Author's Sentiment Prediction

Mohaddeseh Bastan, Mahnaz Koupaee, Youngseo Son, Richard Sicoli and Niranjan Balasubramanian

Stony Brook University

COLING2020

Agenda

❖ Introduction

❖ PerSenT

  ➢ Crowdsourcing

  ➢ Quality Control

  ➢ Dataset Characteristics

❖ Evaluation

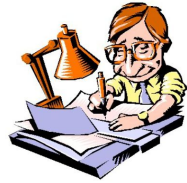❖ Conclusion

Agenda

❖ **Introduction**

❖ PerSenT

  ➢ Crowdsourcing

  ➢ Quality Control

  ➢ Dataset Characteristics

❖ Evaluation

❖ Conclusion

Task definition

Input: a document in news domain

Output: Author's sentiment toward the main entity

Input

Target: Lisa Pratt

Pratt has a "leave no trace" ethic when it comes to the solar system and **she** isn't happy about the idea of red sports cars floating around aimlessly in space.

Despite his enthusiasm Davis didn't get the six figure job. Instead a more seasoned astrobiologist **Lisa Pratt** stepped into the Planetary Protection Officer role earlier this month.

Pratt says **she**'s deeply concerned with how to "safely and ethically" look for life on Mars without accidentally killing any extraterrestrial life that might be lurking out there.

...

"I fully expect we will encounter life in our solar system " **she** said.

There might still be life on Mars and **Pratt** doesn't want us to kill it ...

Pratt's used to looking for life in really harsh places….

...

She's wandered into active gold mines in South Africa to discover what happens in super-hot waters flowing deep below the Earth's surface. **She**'s also ventured to freezing cold mines ..

...

She says any bits of Earthly trash that get deposited could mess up our chances of a human future on Mars.

….

Applications

Document Understanding

News analysis: Journalistic Bias

Sentiment Analysis

Challenges: Irrelevant information

- Non-target information
- Non-author's sentiment

Target: Lisa Pratt

Pratt has a "leave no trace" ethic when it comes to the solar system and **she** isn't happy about the idea of red sports cars floating around aimlessly in space.
Despite his enthusiasm Davis didn't get the six figure job. …

…

Challenges: Discourse Gap

- Long distance discourse connections

<table>
<tr><td align="center">Target: Lisa Pratt

Pratt says she's deeply concerned with how to "safely and ethically" look for life on Mars without accidentally killing any extraterrestrial life that might be lurking out there.</td></tr>
</table>

Paragraph Importance?

- Different levels of importance for different paragraphs

Target: Lisa Pratt

. .

Despite his enthusiasm Davis didn't get the six figure job. Instead a more seasoned astrobiologist **Lisa Pratt** stepped into the Planetary Protection Officer role earlier this month.
Pratt says **she**'s deeply concerned with how to "safely and ethically" look for life on Mars without accidentally killing any extraterrestrial life that might be lurking out there.

...

"I fully expect we will encounter life in our solar system " **she** said.

...

She's wandered into active gold mines in South Africa to discover what happens in super-hot waters flowing deep below the Earth's surface. **She**'s also ventured to freezing cold mines ..

...

Agenda

❖ Introduction

❖ **PerSenT**

➢ Crowdsourcing

➢ Quality Control

➢ Dataset Characteristics

❖ Evaluation

❖ Conclusion

PerSenT: Person SenTiment

News articles manually annotated for opinions, beliefs, emotions, sentiments, speculations on sentence level

TAC 2014 KBP English sentiment slot filling challenge dataset
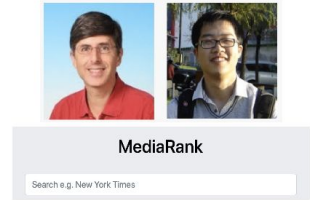
Ranking news websites based on the news contents

Agenda

❖ Introduction

❖ PerSenT

➤ **Crowdsourcing**

➤ Quality Control

➤ Dataset Characteristics

❖ Evaluation

❖ Conclusion

Crowdsourcing: Interface

❖ What is the paragraph-level sentiment?

❖ What is the document-level sentiment?

❖ Who's the main entity?

Crowdsourcing: Interface

paragraph-level sentiment

Taylor might be distancing herself from Kasich now that she's seeking endorsements from Ohio's conservatives for her gubernatorial campaign but that wasn't always the case.
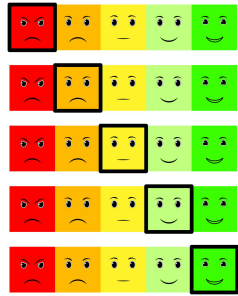
| Negative |
| Slightly Negative |
| Neutral |
| Slightly Positive |
| Positive |

Crowdsourcing: Interface

document-level sentiment

The whole article's view towards Mary Taylor is:

- Negative
- Slightly Negative
- Neutral
- Slightly Positive
- Positive

Crowdsourcing: Interface

Target entity

Is this article about
Mary Taylor ?

| Yes |
| No |

Agenda

❖ Introduction

❖ PerSenT

➢ Crowdsourcing

➢ **Quality Control**

➢ Dataset Characteristics

❖ Evaluation

❖ Conclusion

Inter annotator Agreement?

