

プロジェクト名：  
申請者名：藤巻晴葵

## 概要

単一のハードウェアでユーザーの対面、電子的なコミュニケーションを理解し、ユーザーの継続的なコミュニケーション、創出を活性化させるソフトウェア

# 1. 何を作るのか

## 背景

指標の作成と計測は、人間の行動を促す上で非常に重要である。これは目標設定理論やフィードバックの原則に基づき、具体的な数値目標が行動のモチベーションを高めることを多くの研究が示しているからである。例えば、血圧や血糖値といった健康状態を表す指標は、健康診断を通じて一般的に利用され、病気の早期発見や個人の健康意識の向上に寄与している。

健康状態を示す指標には、身体的指標と心理的、社会的指標の二種類が存在する。身体的指標には体重、血圧、コレステロール値などが含まれ、これらは容易に測定し追跡可能である。一方、社会的な健康を示す指標は、友人や家族との関係性やコミュニティ活動への参加度合いを含むが、これらは測定が難しく、しばしば見過ごされがちである。

その背景には、近年、Apple Watchをはじめとするヘルスケアデバイスの利用者が増加傾向にある。これらのデバイスは主に心拍数や運動量などの身体的な指標を記録し、自己管理を促す目的で利用されている。しかし、心理的、社会的な健康状態の計測に関しては、これらのデバイスによる対応はまだ十分ではない。心理的、社会的な健康状態は、身体的な健康状態と比較して目に見えにくく、自己認識が難しいという特徴があるため、この点での支援は特に重要である。

この問題を解決するために、心理的、社会的な健康状態を計測することができる「SocialHealthシステム」を提案する。このシステムは、ユーザーの社会的関わりを数値化し、把握可能にするものである。具体的には、ユーザーを取り巻く人々との対面または電子的な様々なコミュニケーションを一つのデバイスで記録し、理解することにより、社会的な健康状態を把握する。これにより、ユーザーは自身の社会的健康状態を視覚的に確認し、改善のための行動を促されるだろう。さらに、このシステムは、ユーザーに対してさらに社会的な関わりを深めるための具体的な提案を行う。これには、友人や家族とのコミュニケーションの頻度を増やす、地域のイベントに参加する、趣味のグループに加わるなどの活動が含まれる。これらの提案は、ユーザーがより活動的で社会的に健全なライフスタイルを送るための重要な手段となるだろう。

## 提案するもの

本システムは、ユーザーの対面、電子でのやり取りを把握、記録し、自身を取り巻く人との関係性を可視化し、実績などのゲームのような仕組みを導入し、コミュニケーションを楽しめるようにする。提案するシステムの大きな機能は以下の3つである。

- 単一のハードウェアでのユーザーの対面、電子的なコミュニケーションの記録
- コミュニケーション情報から特徴を抽出し、ユーザーを取り巻く人たちとの関係性を可視化
- ユーザーに合わせた適切な目標の設定、アドバイスによって、ユーザーの社会的な関わりを活性化させる

上記の機能について順に説明する。

### 1. 単一のハードウェアでのユーザーの対面、電子的なコミュニケーションの記録

ユーザーに対して、詳細に自身の社会的健康状態を提示するためにはより詳細にコミュニケーション情報を記録する必要がある。現代において、コミュニケーションの手段は対面だけでない。PCやスマートフォンを利用したテキストチャットやビデオ通話などの電子的なコミュニケーションも対面と同等、それ以上に行われている。そのため、ユーザの社会的な関わりを理解するためには対面でのコミュニケーションだけでなく、電子的なコミュニケーションも記録可能な仕組みを作成する必要がある。

そこで私は、ユーザーの対面、電子的なコミュニケーションを単一のハードウェアで記録する仕組みを開発する。単一のハードウェアで対面と電子的なコミュニケーションの両方を記録を可能にすることによって、ユーザーは各デバイスに対して追加でソフトウェアをインストールをするという煩雑な作業をせずに記録することが可能になる。また、今後Apple Vision Proなどをはじめとするヘッドセットの利用が普及した際にMRシステムとしての移植提供が容易になる。

以下に対面でのコミュニケーション、電子的なコミュニケーションを記録する方法の二つを示す。

#### 対面でのコミュニケーションの記録

私は対面でのコミュニケーションを記録する上で、過去の会話情報との繋がりがわかるような形で記録を行いたいと考えている。つまり、単にある時の会話として扱うのではなく、誰といったどのような会話をしたのかということまで記録したい。この目標を達成するためには、ハードウェアから得る視覚情報、音声情報をもとに話者ごとに顔や音声を分離し、顔画像特徴、音声特徴から一意のIDを付与させ、画像と音声でIDを一致させる必要がある。顔画像、音声に対して一意のID付与ができると、各対面での会話情報に対してユーザーが操作を行うことなく、同一人物との会話情報を蓄積することが可能となる。ここで、視覚情報と音声情報の処理をそれぞれどのように実装するかを説明する。

