Summarization Focusing on Polarity or Opinion Fragments in Blogs

Yohei Seki
Toyohashi University of Technology
(staying at Columbia University as a visiting scholar)
seki@ics.tut.ac.jp

Abstract

We present the *TUT* opinion summarization system which participated in the *TAC 2008*. The system consists of two modules: opinion/polarity automatic annotation module and fragment extraction module for summarization. Our research objective is to estimate the effectiveness of opinion/polarity annotation per sentence units for opinion summarization. The evaluation results showed that the polarity annotation is effective to improve the redundancy elimination and coherence.

1 Introduction

In this paper, we describe the *TUT* opinion summarization system developed in *TAC 2008 opinion summarization pilot track*. Our system is based on *TUT* opinion annotation system developed in the *NTCIR* workshop[1] *MOAT*[2]. There are two new challenging points:

1. The opinionated or polar sentences should be aligned to answer the questions with considering the context information.

2. Opinion annotation and summarization system should be implemented for the Blog test collection. (In the *NTCIR* workshop, the target document genre is newspaper article.)

For the first point, we implemented opinion/polarity fragment extraction system. The details will be explained in this paper. On the other hand, for the second point, we only add two new modules: (A) *body* or *comment* part detection module and (B) author detection module. We did not change the opinion annotation system itself using newspaper articles as a training data this time due to time constraints.

This paper is constructed as follows. In Section 2, we explain our system overview. Section 3 introduces our opinion annotation and polarity annotation approach. Section 4 gives the result in *TAC 2008 opinion pilot* and we discuss our results and clarify our contribution. Finally, we will give our conclusion and improvement points in future in Section 5.

2 System Overview

2.1 Task definition

We summed up the task definition in *TAC 2008 opinion pilot* briefly as follows:

- Generate question-focused summaries from multiple blogs up to 7,000 characters per each question.

- Source documents are from TREC BLOG06 test collection[3] relevant to 25 topics. The average size of document sets is 24.4 documents per topic.

- Answer snippets from *TAC 2008 Opinion QA track* are also provided, but organizers leave the decision to the participants whether we use it or not.

2.2 Our summarization strategy

Basically, we implemented an extractive summarization approach. However, to provide context, we regard up to three consecutive sentences as one unit (*a fragment*) and compute the importance of each unit. Three consecutive sentences are defined as follows:

1. All consecutive sentences should be in the same document.

2. All consecutive sentences should be in the same part (body or comment).

[1] http://research.nii.ac.jp/ntcir
[2] Multilingual Opinion Analysis Task
[3] http://ir.dcs.gla.ac.uk/test_collections/blog06info.html

3. All consecutive sentences should be written by the same author (one blog author or one commenter).

The weighting and redundancy elimination strategies are as follows.

- Three sentence units are weighted with cosine similarity to each question, the blog heading, and the answer snippets in each article.
- Summaries are created by extracting important units up to 7,000 characters.
- Redundant units are removed using the threshold of cosine similarity with other units in the summary.
- All units extracted in the summary are ordered chronologically by each question.

TUT system architecture is described in Figure 1.

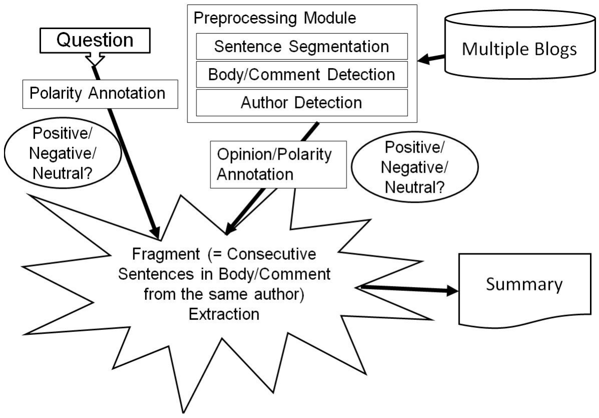

Figure 1: *TUT* System in *TAC 2008*

2.3 Details in each submission

We submitted three results, two results of which are evaluated officially. The details of two submissions are as follows.

