Omuokdlb at the NTCIR-17 QA Lab-PoliInfo-4 Task

Hidenori Yamato, Takaaki Fukunaga, Makoto Okada and Naoki Mori
Graduate School of Informatics, Osaka Metropolitan University

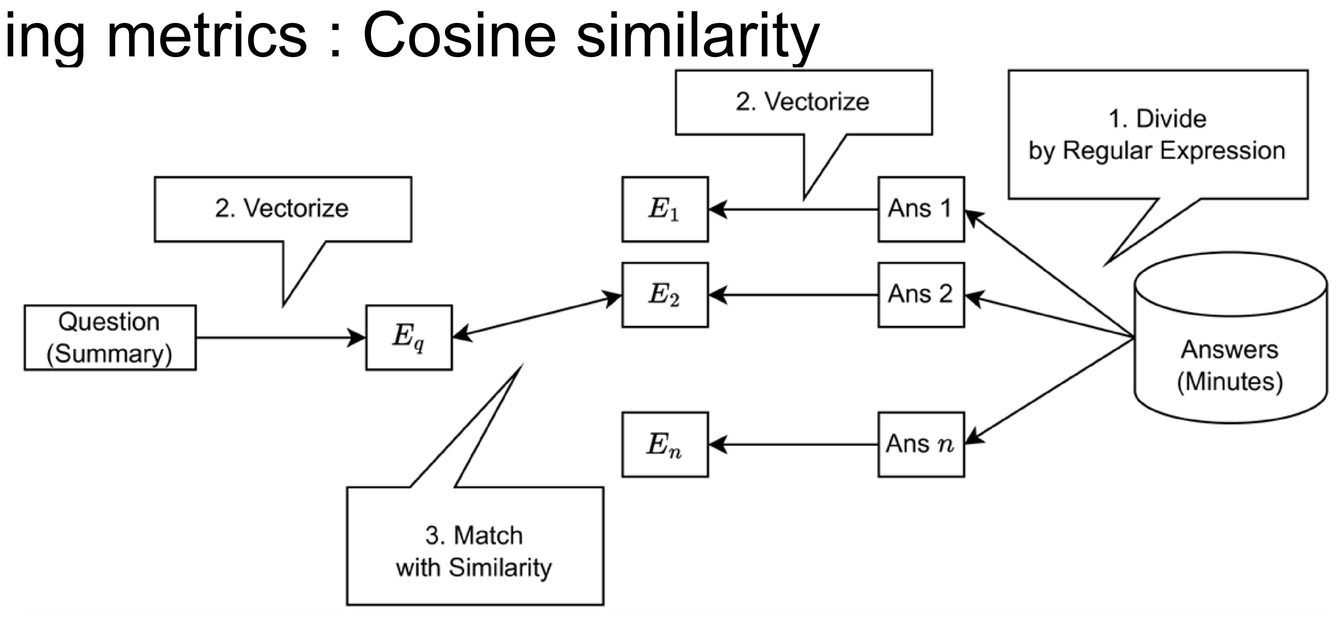
Question Answering-2

Our proposed method are constructed by the following two steps.

Step1. Alignment step

Matching the summary of the question utterances with the relevant part of the answer utterances.

- Paragraph separation method :
  - ditlab's method in the QA of QA Lab-PoliInfo-3[1]
- Model : BERT[2] (fine-tuned by Sentence-BERT[3])
  - Fine-tuning data : QA Alignment of QA Lab-PoliInfo-3
  - Input : Answer candidates or Question summary
  - Output : Sentence embeddings
- Matching metrics : Cosine similarity