まず、視覚情報の処理では、カメラに映る顔の検出とその顔の識別が必要である。検出ではyolo v8のモデルを利用したface detectionモデルが公開されており、初期段階ではそれを用いて実装を行う。また識別に関しては、FaceNetをはじめとする顔画像に対して特徴をベクトルとして扱い顔の識別を行う研究が行われている。ArcFaceは顔の識別タスクに対して比較的高い精度を出しており、初期段階ではそれを用いて実装を行う。

次に、音声情報の処理についてだが、録音された音声に対してどこからどの部分までが誰の音声なのかを分離、識別を行い、その後音声認識によって文字情報に変換する必要がある。まず、音声分離についてはいくつか研究が行われており、pyannoteはPyTorchベースの話者分離を行うツールキットで、これを用いることによって少ない工数で話者分離を実装することができる。また、音声の識別についてはYoungmoon Jungらの研究など音声を識別するための深層学習モデルで一定の成果を得ている。音声から文字に変換する音声認識についてはOpenAI社のWhisperや株式会社レアゾン・ホールディングスによるReasonSpeechなどが挙げられる。そして、顔画像と音声の紐付けは顔画像の検出と同時に発話しているかどうかを判定し、検出された音声のタイミングとマッチングさせる。

## 電子的なコミュニケーションの記録

電子的なコミュニケーションの記録では、自身の発言と他人の発言を認識し、会話の繋がりを検出することが重要である。また、本システムではデバイスに対してスクリーンキャプチャをすることなく、自身の体につけた単一のウェアラブルデバイスによって記録したい。この目標を達成するために、ハードウェアから得る視覚情報からPC、スマホの画面領域の抽出を行う。そして、得た画面領域は解析を行い、発言者と内容を認識し、やり取りの繋がりを検出し、記録する。ここで、具体的な実装方法について説明する。まず、視覚情報からPC、スマホの画面領域の検出についてだが、yolo v8などの物体検出モデルを利用して画面領域部分の抽出を行う。次に抽出された画面領域の解析についてはyolo v8などの一般的な物体検出モデルを用いて行うことも可能だが、画像情報とレイアウト、文字情報を入力として用いてより精度の高い検出を行いたい。そこで、LayoutLMをはじめとするマルチモーダル言語モデルを用いた実装を行うことを目標とする。

## 2. コミュニケーション情報から特徴を抽出し、ユーザーを取り巻く人たちとの関係性を可視化

コミュニケーションの時間などの表面的な部分だけでは、ユーザーの社会的な関わりを深く理解することはできない。コミュニケーションの内容も分析の対象とすることによって、ユーザーがどのような社会的な関わりをしているのかを具体的に理解することが可能となる。提案するシステムでは、ユーザーがコミュニケーションを行う人に関する年齢や立場関係、業務的な会話が多いのか、プライベート的な会話が多いかなどの複数の特徴を抽出し、ユーザーを取り巻く一人一人の関係性や特徴を言語化したい。この目標を達成するために、いくつかのアプローチを取りたいと考えている。まず一つ目として、固定指標の分析である。例として、会話の内容がどのようなジャンルであるか、自身についての話なのか、ニュースに関する話であるのかを判定する。このようなある程度定式化できる判定については、distil-BERTなどの比較的軽量の言語モデルを利用し、分類タスクとして実装する。二つ目のアプローチとして会話と人物についての動的特徴の抽出である。例として、会話の中で、「私は〇〇が好き」という発言

があったとしたら、その発言した人物が好きなものは何なのかを記録したい。これは一つ目の固定指標と比べ、単純な分類タスクに落とし込むことは難しい。私は会話情報を元に会話と人物における知識グラフを構築する。会話特徴、人物特徴をグラフ構造で表現することで視覚的にわかりやすい形で抽象化できる。知識グラフの構築は言語モデルを用いてノードの生成、エッジの生成を行い構築する。これはIgor Melnykの手法を参考にしている。それぞれのアプローチは実際に本システムで記録した会話データや国立国語研究所が公開している日本語日常会話コーパスなどのデータセットを利用し検証を行う。加えて三つ目のアプローチとして、会話が行われている状況を記録したい。そこで、会話が行われた時にハードウェアから得た視覚情報を説明する文章を作成する。ここではBLIP-2などの深層学習モデルを活用して状況を説明する文章と知識グラフを作成し、記録する。

