



RECENT ADVANCES IN AI FOR NATIONAL SECURITY

Analyzing the Trade Space in Multi-lingual Automatic Text Difficulty Estimation – Esther Gupta

Poster Abstract

Assessing the difficulty of text material for foreign language learners remains a challenging task. Advances in the field of natural language processing shows promise for improving the performance of systems designed to evaluate text difficulty. In this work, we examine the impact of using multiple machine learning algorithms, transformer-based embeddings, and machine translation on 20 languages to determine their effectiveness on the text leveling task in an operational environment. We use accuracy and mean squared error (MSE) as our primary figures of merit. We also consider the computational consumption, ease of implementation, and speed of each approach. We find that while sentence embedding features offer some improvements for many languages, an ensemble method of more traditional algorithms gives larger performance gains with lower computational complexity.

About the Author(s)

Esther Gupta is a machine learning researcher at MIT Lincoln Laboratory. She has a background in mathematics and computer systems engineering and has been working mainly in the natural language processing space for the past eight years. Esther is interested in many different aspects of AI, including both the technical aspects and its social, ethical, and environmental implications.

For more information, please contact raains@ll.mit.edu

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

This material is based upon work supported under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. Air Force.

© 2023 Massachusetts Institute of Technology.

Delivered to the U.S. Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (Feb 2014). Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work. This presentation has been approved for public release by the Defense Language Institute Foreign Language Center's Public Affairs Office. For verification please e-mail: mpao@dliflc.edu Contents of this presentation are not necessarily the official views of the Defense Language Institute Foreign Language Center, nor are they endorsed by the Department of the Army, the Department of Defense, or the U.S. Government. All third party products / materials featured in the presentation remain the intellectual property of their respective authors / originators. Use of outside materials is done under the fair use copyright principle, for educational purposes only. The content of this presentation is the sole responsibility of the author(s).