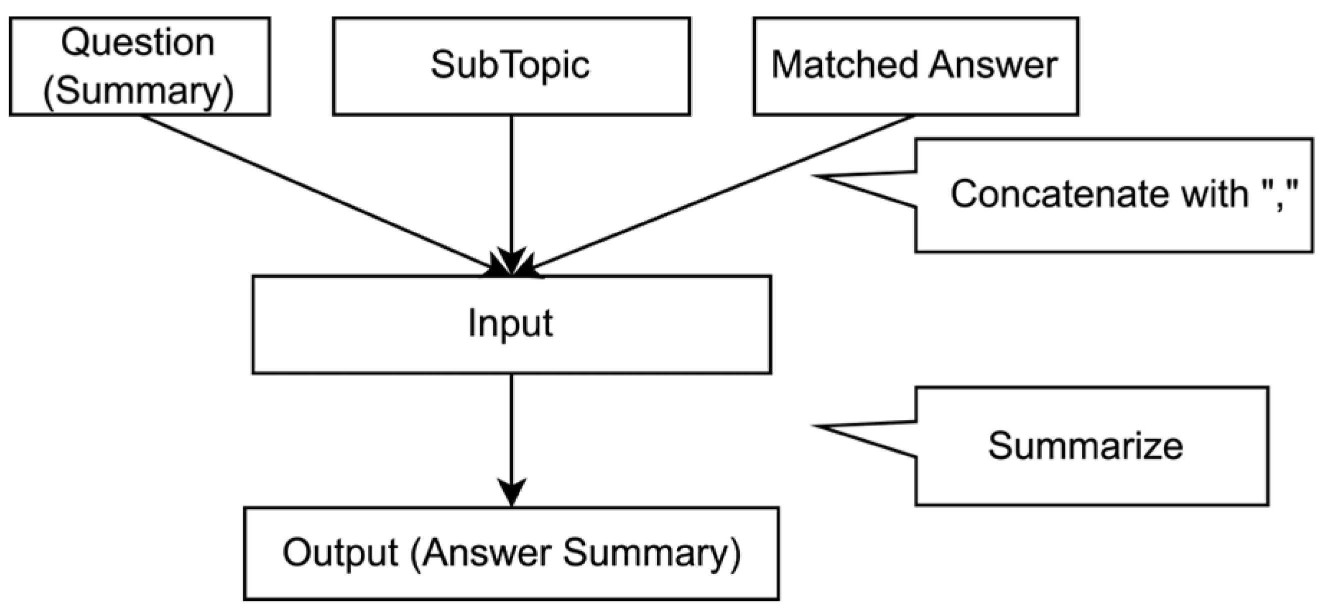


Overview of the alignment step

Step2. Summarization step

Summarizing answer utterances obtained from step 1.

- Model : T5 (Text-to-Text Transfer Transformer)[4]
  - Input : Question, SubTopic, and Matched Answer
  - Output : Summary of Answer
- Same as the baseline method (TO)



