



Multi-XScience

A Large-scale Dataset for Extreme Multi-document
Summarization of Scientific Articles

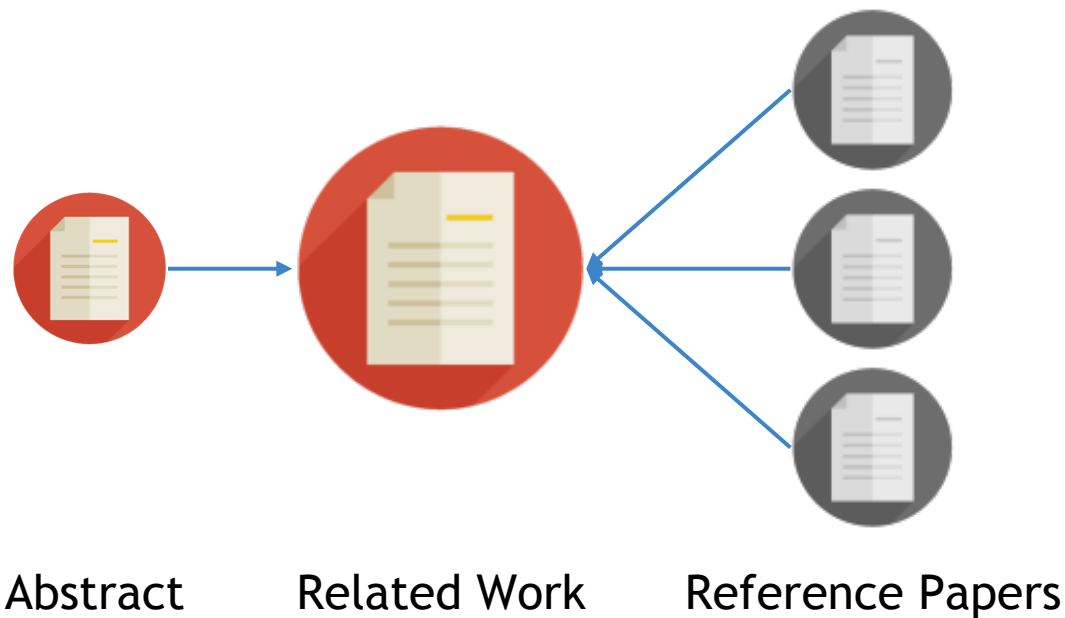
Yao Lu, Yue Dong and Laurent Charlin

lu.yao@ucl.ac.uk, yue.dong.2@mail.mcgill.ca

Dataset available at <https://github.com/yaolu/Multi-XScience>

TL;DR what is Multi-XScience?

- Writing related work



Source 1 (Abstract of query paper)
... we present an approach based on ... lexical databases and ... Our approach makes use of WordNet synonymy information to Incidentally, WordNet based approach performance is comparable with the training approach one.
Source 2 (cite1 abstract)
This paper presents a method for the resolution of lexical ambiguity of nouns ... The method relies on the use of the wide-coverage noun taxonomy of WordNet and the notion of conceptual distance among concepts ...
Source 3 (cite2 abstract)
Word groupings useful for language processing tasks are increasingly available ... This paper presents a method for automatic sense disambiguation of nouns appearing within sets of related nouns ... Disambiguation is performed with respect to WordNet senses ...
Source 4 (cite3 abstract)
In ... word sense disambiguation... integrates a diverse set of knowledge sources ... including part of speech of neighboring words, morphological form ...
Summary (Related work of query paper)
Lexical databases have been employed recently in word sense disambiguation . For example, ... [cite1] make use of a semantic distance that takes into account structural factors in WordNet ... Additionally, [cite2] combines the use of WordNet and a text collection for a definition of a distance for disambiguating noun groupings. ... [cite3] make use of several sources of information ... (neighborhood, part of speech, morfological form, etc.) ...

You should try **Multi-XScience** if you ...

- Need a large-scale Multi-Document Summarization dataset
- Not limited to News and Wikipedia domain

Dataset	# train/val/test	doc. len	summ. len	# refs
Multi-XScience	30,369/5,066/5,093	778.08	116.44	4.42
Multi-News	44,972/5,622/5,622	2,103.49	263.66	2.79
WikiSum	1, 5m/38k/38k	36,802.5	139.4	525

You should try Multi-XScience if you ...