以下に機能の要件を示す。

- 記録した情報からどのような状況での行われたコミュニケーションかを記録
- 会話内容から相手の年齢や関係性と親密度を認識する
- ユーザーのコミュニケーションを合計し、どのようなコミュニケーションを多くするのかどのような人とよく関わる傾向があるのかを分析しレポートする

### 3. ユーザーに合わせた適切な目標の設定、アドバイスによって、ユーザーの社会的な関わりを活性化させる

ユーザーがより円滑に社会的な関わりを持つためには現状のコミュニケーション状況を可視化するだけでなく、目標を設定したり、アドバイスを行うことによってユーザーの積極的な改善行動を促すことができる。以下に機能の要件を示す。

- ユーザーのコミュニケーションに偏りがでないように具体的な目標を設定し、達成できたかを検出する
- ユーザーがより人と深い関係性を持つことのできる機会があれば、その行動を促すようにアドバイスを行う

## 2. どんな出し方を考えているか

私は、このシステムをまずはスマートフォンのアプリケーションで利用できる形として公開したいと考えている。このシステムの売りは単一のハードウェアで対面、電子的なやりとりを記録することによってユーザーの記録するための煩雑さをなくしている。そのため、システムもスマートフォンで利用できるような形にし、ユーザーの利用ハードルを低くしたい。また、私は今後ヘッドセットなどを利用したMRシステムの普及が加速すると考えており、Oculus Quest 3やXREAL Air 2 Ultra, Apple Vision Proなどの利用を想定した、専用のハードを利用せずともコミュニケーションを記録できるシステムとしての提供も行いたいと考えている。

また、このシステムは個人での利用のみならず、企業単位での利用にも発展できると考えている。例えば、接客業や営業においての客との関係性や特徴を記録することによるコンバージョン増加の期待や社内でのコミュニケーションを可視化することによる組織で発生しうる問題の早期発見と予防を期待することができる。本システムをそのままビジネス利用することは難しいが、基礎技術の確立として開発することはとても意義のあるものだと考えられる。

### 3. 新鮮さの主張、期待される効果など

#### 3.1. 既存サービスとの比較

- 既存の社会的な関わりを記録するものは対面で顔の表情や姿勢などを記録し測定したり、単一の電子的なコミュニケーションツールに閉じた解析など、限られた状況下でのコミュニケーションの表面のみを記録するというものが多い。

Sony Lifelogアプリ

スマートバンドを装着することによって

顔数計 一人称ライフログ映像からの顔検出に基づいた社会活動量計

2018年の末踏で採択されたプロジェクト。首元に装着したカメラに映った相手の顔の映り込みの変化を記録し、そのデータを社会活動量として記録するシステム。このシステムは対面での人との関わりを計測するものであり、電子的なやり取りは記録の対象とされていない。また、記録するのは顔の映り込みのみで会話の内容を記録しない。

#### 3.2. 技術面での斬新さ

本システムは以下の3点で技術的に優位性を持っていると考えられる。

- 単一のハードウェアで対面だけでなく、電子的なコミュニケーションも記録できる
- コミュニケーションから関係性や相手の特徴を抽出する
- 膨大なライフログデータを活用できるようなデータ管理

#### 3.3. 期待される効果など

本システムを利用することで以下の3点が効果として期待できる。

- 自身と他人との関係性を可視化することによって、自身の社会関与度を定期的にふりかえり自身の健康状態を理解することができる
- 自分のコミュニケーションの内容まで入り込んだ記録を可能にすることによって、高度な行動提案を可能にすることができる
- 組織内で利用することによって組織のコミュニケーションを可視化することができ、活動を活発化させる

## 4. 具体的な進め方と予算

### 主な開発場所

筑波大学春日エリア 加藤研究室

### 使用する計算機環境

MacBook Pro, GCP, etc.

### 使用する言語・ツール

Python, JavaScript

### 開発計画

### 予算内訳

#### 人件費

まず、提案者である藤巻は1440 [時間] x 2000 [円]=2,880,000 [円]

#### 必要経費

## 5. 提案者の腕前を証明できるもの

### 藤巻晴葵

現在に至るまで、授業、研究、ハッカソン・コンテスト、個人開発、会社で数十のシステムの開発を行ってきた。基本的にはソフトウェアの開発をメインとして行ってきたが、ハードウェアの開発に関してもいくつか作成してきた。また、高専4年次に会社を創業し、その会社での活動がいくつかのメディアで取り上げられている。以下にこれまでの成果を示す。