Different quality control methods for annotation process is described in the paper

Weighted Fleiss's Kappa:

Penalty for each dissimilar classification is based on the distance between two classes

Agreement:

0.79 for paragraph-level annotations

0.81 for document-level annotations

Also in 90% of the instances, at least two out of three annotators agree with each other.
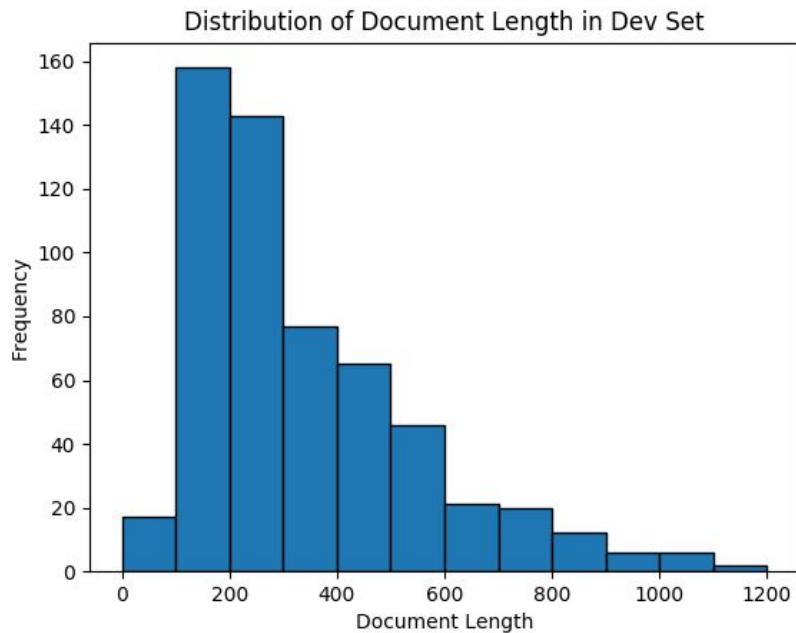
Agenda

❖   Introduction

❖   PerSenT

 ➢   Crowdsourcing

 ➢   Quality Control

 ➢   **Dataset Characteristics**
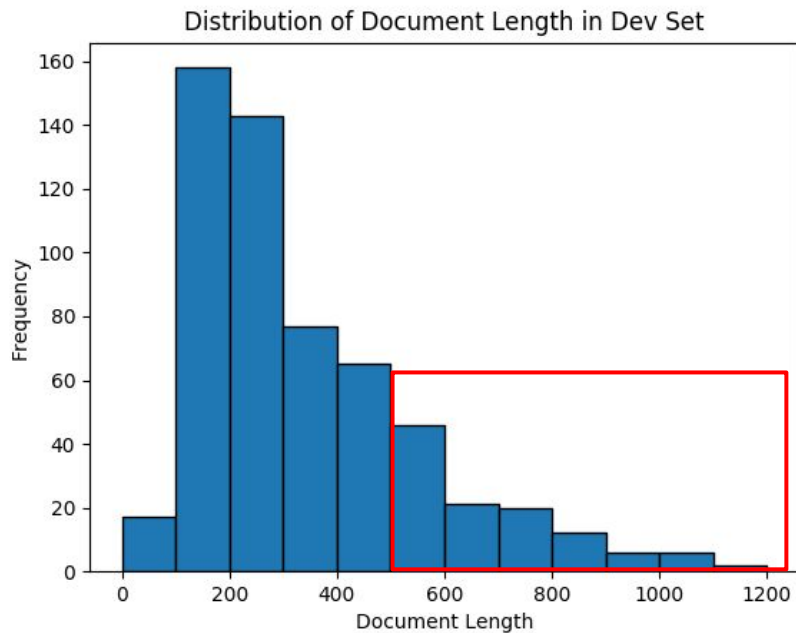
❖   Evaluation

❖   Conclusion

Dataset Characteristics

Few popular entities dominate news collection: 4 entities are main entity in nearly 800 articles

Dataset	Document	Paragraphs	Unique Entities
Train	3355	23953	2376
Dev	578	4169	417
Test Standard	579	4261	426
Test Frequent	827	5969	4
Total	5339	38352	3223

Extremely long documents

Distribution of Document Length in Dev Set

Extremely long documents

Distribution of Document Length in Dev Set
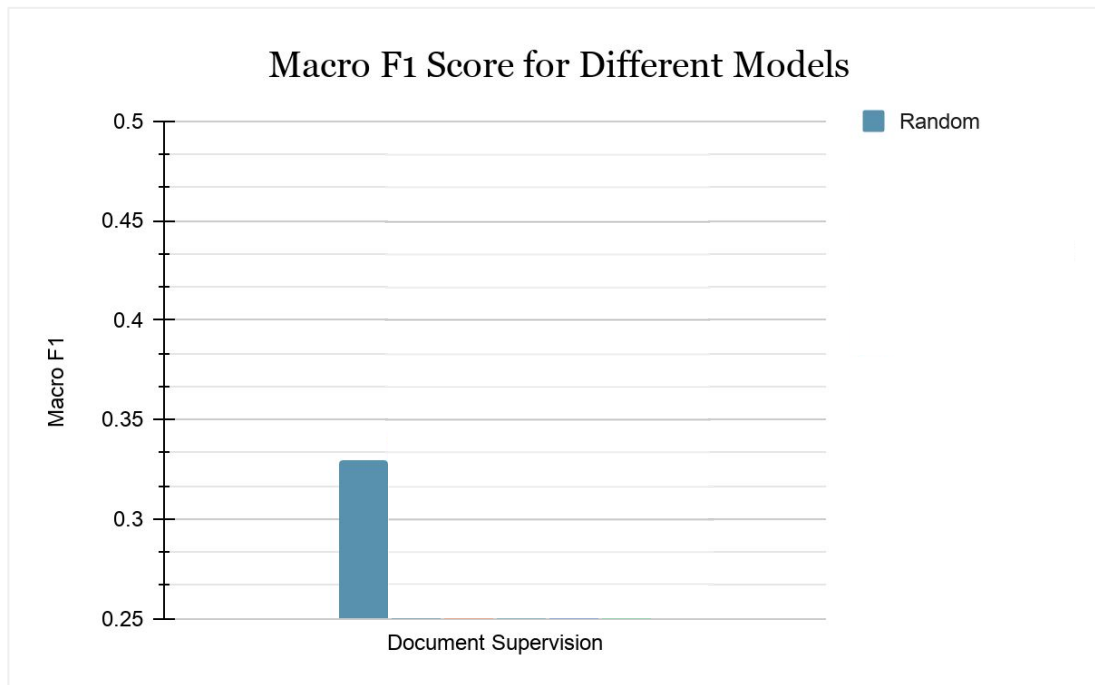
Label agreement

