Oh-My-Papers: a Hybrid Context-aware Paper Recommendation System

Hongjie Fang
Department of Computer Science and Engineering
Shanghai Jiao Tong University
galaxies@sjtu.edu.cn

Zhanda Zhu
Department of Computer Science and Engineering
Shanghai Jiao Tong University
dazz993@sjtu.edu.cn

Haoran Zhao
Department of Computer Science and Engineering
Shanghai Jiao Tong University
zhao.hr@sjtu.edu.cn

Abstract

Currently, many scholarly search engines still use fuzzy-matching-based algorithms to perform academic paper recommendations, while ignoring the power of "jargon" in paper searching. We first notice that we can use citation information to learn "jargon" in order to improve the performance of context-aware academic paper recommendations. Based on this idea, we propose a purely context-aware citation recommendation model, Citation-BERT, and apply it to the task of context-aware paper recommendation, achieving excellent results compared to other scholarly search engines. Our model also outperforms the baseline model, which reflects the effectiveness of our designation. In addition, we have built a structured citation dataset comprising over 360,000 records from approximately 65,000 papers, which is at least 10 times larger than the current dataset. The VGAE-Specter model is also proposed for a related paper recommendation task, which has a relatively great performance. Based on the previous models, we build a hybrid context-aware paper recommendation system *Oh-My-Papers*. Finally, a website has been built for demonstration, which presents our work in a more intuitive way. The full source codes are available at https://github.com/galaxies99/oh-my-papers.

1 Introduction

With the continuous development of science and technology, the number of papers began to grow massively. As a result, academic paper search engines are used by more and more researchers and given more attention in academic community. However, according to our observation, current scholarly search engines are basically based on fuzzy matching for search and recommendations, which limits their performance in real-world applications. Have you ever encountered such a situation, when you type in the keywords like "ResNet", "Yolo" and "ResNeXt" and search them in the leading scholarly search engines like Google Scholar [1], Semantic Scholar [2], *etc.*, you usually cannot find the results you want, as shown in Fig. 1.

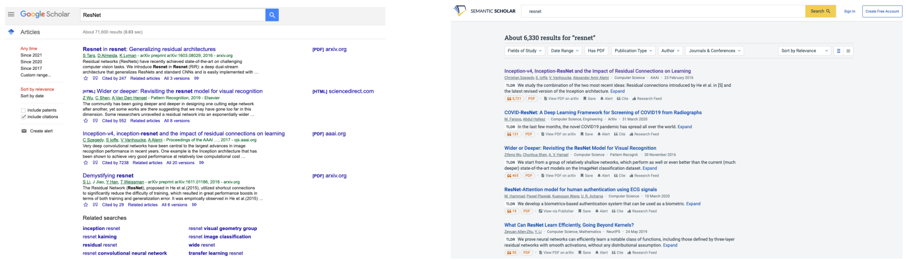
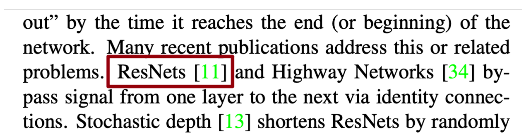


Figure 1: The recommendation results of Google Scholar (left) and Semantic Scholar (right) on keyword "ResNet"

The occurrence of the phenomenon is because that these keywords are so-called "jargon", that is, they are famous abbreviations developed in a certain academic community, such as "ResNet". Usually, these abbreviations do not appear in the title and abstract of the academic papers, which is the reason that many fuzzy-matching-based scholarly search engines can not understand them correctly.

Then, where can we extract the information about jargons? We observe that the citation context information contains this jargon information. When authors want to cite an article, they should summarize and describe the article and thus the citation context can naturally be used as a highly condensed feature of the paper. So one of the most important foundations is to learn jargon information based on how other authors refer that paper in their papers. As shown in Fig. 2, the citation of a paper naturally brings with it a feature extraction and the jargon of that paper, which is the quite high-quality information about the characteristics and jargons of the paper. Nevertheless, that kind of semantic and descriptive information in the citation context is usually under low utilization.



out" by the time it reaches the end (or beginning) of the network. Many recent publications address this or related problems. ResNets [11] and Highway Networks [34] by-pass signal from one layer to the next via identity connections. Stochastic depth [13] shortens ResNets by randomly

Figure 2: Citation context information example of ResNets from the DenseNet paper [3]. Here "[11]" is the citation identifier and 'ResNets' is the part of left citation context.

Therefore, we have developed our *Oh-My-Papers* system, which is able to learn the "jargons" from the citation records and perform satisfactory recommendations when concerning jargons in searching context. Based on the idea, we design more functions of our system, one of which is auto-citation.

Currently, papers are supposed to cite every paper it refers to. Since a paper may refer to more than 50 papers, it takes time and efforts to manually find the papers to be cited and add citations one by one. As Fig. 3 shows, if we can achieve accurate and efficient paper searching based on citation context information, we can also complete auto-citation which means automatically finding and filling in what should be cited in a placeholder. This is a useful feature, as it effectively reduces the burden on researchers when writing their papers.

ResNets [?] ❶ and Highway Networks [?] ❷ bypass signal from one layer to the next via identity connections.

❶ Deep residual learning for image recognition.

❷ Training very deep networks.

Figure 3: Auto-citation example from DenseNet paper [3]. Add placeholders where the citation should be inserted and fill them automatically.

In brief, the contributions of our *Oh-My-Papers* system are listed as follows.

