



擬人化エージェントによる説得に関する研究

川田, 恵

(Degree)

博士 (学術)

(Date of Degree)

2021-03-25

(Date of Publication)

2022-03-01

(Resource Type)

doctoral thesis

(Report Number)

甲第7964号

(URL)

<https://hdl.handle.net/20.500.14094/D1007964>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



博士論文

擬人化エージェントによる説得に関する研究

令和3年1月

神戸大学大学院国際文化学研究科

川田恵

博士論文

擬人化エージェントによる説得に関する研究

所属専攻：グローバル文化専攻

氏名：川田 恵

指導教員氏名：村尾 元

要旨

本研究は、相手に言いにくいことを言わなければならないなど、円滑なコミュニケーションが困難な状況において、これを可能にするように支援を行うコンピュータ・システムについて研究を行う。本研究で提案するシステムは主として、職場や研究室などの場面において利用される事を想定する。これらの場面では、お互いに関係が浅い人とコミュニケーションを取ることは珍しくない。この研究では、そうした関係性での説得という状況を想定している。説得には、説得する側と説得される側の両方に心理的な負担が生じるという問題がある。それに対し、近年、擬人化エージェントを用いて、説得を行うという手法に大きな期待が寄せられている。なかでも、被説得者を直接説得するのではなく、説得者が第三者を説得している様子を被説得者に見せることで間接的に説得するという、オーバーハードコミュニケーションを取り入れた擬人化エージェントの説得システムが効果的であることが報告されている。

本論文では、人間同士の説得というコミュニケーションに介在する擬人化エージェントの必要な要素とその実現技術について研究を進める。従来の研究により、擬人化エージェントが、ユーザに説得を促し意見や態度を変容させるには、ユーザが、エージェントの振る舞いや作業に対する認知性を持つこと、そのために、エージェントにアイコン的なジェスチャをさせる身体性を持たせることが必要である。また、エージェントが知性を備えてユーザとのインタラクションに対応するための社会性や適応性を持たせること、ユーザが

エージェントに対して親密性を抱くためのエージェントの擬人性やエージェントが自律的な活動を行う自律性などの機能が要求されると言われている。本論文では、説得という状況を想定し、ユーザがエージェントに対して擬人性と親密性を感じることを目的とし、それを実現するための2つの機能に関する方法を述べる。

それでは、この説得コミュニケーションに介在する擬人化エージェントを実現するために至要となる要素として、2つのものがある。第一に、ユーザがエージェントに対して、親密性や信頼性を持つこと、第二に、一方的な説得を防ぎ、エージェントがユーザの暗黙的な動作や状況を察知して、それに合わせたリアクションで対応をするための第一歩として、ユーザの動作や状況を認識理解することである。本論文では、第3章で第一の課題、すなわち擬人化エージェントの容姿やコミュニケーション形式について、第4章では、第二の課題、すなわち、擬人化エージェントがユーザの動作を認識理解することに取り組む。

各章の内容は次の通りである。まず、第1章ではコミュニケーションにおける問題と機械による解決策、および先行研究に対する本研究の意義について論じる。第2章では、まず、擬人化エージェントを導入しようとしている試みについての現状とこれらの課題について論ずる。擬人化エージェントによるユーザへの影響力については、題材とするシチュエーションやエージェントの数、エージェントが発する言葉によってユーザの態度の変容に対して影響力に違いがあることが報告されており、擬人化エージェントが説得を施す際の有力なコミュニケーションの形はまだ十分に確立されていないのが現状である。

従来の研究では、ユーザを説得する際に最も影響力のあったエージェントの組み合わせは、男性のエージェントを説得者、女性のエージェントを被説得者として、男性のエージェントが女性のエージェントを説得する場合であった。しかし、彼らの研究ではオーバーハードコミュニケーションについて詳しい言及がされていない。そこで本研究では、男性と女性の擬人化エージェント2体の組み合わせによるオーバーハードコミュニケーションがユーザに対して説得力を持つと仮定し実験を行った。また、実験には、エージェントの髪型や服装、身に着けるアイテムに違いを持たせた、男性のエージェント4体と女性のエ

エージェント2体を用いた。その際、先行研究に倣って、本研究も映画の鑑賞意欲への影響力に関して調査を行う。また、その実験から得ることができた情報を元に、さらに擬人化エージェントの容姿に重点をあて、再び実験を行い、メガネをかけた男性エージェントが映画広告とともに表示された場合に鑑賞意欲の平均値が最も高かった。

第4章では、エージェントがユーザの行動を認識する技術に関して、先行研究に基づいて新しいニューラルネットワークをとして、複数の畳み込みニューラルネットワーク（CNN）を持つニューラルネットワークを提案する。筆者はまず、研究室でRGBカメラを用いて撮影し、「コンピュータの使用」「スマートフォンの使用」「友人との会話」「居眠り」4つの行動について動画像データを取得した。次に、取得した動画像データを3fpsの画像に変換し、それらのデータを用いて2つのデータセットを作成する。1つは、画像から人間領域のみを切り出したデータであり、もう1つは、その人間領域を抽出しさらに姿勢情報を可視化させたデータである。これら2つのデータセットは、どちらも学習モデルを作成するための学習データとモデルの認識精度を評価するためのテストデータに分ける。実験の結果、人間領域のみを切り出したデータを用いて、5つのCNNから構成されるネットワークを学習した場合の成績が最も高く、その認識精度は97.5%であった。しかし、汎用性について比較を行った場合には、人間領域を抽出しさらに姿勢情報を可視化させたデータを用いて、5つのCNNから構成されるネットワークを学習した場合の方がより高い精度を示した。ここから、研究室内という限られた空間で様々な人間が行う生活行動を認識する際は、姿勢情報を可視化させたデータを用いてこのニューラルネットワークで学習させることが有効であると考えられる。

最後に、第5章では、本論文を踏まえた結論について述べる。結論では、説得コミュニケーションに介在する擬人化エージェントを実現するために至要となる2つの機能について、それぞれの知見を得ることができた。まず、1つ目については、説得に適した擬人化エージェントのコミュニケーション形式は、オーバーヘッドコミュニケーションの形式が適していること、また、説得エージェントの違いがユーザの意見変容に影響を与えていることがわかった。擬人化エージェントがユーザを説得する時、ユーザが擬人化エージェントに

対して信頼を置くことが重要であると考え。擬人化エージェントの容姿の違いだけではなく容姿に適したダイアログを提示することで、ユーザは擬人化エージェントに対して信頼性を高め、擬人化エージェントは、ユーザの意見変容に影響を与える。2 つ目については、エージェントがユーザの動作を認識理解する技術についても、ユーザの状態を高い性能で認識できるニューラルネットワークを提案し、実現することができた。

目次

1. 序論.....	9
2 先行研究の整理と本研究の意義.....	11
3. 擬人化エージェントの容姿とコミュニケーション形式.....	17
3.1 関連研究の概要.....	17
3.2 実験①の設定.....	24
3.2.1 実験①の手順.....	24
3.2.2 エージェントの設定.....	26
3.2.3 映画の広告の設定.....	27
3.2.4 ダイアログの設定.....	28
3.3 実験①.....	29
3.3.1 実験①の参加者.....	29
3.3.2 実験①の手順.....	29
3.4 実験①の結果.....	31
3.4.1 説得に適したエージェントパターンについて.....	31
3.4.2 RC と OHC の比較.....	34
3.4.3 OHC の場合の説得エージェントと被説得エージェントの比較.....	36
3.5 実験①の考察.....	37
3.6 擬人化エージェントの要素に関する実験の概要.....	38
3.7 実験②の設定.....	39
3.7.1 実験②の実験用 Web ページ.....	39
3.7.2 エージェントの設定.....	40

3.8 実験②の概要.....	40
3.9 実験②の結果.....	41
3.9.1 映画の広告パターン別比較.....	41
3.9.2 メガネの有無比較.....	42
3.10 まとめ.....	42
4. ディープラーニングを用いた人間の状態認識システム.....	44
4.1 概要.....	44
4.2 関連研究.....	45
4.3 本研究に用いるニューラルネットワーク.....	48
4.3.1 ニューラルネットワークの構成.....	48
4.3.2 ニューラルネットワークの実装.....	49
4.3.3 入力層.....	50
4.3.4 畳み込み層.....	50
4.3.5 プーリング層.....	51
4.3.6 全結合層.....	51
4.3.7 出力層.....	52
4.4 データセット.....	53
4.4.1 動画像データの取得.....	53
4.4.2 動画像データの変換.....	54
4.4.3 データセットの作成.....	54
4.4.4 Detect dataset.....	55
4.4.5 Pose dataset.....	55
4.5 実験.....	57
4.5.1 実験概要.....	57

4.5.2 実験方法	57
4.6 結果.....	58
4.6.1 評価方法	58
4.6.2 被験者 3 人のテストデータに対する結果.....	59
4.6.3 未学習の被験者データに対する推定結果.....	61
4.7 まとめ	62
5. 結論.....	64
付録 A.....	66
付録 B.....	70
謝辞.....	73
業績一覧.....	74
参考文献.....	75

1. 序論

充実した社会生活を送るにあたり、人間関係を他人と築くことは極めて重要なことである。日常生活で感じる喜怒哀楽は、他人と関わっていく中で生まれるものが多く、人間関係が円滑に行われていることで、自分の仕事に対して集中することができたり、学校生活が楽しくなったりと多くの利点がある。こうした中で、他人と円滑なコミュニケーションをとり、人間関係を築くということは人生における重要なテーマである。

良好な人間関係や他人とのコミュニケーションを円滑にするには、コミュニケーションに対する知識や空気を読むといったスキルが必要である。しかし、コミュニケーションにおけるスキルには個人差があり、他人の捉え方も多様に存在する。特に相手の気分を害してしまう可能性があるコミュニケーションは、例えば、他人を直接注意したり、他人の行動に対して批判や指摘をしたりすることである。これらのコミュニケーションには、相手の気分を害さずに自分の意見を伝えるという戦略やコミュニケーションスキルが必要となってくる。しかし、他人の捉え方が多様なことによって、戦略やスキルに明確な正解は今のところ存在しない。

そこで、本研究は、人間同士のコミュニケーションにコンピュータが介入することによって、難しいコミュニケーション、ここでは主に説得という状況を題材として、問題点を解決するためのシステム構築を目的とする。私たち人間は、生きていく上で社会や組織に属しており、その組織内における人と人とのインタラクション、コミュニケーション、人間関係を築くことは不可避的である。コミュニケーションの場には私的なものと公的なものがある。それらの特徴を区別すれば、前者はお互いに対する知識や理解の程度、またお互いでの価値観や文脈などの共有の度合いが大きく、それに対して、後者では、それらの程度や度合いは小さい。本研究では、主として、後者の公的なコミュニケーションの場を想定する。

公的なコミュニケーションの場とは、様々な集団が想定されるが、本論文では、職場や研究室、作業空間といった場面に限定する。この場面でのコミュニケーションは、お互いに

関係が浅い人とコミュニケーションを取ることは珍しくない。また、お互いに年齢が違ったり、お互いの状況がわからなかったり、コミュニケーションを取るにあたり障害が生じる可能性がある。このような人間関係では、こちらと予期する相手の行動と実際に相手が行う行動との相違が生じることも多々ある。このように、相手の実際の行動がこちらの予期していたものとは異なっていたとき、その行動が自分やその場に居合わせる他の人々にとって、不愉快なものであった場合に、相手にその行動を控えさせることが必要である。すなわち、相手に説得を促し、行動を変容させる必要がある。しかし、この説得におけるコミュニケーションには、説得する側と説得される側の両方に心理的な負担が生じる。説得する側には、自分の前提知識や価値観を相手に不用意に押し付けてしまうのではないかという不安や相手との関係が悪化するのではないかという恐れが生じる。また、説得される側には、相手の説得を受け入れることに伴う不満がある。日常生活において、こうした心理的負担が妨げとなって、説得を促すという行為が困難である場面にしばしば遭遇することがある。そこで、本研究は、このような状況にコンピュータシステムを導入することで解決することを目指す。

近年、人間と機械のコミュニケーションは近年、急速にインタラクティブ性のあるものへと変化している。最近、多くみられるものとして、Webサイトにナビゲーション役として擬人化エージェントを導入する情報技術が挙げられる。こうした状況の中、機械による情報提供がユーザに与える影響について関心が高まってきている。中でも、擬人化エージェントは、その言動によって、ユーザの態度や意見に変容を促し、ユーザを説得する可能性を持つ。擬人化エージェントを人間同士のコミュニケーションの一員として導入し、説得におけるコミュニケーションの問題に対する解決方法を示唆すると考える。本研究はこうした状況を背景とする。

2 先行研究の整理と本研究の意義

大学生が苦手とする対人コミュニケーションについて研究した後藤[1]らの研究がある。彼らは、大学生 221 名を対象に苦手とする対人コミュニケーションを検討するため、研究を行った。結果として、大学生は初対面の人との会話、年長者との接し方、顔見知り程度の人とのコミュニケーションに苦手意識を抱いていることが明らかになった。大学生に関わらず、人間同士がコミュニケーションを取る際、初対面の人や、年長者、顔見知り程度の人とコミュニケーションを取ることは珍しくない。その際、他者への配慮や、意思表示の方法など、各個人によって表現方法は様々であり、能力に差が生じる場合も多いと考えられる[2]。そうした時、人間同士のコミュニケーションを円滑に行うことができないという問題が生じる[3]。この問題に関しては、近年人間と人間の間には機械を介在させることによって、こうした心理的葛藤の解決を図ろうとする研究がいくつかなされている[4]。本研究もこの分野における研究の一つである。家族や友人関係における説得に伴う心理的な負担は、以下の点において公的なコミュニケーションの場での説得とは異なる。すなわち、私的なコミュニケーションの場においてはお互いに共有するものが多いために、説得における心理的な負担は、公的な場面に比べて小さいか、あるいはそのために極めて複雑なものになる可能性がある。前者の場合であれば機械の介在は不要であると考えられ、後者の場合であれば機械によって解決はその可能性自体がそもそも望めない。本研究が機械による円滑なコミュニケーションを促進するという問題に関して、公的な場面を想定するのはこのためである。

現在のところ、説得という場面に話を限定せず、広く人間同士のコミュニケーションにおける問題を解決する手法の提案ということに着目した場合の様々な研究が報告されている。例えば、中野ら[5]の研究がある。このシステムが利用に利用が想定される場面は、個人のスペースがパーティションで区切られたオフィス環境である。パーティションという壁があると、個人スペースを訪問する際の心理的障壁が生じてしまい、話しかけるという行為が困難であるという問題がある。そこで、中野らはその心理的負担を軽減させ、個人

