TRoTR: A Framework for Evaluating the Recontextualization of Text

Francesco Periti
University of Milan,
Italy

Pierluigi Cassotti
University of Gothenburg,
Sweden

Stefano Montanelli
University of Milan,
Italy

Nina Tahmasebi
University of Gothenburg,
Sweden

Dominik Schlechtweg
University of Stuttgart,
Germany

https://github.com/FrancescoPeriti/TRoTR

Introduction
Text reuse

- We often reuse someone else's words.
  e.g., proverbs, quotations, reported speech

- There is a growing interest in studying text reuse, i.e., "*the reuse of existing written sources in the creation of a new text*" [(Clough et al., 2002)](#)

- Computational methods focus on the main task of Text Reuse Detection

"To be or not to be, that is the question"
William Shakespeare

"Cogito, ergo sum"
René Descartes

"That's one small step for man, one giant leap for mankind"
Neil Armstrong

"Float like a butterfly, sting like a bee"
Muhammad Ali

"I have a dream"
Martin Luther King

*"Live as if you were to die tomorrow. Learn as if you were to live forever"*
Mahatma Gandhi

Introduction
Text reuse detection

Setting

- Text reuses are all assumed as t*opically related* to the source [(Chiu et al., 2010)](#)

- Boundaries of reused text are unknown

- The goal is to detect text reuse

[…] I have told you this so that my joy may be in you and that your joy may be complete. My command is this: Love each other as I have loved you. Greater love has no one than this: to lay down one's life for one's friends. You are my friends if you do what I command. I no longer call you servants, because a servant does not know his master's business. Instead, I have called you friends, for everything that I learned from my Father I have made known to you. […]

John 15:11-15
NIV Bible

Text
Corpus
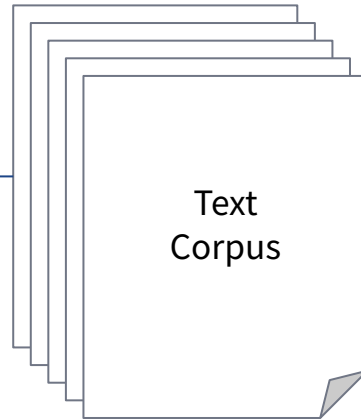
[…] It is to save these people of this suffering, this genocide is the main, core reason, causing motive and goal of the military operation that we've started in Donbass and in Ukraine. Words from the Holy Bible come to my head: there is no greater love than to lay down one's life for one's friends. And we see how heroically our guys act and fight during this operation. […]

Reported Putin's speech

Introduction
Recontextualization

- Topics or contexts of a reused text often differ from the original source

- There is a need of new methods for modeling recontextualization, i.e., "*the dynamic transfer and transformation of a text from one discourse to another*" (Connolly, 2014)

- We propose TRoTR, a framework for evaluating the recontextualization of text

As an example, consider three recontextualizations of the biblical passage *John 15:13* (in bold):

(1) It's the wonderful pride month!! ❤️ 🧡 💛 💚 💙 💜 Honestly pride is everyday! Love is love don't forget I love you ❤️. Remember this! John 15:12-13: "My command is this: Love each other as I have loved you. **Greater love has no one than this: to lay down one's life for one's friends**"

(2) At a large Crimean event today Putin quoted the Bible to defend the special military operation in Ukraine which has killed thousands and displaced millions. His words "**There is no greater love than if someone gives soul for their friends**". And people were cheering him. Madness!!!

