Task Definition and Integration For Scientific-Document Writing Support

Hiromi Narimatsu¹, Kohei Koyama², Kohji Dohsaka³, Ryuichiro Higashinaka¹, Yasuhiro Minami², Hirotoshi 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

(i) Beginning of research

Abstract

(1) Suggest related articles

User

(ii) Mid-point stage of paper writing

Paper draft

(1) Suggest suitable citations for sentence

User

(2) Detect missing reference

(3) Detect inappropriate citation

(4) Suggest suitable citations for sentence

(iii) Finishing paper writing

Paper draft

(5) Write related work section

---

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

Target Paper

| Author1, Author2, … |
|---------------------|
| Extract abst. |
| Ref. DB |

Materials

| Abstract |
|---|
| (1)-1 Abstract and collections of Reference’s abst/full text |
| Title |
| (1)-2 Text removed related work and reference’s abst/full text |
| Related Work |
| (5)-5 Text of Related Work and Reference’s abst/full text |

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 |

---

| Task | Acc. | Prec. | Rec. | F1 | Acc. | Prec. | Rec. | F1 | Acc. |
|---|---|---|---|---|---|---|---|---|---|
| (2) Citation worthiness | Random | 0.499 | 0.408 | 0.500 | 0.449 | Random | 0.489 | 0.492 | 0.490 | Random | 0.280 |
| | BERT | 0.911 | 0.925 | 0.852 | 0.887 | Doc2Vec | 0.558 | 0.541 | 0.763 | 0.633 | Doc2Vec | 0.349 |
| (3)-2 Sent.-cite. pair classification | | | | | | | | | | | |
| | BERT | 0.816 | 0.822 | 0.806 | 0.814 | BERT | 0.747 |
| (2) Integrated task | | | | | | | | | | |
| | BERT | 0.623 | | | | XLNet | 0.844 | 0.846 | 0.841 | 0.843 | XLNet | 0.795 |

Our dataset is available at [https://github.com/citation-minami-lab/citation-dataset](https://github.com/citation-minami-lab/citation-dataset).