1. Opinion-focused Summarization (*TUT2*: the second priority)
   - The system only extracted the units which contains at least one opinionated sentence (*opinion fragments*).
   - Opinionated sentences are automatically annotated using supervised machine learning approach. The details of opinion annotation are written in Section 3.1.
2. Polarity-focused Summarization (*TUT1*: the highest priority)
   - The system only extracted the units which contains at least one polar (= positive/negative) sentence requested by each question (*polarity fragments*) (i.e., "What motivated positive opinions of CARMAX from car buyers?" (= *positive*) or "what motivated negative opinions regarding purchasing a car from CARMAX?" (= *negative*)).
   - Polarities of questions and sentences are judged using several clue weighting learned from the analysis with *MPQA* corpus[4] and *NTCIR-6* English Opinion corpus[5]. The details of opinion annotation is written in Section 3.2.
   - Compared to the first approach, the summary construction is slightly changed to differentiate the polarity from each question when the one questions is positive and the other question is negative.

3 Opinion and Polarity Annotation

3.1 Opinion Annotation

Opinionated sentences are annotated using SVM approach. The original point of our approach is to differentiate (A) opinions written by the *author* and (B) quoted opinions expressed by the other *authority* because their writing styles were different. The selected features are as follows. They are selected based on the analysis with χ-square test using *MPQA* corpus and *NTCIR-6* English corpus, as shown in Table 4.

1. We utilized two type syntactic pairs: (a) subjects and verbs, (b) auxiliary verbs and verbs. Syntactic dependency was checked using Minipar (Lin, 2005).
2. Keyword list features were categorized by nouns, verbs, adjectives and adverbs, any part of speech (anypos) from the entries in the subjective lexicons (Wilson et al., 2005), and several other keywords.
3. We also used polarity term types. These features were determined using adjective entries (Hatzivassiloglou and Wiebe, 2000), which contained 1 914 word entries, and the General Inquirer (Stone, 2000), which contained 1,168 word entries.

The features are shown in Table 4. We clarified the entries in Table 4, as follows.

- The opinion verb types and the verb elements of syntactic pairs were defined based on the generalization using (A) communicative verbs entries in the lexicon

[4] http://www.cs.pitt.edu/mpqa/databaserelease/
[5] http://research.nii.ac.jp/ntcir/permission/ntcir-6/perm-en-OPINION.html

(Bloom et al., 2006) and (B) parts-of-speech with regard to the subjective lexicon (Wilson et al., 2005) and Minipar (Lin, 2005).

- The grammatical subject elements in syntactic pairs were generalized with (C) *ZeroProN* (in case they were missing), (D) named entity types, such as *GPE* or *PERCENT*, (E) case-sensitive pronouns, and (F) parts-of-speech with regard to Minipar.

- We used three count features: *cntopnoun*, *cntopadj*, and *cntopadv* that represented the numbers of the respective subjective nouns, adjectives, and adverbs in the sentence matched with the entries in the subjective lexicon (Wilson et al., 2005).

3.2 Polarity Annotation

Polarites are annotated only for the opinionated sentences using the number of positive/negative clues appeared in the sentence. The clues are selected based on the analysis with χ-square test using *MPQA* corpus and *NTCIR-6* English corpus. They are shown in Table 5. The polarity annotation strategy is as follows:

1. If more than three positive clues and more than three negative clues appeared in the opinionated sentence, we annotate the polarity of the sentence as "*BOTH*".

2. If the number of positive (negative) clues is more then the number of negative (positive) clues in the opinionated sentence, we annotate the polarity of the sentence as "*POS*" (*NEG*).

3. Otherwise, we annotate the polarity of the sentence as "*NEU*".

3.3 Accuracy of Opinion and Polarity Annotation

We experimented to estimate the accuracy of opinion and polarity annotation using newspaper articles in *NTCIR-7 MOAT* (Seki et al., 2008). For the polarity annotation, we implemented slightly different approach in *NTCIR-7 MOAT* using multi-label classification techniques, but we used the same clues in *TAC 2008*. The result is shown in Table 1.

