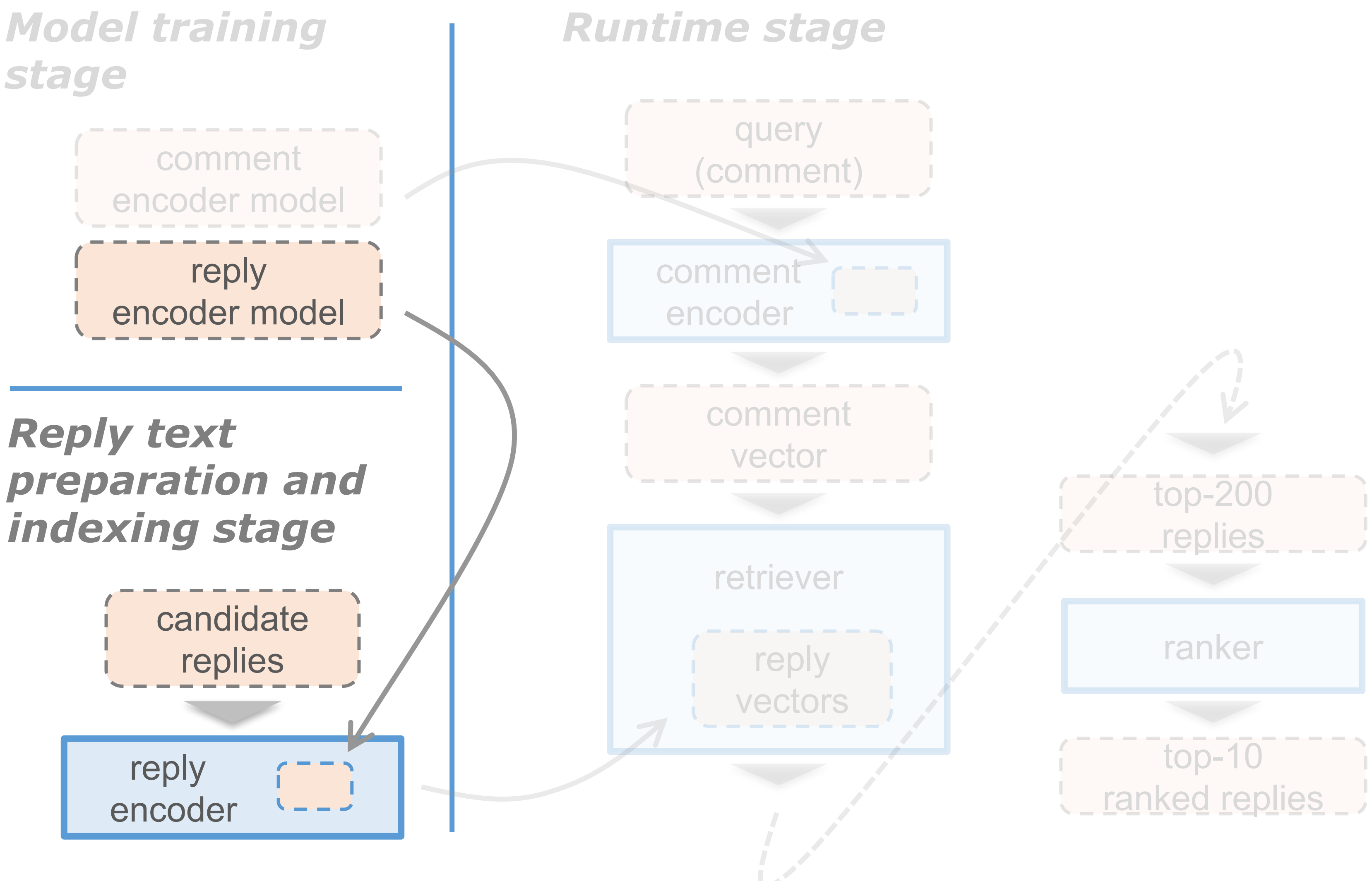


Runtime System Overview: The 1st Stage

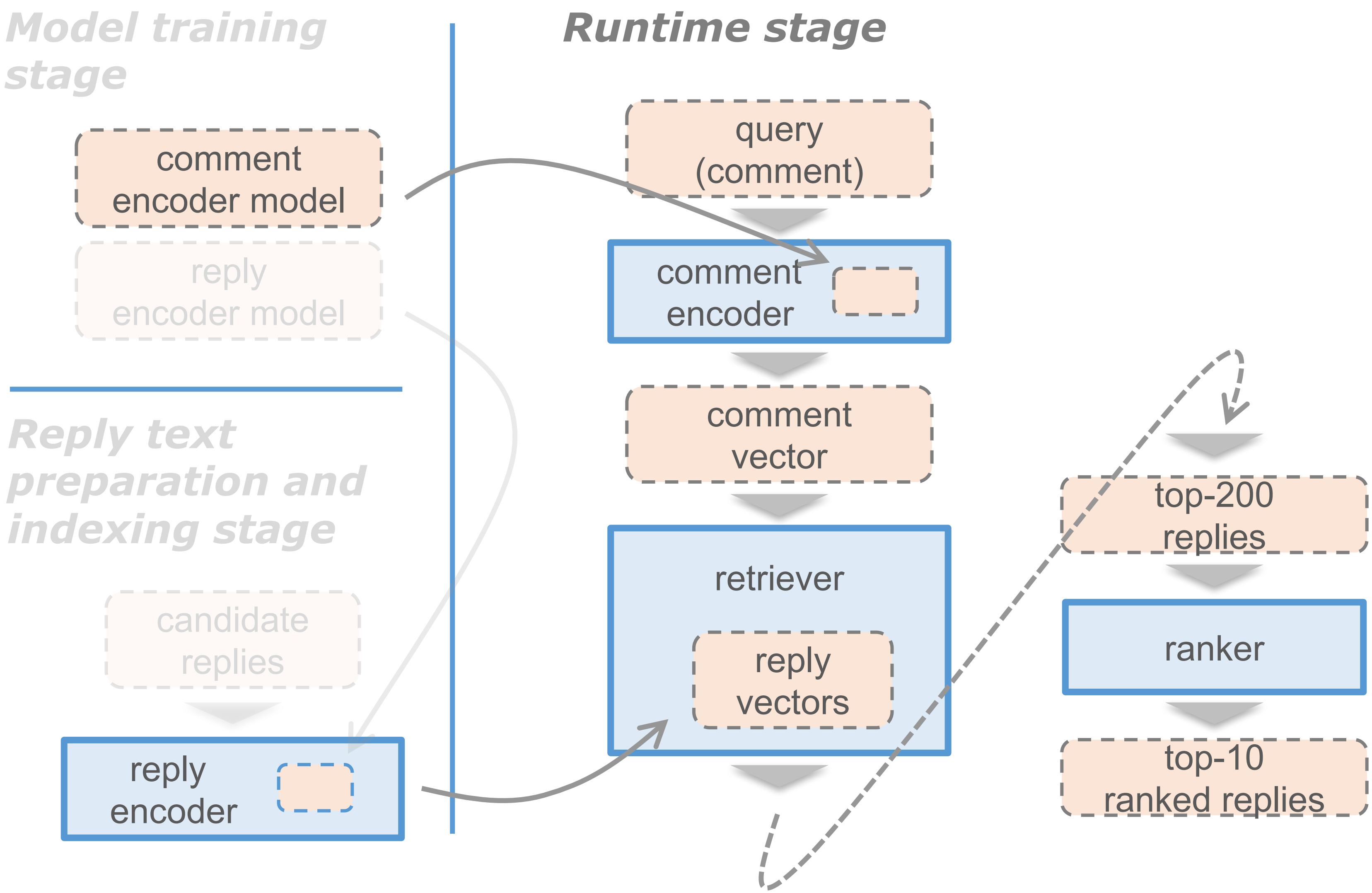


Runtime System Overview: The 2nd Stage

【公開】



Runtime System Overview: The 3rd Stage



Indexer and Retriever

- Generate 1024-element representations of reply candidates by the reply encoder model
- NGT
  - Open source software for graph-based approximate similarity search over dense vectors
    - Developed by M. Iwasaki
    - https://research-lab.yahoo.co.jp/software/ngt/
- Retrieve the nearest 200 reply vectors from a given comment vectors
  - L2-distance, cosine similarity
- Return the list of their texts and metadata

Ranker

- Three tiers for dealing with metadata matching: THEME, GENRE, and OTHER

THEME The Theme is matched btw. the comment and a reply. (At most 3)

GENRE The Genre is matched btw. the comment and a reply (At most 3)

