SANTM: Efficient Self-attention-driven Network for Text Matching

PRAYAG TIWARI, Department of Information Engineering, University of Padova, Italy
AMIT KUMAR JAISWAL, Institute for Research in Applicable Computing, University of Bedfordshire, United Kingdom
SAHIL GARG, École de technologie supérieure, Montréal, QC H3C 1K3, Canada
ILSUN YOU, Department of Information Security Engineering, Soonchunhyang University, Asan 31538, South Korea

Self-attention mechanisms have recently been embraced for a broad range of text-matching applications. Self-attention model takes only one sentence as an input with no extra information, i.e., one can utilize the final hidden state or pooling. However, text-matching problems can be interpreted either in symmetrical or asymmetrical scopes. For instance, paraphrase detection is an asymmetrical task, while textual entailment classification and question-answer matching are considered asymmetrical tasks. In this article, we leverage attractive properties of self-attention mechanism and proposes an attention-based network that incorporates three key components for inter-sequence attention: global pointwise features, preceding attentive features, and contextual features while updating the rest of the components. Our model follows evaluation on two benchmark datasets cover tasks of textual entailment and question-answer matching. The proposed efficient Self-attention-driven Network for Text Matching outperforms the state of the art on the Stanford Natural Language Inference and WikiQA datasets with much fewer parameters.

CCS Concepts: • **Information systems** → **Question answering**; *Document representation*; Information extraction;

Additional Key Words and Phrases: Text matching, deep learning, attention mechanism

ACM Reference format:
Prayag Tiwari, Amit Kumar Jaiswal, Sahil Garg, and Ilsun You. 2021. SANTM: Efficient Self-attention-driven Network for Text Matching. *ACM Trans. Internet Technol.* 22, 3, Article 55 (November 2021), 21 pages.
https://doi.org/10.1145/3426971

Prayag Tiwari and Amit Kumar Jaiswal contributed equally to the article.

This work received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 721321. It was also partially support by the Soonchunhyang University Research Fund (Ilsun You). This work was also supported by the Academy of Finland (grants 336033, 315896), Business Finland (grant 884/31/2018), and EU H2020 (grant 101016775).

Authors' addresses: P. Tiwari, Department of Information Engineering, University of Padova, Via Giovanni Gradenigo 6, 35131 Padova PD), Italy; Department of Computer Science, Aalto University, Espoo, Finland; email: prayag.tiwari@aalto.fi; A. K. Jaiswal, Institute for Research in Applicable Computing, University of Bedfordshire, United Kingdom; School of Mathematics, University of Leeds, Woodhouse, United Kingdom LS2 9JT; email: a.k.jaiswal@leeds.ac.uk; S. Garg, École de technologie supérieure, Montréal, QC H3C 1K3, Canada; email: sahil.garg@ieee.org; I. You (corresponding author), Department of Information Security Engineering, Soonchunhyang University, Asan 31538, South Korea; email: ilsunu@gmail.com.

1533-5399/2021/11-ART55 $15.00
https://doi.org/10.1145/3426971