Agenda

- ❖ Introduction

- ❖ PerSenT
  - ➢ Crowdsourcing
  - ➢ Quality Control
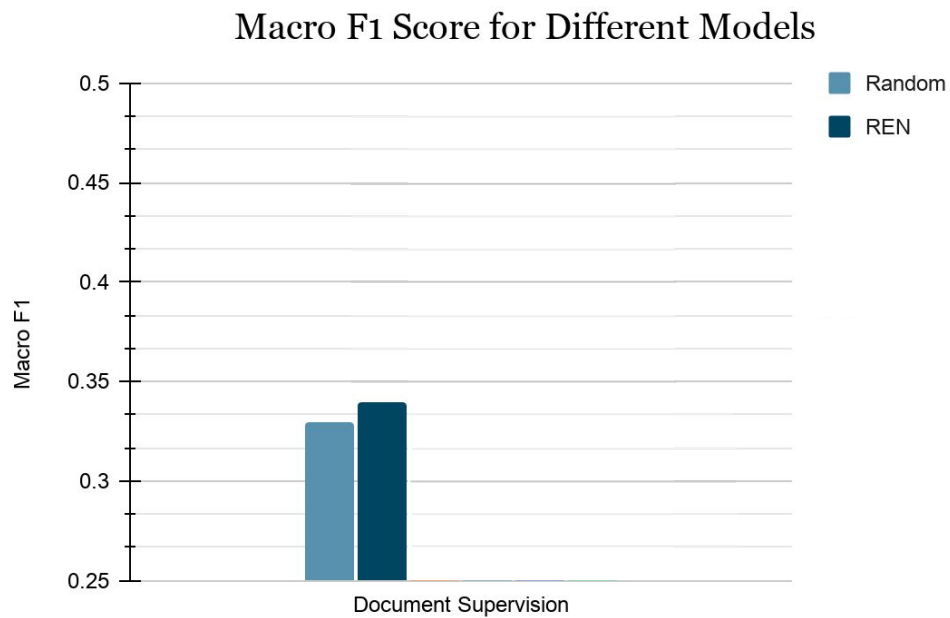 - ➢ Dataset Characteristics

- ❖ **Evaluation**

- ❖ Conclusion

Baselines

❖ Entity-based

  ➢ Modified Recurrent Entity Network

  ➢ Deep Averaging Network

❖ Discourse-based: aggregate information from different parts

  ➢ Hierarchical Discourse-based LSTM
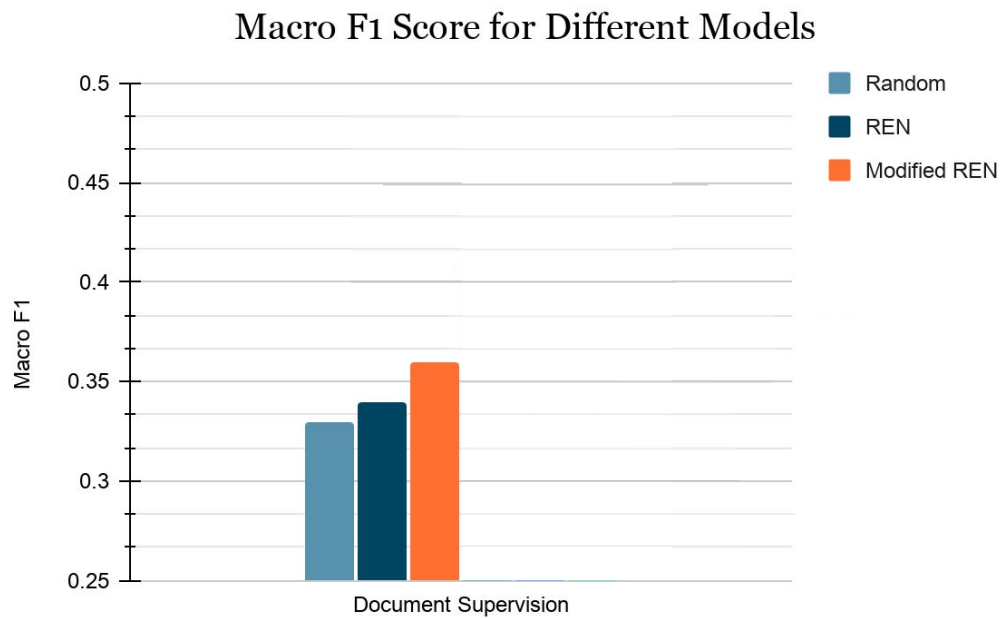
❖ Transformer model

  ➢ BERT w/ top 3 layers fine-tuning.