1. We first point out that we can leverage citation information (especially citation context information) to learn jargons to improve the performance of context-aware paper recommendation. We also discover that we can directly apply the methods context-aware citation recommendation to solve the context-aware academic paper recommendation task.

2. We establish a big, well-documented, structured citation context dataset in the field of computer vision, which is at least 10 times larger than the current datasets that have the same structure to our knowledge.
3. We propose a purely context-aware citation recommendation model Citation-BERT, which outperforms the baseline model in several quantitative metrics. We apply our model to paper recommendation task, and it is able to achieve excellent results comparing to other scholar search engines.
4. We propose a related paper recommendation model VGAE-Specter, which has a relatively great performance in related paper recommendations.
5. We build a website to demonstrate our *Oh-My-Papers* system, which is user-friendly and supports real-time query, to present our system in a more intuitive way.

2 Related Work

2.1 Context-aware Citation Recommendation

Context-aware citation recommendation has been widely explored since it was first proposed by [4]. It is a special area of scientific paper recommendation and its methods can be broadly classified into the following categories[5]: content-based methods, collaboration-based methods, graph-based methods and hybrid methods.

The content-based approach is the first method that has been explored. He *et al.*[4] proposed a non-parametric context-based relevance model that measures the relevance between the given context and candidate documents for global and local recommendations, and they [6] further expanded the models into language model, contextual similarity, topic relevance and dependency feature model. Huang *et al.* [7] modeled the problem as a translation task with contextual references and proposed a contextual translation model which builds a dictionary of referenced papers and uses the expectation-maximization algorithm[8] to solve the problem. Then, they[9] applied neural network to their model and conducted more experiments. Tang *et al.*[10] defined a cross-language citation recommendation task and use embeddings extracted from the context to perform recommendation.

With the proposition of neural citation network[11], a hybrid approach that utilizes both contextual information and graph-based author information, hybrid methods are gradually being explored by an increasing number of researchers. Yang *et al.*[12] proposed an LSTM-based model for a personalized citation recommendation system, which uses the author's other papers to learn the author's personal embeddings for recommendations. Jeong *et al.*[13] proposed the BERT-GCN model, which uses the contextual embedding extracted by BERT[14] and the citation graph embedding extracted by VGAE[15] to make recommendations. However, the existing methods needs many information such as author information as prior knowledge, which is quite impractical for any person to use in real-world settings. Current researches have explored deeper into personal recommendation, without noticing that we can improve the recommendation purely based on the search context.

2.2 Text Representations Extraction

Text representation extraction is a method for extracting representation vectors of natural language objects such as words, sentences and documents using deep neural networks. Bengio *et al.*[16] proposed a neural probabilistic language model, which is the first to apply contemporary distributed representations to statistical language models. Mikolov *et al.* introduced CBOW [17] and Skip-gram[18] models that directly obtain word vectors. Further researches [19] extended the models into the document level. ELMo[20] found a different way to do word embedding. They extracted context-sensitive features from a left-to-right language model and a right-to-left language model. The contextual representation is a concatenation of the representations extracted by the two models. The model outperformed many traditional models in several benchmarks.

Fine-tuning based on pre-trained models has become a new trend in the field of text representation extraction. OpenAI GPT[21] achieved state-of-the-art results for many tasks on many benchmarks at that time, by fine-tuning pre-trained models, which were trained with large amounts of text data. BERT[14] utilized the bidirectional transformer to extract context embeddings and outperformed the models including GPT by a large margin on many tasks of GLUE benchmark[22]. SciBERT[23] and its following SPECTER[24] model applied the idea of BERT to perform document-level embedding extraction from academic papers and reaches satisfactory results.

2.3 Embedding Extraction on Graphs

Graph representation learning has emerged as a powerful strategy for analyzing graph-structured data. It aims to learn an encoding function that transforms nodes into low-dimensional dense embeddings, while preserving the attributes and structural features of the graph. For unsupervised settings, there are no labelled samples, so the loss function should only depend on the information provided by the graph itself, such as the input features or the topology of the graph[25].

A trend for unsupervised graph representation learning is to extend generative models such as variational auto-encoder (VAE) [26] and generative adversarial networks [27] to graph domains. Kipf and Welling [28] proposed Variational Graph Auto-Encoder (VGAE) which first uses GCNs to encode nodes in the graph. Pen *et al.* [29] introduces generative adversarial network (GANs) to regularize a GCN-based graph auto-encoder. Different from recovering the adjacency matrix, MGAE [30] utilizes marginalized denoising auto-encoder to get robust node representation and GALA [31] proposes Laplacian sharpening to build a symmetric graph auto-encoder.

In addition, another important approach is contrastive learning that captures statistical dependencies of interest and those that do not. Traditional unsupervised learning methods like DeepWalk [32] and node2vec [33] follow a contrastive framework. They first sample the neighboring nodes by random walks and minimize the distance between neighboring nodes in embedding space. Hjelm *et al.* proposed Deep InfoMax (DIM) [34] to maximize the mutual information between a high-level global representation and local parts of input. DGI [35] maximizes mutual information between node representations and graph representations. InfoGraph [36] maximizes the mutual information between graph-level representations and the substructure-level representations of different scales including nodes, edges and triangles. GRACE [37] generates two graph views by corruption and learns node representations by maximizing the agreement of node representations in these two views.