(3) "Freeing people from genocide is the reason, motive & goal of the military operation we started in the Donbas & Ukraine", Putin says, then quotes the Bible: "**There is no greater love than to lay down one's life for one's friends**." It's like Billy Graham meets North Korea

TRoTR
Our original contribution

- TRoTR stands for *Topic Relatedness of Text Reuse*

- We provide TRoTR with **two NLP tasks**
  - TRiC
  - TRaC

- We provide TRoTR with a **benchmark**

- The benchmark contains

  - gold labels derived by **human annotation**

  - and a **baseline evaluation** based on SBERT models

Setting
- Boundaries of reused text are known

- The goal is to distinguish reuses of the same text according to their different, latent (i.e., unlabeled) topics
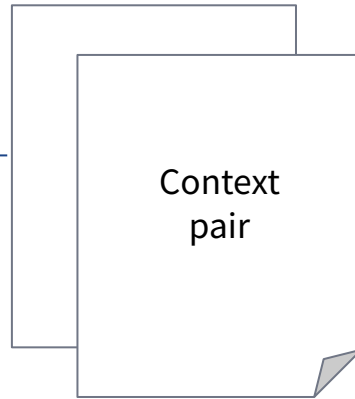
The TRoTR tasks
Text Reuse in Context

Subtask 1 binary classification
Subtask 2 ranking

At a large Crimean event today Putin quoted the Bible to defend the special military operation in Ukraine which has killed thousands and displaced millions. His words "Greater love has no one than this: to lay down one's life for one's friends". And people were cheering him. Madness!!!

Twitter

It's the wonderful pride month!! ❤️🧡💛💚💙💜 Honestly pride is everyday! Love is love don't forget I love you. Remember this!: My command is this: Love each other as I have loved you. Greater love has no one than this: to lay down one's life for one's friends

Twitter

Context pair

A text reuse is framed within two different contexts and the goal is to assess their **topic relatedness,** *i.e. the extent to which two texts share a common topic.*

The TRoTR tasks
Topic variation Ranking across Corpus

At a large Crimean event today Putin quoted the Bible to defend the special military operation in Ukraine which has killed thousands and displaced millions. His words "Greater love has no one than this: to lay down one's life for one's friends". And people were cheering him. Madness!!!

Twitter

A text reuse is framed within a text corpus and the goal is to assess its topic variation, *i.e. the variability in topic usages*

Text Corpus

It's the wonderful pride month!! ❤️🧡💛💚💙💜 Honestly pride is everyday! Love is love don't forget I love you. Remember this!: My command is this: Love each other as I have loved you. Greater love has no one than this: to lay down one's life for one's friends
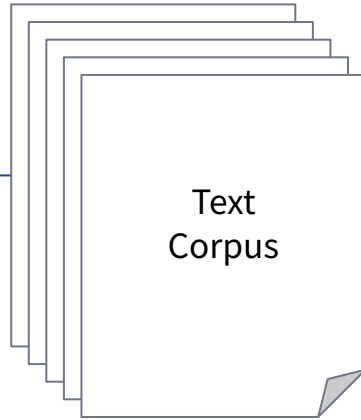
Twitter

The TRoTR annotation

- We collected and curated tweets from Twitter (now X) containing biblical text reuse instances.

- We then incorporated gold labels derived by human annotations.

- In our study, we sidestep the need for annotating topics explicitly

- Our annotation is aligned to the TRiC task

https://github.com/FrancescoPeriti/TRoTR

Annotation guidelines

Your task is to rate the degree of topic relatedness between two texts in which a *text sequence* is used. For instance, presented with a pair as in the below table, you are asked to rate the topic relatedness of the texts in which **Love your neighbor as yourself** is used.

