

Structure-Augmented Keyphrase Generation

Jihyuk Kim*

Myeongho Jeong*

Seungtaek Choi*

Seung-won Hwang**

Yonsei Univ*

Seoul National Univ**



YONSEI
UNIVERSITY



SEOUL
NATIONAL
UNIVERSITY

Target task: keyphrase generation

- Generating keyphrases for a given document.
 - Keyphrases provide main contents of the given document, using only a few words.

Q&A post Is overfitting "better" than underfitting?



I've understood the main concepts behind overfitting and underfitting, even though some reasons as to why they occur might not be as clear to me.

37



But what I am wondering is: *isn't overfitting "better" than underfitting?*

If we have a look at how well each scenario does on the training and test data, it seems that for the overfitting scenario, the model does at least well for the training data.

**Keyphrases
(or hashtags)**

machine-learning

neural-networks

overfitting

bias-variance-tradeoff

Leveraging structure

- The most standard structure: **title-body** structure
 - As in keyphrases, **titles** help to capture essential contents from the **main body**.

[Scientific articles]

Structure-Augmented Keyphrase Generation

This paper studies the keyphrase generation (**KG**) task for scenarios where structure plays an important role. For example, a scientific publication consists of a short title and a long body, where the title can be used for de-emphasizing unimportant details in the body. Similarly, for short social media posts (*e.g.*, tweets), scarce context can be augmented from titles, though often missing. Our contribution is generating/augmenting structure then encoding these information, using existing keyphrases of other documents, complementing missing/incomplete titles. Specifi-

[Q&A post]

Is overfitting “better” than underfitting?

Asked yesterday Active today Viewed 2k times



I've understood the main concepts behind overfitting and underfitting, even though some reasons as to why they occur might not be as clear to me.

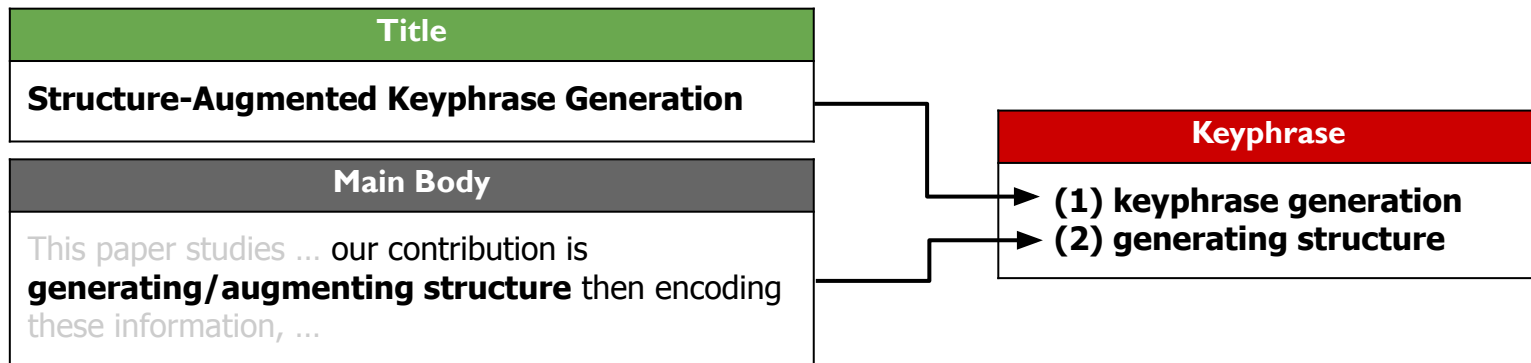
13



But what I am wondering is: *isn't overfitting "better" than underfitting?*

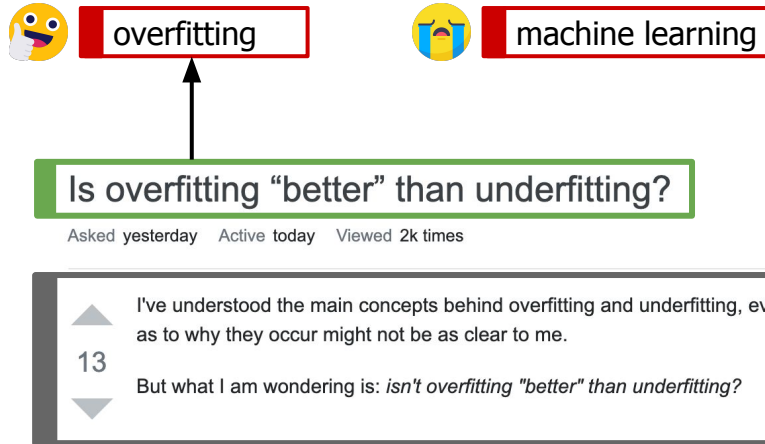
Previous work: title for KG

- Leveraging the title of a scientific paper ^[1].
 - 1) Generating **keyphrases** from the **title**.
 - 2) Generating **keyphrases** from related **body** contents to the **title**.



