FLOWQA:GRASPING FLOW IN HISTORY FOR CONVERSATIONAL MACHINE COMPREHENSION

张璐 吴贺然 2018.12.20

single-turn models

- Question encoding
- Context encoding
- Reasoning
- Answer prediction

Conversational MC

- Existing methods
- just incorporate previous question/answer pairs into the current question and context encoding without modifying higher-level
- FLOWQA
- incorporate the conversation history more comprehensively via a conceptually simple **FLOW mechanism**

FLOW

- **Conversation flow** can be a representation based of the **context tokens** and inferred by **intermediate machine process** for answering previous questions
- Flow builds **information flow** from **the intermediate representation** $C_1^h, \ldots, C_{i-1}^h$ generated for the previous question $Q_1, \ldots, Q_{i-1}$ to the current process for answering Q_i ,for every h and i

- **Context integration**

  pass the current context representation C_i^h for each question i into a **BiLSTM layer**. **All question** i $(1 \leq i \leq t)$are processed **in parallel** during training

$$\hat{C}_i^h = \hat{c}_{i,1}^h, \ldots, \hat{c}_{i,m}^h = \text{BiLSTM}([C_i^h])$$

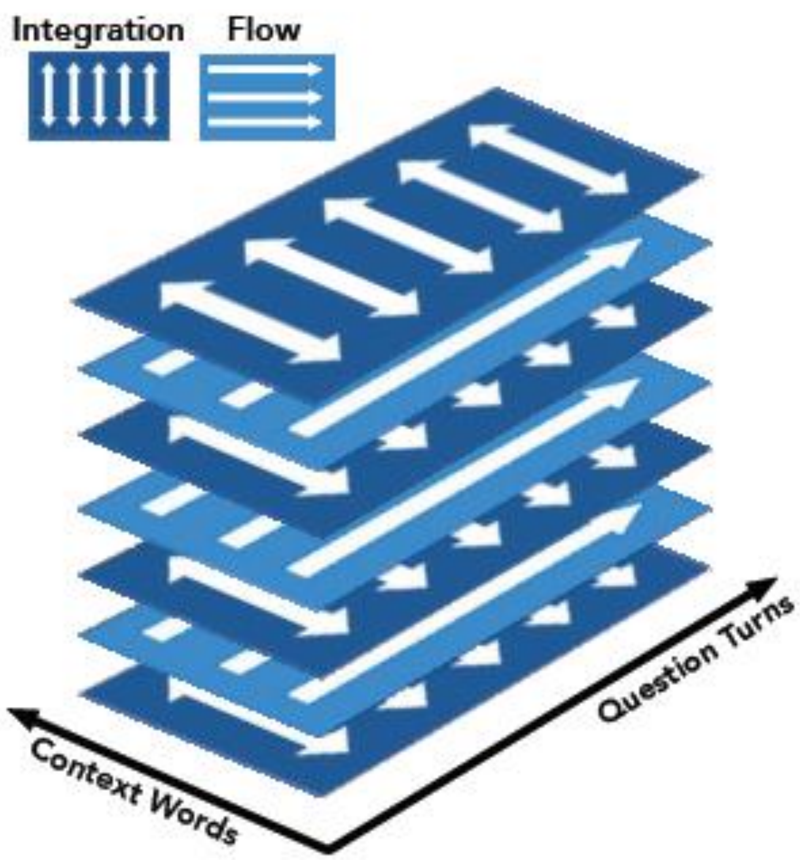

Figure 2: Alternating computational structure between context integration (RNN over context) and FLOW (RNN over question turns).

INTEGRATION-FLOW LAYER

- **FLOW**

  All **context word** $j (1 \le j \le m)$ are processed **in parallel**

$$f_{1,j}^{h+1}, \ldots, f_{t,j}^{h+1} = \mathrm{GRU}(\hat{c}_{1,j}^{h}, \ldots, \hat{c}_{t,j}^{h})$$