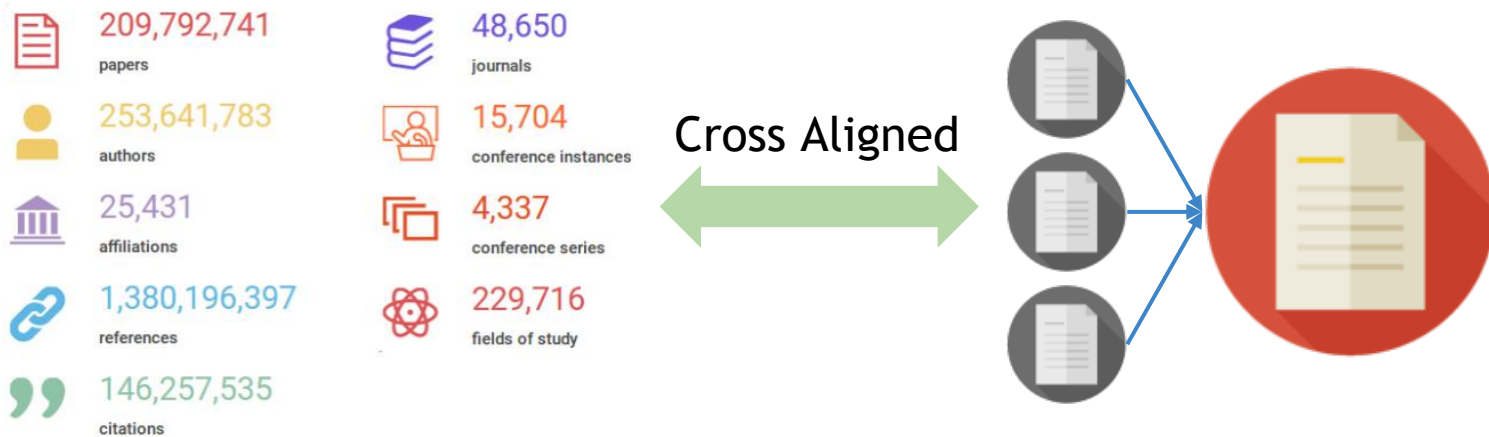
- Care about abstractiveness (extreme summarization)

Datasets	% of novel n -grams in target summary			
	unigrams	bigrams	trigrams	4-grams
CNN-DailyMail	17.00	53.91	71.98	80.29
NY Times	22.64	55.59	71.93	80.16
XSum	35.76	83.45	95.50	98.49
WikiSum	18.20	51.88	69.82	78.16
Multi-News	17.76	57.10	75.71	82.30
Multi-XScience	42.33	81.75	94.57	97.62

Table 3: The proportion of novel n -grams in the target reference summaries across different summarization datasets. The first and second block compare single-document and multi-document summarization datasets, respectively.

You should try Multi-XScience if you ...

- Interested in the intersection of Graph and Summarization



Microsoft Academic Graph

Multi-XScience

You should try Multi-XScience if you ...

- Interested in summarization with explicit supervision

Summary (Related work of query paper)

Lexical databases have been employed recently in **word sense disambiguation**. For example, ... [cite1] make use of a **semantic distance** that takes into account structural factors in WordNet ... Additionally, [cite2] ~~combines the use~~ of WordNet and a text collection for a definition of a distance for disambiguating noun groupings. ... [cite3] make use of several sources of information ... (neighborhood, part of speech, morfological form, etc.) ...

Source 2 (cite1 abstract)

This paper presents a method for the resolution of lexical ambiguity of nouns ... The method relies on the use of the wide-coverage **noun taxonomy of WordNet and the notion of conceptual distance among concepts** ...

Source 3 (cite2 abstract)

Word groupings useful for language processing tasks are increasingly available ... This paper presents a method for **automatic sense disambiguation of nouns appearing within sets of related nouns** ... Disambiguation is performed with respect to WordNet senses ...

Source 4 (cite3 abstract)

In ... **word sense disambiguation**... integrates a diverse set of knowledge sources ... including part of speech of neighboring words, morphological form ...



You should try **Multi**-XScience if you ...

- Interested in modeling cross-document relationship in summarization

Source 1 (Abstract of query paper)

... we present an approach based on ... **lexical databases** and ... Our approach makes use of WordNet synonymy information to Incidentally, WordNet based approach performance is comparable with the training approach one.