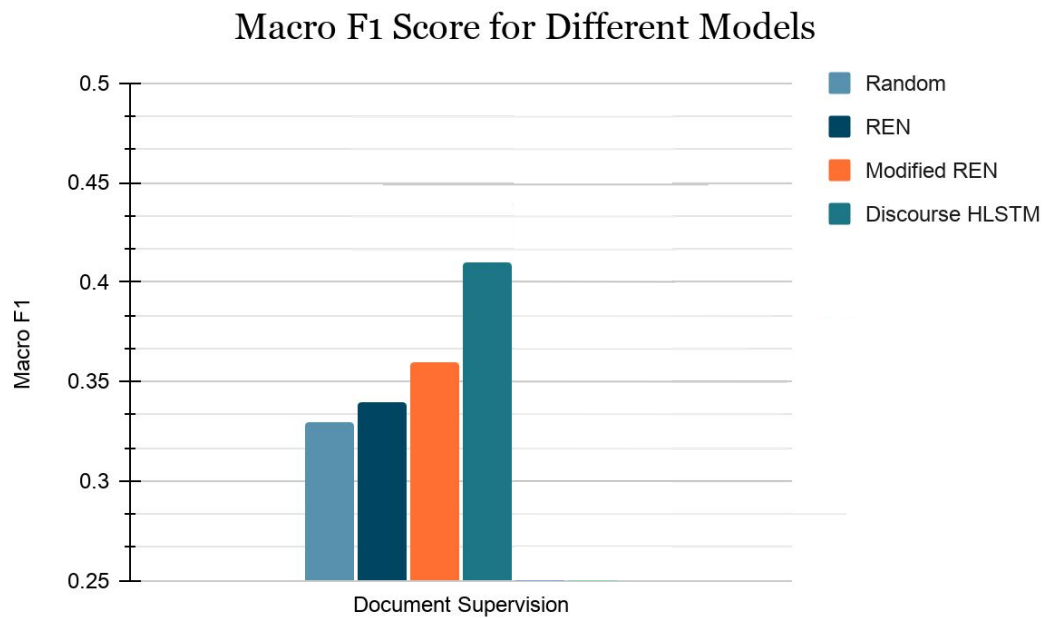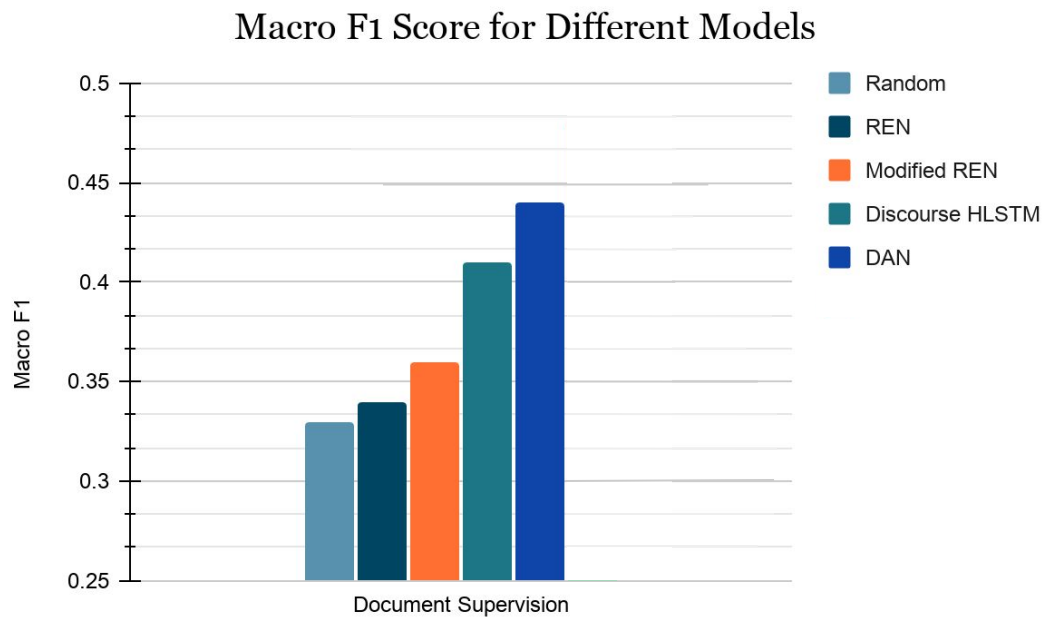# Results