同士が話しやすい環境を作るためにこのシステムを提案した。具体的には、個人スペースに置かれたマグカップの下に圧力センサを設置し、オフィス全員の飲み物の残量情報を取得しパソコン上のサーバで管理する。サーバでは、全員の残量履歴情報が管理されているため、全員が参照することが可能となっている。そして、サーバから情報を得た訪問者は「コーヒーを注ぎに来た」という口実で個人スペースに訪問することが可能となる。このシステムをオフィスに導入し、検証した結果、個人スペースに訪問する心理的障壁が軽減され、またお互いに「訪問すること」「訪問されること」に対する肯定的な意識を作り出すことができた。このように、話しかけるといふ行為における心理的負担を軽減する研究は他にもある。椎尾ら[6]の研究である。これは、個人が隔離されたオフィスで働く人々の対話を積極的に支援するための研究である。椎井らは、オフィスの休憩スペースに設置されたコーヒーマーカーに着目し、休憩に集まった人々がコーヒーマーカーのスイッチを押すと、コーヒーの香りがオフィス全体に伝達する Meeting Pot というシステムを設計した。このシステムは、コーヒーの抽出と共に、コーヒーのアロマを同時に発生させ、休憩の時間が始まることをオフィスにいる人々の嗅覚を刺激することで伝達するという流れになっている。このように、自分の状態や環境の変化をシステムや媒体に乗せて相手に伝達し、それによって話しかけるといふ行為や話す場を設けることへの心理的負担を軽減させる手法は多く提案されている。このような研究を踏まえると、人間同士のコミュニケーションにシステムを介在させることは、コミュニケーションに対して苦手意識のある人を支援したり、オフィスの環境を改善し、仕事の生産性を高めたりすることに役立っていると言える。しかし、前に紹介した2つの先行研究は主として、コミュニケーションの開始に伴う心理的負担の軽減を目的とするものである。本論文の主題である、相手を説得するという場面を想定とした場合、これらとはまた別の技術が必要となる。

相手に説得をするという場面に関して、オーバーハードコミュニケーションに関する研究は、説得によって相手の行動の変容を促すという点で大きな期待を寄せられている[7]。オーバーハードコミュニケーションによる説得とは、被説得者を直接説得するのではなく、説得者が第三者を説得している様子を被説得者に見せることで間接的に説得する方法であ

る[8]。しかしこれに関して、藤原[9]らによる研究がある。彼らは、擬人化エージェントがユーザを説得する際に、エージェントの数によるユーザへの影響力の違いについて研究を行った。エージェントの数が1体のときは、ユーザに直接説得を行うレギュラーコミュニケーションという形をとり、エージェントの数が2体、3体と複数の場合には、オーバーヘッドコミュニケーションの形をとった。これらのエージェントを用いて、ユーザに与える説得効果を検証した結果、説得成功回数が最も多かったのはエージェントが1体でユーザに対してレギュラーコミュニケーションを行った場合であった。これは、オーバーヘッドコミュニケーションよりもレギュラーコミュニケーションの方がユーザの説得に適しているという結果を示す報告である。一方、齋藤[10]らの擬人化エージェントによるオーバーヘッドコミュニケーションに着目した研究がある。そこでは、擬人化エージェントを利用したオーバーヘッドコミュニケーションによるユーザへの説明や説得が、効果的であるという可能性が示されている。齋藤らは擬人化エージェントを説得者と被説得者の立場に用いて、第三者であるユーザに向けてオーバーヘッドコミュニケーションを行い、被説得エージェントの反応の違いが、ユーザの意見変容とエージェントの印象形成に与える変化について検討した。その結果、被説得エージェントがユーザの意見に同意した場合は、説得エージェントと共にユーザにより大きな信頼感を与えることが明らかになり、擬人化エージェントを用いることで人間の説得や説明を有効的に行うことができるという可能性が示された。その他にも、ユーザを説得する擬人化エージェントとして適した容姿やテキスト、単数で説得する場合と複数で説得する場合のユーザへの影響を比較し検討した Kantharaju ら[11]の研究がある。彼らは、擬人化エージェントの数、性別、発言内容や服装、ジェスチャの違いがユーザに与える影響について研究を行った。彼らは、映画を題材にして、男性2体と女性2体の擬人化エージェントを準備し、ユーザに対して映画を観るよう説得を行った。ユーザには事前にその映画についてどれほど観たいかを調査しておき、その後、擬人化エージェントの説得によってどのように映画への鑑賞意欲が変化するかということと、擬人化エージェントへの信頼度を調査した。その結果、擬人化エージェントが複数である場合に、また、男性と女性の擬人化エージェントが2体であった場

合にユーザに影響力があつた。その中で、最も影響があつたのが、男性の説得エージェントが女性の説得エージェントを映画に対して鑑賞意欲を高めるような説得をしたパターンである。このような結果が出たが、この研究は、擬人化エージェントが単数の場合と複数の場合のどちらがユーザに影響力があるかを調査した実験であるため、オーバーヘッドコミュニケーションに関する詳しい言及はされていない。とはいえ、単数のエージェントよりも複数のエージェントによる説得の効果を示しているという点で、それはオーバーヘッドコミュニケーションの有効性を支持する研究であると言えよう。このように擬人化エージェントがユーザへの説得を行う研究はいくつか報告されており、これらの研究を踏まえると、擬人化エージェントを用いることで、ユーザに説得を促すことは可能であることが示されている。しかし、題材とするシチュエーションやエージェントの数、エージェントが発する言葉によってユーザの態度の変容に対する影響に違いがあり、擬人化エージェントが説得を施す有力なコミュニケーションの形はまだ十分に確立されていないのが現状である。

そこで、本研究では、人間同士の説得というコミュニケーションに介在する擬人化エージェントの必要な要素について研究を進める。まず、説得コミュニケーションの必要な要素として信頼、好感、情熱、共感、論理などがあげられる。これらの要素を取り入れて、説得コミュニケーションを行う時、相手からの信頼や好感を得るために人間関係を構築し、論理的に説得を行うためにコミュニケーションスキルを学ぶ必要がある。そのため、時間や訓練を要す。そこで、擬人化エージェントが人間の代わりに説得コミュニケーションのスキルの能力を身につけることで、人間が時間や訓練をせずとも、意見変容が起りやすい環境を作ることができると考えられる。説得を促し、意見や態度を変容させる能力のある擬人化エージェントを設計する場合に必要な要素はいくつか挙げられる。間瀬[12]によれば、エージェントに対する認知性やエージェントの操作性、エージェントの行動と社会性、適応性、親密性、自律性などの機能が要求されるといわれている。

その中でも説得という状況を考えれば、ユーザが擬人化エージェントの容姿から親密性を感じることで、ユーザが擬人化エージェントに適応し、その擬人化エージェントとのイ

インタラクティブなコミュニケーションが実現すると考える。すなわち、ユーザは擬人化エージェントに対して、親密性や友好的な印象を持つことで、擬人化エージェントによる意見変容の促しを受け入れることが可能となる。また、擬人化エージェントが一方的にユーザに対して説得を行うことはユーザが不快感を抱く可能性があるため、ユーザのリアクションに対応した説得を行う適応性の機能を実現することが必要であると考え[13]。

これにより、説得力のある擬人化エージェントを設計するにあたり、具体的に必要な要素は2つあると考える。1点目は、ユーザがエージェントの容姿に対して、親密性や信頼性を持たせることである。そして2点目は、一方的な説得を防ぐために、擬人化エージェントがユーザの暗黙的な動作や状況を察知して、それに合わせた対応をするという、人間の動作認識理解をすることである。この論文では第3章で説得力のある擬人化エージェントの容姿とコミュニケーション形式の課題に、第4章では、ユーザの状況や動作を理解する方法について論ずる。

以下に続く各章の内容は次の通りである。まず、第3章では、擬人化エージェントを導入している事案に関する現状とこれらの課題について論ずる。次に、Kantharaju らの研究で行われた実験を参考に、擬人化エージェントの容姿や性別、複数のエージェントの関係性の違いがユーザを説得する際にどのような影響を与えるのかということについて調査する。彼らの研究は、擬人化エージェントの数による影響の相違を主題としており、オーバーハードコミュニケーションを主題的に扱っていない。それに対して、本論文では、オーバーハードコミュニケーションが人間の第三者に与える影響力を取り上げる。その際、彼らが映画を題材としていたことに倣って、本論文も映画の鑑賞意欲に対してそれがどのような影響力を与えるのかということに関して調査を行う。これは、彼らの実験をモデルとしつつ、オーバーハードコミュニケーションの影響力という問題に焦点を合わせるためである。彼らの実験と比べて、本論文の調査がどのような点でオーバーハードコミュニケーションの影響力という問題をより直接的に取り上げているのかについての詳細は第3章に委ねる。第4章では、擬人化エージェントはユーザとインタラクティブにコミュニケーションをとることをも課題としているため、その第一段階の技術として先行研究に基づいて人

人間の行動認識技術に対する新しいニューラルネットワークを提案する。人間の生活行動を認識するために、本章では、複数の畳み込みニューラルネットを持つニューラルネットワークについて論ずる。筆者はまず、学習データとして研究室で RGB カメラを用いて撮影し、動画像データを取得した。筆者は、機械が人間の行動を認識する際には、人間の姿勢データを可視化させた方が、動画像データに含まれる服装や顔などの情報に影響されることなく、人間の行動認識が可能となるという仮説を立て、また、その仮説を検証するための比較対象として2つのデータセットを作成した。1つ目のデータセットは Detect dataset であり、これは画像から人間領域のみを切り出したデータである。2つ目のデータセットは Pose dataset であり、Detect dataset を作成する際に抽出した人間領域に、さらに人間の姿勢情報を可視化させたものである。モデルの認識精度評価では、この2つのデータセットを用いて行う。その結果、一番性能がよかったモデルは、Detect dataset 用いた学習で、CNN を5つ用いた場合のものであり、認識精度は97.5%と高い精度を示した。しかし、汎用性の高さで比較すると、Pose dataset を用いた学習で CNN を5つ用いた場合が高いという結果になった。このように、研究室内という限られた空間で人間の行動認識を行う際は、このニューラルネットワークが有効であり、人間の姿勢情報を可視化させた画像を学習した場合に汎用性が高くなるということが実証された。最後に、第5章では、結論を述べる。

3. 擬人化エージェントの容姿とコミュニケーション形式

3.1 関連研究の概要

擬人化エージェントの技術は、近年、様々なシチュエーションで人間の生活に導入されている。インターネット上のオンラインショッピングにて商品を推薦する擬人化エージェントを導入したり[14]、本来、人間が担当してきたウェブ上のカスタマーサービスを擬人化エージェントが業務を行ったり[15]、医療や治療に関わる患者をサポートするために擬人化エージェントを導入したり[16]、福祉施設などでのリハビリ[17]に擬人化エージェントの導入をしたり、するなど様々である。例えば、会話エージェントとして人型のエージェントを用いて、認知症患者とコミュニケーションを試みる中野ら[17]の研究がある。そこからは、擬人化エージェントのもつ心理的負担の軽減という効果が特に明らかに見て取れる。彼らの研究は、認知症患者の数に対する介護者が足りないという問題において、擬人化エージェントとの会話を通して患者の精神安定、介護者の負担軽減を試みた研究である。彼らが提案したエージェントは、あらかじめ、入力されていたテキストを、音声によって発する。また、中野らは、ユーザの話しが終わってから次の質問に進む、相槌を打つ、ユーザの発言を受けて頷くといったジェスチャの機能も付随させたいと考えた。しかし、ユーザに対する次の質問への切り替えや相槌の心地よいタイミングはまだ解明されていない。そこで彼らは、相手の発話に反応し、6秒以上発話音声がない場合に次の質問を投げかける協調的エージェントと相手の発話に関わらず、10秒間隔で質問を発する非協調的エージェントを用いて、実際に、認知症患者と会話させ、認知症患者の発話量が最も増える擬人化エージェントの特徴は何か調査した。その結果、認知症患者の協調的エージェントに対する発話量と非協調的エージェントに対する発話量に大きな差は見られなかった。しかし、彼らが、実験中の被験者の反応を観察したところ、被験者は興味のある話題とそうでない話題とで、発話量が大きく異なることがわかった。また、機械的に質問するだけでなく、エージェントが自分の話を交えながら質問する方が、より相手からの発話を引き出

しやすいくともわかった。これらの結果を踏まえて中野らは、今後は、ユーザの性別や趣味などもあらかじめ選別しておく必要があると考えた。このように、人間に寄り添う振る舞い、あるいは、人間らしいとユーザに印象を当てる要素を擬人化エージェントに加えることは、私たちの生活をサポートし、私たちが負担であると感じる障壁を取り除く可能性を持っている。こうした状況から、擬人化エージェントに関する研究は、擬人化エージェントの容姿や、動作、エージェントとユーザの関係によって、ユーザに対してどのような影響を与えるのか着眼させたものが報告されてきた。まず、擬人化エージェントはユーザに対して、親近感の湧く容姿をもち、友好的なアクションを行い、専門的なテキスト出力をしなければ、影響を与えることはできないとされている[16]。エージェントの容姿に関する研究として梁ら[18]の研究がある。梁らは、インターネット上のオンラインショッピングにおいて、商品推薦に効果的な擬人化エージェントの外見の性質を解明するために研究を行った。そこでは、比較対象として、人型エージェント、ロボットエージェント、犬エージェント、実写ビデオ、仏像エージェント、テキストのみ、という6つの異なる容姿のエージェントが用意された。実験の結果は、人型エージェントが商品推薦をした場合に、ユーザは最も購買意欲が強くなり、親しみやすさも高いという結果が得られた。一方で、仏像エージェントやテキストのみの商品推薦の場合は購買意欲が低く、エージェントへの親しみやすさも低いという評価であった。この研究の結果から、エージェントが人型の外見を持つということが、親しみやすい、信頼できるという印象をユーザに与えるということがわかる。梁らの研究では、エージェントの容姿に着目した研究が行われていたが、ユーザを説得するのにより効果的な方法として、専門的な知識を付加させる必要があるという研究がある[19]。これは、松井らの[19][20]研究で実証されている。彼らは梁らと同じくオンラインショッピング上でユーザに商品を推薦する際にエージェントを導入することを想定した。そうして、ユーザが信頼感を抱く擬人化エージェントには何の要素が必要かという実験を行い、検証した。彼らは、擬人化エージェントによる推薦がユーザの購買意欲を向上させるためには、ユーザが擬人化エージェントを信頼することが必要[21]であるという前提、および、ユーザがよりポジティブな感情のときに擬人化エージェントに対して

信頼感が増す[15]という知見をもとに、システムの設計を行った。そのため、彼らは、エージェントが知的であるという印象をユーザに与えるため[21]、また、ユーザをポジティブな気分させる雰囲気づくりをするために2つの要素をとり入れて設計を行った。1つ目に、推薦する商品に関する詳しい専門的な説明を取り入れ、2つ目は、エージェントが明るく振る舞うジェスチャや表情を取り入れた。実験では、被験者に女性の容姿を持つエージェントが商品の推薦を行う推薦動画を見せ、被験者がエージェントに対してどのような印象を持ったのか、アンケートを行った。その結果、ユーザーは、擬人化エージェントの笑う表情や大きなジェスチャを行うことに対して、ポジティブな感情を抱き、ユーザの推薦動画を見ている時間を楽しいと感じた。しかし、擬人化エージェントが正しい知識を持っているかという問いに対しては、擬人化エージェントの動作や表情に関係なく大きな差がなかった。つまり、これらの結果から、擬人化エージェントが明るい表情や大きなジェスチャをすることに対しては、ユーザはポジティブな印象を持つことがわかる。しかし、商品説明をより詳しく行う。専門的な知識を持った擬人化エージェントに対しては、表情やジェスチャの違いに関係なく、信頼度が高い傾向にあるということがわかった。これらの研究を踏まえると、商品を推薦する場合に、ユーザの意思や意見に影響を与えるには、擬人化エージェントへの信頼度、すなわち、専門的な知性に基づいてテキストを提示することが必要であることがわかる。以上の研究は主にレギュラーコミュニケーションに関するものであった。以下では、擬人化エージェントのオーバーハードコミュニケーションに関する先行研究について整理を行い、本研究で行う実験について詳しく論じる。

