A Co-Matching Model for Multi-choice Reading Comprehension

Speaker: Liu Ao

Machine Comprehension

Passage (*P*) + Question (*Q*) ⟶ Answer (*A*)

P

Alyssa got to the beach after a long trip. She's from Charlotte. She traveled from Atlanta. She's now in Miami. She went to Miami to visit some friends. But she wanted some time to herself at the beach, so she went there first. After going swimming and laying out, she went to her friend Ellen's house. Ellen greeted Alyssa and they both had some lemonade to drink. Alyssa called her friends Kristin and Rachel to meet at Ellen's house…….

Q What city is Alyssa in?

A Miami

Datasets

- Cloze-style datasets:
  - CNN/Daily Mail
  - Childrens Book Test (CBT)
  - Who Did What (WDW)
- Span-based datasets:
  - SQuAD
  - MS MARCO
- Multi-choice datasets:
  - MCTest
  - RACE (focused in this paper)

CNN/Daily Mail Datasets

P

(@entity4) if you feel a ripple in the force today , it may be the news that the official @entity6 is getting its first gay character . according to the sci-fi website @entity9 , the upcoming novel " @entity11 " will feature a capable but flawed @entity13 official named @entity14 who " also happens to be a lesbian . " the character is the first gay figure in the official @entity6 -- the movies , television shows , comics and books approved by @entity6 franchise owner @entity22 -- according to @entity24 , editor of " @entity6 "