Challenge I.

- Titles are short!



overfitting machine learning

Is overfitting “better” than underfitting?

Asked yesterday Active today Viewed 2k times

13

I've understood the main concepts behind overfitting and underfitting, even though some reasons as to why they occur might not be as clear to me.

But what I am wondering is: *isn't overfitting "better" than underfitting?*

Challenge 2.

- Titles may not exist at all!

Is overfitting “better” than underfitting?



Asked yesterday Active today Viewed 2k times



13

I've understood the main concepts behind overfitting and underfitting, even though some reasons as to why they occur might not be as clear to me.

But what I am wondering is: *isn't overfitting "better" than underfitting?*



(Title does not exist in tweets)



EMNLP 2021 @emnlpmeeting ...

The wait's over! Final decisions on your papers should now be visible to authors via softconf and email notifications are slowly going out.
[#EMNLP2021](#)

Goal

Our goal is **generating/augmenting structures** for KG.

- generating title-like structures when titles are not available,
- and augmenting structures to complement incomplete titles.

Proposal: leveraging existing keyphrases

- Augmenting structures, using **existing keyphrases**.

An existing post with related keyphrases:

What does one imply by the term "overgeneralization" in machine learning?

Asked 4 months ago · Active 3 months ago · Viewed 75 times

▲ I know overfitting and underfitting in machine learning context, and what generalisation means as well. But, recently I was introduced to an uncommon terminology "overgeneralization" in context of fitting. What should this term relate to? Underfitting? Overfitting? Something else completely?

2

▼

machine-learning terminology overfitting generalization

Gold keyphrases:



overfitting



machine learning

Structure-augmented document:

Is overfitting "better" than underfitting?

Asked yesterday · Active today · Viewed 2k times

13

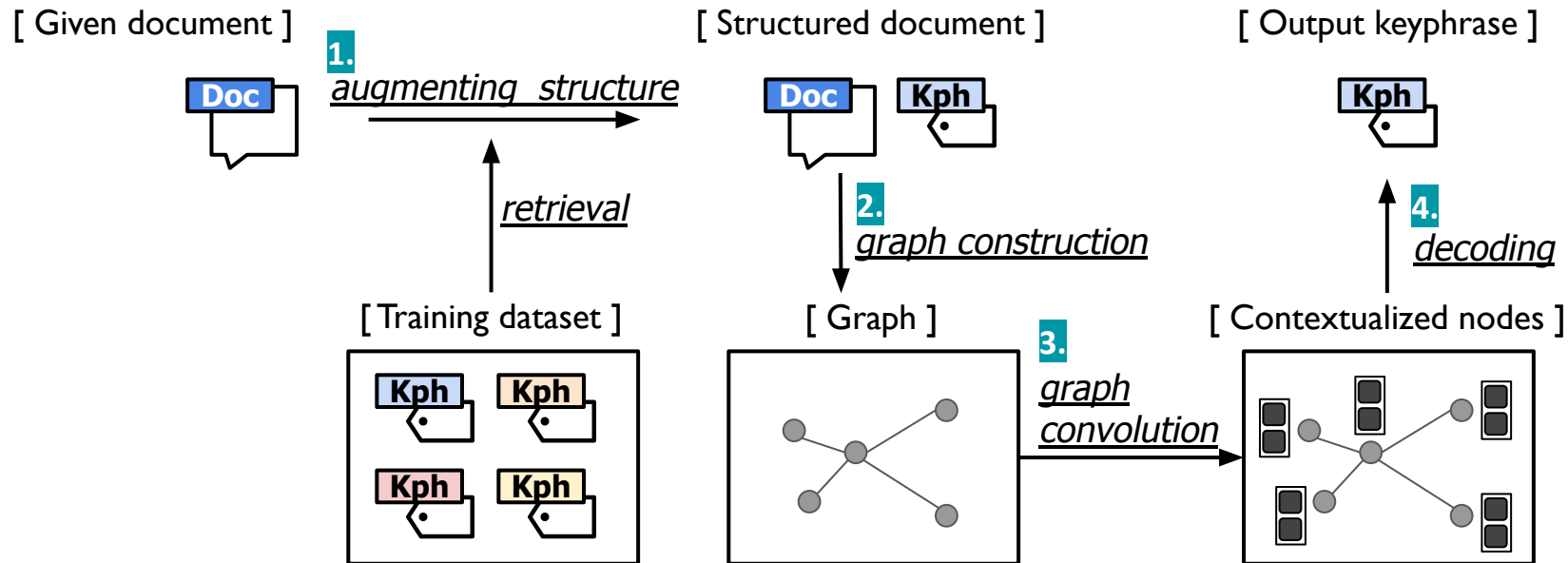
▲ I've understood the main concepts behind overfitting and underfitting, even though some reasons as to why they occur might not be as clear to me.