序論でも触れたようにオーバーハードコミュニケーションに関する研究は、説得によって相手の行動の変容を促すという点で大きな期待を寄せられている。以下では、このオーバーハードコミュニケーションに擬人化エージェントを導入したいくつかの研究を紹介したい。まず、擬人化エージェントがユーザを「説得する」メディアとして強い影響力があると報告した鈴木らの研究がある[22]。彼らはオンラインショッピングにおいて、説得エージェントと被説得エージェントによるオーバーハードコミュニケーションと説得エージェントがユーザに直接情報を提示するレギュラーコミュニケーションのそれぞれの影響の

相違を調べた。そのために彼らは、それぞれのコミュニケーション形式に関して、ユーザに説得を施した後、ユーザの商品に対する購買意欲、エージェントに対する魅力、エージェントに対する信頼性についてアンケート調査を行い比較した。実験の参加者は大学生・大学院生・社会人の24名が参加した。実験はwebブラウザ上で行われた、画面にある雑貨について擬人化エージェントがレギュラーコミュニケーション、あるいはオーバーハードコミュニケーションの形式で商品の紹介を行う様子が映し出され、商品紹介が終わった後に、10段階で購入意欲についてアンケートを行うという流れであった。レギュラーコミュニケーションの条件では、説得エージェントのみが画面に登場し、説得エージェントが被験者に視線を向けて直接商品説明を行う様子を提示した。一方で、オーバーハードコミュニケーションの条件では、説得エージェントと被説得エージェントが画面上に登場し、説得エージェントが被説得エージェントに視線を向けて、被説得エージェントに対して商品の説明を行う様子を提示した。また、オーバーハードコミュニケーションの条件では、被説得エージェントが説得エージェントの発話に対して「ほう」「なるほど」「ふむふむ」などの相槌を打つというパフォーマンスが取り入れられている。この実験を行った結果、購買意欲の評価に関しては、擬人化エージェントによるオーバーハードコミュニケーションの方がレギュラーコミュニケーションに比べて、より効果的であるという結果が得られた。しかし、「商品説明に対する専門性」や「商品説明に対する信頼性」については、オーバーハードコミュニケーションであってもレギュラーコミュニケーションであっても2つの条件の間で有意差は見られなかった。

この研究を踏まえて、被説得エージェントの返事や表情などの反応の違いによってユーザが抱く擬人化エージェントへの印象が変化し、説得に対する影響も変化が生じる可能性があるとする齋藤の研究がある[10]。この研究では、鈴木ら[22]と同様に説得におけるオーバーハードコミュニケーションに着目する。しかし、それに加えて、被説得エージェントは相槌だけでなく、3つの反応パターンを実装された。すなわち、齋藤は、説得エージェントの意見をすべて否定する場合、すべて肯定する場合と同調をする場合の3つの条件を設定し、比較実験を行った。実験の参加者は大学生45人として、肯定条件、同調条件、

否定条件の3条件にわけ、実験を行った。エージェントは2体用意し、参加者は事前に説得話題に関して同意するか否定するかをアンケートで答えた。そして、実験の後に、事前に行ったアンケートと同様の内容について回答した。また、その際に、説得エージェントと被説得エージェントに対する印象もアンケートに答えてもらった。実験で取り上げた話題としては、車の後部座席のシートベルトの重要性、夜食の危険性、傘さし運転の危険性、エアコン暖房についての4つである。この実験を行った結果、どの条件においても、事前から事後にかけて各話題の説得内容に肯定的な意見への変容が見られた。特に、同調条件と否定条件に関して、同調条件が有意に大きいことが明らかになった。肯定条件については、2体に説得されていると感じたという意見もあるため、あまりにも、自分の意見とかけ離れた場合に肯定してしまうことは、結果として押し付けがましさを与えてしまうと考えられる。また、説得エージェントと被説得エージェントに対する印象については、被説得エージェントについては「親切」「友好的」「好感が持てる」「温かい」「客観的」の項目が否定条件において低くなることが明らかになり、説得エージェントにおいては、どの条件においても「信頼できる」という評価を得た。これらの結果を踏まえて、同調条件においては、被説得エージェントが肯定的な意見だけを述べるのではなく、否定的な意見を述べることで、客観的に2体のエージェントの会話を捉えることができ、押し付けがましさを軽減できる。また、被説得エージェントがユーザの意見に同意した場合、説得エージェントと共にユーザにより信頼感を与えることが明らかになった。したがって、被説得者に「人らしさ」を感じさせることが魅力的に説得できるという可能性を示された。

他にも、この「人間らしさ」を擬人化エージェントの要素として付け加え、ユーザに対する影響力を調査した研究がある。性別の違いや表情、ジェスチャー、服装、専門的な用語を使うか使わないか、エージェントの数といった様々な要素を組み合わせ、ユーザに対して影響力のあるエージェントが何かということについて、比較を行った Kantharaju ら [11]の研究である。彼らは、まず擬人化エージェントとして、女性 (F: Female) と男性 (M: Male) の2体を用意した。さらに、専門的な知識を有する Authoritative (A) と学生のようなカジュアルな表現を用いる Peer (P) の要素をそれぞれに与え、4体の擬人化

エージェント（MA、FA、MP、FP）を用いて比較実験を行った。MA と FA は年齢が 40 代ぐらいのエージェントとして設定されているため、服装はフォーマルなもの、表情やジェスチャーなどは落ち着いたものが表現として用いられている。一方、MP と FP はカジュアルな服装であり、髪色も明るく若者のエージェントとして設定されている。実験で扱う題材としては、年代別や性別の違いで偏見を持たないことを理由として映画が取り上げられた。彼らは、まず、4 体の擬人化エージェントがユーザに対して、それぞれの映画を観るように説得を行ったビデオクリップを作成した。ビデオクリップは大きく 3 つのパターンがある。4 体それぞれのエージェントが単数で被験者に直接説得するレギュラーコミュニケーションのパターン、2 体のエージェントが被験者に対して直接説得を行うパターンと説得エージェントと被説得エージェントの役目を持つ 2 体のエージェントが対話する様子を被説得者に見せるオーバーヘッドコミュニケーションのパターンである。さらに、それぞれのパターンに 4 つ説得の種類があり、合計で 12 この説得パターンがある。次に、彼らは、被験者に対して事前アンケートとして、歴史映画、コメディ映画、犯罪映画それぞれ 3 つの映画について、どれほどの鑑賞意欲があるのかアンケートを行った。そして、彼らは、被験者が一番低い評価をした映画に関して擬人化エージェントが説得を行う対話ビデオクリップをみせた。最後に、改めて、評価が低かった映画の鑑賞意欲についてアンケートを行い、ユーザの評価がどのように変化したのか調査を行った。この実験の結果、擬人化エージェントの性別に有意差はなく、Authoritative に信頼性を感じ、説得されやすいというユーザの評価があった。また複数のエージェントで説得された方が意見の変容が起きやすいという結果になった。最も説得力があったエージェントのパターンは、MA が FA を説得するビデオクリップをみた場合と FA と MP が被験者に対して説得を行うビデオクリップをみた場合であった。それに次いで説得力があったのは、FA と MA の両者が被験者に対して説得したビデオクリップを観たパターンのものであった。これらすべての研究は、説得の場を商品や映画の推薦という特定のシチュエーションに限定しているため、擬人化エージェントによるユーザへの影響について 1 つの知見を示したに過ぎない。しかし、Kantharaju ら[11]の研究では、擬人化エージェントの性別にはユーザに対して影

響があまりないとされるが、男性と女性が2体のペアである場合に、ユーザに説得力を与えている可能性があると考えられる。

そこで、本研究では、男性と女性の擬人化エージェント2体の組み合わせによるオーバーヘッドコミュニケーションがユーザに対して影響力を持つ仮定し、説得エージェントと被説得エージェントの性別の違い、外見やセリフの違いがユーザの意見の変容にどのように影響力を与えるのかを検討する。また、単数のエージェントが被験者(以下:ユーザ)に対して直接説得を行う場合のパターンも用意し、説得エージェントと被説得エージェントと擬人化エージェントが複数で説得する場合とで比較実験を行う。また、その実験から得た考察を元に、さらに擬人化エージェントの容姿に重点を当てた実験を行い、比較を行う。

3.2 実験①の設定

3.2.1 実験①の手順

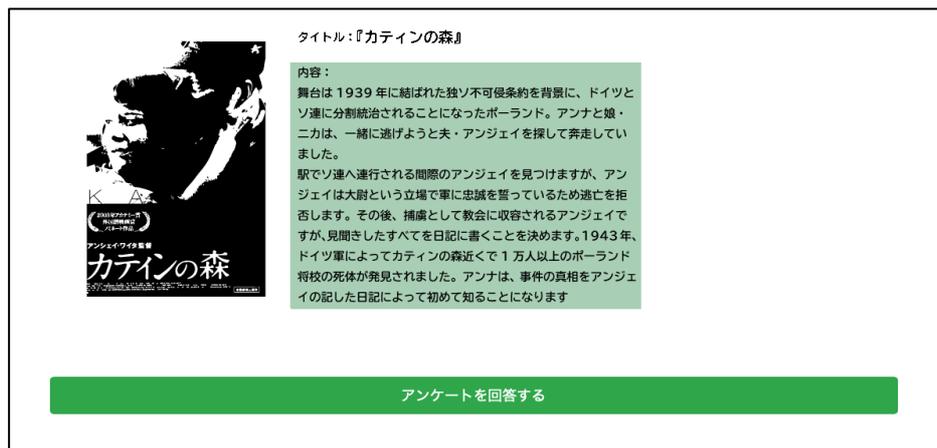


図 1 映画の広告文章



図 2 エージェントの説得文章

本研究では、ユーザに対して効果的な説得を行うエージェントの外見、およびそれら2体の組み合わせとオーバーヘッドコミュニケーションの影響との関係性を調査するために、性別や服装、髪型、身につけているアイテムに違いを持たせた6体のエージェントを用意した。6体のエージェントは、被説得エージェントと説得エージェントの役割に振り分けられ、22パターンの組み合わせで比較実験を行う。

Kantharaju らの実験に倣って、エージェントやエージェントが発言するダイアログ、説得する話題を設定した。ユーザは犯罪映画、コメディ映画、歴史映画、それぞれのジャンルに対して3つの映画、合計9つの映画のなかからランダムに選別された広告を見て、その映画に対する鑑賞意欲についてアンケートに答える。映画の広告文章を表示する Web ページを図 1 に示す。続いて、単数のエージェント、あるいは、2 体のエージェントが表示され、ユーザに対して説得を行うページが表示される。2 体のエージェントによる説得の例を図 2 に示す。映画の広告文章を表示する Web ページでは、実際に映画のジャケットとして使われた画像をモノクロ加工し、映画に関する広告文章を掲載した。広告文章は 200 字から 300 字としたが、予備実験を行った際に、映画の広告文章が長い、文章を読み終わるまで時間がかかるといった指摘があったため、アンケート画面に移行する際は、各個人のタイミングで切り替えることができるように設計した。図 1 の Web ページの下部にある緑色のボタンをクリックすることで、次のアンケート画面に切り替わる。次に、アンケート画面では、提示された映画に対する鑑賞意欲を 5 段階評価（1：全然観たくない～5：とても観たい）で回答するよう求められる。ユーザが、アンケートに回答した後、図 2 のようなエージェントによる説得が行われる Web ページに画面が移行する。エージェントが単数の場合は Web ページにエージェント表示され、エージェントの右側にテキストが表示される。また、エージェントが複数の場合には、図 2 のように、Web ページの左側に説得エージェント、右側に被説得エージェントとして擬人化エージェントが表示される。表示されるエージェントは、単数の場合であっても、複数の場合であっても、6 体のエージェントがランダムに表示される。テキストは、エージェントの間に表示されるが、吹き出しで囲み、さらに文章の位置を左揃えと右揃えにすることで、それがどちらによる発言か分かるように表示した。具体的な説得文章については 4.3.5 で説明する。その後、再び、アンケートの画面に表示され、エージェントによる説得文章を読んで、映画に対する鑑賞意欲について 5 段階評価（1：全然見たくなくなった～5：とても見たくなくなった）で回答する。また、説得を行ったエージェントに対する信頼度についても 5 段階評価（1：全然信用でき

ない～5：とても信用できる) で回答を求める。この実験をユーザ 1 名に対して 4 回繰り返し行う。

3.2.2 エージェントの設定

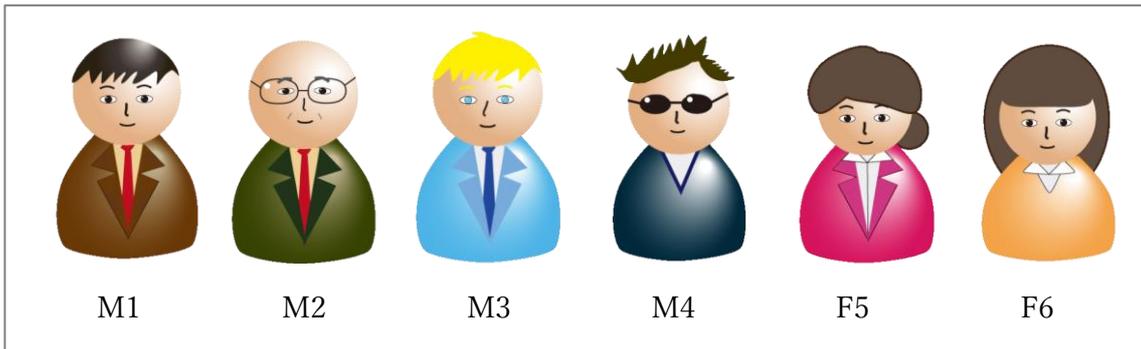


図 3 エージェント

表 1 説得パターン

パターン	説得者	被説得者	パターン	説得者	被説得者
0	M1	ユーザ	11	M2	F6
1	M2	ユーザ	12	M3	F6
2	M3	ユーザ	13	M4	F6
3	M4	ユーザ	14	F5	M1
4	F5	ユーザ	15	F5	M2
5	F6	ユーザ	16	F5	M3
6	M1	F5	17	F5	M4
7	M2	F5	18	F6	M1
8	M3	F5	19	F6	M2
9	M4	F5	20	F6	M3
10	M1	F6	21	F6	M4

