

未来社会創造事業 探索加速型  
「共通基盤」領域  
終了報告書(探索研究期間)

令和3年度  
研究開発終了報告書

令和元年度採択研究開発代表者

[研究開発代表者名：香川 璃奈]

[筑波大学医学医療系・講師]

[研究開発課題名：質的な知を客体化するドキュメンテーション基盤技術]

実施期間：令和元年11月1日～令和4年3月31日

## § 1. 研究実施体制

### (1)「医療」グループ(国立大学法人 筑波大学)

① 研究開発代表者:香川 璃奈 (筑波大学医学医療系、講師)

#### ② 研究項目

- ・医療領域の実験
- ・料理レシピの実験

### (2)「応用」グループ(国立大学法人 筑波大学)

① 主たる共同研究者:松原 正樹 (筑波大学図書館情報メディア系、准教授)

#### ② 研究項目

- ・芸術領域の実験
- ・各応用領域に関する知見に基づく助言

### (3)「基礎」グループ(国立大学法人 筑波大学)

① 主たる共同研究者:馬場 雪乃 (筑波大学システム情報系、准教授)

#### ② 研究項目

- ・「医療」グループおよび「応用」グループにおける解析手法の提案と助言

### (4)「認知的モデリング」グループ(追手門学院大学)

① 主たる共同研究者:本田 秀仁 (追手門学院大学心理学部、准教授)

#### ② 研究項目

- ・研究用データの質と量を増加させる入力フォーマットの開発
- ・タスクに応じた背景知識の計測手法の開発

### (5)「数理的モデリング」グループ(国立大学法人東海国立大学機構 名古屋大学)

① 主たる共同研究者:宮田 玲 (名古屋大学工学系研究科、助教)

#### ② 研究項目

- ・図書館情報学の観点に則った「医療」グループおよび「応用」グループから得られた成果のとりまとめ
- ・自然言語処理技術の知見に基づく「医療」グループおよび「応用」グループから得られた成果のとりまとめ

## § 2. 研究実施の概要

本研究開発では、文書というかたちでしか記録できない質的な知を、書き手に負担のない範囲でより多く収集すること支援し、さらに収集された文書から読み手が利用したい情報をより多く収集することを目標とする。その基盤として、文書特徴を明らかにする集合知フレームワークを開発し、文書タイプおよび文書が持つコミュニケーション上の目的を自由に指定した上で、誰でも利用できる要素技術として提案した。

開発した要素技術を、以下の3つの応用課題に対して適用し、それぞれ成果を得た。

### 1. 読み手にとって有用な文書の定量的評価 [1, 2]

料理レシピにおいて、上記手法を適用し、全くの経験的な知見であった読み手に有用な記載内容を6種類に集約し、さらにそのうちの一部が読み手が感じる有用性を有意に増加させることを示した。

料理レシピ以外の、カルテ文書や楽器演奏指導文書でも同様に、読み手が感じる有用性を有意に増加させる特徴を明らかにできることを示し、教育応用の準備を進めている。

2. 自然言語処理技術を利用した校正支援システムに対する、読み手に有用な記載内容の判別の組み込み  
料理レシピは web 上に大規模なテキストが公開されている。そこで、料理レシピを対象として、先の結果に基づき、読み手に有用なレシピの記載内容を 6 種類のいずれであるか(あるいは、いずれでもないか)を自動識別する技術を実装した。レシピ管理アプリに搭載する準備を進めている。なおこの 6 種類の内容は、いずれも、動画や画像、あるいは将来的に音や香りなどの知覚可能なデータの共有が進んだとしても、言語情報を使わずに共有することが極めて難しいと想定される知である。
3. 文書の定性的かつ経験的な特徴の顕在化と、その特徴を備えた教育研究用データセットの作成と公開 [3]  
病院で、患者の診察結果だけでなく医師の思考過程が日々蓄積される電子カルテのテキストのデータは、プライバシー保護の観点から、教育・研究用に公開・共有することができない。これは、昨今の人工知能・自然言語処理技術の発展の中で、医療データの分析が大きく遅れをとる主たる原因となっている。その一方で、そもそも現実の(いわば”汚い”)カルテ文書が持つ特徴は明らかになっていない。  
我々が作成した要素技術は、「読み手の意思決定に有用」という以外の観点でも、任意の目的に応じて、文書に蓄積されている暗黙的な知見を明らかにできることに利点がある。そこで、実際のカルテ文書が持つ特徴を明らかにし、その特徴を備えた擬似的なカルテ文書データセットを公開した。公開カルテ文書データセットとして、文書数は言語を問わず世界最大であり、利用申請が一切不要なカルテ文書データセットは世界で唯一である。なお、実験前に全てのカルテ文書に対して実験用の匿名化処理を行った。

今後の課題としては、以下があげられる。

(a) 文書の有用性の自動予測手法の開発

将来的な課題として、文書が、コミュニケーション上のゴール、たとえば文書が意思決定に有用である、ということ自動予測するための技術開発を行う必要がある。しかし、テキストのいわば表層的な特徴に基づいて、直接、有用かどうかを判断することは困難である。そこで、「有用性」を要素(自動予測の可能性がより高い要素)に分割して、現実的な技術開発課題に落とし込むことを目的として、萌芽的な成果を得ている。

(b) タスクに応じた背景知識の推測と回答傾向の認知特性の解明

集合知フレームワークに応用するために、より基盤的な回答者の認知特性(特に、解答欄および回答時間に関して)の認知特性の解明について、萌芽的な成果を得た(論文投稿中)。研究期間の終了後も研究を継続し、本研究期間で開発した集合知フレームワークの改善に応用する。

全ての研究は、各研究機関における倫理面の審査を経た上で、適切な倫理的配慮のもとに実施された。

[1] Matsubara, M., Kagawa, R., (co-first) et al., Analysis of Usefulness of Critique Documents on Musical Performance: Toward better Instructional Document Format. The proceedings of 23rd International Conference on Asia-Pacific Digital Libraries, 344-353, 2021 [2] Kagawa, R., and Tsurushima, H.: How the Content of Progress Notes Affects Readers' Perceptions of Their Usefulness and the Burden Placed on Writers: Improving Interprofessional Communication, MEDINFO2021 (in press) [3] Kagawa, R., Baba, Y., and Tsurushima, H. A practical and universal framework for generating publicly available medical notes of authentic quality via the power of crowds. The proceedings of 2021 IEEE International Conference on Big Data ,3534-3543, 2021