

矯正歯科治療における所見文書からの自動診断に向けて

西原 大貴[†] 梶原 智之[‡] 谷川 千尋^{*} 清水 優仁^{*} 長原 一[‡]

[†]大阪大学大学院情報科学研究科 [‡]大阪大学データビリティフロンティア機構 ^{*}大阪大学歯学部附属病院

1 はじめに

歯科医師が適切な診断を行うには、豊富な知識と長年の経験が必要であり、経験の浅い歯科医師には問題の見落としがしばしば見られる。専門医の思考過程を模倣するシステムを実現できれば、経験の浅い歯科医師にとっては問題の見落としの防止、専門医にとっては大きな負担軽減につながり、根拠に基づく安心安全な医療を患者に広く提供する上で大きな意義を有する。

このような背景のもと、自動診断は古くから試みられてきた。矯正歯科治療においても、抜歯・非抜歯を選択するファジー理論に基づく矯正歯科診断支援システム[1]や、矯正装置の選択支援システム[2]が開発されている。しかし、これらは 2-3 種類の治療方針の選択という単純な問題に対して機械学習を適用しているにすぎず、矯正歯科臨床において応用可能な、包括的な自動診断システムの開発には至っていない。

本研究では、自然言語処理の技術を用いて、所見文書から特徴表現を抽出し、診断を行った。専門医による評価の結果、実際の利用に耐える品質で自動診断を実現できることを確認した。

2 タスク設計：矯正歯科診断の自動化

本研究では、大阪大学歯学部附属病院に所蔵されている治療計画書のデータを用いて、臨床所見から診断を行う自然言語処理モデルを訓練する。本節では、データとタスクを概説する。

図 1 に示すように、我々の治療計画書コーパスには、診断結果と治療計画に加えて、年齢や性別などの患者属性情報、顔画像・X 線画像・模型に基づく所見テキストが含まれる。本研究では、967 件の治療計画書を用いて、患者属性情報および所見テキストから診断結果を推定する。診断結果は、自由記述の病状を列挙したリストとして提供されている。しかし、この規模のデ



図 1：治療計画書コーパスの概要

ータから自由記述のテキストを生成することは困難である¹ため、意味的に等価な病状をまとめ、422 種類のラベルからなるマルチラベル分類問題としてタスクを設計する。基本的な統計としては、所見テキストは 1 件あたり平均 2,176 単語、所見テキストの語彙サイズは全体で 9,531 単語、病状リストは 1 件あたり平均 13.2 ラベルである。

3 マルチラベル文書分類モデル

一般的な文書分類モデルを適用するために、Binary Relevance[3]の手法 (BR 法) を用いてマルチラベル分類をシングルラベル分類タスクに変換して解く。BR 法は、ラベルごとに 2 値分類器を訓練するアプローチであり、ラベル間の関係を考慮しない代わりに計算量が小さいという利点を持つ。なお、ラベル間の関係を考慮できる Classifier Chain[4]や Label Powerset[5]の手法も検討したが、本稿では予備実験においてより高い性能を達成した BR 法について報告する。

所見文書からの素性抽出には、Bag of Words モデル (BoW) および単語を句に拡張した Bag of Phrases モデル (BoP) を用いる。これらは、文書中に出現した単語を 1、その他の単語を 0 で表現する語彙サイズ次元のベクトルを用いる素性抽出手法である。BoP では、「咬合 平面 板」など名詞の連続を句として用いる。これによって、専門用語を考慮でき、性能の改善を期待できる。なお、USE²や BERT³などの深層学習に基づく素性

Toward Automated Orthodontic Diagnosis from Findings

Daiki Nishihara[†] (nishihara.daiki@ist.osaka-u.ac.jp)

Tomoyuki Kajiwara[‡] (kajiwara@ids.osaka-u.ac.jp)

Chihiro Tanikawa^{*} (ctanika@dent.osaka-u.ac.jp)

Yuuji Shimizu^{*} (yjshimizu@dent.osaka-u.ac.jp)

Hajime Nagahara[‡] (nagahara@ids.osaka-u.ac.jp)

[†] Graduate School of Information Science and Technology, Osaka University

[‡] Institute for Datability Science, Osaka University

^{*} Graduate School of Dentistry, Osaka University

¹ 機械翻訳などのテキスト生成タスクでは、数十万から数百万文の規模の訓練データを用いるのが一般的である。

² <https://tfhub.dev/google/universal-sentence-encoder-multilingual/3>

³ [https://github.com/google-](https://github.com/google-research/bert/blob/master/multilingual.md)

[research/bert/blob/master/multilingual.md](https://github.com/google-research/bert/blob/master/multilingual.md)