エージェントの容姿、男性、女性の組み合わせによって、ユーザの説得効果にどのような影響があるのかを調査するために、図 3 のような、男性エージェント 4 体と女性エージェント 2 体の 6 体を用意した。容姿や性別に大きな違いが出るように、フリー素材である男性サラリーマンの画像と女性サラリーマンの画像を参考にし、Adobe Illustrator を用いて筆者自身が制作した。また、髪型と髪色の違いや、年齢の違い、身につけているアイテムの違いなど、できるだけエージェント同士に容姿の違いが出るように設計した。エージェントの職業や名前などの詳細な設定はなく、左から M1、M2、M3、M4、F5、F6 とする。エージェントの職業を設定しなかった理由としては、エージェントの職業によって信頼度に偏りが出ないようにしたためである。この 6 体を用いて、単数のエージェントで説得する場合と複数のエージェントで説得する場合の 2 条件で比較を行う。その中で、組み合わせをさらに細分化し 22 個の説得パターンで比較実験を行う。単数のエージェントの場合はユーザに直接説得を行うレギュラーコミュニケーションの形をとり、複数のエージェントで説得を行う場合は、1 体を説得エージェント、他の 1 体を被説得エージェントの役に振り分け、説得エージェントが被説得エージェントを説得するというオーバーヘッドコミュニケーションの形をとる。

3.2.3 映画の広告の設定

個人的な偏見を避けるために、説得を行う話題は可能な限り中立であり、幅広い人々が関心を持つ映画について取り上げた[23]¹。映画のジャンルは性別によって人気や不人気を問わない犯罪映画、コメディ映画、歴史映画の 3 つのジャンルを選択した。それぞれのジャンルにおいて、実際に放映された作品の中から 3 つずつ選んだ。映画は実際に放映されたものであるが、映画の広告文章については筆者が独自に作成したものである。実際にイン

¹映画という主題選択においては十分に再考の余地があるが、今回は Kantharaju らの実験設定に倣うこととする。

ターネット上にあるこれらの映画に関する説明文章や様々な人によって投稿されているレビューや口コミを参考に、できる限り中立的な立場から、映画の広告文章を作成した。広告の文章は大体 200 字から 300 字程度とし、映画ごとに 1 つの広告文章を用意した。説得システムで扱った映画のタイトルと説明文を以下の付録 A に示す。実験では、ユーザがこの映画の広告文章を読み、この映画に対する鑑賞意欲をアンケートに 5 段階評価 (1: 全然見たくない~5: とても観たい) で回答してもらう。

3.2.4 ダイアログの設定

ダイアログは単数のエージェントが説得する場合と、複数のエージェントが説得する場合の 12 パターンとなっている。エージェントが単数の場合、つまり、レギュラーコミュニケーションの形をとる場合のダイアログは、専門的な知識に基づいて説得を行う文章 (パターン: a・b・e) とカジュアルに説得を行う文章 (パターン: c・d・f) の 2 つのパターンを用意した。一方、エージェントがオーバーハードコミュニケーションの形をとる場合のダイアログは、被説得エージェントが説得エージェントに発するテキストに対して様々な反応を示すものを用意した。すなわち、説得エージェントに対して否定的な意見を持つが、意見が変容している様子がある (パターン: g・h・l) ものと被説得エージェントが説得エージェントに対して同意、肯定的な反応を示しながら説得される様子を表す (パターン: h・i・j) ものを用意した。具体的に表示されるテキストを付録 B に示す。A が説得エージェントのセリフであり、B が被説得エージェントのセリフである。

3.3 実験①

3.3.1 実験①の参加者

121名の神戸大学国際人間科学部の学生に実験への協力をしてもらった。121名のうち3名のユーザは1回の参加で5回、他のユーザは設定通り4回の実験を受けており、全体として487個のデータを収集した。

3.3.2 実験①の手順

ユーザには実験を行う際にIDとして学籍番号を入力してもらいが、このIDはアンケートの回答が同一人物であるものごとを確認する目的のみとして利用され、保存されることや再利用することは一切ないことを注意点として実験前に伝えた。この注意点に加えて、以下の教示を与えた。

- この研究では、エージェントによる映画オススメシステムの評価と検証を行います。
- 最初に映画の広告文章が表示され、その後、その映画に対する鑑賞意欲についてアンケートに答えてもらいます。続けて、その映画に関してエージェントがオススメ（説得文章）をテキスト表示によって行うので、その文章を読んでもらいます。その後、再びアンケート画面に切り替わりますので、映画に対する鑑賞意欲がどう変わったのかについてとオススメを行ったエージェントに対する信頼度についてアンケートに回答してください。この映画オススメシステムが4回繰り返し行われます。
- 実験中は、リロードボタンやブラウザバック、[戻る]のボタンを押さないでください。
- 実験は10分弱を予定しています。
- 何か、不明な点、システムの不具合があれば、実験者に質問してください。

まず、ユーザには、このWebシステムの流れとアンケートの回答方法に関する説明文章を読んでもらう。そして、映画の広告文章を読み、エージェントの説得を読んで、鑑賞意欲

にどのような変化が出たのかをアンケートに回答してもらうという作業を4回繰り返した。この Web システムは、映画の広告文章や説得を行うエージェントのパターン、ダイアログの全てに関して重複しないように、かつ、ランダムで行われているように設計されている。ユーザは、映画の広告文章を読んだ後、その映画に対する鑑賞意欲について5段階評価（1:全然観たくない～5:とても観たい）でアンケートに回答する。次に、エージェントによる映画を観るように促す説得があり、エージェントの説得文章を読んでから、どのように映画に対する鑑賞意欲が変わったのかを5段階評価（1:全然観たくなかった～5:とても見たくなかった）でアンケートに回答する。また、エージェントによる説得後に行われるアンケートでは、説得を行ったエージェントに対する信頼度についても5段階評価（1:全然信用できない～5:とても信用できる）で回答してもらう。

3.4 実験①の結果

3.4.1 説得に適したエージェントパターンについて

結果として Rate change と Credit を求め、分析した。Rate change とは、ユーザの鑑賞意欲の変化について示し、エージェントの説得後の鑑賞意欲からエージェントの説得前の鑑賞意欲を減じた値を、説得パターンまたはダイアログの種類ごとに集計、平均したモノである。Credit とは、エージェントに対する信頼度に関するアンケート結果を平均した値を示す。

それでは、エージェントのパターンとダイアログのパターンのどちらがユーザの意見変容に影響を与えるのかを明らかにするため、エージェントパターンとダイアログパターンの Rate change をもとに回帰分析を行った。今回は1つの目的変数を複数の説明変数で予測するため、重回帰分析を用いる。重回帰分析の式は以下に示す。

$$y = a x_1 + b x_2 + c \quad (1)$$

目的変数 y と Rate change を代入し、 x にはエージェントのパターンとダイアログのパターンを正規化した値をそれぞれ説明 x_1 、 x_2 として係数 a 、 b 、 c を求めた。結果を(2)式に示す。

$$\mathbf{Rate\ change = -0.21 \times Agent + -0.01 \times Dialog + 1.26} \quad (2)$$

(2)の式の結果から、エージェントパターンの違いがダイアログパターンの違いよりユーザの意見変容に影響を与えている可能性があることが示された。

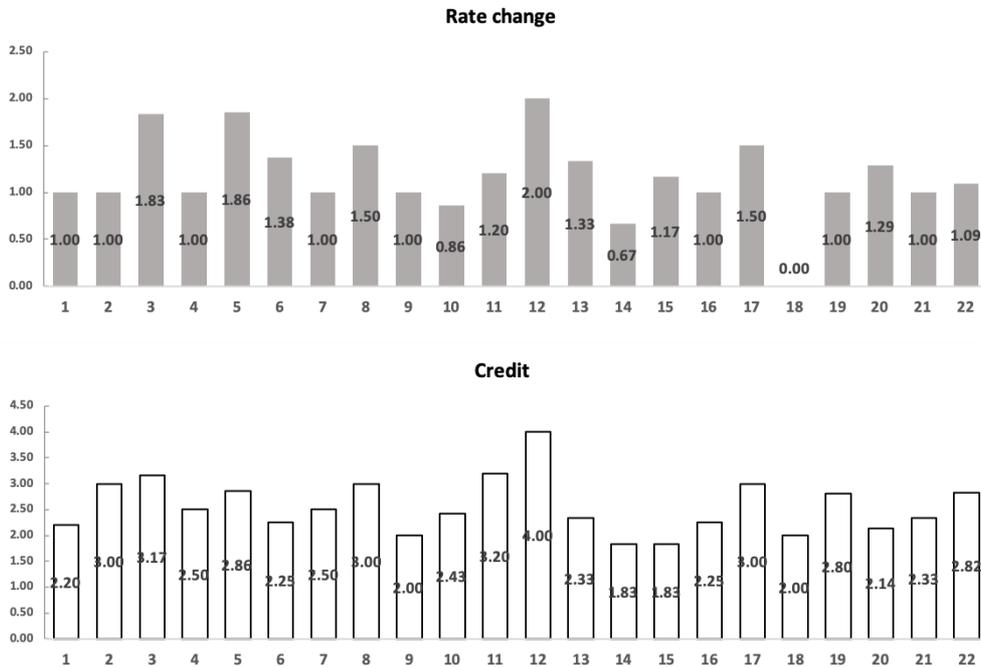


図 4 エージェントのパターン別 Rate change(上)と Credit(下)の比較

ユーザがエージェントによる説得前の鑑賞意欲が1または2の場合について分析を行う。まず、説得パターンを図4に示す。図4の横軸1~22はエージェントのパターンを示し、上の図がRate changeの平均、下の図がCreditの平均を示したものである。Rate changeとCreditの図はそれぞれ、エージェントによる説得後の変化について平均を集計したものである。また、Rate changeとCreditの相関関係数は0.7であり強い相関があることがわかった。これは、ユーザがエージェントに対して信頼度が高いと、意見が肯定的に変容する可能性が高いということを示す。また、パターン別で平均値を比較したところ、最も鑑賞意欲を高め、信頼度があったパターンは12であった。これはM2の男性エージェントがF6の女性エージェントを説得しているパターンである。他のパターンと平均の差を比較するため有意水準を0.05とし、t検定を行ったが、有意な差はみられなかった。

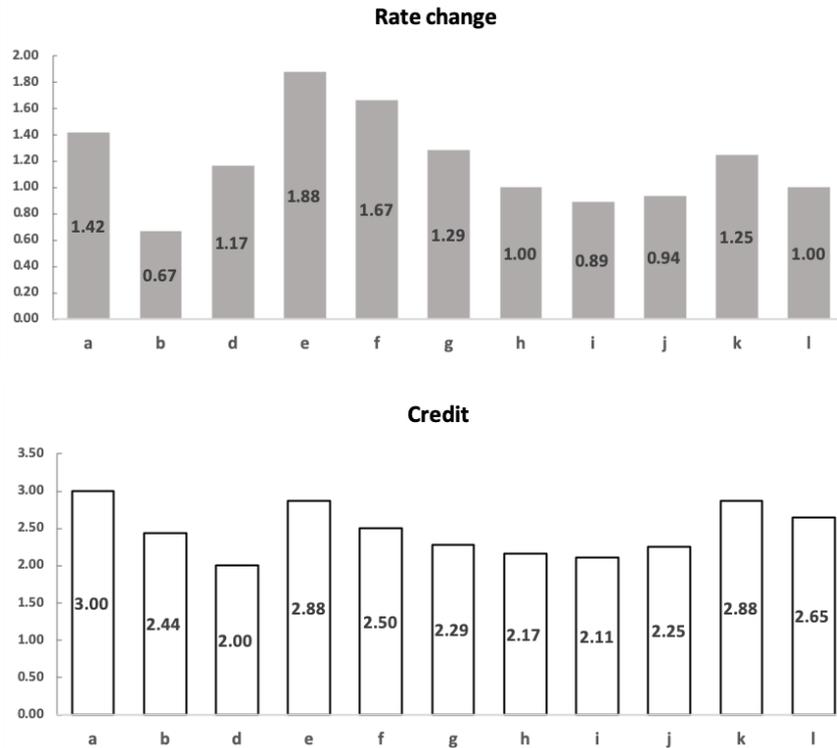


図 5 ダイアログパターン別 Rate change(上)と Credit(下)の比較

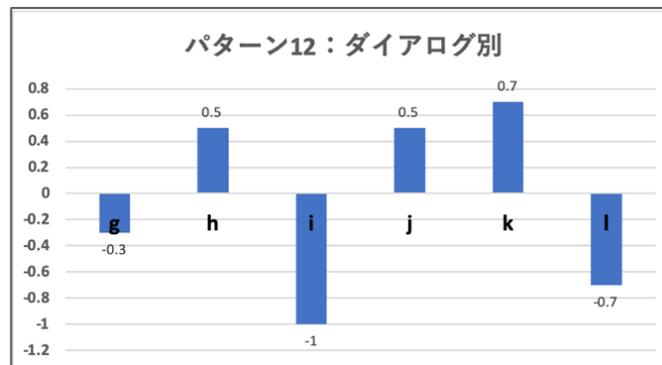


図 6 パターン 12 のダイアログ

図 5 は、ダイアログのパターンごとに Rate change と Credit をまとめたものである。ダイアログのパターンは a~l の 12 個あり、a~f がレギュラーコミュニケーションの場合のダイアログ、g~l がオーバーハードコミュニケーションの場合のダイアログである。鑑賞意欲を高めたダイアログパターンは e であり、信頼度を与えたダイアログパターンは a であったが、有意な差はなかった。この実験では、エージェントパターンが 12 の場合で最もユーザーに意見変容の影響を与えたのは k のダイアログパターンであることがわかった。以

上の結果から、エージェントパターンが 12 のダイアログが k の場合に、ユーザの鑑賞意欲が低かった映画に対して、鑑賞意欲が上がるような説得を行うことができるということが明らかになった。

3.4.2 RC と OHC の比較

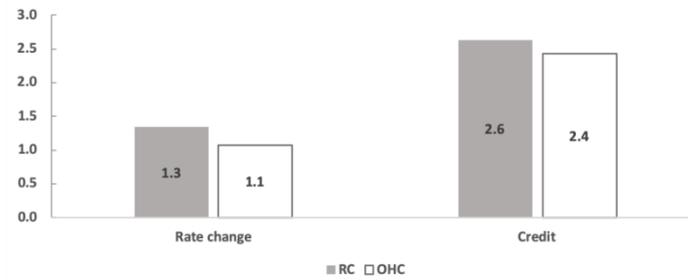


図 7 RC と OHC の Rate change 平均と Credit 平均の比較

レギュラーコミュニケーション(RC)とオーバーヘッドコミュニケーション(OHC)で比較すると、レギュラーコミュニケーションの Rate change と Credit のどちらもオーバーヘッドコミュニケーションより高い数値を示した。この比較から、この実験の場合には、レギュラーコミュニケーションの方がユーザに対して意見の変容を与え、Credit も高いという結果になった。

