Evaluating Medical Lexical Simplification: Rule-Based vs. BERT

Linh TRAN^{a,1}, Erick VELAZQUEZ^b, Robert-Jan SIPS^b and Victor DE BOER^a

^a *Vrije Universiteit, Amsterdam, The Netherlands*

^b *myTomorrows, Amsterdam, The Netherlands*

Abstract. Lexical simplification (LS) can decrease the communication gap between medical experts and laypeople by replacing medical terms with layperson counterparts. In this paper, we present: 1) a rule-based approach to LS using a consumer health vocabulary, and 2) an unsupervised approach using BERT to generate word candidates. Human evaluation shows that the unsupervised model performed better for simplicity and grammaticality, while the rule-based method was better at meaning preservation.

Keywords. Health Vocabulary, Lexical Simplification, Machine Learning

1. Introduction

The task of Lexical Simplification (LS) (1) in the medical domain is to replace complex terms by simpler ones. Recently, Qiang et al. (2) improved on the state-of-the-art by using pre-trained transformers to predict simplification. In this paper, we compare this method to a rule-based LS method for simplifying medical conditions at the word-level.

2. Methods

CHV Rule-based lexical simplification. This method uses the BERN (3), a named entity recognition tool, to identify a disease-related concept. These are then mapped to the UMLS (and thereby to CHV²) through 1) exact string matching, 2) looking up the MeSH IDs returned by BERN, and 3) approximate string matching using QuickUMLS (4). Finally, we rank the returned CHV terms on their annotated frequency scores and occurrence counts in the CHV-term generation method.

BERT-based lexical simplification. For this method, we follow the methodology of Qiang et al. (2) with the ClinicalBERT (5) model. Using the Masked Language Model (MLM), we find the closest predictions in BERT's vocabulary to the target word. Via the rank averaging strategy with the features: prediction probability, semantic similarity, word frequency, and n-gram language model probabilities, the best substitution candidate is selected. An experiment with BERT's feature-based approach to extract

¹ Corresponding Author, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands; E-mail: trplinh@gmail.com.

² <https://consumerhealthvocab.org/>

embeddings from a colloquial medical text is set up. To this end, we constructed a set of 814 articles from Simple Wikipedia describing medical conditions, accounting for 12,694 sentences and 204,938 words. We extract word embeddings for every word in this dataset using ClinicalBERT, resulting in a search index for finding the k most similar words to the target word. The same ranking method is used for substitution selection.

3. Results

84 Human judges on the Amazon Mechanical Turk platform rated 100 simplified sentences (50 CHV, 50 MLM) in three dimensions: *grammaticality*, *meaning preservation*, and *simplicity*. Responses were measured on a 1 to 5 Likert scale. Six unique individuals evaluated each sentence. The inter-annotator agreements for all scores were above 0.75. The results show that the average grammar score for the CHV simplified sentences is 4.17 ± 1.06 , and thus, a little lower than the grammar score for the MLM simplified sentences 4.36 ± 0.96 . In terms of meaning preservation, the CHV method scores 4.18 ± 0.97 on average, which is a lot higher in comparison to the average score of 3.28 ± 1.34 achieved with the MLM method. Lastly, the CHV method scores slightly lower in terms of simplicity compared to the MLM method, 3.55 ± 1.11 and 3.66 ± 0.96 , respectively. An independent-sample t-test shows a significant difference between the grammaticality and the meaning preservation scores.

4. Discussion

The human evaluation shows encouraging results for the BERT language model on the LS task for simplicity and grammaticality. However, using the feature-based approach, we observe that the ClinicalBERT model often suggests related candidates (hypernyms, hyponyms, related symptoms, etc.) rather than synonyms, thereby not preserving the exact meaning of the sentence. Therefore, it falls short of more traditional approaches in meaning preservation, which hampers real-world applications. For future work, we plan to improve unsupervised lexical simplification in the medical domain by complementing BERT's distributional representations with external lexical knowledge based on this method (6) and taking synonymous pairs from the UMLS as the semantic similarity constraints. This way, semantically similar candidates will be generated, and thus, meaning preservation can be improved.

References

- [1] Paetzold GH, Specia L. A survey on lexical simplification. *J Artif Intell Res.* 2017;60:549–93.
- [2] Qiang J, Li Y, Zhu Y, Yuan Y, Wu X. Lexical Simplification with Pretrained Encoders. *Proc AAAI Conf Artif Intell.* 2020 Apr 3;34(05):8649–56.
- [3] Kim D, Lee J, So CH, Jeon H, Jeong M, Choi Y, et al. A Neural Named Entity Recognition and Multi-Type Normalization Tool for Biomedical Text Mining. *IEEE Access.* 2019;7:73729–40.
- [4] Soldaini L, Goharian N. QuickUMLS: a fast, unsupervised approach for medical concept extraction. *MedIR Workshop Sigir.* 2016 Jul;1–4.
- [5] Alsentzer E, et al. Publicly Available Clinical BERT Embeddings. *ArXiv190403323 Cs [Internet].* 2019 Jun 20 [cited 2020 Jan 19]; Available from: <http://arxiv.org/abs/1904.03323>.
- [6] Lauscher A, et al. Specializing Unsupervised Pretraining Models for Word-Level Semantic Similarity. *ArXiv190902339 Cs [Internet].* 2020 Apr 20; Available from: <http://arxiv.org/abs/1909.02339>.