表 1: 所見文書のマルチラベル分類の実験結果

	Precision	Recall	F1
BoW + SVM	0.521	0.360	0.426
BoP + SVM	0.660	0.319	0.431
BoW + LGBM	0.667	0.363	0.470
BoP + LGBM	0.704	0.389	0.501

抽出手法も検討したが、本稿では予備実験においてより高い性能を達成した BoW および BoP について報告する。

分類器には、Support Vector Machine (SVM) [6] および Light Gradient Boosting Machine (LGBM) ⁴[7]を用いる。SVM はマージン最大化に基づく手法、LGBM は勾配ブースティングに基づく手法であり、いずれも文書分類タスクで広く用いられている。なお、Random Forest や Multi-Layer Perceptron の手法も検討したが、本稿では予備実験においてより高い性能を達成した SVM および LGBM について報告する。

4 実験設定

967 件の治療計画書を、800 件の訓練用データ、67 件の検証用データ、100 件の評価用データに無作為に分割して用いた。単語分割には MeCab (IPADIC-2.7.0) ⁵を使用した。

BR 法は scikit-multilearn⁶ の実装、SVM は scikit-learn⁷ の実装を用いた。SVM には RBF カーネルを使用した。

各手法の性能は、Precision、Recall、および F1 スコアによって自動評価した。また、自動評価において最も高い性能を達成した手法について、人手評価も実施した。人手評価には、評価用データから無作為抽出した 40 件について、2 人の専門医が以下の基準で 4 段階評価を行った。

1. 許容できない
2. 許容できる
3. 良い
4. 理想的

5 実験結果

自動評価の結果を表 1 に示す。Precision と Recall の調和平均である F1 スコアを比較すると、分類器によらず一貫して BoW よりも BoP が高い性能を達成することがわかる。これは我々の期待通り、名詞句の使用によって専門用語を考慮することの有効性を示すものである。分類器に

注目すると、SVM よりも LGBM が高い性能を達成している。本実験では小規模なデータセットを用いたため、LGBM の弱分類器によるアンサンブル学習が有効であったと考えられる。

表 1 において最高性能を達成した BoP+LGBM の手法を、2 人の専門医が人手評価した。人手評価では、4 点満点のうち平均で 3.4 点の評価が得られ、実際の利用に耐えうる品質で矯正歯科治療の自動診断を実現できることを確認できた。

6 おわりに

本研究では、矯正歯科治療における所見文書からの自動診断を目的に、967 件の症例にアノテーションを施した。マルチラベル文書分類タスクとして自動診断の問題を設計し、専門医から高い評価を得る自然言語処理モデルを構築した。

今後は、本研究で用いていない治療計画の部分にもアノテーションを追加し、自動診断に加えて治療計画の自動生成にも取り組みたい。また、対応する顔画像や X 線画像を組み合わせたマルチモーダルなアプローチにも取り組みたい。

参考文献

- [1] J. H. Sims-Williams, I. D. Brown, A. Matthewman, C. D. Stephens: A Computer-Controlled Expert System for Orthodontic Advice, *British Dental Journal*, Vol.163, No.5, pp.161-166 (1987)
- [2] C. D. Stephens, N. Mackin, J. H. Sims-Williams: The Development and Validation of an Orthodontic Expert System, *British Journal of Orthodontics*, Vol.23, No.1, pp.1-9 (1996)
- [3] Grigorios Tsoumakas, Ioannis Katakis: Multi-Label Classification: An Overview, *International Journal of Data Warehousing and Mining*, Vol.2007, pp.1-13 (2007).
- [4] Jesse Read, Bernhard Pfahringer, Geoff Holmes, Eibe Frank: Classifier Chains for Multi-Label Classification, *In Proc. of PKDD*, pp.254-269 (2009)
- [5] Everton Alvares Cherman, Maria Carolina Monard, Jean Metz: Multi-Label Problem Transformation Methods: a Case Study, *CLEI Electronic Journal*, Vol.14, No.1, pp.1-10 (2011)
- [6] Bernhard E. Boser, Isabelle M. Guyon, Vladimir N. Vapnik: A Training Algorithm for Optimal Margin Classifiers, *In Proc. of COLT*, pp.144-152 (1992)
- [7] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu: LightGBM: A Highly Efficient Gradient Boosting Decision Tree, *In Proc. of NIPS*, pp.3149-3157 (2017)

⁴ <https://github.com/microsoft/LightGBM>

⁵ <https://taku910.github.io/mecab/>

⁶ <http://scikit.ml/>

⁷ <https://scikit-learn.org/>