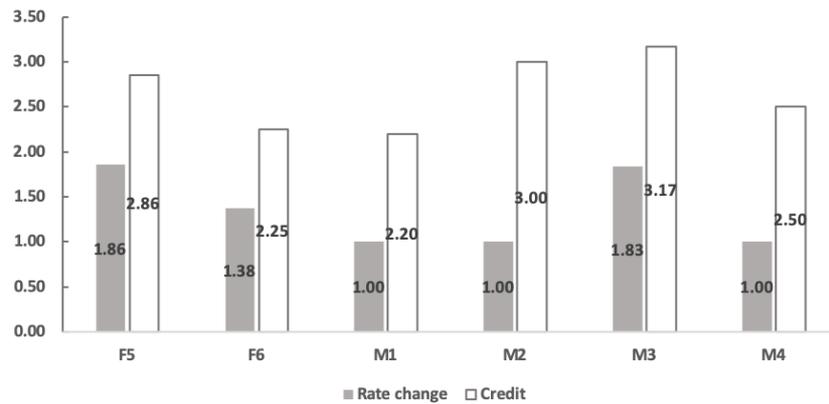


図 8 RC エージェント別 Rate change と Credit の比較

次に、図 8 ではレギュラーコミュニケーションのコミュニケーション形式における、エージェント別の比較を行う。実感参加者の鑑賞意欲に対して、最も意見変容を与えたエージェントは、F5 であった。そこで、F5 と他のエージェントに有意な差があるのか、t 検定を行ったところ、有意差が見られたのは、F5 と M1 のエージェントの間($p < .05$)と F5 と M4 のエージェントの間($p < .05$)であった。したがって、この結果から、M1 と M4 のエージェントと F5 のエージェントの説得効果を比較した場合に、M1 と M4 のエージェントに比べて、F5 のエージェントが説得に適していることが示された。

3.4.3 OHC の場合の説得エージェントと被説得エージェントの比較

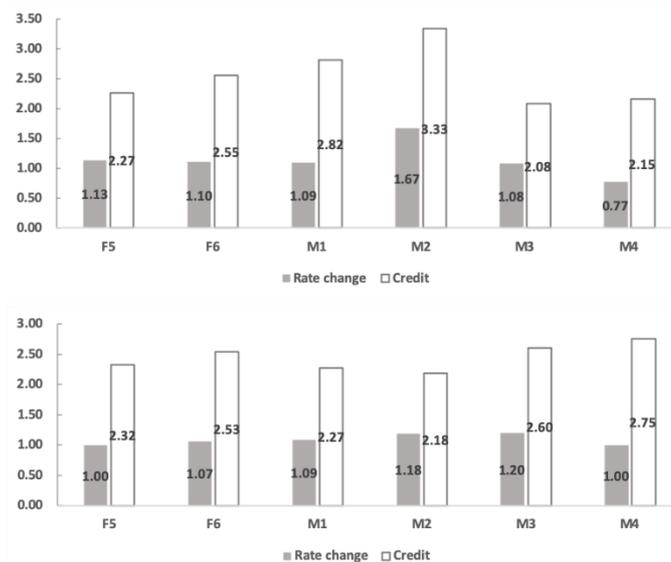


図 9 説得エージェント別(上)比較・被説得エージェント別(下)比較

説得エージェント別と被説得エージェント別に Rate change と Credit の平均を示したのが図 9 である。説得エージェントとは、オーバーハードコミュニケーションの場合に説得する側のエージェントのことを指し、被説得エージェントとは、説得される側のエージェントのことを指す。説得エージェント別に比較をすると、M2 が説得エージェントの場合に、最も Rate change が高く、Credit が高い。一方で、説得エージェントが M4 の場合、最も Rate change が低く、Credit の評価も低いという結果になった。しかし、6 体それぞれのエージェント間で t 検定を行ったところ、有意な差は見られなかった。

次に、被説得エージェント別に Rate change と Credit の平均を比較すると、相関もなく、エージェント別に大きな評価の差がなかった。また、被説得エージェントの間にも有意な差がなく、被説得エージェントとして最も効果のあるエージェントはいなかった。これらの結果から、オーバーハードコミュニケーションのコミュニケーションの場合、ユーザを

説得する際に、説得エージェントの違いがユーザに影響を与えている可能性があるということがわかった。

3.5 実験①の考察

本章では、擬人化エージェントに容姿や性別の違いを持たせ、ユーザを効果的に説得する説得エージェントと被説得エージェントの組み合わせ、エージェントの容姿についての検証を目的とした。実験は、Web アプリケーションを作成し実施した。

実験の結果については、まず、オーバーハードコミュニケーションの場合、パターン 12 である M2 の男性エージェントが F6 の女性エージェントを説得しているものがユーザに対して、映画の意見変容を促した。

次に、レギュラーコミュニケーションとオーバーハードコミュニケーションを比較した場合には、レギュラーコミュニケーションの方がオーバーハードコミュニケーションよりユーザの説得については効果的であった。また、レギュラーコミュニケーションの中で有意な差が見られたのは M1 と M4 のエージェントと F5 を比較した場合であった。これは、M1 と M4 の男性エージェントと比較した場合に、F5 の女性エージェントが説得に適しているということがわかった。

さらに、エージェントとダイアログの比較を行った場合、エージェントの方がパターン別で Rate change と Credit に違いが生じた。これは、エージェントの容姿や性別の違いがユーザに影響を与えている可能性が高い。また、オーバーハードコミュニケーションの場合、特に説得を行うエージェントの容姿や性別がユーザに影響を与えていることがわかった。しかし、6 体のエージェントの中で特に影響力を持つエージェントは見られなかった。また、ユーザが Credit が高いと評価したエージェントパターンは、映画に対する鑑賞意欲を説得前よりも説得後に高い評価に変化させた。つまり、ユーザが Credit が高いと評価をした説得エージェントは、ユーザへ説得を促し、映画に対する鑑賞意欲を高めるという効果

がある。これらの結果を踏まえて、エージェントのパターンを減らし、ユーザに対して説得を行うエージェントの容姿に重点をあて次の実験を行った。

3.6 擬人化エージェントの要素に関する実験の概要

前節の実験を踏まえて、ユーザに対して説得力を持つ擬人化エージェントの容姿の具体的な要素を明らかにするために実験を行った。前節の実験では、擬人化エージェントのパターンやダイアログのパターン、映画広告のパターンが多いことが問題となり、説得力を持つ擬人化エージェントの容姿について具体的な要素を明確にすることが困難であった。そこで、本節の実験では、より擬人化エージェントの容姿に重点をあてた。具体的には、擬人化エージェントの容姿にメガネの有無といった明確な違いを持たせた。擬人化エージェントの違いをユーザの意見変容に影響に違いがあるのかを検討し、擬人化エージェントに説得力を持たせるための具体的な要素を明らかにする。また、この実験では、オーバーハードコミュニケーションやレギュラーコミュニケーションの組み合わせや説得を促すようなダイアログを用いない。本節の実験では、映画の説明文章のみの場合と容姿の違いを持たせた擬人化エージェントがそれぞれ、映画をオススメするような場合で比較を行う。本節においても、実験用の Web アプリケーションを作成した。Web アプリケーションでは、ユーザが映画の広告をみて、その映画に対する鑑賞意欲をアンケート形式で回答するというようなシステムになっている。また、本節の実験に用いる映画は、前回の実験で最も低評価を受けた映画広告 1 つのみである。

3.7 実験②の設定

3.7.1 実験②の実験用 Web ページ

実験用の Web アプリケーションに表示される映画の広告は以下の図 10 のようになっており、表示される映画の広告の種類は 5 つである。映画の紹介文章のみが表示される広告が 1 パターンと、男女のエージェント 4 体がそれぞれ映画の紹介文章と共に表示される広告が 4 パターンあり、合計 5 パターンの映画の広告がランダムに表示される。

本節の実験では、擬人化エージェントによるユーザに対する説得ダイアログはなく、図 11 のような映画の広告を使用する。映画の説明文章の右下に擬人化エージェントと吹き出しに「オススメ」と書かれたものを使用し、映画の広告 5 つのパターンで比較を行う。そして、映画の広告は、前節の実験で最も低評価が多かった歴史映画のジャンルである「トゥルー・ヌーンイワノビッチ村」を用いる。

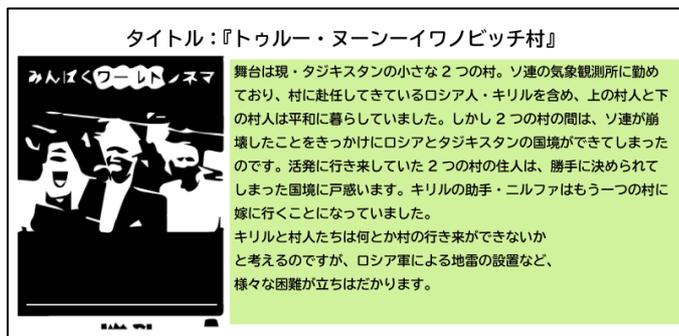


図 10 映画広告

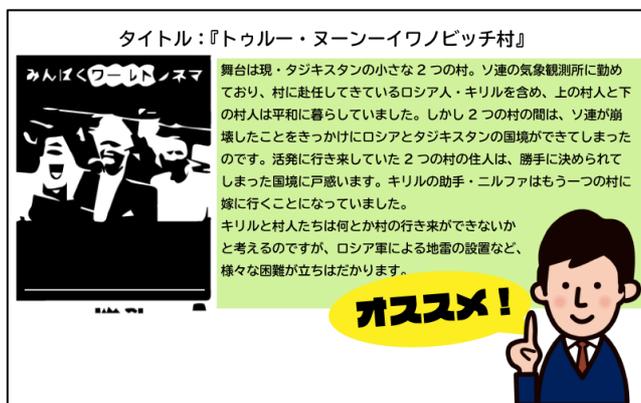


図 11 映画広告エージェントあり

3.7.2 エージェントの設定

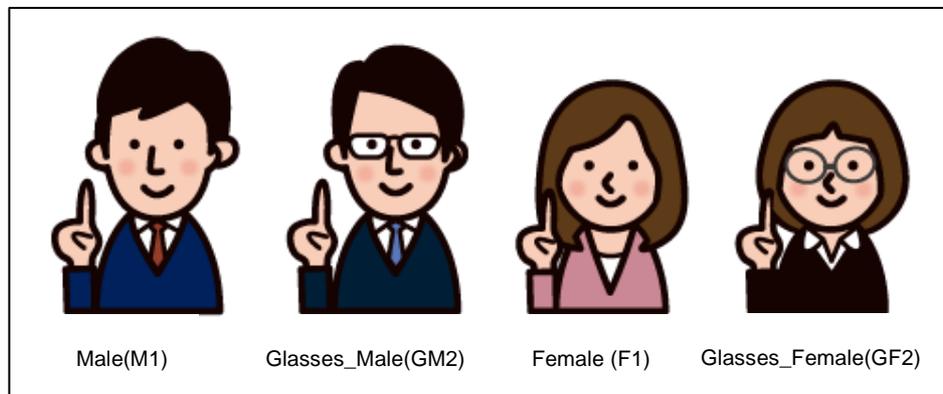


図 12 容姿に違いを持たせたエージェント 4 体

擬人化エージェント 4 体は図 12 に示す。Male(M1)、Glasses_Male(M2)、Female(F1)、Glasses_Female(GF2)の男性と女性のエージェントをそれぞれ 2 体ずつ用意した。前回の擬人化エージェントとは違い、メガネの有無で明確に容姿の違いを持たせた。実験用の Web アプリケーションでは、これら 4 体のエージェントが映画の広告の右下に「オススメ」とともにランダムで表示される仕様となっている。

3.8 実験②の概要

ユーザは、神戸大学に所属する学生や卒業した社会人、他の大学に在籍する学生に協力してもらい、合計で 122 名であった。ユーザには実験用の Web アプリケーションに接続できる URL を配布し、各自で実験の参加をお願いした。ユーザは、各自で実験に関する説明の文章を読み、映画の広告文章を読んだ後、その映画に対する鑑賞意欲について 5 段階(1：全然観たくない～5：とても観たい) で評価するという手順になっている。

3.9 実験②の結果

3.9.1 映画の広告パターン別比較

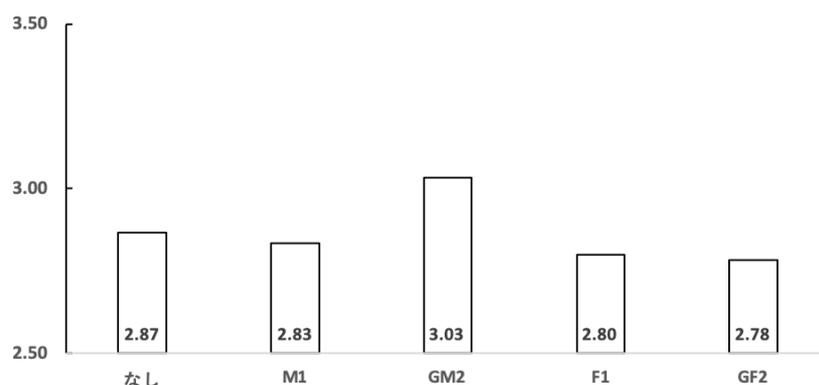


図 13 映画の広告パターン別鑑賞意欲の比較

変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値
グループ間	1.07977261	4	0.26994315	0.29836508	0.87849672	2.45057052
グループ内	104.045227	115	0.90474111			
合計	105.125	119				

分散分析表						
変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値
グループ間	1.07977261	4	0.26994315	0.29836508	0.87849672	2.45057052
グループ内	104.045227	115	0.90474111			
合計	105.125	119				

表 2 分散分析結果

図 13 は、映画の広告パターン別に評価の平均を比較したグラフである。最も平均値が高かったのがエージェント M2 のメガネをかけた男性が表示された場合であった。また、最も平均値が低かったのがエージェント F2 のメガネをかけた女性が表示された場合であった。映画広告のパターン 5 つの平均値を比較するために一元配置分散分析を行ったところ、表 2 のようになった。映画広告のパターン別に有意な差はみられなかった。

3.9.2 メガネの有無比較

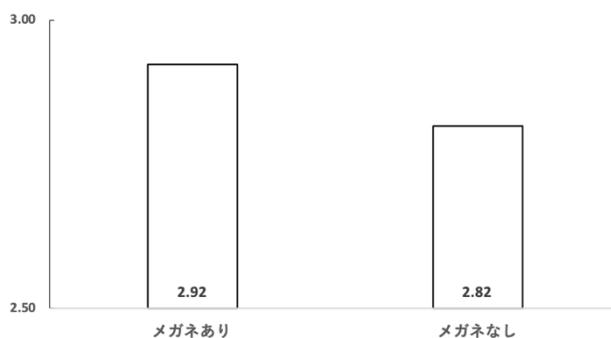


図 14 メガネの有無鑑賞意欲比較

メガネの有無で映画の鑑賞意欲を比較した結果が図 14 である。メガネありとメガネなしでは 0.1 の差でメガネありのエージェントがメガネなしのエージェントを上回るが、有意な差は見られなかった。したがって、メガネの有無がユーザに意見変容に影響を与えるとは限らないという結果になった。

3.10 まとめ

本節の実験では、前節の実験の考察を元に、ユーザに対して説得力を持つ擬人化エージェントの容姿の具体的な要素を明確にすることを目的に実験を行った。映画の説明文章のみの場合とメガネの有無の違いを持たせた男性と女性の擬人化エージェント 4 体が映画の説明文章に対して「オススメ」という文字とともに表示される場合でユーザの映画に対する鑑賞意欲について比較を行った。その結果、有意な差は見られなかったが、メガネ有りの男性エージェントの評価が最も高く、メガネ有りの女性エージェントの評価が最も低いという結果になった。メガネの有無に関しても有意な差は見られず、映画の広告に対して文章のみの場合と擬人化エージェントがいる場合の比較についても大きな差は見られなかった。有意な差が出なかった原因としては、データの数が 122 と少なかったことが挙げられる。また、これらの結果から、擬人化エージェントがユーザを説得するには、擬人化エ