Table 1: Accuracy of Opinion and Polarity Annotation in *NTCIR-7 MOAT*

	Precision	Recall	F-value
Opinion	0.3185	0.4092	0.3582
Polarity	0.1948	0.1830	0.1885

We used these classification clues with not changing for the blogs this time due to time constraints. In future, we should improve the accuracy for the opinion and polarity annotation in Blog data.

4 Results and Discussion

4.1 Results in *TAC 2008*

We have shown the result from *TAC 2008* organizer in Table 2. We also show the results by topics in Table 3.

Table 2: *TAC2008* Results

TeamID	F-score		Grammaticality		Non-Redundancy	
	Score	Rank	Score	Rank	Score	Rank
TUT1	0.132	29	5.591	10	6.545	8
TUT2	0.133	27	5.545	12	6.045	16
	Structure/Coherence		Fluency/Readability		Responsiveness	
	Score	Rank	Score	Rank	Score	Rank
TUT1	2.409	24	3.545	22	2.818	21
TUT2	2.318	29	3.591	18	3	16

The low result of F-score partially came from the misunderstanding of task definition. We created the summary based on the maximum length (7,000 characters by questions), but this seems too long. The precision seems low compared to other systems, as shown in Table 3. This defeat could be improved using threshold to include the fragments into the summary.

On the other hand, the *grammaticality* and *non-redundancy* evaluation results are above average. This proved that the sentence segmentation and redundancy elimination modules implemented well to some extent. For *non-redundancy* evaluation, *TUT1* with polarity annotation approach is quite effective and better than *TUT2* with the opinion annotation approach. This proves that polarity annotation is effective to eliminate redundant information from the summary.

Table 3: TAC2008 Results by Topics

Target ID	Topic	TUT1											TUT2										
		Precision		Recall		F-value		G	NR	S/C	R/F	OR	Precision		Recall		F-value		G	NR	S/C	R/F	OR
		score	rank	score	rank	score	rank	score					score	rank	score	rank	score	rank	score				
1001	Carmax	0.07	27	0.402	22	0.119	27	9	9	1	3	3	0.07	27	0.402	22	0.119	27	6	7	4	3	2
1003	Jiffy Lube	0.017	32	0.059	32	0.027	32	5	4	2	2	1	0.025	29	0.206	19	0.044	28	4	3	1	3	1
1004	Starbucks coffee shops	0.029	29	0.101	27	0.045	30	1	10	3	4	1	0	35	0	35	0	35	4	8	1	3	1
1005	Windows Vista	0.16	24	0.486	12	0.24	19	7	6	4	7	5	0.155	26	0.49	11	0.235	21	7	6	4	7	5
1008	UN Commission on Human Rights	0.064	34	0.373	22	0.11	33	2	3	2	3	2	0.048	35	0.296	26	0.083	35	2	3	2	2	1
1009	architecture of Frank Gehry	0.081	18	0.228	11	0.12	15	6	4	4	4	3	0.067	20	0.196	14	0.1	16	5	3	4	5	2
1010	Picasa	0.174	25	0.51	10	0.26	22	4	9	1	4	5	0.188	22	0.517	9	0.276	17	5	9	2	4	4
1018	MythBusters	0.181	26	0.467	14	0.261	21	5	7	3	3	3	0.102	33	0.207	27	0.136	32	3	5	2	3	2
1019	China one-child per family law	0.099	22	0.477	13	0.164	21	8	4	1	2	4	0.089	23	0.43	15	0.147	22	8	4	2	3	5
1021	Sheep and Wool Festival	0.096	34	0.343	29	0.15	35	7	9	5	6	3	0.18	26	0.682	9	0.285	21	7	9	5	6	6
1022	Subway Sandwiches	0.017	32	0.066	29	0.027	31	6	8	3	3	1	0.085	23	0.276	14	0.129	19	5	7	1	2	3
1024	Zillow	0.129	27	0.51	16	0.206	24	3	8	3	3	3	0.172	19	0.615	14	0.269	18	5	5	3	5	6
1026	criminalizing flag burning	0.058	28	0.185	23	0.088	27	5	6	2	4	3	0.066	26	0.235	19	0.104	22	5	6	4	6	4
1027	NAFTA	0.034	32	0.252	22	0.059	29	2	2	2	3	1	0.034	32	0.252	22	0.059	29	4	4	3	4	1
1030	System of a Down	0.205	16	0.654	14	0.312	12	6	7	2	3	4	0.203	17	0.63	15	0.307	13	8	9	3	3	4
1033	World Bank	0.033	35	0.065	33	0.044	33	8	8	6	4	2	0.041	33	0.111	27	0.06	31	6	5	2	2	4
1043	A Million Little Pieces	0.094	17	0.247	12	0.136	13	4	6	1	2	4	0.051	29	0.161	20	0.078	24	7	7	1	1	3
1044	talk show hosts	0.051	29	0.169	20	0.078	28	6	4	3	6	2	0.041	33	0.153	22	0.065	30	6	2	2	5	1
1045	women on Numb3rs	0.086	17	0.556	12	0.148	17	9	10	1	4	4	0.061	19	0.458	14	0.108	18	8	10	1	4	4
1047	Trader Joe's	0.038	31	0.077	29	0.051	30	6	6	2	2	1	0.052	30	0.111	25	0.071	27	6	6	2	2	1
1049	YouTube	0.101	29	0.451	18	0.165	24	7	9	1	4	4	0.119	23	0.511	15	0.193	18	5	9	1	4	5
1050	George Clooney	0.066	28	0.178	18	0.097	25	7	5	1	2	3	0.042	30	0.085	28	0.056	29	6	6	1	2	1
Avg.		0.086	30	0.312	18	0.132	29	5.6	6.5	2.4	3.5	2.8	0.086	29	0.319	16	0.133	28	5.5	6.0	2.3	3.6	3.0