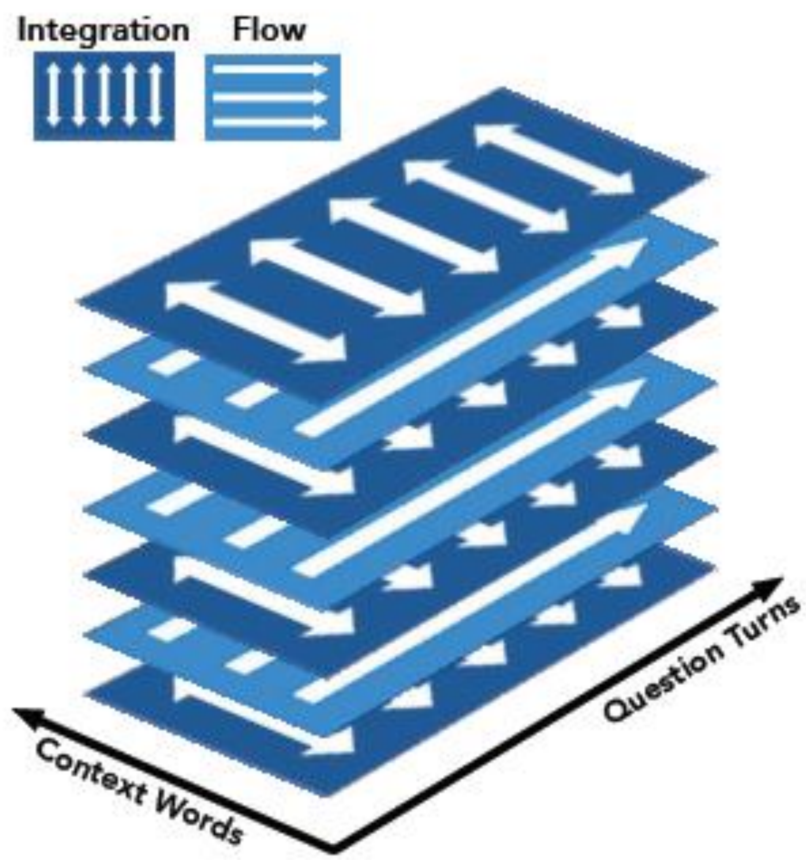

Figure 2: Alternating computational structure between context integration (RNN over context) and FLOW (RNN over question turns).

INTEGRATION-FLOW LAYER

- Reshape the outputs from the Flow layer back, and concatenate them to the output of the integration

$$F_i^{h+1} = \{f_{i,1}^{h+1}, \dots, f_{i,m}^{h+1}\}$$
$$C_i^{h+1} = c_{i,1}^{h+1}, \dots, c_{i,m}^{h+1} = [\hat{c}_{i,1}^{h}; f_{i,1}^{h+1}], \dots, [\hat{c}_{i,m}^{h}; f_{i,m}^{h+1}]$$

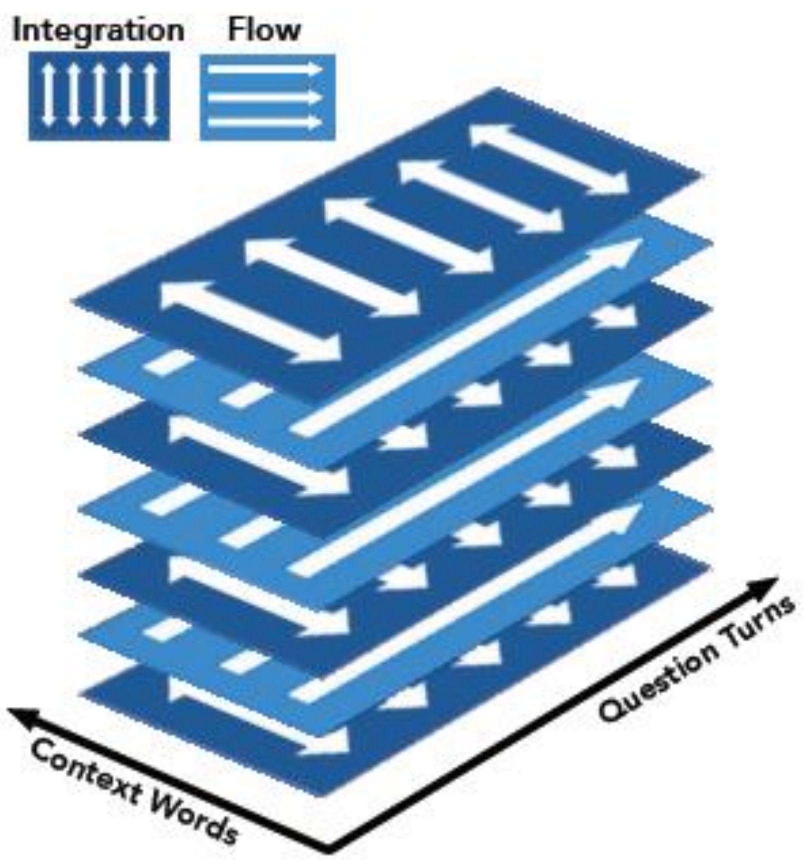

Figure 2: Alternating computational structure between context integration (RNN over context) and FLOW (RNN over question turns).