エージェントの容姿に適したダイアログを設定する必要があると考えられる。つまり、ユーザに対して擬人化エージェントを提示するだけでなく、擬人化エージェントに対してユーザは信頼性を持つ要素が必要である。以上から、ダイアログがあった場合には、説得の擬人化エージェントに対して信頼性を持ち、ユーザの意見変容に影響を与えるが、「オススメ」といった簡単な文章と擬人化エージェントを表示させるだけでは、ユーザに対して説得を促すことは容易ではなく、擬人化エージェントの説得は効果を発揮しない可能性があるということがわかった。

4. ディープラーニングを用いた人間の状態認識システム

4.1 概要

第4章では、擬人化エージェントがユーザに不愉快な印象を与えず、ユーザのリアクションに対応した反応を行うことができる技術の第一段階として、動作認識理解の手法について論ずる。具体的には、研究室で撮影した動画データにタグをつけ、ニューラルネットワークを用いて学習し、ユーザの状態を認識するという手法である。まず、動画データを学習し分類させるためのディープラーニングの手法として代表的なものはLSTM[24]や3-DCNN[25]などが挙げられる。これらは時系列データを扱うため、長い時間撮影したデータを学習することで任意精度が上がる特徴を持つ。そのため、認識精度をあげるためには、大量の学習データが必要となり、パラメータの複雑な調整が必要となる。そこで、本研究では少ない動画データで人間の生活行動を認識可能とする、複数の畳み込みニューラルネットワークを持つニューラルネットワークを提案する。そして、2つのデータセットを用いて学習し、作成した学習モデルの認識精度の評価を行なう。2つのデータセットのうち、1つは Detect dataset であり、画像に含まれる人間領域のみをYOLO[26]を利用して抽出したものである。もう1つは、Pose dataset として、抽出した人間領域のみの画像を利用し、さらに Open Pose[27]の姿勢の可視化を用いて、ポーズ情報を可視化したものである。研究室内でよく見られる生活行動として「コンピュータの使用」、「スマートフォンの使用」、「友人との会話」、「居眠り」という4つの行動タグを用意し、これらを認識させる。実験では、4人の被験者にこれら4つの行動に対応するような振る舞いを実行するように依頼する。まず、4人の被験者の動きをRGBカメラで30fpsで撮影を行った。次に、取得した動画データを3fpsの画像に変換し、そこから2つの異なるデータセットを作成した。そして、この2つのデータセットを用いてモデル作成し評価を行った。その結果、このニューラルネットワークは限られた空間での人間の生活行動を行う際

は、動画画像データの連続性に依存することなく、認識することが可能であるということが実証された。

4.2 関連研究

人間同士のコミュニケーションを機械に分析させるにあたって、主に用いられる技術は機械学習である。感情を含む人間の状態を機械に認識させるために用いられるデータは、主に画像や動画、音声や心拍数などであり、機械のコミュニケーション能力の向上のための方法としては、それらを機械学習させることが一般的な手法である。その中でも、画像認識において、特徴量を深いレベル、かつ高い精度で取得可能であるディープラーニングの技術が最も活用されている。画像認識の分野では主に畳み込み層とプーリング層で構成された畳み込みニューラルネットワークが優れた性能を発揮している。代表的なものとしては、Redmanら[26]が発表したYOLOという画像認識のアルゴリズムやLiuら[28]が発表したSSDの研究が挙げられる。それらはどちらも物体検出を行うためのアルゴリズムとして提案がされている。今までのR-CNN[29]やFast R-CNN[30]では、画像における似た特徴を持つ小さい領域を統合させていくことで、物体が存在しそうな領域に当たり判定をつけるという選択的検索法が用いられていた。しかし、この手法では実行時間が長く、パラメータの設定が複雑であるという問題があった。そこで登場したYOLOやSSDでは、画像をグリッド分割し、それぞれのグリッドに対して固定された複数のバウンディングボックスから当てはまるものを探すという手法が用いられた。YOLOでは、画像認識を回帰問題とし、画像の領域推定と分類を同時に行うことが実現された。さらに、YOLOのアルゴリズムは1つのCNNで完結するため、シンプルであり、処理の高速化が実現した。また、YOLOでは画像全体の情報から学習や検証を実施することができるので、誤検出がFast R-CNNの半分以下となった。SSDでは、Extra Feature Layersという畳み込み層を挿入することで、ニューラルネットの後段に向かうほど特徴マップの分割領域数を少なくさせることが可能となった。他にも、2017年に行われた画像認識コンペImageNet Large Scale Visual Recognition Challenge(INLAVRC)で優勝した技術がある。ISCASの研究である[31]。通常のCNNの畳み込み層では、入力チャンネルが畳み込まれた後に、全て均等にチャンネルが出力されるが、ISCASが提案したSENNetsは畳み込み層の各チャンネルを均等に出力せず、適応的

に重みをかけていることが特徴である。このように物体検知や画像認識の分野において、CNNの技術は次々に新しい技術が開発され、改良され、より高速かつ認識精度の高いものが発表されている。

そして、この物体検知の技術を応用して、2Dの画像から3D物体を検知する技術もいくつか存在する。それは、2DのRGB画像から3Dの物体検知を行うLahoudら[32]の研究である。手法としては、まず、Faster R-CNNを利用し、2Dの画像から物体検出を行う。次に、3DのデータをMFE手法によってオリエンテーションを推定する。そして、物体間の関係をRGB-D画像から物体の3Dバウンディングボックスを作成するという手法を提案した。この研究は、SUN RGB-Dデータセットによって高い精度と最も早い処理速度を示した。処理速度が1枚の画像に対して4.1秒であり、これは、従来の3D物体検出の処理速度に比べて3倍速い。Lahoudらの研究によって、2Dでの物体検出を3Dデータに利用することの有効性が示された。次に、Martinezら[33]の研究がある。この研究は、2Dの姿勢から3Dの姿勢に変換する手法を提案した。そのネットワークはシンプルに構成されていて、2Dの座標を入力することで3Dの座標が出力される。このネットワークでは、画像ではなく関節の座標点を学習に用いるので、軽量のモデルで実現することができる点の特徴である。また、Residualコネクションやbatch normalizationやドロップアウトを駆使し、軽量化・高速化が実現された。3Dの関節座標の情報も含むHuman3.6Mのデータセットを用いて精度の評価を行ったところ、エラー率は30%と、今まで提案されてきたモデルに対して最高水準を記録した。このように、2Dの画像から人間の姿勢を推定する試みや、3Dの物体を検知する手法が発表されている。

動画における人間行動の認識においては、大和ら[34]が、画像中の人物領域のメッシュ特徴をベクトル量子化によりシンボルに変換し、隠れマルコフモデルを用いて学習させる手法を提案している。これにより、教師付き学習による動作カテゴリー認識が実現した。それを用いてテニスの動作を例に、複数の被験者の動作認識実験を行った結果、90%以上の認識率が得られた。また、産業技術総合研究所の小川ら[35]の、リアルタイムに動画から異常値を検出する技術を検討した研究がある。この研究は、Jubatus というオンライン機械学習向け分散処理フレームワークを主に用いている。まず、小川ら[35]は、文字列しか扱うことができない Jubatus の特徴ベクトルコンバータを拡張し、画像や動画などの

マルチメディアデータを扱うことを可能とした。Juatus は、外れ値の検知する jubaanomaly という検知器を提供しており、小川ら[35]はこれを主に利用する。この jubaanomaly は、クライアントから特徴ベクタを add すると特徴点の追加を行うとともにその点の異常度を算出して返し、clac_score することで特徴点の異常度を算出して返すことが可能となる。この機能を用いて、ビデオカメラから得た画像をグレースケール化しフレーム間の差分と自動閾値選定による二値化処理を行う。次にその二値差分画像から CHLAC 特徴を算出する。そして、特徴ベクタを因数に jubaanomaly に clac_score を行い、特徴ベクタの異常度を算出するというシステムとなっている。このシステムを評価した結果、学習データに基づいた異常値検知はリアルタイムに実現できたが、学習フェーズは学習データの増加とともにリアルタイムに処理することに問題点が残った。

近年のインターネットの発達により、画像や動画像、それに付随するタグなどの情報が容易に入手可能となり、様々な研究に応用されている。こうした状況の中、機械学習に用いるデータの量は多く、質も高いものが求められる。画像のアノテーション性能に関してもデータセットに高い質が求められている。これに関して、これまでの機械学習を用いた人間の行動認識手法やデータセットを様々な角度から分析したカーネギーメロン大学の Gunnar ら[36]の研究がある。彼らによれば、今まで提案されてきたデータベース KTH[37] や UCF101[38]や HMDB51[39]や Charades[40]は、活動の時間に違いがあるが、現在のデータセットとしては効果的なベンチマークを可能にすることが示された。またそこには、機械学習を用いて人間の行動認識をするにあたって今後の研究に必要なことが述べられていた。それは、機械が画像に含まれた個々のオブジェクトや人間のポーズレベルで、動画像に描かれているシーンを認識する必要があるという意見である。また、人間が動画像に対してアノテーションを行う際に、そこでのタグ付けが曖昧であったり、人それぞれ意見が異なっていたりしていても学習には影響がないということがわかった。そこで、本研究では、Openpose[41]の技術を利用して動画像にアノテーションを加える。これにより、動画像に含まれる人物の動きの特徴を可視化させることが可能となる。また、この研究の最終目標として動画像データに含まれた人間の行動情報のみだけでなく、周囲の状況など

の分析も可能にしたい。そのため、連続データの時間という部分を拡張するのではなく、入力画像のアノテーションやデータにどのような注釈をつけて読み込むかという点がポイントとなってくると考える。したがって、本研究では、画像の特徴を抽出することを得意とする畳み込みニューラルネットワークを用いて機械に学習させる。

4.3 本研究に用いるニューラルネットワーク

4.3.1 ニューラルネットワークの構成

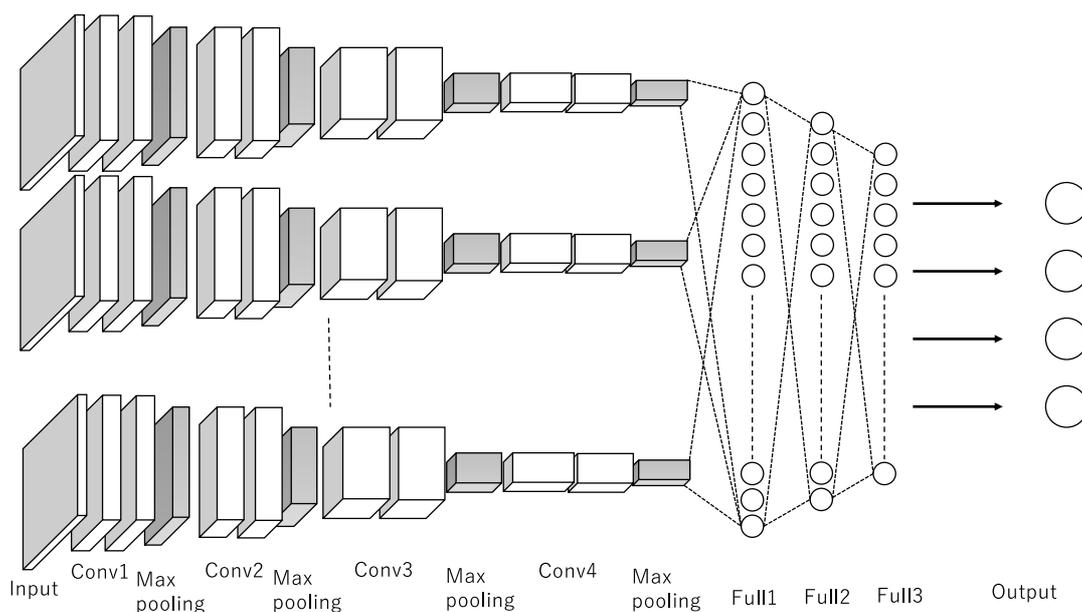


図 15 本研究で提案するニューラルネットワーク

本研究で用いられるニューラルネットワークは、入力層、畳み込み層 4 層とそれぞれに隣接するプーリングが 4 層、全結合層 3 層から構成されている(図 15)。特徴としては、畳み込みニューラルネットワークが複数ある点である。これによって、動画データにおける連続する画像を複数枚入力することが可能となる。この畳み込みニューラルネットワーク

の数を N とする。例えば $N=3$ の場合は、このネットワークに入力されたデータを、3つの畳み込みニューラルネットワークの出力を1つのデータとして結合し全結合層へと入力し、最終的に、4つのラベルを出力する。畳み込みニューラルネットワーク内のニューロン活性化関数として ReLU を設置している。ReLU は、畳み込みフィルターと全結合層の後に画像データに対して活性化関数を適応させるという役目を果たす。つまり、これは 0 未満の出力値を全て 0 にする関数であり、ある閾値以上の部分だけを意味のある画像として次の層に送り、抽出された特徴をより強調する働きがある。さらに、これらの ReLU に加えてオーバーフィッティングを避けるためにドロップアウト層も全結合層の後に設置している。この層は、過学習を防ぐ働きがある。以下の各節では入力層、畳み込み層、プーリング層、全結合層、出力層のそれぞれの層についての詳細と機能を論じる。

4.3.2 ニューラルネットワークの実装

本研究では、Chainer を用いてニューラルネットワークを実装した。Chainer は、Preferred Networks が開発した、ニューラルネットワークを実装するためのフレームワークである。Chainer の特徴として「高機能」「柔軟性」「直感的」の3点があげられる。インストールが簡単なことや GPU による演算に対応していること、柔軟な記法で様々な事例に対応できることにより、Python の深層学習のフレームワークとして幅広い分野で利用されている。Chainer は“Define-and-Run”という仕組みを採用している。これは、順伝播計算を行うときに初めてネットワークを定義するというものである。ニューラルネットでは、Weight という結合の重みと Bias の2つのパラメータを使って目的の値を出力する関数に近いものを作り、この目的関数を目印にパラメータを更新していくことがニューラルネットの学習を意味する。Chainer では、これらのパラメータを持ったニューラルネットの関数を Link と呼び、import の際に L という名前指定が必要がある。また、入力された信号を受けて活性化関数を使う。このような関数を Chainer では Function としてまとめていて、これも import の際には F という名前指定が必要がある。そして、これらの

Link と Function をまとめて管理するのが Chainer であり、これを継承することによってニューラルネットを実装する際に便利な機能を様々な使用することが可能となる。また、Chainer では、Variable という機能を使うことでニューラルネットにデータが入力されてから出力を返すまでの計算の流れ(フォワードプロパゲーション)を計算グラフとして記録し、これを逆に辿ることで自動的にパラメータを更新している。そのため、畳み込みニューラルネットワーク、フィードフォワードニューラルネットワーク、リカレントニューラルネットワークなど様々なネットワークに対応しており、画像処理や自然言語処理、ロボット制御など多岐に渡る分野で使われている。

