nagoy Team's Summarization System

at the NTCIR-14 QA Lab-PoliInfo

Yasuhiro Ogawa, Michiaki Satou, Takahiro Komamizu, Katsuhiko Toyama
Nagoya University

Motivation

Summarization for Japanese statutes

- using Random Forest
- based on sentence extraction

good result

Our method is also useful for assembly member speeches?

different imbalances

Proposed method

sentence extraction + sentence reduction

Features of Random Forest

- sentence position
- sentence length
- presence of a word
  - noun in the summary
  - occurs more than once
  - not top 20 in the source

活気の	ある	社会を	構築すべきと	考えますが	知事の	所見を伺います。
100	200	300	400	500	500	dependency depth
40	0	20	0	0	40	case information
0	0	13	108	0	61	frequency in all summaries
140	200	333	508	500	601	

Progressive Ensemble Random Forest

Document ID	111	106	23	19	92
# of sentences	45	11	34	8	13
N N N N N P x 5	0	0	0	0	1
N N N N N P x 4	0	0	0	1	1
N N N N P x 3	0	2	1	3	1
N N N P x 2	0	5	2	3	3
N P x 1	1	9	3	7	5

Evaluation results

ROUGE Scores

	recall							F-measure						
	N1	N2	N3	N4	L	SU4	W1.2	N1	N2	N3	N4	L	SU4	W1.2
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

Quality Question Scores

Training: 9,979 sentences

	all-topic				single-topic				multi-topic			
	content		formed	total	content		formed	total	content		formed	total
	X=0	X=2			X=0	X=2			X=0	X=2		
KitAi-01	0.856	1.134	1.732	0.912	0.953	1.170	1.660	0.995	0.745	1.092	1.815	0.815
KitAi-02	0.788	1.035	1.308	0.667	0.849	1.028	1.340	0.722	0.717	1.043	1.272	0.603
TTECH-01	0.290	0.644	1.783	0.402	0.274	0.575	1.755	0.401	0.310	0.723	1.815	0.402
nagoy-01	0.886	1.104	1.619	0.899	0.953	1.179	1.642	1.028	0.810	1.016	1.592	0.750
akbl-01	0.722	1.005	1.833	0.826	0.708	1.009	1.844	0.849	0.739	1.000	1.821	0.799
akbl-02	0.707	1.000	1.837	0.793	—	—	—	—	0.707	1.000	1.837	0.793
KSU-01	0.043	0.043	1.955	0.048	0.052	0.052	1.934	0.057	0.033	0.033	1.978	0.038
KSU-02	0.076	0.121	1.745	0.071	0.080	0.156	1.722	0.104	0.071	0.082	1.772	0.033
KSU-03	0.091	0.157	1.715	0.104	0.104	0.179	1.731	0.156	0.076	0.130	1.696	0.043
KSU-04	0.111	0.167	1.419	0.093	0.118	0.193	1.420	0.132	0.103	0.136	1.418	0.049
KSU-05	0.048	0.078	1.692	0.048	0.057	0.085	1.726	0.057	0.038	0.071	1.652	0.038
KSU-06	0.078	0.169	1.535	0.091	0.085	0.151	1.542	0.094	0.071	0.190	1.527	0.087
LisLb-01	0.720	0.942	1.237	0.591	0.722	0.920	1.349	0.684	0.717	0.967	1.109	0.484
TO-01	0.504	0.846	1.763	0.551	0.464	0.794	1.778	0.521	0.550	0.905	1.746	0.586
average	0.423	0.603	1.655	0.435	0.387	0.535	1.532	0.414	0.406	0.599	1.646	0.394

Comparison with RF with probability

	Precision				Recall				F-measure			
	closed	open			closed	open			closed	open		
		all	single	multi		all	single	multi		all	single	multi
proposed	0.963	0.446	0.481	0.417	0.967	0.523	0.526	0.520	0.965	0.481	0.503	0.463
1	0.860	0.465	0.482	0.450	0.785	0.437	0.432	0.441	0.821	0.450	0.456	0.446
2	0.967	0.471	0.464	0.477	0.893	0.406	0.411	0.402	0.929	0.436	0.436	0.436
3	0.973	0.520	0.560	0.483	0.875	0.452	0.495	0.412	0.921	0.484	0.525	0.444
4	0.983	0.526	0.588	0.466	0.886	0.462	0.526	0.402	0.932	0.492	0.556	0.432
5	0.987	0.511	0.553	0.473	0.886	0.457	0.495	0.422	0.933	0.483	0.522	0.446
all	1.000	0.523	0.571	0.477	0.896	0.457	0.505	0.412	0.945	0.488	0.536	0.442

Future works

- Improve sentence reduction
- Investigate the relationship between the proposed method and the RF with probability