3 Dataset

3.1 Overview

We constructed a new dataset for a context-aware citation recommendation task that consists of approximately 65,000 papers and over 360,000 structured citation records. The citation dataset includes a variety of citation metadata and it is applicable to all citation recommendation methods as it provides all information to describe a citation in the level of paper. To our knowledge, it is the largest available structured citation dataset currently, which is at least 10 times larger than previous datasets; it is also the first citation dataset for papers in the field of computer vision. Our dataset has been publicly available and can be downloaded for academic use.

3.2 Data Acquisition and Dataset Statistics

We use the recent papers in the field of computer vision to construct our dataset, including several top conferences such as CVPR (2013-2020), ICCV(2013-2019), ECCV (2018), ICML (2013-2020), AAAI (2013-2020) and NIPS (2013-2020). We download the papers and use Science-Parse[38] to parse raw PDF files into structured files in JSON format.

After parsing, we can extract information from the structured JSON file. We first find the location of the citation placeholder and extract the context to the left and right of the placeholder respectively. We then extract information about the current paper, such as title, author, abstract, venue and year. Finally, we complete a citation record by searching the database for the cited paper and obtaining information including title, author, abstract, etc.

The dataset comparison between our dataset and other well-built datasets like datasets constructed by Jeong *et al.*[13] is shown in Tab. 1. As we can see from the table, the dataset we have created is approximately 10 times larger than the original dataset, which makes the subsequent related researches more convenient.

Table 1: The statistics of datasets

Statistics	# of total papers	# of citation records	Paper published year
AAN[13]	7,073	12,125	1965-2015
FullPeerTextRead[13]	4,898	17,247	2007-2017
Ours	64,403	366,183	2013-2020

3.3 Citation Records

For every citation record, we provide attributes listed in Tab. 2 in a structured format in our dataset.

Table 2: The attributes of our dataset

Attribute	Explanation
`index`	the index of citation records
`left_context`	the left context before citation
`right_context`	the right context after citation
`src_id`	the index of the source paper
`src_title`	the title of the source paper
`src_author`	the authors of the source paper
`src_venue`	the venue of the source paper
`src_year`	the published year of the source paper
`src_abstract`	the abstract of the source paper
`ref_id`	the index of the referenced paper
`ref_title`	the title of the referenced paper
`ref_author`	the authors of the referenced paper
`ref_venue`	the venue of the referenced paper
`ref_year`	the published year of the referenced paper
`ref_abstract`	the abstract of the referenced paper

Here is a specific example of the records and the process of collecting records in our dataset. First, we choose a paper in our dataset, for example, "Learning to Compare: Relation Network for Few-Shot Learning"[39]. Then we find the location of the citation placeholder as shown in Fig. 4, and extract the context before and after the placeholder as follows.

- **Left Context**: Unless otherwise specified, we use InceptionV2 as the query image embedding DNN in the old and conventional setting and ResNet101
- **Right Context** for the GBU and generalised setting, taking the top pooling units as image embedding with dimension D = 1024 and 2048 respectively.

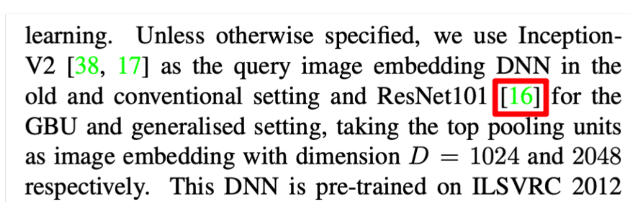

learning. Unless otherwise specified, we use Inception-V2 [38, 17] as the query image embedding DNN in the old and conventional setting and ResNet101 [16] for the GBU and generalised setting, taking the top pooling units as image embedding with dimension $D = 1024$ and 2048 respectively. This DNN is pre-trained on ILSVRC 2012

Figure 4: A paper in our dataset and the citation placeholder in it

After that, we collect the basic information of the paper such as title, authors, venue and year from the metadata. In this specific example, we extract the following information.

- **Source Paper Title**: Learning to Compare: Relation Network for Few-Shot Learning;
- **Source Paper Authors**: Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M. Hospedales;
- **Source Paper Venue**: CVPR;
- **Source Paper Year**: 2018;

- **Source Paper Abstract**: We present a conceptually simple, flexible, and general framework for few-shot learning, where a classifier must learn to recognise new classes given only few examples from each. Our method, called the Relation Network (RN), is trained end-to-end from scratch. During meta-learning, it learns to learn a deep distance metric to compare a small number of images within episodes ...

Then, we search our database for the referenced paper, and collect the basic information of the referenced paper as follows.

- **Referenced Paper Title**: Deep Residual Learning for Image Recognition
- **Referenced Paper Authors**: Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun;
- **Referenced Paper Venue**: CVPR;
- **Referenced Paper Year**: 2016;
- **Referenced Paper Abstract**: Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the layer ...

Finally, we assign a unique identifier to each paper, and include the identifiers in the citation records, which completes a citation record in our dataset.

4 Models

4.1 Overview

As mentioned in Section. 2.1, our question is to improve the academic paper recommendation purely based on the search context, without other additional information.

Our question only makes use of the input context of users without using any other information about the user, which reduces the burden on the user to provide more information. From another perspective, our problem setting also preserves the user's privacy. To the best of our knowledge, the exploration of our questions[4, 6, 7, 9, 10] currently stays before the advent of modern embedding extraction techniques. Recent approaches[11, 12, 13] all require a priori information, such other papers of the user, the venue of the context, etc., which is impractical to use in real-world applications. Therefore, we abandon those methods since they do not match the conditions of our proposed problem. Instead, we build a baseline model by ourselves and propose our new Citation-BERT model to improve the performance of the baseline model, which is also the core model of our *Oh-My-Papers* system.

