

令和 5 年 2 月 7 日

工学研究科委員長
掛下 知行 殿

審査委員会報告書

審査委員	
(主査) 教授 芥子 育雄	
教授 八木 哲也	
教授 大下 福仁	

学位論文提出者氏名
辻岡 和孝

学位論文題目

自然言語処理を利用した本邦版 Computer Assisted Coding 構築手法の提案と性能評価

学位申請受理年月日

令和 5 年 1 月 25 日

1. 学位論文の内容の要旨

別紙論文要旨のとおり

2. 学位論文審査結果の要旨

米国では、退院サマリから診断病名のコーディングを自然言語処理により支援する Computer Assisted Coding (CAC) の普及が進む。CAC は診療情報管理士の業務支援に加えて、医師の負担軽減や診療支援にも活用できる可能性があり、国際的に重要なテーマである。しかし、日本では、医師による自由記述が行われる退院サマリの経過要約には、略語や表記揺れも多く、また、退院サマリは要配慮個人情報であることから、一部大学の研究者を中心に自然言語処理による病名推定の研究がされるに留まっている。米国の電子カルテは主にテンプレート入力方式であり、コード入力も多く、表記揺れは少ない。

本論文では、非構造化テキストを意味的表現からベクトル化して解析する手法である意味表現学習を適用した日本版 CAC 構築手法を提案している。経過要約を意味表現学習することで得られた「264 個の特徴単語ベクトル値」と「年齢」「性別」「診療科名」を説明変数、「診断病名コード」を目的変数とした、「意味表現学習+サポートベクターマシン (SVM)」による CAC を構築し、以下に示す 3 種類の評価を元に先行研究に対

する優位性を明らかにしている。尚、6つの医療機関の退院サマリを対象にクロス評価を行った千葉大学の鈴木らの研究を先行研究としている。但し、鈴木らの研究では評価に正答率を用いているが、本研究では、結果に至った理由をより詳細に捉えるため、正答率ではなく再現率と適合率の調和平均であるマクロ平均 F 値により評価を行っている。

評価 1 は自施設内において他施設のデータにも病名推定が適用できるかどうかの汎用性の評価（クロス評価）を試みたもので、先行研究と同等レベルのマクロ平均 F 値を得ている。

評価 2 では経過要約の大規模コーパスによる事前学習モデルを前提にクロス評価を試みたもので、約 0.75 マクロ平均 F 値を得ている（先行研究は 6 医療機関の統合データにてモデルを作成し正答率 70%）。

評価 3 では自施設内データによる評価に対応し約 0.87 のマクロ平均 F 値を得ている（先行研究の正答率 72%）。本研究で提案する「意味表現学習+SVM」の組み合わせが CAC として安定して動作するかを検証したものである。これにより、診断病名コードに詳細不明コード「(.9)」が付与された状態で学習モデルを作成すると、詳細分類毎の F 値に負の影響を与えているが、詳細分類コードまで入力することで、F 値の改善が示唆されている。

本論文の貢献は以下の 3 点である。第 1 に「意味表現学習+SVM」の組み合わせによる CAC 構築手法は、予測値を特徴単語と経過要約との関連性を数値化して説明することができるため、説明責任を重要視する医療分野において親和性のある手法である。第 2 に機械学習時の説明変数の追加が任意に設定できるため、混合行列の結果より、診断病名コーディングの専門家である診療情報管理士の知見を活かした予測モデルが構築できることである。第 3 にデータ分布が異なる 2 種類の電子カルテシステムの退院サマリを用い、病名推定のための機械学習モデルの汎用的な評価が行えるベンチマークを確立したことである。

本研究に関して、2 編の原著論文と 1 編の速報論文、3 編の学会発表を行っており、これまでの種々の研究業績を総括すると 8 編の学術論文（1 編の速報論文含む）、32 編の査読なし発表を行っていることから、高く評価できる。

上記の研究内容を審査委員会で慎重に検討した結果、辻岡和孝氏は博士研究者として十分な能力を有しており、また本論文は博士（工学）の論文として学術上、価値のあるものと認める。

3. 公聴会の日時

令和 5 年 2 月 2 日 14:00 ~ 15:15

4. 最終試験結果の要旨

令和 5 年 2 月 7 日、論文内容およびそれに関連した事項、学科内での予備審査及び公聴会での質疑に対する応答などについて試問を行い、審議の結果合格と認めた。

5. 審査委員会の所見

審査の結果および最終試験結果等を考慮して、申請者は博士（工学）の学位を授与される資格を有するものと認める。