Overview of the NTCIR-14 QA Lab-PoliInfo Task

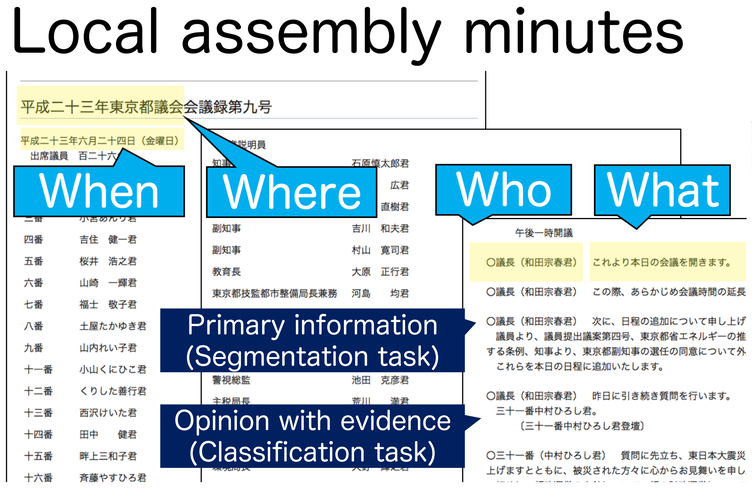
Yasutomo Kimura(OUC, RIKEN), Hideyuki Shibuki(YNU NII), Hokuto Ototake(FU),
Yuzu Uchida(HGU), Keiichi Takamaru(UKU), Kotaro Sakamoto(YNU NII), Madoka Ishioroshi(NII),
Teruko Mitamura(CMU), Noriko Kando(NII, SOKENDAI),
Tatsunori Mori(YMU), Harumichi Yuasa(IISEC), Satoshi Sekine(RIKEN) and Kentaro Inui(RIKEN)

What is NTCIR14 QALab-PoliInfo ?

The NTCIR-14 QA Lab-PoliInfo aims at real-world complex Question Answering (QA) technologies using Japanese political information such as local assembly minutes and newsletters.

	ID	Dry Run			Formal Run		
		Segmentation	Summarization	Classification	Segmentation	Summarization	Classification
1	FU01	-	-	1	-	-	3
2	FU02	-	-	1	-	-	2
3	KitAi	-	-	-	-	2	-
4	TTECH	-	1	4	-	1	10
5	nami	11	-	-	11	-	-
6	nagoy	-	1	-	-	1	-
7	akbl	1	2	1	3	2	1
8	ibrk	-	-	1	-	-	2
9	RICT	1	-	1	5	-	7
10	STARS	-	-	4	-	-	4
11	tmcit	-	-	1	-	-	6
12	KSU	2	1	-	4	6	8
13	CUTKB	-	-	-	-	-	1
14	LisLb	-	-	-	-	1	1
15	TO	1	1	-	1	1	-
	Sum	16	6	14	24	14	45

Local assembly minutes



Newsletters

2018 / 2019

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
	Kickoff		1st Meeting	2nd		Dry Run		3rd Meeting	4th	Formal Run				5th Meeting			

Dry run

Input	The minutes and a summary of an assembly member speeches
Output	The first and the last sentences of the original speech corresponding to each summary
Evaluation	Recall, precision, and F-measure of the concordance rate of the first and last sentences

Segmentation task

Dataset	Segmentation	Segmentation
Training	67 set	298 set
Test	50 set	83 set

	R	P	F
nami-01	0.464 (311/670)	0.342 (311/909)	0.394
nami-02	0.458 (307/670)	0.339 (307/905)	0.390
nami-03	0.391 (262/670)	0.373 (262/702)	0.382
nami-04	0.479 (321/670)	0.304 (321/1,057)	0.372
nami-05	0.473 (317/670)	0.301 (317/1,053)	0.368
nami-06	0.396 (265/670)	0.354 (265/748)	0.374
nami-07	0.509 (341/670)	0.283 (341/1,203)	0.364
nami-08	0.503 (337/670)	0.281 (337/1,199)	0.361
nami-09	0.416 (279/670)	0.342 (279/815)	0.375
nami-10	0.370 (248/670)	0.420 (248/591)	0.393
nami-11	0.582 (390/670)	0.270 (390/1,444)	0.369
akbl-01	0.703 (471/670)	0.390 (471/1,207)	0.502
RICT-01	0.484 (324/670)	0.694 (324/467)	0.570
KSU-01	0.399 (267/670)	0.859 (267/311)	0.545
KSU-02	0.391 (262/670)	0.856 (262/306)	0.537
TO-01	0.267 (179/670)	0.056 (179/3,195)	0.093