Since our *Oh-My-Papers* system is a hybrid context-aware paper recommendation system, we need to extract the relationship between papers from our citation dataset. Hence, we also propose a new relation paper recommendation model based on the citation graph and document-level academic paper embeddings.

4.2 Baseline Model

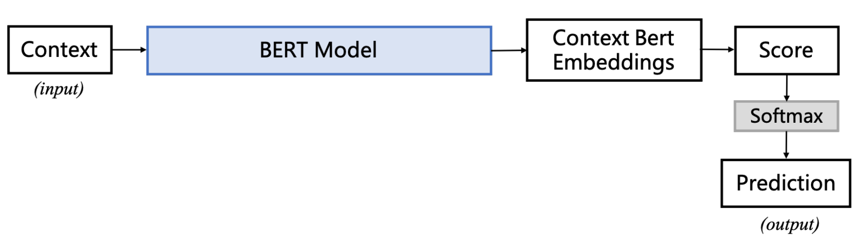


Figure 5: The architecture of the baseline model

Our baseline model is simply based on the BERT model[14], as shown in Fig. 5. The model uses BERT to extract contextual embeddings from the input context and then uses a feed-forward neural network to predict the score for each paper. Finally, the model generates output recommendations based on the results after a soft-max layer of scores.

4.3 Our Proposed Citation-BERT Model

Our baseline model only extracts information from the input context without exploring the similarity between the input context and the candidate papers. Therefore, we propose our Citation-BERT model.

As shown in Fig. 6, our Citation-BERT model consists of two branches, one of which is the same as the baseline model. The other branch uses a fixed pre-trained SPECTER model[24] to extract document-level paper embeddings for each of the candidate papers using titles and abstracts; and it also utilizes a new SPECTER model[24] to retrieve the representations of the context for fine-tuning, where we regard the full context as "title" of a virtual paper. Then, we use the cosine similarity to measure the similarity between context and the papers. Finally, the scores of the two branches are combined together after layer normalizations, and a softmax layer is used to produce the output recommendations based on the combined scores.

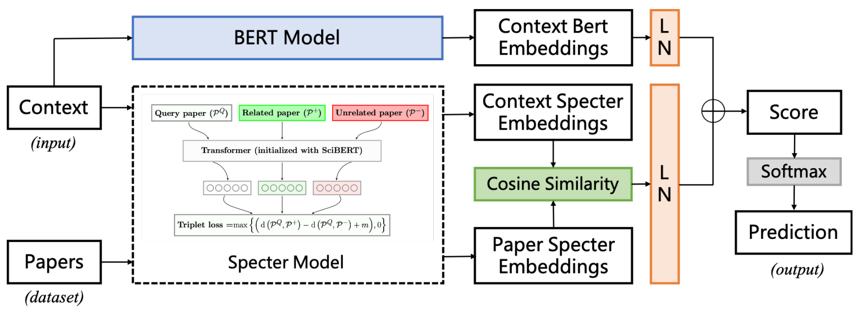


Figure 6: The architecture of our proposed Citation-BERT model

4.4 Our Proposed VGAE-Specter Model

To extract the relationship between papers from our citation dataset, we propose our VGAE-Specter model based on the citation graph and document-level paper embeddings extracted by SPECTER model[24] as a related paper recommendation model.

The main architecture of our VGAE-Specter model is based on unsupervised variational graph auto-encoder (VGAE) [15], with paper embeddings extracted by SPECTER model[24] as the node features in paper citation network. Then, we apply cosine similarity on the extracted embeddings from the candidate paper and the given paper to determine the relationship between the two papers. Our model will select the top-ranked papers with the highest similarity scores as the recommendation results for the given paper.

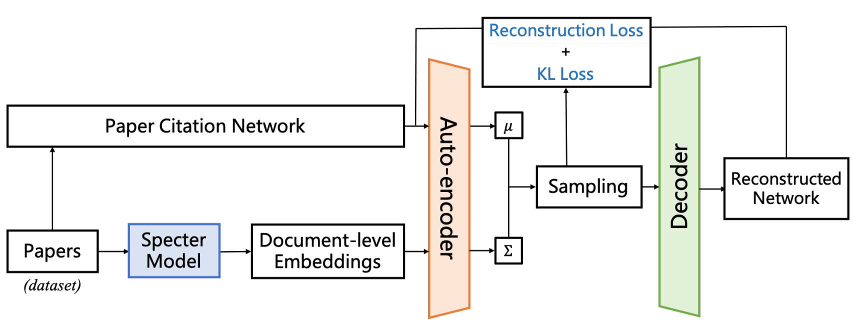


Figure 7: The architecture of our proposed VGAE-Specter model

5 Experiments

5.1 Context-aware Citation Recommendation Models

We divide our dataset as training set and testing set in a rough proportion of 3:1. Specifically, we use the citation records before 2020 for training, and use the citation records in 2020 for testing. For our baseline model and Citation-BERT model, we set the dimensions of the embedding to 768 and the maximum length of the context to 128. We use pre-trained BERT model[14] on uncased or cased pre-training corpus, and pre-trained SPECTER model[24] on scientific papers as initial. Then, we use AdamW optimizer[40] with learning rate 0.00005 and batch size 16 to train and fine-tune our models based on our dataset.