DETAILED DESIGN OF FLOWQA

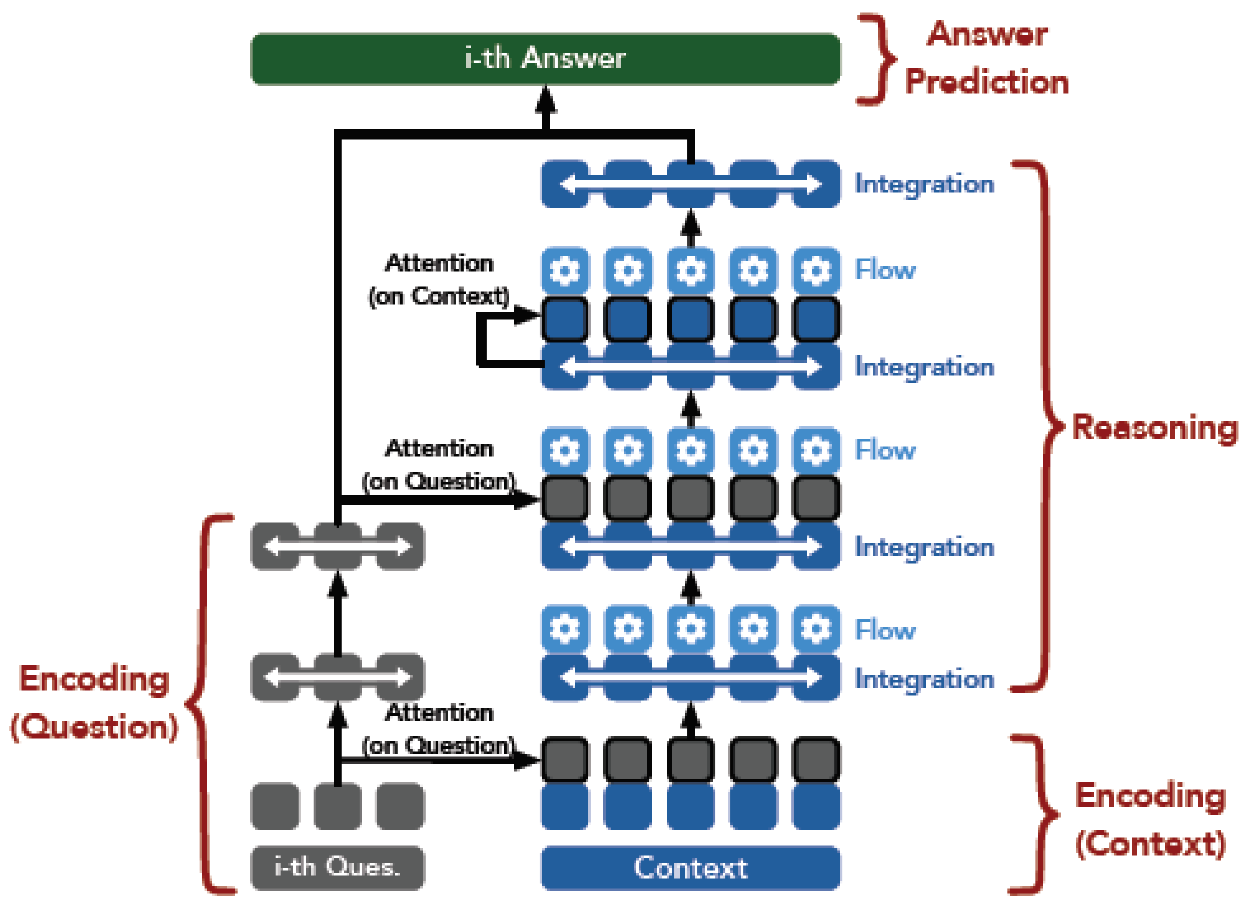

Figure 3: An illustration of the architecture for FLOWQA.