G = Grammaticality
NR = Non-Redundancy
S/C = Structure/Coherence
R/F = Readablity/Fluency
OR = Overall Responsiveness

4.2 Discussion

We found 1018 as the most improved target with polarity annotation (*TUT1* over *TUT2*) and 1021 as the most degraded one. By focusing these targets, we investigated the difference using polarity annotation in our approach.

We found the different evaluation results sometimes caused by the judgment error of nuggets from the assessors, although both *TUT1* and *TUT2* summaries contain the same fragments relevant to the same pyramid nuggets. We also found the different results came from the failure of polarity annotation for sentences written in colloquial style such as tag questions, which sometimes written in blogs, but not contained in the newspaper articles.

5 Conclusion

We described our opinion summarization system based on the opinion and polarity annotation system. We have proved that polarity annotation is effective to eliminate the redundancy.

Obviously, we have several improvement points. The first point is that summaries seem to contain slightly off-topic fragments and must be combined with QA system. The second point is to improve fluency considering discourse structure, such as question-answering pairs used in e-mail summarization (McKeown et al., 2007). We also should estimate the threshold to create the proper amount of summary from multiple blogs. Finally, we also plan to improve the accuracy of opinion and polarity annotation by creating the training dataset using Blog data.

Acknowledgments

This work was conducted when the author visited in Prof. Kathleen McKeown at Columbia University and I appreciate her precious advice for the improvements in future. Note that the author is responsible for all the results this time.

This work is partially supported by the Overseas Advanced Research Practice Support Program from the Ministry of Education, Culture, Sports, Science and Technology, Japan. This work was also partially supported by the Artificial Intelligence Research Promotion Foundation in Japan.

References

K. Bloom, N. Garg, and S. Argamon. 2006. Extracting Appraisal Expressions. In *Proc. of the Human Language Technology Conf. of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL 2007)*, pages 308–315, Rochester New York, USA, April.

V. Hatzivassiloglou and J. M. Wiebe. 2000. Lists of manually and automatically identified gradable, polar, and dynamic adjectives. gzipped tar file. [cited 2005-8-26]. Available from: <http://www.cs.pitt.edu/ wiebe/pubs/coling00/coling00adjs.tar.gz>.

D. Lin. 2005. MINIPAR Home Page [online]. [cited 2005-8-26]. Available from: <http://www.cs.ualberta.ca/˜lindek/minipar.htm>.

Kathleen McKeown, Lokesh Shrestha, and Owen Rambow. 2007. Using question-answer pairs in extractive summarization of email conversations. In *Proceedings of CICLing*, volume 4394, pages 542–550.