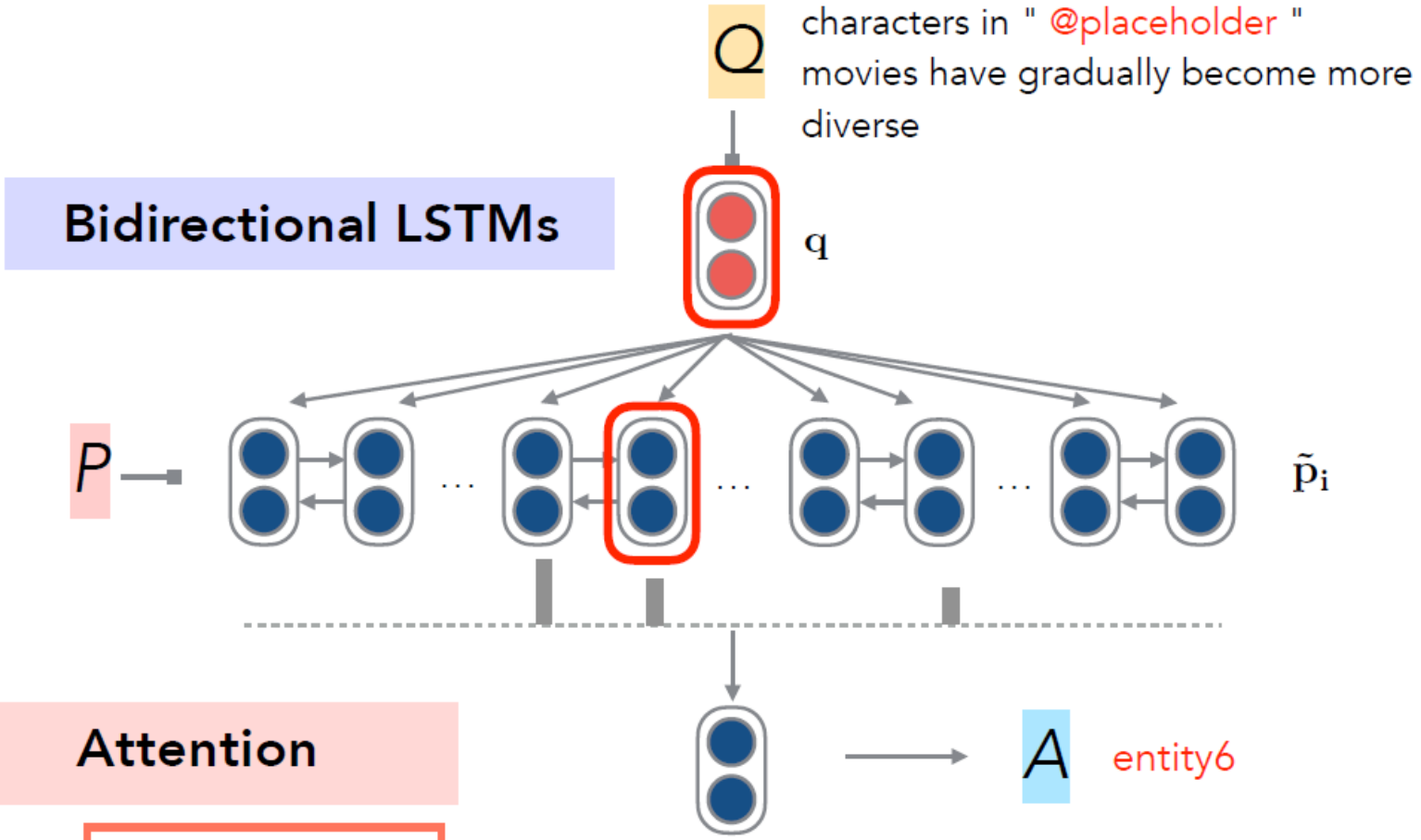
Q

characters in " @placeholder " movies have gradually become more diverse

A

@entity6

Stanford Attentive Reader



$$\alpha_i = \underset{i}{\text{softmax}}\left(\mathbf{q}^\intercal \mathbf{W}_s \tilde{\mathbf{p}}_i\right) \qquad \mathbf{o} = \sum_i \alpha_i \tilde{\mathbf{p}}_i \qquad a = \arg\max_{a \in p \cap E} W_a^\intercal \mathbf{o}$$

Stanford Question Answering Dataset (SQuAD)

- **Passage + Question -> Answer**

Who did **Genghis Khan** **unite** **before** **he** began **conquering** the rest of **Eurasia**?

He came to power by **uniting** many of the nomadic tribes of Northeast Asia. **After** founding the Mongol Empire and being proclaimed "**Genghis Khan**", he started the Mongol invasions that resulted in the **conquest** of most of **Eurasia**. These included raids or invasions of the Qara Khitai, Caucasus, Khwarezmid Empire, Western Xia and Jin dynasties. These campaigns were often accompanied by wholesale massacres of the civilian populations – especially in the Khwarezmian and Xia controlled lands. By the end of his life, the Mongol Empire occupied a substantial portion of Central Asia and China.

Multi-choice Dataset: RACE

- MCTest:
  - Small size: 2640 questions.
  - designed for 7-year-old children.
- RACE:
  - 28,000 passages and near 100,000 questions.
  - Collected from English exams for middle and high school Chinese students of age 12 – 18.

Alyssa got to the beach after a long trip. She's from Charlotte. She traveled from Atlanta. She' s now in Miami. She went to Miami to visit some friends

...

The girls went to a restaurant for dinner. The restaurant had a special on catfish. Alyssa enjoyed the restaurant's special. Ellen ordered a salad. Kristin had soup. Rachel had a steak.

1: one: Why did Alyssa go to Miami?
A) swim
B) travel
*C) visit friends
D) laying out

2: multiple: What did Alyssa eat at the restaurant?
A) steak
B) soup
C) salad
*D) catfish

An example of RACE dataset

Passage:

In a small village in England about 150 years ago, a mail coach was standing on the street. It didn't come to that village often. People had to pay a lot to get a letter. The person who sent the letter didn't have to pay the postage, while the receiver had to.
"Here's a letter for Miss Alice Brown," said the mailman.
" I'm Alice Brown," a girl of about 18 said in a low voice.
Alice looked at the envelope for a minute, and then handed it back to the mailman.
"I'm sorry I can't take it, I don't have enough money to pay it", she said.
A gentleman standing around were very sorry for her. Then he came up and paid the postage for her.
When the gentleman gave the letter to her, she said with a smile, " Thank you very much, This letter is from Tom. I'm going to marry him. He went to London to look for work. I've waited a long time for this letter, but now I don't need it, there is nothing in it."
"Really? How do you know that?" the gentleman said in surprise.
"He told me that he would put some signs on the envelope. Look, sir, this cross in the corner means that he is well and this circle means he has found work. That's good news."
The gentleman was Sir Rowland Hill. He didn't forgot Alice and her letter.
"The postage to be paid by the receiver has to be changed," he said to himself and had a good plan.
"The postage has to be much lower, what about a penny? And the person who sends the letter pays the postage. He has to buy a stamp and put it on the envelope." he said . The government accepted his plan. Then the first stamp was put out in 1840. It was called the "Penny Black". It had a picture of the Queen on it.