- FLOWQA is based on the single-turn MC structure with **fully-aware attention**

$$S(x, y) = \mathrm{ReLU}(\mathbf{U}x)^T \mathbf{D}\, \mathrm{ReLU}(\mathbf{U}y)$$

- **Initial encoding**
- **Reasoning**
- **Answer prediction**

QUESTION/CONTEXT ENCODING

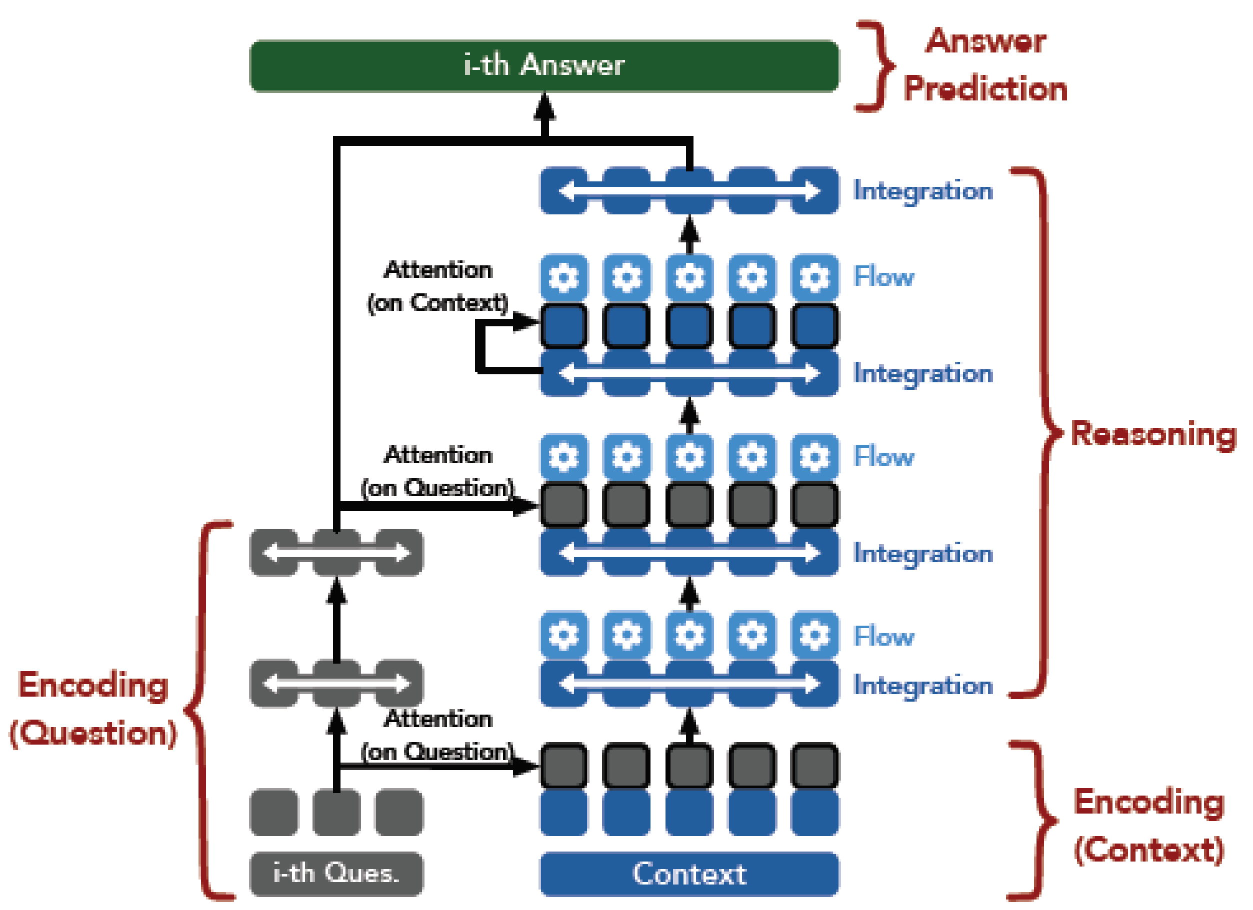

Figure 3: An illustration of the architecture for FlowQA.