Field name	Explanation	Dry run	Formal run
ID	Identification code	○	○
Prefecture	Prefectural name	○	○
Date	*According to the Japanese calendar*	○	○
Meeting	*According to Togikai dayori*	○	○
MainTopic	*According to Togikai dayori*	○	○
SubTopic	*According to Togikai dayori*	○	○
Speaker	*Name of assembly member*	○	
Summary	*Description in Togikai dayori*	○	
QuestionSpeaker	*Name of assembly member*		○
QuestionSummary	*Description in Togikai dayori*		○
AnswerSpeaker	*Name of assembly member*		○
AnswerSummary	*Description in Togikai dayori*		○
StartingLine	Answer section	○	
EndingLine	Answer section	○	
QuestionStartingLine	Answer section		○
QuestionEndingLine	Answer section		○
AnswerStartingLine	Answer section		○
AnswerEndingLine	Answer section		○

Summarization task

Input	A speech of a member of assembly in the minutes and a limit length of the summary
Output	A summary corresponding to the speech
Evaluation	ROUGE scores and participants assessment in terms of content, formedness and total.

Dataset	Summarization	Summarization
Training	67 set	598 set
Test	50 set	146 set

The quality questions were assessed by a three-grade evaluation (i.e., A to C) from viewpoints of content, formedness and total.

$$QQ(v) = \frac{\sum_{s \in S} g(s,v)}{|S|} \qquad g(s,v) = \begin{cases} 2\ (gradeA) \\ 1\ (gradeB) \\ 0\ (gradeC) \\ a\ (gradeX) \end{cases}$$

Field name	Explanation	Dry run	Formal run
ID	Identification code	○	○
Prefecture	Prefectural name	○	○
Date	*According to the Japanese calendar*	○	○
Meeting	*According to Togikai dayori*	○	○
Speaker	Name of assembly member	○	○
StartingLine	The number of first sentence	○	○
EndingLine	The number of last sentence	○	○
Main topic	*According to Togikai dayori*	○	○
Sub topic	*According to Togikai dayori*	○	○
Summary	Answer section	○	○
Length	Limit length	○	○
Source	Speech of assembly member	○	○

		content		formed	total
		X=0	X=2		
TTECH-01		0.556	0.804	1.168	0.532
nagoy-01		0.156	0.204	0.856	0.168
akbl-01		0.644	1.036	1.656	0.784
akbl-02		0.608	0.968	1.664	0.744
KSU-01		0.000	0.000	0.064	0.000
TO-01		0.276	0.516	1.396	0.340
average		0.373	0.588	1.134	0.428

		recall							F-measure						
		N1	N2	N3	N4	L	SU	W1.2	N1	N2	N3	N4	L	SU4	W1.2
surface form	TTECH-01	0.363	0.114	0.072	0.045	0.322	0.157	0.161	0.261	0.075	0.044	0.027	0.226	0.102	0.148
	nagoy-01	0.131	0.031	0.013	0.003	0.115	0.047	0.059	0.116	0.021	0.009	0.003	0.101	0.038	0.063
	akbl-01	0.388	0.145	0.092	0.060	0.351	0.182	0.173	0.317	0.114	0.071	0.046	0.283	0.141	0.180
	akbl-02	0.373	0.127	0.076	0.046	0.335	0.166	0.165	0.307	0.100	0.059	0.036	0.272	0.130	0.172
	KSU-01	0.106	0.007	0.001	0.000	0.095	0.024	0.048	0.135	0.009	0.001	0.000	0.121	0.031	0.070
	TO-01	0.207	0.051	0.022	0.013	0.186	0.070	0.093	0.189	0.046	0.021	0.013	0.167	0.062	0.105

Classification task

Input	A political topic and a sentence in the minutes
Output	A class (support with fact-checkable reasons, against with fact-checkable reasons or other)
Evaluation	Accuracy of all classes, recall of each class, precision of each class and F-measure of each class.