Macro F1 Score for Different Models

Results

Macro F1 Score for Different Models

Results

Macro F1 Score for Different Models

Legend:
- Random
- REN
- Modified REN

Y-axis: Macro F1 (0.25 to 0.5)
X-axis: Document Supervision

Results

Macro F1 Score for Different Models

Results

Macro F1 Score for Different Models

Results

Macro F1 Score for Different Models
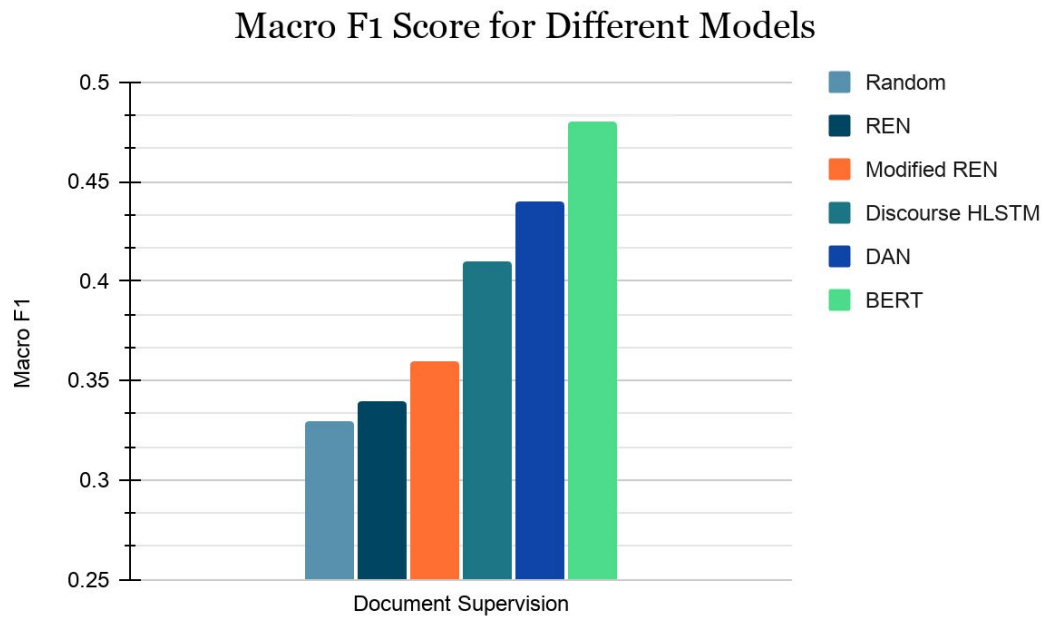
Focused Representations

Paragraph-level supervision
- Fine-tune BERT on paragraph-level task first.
- Fine-tune further on the document-level task.
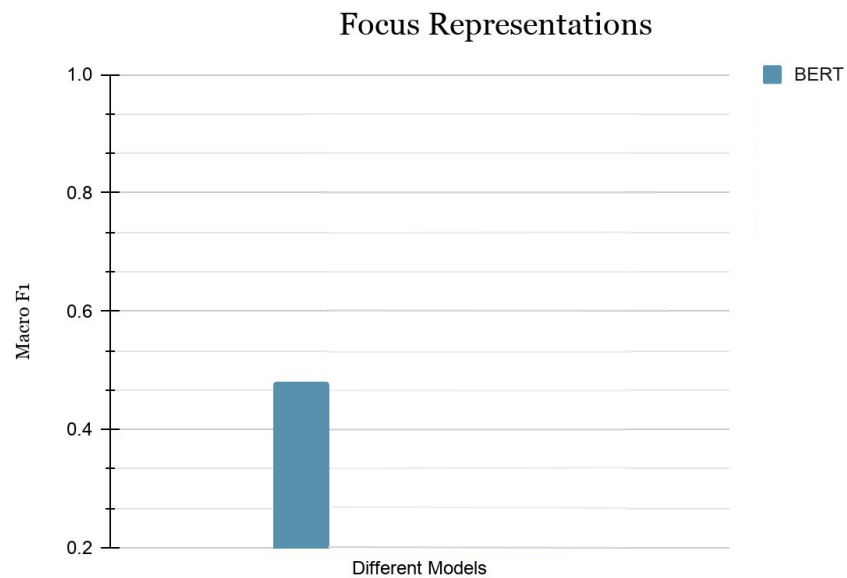
Focused Representations

Paragraph-level supervision
- Fine-tune BERT on paragraph-level task first.
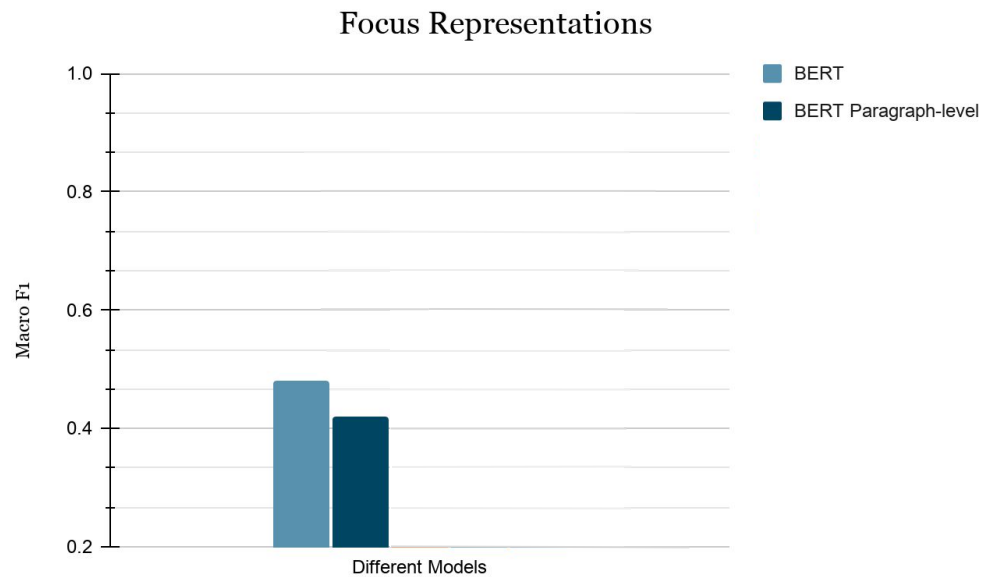- Fine-tune further on the document-level task.