Overview of the summarization step

Result

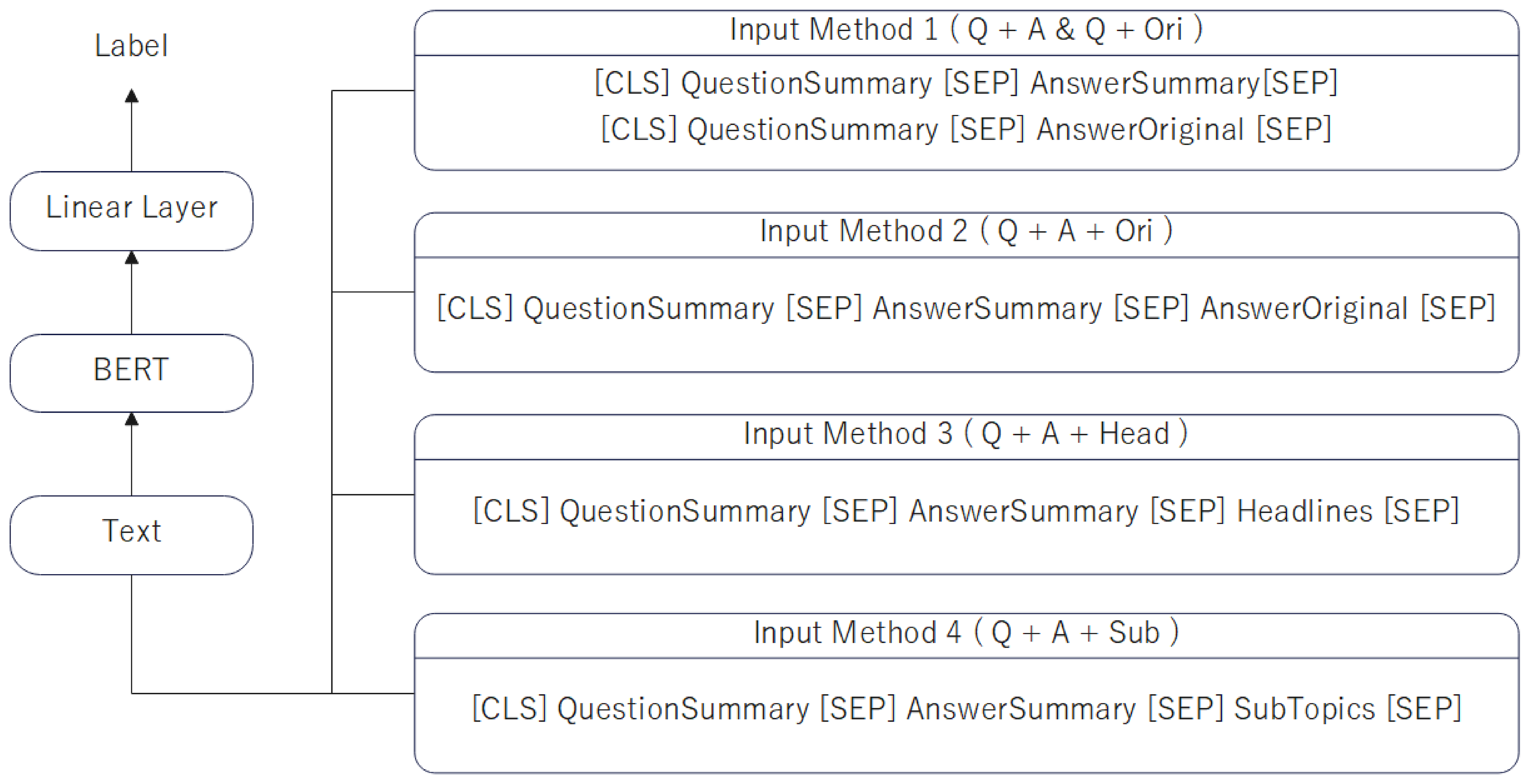
Result of the human evaluation

ID	Team	Correspondence				Content				Well-formed				Overall			
		A	B	C	Score	A	B	C	Score	A	B	C	Score	A	B	C	Score
Gold		93	6	1	192	47	47	6	141	96	3	1	195	69	28	3	166
153	ditlab	94	5	1	193	46	48	6	140	92	8	0	192	67	22	11	**156**
174	ditlab	94	6	0	**194**	47	48	5	**142**	84	12	4	180	65	25	10	155
116	IKM23	93	6	1	192	34	46	20	114	94	5	1	**193**	54	26	20	134
204	HUKB	85	10	5	180	23	61	16	107	87	9	4	183	46	39	15	131
134	omuokdlb	84	12	4	180	32	58	10	122	86	10	4	182	49	32	19	130
101	TO	86	8	6	180	35	49	16	119	89	6	5	184	48	29	23	125
192	AKBL	41	18	41	100	10	17	73	37	25	27	48	77	8	11	81	27

