

博士学位論文 審査結果の要旨

芝浦工業大学大学院 理工学研究科 博士（後期）課程

博士学位論文審査委員会

主 査 木村昌臣

審査委員 杉本 徹

審査委員 菅谷みどり

審査委員 村上嘉代子

審査委員 Mohammad Iqbal

*審査委員

氏 名	GLORIANA MONKO JOSEPH
論文題目	A Multi-Stage Framework for Adverse Drug Reaction Detection: Integrating SS-DBSCAN Clustering, Human-in-the-Loop Labeling, and Transformers
〔論文審査の要旨〕	
<p>医薬品の有害事象は、患者の安全に対する大きな脅威である。しかし、現在の医薬品安全性監視システムには、医療従事者による報告の不足、データ品質の低さ、有害事象検出に適した枠組みの欠如といった深刻な問題がある。一方で、電子カルテの利用が進むことにより、機械学習や自然言語処理技術を用いて有害事象の検出を自動化する仕組みを構築することが期待される。特に発展途上国においては医療従事者が極めて不足しており、有害事象の報告に割ける時間・労力がないため有害事象が十分に報告できないという背景もある。そのため、電子カルテに入力されたデータから有害事象を効率よく抽出し医療従事者を支援するシステムが必要とされる。</p> <p>申請者は本論文にて、電子カルテに記録されたテキスト形式の臨床データから有害事象情報を抽出するための堅牢でスケーラブルな手法を提案した。具体的には、電子カルテデータに含まれる複雑な構造やノイズを適切に扱うため、密度ベースのクラスタリング手法として有名な DBSCAN を改良した層化サンプリング DBSCAN (SS-DBSCAN) を提案した。この手法では、周辺データの探索範囲を示すパラメータ ϵ (イプシロン) の動的な推定に層化サンプリング戦略を用い (SS-DBSCAN)、周辺データの最小個数を指定する MinPoints パラメータの最適化には探索の無駄を省く高速化したグリッドサーチを提案することによる改良を行っている。さらに、Human-in-the-Loop 手法 (HITL) を適用し、専門家のフィードバックを適切に反映してラベル付けの精度を段階的に向上させるデータへのラベル付加手法を実現している。これにより一部にしかラベルが付与されていないデータセットに対しても高速で高精度なラベル付けを可能にした。また、Sentence-BERT を用い電子カルテの文書内に見られる意味的關係を捉えて有害事象の予測を行う。主成分分析や t-SNE という次元削減技術も導入し、特徴表現の精度向上とクラスタ結果の可視化を実現した。結論として、提案フレームワークは SS-DBSCAN、HITL などを活用することにより高次元かつノイズの多いデータセットにおいてデータへのラベルの付加と有害事象検出について高い性能を持つことを示した。また、さらに電子カルテの入力を受け付け、有害事象検出の結果を患者が服用した医薬品の情報と関係づけて表示・ファイル生成を行うでもプログラムも作成し、現存のシステムとの連携の仕方についてのデモンストレーションを行った。</p> <p>最終審査においては、一時間の研究内容の発表ののち、約一時間の質疑が行われた。電子カルテデ</p>	

ータをベクトル空間に埋め込んだときのデータ分布からみた SS-DBSCAN の必要性や、他のクラスタリング手法との比較、人手で作成した少数のデータを学習データとした有害事象が含まれているかどうかを判定する分類モデルと比較したときの本研究の優位性などについての質問があり、概ね必要な回答を行うことができおり、その結果、博士論文審査委員会としては合格に足るものであると判断した。