Dataset	Classification	Classification
Training	344 set	10,201 set
Test	344 set	3,412 set

Field name	Explanation	Dry run	Formal run
ID	Identification code	○	○
Topic	Political topic	○	○
Utterance	A sentence in the minutes	○	○
Relevance	Answer section	-	○
Fact-checkability	Answer section	-	○
Stance	Answer section	-	○
Class	Answer section	○	○

	A	support			against			other		
		R	P	F	R	P	F	R	P	F
FU01-01	0.326	0.811	0.130	0.224	0.292	0.292	0.292	0.265	0.833	0.402
FU02-01	0.410	0.351	0.105	0.105	0.292	0.103	0.152	0.428	0.807	0.559
TTECH-01	0.642	0.405	0.278	0.330	0.667	0.200	0.308	0.671	0.905	0.771
TTECH-02	0.494	0.541	0.392	0.455	0.708	0.113	0.195	0.470	0.930	0.624
TTECH-03	0.712	0.270	0.400	0.322	0.583	0.215	0.314	0.781	0.870	0.823
TTECH-04	0.497	0.514	0.373	0.432	0.583	0.103	0.175	0.488	0.879	0.628
akbl-01	0.762	0.216	0.205	0.210	0.125	0.375	0.188	0.887	0.845	0.865
ibrk-01	0.823	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.823	0.903
RICT-01	0.820	0.000	NaN	NaN	0.042	0.333	0.075	0.993	0.824	0.901
STARS-01	0.823	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.823	0.903
STARS-02	0.823	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.823	0.903
STARS-03	0.823	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.823	0.903
STARS-04	0.823	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.823	0.903
tmcit-01	0.727	0.000	0.000	NaN	0.167	0.087	0.114	0.869	0.834	0.851

The best accuracy (i.e. 0.823) was achieved by ibrk-01 and all STARS.

For support, the best recall was 0.811 of FU01-01, the best precision was 0.400 of TTECH-03, and the best F-measure was 0.455 of TTECH-02.

For against, the best recall was 0.708 of TTECH-02, the best precision was 0.375 of akbl-01, and the best F-measure was 0.314 of TTECH-03.

For other, the best recall was 1.000 of ibrk-01 and all STARS, the best precision was 0.930 of TTECH-02, and the best F-measure was 0.903 of ibrk-01 and all STARS.

Formal run

Segmentation task

Input	The minutes and a pair of summaries of a question and the answer. An assembly member asks the question. A governor or staff answer the question.
Output	The first and the last sentences of the original speech corresponding to each summary
Evaluation	Recall, precision, and F-measure of the concordance rate of the first and last sentences

The best recall was 1.000 of nami-11, the best precision was 0.940 of nami-01, and the best F-measure was 0.895 of RICT-01.

	R	P	F
nami-01	0.814 (1,433/1,761)	0.940 (1,433/1525)	0.872
nami-02	0.864 (1,521/1,761)	0.851 (1,521/1,788)	0.857
nami-03	0.984 (1,733/1,761)	0.499 (1,733/3,475)	0.662
nami-04	0.639 (1,125/1,761)	0.805 (1,125/1,398)	0.712
nami-05	0.553 (973/1,761)	0.931 (973/1,045)	0.694
nami-06	0.655 (1,153/1,761)	0.657 (1,153/1,754)	0.656
nami-07	0.797 (1,404/1,761)	0.933 (1,404/1,505)	0.860
nami-08	0.831 (1,464/1,761)	0.932 (1,464/1,570)	0.879
nami-09	0.875 (1,541/1,761)	0.843 (1,541/1,827)	0.859
nami-10	0.993 (1,749/1,761)	0.464 (1,749/3,769)	0.632
nami-11	1.000 (1,761/1,761)	0.112 (1,761/15,765)	0.201
akbl-01	0.768 (1,352/1,761)	0.538 (1,352/2,515)	0.633
akbl-02	0.847 (1,492/1,761)	0.455 (1,492/3,282)	0.592
akbl-03	0.656 (1,155/1,761)	0.519 (1,155/2,227)	0.580
RICT-01	0.882 (1,554/1,761)	0.909 (1,554/1,709)	0.895
RICT-02	0.856 (1,507/1,761)	0.889 (1,507/1,695)	0.872
RICT-03	0.853 (1,503/1,761)	0.780 (1,503/1,926)	0.815
RICT-04	0.780 (1,374/1,761)	0.746 (1,374/1,842)	0.763
RICT-05	0.936 (1,648/1,761)	0.712 (1,648/2,314)	0.809
KSU-01	0.779 (1,372/1,761)	0.243 (1,372/5,643)	0.370
KSU-02	0.759 (1,337/1,761)	0.268 (1,337/4,998)	0.396
KSU-03	0.820 (1,444/1,761)	0.661 (1,444/2,185)	0.732
KSU-04	0.797 (1,403/1,761)	0.922 (1,403/1,521)	0.855
TO-01	0.354 (623/1,761)	0.898 (623/694)	0.508