4.3.3 入力層

入力層では、入力された RGB からなる 3 チャンネルの画像を学習データとして、これを 128×128 のサイズにリサイズする。本論文で提案するニューラルネットワークは入力層の数を、入力する画像の数によって複数用意することが可能である。ここで、入力層の数は N とする。例えば、連続した 3 枚の画像データを入力する場合は、入力層が 3 つ用意されるので、 $N3$ とする。また、連続した画像の枚数が 5 枚、7 枚となれば、 $N5$ 、 $N7$ と表現することができる。本論文の学習データは、取得した動画データ全てのデータにあらかじめラベル付けを行い、それらの 80% をランダムに選出し、残りの 20% はテストデータとして使用する。

4.3.4 畳み込み層

畳み込み層では、畳み込みフィルターと呼ばれる小さな特徴検出器の集合体を用いられる。この畳み込みフィルターを用いて、重み付き和の計算を画像の部分ごとに行う。2 次元配列である画像のデータ $x(i, j)$ があるとした時、 i, j が画像のピクセルの位置を指定し、そ

の位置 (i, j) における画素値を $x(i, j)$ とする。そしてこのデータに対して畳み込みフィルタ $h(p, q)$ を適用する場合、計算は以下ようになる。

$$u(i, j) = \sum_m \sum_n x(i + m, j + n)h(m, n) \quad (3)$$

$u(i, j)$ は処理が施された後の画像の位置 i, j の画素値を示す。畳み込みフィルタは、特徴がどこにあっても抽出することができる移動不変性の性質をもち、畳み込みフィルタは a 値もまた学習により獲得する。本論文で用いた畳み込み層は4層存在し、それぞれ図1では、Conv1、Conv2、Conv3、Conv4と表す。Conv1では特徴マップのピクセルを 128×128 と設定し、Conv2では 64×64 、Conv3では 32×32 、Conv4では 16×16 とする。それぞれのフィルタのチャンネル数は3に設定している。

4.3.5 プーリング層

プーリング層は、特徴マップの解像度を下げるために使用される。この処理によって、入力画像よりサイズが小さくなってしまいが、この層では、情報が失われることを最小限にし、データの量を減らす。また、位置のずれに対してもある程度の不変性を実現できる。他にも、ある程度の過学習も抑制するという特徴などがあげられる。本研究では、 2×2 のピクセルの最大値をとって結合するMax poolingを採用し設定した。このMax pooling層を畳み込み層4層にそれぞれ隣接させ、畳み込み層とプーリング層でデータの特徴を抽出する。

4.3.6 全結合層

全結合層では、画像の識別を行うために特徴量に基づいた分類を行う。畳み込み層とプーリング層を通して特徴部分が取り出された画像データを1つのノードに結合し、活性化

関数によって変換された特徴変数を出力する。ノードの数が増えると特徴量空間の分割数が増え、各領域を特徴付ける特徴変数の数が増える。この特徴を持った全結合層を本研究では、Full1、Full2、Full3の3層を用意した。Full1のニューロン数はChainerが出力してくれるため、自ら設定しなければいけないのはFull2、Full3のニューロン数のみとなる。したがって、Full1では12000のニューロンを持ち、Full2は4096、Full3では1024のニューロンを持つように設定した。

4.3.7 出力層

出力層では、全結合層から特徴変数を元に、ソフトマックス関数を用いて確率に変換し、それぞれの領域に正しく分類される確率を最大化することによって分類を行う。ソフトマックス関数は分類問題をニューラルネットワークで解く場合に、出力層の活性化関数として用いられる関数であり、数式は以下で表される。

$$y_i = \frac{\exp(a_j)}{\sum_j^D \exp(a_j)} \quad (4)$$

ここで y_i というのは出力層の i 番目のユニットの出力、 D は出力層のユニット数を示す。分母に和の式を導入することで、出力層の全てのユニットの出力の和をとると、

$$\sum_j^D y_i = \frac{\sum_j^D \exp(a_i)}{\sum_j^D \exp(a_i)} = \mathbf{1} \quad (5)$$

となる。

これを出力層の活性化関数に用いることで、ニューラルネットワークの出力がクラスに属する確率を算出し、それを学習させることが可能となる。本論文では、0から3の4つのクラスに分類する。

4.4 データセット

4.4.1 動画像データの取得



図 16 撮影の様子

撮影には、iPad を用いる。本研究では、研究室における人間の状態認識を行うため、被験者は研究室のデスクに着き、そこでの様子を撮影する。カメラは被験者の側面に配置し、被験者の頭から足先の全身が映るように設置する。図 16 に示す。今回は iPad Pro を三脚に固定し、被験者の撮影を試みた。カメラは設置された位置から被験者の研究室での作業様子を一定時間撮影し、30fps で動画像データを取得する。そして、動画像データの中から特徴が現れた部分を選択する。本研究では「コンピュータの使用」、「スマートフォンの使用」、「友人との会話」、「居眠り」という 4 つの部分トリミングし保存する。トリミングには、iMovie を利用した。

4.4.2 動画像データの変換

RGB カメラから取得した 30fps の動画像データを 3fps に変換する際には、OpenCV[27]の技術を用いる。OpneCV[27]は、Intel によって開発されたライブラリであり、コンピュータで画像や動画を処理するために必要な機能が様々に実装されている。また、Python や C 言語、C++によって記述することが可能であり、Windows、Linux など複数のプラットフォームに対応している。さらに BSD ライセンスに基づくオープンソースソフトウェアとして提供されているため、商用・非商用問わず誰もが無償で利用することが可能である。OpenCV を利用することで、物体の認識や、パターン認識、フィルター処理や行列演算、カメラのキャリブレーションなどコンピュータビジョンに関する高度な画像処理機能を容易に利用することが可能となる。本研究では Python と OpenCV を用いて、カメラから得られた動画像ファイルのフレームレートとフレーム数を求めるプログラムを使い、出力されたフレームを 10 枚飛ばしで保存するシステムを構築した。これにより、1 秒間 30 の画像データを 3 枚の画像データとして変換する。飛ばすフレーム数を変更することで多様なフレームレートを容易に指定できる。

4.4.3 データセットの作成

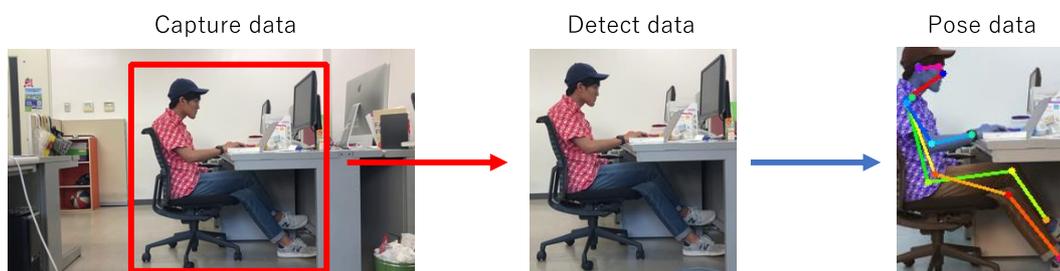


図 17 データセット

本研究では、YOLO を用いて人物領域を切り出したデータセット (Detect dataset) と、OpenPose を用いた人間のポーズ推定情報を可視化させたデータセット (Pose dataset) と 2 つのデータセットを作成した。4 人の学生 (男 : 2、女 : 2) に協力してもらい、研究室でデスクに向かって作業している様子を撮影した。これら 2 つのデータセットを作成した理由は 2 つある。理由の 1 つ目は、人間の姿勢データを可視化させた方が、動画データに含まれる服装や顔などの情報に影響されることなく、人間の行動認識が可能となるという仮説を立てたからである。また、2 つ目の理由としては、その仮説の検証を行う際に、比較となる対象が必要であったためである。

4.4.4 Detect dataset

図 3 に示したように、Detect dataset は画像データから人物領域のみを抽出する。その際、動画データを 1 秒 3 枚の画像データとして変換し、その画像データから YOLO というオブジェクト検出の技術を用いて人物領域を抽出する [41]。人物検出の閾値は 0.6 と設定し、検出を行う。データセットには男女 4 人の被験者のすべてのデータ約 3,000 枚の画像が含まれている。そのうち、1 人の約 750 枚の画像データはすべてテストデータとして用いる。それ以外 3 人の約 2,300 枚の画像のうち 80% を学習データとして、残りの 20% のデータをテストデータとする。

4.4.5 Pose dataset

図 3 に示したように、Pose dataset は Detect dataset の画像データから人物領域のみを抽出した画像に、さらに人物のポーズを OpenPose を用いて可視化させる。ここでは、OpenPose の技術を応用した Zhe [41] の技術を用いる。これは、2D の静止画や動画において人物の姿勢推定を行うモデルである。入力画像から部位の位置を推定し、その部位との関

連を表す 2D ベクトルを算出し、その 2 つの情報から同じ人物の部位を組み合わせ、姿勢情報を出力するというシステムになっている。Detect dataset の場合と同じようにこのデータセットにも男女 4 人の被験者のすべてのデータ約 3,000 枚の画像が含まれている。そのうち、1 人の約 750 枚の画像データはすべてテストデータとして用いる。それ以外の 3 人の約 2,300 枚の画像のうち 80%を学習データとして、残りの 20%のデータをテストデータとする。

4.5 実験

4.5.1 実験概要

実験では、被験者 4 人の画像データから生成した Detect dataset と Pose dataset の 2 つのデータセットを用いる。どちらのデータセットの場合も、被験者 4 人のうち被験者 1 人の画像データは、モデルの学習時に用いずテストデータとしてのみ使用することとする。よって、被験者 3 人のデータセットを学習データとテストデータの 2 つに分け、残りの被験者 1 人のデータセットを、モデルの精度評価時のみに使用することとする。学習時には、学習データのうちの 80% の画像をランダムに学習し、残りの 20% を学習モデルの検証データとして用いる。そして、モデルの精度を評価する際には、テストデータを用いる。

4.5.2 実験方法

本研究のニューラルネットワークは、連続した画像データを入力するため、入力層を複数もつ。この入力層の数を 1 (N1)、3 (N3)、5 (N5) とするモデルをそれぞれ 5 個作成する。学習時のパラメータは、エポック数を 300epoch、バッチサイズを 100 として、これを 5 回繰り返す。そして、その 5 つのモデルを用いて、テストデータに対する正解率、適合率、再現率、F 値を計算し、それらの平均値を結果とする。本実験環境(表 3)は以下に示す通りである。

表 3 実験環境

GPU	NVIDIA : GTX 1080	RTX 2080
CPU	Core i7-9700K	Core i7-7700K
メモリ	3.6GB	4.2GB

4.6 結果

4.6.1 評価方法

まず、2値分類の予測をする際の指標について説明する。実験結果として、テストデータに対して、Accuracy(正解率)、Precision(適合率)、Recall Rate(再現率)、Fmeasure(F値)の4つの評価尺度でモデルが出力した予測結果の精度を比較する。

表 4 評価指標

		真の結果	
		正	負
予測の結果	正	TP	FP
	負	FN	TN

まず、正解率を示す Accuracy は、予測結果全体と答えがどの程度一致しているのかを判断する指標であり、計算式は以下のように示す。

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

次に、適合率を示す Precision は、正と予測したデータのうち、実際に正であるものの割合のことであり、計算式は以下のように示す。

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

再現率を示す Recall_rate は、実際に正であるもののうち、正であると予測されたものの割合のことであり、計算式は以下のように示す。

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

そして、適合率と再現率はどちらかが高くなると、どちらかが低くなるというトレードオフの関係にあるので、調和平均を取った指標を F 値といい、計算式は以下のように示す。

$$F_{measure} = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision} \quad (9)$$

これらの式に基づいて計算された結果を各パターン、クラス別に分類し以下の表に示す。

4.6.2 被験者 3 人のテストデータに対する結果

表 5 被験者 3 人データに対する推定結果

Detect dataset						Pose dataset					
	class	accuracy	precision	recall_rate	Fmeasure		class	accuracy	precision	recall_rate	Fmeasure
N1	0	1.00	0.99	1.00	0.99	N1	0	1.00	1.00	1.00	1.00
	1	0.98	0.95	0.98	0.97		1	0.96	0.98	0.87	0.92
	2	0.93	0.97	0.74	0.84		2	0.96	0.88	0.98	0.93
	3	0.95	0.83	0.99	0.90		3	1.00	1.00	1.00	1.00
N3	0	1.00	0.99	0.99	0.99	N3	0	1.00	1.00	0.99	1.00
	1	0.99	0.96	0.99	0.98		1	0.97	0.96	0.92	0.94
	2	0.93	0.98	0.73	0.83		2	0.97	0.92	0.96	0.94
	3	0.94	0.80	0.99	0.89		3	1.00	1.00	1.00	1.00
N5	0	0.99	0.98	0.99	0.98	N5	0	1.00	1.00	0.98	0.99
	1	0.96	0.92	0.92	0.92		1	0.98	0.97	0.96	0.96
	2	0.90	0.89	0.71	0.79		2	0.98	0.95	0.97	0.96
	3	0.94	0.81	0.98	0.89		3	1.00	0.99	1.00	0.99

表 6 CNN の数別の推定結果

A.Detect data

	accuracy	precision	recall_rate	Fmeasure
N1	0.98	0.96	0.96	0.96
N3	0.98	0.98	0.97	0.97
N5	0.99	0.98	0.98	0.98

B.Pose data

	accuracy	precision	recall_rate	Fmeasure
N1	0.96	0.93	0.93	0.92
N3	0.96	0.94	0.93	0.92
N5	0.95	0.90	0.90	0.89

被験者 3 人の Detect dataset と Pose dataset の 2 つのデータセットを用いて、モデルを作成し、テストデータに対して、それぞれの精度を示したものを表 5 に示す。精度が高い物は赤色のセルとなっており、精度が低いものが黄色セルとなっている。表 6 には、クラス 0～3 の値を平均した精度評価を表にまとめた。赤色のセルが精度評価を高いものを示している。Detect dataset では、4 クラス全てにおいて、Accuracy が 90%以上を示す評価となり、非常にモデルの精度が良いことがわかる。表 6 の平均値をまとめた表を見ると、特に入力層 5 つを用いた場合の N5 のモデルが最も精度が高いことがわかる。Pose dataset で生成したモデルの場合、入力層が 1 つの N1 のモデルが最も精度が高いという結果になった。2 つのデータセットを比較すると、Detect dataset を用いて学習したモデルの方が Accuracy、Precision、Recall_rate、F_measure のすべての精度評価において高い能力を発揮していることがわかる。

4.6.3 未学習の被験者データに対する推定結果

表 7 未学習の被験者データに対する推定結果

Detect dataset						Pose dataset					
	class	accuracy	precision	recall_rate	Fmeasure		class	accuracy	precision	recall_rate	Fmeasure
N1	0	0.724	0.629	0.248	0.340	N1	0	0.827	0.888	0.288	0.396
	1	0.674	0.136	0.010	0.018		1	0.709	0.186	0.039	0.062
	2	0.454	0.372	0.860	0.518		2	0.339	0.225	0.654	0.333
	3	0.895	0.000	0.000	0.000		3	0.628	0.047	0.012	0.020
N3	0	0.729	0.579	0.504	0.537	N3	0	0.832	0.912	0.317	0.441
	1	0.678	0.112	0.006	0.011		1	0.718	0.551	0.108	0.163
	2	0.624	0.479	0.889	0.620		2	0.434	0.291	0.750	0.417
	3	0.895	0.001	0.008	0.002		3	0.730	0.277	0.221	0.229
N5	0	0.676	0.488	0.445	0.455	N5	0	0.783	0.529	0.126	0.199
	1	0.635	0.344	0.063	0.082		1	0.767	0.441	0.075	0.121
	2	0.682	0.561	0.854	0.664		2	0.271	0.230	0.741	0.351
	3	0.908	0.000	0.000	0.000		3	0.674	0.114	0.016	0.028