Focusing towards the target entity
- Mask entities (target and non-target) and ask BERT to predict if MASK is target.
- Fine-tune pre-trained BERT on this ELM task.
- Sequential: Fine-tune further on the document-level task.
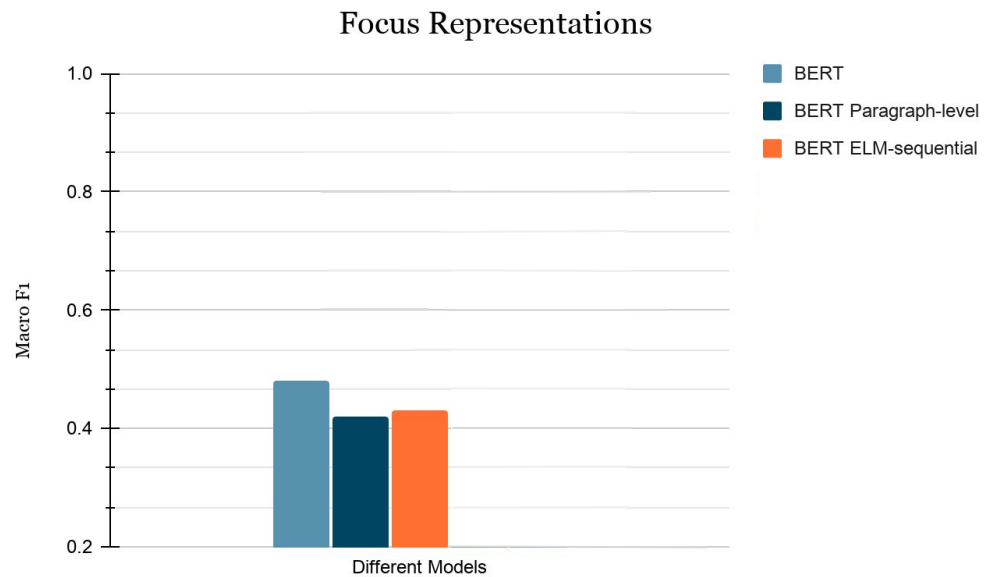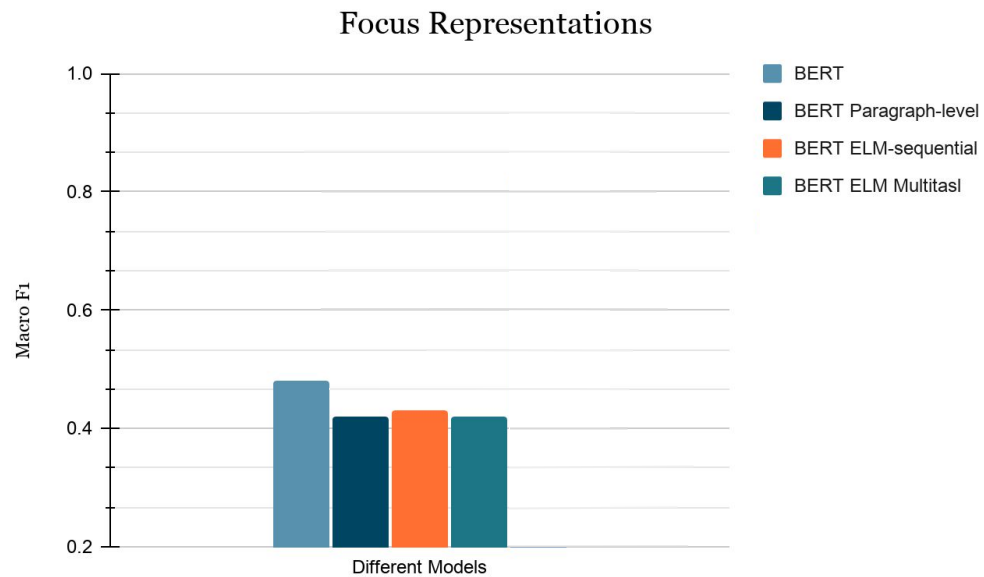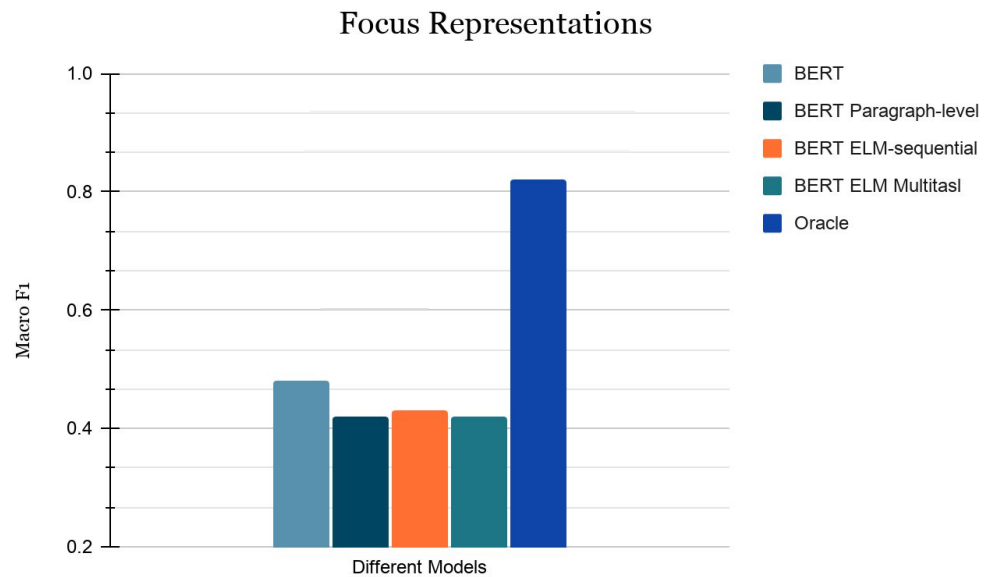- Multi-task: Use ELM task while fine-tuning on document-level task.

