

学位論文抄録

An alternative application of Natural Language Processing to express a characteristic feature
of diseases in Japanese medical records
(新しい自然言語処理を用いて日本語診療記録から疾患特徴表現を抽出する)

山ノ内 祥訓

熊本大学大学院医学教育部博士課程医学専攻医療情報学

指導教員

宇宿 功市郎 前教授

熊本大学大学院医学教育部博士課程医学専攻医療情報学

中村 太志 教授

熊本大学大学院医学教育部博士課程医学専攻医療情報学

Abstract of the Thesis

Background: Owing to the linguistic situation, Japanese natural language processing (NLP) requires morphological analyses for word segmentation using dictionary techniques.

Objective: We aimed to clarify whether it can be substituted with an open-end discovery-based NLP (OD-NLP), which does not use any dictionary techniques.

Methods: Clinical texts at the first medical visit were collected for comparison of OD-NLP with word dictionary-based NLP (WD-NLP). Topics were generated in each document using a topic model, which later corresponded to the respective diseases determined in international statistical classification of diseases and related health problems (ICD)-10. The prediction accuracy and expressivity of each disease was examined in equivalent number of entities/words after filtration with either term frequency and inverse document frequency (TF-IDF) or dominance value (DMV).

Results: In documents from 10,520 observed patients, 169,913 entities and 44,758 words were segmented using OD-NLP and WD-NLP, simultaneously. Without filtering, accuracy and recall levels were low, and there was no difference in the harmonic mean of the F-measure between NLPs. However, physicians reported OD-NLP contained more meaningful words than WD-NLP. When datasets were created in an equivalent number of entities/words with TF-IDF, F-measure in OD-NLP was higher than WD-NLP at lower thresholds. When the threshold increased, the number of datasets created decreased, resulting in increased values of F-measure, although the differences disappeared. Two datasets near the maximum threshold showing differences in F-measure were examined whether their topics were associated with diseases. The results showed that more diseases were found in OD-NLP at lower thresholds, indicating that the topics described characteristics of diseases. The superiority remained as much as that of TF-IDF when filtration was changed to DMV.

Conclusion: The current findings prefer the use of OD-NLP to express characteristics of diseases from Japanese clinical texts and may help in the construction of document summaries and retrieval in clinical settings.

学位論文抄録

[背景と目的] 日本語の自然言語処理(Natural Language Processing: NLP)では、その言語的特性から単語辞書を用いて文を単語の並びに分割する形態素解析が必要である。しかし、この手法は単語辞書の充実度に強く依存し、未知の単語や略語等が日々発生する分野においては十分な性能を得られない。そこで、単語辞書を用いないオープンエンド発見手法型 NLP(Open-end Discovery-based NLP: OD-NLP)を用いた自然言語処理で代替可能かどうか明らかにすることを目的とした。

[方法] OD-NLP と単語辞書型 NLP(Word Dictionary-based NLP: WD-NLP)の二つの自然言語処理手法を比較するために、2015 年から 2017 年の間に熊本大学病院を受診した患者で初診時に自由記載文で記録された診療記録を使用した。各診療記録は Latent Dirichlet allocation(LDA)によるトピックモデルを用いてトピックを生成し、各トピックは国際疾病分類第 10 版(ICD-10)で分類された疾患に対応させた。各疾患の予測精度と特徴表現は、単語頻度・逆文書頻度(Term Frequency - Inverse Document Frequency: TF-IDF)またはドミナンス値によるフィルタリング前後のエンティティ/単語を使用した。評価はトピックモデルで分類したトピックと実際の疾患の対応を解析した 2×2 の分割表と、各トピックで重要度の高い単語が疾患の特徴をとらえているかを複数の医師が判定することとした。

[結果] 10,520 人の患者の診療記録から、169,913 のエンティティと 44,758 の単語が OD-NLP と WD-NLP によってそれぞれ分割された。フィルタリングを行わない場合、適合率と再現率は低く、調和平均である F 値は両手法間で差がなかった。しかし、医師による評価を行ったところ、OD-NLP の方が WD-NLP よりも意味のある単語を多く含んでいると報告した。次に、TF-IDF で同数のエンティティ/単語でデータセットを作成した場合、低い閾値では OD-NLP の F 値は WD-NLP より高かった。しかし、閾値が大きくなると作成されるトピック数が減少し、その結果、差はなくなるものの F 値の値が大きくなった。最後に、F 値に差がある最大閾値付近の 2 つのデータセットについて、そのトピックが疾患と関連しているかどうかを確認した。その結果、OD-NLP では閾値が低いほど多くの疾患が特定され、トピックが疾患の特徴を表していることがわかった。また、フィルタリングをドミナンス値に変更しても、TF-IDF と同程度の優位性が保たれた。

[結論] 本研究では、日本語の診療記録から疾患の特徴を表現するための自然言語処理について、WD-NLP よりも OD-NLP を使用することが望ましいことを示した。今後この技術は、臨床現場における文書要約の作成や類似文書の検索など、医師の業務負担を低減させるアプリケーションの開発に役立つと考えられる。