We first conduct experiments to verify the jargon recognition of our models by searching "Transformer"[41] and "DenseNet"[3] in our models and other scholar search engine. N From Tab. 3 and Tab. 4 we can see that, our Citation-BERT model can easily recognize those jargons and provide the correct papers, while other search engines may provide many irrelevant results. Notice that neither "Transformer"[41] nor "DenseNet"[3] has their names in the title of their original paper, which might be the reason of the bad performance of the current fuzzy-matching-based scholarly search engines, such as Google Scholar[1] and DBLP[42].

Table 3: Result of searching "Transformer"[41] in our system and other scholar search engines

Engine	Top-3 Results
Ours	**Attention is All You Need** End-to-end Dense Video Captioning with Masked Transformer Spatial Transformer Networks
Google Scholar	Spatial Transformer Networks Image Transformer Transformer in Transformer
Semantic Scholar	Exploring the Limits of Transfer Learning with a Unified Text-to-Text ... Spatial Transformer Networks Longformer: The Long-Document Transformer
DBLP	The Impact of Geomagnetically Produced Negative-Sequence ... Understanding the Influence of Power Transformer Faults on the ... Dynamic Thermal Model for Power Transformers

Table 4: Result of searching "DenseNet"[3] in our system and other scholar search engines

Engine	Top-3 Results
Ours	**Densely Connected Convolutional Networks** Condensenet: An Efficient DenseNet Using Learned Group Convolutions Deep Residual Learning for Image Recognition
Google Scholar	Densenet: Implementing Efficient Convnet Descriptor Pyramids Condensenet: An Efficient DenseNet Using Learned Group Convolutions Densenet for Dense Flow
Semantic Scholar	Condensenet: An Efficient DenseNet Using Learned Group Convolutions COVID-DenseNet: A Deep Learning Architecture to Detect COVID-19 ... Coverless steganography based on image retrieval of DenseNet features ...
DBLP	A Selective Mitigation Technique of Soft Errors for DNN Models Used ... Dermoscopy Image Classification Based on StyleGAN and DenseNet201 Survival prediction of patients suffering from glioblastoma based on ...

The goal of the context-aware citation recommendation is to recommend more relevant reference papers to the given citation context. Therefore, we use the following common metrics to evaluate our models quantitatively.

- **Recall@**N: It is defined by calculating the percentage of the original reference papers in the top recommended papers of the given scientific paper. We select $N = \{5, 10, 30, 50, 80\}$ to evaluate our models. We use the abbreviation "Rec@N" to represents the metric with parameter N in the following experiments.

- **Mean Reciprocal Rank (MRR)**: It measures how far from the top-ranked result to the first relevant reference papers, which is defined as

$$MRR = \frac{1}{|\mathcal{T}|} \sum_{(p_{src}, p_{ref}) \in \mathcal{T}} \frac{1}{\text{rank}(p_{ref}; p_{src})}$$

where $\mathcal{T}$ is the dataset we used, and $\text{rank}(p_{ref}; p_{ref})$ indicates the position of the ground truth referenced paper p_{ref} in the recommendation list of paper p_{src}

The evaluation results on testing set is shown in Tab. 5, where the suffix "-cased" means the model is cased-sensitive, while the suffix "-uncased" means the model is cased-insensitive.

Table 5: The evaluation results of several metrics on testing set

Method	MRR	Rec@5	Rec@10	Rec@30	Rec@50	Rec@80
baseline-cased (baseline)	0.1719	0.3562	0.4127	0.4901	0.5247	0.5545
Citation-BERT-cased (ours)	**0.1733**	**0.3573**	**0.4186**	**0.5064**	**0.5448**	**0.5786**
baseline-uncased (baseline)	0.1737	0.3590	0.4164	0.4968	0.5315	0.5623
Citation-BERT-uncased (ours)	**0.1761**	**0.3633**	**0.4204**	**0.5014**	**0.5372**	**0.5682**

From the results we can see that, our model outperforms the baseline model, which shows the effectiveness of taking the similarity between context and papers into consideration in our model.

5.2 Related Paper Recommendation Models

For related paper recommendation model, we train our VGAE-Specter model on the whole dataset since it is an unsupervised representation extraction model. We set the dimensions of the paper embeddings to 768 in VGAE-Specter model, and use a pre-trained SPECTER model[24] to extract document-level paper embeddings as node features. AdamW optimizer[40] with learning rate 0.01 is used to train our model for 2000 epochs.

We conduct experiments to evaluate the related paper recommendation results by searching the related papers of "Deep residual learning on image recognition"[43].

Table 6: Result of searching the related papers of "Deep residual learning on image recognition"[43] in our VGAE-Specter model

Recommendation Engine	Results of Related Papers Recommendation
Our VGAE-Specter Model	Adversarial Examples for Semantic Segmentation and Object ... Batch Normalization: Accelerating Deep Network Training by ... Going Deeper with Convolutions Aggregated Residual Transformations for Deep Neural Networks Densely Connected Convolutional Networks

From Tab. 6 we can observe that, our models recommends related papers of "Density Adversary Generation"[44], "Batch Normalization"[45], "GoogLeNet"[46], 'ResNeXt"[47] and "DenseNet"[3], all of which have strong connections with the original ResNet paper.

The goal of the related paper recommendation is to recommend more relevant papers to the given paper. By regarding the related paper recommendation task as the link prediction task on citation graphs, we can use the following common metrics to evaluate our models quantitatively.