Summarization task

Input	A speech of a member of assembly in the minutes and a limit length of the summary
Output	A summary corresponding to the speech
Evaluation	ROUGE scores and participants assessment in terms of content, formedness and total.

		recall							F-measure						
		N1	N2	N3	N4	L	SU4	W1.2	N1	N2	N3	N4	L	SU4	W1.2
Surface Form	KitAi-01	0.440	0.185	0.121	0.085	0.375	0.217	0.179	0.357	0.147	0.096	0.067	0.299	0.168	0.188
	KitAi-02	0.390	0.174	0.113	0.078	0.320	0.200	0.154	0.343	0.154	0.101	0.069	0.281	0.173	0.176
	TTECH-01	0.278	0.060	0.035	0.020	0.216	0.092	0.096	0.240	0.055	0.031	0.018	0.187	0.079	0.111
	nagoy-01	0.459	0.200	0.131	0.089	0.394	0.229	0.186	0.361	0.151	0.097	0.064	0.305	0.169	0.192
	akbl-01	0.400	0.173	0.113	0.076	0.345	0.189	0.157	0.361	0.156	0.102	0.068	0.310	0.167	0.185
	akbl-02	0.326	0.124	0.080	0.057	0.269	0.147	0.112	0.320	0.119	0.077	0.055	0.262	0.141	0.144
	KSU-01	0.158	0.028	0.009	0.002	0.147	0.043	0.071	0.210	0.039	0.013	0.004	0.196	0.059	0.107
	KSU-02	0.185	0.043	0.021	0.014	0.167	0.063	0.080	0.230	0.056	0.027	0.017	0.209	0.080	0.116
	KSU-03	0.172	0.036	0.008	0.002	0.157	0.050	0.075	0.211	0.043	0.011	0.003	0.192	0.062	0.106
	KSU-04	0.171	0.044	0.013	0.002	0.153	0.055	0.072	0.219	0.056	0.017	0.003	0.195	0.072	0.106
	KSU-05	0.227	0.029	0.010	0.002	0.195	0.064	0.089	0.231	0.029	0.010	0.003	0.196	0.065	0.110
	KSU-06	0.221	0.038	0.013	0.004	0.187	0.065	0.086	0.230	0.038	0.012	0.004	0.192	0.067	0.108
	LisLb-01	0.251	0.120	0.079	0.058	0.211	0.132	0.103	0.226	0.107	0.071	0.051	0.188	0.115	0.118
	TO-01	0.267	0.093	0.061	0.045	0.230	0.117	0.105	0.272	0.086	0.052	0.036	0.233	0.110	0.133

For ROUGE scores, nagoy-01 achieved the best scores except some cases.

Classification task

Input	A political topic and a sentence in the minutes
Output	A relevance (existence or absence), A fact-checkability (existence or absence), A stance (agree, disagree or other) A class (support with fact-checkable reasons, against with fact-checkable reasons or other)
Evaluation	Accuracy of all classes, recall of each class, precision of each class and F-measure of each class.

The best accuracy was 0.942 of TTECH-07, -08 and -10.

For support, the best recall was 0.731 of FU01-02, the best precision was 0.738 of KSU-03, -04, -07 and -08, and the best F-measure was 0.256 of TTECH-02.

For against, the best recall was 1.000 of CUTKB-04, the best precision was 0.207 of TTECH-05, and the best F-measure was 0.216 of TTECH-05.

For other, the best recall was 1.000 of TTECH-07, -08, -10, RICT-01, -05, -06 and STARS-01, the best precision was 0.947 of TTECH-02 and -05, and the best F-measure was 0.970 of TTHECH-07, -08 and -10.