Results

Focus Representations
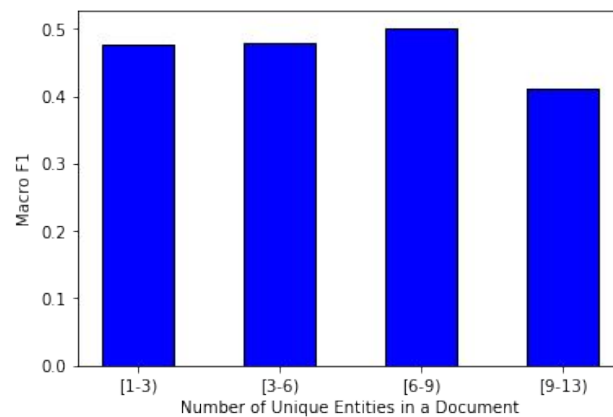
Macro F1 — Different Models — BERT

Results

Focus Representations

BERT
BERT Paragraph-level

Results

Focus Representations

Legend:
- BERT
- BERT Paragraph-level
- BERT ELM-sequential

Y-axis: Macro F1
X-axis: Different Models

Results

Focus Representations

- BERT
- BERT Paragraph-level
- BERT ELM-sequential
- BERT ELM Multitasl

Macro F1 / Different Models

Results

Focus Representations

- BERT
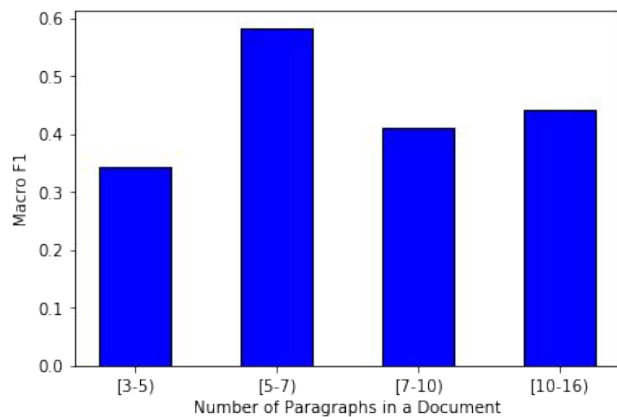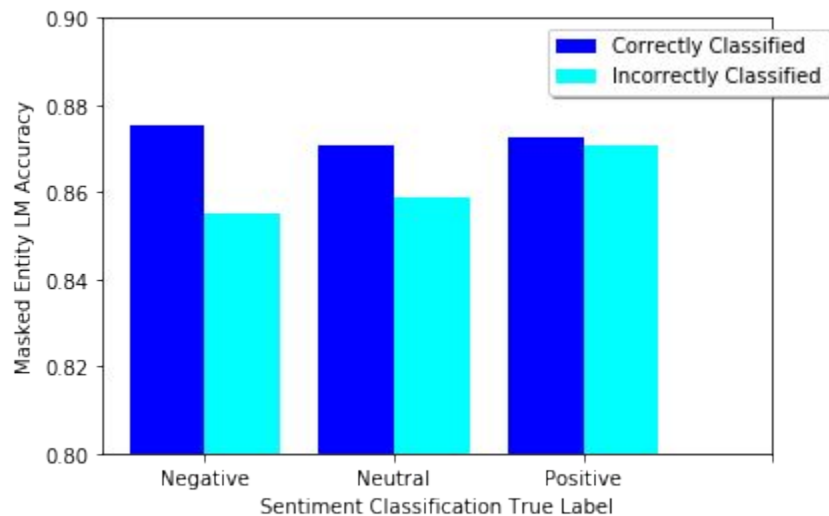- BERT Paragraph-level
- BERT ELM-sequential
- BERT ELM Multitasl
- Oracle

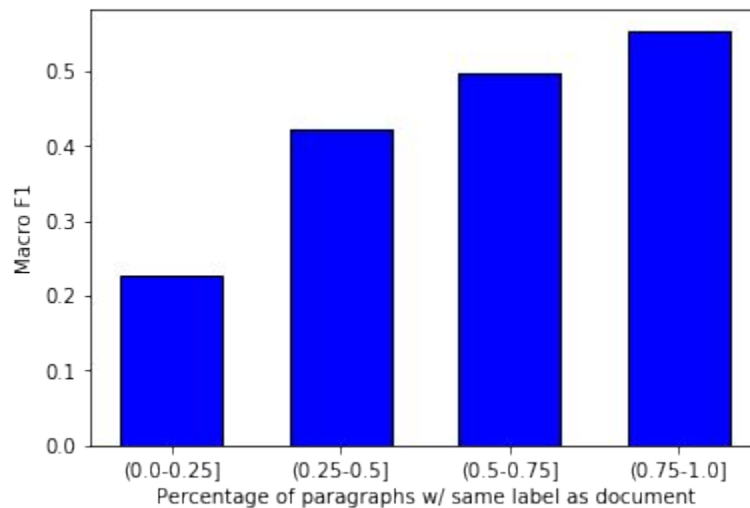# Effect of # of paragraphs and unique entities

Does Entity-LM performance correlate with task accuracy?