Yohei Seki, David Kirk Evans, Lun-Wei Ku, Le Sun, Hsin-Hsi Chen, and Noriko Kando. 2008. Overview of Multilingual Opinion Analysis Task at NTCIR-7. In Noriko Kando, editor, *Proceedings of the Seventh NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering, and Cross-Lingual Information Access*, page (forthcoming), Tokyo 101-8430, Japan, Dec. National Institute of Informatics.

P. J. Stone. 2000. The General-Inquirer [online]. [cited 2005-8-26]. Available from: <http://www.wjh.harvard.edu/˜inquirer/spreadsheet_guide.htm>.

T. Wilson, J. Wiebe, and P. Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proc. of the 2005 Human Language Technology Conf. and Conf. on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, pages 347–354, Vancouver, B. C.

Appendix

Features used in opinion annotation

Table 4: Syntactic Pairs, Polarity Term Lists, and Keywords Clues Used in Author and Authority Opinion Extraction

Feature Type	Author Clues			Authority Clues		
"auxiliary verb" – "verb"	will	–	have	do	–	declare
	cannot	–	SbjVerb	to	–	be
	can	–	say	could	–	SbjVerb
	may	–	be	to	–	SbjVerb
"subject" – "verb"	WDT	–	SbjVerb	POS	–	NN
	NN	–	say	they	–	attitude
	I	–	VB	NNS	–	SbjVerb
	NN	–	VBZ	IN	–	judgment
	ZeroProN	–	conjecture	I	–	declare
	It	–	VBZ	*GPE*	–	VB
	it	–	JJ	*GPE*	–	VBG
	ZeroProN	–	declare	*ZeroProN*	–	SbjAdj
	NNS	–	VBD	I	–	admire
	they	–	VBP	We	–	VBP
	NNP	–	say	NN	–	SbjVerb
	WDT	–	VB	he	–	SbjVerb
	He	–	say	I	–	SbjVerb
	NNP	–	VBD	NNS	–	attitude
	it	–	VBZ	NNS	–	judgment
	ZeroProN	–	JJ	NNP	–	SbjVerb
	ZeroProN	–	VB	*PERCENT*	–	VBD
	DT	–	VBZ	*GPE*	–	SbjVerb
	ZeroProN	–	SbjVerb	he	–	declare
	It	–	VB	we	–	SbjAdj
	it	–	SbjVerb	he	–	SbjAdj
	—			we	–	VB
	—			NNS	–	say
	—			they	–	SbjVerb
	—			he	–	judgment
	—			IN	–	SbjVerb
	—			DT	–	SbjVerb
	—			I	–	VBP
	he–VBD,he–say,NN–VB,NN–SbjAdj					
subjective verb type	meet,include,demonstrate,SbjVerb,make, prevent,appear,be,seem,SbjNoun,become,were			judgment,express,denied,declare,tell,characterize, admire,advise,have,apologize,voice,expand		
	add,say					
subjective adjective/adverb	*cntopadj*,*cntopadv*,tragic,vicious,open,worse			unfair,angry,firmly		
subjective noun	*cntopnoun*,virtue,propaganda,failure,diplomacy, power,influence, enemy,doubt,right,humanity,resistance,excuse, stability			harassment,fear,opposition —		
subjective anypos	must,certainly,should,merely,unfortunately, real,perhaps,rather,seem,however			condemn —		
polarity term type	humaneness,education,defense,thing			report		
other keywords	",content,display,perpetrate,agency,discuss			relationship,century,spokesman,",ministry		

Table 5: Syntactic Pairs, Polarity Term Lists, and Keywords Clues Used in Positive and Negative Polarity Judgment