▼ But what I am wondering is: *isn't overfitting "better" than underfitting?*

machine-learning terminology overfitting generalization

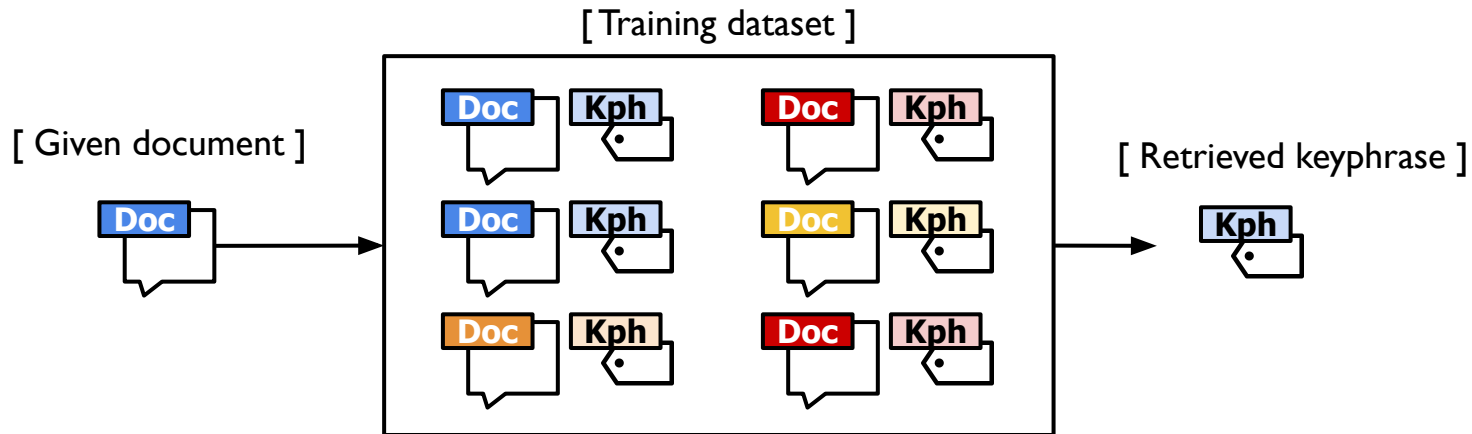
Overall process

- Structure-augmented keyphrase generation.



Retrieving relevant keyphrases

- Motivation: "Similar documents tend to have similar keyphrases."
 - We retrieve keyphrases of similar documents from training dataset.



Closed/open set scenarios

Closed set

(e.g., social media posts)

: Trending hashtags are frequently reused.

Open set

(e.g., scientific publications)

: New terms are introduced continuously.

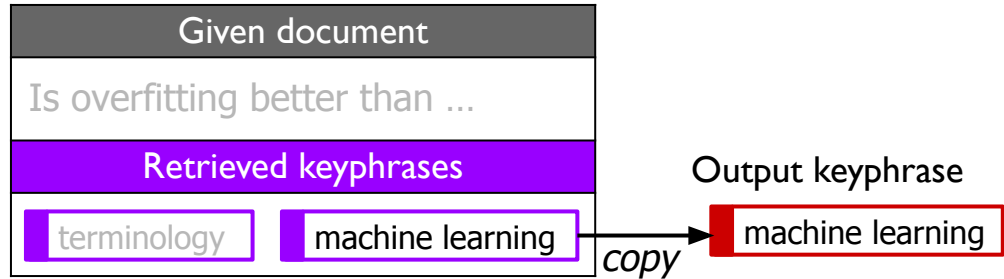
Dataset	% of unobserved kph
social media posts	4.1% (95.9% kphs are reused)
sci-publications	18.1% (18.1% kphs are newly introduced)

