SolutionTailor: Scientific Paper Recommendation Based on Fine-Grained Abstract Analysis

Tetsuya Takahashi
Marie Katsurai
Doshisha University
katsurai@mm.doshisha.ac.jp

Approaches of Scientific Paper Recommendation

* Content-based document similarity calculation
  - Textual features (e.g., TFIDF and BERT)
* Collaborative filtering and graph-based algorithms
  - User's past research history (e.g., co-authorship and citations)

recommended papers

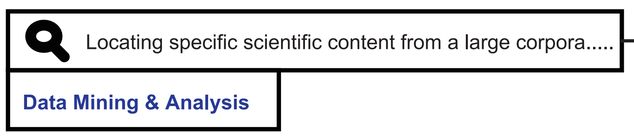
May overlap with search engine's results

Why and how similar to the user's query?
--- Unknown

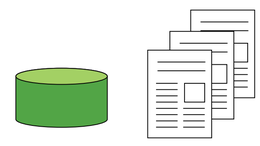
Proposed Framework

... Recommending literature on diverse research methodologies in a specific research objective

Input:
- Query abstract
- Target research field



Dataset:
- Publications of top venues in the specified field
- Downloaded from SemanticScholar



Step A) Sentence labeling
- Sequential sentence classification [Cohan+ 2019]
- Detect sentences corresponding to "Background/Objective (BO)" or "Method (M)"

BO and M sentences of the query abstract

BO and M sentences of the abstract in the dataset

Step B) Sentence embedding
- SciBERT [Beltagy+ 2019]
- Calculate the embeddings of BO and M sentences, respectively

BO and M embeddings of the query abstract

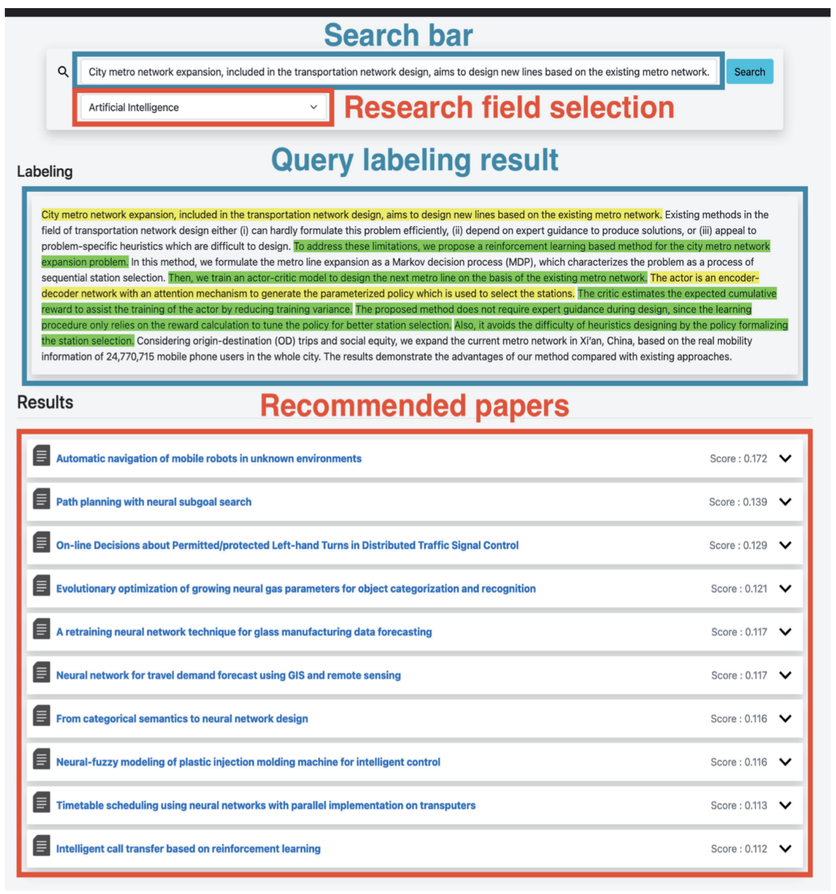
BO and M embeddings of the abstract in the dataset

Step C) Similarity score calculation
- Calculate the cosine similarities between BO or M embeddings, respectively
- The final score: $score = \cos_{BO} - \cos_{M}$

Output:
- Top-10 papers having the highest scores

Recommendation Interface

https://mm.doshisha.ac.jp/sci2/SolutionTailor.html



Evaluation

Overlap with citations:

- Does the system provide new insights for the user into a specific research background?
- We calculate the overlap between each of 100 query papers and the top-10 recommended papers
- The lower MAP@10 indicates that the system provides literature beyond the query paper's knowledge

Method	MAP@10
Proposed	**0.003**
Whole-text embedding	0.076
BM25	0.049

Similarity of objectives:

- Does the similarity measure between papers actually find the same objective papers?
- We used the 11 task description papers of SemEval-2021 and evaluated whether the recommended papers include their system description papers

Method	MAP
Proposed	**0.141**
Whole-text embedding	0.115

Future Work

- Extending the dataset
- Subjective evaluation of the proposed system

同志社大学
Doshisha University