How useful is paragraph-level agreement?
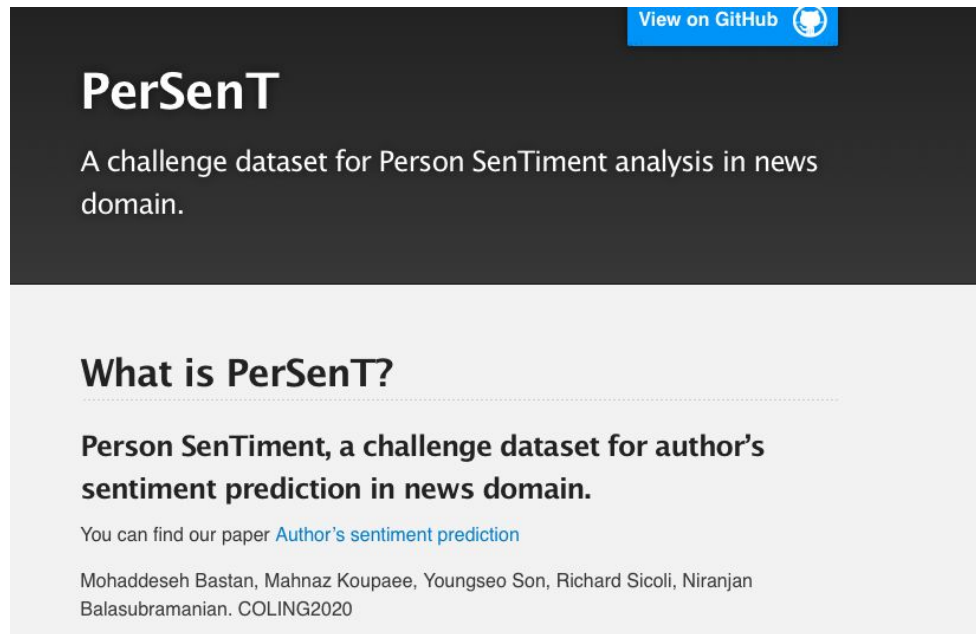
Distribution of Errors in Different Categories
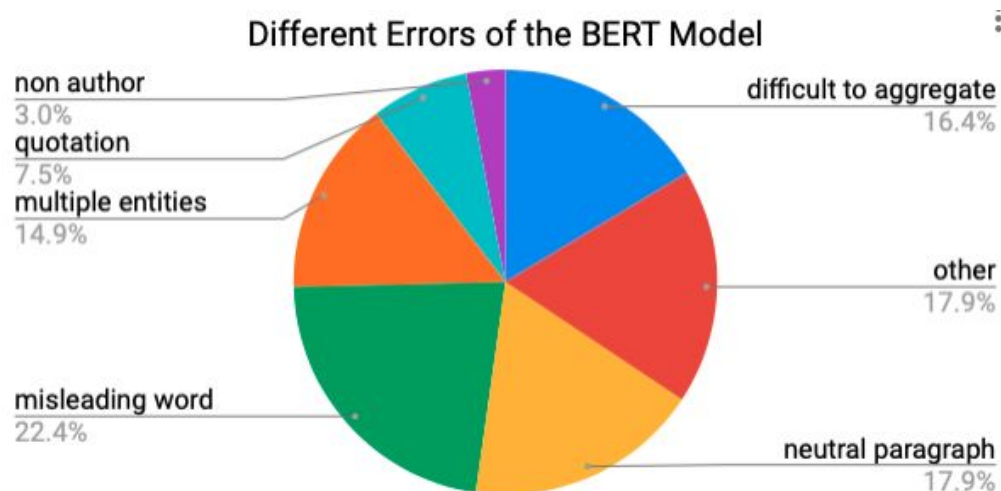
Different Errors of the BERT Model

non author 3.0%
quotation 7.5%
multiple entities 14.9%
misleading word 22.4%
difficult to aggregate 16.4%
other 17.9%
neutral paragraph 17.9%

Agenda

❖ Introduction

❖ PerSenT

  ➢ Crowdsourcing

  ➢ Quality Control

  ➢ Dataset Characteristics

❖ Evaluation

❖ **Conclusion**

Conclusion

❖ Introduced a new task: author sentiment inference in the news domain

❖ Released a challenge dataset

❖ Benchmarks suggest need for further research.

➢ models that operate over global representation

➢ models that aggregate local decisions

Code and dataset is available in our website...

https://stonybrooknlp.github.io/PerSenT/

PerSenT

A challenge dataset for Person SenTiment analysis in news domain.

What is PerSenT?

Person SenTiment, a challenge dataset for author's sentiment prediction in news domain.

You can find our paper Author's sentiment prediction

Mohaddeseh Bastan, Mahnaz Koupaee, Youngseo Son, Richard Sicoli, Niranjan Balasubramanian. COLING2020