- **Area Under the Curve (AUC)**: It measures the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. For a binary predictor (here refer to our link predictor) f, an unbiased estimator of its AUC can be expressed as follows.

$$AUC(f) = \frac{\sum_{t_0 \in \mathcal{D}^0} \sum_{t_1 \in \mathcal{D}^1} [f(t_0) < f(t_1)]}{|\mathcal{D}^0| \cdot |\mathcal{D}^1|}$$

where $\mathcal{D}^0$ is the set of negative samples, and $\mathcal{D}^1$ is the set of positive samples, while $[\cdot]$ is the indicator function.

- **Average Precision (AP)**: Precision and recall are single-value metrics based on the whole list of documents returned by the system. For systems that return a ranked sequence of documents, it is desirable to also consider the order in which the returned documents are presented. By computing a precision and recall at every position in the ranked sequence of documents, one can plot a precision-recall curve, plotting precision $p(r)$ as a function of recall r. Then, average precision is defined as follows.

$$AP = \int_0^1 p(r)\mathrm{d}r$$

Our model can reaches **0.8433 AUC** score and **0.7748 AP** score, which shows we extracted accurate and effective embeddings of papers in our dataset. It also reflects the effectiveness of our model.

6 Demo

6.1 Overview

The motivation of *Oh-My-Papers* derives from a problem close to our life, therefore, our work should also try to solve this problem in real life. That motivation inspires us to combine our research to engineering. Therefore, we develop an *Oh-My-Papers* website for demonstration.

Our back-end has already implemented very strong and useful paper recommendation system, so we aim to build a front-end website, which enables users to get access to our powerful system conveniently.

Apart from the functions provided by the back-end, our front-end should be user-friendly and practical, and we should make its speed as fast as possible. In a word, we aim to build an industrial product, which can be widely applied during our study and improve the efficiency of searching.

6.2 Implementation

We firstly deploy the program to our local website. We design a user interface based on our functions, which is practical and friendly. We implement three functions: context inference, relation inference and citation inference, and they will be introduced in detail later. Users can change to different functions conveniently through the top bar, which is shown in Fig. 8.

Figure 8: Website with different functions

We utilize websockets to communicate between the front-end and the back-end, which is a computer communications protocol, providing full-duplex communication channels over a single TCP connection. The process of websockets is shown in Fig. 9. The back-end and the-front shake hands for only once. Applying different websockets, more than one client can access our system at the same time.

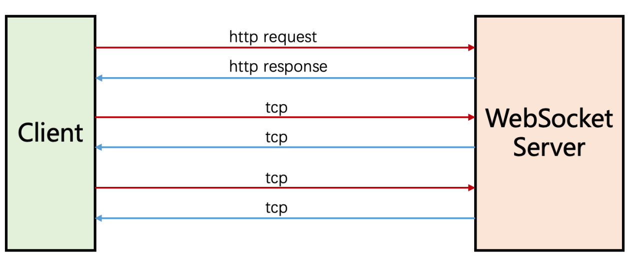


Figure 9: Process of websockets

We try a lot of techniques to implement real-time reaction. When the back-end of our system starts, we preload the trained models. Then the back end can react immediately it receives a request from the front end. This makes our search engine fast and efficient.

6.3 Function 1: Context Inference

Our first function is context inference, based on our context-aware citation recommendation model.

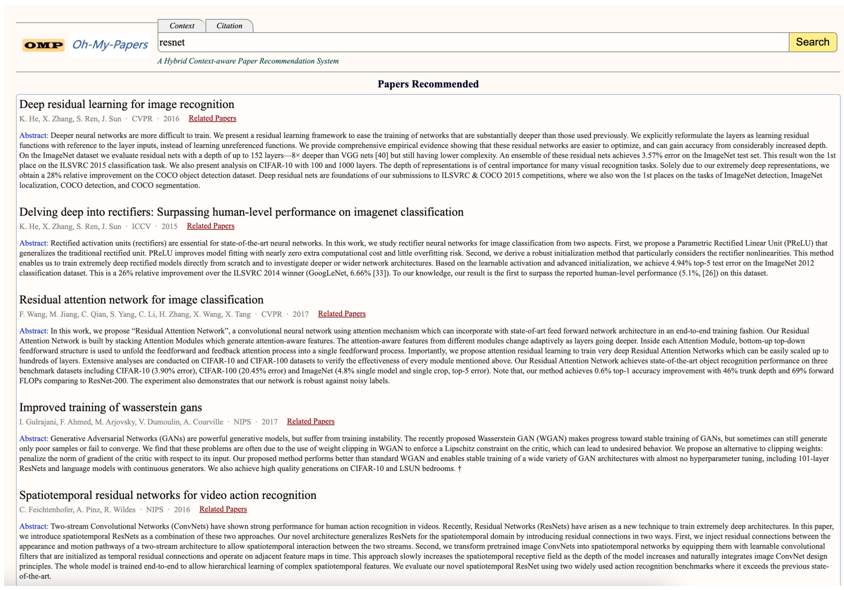


Figure 10: Result of Context Inference

As is shown in Fig. 10, the input is a context (for example, “resnet”). Our front-end reads in the context, organizes the information into a JSON file and sends it to our back-end. Then the front-end receives the organized result from back-end and displays the content in the website. In Fig. 10, we can see that the top recommended paper is “Deep residual learning for image recognition”, which satisfies our expectation.