- **Word embedding**

  Embed the context an question into a sequence of vectors with **pretrained GloVe, CoVE and ELMo**

$$C = \{c_1, \ldots, c_m\}$$

$$Q_i = \{q_{i,1}, \ldots, q_{i,n}\}$$

- **Attention (on Question)**

  in word level

$$g_{i,j} = \sum_k \alpha_{i,j,k}\, \mathrm{g}^Q_{i,k}, \quad \alpha_{i,j,k} \propto \exp(\mathrm{ReLU}(W\, \mathrm{g}^C_j)^T\, \mathrm{ReLU}(W\, \mathrm{g}^Q_{i,k})),$$

- **Word embedding**
- **A binary indicator**
- **Output from the attention**

$$C_i^0 = [c_1; \mathrm{em}_{i,1}; g_{i,1}], \ldots, [c_m; \mathrm{em}_{i,m}; g_{i,m}]$$

QUESTION/CONTEXT ENCODING

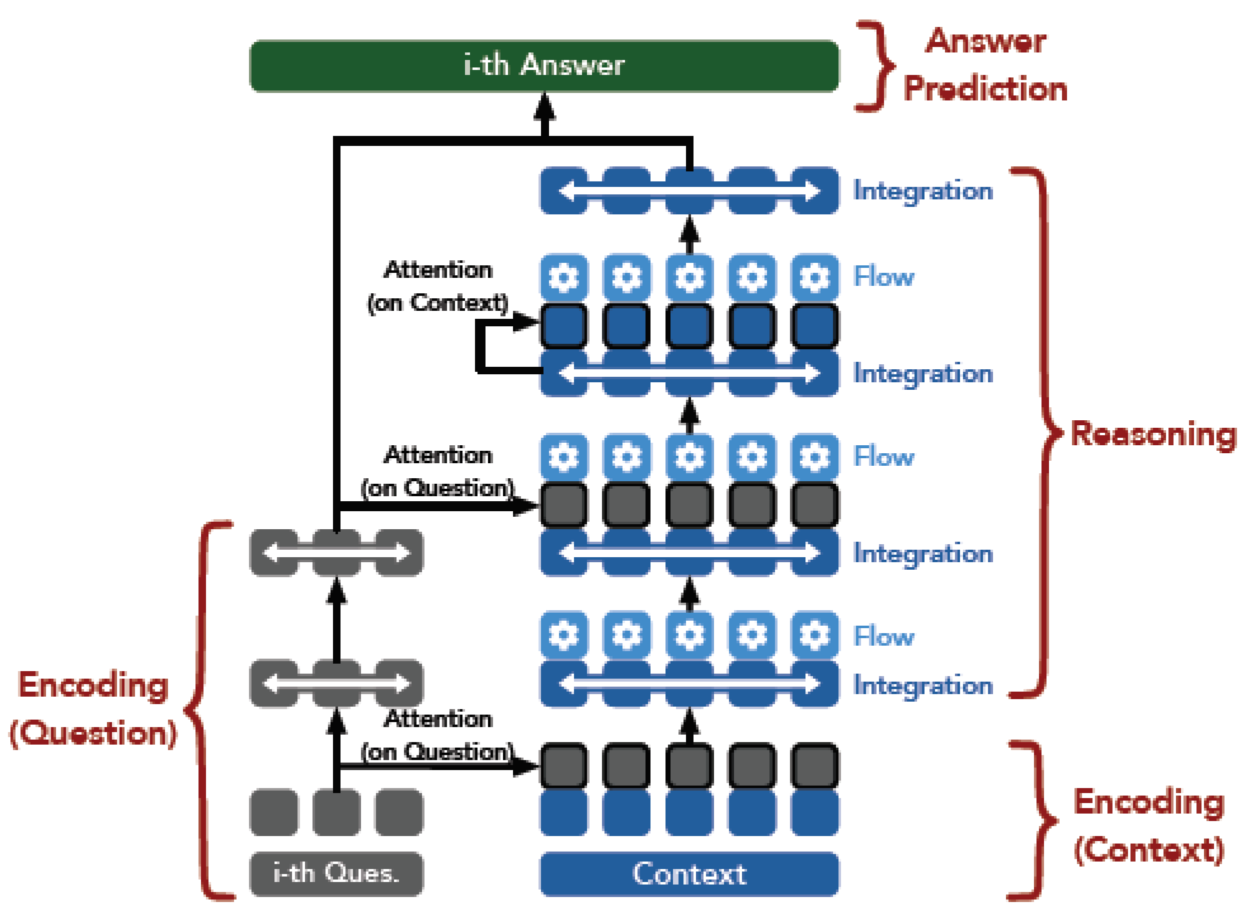

Figure 3: An illustration of the architecture for FLOWQA.