Questions:

1): The first postage stamp was made _.
A. in England B. in America C. by Alice D. in 1910

2): The girl handed the letter back to the mailman because _ .
A. she didn't know whose letter it was
B. she had no money to pay the postage
C. she received the letter but she didn't want to open it
D. she had already known what was written in the letter

3): We can know from Alice's words that _ .
A. Tom had told her what the signs meant before leaving
B. Alice was clever and could guess the meaning of the signs
C. Alice had put the signs on the envelope herself
D. Tom had put the signs as Alice had told him to

4): The idea of using stamps was thought of by _ .
A. the government
B. Sir Rowland Hill
C. Alice Brown
D. Tom

5): From the passage we know the high postage made _ .
A. people never send each other letters
B. lovers almost lose every touch with each other
C. people try their best to avoid paying it
D. receivers refuse to pay the coming letters

Answer: ADABC

Method: Hier-Co-Matching

- Inputs:
  - Passage: $\mathbf{P} \in \mathcal{R}^{d \times P}$ (get $P_1, P_2, \ldots P_n$ for each sentence in P)
  - Question: $\mathbf{Q} \in \mathcal{R}^{d \times Q}$
  - A candidate answer: $\mathbf{A} \in \mathcal{R}^{d \times A}$ (4 candidate answers in RACE dataset)
- Target:
  - Select a correct choice out of 4 candidates.
- Two phases:
  - **Co-Matching**: Match each sentence of P with Q and with A simultaneously to construct the vector for answer selection
  - **Hierarchical aggregation**: To connect all sentence-level matching vectors together.