Closed/open set scenarios

Closed set

(e.g., social media posts)

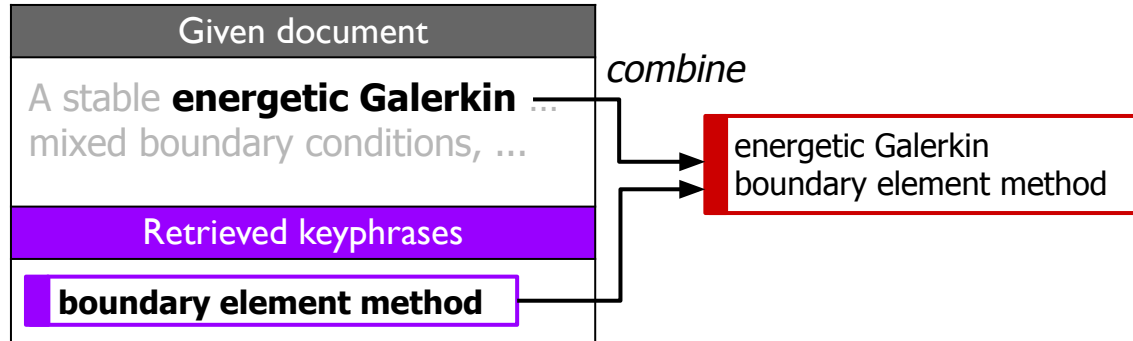
: Trending hashtags are frequently reused.



Open set

(e.g., scientific publications)

: New terms are introduced continuously.



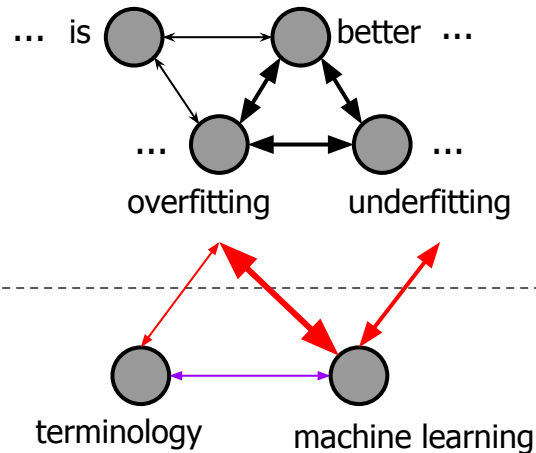
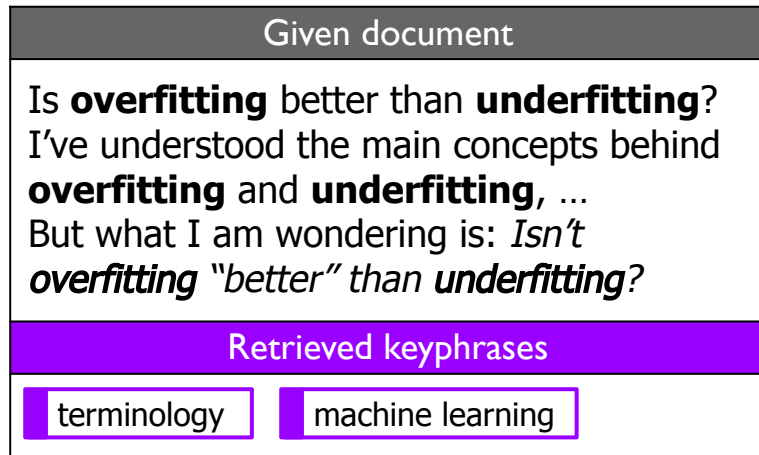
Graph construction

- **Two principles for graph construction**

1. Relevant nodes should be merged or connected with each other, and irrelevant nodes should be disconnected.
2. Relevant contexts between the given document and the retrieved keyphrases should be exchanged to each other.

