THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS

Master's Thesis Presentation

Junchi Zhang
Department of Statistics
The University of Chicago

"NLP Application in Clinical Records — Using BERT and Edit Distance for Spell Check Pipeline"

Thursday, **January 18,** 2024, at 2:00 PM, **via Zoom**
Zoom information provided in email announcement.

Abstract

Given the crucial role that clinical notes play in ensuring optimal patient care, it is imperative to maintain their accuracy, as even minor inaccuracies can have a significant impact on patient treatment and outcomes. Spelling correction, especially within the specialized domain of clinical settings, remains a significant challenge due to the plethora of medical terminologies and abbreviations. In this paper, we will explore the utilization of BERT-based models and Damerau-Levenshtein distance, combining BERT's predictive power with the accuracy of algorithm-based spelling correction using edit distance to amend misspellings in clinical notes. A core aspect of our research involves delving deeply into the mathematical intricacies of both the BERT model and the edit distance algorithm. Additionally, we aim to emphasize the extra benefits gained from meticulously fine-tuning the model with data specific to the medical domain. We will demonstrate how our spellcheck pipeline performs using real-world clinical notes, highlighting its practical applications in healthcare settings.

Key words: NLP, BERT model, Edit Distance, Spelling Correction, Clinical notes.