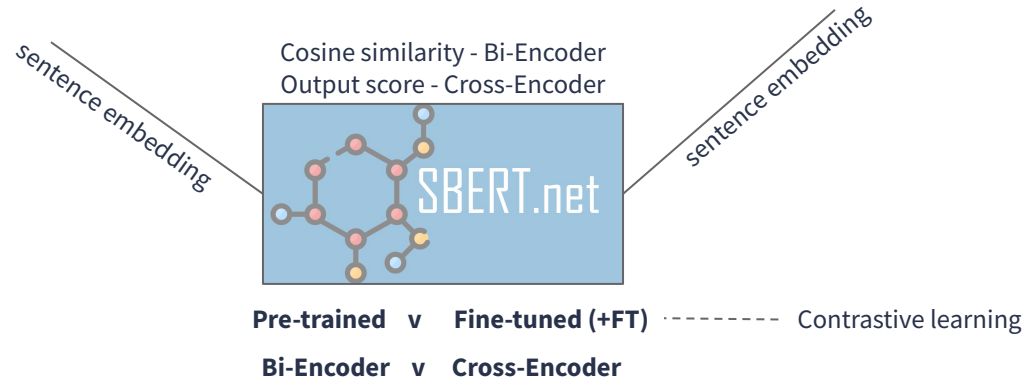| Text 1	Text 2
Love your neighbor as yourself. There is no commandment greater than these. You're a hypocritical Christian who ignores the greatest commandment because you're a bigot.	Jesus didn't tell you to be a bigot! Jesus had nothing to say about LGBTQIA+ people, but he did say to **love your neighbor as yourself.** #loveislove ❤️🧡💛💚💙💜

What is a topic?

The topic of a text answers the question "*What is this text about?*"

Evaluation setup
setting

At a large Crimean event today Putin quoted the Bible to defend the special military operation in Ukraine which has killed thousands and displaced millions. His words "Greater love has no one than this: to lay down one's life for one's friends". And people were cheering him. Madness!!!

It's the wonderful pride month!! ❤️🧡💛💚💙💜 Honestly pride is everyday! Love is love don't forget I love you. Remember this!: My command is this: Love each other as I have loved you. Greater love has no one than this: to lay down one's life for one's friends

sentence embedding

Cosine similarity - Bi-Encoder
Output score - Cross-Encoder

SBERT.net

sentence embedding

Pre-trained v **Fine-tuned (+FT)** --------- Contrastive learning

Bi-Encoder v **Cross-Encoder**
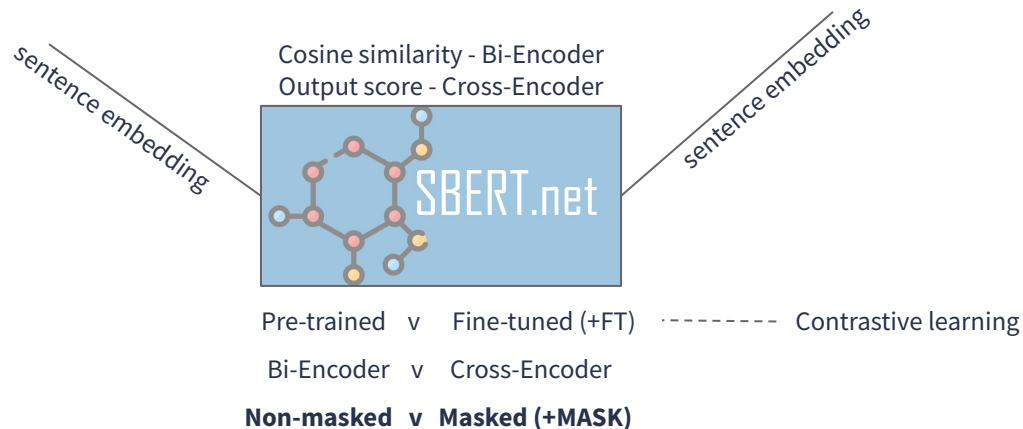
Evaluation setup
Masked setting

At a large Crimean event today Putin quoted the Bible to defend the special military operation in Ukraine which has killed thousands and displaced millions. His words ''

'':

And people were cheering him. Madness!!!