Overview of the model

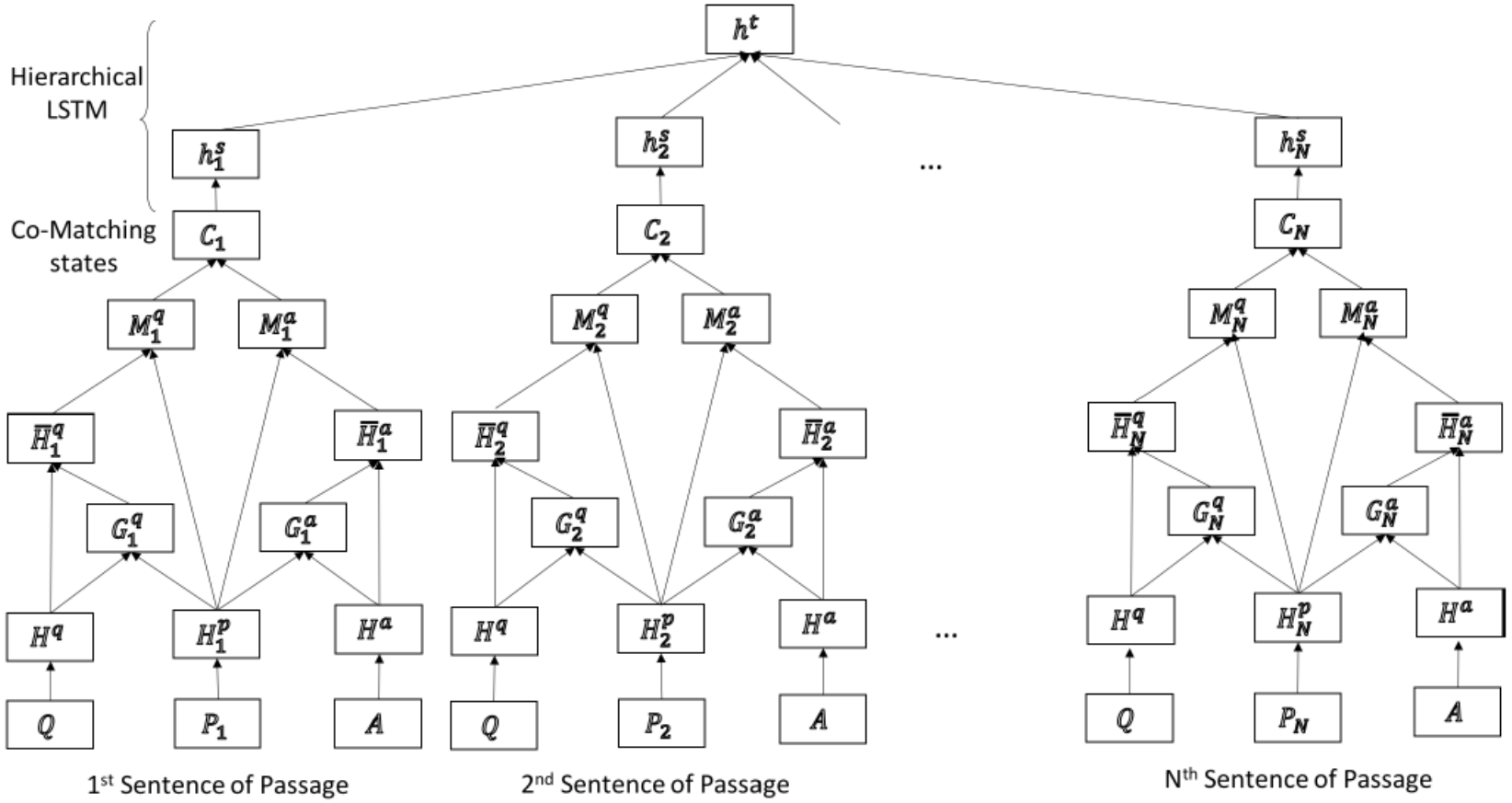


Figure 1: An overview of the model that builds a matching representation for a triplet $\{\mathbf{P}, \mathbf{Q}, \mathbf{A}\}$ (*i.e.*, passage, question and candidate answer).

Co-Matching

1. Use bidirectional LSTM to preprocess the text vectors.
   - $\mathbf{H}^{\mathrm{p}} = \mathrm{Bi} - \mathrm{LSTM}(\mathbf{P}); \mathbf{H}^{\mathrm{q}} = \mathrm{Bi} - \mathrm{LSTM}(\mathbf{Q}); \mathbf{H}^{\mathrm{a}} = \mathrm{Bi} - \mathrm{LSTM}(\mathbf{A})$
   - $\mathbf{H}^{\mathrm{p}} \in \mathcal{R}^{l \times P}; \mathbf{H}^{\mathrm{q}} \in \mathcal{R}^{l \times Q}; \mathbf{H}^{\mathrm{a}} \in \mathcal{R}^{l \times A}$.
   - The outputs are the hidden states generated from Bi-LSTMs.
2. Compute the **passage-question** and **passage-answer** attention vectors by attention mechanism.
   -
$$
\begin{aligned}
\mathbf{G}^{\mathrm{q}} &= \mathrm{SoftMax}\left((\mathbf{W}^{\mathrm{g}}\mathbf{H}^{\mathrm{q}} + \mathbf{b}^{\mathrm{g}} \otimes \mathbf{e}_Q)^{\mathrm{T}}\mathbf{H}^{\mathrm{p}}\right) \\
\mathbf{G}^{\mathrm{a}} &= \mathrm{SoftMax}\left((\mathbf{W}^{\mathrm{g}}\mathbf{H}^{\mathrm{a}} + \mathbf{b}^{\mathrm{g}} \otimes \mathbf{e}_Q)^{\mathrm{T}}\mathbf{H}^{\mathrm{p}}\right) \\
\overline{\mathbf{H}}^{\mathrm{q}} &= \mathbf{H}^{\mathrm{q}}\mathbf{G}^{\mathrm{q}}, \\
\overline{\mathbf{H}}^{\mathrm{a}} &= \mathbf{H}^{\mathrm{a}}\mathbf{G}^{\mathrm{a}},
\end{aligned}
$$
   - $\overline{\mathrm{H}}^{\mathrm{q}}$ / $\overline{\mathrm{H}}^{\mathrm{a}}$ is the weighted sum of all the question/answer hidden states.
   - $\mathrm{W}^{\mathrm{g}} \in \mathcal{R}^{l \times l}$ and $\mathrm{b}^{\mathrm{g}} \in \mathcal{R}^{l}$ are learned parameters.

