

様 式 F - 7 - 2

科学研究費助成事業（学術研究助成基金助成金）実績報告書（研究実績報告書）

1. 機関番号

1	4	6	0	3
---	---	---	---	---

 2. 研究機関名 奈良先端科学技術大学院大学
3. 研究種目名 挑戦的萌芽研究 4. 補助事業期間 平成25年度～平成27年度
5. 課題番号

2	5	5	4	0	0	9	6
---	---	---	---	---	---	---	---
6. 研究課題名 テキストの安全な匿名化に関する研究

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
7 0 4 0 1 0 7 3	アラマキ エイジ	研究推進機構	准教授
	荒牧 英治		

8. 研究分担者

研究者番号	研究分担者名	所属研究機関名・部局名	職名
0 0 5 1 9 3 1 6	モリタ ミズキ	岡山大学・医歯（薬）学総合研究科	准教授
	森田 瑞樹		

9. 研究実績の概要

現在、医療のIT化が進み、その結果、かつてない大量の臨床データが電子化された状態でストックされつつある。このデータを有効に利用することができれば患者の生活習慣と疾患の相関（例えば、喫煙と癌）や、薬品とその副作用の相関（タミフルと精神障害）について過去類を見ない大規模な調査が可能となり、臨床研究が加速的に進展するとして高い期待が寄せられている。しかし、単にデータを電子化しただけで、大規模な調査が可能となるわけではない。実際に大規模にカルテを共有し解析するためにはこれらを匿名化する必要がある。さらに、電子カルテには自然言語で入力される箇所が相当な割合で存在するため、データをフル活用するためには、まず文章を匿名化する必要がある。

我々も2006年から匿名化の研究を開始し、2007年ではF値0.98という高い精度で、個人情報除去に成功している。しかし、このような高精度であってもカルテの提供にあたっては問題が生じる。例えば、「2012年の小児の移植ドナー」など年間に数例しか行われない治療であった場合、個人情報の範囲外である単なる年代と術名の組み合わせから個人を特定可能な場合がある。また、筆記上の特徴から医師が判明可能である可能性もある。

このような個人情報の削除だけでは十分な匿名化と言えない場合に対応するため、本研究では、文章固有の表現を削除することで、文章を特定不可能とする新しいタイプの匿名化を提案した。これは、任意の文字列が最低k回以上出現するようにテキストの一部を削除するものであり、筆者らはこれを「テキストのk匿名化」とよび、そのアルゴリズムについて発表を行った。

さらに、「テキストのk匿名化」を実用化するために必須となる高速化について研究を行った。これらはする論文発表（国内2本）され、さらに、提案する匿名化手法を実装したデバイス（匿名コピー）を開発した。

10. キーワード

- (1) 自然言語処理 (2) 医療情報学 (3) 匿名化 (4) プライバシー保護
- (5) _____ (6) _____ (7) _____ (8) _____

（注）・印刷に当たっては、A4判（縦長）・両面印刷すること。

11. 研究発表

(雑誌論文) 計(0)件/うち査読付論文 計(0)件 (最終年度分)

/うち国際共著論文 計(0)件 (最終年度分) /うちオープンアクセス 計(0)件 (最終年度分)

著者名		論文標題				
雑誌名	査読の有無	巻	発行年	最初と最後の頁	国際共著	
掲載論文のDOI(デジタルオブジェクト識別子)						
オープンアクセス						

(学会発表) 計(2)件/うち招待講演 計(2)件 (最終年度分) /うち国際学会 計(0)件 (最終年度分)

発表者名		発表標題	
荒牧英治, 増川佐知子, 宮部真衣, 森田瑞樹		テキストデータを対象にした医療に関するビッグデータ解析の動向	
学会等名	発表年月日	発表場所	
医療の質安全学会・学際的医療研究入門WS(招待講演)	2015年09月05日	東京都, 品川区, 赤坂Bizタワー	

発表者名		発表標題	
荒牧英治		言語処理がつくる医療と情報のシナジー	
学会等名	発表年月日	発表場所	
Synergy!LIVE2015(招待講演)	2015年12月08日	東京都, 品川区, 品川インターシティ	

(図書) 計(0)件 (最終年度分)

著 者 名		出 版 社	
書 名		発行年	総ページ数

12. 研究成果による産業財産権の出願・取得状況

(出願) 計(0)件 (最終年度分)

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	出願年月日	国内・外国の別

(取得) 計(0)件 (最終年度分)

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	取得年月日	国内・外国の別
				出願年月日	

13. 科研費を使用して開催した国際研究集会

(国際研究集会) 計(0)件 (最終年度分)

国際研究集会名	開催年月日	開催場所

14. 本研究に関連して実施した国際共同研究の実施状況

(1) 国際共同研究: -

15. 備考

ソーシャル・コンピューティング研究室
<http://sociocom.jp>