OTHER No metadata match. (No limitation of number)

The final top-10 replies

- reply 1
- reply 2
- reply 3
- reply 4
- reply 5
- reply 6
- reply 7
- reply 8
- reply 9
- reply 10

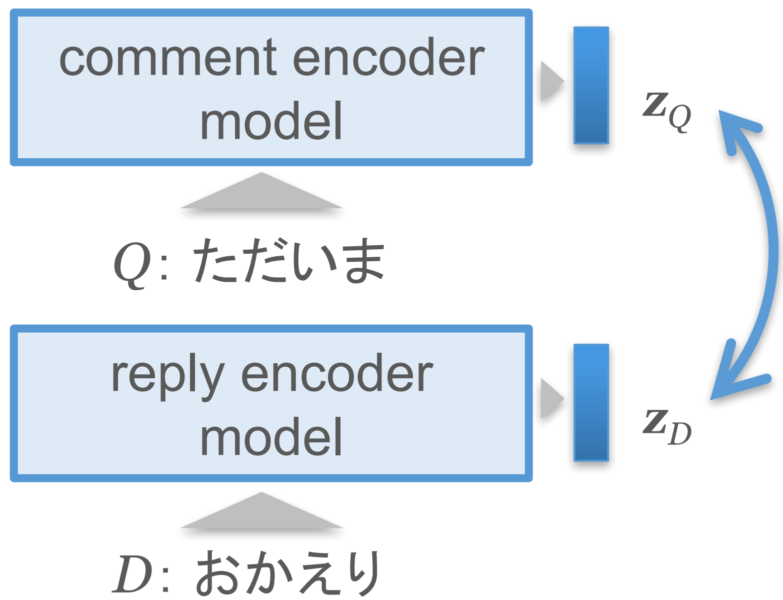
Model, Data, and Training

Comment/Reply Encoder Model

- 3-layer LSTM RNN
  - Formulation: Graves, 2013
  - LSTM's hidden layer size: 1024 (for all the
  - Embedding layer size: 256
  - Representation size: 1024

Comment/Reply Encoder Model

- Training



$$R_{\boldsymbol{\Theta}}(Q, D) = \frac{\mathbf{z}_Q^{\mathrm{T}} \mathbf{z}_D}{\|\mathbf{z}_Q\| \|\mathbf{z}_D\|}$$

  – Consider this as a classification problem and maximize the probability for the right choice over a given dataset

$$P_{\boldsymbol{\Theta}}(D_i^k | Q_i) = \frac{\exp(\beta R_{\boldsymbol{\Theta}}(Q_i, D_i^k))}{\sum_{j=1}^{5} \exp(\beta R_{\boldsymbol{\Theta}}(Q_i, D_i^j))}$$

Comment/Reply Encoder Model

YAHOO! JAPAN

- Training cont'd

run	model type	data name	records comsumed
YJTI-J-R1	DSSM	Twitter conversation	135.0M
YJTI-J-R2	LM	Y! Chiebukuro LM	171.5M
	DSSM	Twitter conversation	85.8M
	DSSM	Y! Chiebukuro QA	42.9M

Data for Model Training

name	type	no. of records
Twitter LM	posts	100.0M
Twitter conversation	pairs	65.1M
Y! Chiebukuro LM	posts	202.0M
Y! Chiebukuro QA	pairs	66.3M

Results

Analysis and Results

- Performances measured by the validation data

matching task	YJTI-J-R1	YJTI-J-R2
Twitter conversation	**0.835**	0.759
Chiebukuro QA	0.864	**0.967**

Analysis and Results

- The official results under Rule-2

metric	YJTI-J-R1	YJTI-J-R2
Mean $nG@1$	0.4171	**0.4726**
Mean $nERR@2$	0.4544	**0.5288**
Mean $Acc_{L2}@1$	0.1860	**0.2040**
Mean $Acc_{L2}@2$	0.1490	**0.2030**
Mean $Acc_{L1,L2}@1$	0.6100	**0.7200**
Mean $Acc_{L1,L2}@2$	0.5750	**0.6900**

Conclusions

YAHOO! JAPAN

- Effectiveness of the overall approach:
  - Retrieval-based system
  - DSSM-like matching powered by LSTM-RNNs trained over a large amount of linguistic resources
- Social QA data was surprisingly useful for modeling topic-oriented conversations seen in this Yahoo! News comments data