6.4 Function 2: Relation Inference

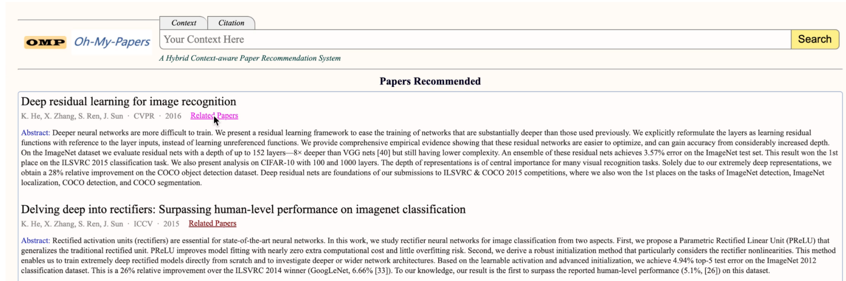
Our second function is relation inference, based on our related paper recommendation model.

As is shown in Fig. 11, a “Related Papers” link will be attached to each paper recommended. Clicking the link, our front-end will organize all the information about this paper (especially its “id”) and send it to back-end. Our back-end can then provide several papers related to the specified paper. In Fig. 11, we can see that the papers recommended are actually highly related to “ResNet”.

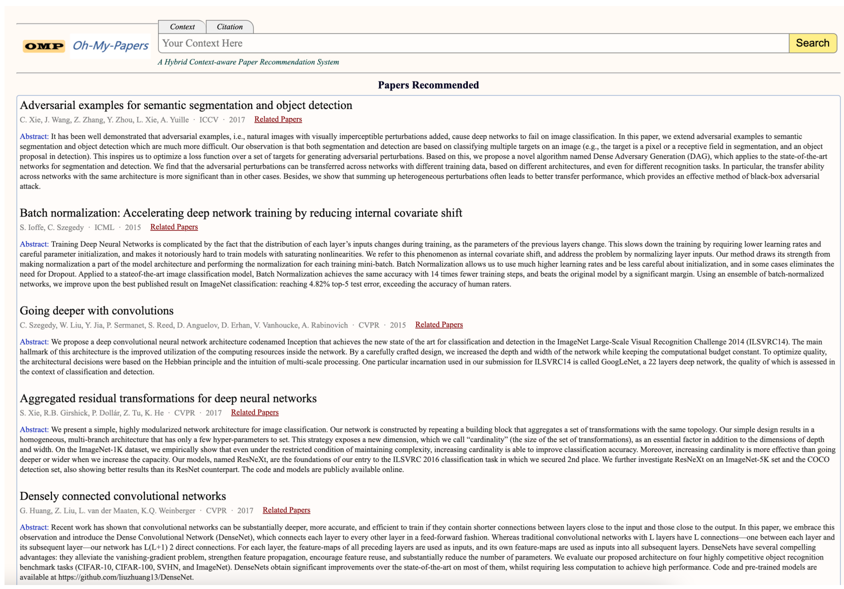
6.5 Function3: Citation Inference

Our last function is citation inference, based on our context-aware citation recommendation model.

As is shown in Fig. 12, the input is a context with some placeholders. Our front-end reads in the context, extracts the context aside the placeholder and sends the information to back-end. Our back-end will then give the top recommended paper for each placeholder separately. Thanks to our powerful model, the top recommended paper is often the targeted one, then we can implement auto citation. In Fig. 12, we can see that our program precisely cites “ResNet”, “YOLO” and “AlphaPose”. This function will help us a lot when we are writing papers and making citations.



(a) Before clicking "Related Papers"



(b) After clicking "Related Papers"

Figure 11: Result of Relation Inference

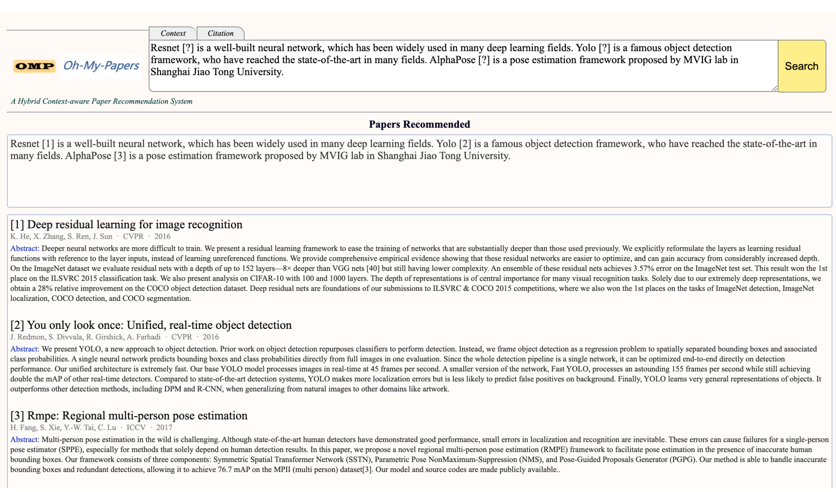


Figure 12: Result of Citation Inference

7 Conclusion

Our *Oh-My-Papers* context-aware academic paper recommendation system leverages citation information to achieve paper searching, auto-citation and paper recommendation functions. The core of our system is the pure context-aware citation recommendation model Citation-BERT, which utilizes citation context information that is easy to be ignored previously to learn the jargons of papers. In terms of the metrics, it outperforms previous methods. And in practical use, it achieves better results compared to other scholar search engines, which shows the power of "jargon" learning. Our VGAE-Specter model, which combines citation graph representations and meta data of papers to recommend related papers, also achieves relatively great performance on related paper recommendation task. We also build a demo website for our *Oh-My-Papers* system, which integrates these two models to provide real-time and user-friendly query. The source codes of our *Oh-My-Papers* system has been publicly available at https://github.com/galaxies99/oh-my-papers.