Structure-aware graph for closed set

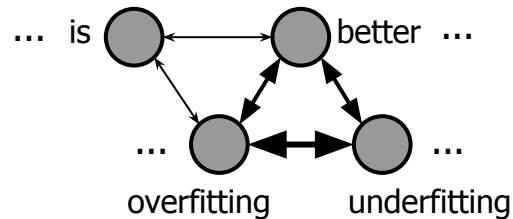
- **Connecting** the two graphs, using **inter-field edges**:
 - 1. graph for the given document.
 - 2. graph for the retrieved keyphrases.



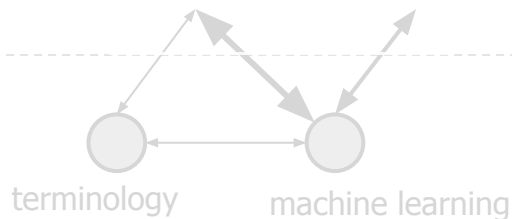
Structure-aware graph for closed set

- Graph for the given document:

Given document
Is overfitting better than underfitting ? I've understood the main concepts behind overfitting and underfitting , ... But what I am wondering is: <i>Isn't overfitting "better" than underfitting?</i>
Retrieved keyphrases
terminology machine learning



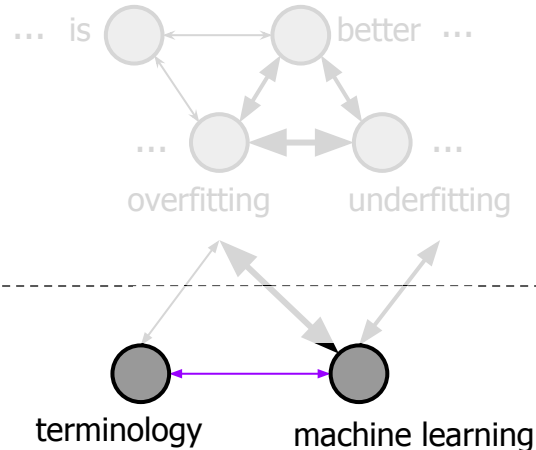
- **nodes**
: words in doc
- **edge weights**
: reciprocal of distance
(position-based)



Structure-aware graph for closed set

- Graph for the retrieved keyphrases:

Given document	
Is overfitting better than underfitting ? I've understood the main concepts behind overfitting and underfitting , ... But what I am wondering is: <i>Isn't overfitting "better" than underfitting?</i>	
Retrieved keyphrases	
terminology	machine learning

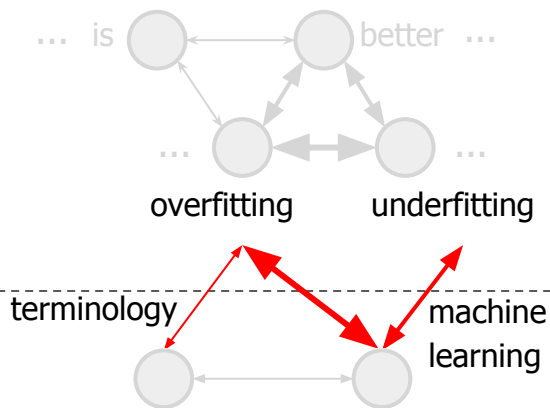


- **nodes**
: keyphrases
- **edge weights**
: co-occurrence
between keyphrases

Structure-aware graph for closed set

- Connecting edges between nodes in the two graph:

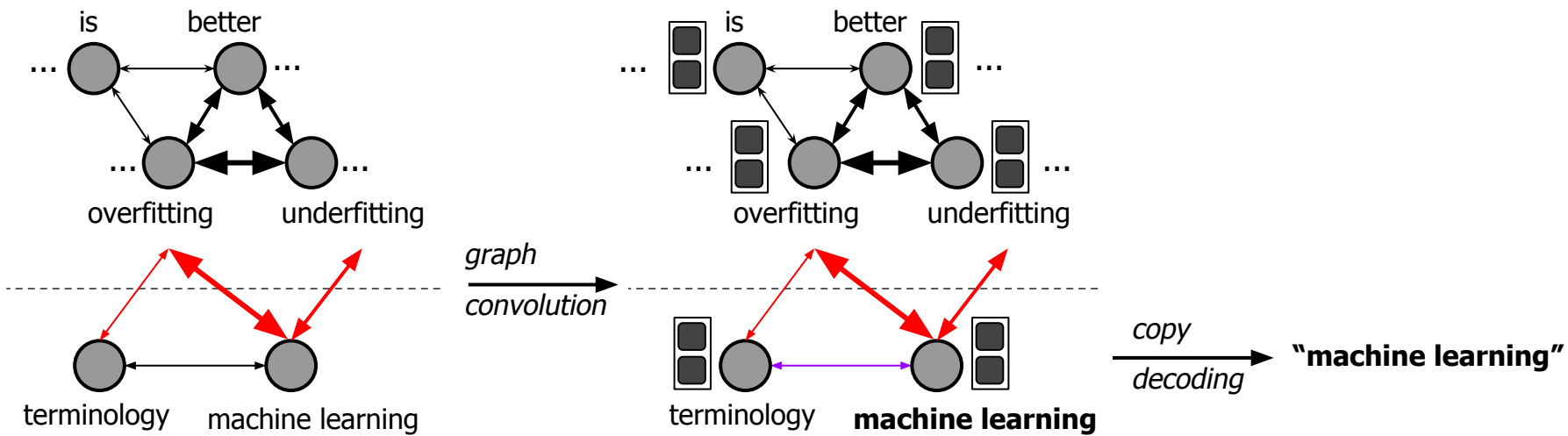
Given document	
Is overfitting better than underfitting ? I've understood the main concepts behind overfitting and underfitting , ... But what I am wondering is: <i>Isn't overfitting "better" than underfitting?</i>	
Retrieved keyphrases	
terminology	machine learning



- **edge weights**
: co-occurrence
between doc word
and keyphrase.

Decoding for closed set

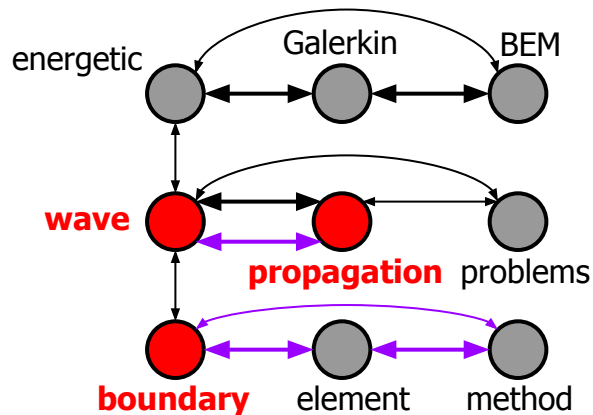
- Copying keyphrases based on keyphrase node representations.



Structure-aware graph for open set

- **Merging** the two graphs,
using the **same word nodes** with different edges (i.e., multi-graph):

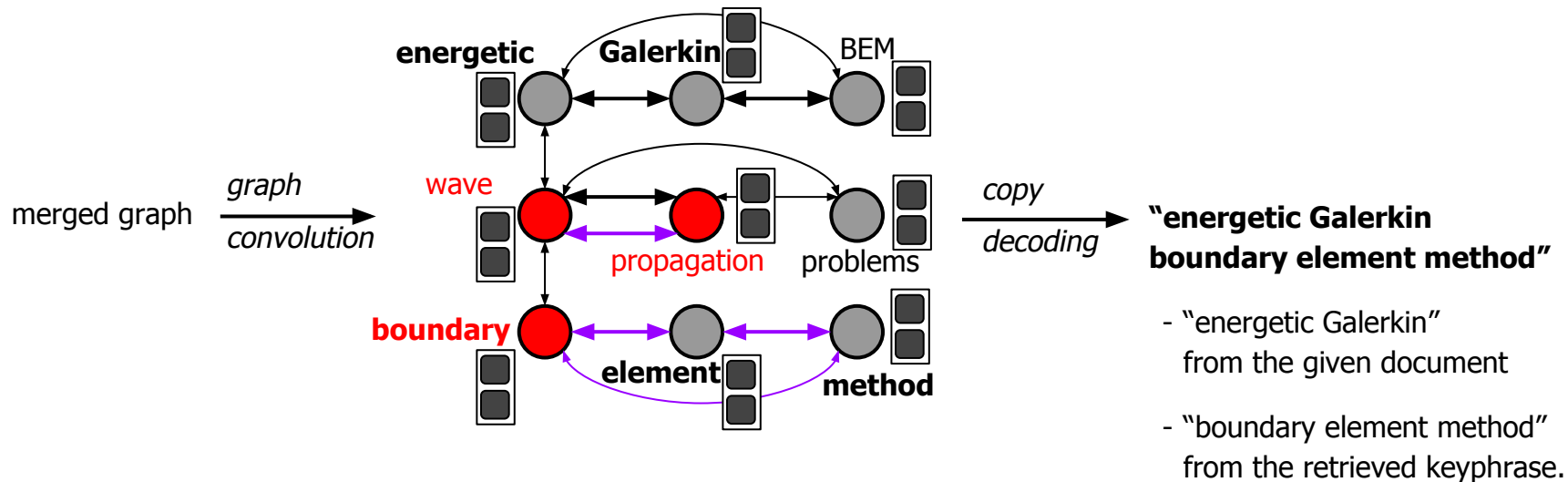
Given document
A stable 3D energetic Galerkin BEM approach for wave propagation problems. ... mixed boundary conditions, ... boundary integral ...
Retrieved keyphrases
wave propagation
boundary element method