- **Question Integration with QHierRNN two layers BiLSTM**

$$Q_i^1 = q_{i,1}^1, \ldots, q_{i,n}^1 = \mathrm{BiLSTM}(Q_i)$$

$$Q_i^2 = q_{i,1}^2, \ldots, q_{i,n}^2 = \mathrm{BiLSTM}(Q_i^1)$$

- **Answer pointer vectors used in answer prediction layer**

$$\tilde{q}_i = \sum_{k=1}^{n} \alpha_{i,k} \cdot q_{i,k}^2, \ \alpha_{i,k} \propto \exp(w^T q_{i,k}^2)$$

$$p_1, \ldots, p_t = \mathrm{LSTM}(\tilde{q}_i, \ldots, \tilde{q}_t)$$

REASONING

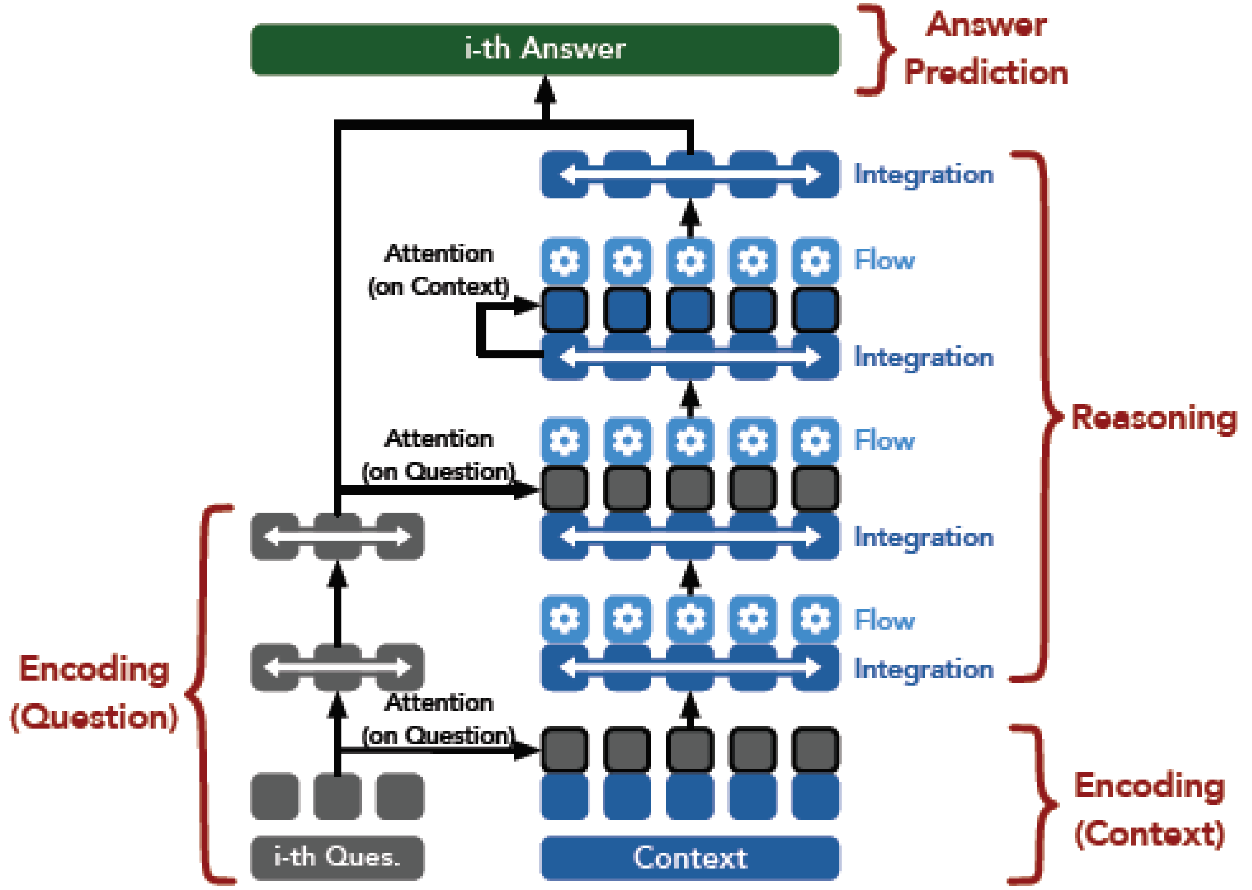

Figure 3: An illustration of the architecture for FlowQA.