Acknowledgments

We would like to express our deepest gratitude to Prof. Luoyi Fu for the detailed introduction about the mobile internet field, from which we learnt lots of knowledge that is useful in further researches.

We would like to thank teaching assistants for their dedications and supports.

References

[1] "Google scholar." `https://scholar.google.com`.

[2] "Semantic scholar." `https://research.semanticscholar.org/publications`.

[3] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017.

[4] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles, "Context-aware citation recommendation," in *Proceedings of the 19th international conference on World wide web*, pp. 421–430, 2010.

[5] X. Bai, M. Wang, I. Lee, Z. Yang, X. Kong, and F. Xia, "Scientific paper recommendation: A survey," *IEEE Access*, vol. 7, pp. 9324–9339, 2019.

[6] Q. He, D. Kifer, J. Pei, P. Mitra, and C. L. Giles, "Citation recommendation without author supervision," in *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 755–764, 2011.

[7] W. Huang, S. Kataria, C. Caragea, P. Mitra, C. L. Giles, and L. Rokach, "Recommending citations: translating papers into references," in *Proceedings of the 21st ACM international conference on Information and knowledge management*, pp. 1910–1914, 2012.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[9] W. Huang, Z. Wu, C. Liang, P. Mitra, and C. Giles, "A neural probabilistic model for context based citation recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, 2015.

[10] X. Tang, X. Wan, and X. Zhang, "Cross-language context-aware citation recommendation in scientific articles," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pp. 817–826, 2014.

[11] T. Ebesu and Y. Fang, "Neural citation network for context-aware citation recommendation," in *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval*, pp. 1093–1096, 2017.

[12] L. Yang, Y. Zheng, X. Cai, H. Dai, D. Mu, L. Guo, and T. Dai, "A lstm based model for personalized context-aware citation recommendation," *IEEE access*, vol. 6, pp. 59618–59627, 2018.

[13] C. Jeong, S. Jang, E. Park, and S. Choi, "A context-aware citation recommendation model with bert and graph convolutional networks," *Scientometrics*, vol. 124, no. 3, pp. 1907–1922, 2020.

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[15] T. N. Kipf and M. Welling, "Variational graph auto-encoders," *arXiv preprint arXiv:1611.07308*, 2016.

[16] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, "A neural probabilistic language model," *The journal of machine learning research*, vol. 3, pp. 1137–1155, 2003.

[17] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[18] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *arXiv preprint arXiv:1310.4546*, 2013.

[19] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International conference on machine learning*, pp. 1188–1196, PMLR, 2014.

[20] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," *arXiv preprint arXiv:1802.05365*, 2018.

[21] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding with unsupervised learning," 2018.

[22] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "Glue: A multi-task benchmark and analysis platform for natural language understanding," *arXiv preprint arXiv:1804.07461*, 2018.

[23] I. Beltagy, K. Lo, and A. Cohan, "Scibert: A pretrained language model for scientific text," *arXiv preprint arXiv:1903.10676*, 2019.

[24] A. Cohan, S. Feldman, I. Beltagy, D. Downey, and D. S. Weld, "Specter: Document-level representation learning using citation-informed transformers," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 2270–2282, 2020.

[25] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, and M. Sun, "Graph neural networks: A review of methods and applications," *ArXiv*, vol. abs/1812.08434, 2018.

[26] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[27] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[28] T. Kipf and M. Welling, "Variational graph auto-encoders," *ArXiv*, vol. abs/1611.07308, 2016.

[29] S. Pan, R. Hu, G. Long, J. Jiang, L. Yao, and C. Zhang, "Adversarially regularized graph autoencoder for graph embedding," in *IJCAI*, 2018.

[30] C. Wang, S. Pan, G. Long, X. Zhu, and J. Jiang, "Mgae: Marginalized graph autoencoder for graph clustering," *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017.

[31] J. Park, M. Lee, H. J. Chang, K. Lee, and J. Y. Choi, "Symmetric graph convolutional autoencoder for unsupervised graph representation learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6519–6528, 2019.

[32] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: online learning of social representations," *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014.

[33] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[34] R. D. Hjelm, A. Fedorov, S. Lavoie-Marchildon, K. Grewal, A. Trischler, and Y. Bengio, "Learning deep representations by mutual information estimation and maximization," *ArXiv*, vol. abs/1808.06670, 2019.

[35] P. Velickovic, W. Fedus, W. L. Hamilton, P. Lio', Y. Bengio, and R. D. Hjelm, "Deep graph infomax," *ArXiv*, vol. abs/1809.10341, 2019.

[36] F.-Y. Sun, J. Hoffmann, and J. Tang, "Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization," *ArXiv*, vol. abs/1908.01000, 2020.

[37] Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang, "Deep graph contrastive representation learning," *ArXiv*, vol. abs/2006.04131, 2020.

[38] "Science-parse." `https://github.com/allenai/science-parse`, 2019.

[39] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1199–1208, 2018.

[40] I. Loshchilov and F. Hutter, "Fixing weight decay regularization in adam," 2018.

[41] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *ArXiv*, vol. abs/1706.03762, 2017.

[42] "Dblp." `https://dblp.org/`.

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[44] C. Xie, J. Wang, Z. Zhang, Y. Zhou, L. Xie, and A. Yuille, "Adversarial examples for semantic segmentation and object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1369–1378, 2017.

[45] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, pp. 448–456, PMLR, 2015.

[46] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[47] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1492–1500, 2017.