It's the wonderful pride month!! ❤️🧡💛💚💙💜 Honestly pride is everyday! Love is love don't forget I love you. Remember this!: My command is this: Love each other as I have loved you.

sentence embedding

Cosine similarity - Bi-Encoder
Output score - Cross-Encoder

SBERT.net

sentence embedding

Pre-trained   v   Fine-tuned (+FT)   --------   Contrastive learning

Bi-Encoder   v   Cross-Encoder

Non-masked v Masked (+MASK)

Evaluation setup
TRiC and TRaC

TRiC Subtask 1 binary classification
Threshold classifier based on similarities between sentences – **F1 score**

TRiC Subtask 2 ranking
Raw similarities between sentences – **Spearman correlation**

TRaC ranking
Average similarity for each target reuse – **Spearman correlation**

Evaluation results
Insights

- We considered 36 SBERT models

- Fine-tuned models outperformed pre-trained models

- Bi-Encoder models outperformed Cross-Encoder models

TRiC Subtask 1   TRiC Subtask 2

Models	All	
	F1	SP
ADR	.61±.10	.55±.09
+FT	.71±.10	.66±.07
+MASK	.82±.03	.67±.06
+FT+MASK	.85±.04	.71±.05
DBM	.43±.12	.54±.09
+FT	.61±.13	.64±.07
+MASK	.81±.03	.64±.04
+FT+MASK	.83±.03	.66±.04
PAM	.61±.07	.58±.08
+FT	.70±.09	.66±.06
+MASK	.83±.03	.67±.04
+FT+MASK	.86±.03	.69±.04
PAR	.56±.07	.56±.09
+FT	.71±.07	.66±.06
+MASK	.83±.03	.68±.03
+FT+MASK	.86±.03	.70±.04
MQA	.58±.09	.55±.09
+FT	.72±.09	.68±.06
+MASK	.83±.04	.68±.05
+FT+MASK	.86±.03	.72±.04

Evaluation results
Insights

- We considered 36 SBERT models

- Fine-tuned models outperformed pre-trained models

- Bi-Encoder models outperformed Cross-Encoder models

- Performance increased in the masked setting

  models exhibit a greater sensitivity to semantic similarity rather than topic relatedness

TRiC Subtask 1   **TRiC Subtask 2**

Models	All	
	F1	SP
ADR	.61±.10	.55±.09
+FT	**.71±.10**	**.66±.07**
+MASK	*.82±.03*	*.67±.06*
+FT+MASK	*.85±.04*	*.71±.05*
DBM	.43±.12	.54±.09
+FT	**.61±.13**	**.64±.07**
+MASK	*.81±.03*	*.64±.04*
+FT+MASK	*.83±.03*	*.66±.04*
PAM	.61±.07	.58±.08
+FT	**.70±.09**	**.66±.06**
+MASK	*.83±.03*	*.67±.04*
+FT+MASK	*.86±.03*	*.69±.04*
PAR	.56±.07	.56±.09
+FT	**.71±.07**	**.66±.06**
+MASK	*.83±.03*	*.68±.03*
+FT+MASK	*.86±.03*	*.70±.04*
MQA	.58±.09	.55±.09
+FT	**.72±.09**	**.68±.06**
+MASK	*.83±.04*	*.68±.05*
+FT+MASK	*.86±.03*	*.72±.04*

TRaC

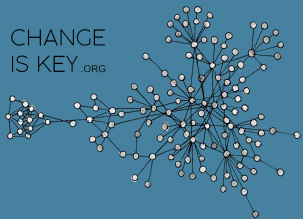| Models	ADR +MASK	DBM +MASK	PAM +MASK	PAR +MASK	MQA +MASK
Spearman	.72	.66	.66	.73	.65
	.84	.80	.81	.76	.80

13

Conclusion

- TRoTR: a framework for evaluating the recontextualization of text

- TRoTR consists of two tasks and a benchmark

  TRoTR can be used for any kind of text recontextualization / reuse
  (e.g., proverbs, quotations, citations)

- We provide a baseline by evaluating 36 SBERT models

- Our evaluation indicates that these models exhibit a bias towards their typical pre-training focus, namely *semantic similarity*

TRoTR: A Framework for Evaluating the Recontextualization of Text

Francesco Periti
University of Milan,
Italy

Pierluigi Cassotti
University of Gothenburg,
Sweden

Stefano Montanelli
University of Milan,
Italy

Nina Tahmasebi
University of Gothenburg,
Sweden

Dominik Schlechtweg
University of Stuttgart,
Germany