

佐藤 真一

国立情報学研究所 コンテンツ科学研究系  
教授

## 精神医学×メディア解析技術による心の病の定量化・早期発見と社会サービスの創出

### § 1. 研究成果の概要

精神疾患は罹患率が高く罹病期間も長いこと、人類の生活の質（QOL）を低下せしめるものとして重要な疾患群である。精神科領域は、最も重要な医学領域と言っても過言ではないが、残念ながら現状はこれら精神疾患の克服には程遠い。その原因の一つに、精神疾患におけるその診断や重症度評価に適したバイオマーカーの不足が挙げられる。現状の精神科医の経験に大きく依存する方法は客観性や定量性に乏しく、結果的に診断の不一致、治療開始基準がないことなど、様々な問題につながっている。精神疾患の診断の困難さのもう一つの側面として、精神疾患と正常な精神状態との連続性があげられる。本研究では、前述の精神疾患の診断や重症度評価の困難さに対し、心の状態を可視化・定量化するシステム、いわば「心のバロメーター」として機能するシステムを作る。正常から、未病、さらにうつや認知症等の幅広い精神疾患にわたるスペクトラムを、きめ細かく計測できる技術を実現する。リスクある状況に対しては、ストレスを和らげるアプリの開発、うつ症状を防ぐための認知トレーニングを行うアプリの開発など、未然に精神疾患の発生を抑制するような心の健康ビジネスを創出する。また、精神疾患を患った人や（精神疾患や認知機能低下に関連した）社会的弱者の支援につながるサービスを創出する。2019年度の主たる研究成果を以下に示す。

## 1. 精神症状に精緻に紐づけされたデータ収集と解析

2019年度末までに、患者・健常者のインタビューを通じた音声データ約 480 時間分、SNS による患者テキストデータ約 860 万文字分のデータ収集を完了し、2019年度の目標値(音声データ 350 時間・SNS テキスト 700 万文字)を超えるデータを取得した。

表 1.1 リクルート数およびデータ取得数(2020年3月末時点)

Study	疾患名	同意取得数	データセット数
Study1 インタビューを通じた患者音声解析	うつ病	55	119
	双極性障害	38	94
	不安症	25	47
	統合失調症	80	169
	認知症	48	86
	健常者	45	87
	合計	291	602 (合計約 480 時間)
Study2 SNS を用いた患者テキストデータ解析	うつ病	14	17
	双極性障害	15	18
	不安症	4	6
	統合失調症	5	5
	健常者	10	15
	合計	48	61 (合計約 860 万文字)

収集された音声データに対し文字起こしを行い、一部データに対しては音声・言語学的アノテーションの付与作業を行った。解析の進捗に伴い、アノテーション付与基準の精密化を行い、過去のアノテーションについても修正作業を行った。これらデータを用い各疾患についておおむね 7 割～8 割前後程度の推定性能が得られているが、疾患によりサンプル数と性能にばらつきがあること、また絶対的なデータ量の不足から、必ずしも安定した結果ではなく、さらにデータの追加と精密化を行っていく。

## 2. 分布外データの高精度検出:対象外データを混在するデータの機械学習

深層学習は画像認識、自然言語処理などの特定課題で人をも上回る性能を示し、様々な現実の課題に適用されるに至っている。実応用において、深層学習に基づく認識器が安定して高い性能を示すためには、入力データはトレーニング時のデータと同じ性質を持つデータである必要がある。すなわち、対象としないデータが入力された場合など、トレーニングデータから大きく外れた分布の入力に対しては、望ましい応答を返さない。このため、それらの対象外データを検出する必要がある。

本研究では、対象外データを検出するための新しい手法を提案した。その手法では、二つの分類器を最終段に有するネットワーク構造を示し、対象内データに対しては、分類器の出力が類似し、対象外データに対しては異なるという傾向を促進するアルゴリズムを提案した。異なるデータセットを混在させたデータでの評価により、提案手法は、state-of-the-art を大きくしのぐ分離性能を出

すことを示した。また、食事か食事でないか、服か服でないかというような実应用到に近いセッティングに対して、ほぼ 100% の分離性能を示すことを明らかとした。

本成果は、実際の混在するデータのクリーンアップに用いることで、よりよいデータ整備への活用が期待される。

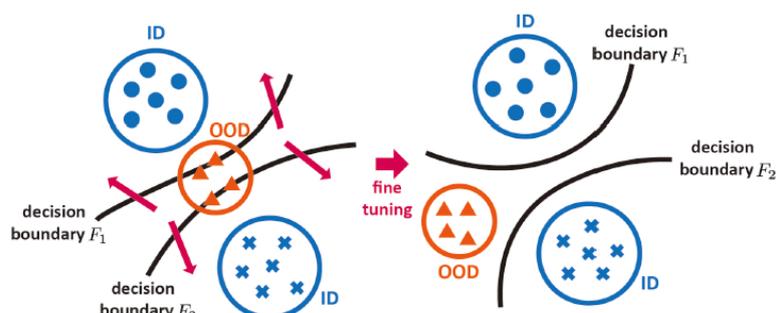


図 1.1 分布外データの検出手法

#### 【代表的な原著論文】

1. Masahito Sakishita, Taishiro Kishimoto, Akiho Takinami, Yoko Eguchi and Yoshinobu Kano. Large-scale Dialog Corpus Towards Automatic Mental Disease Diagnosis. Precision Health and Medicine – A Digital Revolution in Healthcare, Studies in Computational Intelligence, volume 843, pp.111-118. Springer. 2019/8/24
2. Qing Yu, Kiyoharu Aizawa, Unsupervised Out-of-Distribution Detection by Maximum Classifier Discrepancy, Int. Conf. Computer Vision (ICCV) 2019, pp.9518-9526, Oct.27-Nov.2, 2019, Seoul, Korea

## § 2. 研究実施体制

### (1) 佐藤グループ

- ① 研究代表者: 佐藤真一 (国立情報学研究所コンテンツ科学研究系 教授)
- ② 研究項目
  - ・メディア解析技術を応用した多面的な精神疾患診断技術の開発

### (2) 岸本グループ

- ① 主たる共同研究者: 岸本泰士郎 (慶應義塾大学医学部 専任講師)
- ② 研究項目
  - ・精神症状、心理状態に精緻に紐づけられたデータ収集および解析支援

### (3) 相澤グループ

- ① 主たる共同研究者: 相澤清晴 (東京大学大学院情報理工学系研究科 教授)
- ② 研究項目
  - ・精神・心理・認知・フレイル状態の把握のためのライフログの収集、解析

### (4) 山崎グループ

- ① 主たる共同研究者: 山崎俊彦 (東京大学大学院情報理工学系研究科 准教授)
- ② 研究項目
  - ・精神状態推定のためのマルチモーダル処理技術の研究とデータセット整備

### (5) 狩野グループ

- ① 主たる共同研究者: 狩野芳伸 (静岡大学大学院情報学領域 准教授)
- ② 研究項目
  - ・自然言語処理による音声・テキストからの精神状態自動推測