

Task Definition and Integration For Scientific-Document Writing Support

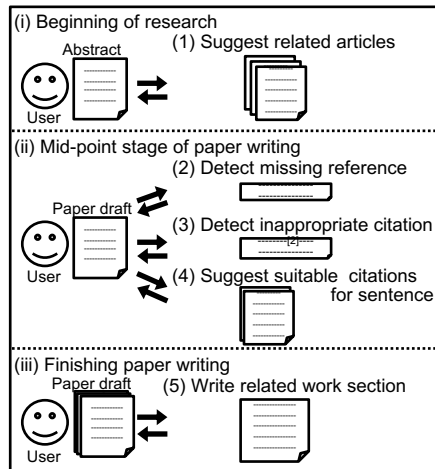
Hiromi Narimatsu¹, Kohei Koyama², Kohji Dohsaka³, Ryuichiro Higashinaka¹, Yasuhiro Minami², Hirotoishi Taira⁴

¹NTT Communication Science Laboratories, ²The University of Electro-Communication, ³Akita Prefectural University, ⁴Osaka Institute of Technology

Abstract

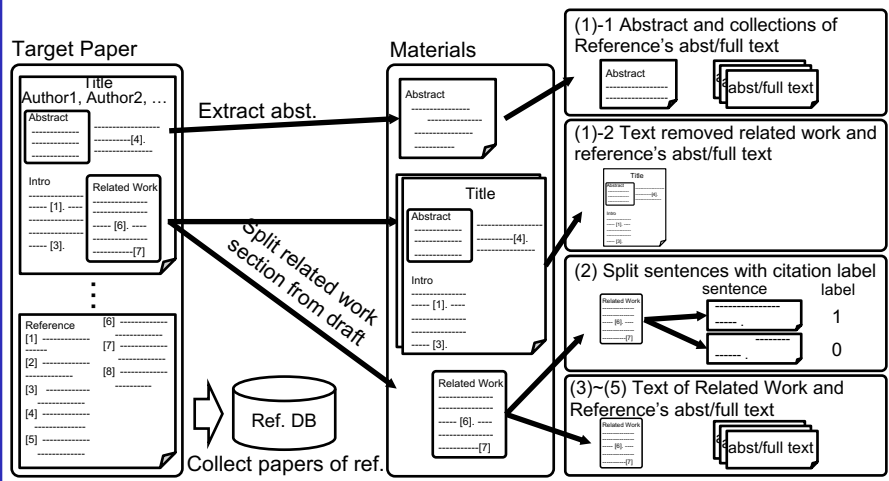
To achieve better support of scientific-paper writing, we (i) defined a series of tasks that can be pipelined, (ii) created a dataset of academic papers for the evaluation of each task as well as an integrated series of the tasks, and (iii) performed the evaluation of integrated tasks.

Task Definition



Task	Definition
(1)-1 Citation extraction	For an abstract, retrieve relevant papers from a large DB of scientific documents.
(1)-2 Citation recommendation for draft paper	For a draft paper, retrieve relevant papers from a large DB of scientific documents.
(2) Citation worthiness	Detect whether a sentence needs citations.
(3)-1 Citation allocation	Allocate appropriate papers to the sentences which need citation from a small DB of scientific documents.
(3)-2 Sentence-citation pair classification	Classify whether the allocation of the citation is appropriate or not. (sub-task of (3)-1)
(4) Citation recommendation for sentence	For the sentences which need citation in tasks (2) and (3), retrieve and allocate appropriate citations from a large DB of scientific documents.
(5)-1 Citation categorization	Categorize the cited papers on the basis of their topics.
(5)-2 Citation sentence generation	Suggest alternative citation sentences to achieve better clarity and fluency.
(5)-3 Citation text generation	Suggest alternative citation text for the content.

Data Creation



Evaluation

	# Examples	# Positive examples	# Negative examples
(2) Citation worthiness	1,137,168	461,882	675,286
(3)-1 Citation allocation	586	N/A	N/A
(3)-2 Sentence-citation pair classification	41,582	20,791	20,791

(2) Citation worthiness

	Acc.	Prec.	Rec.	F1
Random	0.499	0.408	0.500	0.449
BERT	0.911	0.925	0.852	0.887

(3)-2 Sent.-cite. pair classification

	Acc.	Prec.	Rec.	F1
Random	0.488	0.489	0.492	0.490
Doc2Vec	0.558	0.541	0.763	0.633
BERT	0.816	0.822	0.806	0.814
XLNet	0.844	0.846	0.841	0.843

(3)-1 Cite. allocation

	Acc.
Random	0.280
Doc2Vec	0.349
BERT	0.747
XLNet	0.795

(2)~(3) Integrated task

	Acc.
BERT	0.623

Our dataset is available at <https://github.com/citation-minami-lab/citation-dataset>.