Summary (Related work of query paper)

Lexical databases have been employed recently in **word sense disambiguation**. For example, ... [cite1] make use of a semantic distance that takes into account structural factors in WordNet ... Additionally, [cite2] combines the use of WordNet and a text collection for a definition of a distance for disambiguating noun groupings. ... [cite3] make use of several sources of information ... (neighborhood, part of speech, morfological form, etc.) ...

Source 2 (cite1 abstract)

This paper presents a method for the resolution of lexical ambiguity of nouns ... The method relies on the use of the wide-coverage **noun taxonomy of WordNet and the notion of conceptual distance among concepts** ...

Source 3 (cite2 abstract)

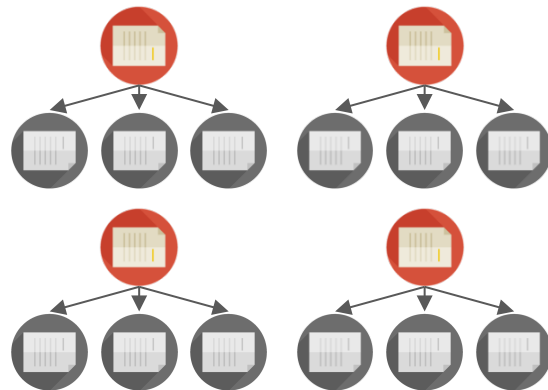
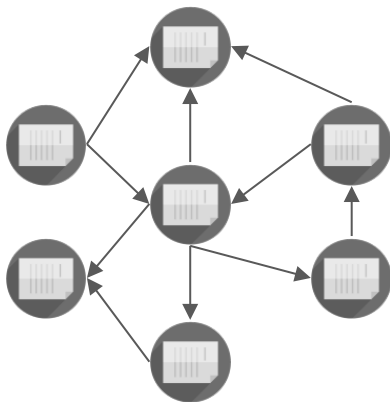
Word groupings useful for language processing tasks are increasingly available ... This paper presents a method for **automatic sense disambiguation of nouns appearing within sets of related nouns** ... Disambiguation is performed with respect to WordNet senses ...

Source 4 (cite3 abstract)

In ... **word sense disambiguation**... integrates a diverse set of knowledge sources ... including part of speech of neighboring words, morphological form ...

Dataset Construction

- Data sources: arXiv.org and Microsoft Academic Graph



arXiv (Source Papers) + MAG (Reference Papers) = Multi-XScience

Models

Extractive Models

- Lead Baseline
- LexRank
- TextRank
- Extractive Oracle

Multi-Doc Abstractive Models (fusion)

- HierSumm
- HIMAP

Multi-Doc Abstractive Models (concat)

- BertABS
- BART
- SciBertABS
- Pointer-Generator

Model Performance (ROUGE Score)

The dataset is challenging.

According to ROUGE-L:

Abstractive > Extractive

Models	ROUGE-1	ROUGE-2	ROUGE-L
Multi-doc Extractive			
LEAD	27.46	4.57	18.82
LEXRANK	30.19	5.53	26.19
TEXTRANK	31.51	5.83	26.58
EXT-ORACLE	38.45	9.93	27.11
Multi-doc Abstractive (Fusion)			
HIERSUMM(MULTI)	30.02	5.04	27.60
HIMAP(MULTI)	31.66	5.91	28.43
Multi-doc Abstractive (Concat)			
BERTABS	31.56	5.02	28.05
BART	32.83	6.36	26.61
SCIERTABS	32.12	5.59	29.01
POINTER-GENERATOR	34.11	6.76	30.63

Model Performance (Novel N-gram)

The dataset is highly
abstractive.

Models	% of novel n-grams in generated summary			
	unigrams	bigrams	trigrams	4-grams
PG (CNNDM)	0.07	2.24	6.03	9.72
PG (XSUM)	27.40	73.33	90.43	96.04
PG	18.82	57.54	80.22	89.32
HIERSUMM	27.52	77.16	95.03	98.51
HiMAP	23.13	63.58	86.50	94.15
BART	8.15	30.13	44.53	51.75
BERTABS	34.18	81.99	95.70	98.64
SCIBERTABS	46.57	89.05	97.92	99.31

Table 7: The proportion of novel n-grams in generated summary. PG (CNNDM) and PG (XSUM) denotes the pointer-generator model performance reported by papers ([See et al., 2017](#); [Narayan et al., 2018b](#)) trained on different datasets. All the remaining results are trained on Multi-XScience dataset.

Thank You!

Please come to our Q&A session for more discussions