### 2019年度 全国高専プログラミングコンテスト 課題部門

- 最優秀賞・文部科学大臣賞
- :::doc (てんどっく)
  - 自動点字墨字相互翻訳システム

## 2020年度 全国高専プログラミングコンテスト 課題部門

- 最優秀賞・文部科学大臣賞
- ぷらんとこれくしょん
  - 小学生の観察学習での利用を目的としたアプリケーション
  - 児童たちが植物や昆虫を撮影し、名前を自動で判別しマップにプロットされ、季節ごとや他学校とのデータを利用し、植物や昆虫の生息場所などを比較しながら学習することができる。

## 2020年度 高専ディープラーニングコンテスト 2020

- 最優秀賞
- :::doc (てんどっく)
  - toB向けに自動点字翻訳システムを開発し、情報の発信者が手軽に文書を点字などの視覚障害者が読める形に変換し提供することができる。

## 2022年度 情報処理学会 コンシューマ・デバイス&システム (CDS) 研究会 CDS36

- 優秀発表賞
- レイアウト付き文書に対応したクラウド型点字翻訳システムの実用化と深層学習による半自動化

## 2022年度 総務省 異能vation

- 異能ジェネレーションアワード受賞

## その他の制作物

- ワンタイムQRコードによる不正防止モバイルスタンプラリーシステム
- 画像認識による三次元的な駅構内の位置特定とARによる案内システム
- 紙媒体、デジタルカレンダーの情報を相互に共有できるシステム
- 画像認識を用いた、簡単に利用できる廃棄野菜のオークションシステム
- BLEビーコンを用いた、視覚障害者への駅構内に関するリアルタイム情報提供システム

## 6. プロジェクト遂行にあたっての特記事項

現在、藤巻は筑波大学情報学群知識情報・図書館学類3年に所属しており、来年度4年に進学する。藤巻は大学の授業をほとんど取り切っており、4年では卒業研究がメインとなる。研究室の指導教員から、未踏の応募について容認されており、プロジェクト遂行に当たって問題になる可能性は極めて低い。

## 7. ソフトウェア作成以外の勉強、特技、生活、趣味など

### 藤巻晴葵

趣味はいくつかあるが、主に二つある。一つ目はサイクリングである。筑波大学へ編入する前の高専での在学中は家から高専まで、アップダウンのある片道10kmをクロスバイクで登校していた。そのため、自身の中でのクロスバイクでの移動は普通であり、体を動かしたくなった際には100km未満の距離のサイクリングをよく行なっている。二つ目はアニメ、ドラマ、映画などのPGC系動画コンテンツの視聴である。特にアニメは毎シーズン必ず何本かのアニメを見るようにしている。好きな内容の傾向としては、バトル系というよりも、心理描写がしっかりと描かれているヒューマンドラマなどを好みとしている。

また、大学のサークルなどを通してイベントの運営などを行なっている。TEDxUTsukubaという学生団体では、毎年100名規模のイベントを開催している。このイベントではワークショップや6名のスピーカーによる登壇セッションなどがあり、自身は協賛していただけるパートナーとの交渉や、翻訳システムの開発に携わった。また、エンジニアを対象としたミートアップ運営に携わっている。イベントを通じて、様々な背景を持った人たちと関わることに喜びを感じている。

## 8. 将来のITについて思うこと・期すること

近年、LLMが話題になったことによって、エンジニアの業務を改善するためのツールが多くリリースされ、普及も進んでいる。例として、CursorやGithub Copilotがあげられる。どちらもコーディングを支援するツールである。今までコードエディターに搭載されていたルールベースのコード補完とは違い、先ほどのツールはプログラマーの意図を高度に先読みし、ほとんどのコードを補完してくれる。私はこのツールの中でもGithub Copilotを利用しているが、体感として、自分がコードを書くという作業が1/2になったのではないかと感じている。それほどに強力なツールである。ここで一つ思うことがあるかもしれない。「このツールが発展していく、AIが発展していくことでエンジニアがいらなくなるのではないか」と。私は、エンジニアという存在がなくなることはないと感じている。一方、AIの進化によって、より開発することのコストは小さくなっていくと容易に想像がつく。今、あるサービスを開発するのに1年かかったとして、10年後には1週間、1日、1時間で作れるようになると考えられる。だが、考えてみてほしい。プログラミングに関しては、過去10年間の間で新しいプログラミング言語やフレームワーク、ライブラリの登場によってどんどんサービス開発等のコストが下がっていることに。しかし、今もプログラマーという需要は大きい。結論として、技術的なパラダイムシフトがあっても本質的に人間が行う活動というのはなくならないのだ。確かにプログラマーという存在はなくなるかもしれない。だが、人間が思考するというのは自分のクローンが作られるまで続く。エンジニアという存在も時の流れによって、当時やっていたことをしなくなり、また違う部分に焦点を当てて活動するようになる。私は将来のITは日々新しいことに取り組み、それぞれが常に誰かに影響を及ぼすよりカオスな状態となると考える。