Feature Type	Positve Clues	Negative Clues
“auxiliary verb” – “verb”	to – promote	do – SbjVerb
	to – attract	do – admire
	to – set up	to – cover
	will – continue	to – remain
“subject” – “verb”	He – VBD	GPE – VBD
	I – VB	NN – SbjVerb
	I – VBN	EX – VBD
	NNP – VBZ	GPE – characterize
	PERSON – SbjAdj	GPE – say
	he – VBD	IN – characterize
	wood – say	IN – conjecture
	GPE – admire	IN – judgment
	GPE – judgment	JJ – VBD
	I – SbjAdj	NN – VBD
	I – VBP	NN – characterize
	NN – admire	NN – judgment
	NN – contribute to	NN – say
	NN – judgment	NNP – VBD
	NNP – SbjAdj	NNP – VBG
	NNP – VB	NNS – judgment
	NNP – judgment	NNS – say
	NNP – say	One – VBZ
	NNS – NN	PERSON – SbjVerb
	NNS – judgment	POS – NN
	PERSON – say	POS – NNP
	he – SbjAdj	She – say
	he – VBZ	WDT – SbjVerb
	he – judgment	WP – SbjVerb
	she – SbjAdj	ZeroProN – judgment
	—	she – say
subjective verb type	have,call,continue,play,bring,promote,strengthen,act, contribute,demonstrate,own,generate,broaden,be,admire, judgment,tell,express,contain,reduce,attract,voice,alter	were,advise,cover,pose,deliver,whitewash,SbjVerb,have, say,characterize,judgment,order,release,charge,draw, complain,plunge,gather,deem,term,notice,label,rely
subjective adjective/adverb	*cntopadj*,able,balanced,well,wonderful,ambitious,bright, colorful,confident,cooperative,credible,exemplary,glad,grateful,great, happy,jubilant,optimistic,peaceful,pleased,popular,positive	controversial,harmful,negative,wrong,antiAmerican,bad,cautious,central,disadvantageous, erroneous,evil,exclusive,hardline,illegitimate,impartial,intense,leftleaning,massive,odd,opportunistic, relevant,systematic,unfair,unfounded,unpopular,unrealistic,unreasonable,wary,widespread,firmly
subjective noun	breakthrough,comment,dream,genius,peace,persistence, player,pleasure,reconciliation,remark,respect,appreciation,approval, champion,cooperation,confidence,contribution,esteem,friendship, goodwill,gratitude,hope,knock,pledge,praise,recognition,reform, resolve,restoration,significance,split,support,supporter,understanding	danger,impression,lack,mistake,nature,reaction,sentiment,thought,abuse,accusation, activist,anger,blame,condemnation,constraint,critic,criticism,denunciation, destruction,discontent,dissatisfaction,fear,frustration,gaffe,harm,interference, intimidation,irregularity,motive,objection,opposition,outcry,protest,refusal,reluctance, shock,sorrow,starvation,suspicion,terrorism,threat,treason,violation,wrath,*cntopnoun*
subjective anypos	achievement,good,really,wonderful,although, champion,grateful,sensible,show,welcome	claim,furthermore,seriously,wrong,against,angrily,besides, condemn,critical,disapprove,erroneous,odd,too,unreasonable
polarity term type	*IPS*,quality,inhabitant,improvement,label,phenomenon,archetypal,order,Asian, transport,orientation,maneuver,contestant,compete,association,grow,right, speech,section,imagination,northbound,*POLP*,activity,capacity,clergyman, affect,argumentation,gathering,convey,unit,continent,decrease,degree,talk, advocate,support,agreement,approval,conversation,measure,wish,drive,feeling, cooperation,knowing,anticipation,meet,capitalist,keep,acceptance,equivalent, energy,union,furniture,affair,presentation,weekday,arouse,applaud,executive —	instrumentality,priesthood,substance,male,politician,affirm,Sinitic,note,kill,response, motion,attitude,island,damage,*INS*,express,*GRAP*,polity,state,*POLM*, organization,abstraction,government,vote,document,title,associate,composer,action,statement, coercion,judgment,charge,choice,division,care,spokesperson,race,disapproval,objection, decision,react,professional,hominid,reformer,comment,press,crime,attach,aggression,neckwear, put,anthropologist,reject,whole,hit,resident,knock,watch,designate,complain,emotion,accusation, ellipse,display,anger,beat,denial,Russian,prejudice,penetrate,poverty,weekday,encase,larceny, dismiss,disapprove,formulation,discontentment,handwear,reorient,misbehavior,disappointment