- **nodes**
: words in doc
and words in kph
(the same word nodes
are merged)
- **edges from doc**
and **edges from kph**
: reciprocal of distance
(position-based)

Decoding for open set

- Combining keywords from the given document and the retrieved keyphrases.



Experiments

- Dataset.
 - StackExchange (social Q&A posts) for closed set scenario.
 - KP20k (scientific publication) for open set scenario.

title [StackExchange; social Q&A post]

Is overfitting “better” than underfitting?

body

▲ I've understood the main concepts behind overfitting and underfitting, even though some reasons as to why they occur might not be as clear to me.

13

▼ But what I am wondering is: *isn't overfitting "better" than underfitting?*

keyphrase (annotated by the author)

machine-learning

neural-networks

overfitting

bias-variance-tradeoff

title [KP20k; scientific publications]

A stable 3D energetic Galerkin BEM approach for wave propagation interior problems

body

A B S T R A C T

We consider 3D interior wave propagation problems with vanishing initial and mixed boundary conditions, reformulated as a system of two boundary integral equations with retarded potentials. These latter are then set in a weak form, based on a natural energy identity satisfied by the solution of the differential problem, and discretized by the energetic Galerkin boundary element method. Numerical results are presented and discussed in order to show the stability and accuracy of the proposed technique.

keyphrase (annotated by the authors)

Keywords:

Wave propagation

Boundary integral equation

Energetic Galerkin boundary element method

Experiments

- **Baselines:**
 - CopyRNN^[1]: using plain texts without structures.
 - TGNet^[2]: using title-body structures.

[1] Deep Keyphrase Generation. Meng et al., ACL 2017

[2] Title-Guided Encoding for Keyphrase Generation. Chen et al., AACL 2019

Experiments

- Evaluation metrics: **F1 score on top-k keyphrase predictions.**

Beam search
decoding



Prediction = top-k keyphrases, regarding likelihood.

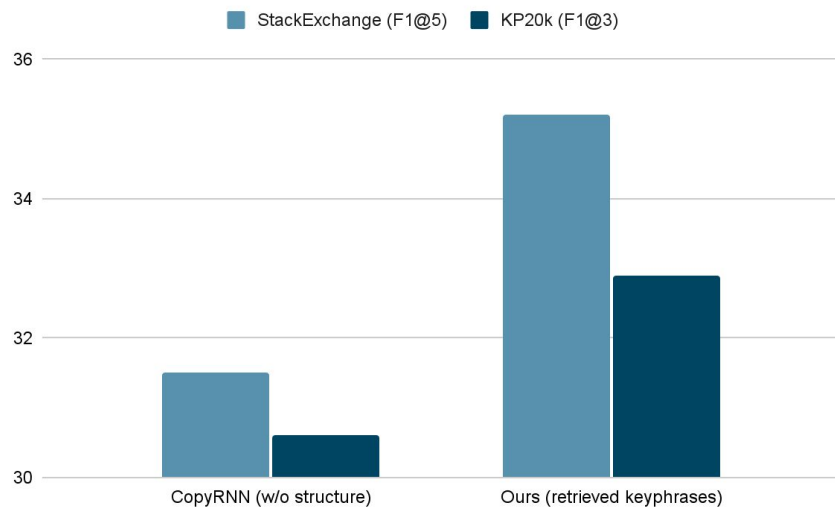
Precision@k = # of correct predictions / k

Recall@k = # of correct predictions / # of author-annotated keyphrases

F1@k = (2 x precision@k x recall@k) / (precision@k + recall@k)