Co-Matching

3. Co-Match:

$$\mathbf{M}^{\mathrm{q}} = \mathrm{ReLU}\left(\mathbf{W}^{\mathrm{m}}\begin{bmatrix}\overline{\mathbf{H}}^{\mathrm{q}} \ominus \mathbf{H}^{\mathrm{p}} \\ \overline{\mathbf{H}}^{\mathrm{q}} \otimes \mathbf{H}^{\mathrm{p}}\end{bmatrix} + \mathbf{b}^{\mathrm{m}}\right)$$

$$\mathbf{M}^{\mathrm{a}} = \mathrm{ReLU}\left(\mathbf{W}^{\mathrm{m}}\begin{bmatrix}\overline{\mathbf{H}}^{\mathrm{a}} \ominus \mathbf{H}^{\mathrm{p}} \\ \overline{\mathbf{H}}^{\mathrm{a}} \otimes \mathbf{H}^{\mathrm{p}}\end{bmatrix} + \mathbf{b}^{\mathrm{m}}\right)$$

$$\mathbf{C} = \begin{bmatrix}\mathbf{M}^{\mathrm{q}} \\ \mathbf{M}^{\mathrm{a}}\end{bmatrix},$$

- $\mathrm{W}^{\mathrm{m}} \in \mathcal{R}^{l \times 2l}$ and $\mathrm{b}^{\mathrm{m}} \in \mathcal{R}^{l}$ are learned parameters.
- $\mathrm{M}^{\mathrm{q}} \in \mathcal{R}^{l \times P}$ matches passage vector H^{p} and attended question vector $\overline{\mathrm{H}}^{q}$.
- $\mathrm{M}^{\mathrm{a}} \in \mathcal{R}^{l \times P}$ matches passage vector H^{p} and attended answer vector $\overline{\mathrm{H}}^{a}$.
- M^{a} and M^{q} are concatenated to get the co-matching state C.

Hierarchical Aggregation

- Split the passage P into N sentences $P_1, P_2, \ldots, P_N$.
- For each triplet $\{P_n, Q, A\}, n \in [1, N]$, we achieve a co-matching state C_n.
- Use a Bi-LSTM for each C_n and apply max-pooling
  - $\mathbf{h}_n^s = \text{MaxPooling}(\text{Bi-LSTM}(\mathbf{C}_n))$
  - $h_n^s \in \mathcal{R}^l$
- Concatenate all h_ns together:
  - $\mathbf{H}^s = [\mathbf{h}_1^s; \mathbf{h}_2^s; \ldots; \mathbf{h}_N^s]$
  - $H^s \in \mathcal{R}^{l \times N}$
- Use a higher-level LSTM to get the final representation vector:
  - $\mathbf{h}^t = \text{MaxPooling}(\text{Bi-LSTM}(\mathbf{H}^s))$
  - $h^t \in \mathcal{R}^l$

Answer Prediction

- For each candidate answer A_i, we can build a representation vector h_i^t.
- Use a softmax classifier to predict the answer. The loss function is as follows:

$$L(\mathbf{A}_i|\mathbf{P},\mathbf{Q}) = -\log \frac{\exp(\mathbf{w}^T\mathbf{h}_i^t)}{\sum_{j=1}^{4}\exp(\mathbf{w}^T\mathbf{h}_j^t)}.$$