- **Integration-FLOW *2**

$$C_i^1 = \mathrm{IF}(C_i^0)$$
$$C_i^2 = \mathrm{IF}(C_i^1)$$

- **Attention(on Question)**

$$\hat{q}_{i,j} = \sum_{k=1}^{n} \alpha^{i,j,k} \cdot q_{i,k}^2, \quad \alpha^{i,j,k} \propto \exp(S([c_i; c_{j,i}^1; c_{j,i}^2], [q_{j,k}; q_{j,k}^1; q_{j,k}^2]))$$

- **Integration-FLOW**

$$C_i^3 = \mathrm{IF}([c_{i,1}^2; \hat{q}_{i,1}], \ldots, [c_{i,m}^2; \hat{q}_{i,m}])$$

- **Attention(on Context)**

$$\hat{c}_{i,j} = \sum_{k=1}^{m} \alpha^{i,j,k} \cdot c_{j,k}^3, \quad \alpha^{i,j,k} \propto \exp(S([c_{i,j}^1; c_{i,j}^2, c_{i,j}^3], [c_{k,j}^1; c_{k,j}^2, c_{k,j}^3]))$$

- **Integration**

$$C_i^4 = \mathrm{BiLSTM}([c_{i,1}^3; \hat{c}_{i,1}], \ldots, [c_{i,m}^3; \hat{c}_{i,m}])$$