Answer Verification

Experimental method

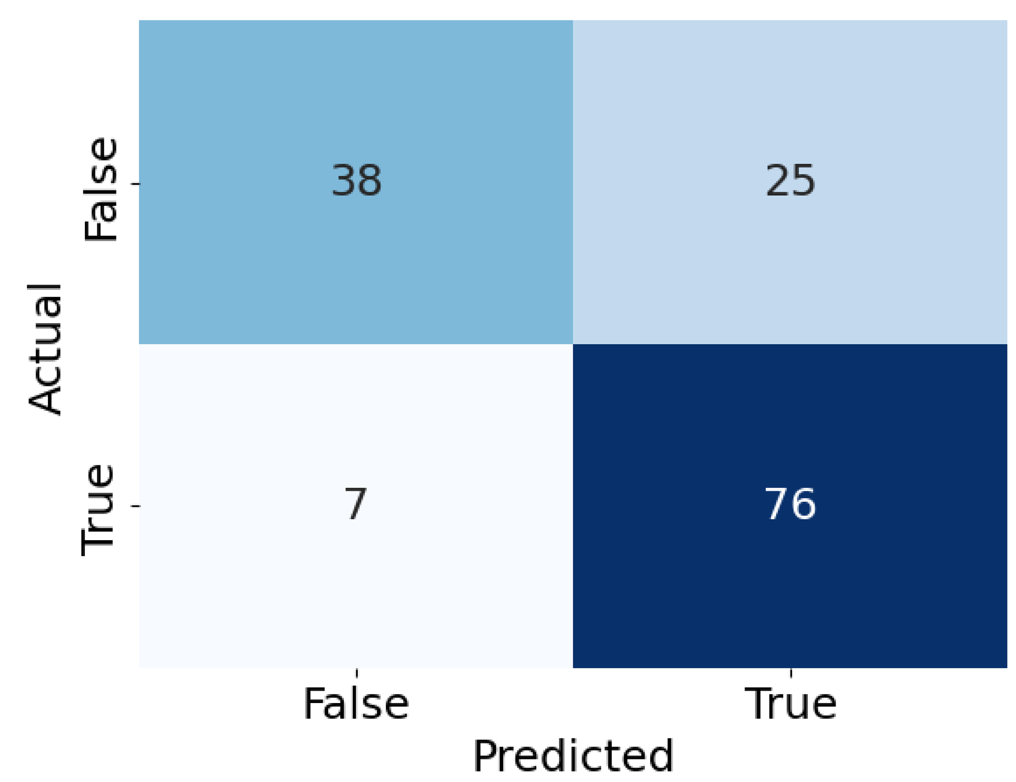
- Give minutes data to pre-trained BERT[2]
- Fine-tuning to determine correctness
- Split training data: test data= 4 : 1
- Four cross-validation with the training data



Result

Result for each input method

Input Method	Q + A & Q + Ori	Q + A + Ori	Q + A + Head	Q + A + Sub
$\text{Accuracy}_{\text{mean}}$	0.7483	0.7466	0.6969	0.7021
$\text{Accuracy}_{\text{variance}}$	4.446×10^{-4}	6.568×10^{-4}	1.994×10^{-4}	9.539×10^{-4}
F1_{mean}	0.7983	0.8059	0.6903	0.6820
$\text{F1}_{\text{variance}}$	7.289×10^{-4}	3.978×10^{-4}	2.893×10^{-4}	9.892×10^{-5}
$\text{Precision}_{\text{mean}}$	0.7432	0.7153	0.8618	0.8963
$\text{Precision}_{\text{variance}}$	8.945×10^{-4}	1.583×10^{-3}	6.594×10^{-4}	9.507×10^{-3}
$\text{Recall}_{\text{mean}}$	0.8594	0.9380	0.7628	0.7695
$\text{Recall}_{\text{variance}}$	7.452×10^{-4}	6.347×10^{-3}	9.743×10^{-5}	1.565×10^{-3}



Confusion matrix of prediction results at Q + A, Q + Ori
(False: Fake data, True : Fact data)

- To train using AnswerOriginal (Ori) improved the models' performance
- Precision of fake data is tend to be lower than one of fact data

1. Yuuki Tachioka and Atsushi Keyaki. 2022. ditlab at the NTCIR-16 QA Lab-Poliinfo3. proceeding of The 16th NTCIR Conference
2. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT:Pre-training of Deep Bidirectional Transformers for Language Understanding. Association for Computational Linguistics
3. Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In EMNLP-IJCNLP. Association for Computational Linguistics
4. Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. J. Mach. Learn. Res. 21