Overview of the model

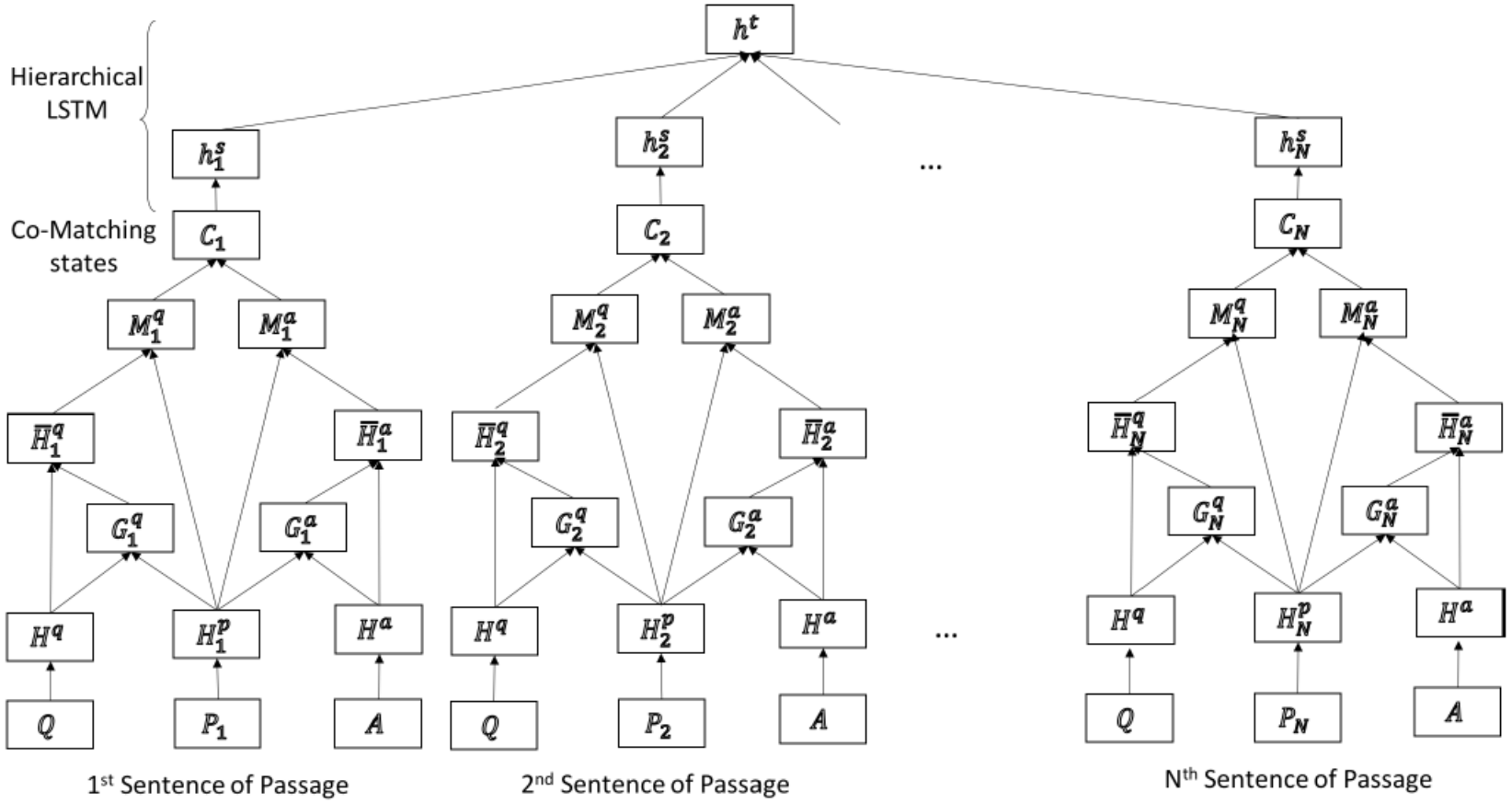


Figure 1: An overview of the model that builds a matching representation for a triplet $\{\mathbf{P}, \mathbf{Q}, \mathbf{A}\}$ (*i.e.*, passage, question and candidate answer).

Experiment results

- RACE dataset of three splits:
  - RACE-M: questions from middle school exams.
  - RACE-H: questions from high school exams.
  - RACE: The full dataset.
- Ablation study:
  - Remove Co-Matching module and Hier-Aggregation module respectively to see if the performance decreases.

Experiment results

	RACE-M	RACE-H	RACE
Random	24.6	25.0	24.9
Sliding Window	37.3	30.4	32.2
Stanford AR	44.2	43.0	43.3
GA	43.7	44.2	44.1
ElimiNet	-	-	44.7
HAF	45.3	47.9	47.2
MUSIC	51.5	45.7	47.4
Hier-Co-Matching	**55.8***	**48.2***	**50.4***
- Hier-Aggregation	54.2	46.2	48.5
- Co-Matching	50.7	45.6	46.4
Turkers	85.1	69.4	73.3
Ceiling	95.4	94.2	94.5

References

1. “Towards the Machine Comprehension of Text” by Danqi Chen, 2017.
2. Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy.RACE: Large-scale reading comprehension dataset from examinations. EMNLP 2017.
3. Wang S, Yu M, Chang S, et al. A Co-Matching Model for Multi-choice Reading Comprehension[J]. arXiv preprint arXiv:1806.04068, 2018.

SMILE LAB
Statistical Machine Intelligence & Learning

Thank you!