ANSWER PREDICTION

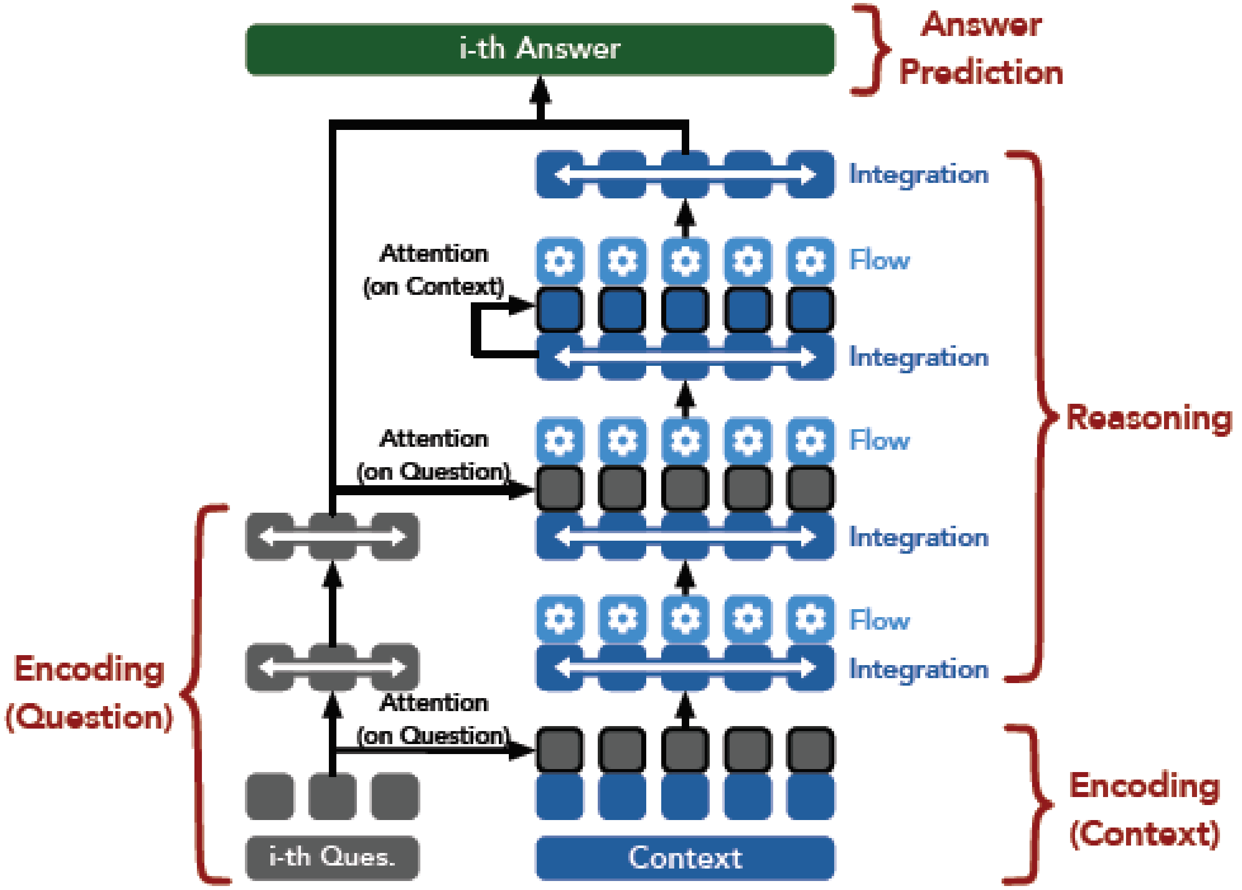

Figure 3: An illustration of the architecture for FlowQA.

- **Estimate the start and end probabilities**

$$P^S_{i,j} \propto \exp([c^4_{i,j}]^T W_S p_i), \quad \tilde{p}_i = \mathrm{GRU}(p_i, \sum_{i,j} P^S_{i,j} c^4_{i,j}), \quad P^E_{i,j} \propto \exp([c^4_{i,j}]^T W_E \tilde{p}_i)$$

- **Unanswerable question**

$$P^{\emptyset}_i \propto \exp\left(\left[\sum_{j=1}^{m} c^4_{i,j};\ \max_j c^4_{i,j}\right]^T W p_i\right)$$

EXPERIMENTS

	Child.	Liter.	Mid-High.	News	Wiki	Reddit	Science	Overall
PGNet (1-ctx)	49.0	43.3	47.5	47.5	45.1	38.6	38.1	44.1
DrQA (1-ctx)	46.7	53.9	54.1	57.8	59.4	45.0	51.0	52.6
DrQA + PGNet (1-ctx)	64.2	63.7	67.1	68.3	71.4	57.8	63.1	65.1
BiDAF++ (3-ctx)	66.5	65.7	70.2	71.6	72.6	60.8	67.1	67.8
FlowQA (1-Ans)	**73.7**	**71.6**	**76.8**	**79.0**	**80.2**	**67.8**	**76.1**	**75.0**
Human	90.2	88.4	89.8	88.6	89.9	86.7	88.1	88.8

Table 1: Model and human performance (% in F_1 score) on the CoQA test set. (N-ctx) refers to using previous N QA pairs. (N-Ans) refers to providing previous N gold answers.

EXPERIMENTS

	F_1	HEQ-Q	HEQ-D
Pretrained InferSent	20.8	10.0	0.0
Logistic Regression	33.9	22.2	0.2
BiDAF++ (0-ctx)	50.2	43.3	2.2
BiDAF++ (1-ctx)	59.0	53.6	3.4
BiDAF++ (2-ctx)	60.1	54.8	4.0
BiDAF++ (3-ctx)	59.5	54.5	4.1
FLOWQA (2-Ans)	**64.1**	**59.6**	**5.8**
Human	80.8	100	100

Table 2: Model and human performance (in %) on the QuAC test set. (baselines from (Choi et al., 2018))

	CoQA	QuAC
Prev. SotA (Yatskar, 2018)	70.4	60.6
FLOWQA (0-Ans)	75.0	59.0
FLOWQA (1-Ans)	**76.2**	64.2
- FLOW	72.5	62.1
- QHierRNN	76.1	64.1
- FLOW - QHierRNN	71.5	61.4
FLOWQA (2-Ans)	76.0	**64.6**
FLOWQA (All-Ans)	75.3	**64.6**

Table 3: Ablation study: model performance on the dev. set of both datasets (in % F_1).

FLOWQA:GRASPING FLOW IN HISTORY FOR CONVERSATIONAL MACHINE COMPREHENSION

张璐 吴贺然 2018.12.20