	A	support			against			other		
		R	P	F	R	P	F	R	P	F
FU01-01	0.624	0.417	0.057	0.100	0.076	0.041	0.053	0.648	0.938	0.766
FU01-02	0.373	0.731	0.057	0.106	0.183	0.045	0.072	0.362	0.943	0.523
FU01-03	0.909	0.089	0.164	0.115	0.008	0.020	0.011	0.970	0.936	0.953
FU02-01	0.842	0.027	0.040	0.032	0.095	0.033	0.049	0.899	0.933	0.916
FU02-02	0.840	0.073	0.063	0.068	0.069	0.030	0.042	0.895	0.933	0.914
TTECH-01	0.923	0.046	0.163	0.072	0.015	0.133	0.027	0.987	0.935	0.960
TTECH-02	0.896	0.260	0.252	0.256	0.221	0.199	0.209	0.943	0.947	0.945
TTECH-03	0.919	0.116	0.254	0.159	0.069	0.200	0.103	0.978	0.938	0.958
TTECH-04	0.921	0.043	0.134	0.065	0.015	0.133	0.027	0.985	0.934	0.959
TTECH-05	0.897	0.251	0.251	0.251	0.225	0.207	0.216	0.944	0.947	0.945
TTECH-06	0.918	0.132	0.269	0.177	0.080	0.206	0.115	0.976	0.939	0.957
TTECH-07	0.942	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.942	0.970
TTECH-08	0.942	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.942	0.970
TTECH-09	0.926	0.000	0.000	NaN	0.000	NaN	NaN	0.982	0.941	0.961
TTECH-10	0.942	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.942	0.970
akbl-01	0.923	0.118	0.344	0.176	0.034	0.097	0.050	0.983	0.939	0.960
ibrk-01	0.731	0.178	0.063	0.093	0.202	0.045	0.074	0.770	0.934	0.844
ibrk-02	0.731	0.178	0.063	0.093	0.202	0.045	0.074	0.770	0.934	0.844
RICT-01	0.933	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.933	0.965
RICT-02	0.932	0.002	0.091	0.004	0.004	0.111	0.008	0.998	0.933	0.964
RICT-03	0.893	0.118	0.145	0.130	0.111	0.117	0.114	0.949	0.940	0.944
RICT-04	0.894	0.114	0.143	0.127	0.111	0.117	0.114	0.950	0.939	0.944
RICT-05	0.933	0.000	NaN	NaN	0.000	0.000	NaN	1.000	0.933	0.965
RICT-06	0.933	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.933	0.965
RICT-07	0.932	0.084	0.440	0.141	0.042	0.407	0.076	0.994	0.937	0.965
STARS-01	0.933	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.933	0.965
STARS-02	0.889	0.002	0.002	0.002	0.000	NaN	NaN	0.953	0.933	0.943
STARS-03	0.889	0.002	0.002	0.002	0.000	NaN	NaN	0.953	0.933	0.943
STARS-04	0.889	0.002	0.002	0.002	0.000	NaN	NaN	0.953	0.933	0.943
tmcit-01	0.875	0.282	0.139	0.186	0.000	NaN	NaN	0.925	0.943	0.934
tmcit-02	0.893	0.239	0.160	0.192	0.000	NaN	NaN	0.946	0.942	0.944
tmcit-03	0.873	0.296	0.142	0.192	0.000	NaN	NaN	0.922	0.943	0.932
tmcit-04	0.879	0.319	0.161	0.214	0.000	NaN	NaN	0.928	0.944	0.936
tmcit-05	0.898	0.267	0.189	0.221	0.000	NaN	NaN	0.950	0.942	0.946
tmcit-06	0.878	0.292	0.148	0.196	0.000	NaN	NaN	0.927	0.943	0.935
KSU-03	0.934	0.071	0.738	0.130	0.008	0.083	0.015	0.998	0.937	0.967
KSU-04	0.934	0.071	0.738	0.130	0.008	0.083	0.015	0.998	0.937	0.967
KSU-05	0.932	0.075	0.579	0.133	0.019	0.111	0.032	0.995	0.937	0.965
KSU-06	0.932	0.071	0.689	0.129	0.019	0.088	0.031	0.995	0.937	0.965
KSU-07	0.934	0.071	0.738	0.130	0.011	0.100	0.020	0.997	0.937	0.966
KSU-08	0.934	0.071	0.738	0.130	0.011	0.100	0.020	0.997	0.937	0.966
CUTKB-04	0.025	0.000	NaN	NaN	1.000	0.025	0.049	0.000	NaN	NaN
LisLb-01	0.914	0.021	0.065	0.032	0.037	0.080	0.051	0.976	0.935	0.955

We conducted in a dry run and a formal run,
which are including the segmentation, summarization, and classification tasks. Fifteen teams submitted 119 runs in total.