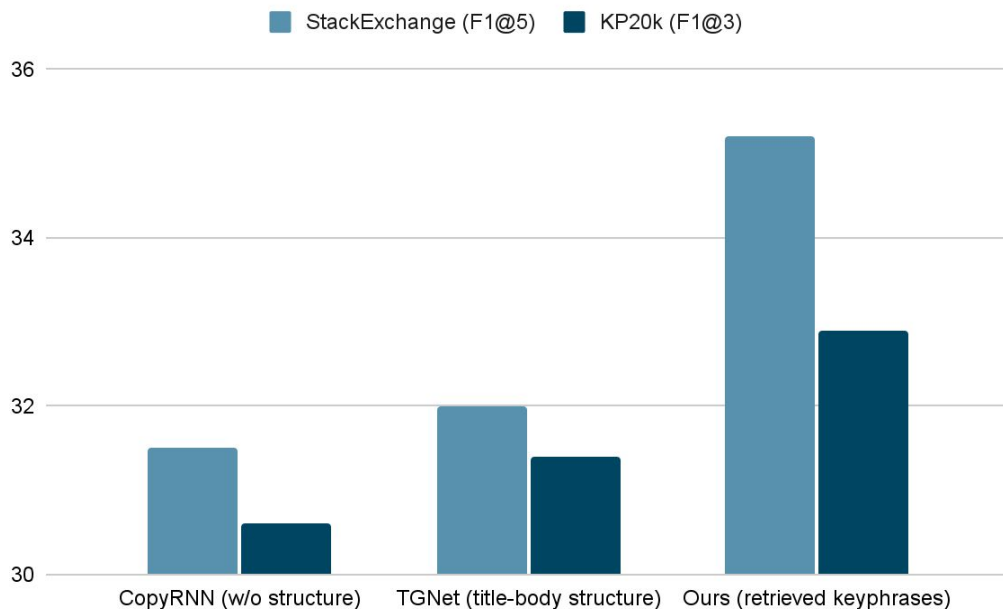
RQ I.

- Whether leveraging existing keyphrases as structures improves performance.
 - CopyRNN w/o structures < Ours w/ augmented structures.



RQ 2.

- Whether retrieved keyphrases are better than given titles.
 - TGNNet using given titles < Ours using retrieved keyphrases.

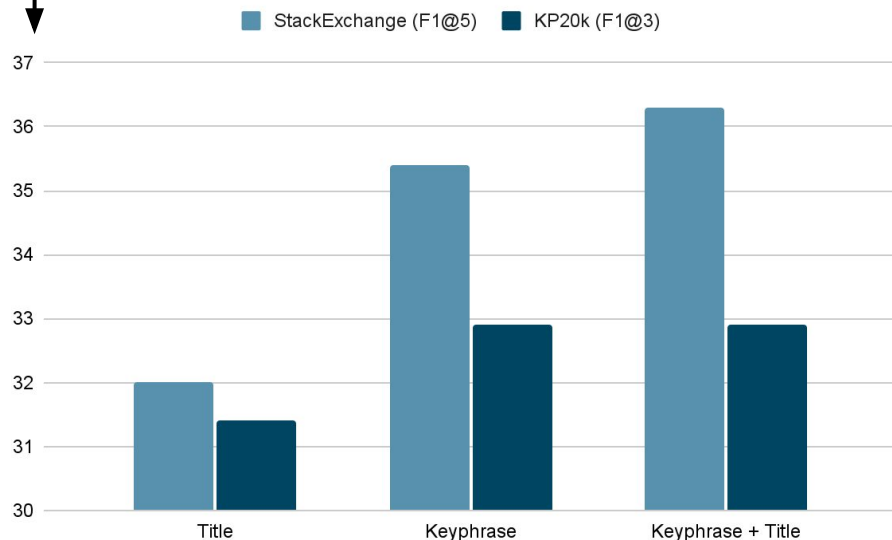


RQ 3.

- Whether titles and retrieved keyphrases are complementary to each other.

When retrieved keyphrases are less relevant (e.g., StackExchange), titles complement the retrieved keyphrases.

Dataset	F1 of retrieved keyphrases
StackExchange	11.5
KP20k	22.3



Conclusion

- We studied augmenting structure for keyphrase generation task.
- We leverage existing keyphrases to augment or generate structures.
- Our proposed methods work for both closed/open set scenarios.