表 8 CNN の数別の推定結果

A.Detect data				
	accuracy	precision	recall_rate	Fmeasure
N1	0.626	0.336	0.248	0.203
N3	0.679	0.508	0.349	0.312
N5	0.624	0.329	0.240	0.175

B.Pose data				
	accuracy	precision	recall_rate	Fmeasure
N1	0.687	0.284	0.280	0.219
N3	0.732	0.293	0.352	0.292
N5	0.725	0.348	0.340	0.301

この節では、学習時に用いなかった被験者 1 人のデータをすべてテストデータとして用いて、被験者 3 人のデータを用いて学習したモデルの精度を評価する。表 7 には、すべてのクラスについて、それぞれの Accuracy、Precision、Recall_rate、F measure をまとめる。赤色のセルが数値が高く、緑色のセルが低い数値となっている。全体的に、学習時に用いなかったテストデータで評価すると、認識精度は低いことがわかる。Detect

dataset では、ラベルに対する正解率は、約 60%という数値になり、作成したモデルは高い汎用性があるとは言えない結果になった。しかし、Pose dataset に関しては、正解率は約 70%となっており、ポーズを可視化させた場合の方が約 10%正解率が高いことがわかる。これは、画像データに含まれた人間の情報における服装や体型や顔に依存することなく、ポーズデータを可視化させることが人間の行動認識において役に立っているということである。また、表 8 には、0~3 のクラスの平均値を示したものを表にまとめた。白色のセルが精度が低いことを示し、赤色のセルが精度が高いことを示している。この表を見ると、Pose dataset を用いた入力層 5 つ、つまり、N5 のモデルが最も高い精度を示した。

4.7 まとめ

本論文では、人間の行動を認識することを目的として、入力層を複数持つニューラルネットワークを提案した。研究室で人間の行動認識をする場合、映像から人物を切り出し、連続する 5 枚のデータを学習させることで、高い性能を持つモデルを作成することができるということがわかった。しかし、動画データを取得する際は、画像データに衣服の色や体型、顔といった情報が含まれてしまう。このような場合であっても、画像に含まれた色や人物に依存することなく、様々な人間がとる行動を機械が認識するには、人間のポーズ情報を可視化させた方が有効であると考えられる。実際に、学習時に用いなかった 1 人のテストデータに関して精度評価を行った実験では、Pose dataset の方が Detect dataset よりもわずかに認識精度が高いという結果が出た。また、この場合においても入力層を 5 つ用いた N5 で作成したモデルの精度が高いという結果が出ている。これを踏まえると、筆者が立てた仮説が支持されたのではないかと思われる。このようにして、本研究が提案したニューラルネットワークでは、少ない動画データから、人間の行動を認識することが可能となった。今後としては、今のところ、人間の行動を動画データで取得する際に 1 つの角度から行っているが、側面に加えて、正面、背面と様々な角度から動画データを取得し、学習時に用いたいと考えている。また、モデルの汎用性を高めるために、Pose

dataset を用いて、学習に用いなかった人物に対しても高い認識精度を出せるように検討する必要がある。

5. 結論

本論文では、人間同士の説得というコミュニケーションに介在する擬人化エージェントの必要な要素について研究を進めた。擬人化エージェントが人間のコミュニケーションに介在するにあたり、必要な要素として、擬人化エージェントに対する認知性、親密性、擬人化エージェントの行動の社会性、自律性などが挙げられる。これを踏まえて、説得というコミュニケーションに介在する擬人化エージェントに必要な機能として、ユーザが擬人化エージェントに親密性を持つという点に着目し研究を行った。また、この点に着目した理由としては、人間同士のコミュニケーションにおいて、人間関係を構築する、相手に親密性を感じるには時間を要し、人間関係を構築する過程で心理的負担が生じる可能性が高いからである。そこで、擬人化エージェントに親密性を感じることで、相手を注意するコミュニケーション、説得するコミュニケーションを人間の代わりに擬人化エージェントが行うことができると考えた。そして、この擬人化エージェントを設計するための段階として必要となる、2つの点について検討した。第一に、ユーザが擬人化エージェントの容姿に対して、親密性や信頼性を持たせること、第二に、一方的な説得を防ぐために、擬人化エージェントがユーザの暗黙的な動作や状況を察知して、それに合わせた対応をするという技術として、その第一段階である人間の動作認識理解をすることである。

まず、第一のユーザの意見を変容させるような働きかけを行う擬人化エージェントの容姿とコミュニケーション形式については、レギュラーコミュニケーションとオーバーハードコミュニケーションの比較についてはレギュラーコミュニケーションがより説得力があるという結果であった。しかし、エージェントパターン別に比較すると、レギュラーコミュニケーションの中で最も説得力があった M2 のパターンより、OHC の中で最も説得力があった M2 が F6 を説得しているパターンの方が、より高い説得力があることを示した。容姿に関する具体的な要素は明らかにするため、メガネの有無で違いを持たせたエージェント 4 体の比較を行ったが、メガネをかけた男性のエージェントが最もユーザの映画に対する鑑賞意欲で高い評価を受けたが、メガネの有無、パターン別に有意な差は見られなか

った。これらを踏まえて、擬人化エージェントがユーザを説得するには、擬人化エージェントの容姿も含め、ユーザに信頼性を与えることが重要である可能性が高い。エージェントの信頼度が高いと意見が肯定的に変容するという傾向から、ユーザに信頼を与える要素についてさらに検討する必要がある。

次に、第二の擬人化エージェントがユーザの動作認識理解を行うための技術についての結論を述べる。人間の行動を認識することを目的として、入力層を複数持つニューラルネットワークを提案した結果、研究室内で人間の行動認識をする場合、映像から人物を切り出し、連続する5枚のデータを学習させることで、高い性能を持つモデルを作成することができるということがわかった。したがって、ユーザの動作を認識理解する機能として十分に利用できることが明らかになった。

以上より、研究により、説得というコミュニケーションに介在する擬人化エージェントとして必要な2つの技術について、実現することができた。

付録 A

映画の説明文²

コメディ映画

1. 『なんちゃって家族』

ちまちまとマリファナ売人をやっているデヴィッドは独身で、自由きままに暮らしていた。しかし、路上でもめていたホームレス少女を助ける際、逆にギャングたちにありったけの麻薬と稼いだお金を奪われてしまう。組織のボスに呼び出されたデヴィッドは全てをチャラにする代わりにブーツをメキシコから密輸する、と命じられる。デヴィッドは疑似家族の観光を装うことで怪しまれることなく越境できると思いつき、貧乏ストリッパー、冒険に夢見てる童貞君、ゴスメイクなホームレス娘らちょっとワケありの隣人たちを巻き込んでなんちゃって家族を結成。キャンピングカーに乗り込んで密輸の旅に出るのだが…。

2. 『おバカんす家族』

主人公のラスティは、妻デビーと息子 2 人を持つ父親。家族生活のマンネリをバカンスで解消し、家族の絆を取り戻そうと一念発起したラスティは、新たな家族旅行を計画。最新式 (!?) の新車を買ひ、妻と息子を連れ、国民的ファミリー・テーマパーク、ワリー・ワールドを目指してアメリカ大陸横断の旅に出発する！ しかし、楽しい家族旅行になるはずが、行く先々で“お！バカすぎる”トラブルが続出。ラスティは家長の威厳と家族愛を取り戻せるのか!?

3. 『噂のモーガン夫妻』

マンハッタンで数千億円の物件ばかりを扱う不動産のトップセールス・ウーマン、メリル・モーガンと、全米でも 10 本の指に入る敏腕弁護士のポール・モーガンは、誰もがうらやむ完璧な超セレブカップル。だが、ポールの浮気が発覚し、メリルの気持ちはすっかり冷め切ってしまっていた。そんなある日、2 人は図らずも殺人事件に遭遇してしまう。犯人に顔を見られた 2 人は、証人保護プログラ

² 映画の説明文は以下のホームページを参考に作成

(<https://eigahitotobi.com/article/63195/>)(<https://curazy.com/cinema/archives/1331>)

ムにより身分を隠してワイオミングへ送られる。人間よりも牛や馬のほうが多いド田舎で2人きりになったこの夫婦は、この先いったいどうなるのか…？

歴史映画

1. 『カティンの森』

舞台は1939年に結ばれた独ソ不可侵条約を背景に、ドイツとソ連に分割統治されることになったポーランド。アンナと娘・ニカは、一緒に逃げようと夫・アンジェイを探して奔走していました。

駅でソ連へ連行される間際のアンジェイを見つけますが、アンジェイは大尉という立場で軍に忠誠を誓っているため逃亡を拒否します。その後、捕虜として教会に収容されるアンジェイですが、見聞きしたすべてを日記に書くことを決めます。1943年、ドイツ軍によってカティンの森近くで1万人以上のポーランド将校の死体が発見されました。アンナは、事件の真相をアンジェイの記した日記によって初めて知ることになります

2. 『グッバイ！レーニン』

東ドイツに住んでいたアレックス。母・クリスティアーネは、熱狂的と言えるほどに社会主義を支持しています。それはアレックスの父・ローベルトが西ドイツに亡命したことがきっかけでした。そんなクリスティアーネは、東ドイツ建国記念日にアレックスが反体制デモに参加しているのを見て、ショックのあまり心臓発作を起こしてしまいます。

その後、奇跡的に目が覚めたクリスティアーネでしたが、すでにベルリンの壁は崩壊。次に大きなショックがあった時にはどうなるか分からない、と医師に宣告されてしまいます。アレックスは、必死で東西ドイツ統合の事実をクリスティアーネから隠そうと奔走することになるのです。

3. 『トゥルー・ヌーンーイワノビッチ村』

舞台は現・タジキスタンの小さな2つの村。ソ連の気象観測所に勤めており、村に赴任してきているロシア人・キリルを含め、上の村人と下の村人は平和に暮らしていました。しかし2つの村の間に突

如、鉄条網が引かれてしまいます。ソ連が崩壊したことをきっかけにロシアとタジキスタンの国境ができてしまったのです。活発に行き来していた2つの村の住人は、勝手に決められてしまった国境に戸惑います。キリルの助手・ニルファはもう一つの村に嫁に行くことになっていました。

キリルと村人たちは何とか村の行き来ができないかと考えるのですが、ロシア軍による地雷の設置など、様々な困難が立ちまわります。

犯罪映画

1. 『逃亡者』

シカゴで働く優秀な外科医リチャード・キンブル。ある夜帰宅すると妻が何者かに襲われて、殺されそうになっていた。慌てて犯人を追い払うが、妻の状態はもう手遅れで死に瀕していた。しかしキンブル医師は状況的に濡れ衣を着せられ、妻殺しの犯人として逮捕されてしまう。彼は無実が証明できずに死刑判決を受け、刑務所に移送されることに。しかし護送車の中で他の囚人たちが暴れ出し、車が横転した際に彼はうまく逃げ出す。逃亡者が出たことに気づき、追跡を始める連邦保安官。自身の無実を証明するために逃亡を続けるキンブル医師は、真犯人にたどり着けるのか?! 決死の逃亡劇が幕を開ける。

2. 『その男、身も心も崖っぷち』

ニューヨークの高層ビルの21階、窓の縁に立つ男、それはニック・キャッシュ。4千万ドルもの価値を持つダイヤモンドを盗んだ罪で、刑務所に入れられていた元刑事だった。父の葬儀の最中に、騒ぎを起こして逃げてきた。窓の縁に立つ男を見て通りの人が騒ぎ出し、付近ではすでに大騒ぎになりはじめていた。元刑事であり犯罪者であるニックは、なぜそこに立つのか?! 単純に自殺?? いや、ニックには果たさなければならない目的があったのだ。果たしてその目的とは、ニックの運命はどうなるのか?!

3. 『レザボア・ドッグス』

宝石強盗をするために集まった、6人と2人。ダイナーで朝食をとりながらあいもない話をしていく。実行犯は6人。その6人は葬式でも行くのかのように黒いスーツを身にまとっている。強盗を終

え集合場所に辿り着くオレンジとホワイト。しかし、警察に撃たれてオレンジは負傷している。そこにピンクがやって来ると、警察が来るのが早すぎると、裏切り者の存在を疑うのであった……はたして、他のメンバーはどうなったのか?! ほんとうに裏切り者はいるのか?! 事の顛末は? 男たちの止まらぬ疑いが新たな展開を迎える!!

付録 B

実験で用いたダイアログリスト

パターン	ダイアログ
a	A: この映画は星 4.3 を獲得したんだ A: 映画評論家のアレックスガーナーも最高評価を与えていたよ A: 絶対この映画は見るべきだとオススメするよ
b	A: この映画は、時々眠くなったけど、最後は本当に面白かった A: 映画評論家の中でも、最後のシーンは高い評価を受けてるみたい A: ぜひ、みて欲しい。
c	A: ドキドキする映画は好き? A: この映画はドキドキもするし、妙に納得するストーリーになっているよ A: オススメするよ!!
d	A: この映画、観て本当によかった A: ちょっと理解が難しいところもあるけど、面白かったな
e	A: この監督は去年、賞を受賞したんだよ A: 脚本も良いし、演出もとてもよくできているから、是非観て欲しい!
f	A: この作品は、予告編を観たときちょっと期待してなかったんだけど、 A: 見始めたら、最後まであつという間だったよ A: この映画はよく楽しめたから、是非観て欲しいな。
g	A: この前に、この映画を見たんだけど、本当に素晴らしかったよ B: そうなの? そうは思えないけど A: 最初は暇つぶしに作業しながら、みようかなあと思ったんだけど、いつの間にか集中しちゃったよ! B: そんなにいい映画なの? A: うん、見る価値は大いにあるよ!
h	A: 先日、ラジオを聞いていたときに映画のコーナーがあったんだよね B: うんうん

	<p>A：それで、この映画について話があったんだけど、今年一番の映画だと話していたよ</p> <p>A：そうなの？それはなんで？</p> <p>B：最初から最後まで、人の気持ちをドキドキさせる映画だと話していたよ。この映画を見たけど、見る価値は大いにあるよ。</p> <p>A：そうなんだ。チェックしてみようかな！</p>
i	<p>A：この映画は絶対に見るべきだよ</p> <p>B：映画館に行く時間がないんだよね</p> <p>A：平日にやることをできるだけ終わらせて、休日に行くと良いんじゃない！？</p> <p>B：そうかもね</p> <p>A：支出や時間に見合う価値は必ずあるよ。</p> <p>B：わかった。チェックしてみるよ！</p>
j	<p>A：この映画知ってる？</p> <p>B：今初めて知った</p> <p>A：もし、今観たい映画がないなら1番にオススメするよ</p> <p>B：そうなの？</p> <p>A：ストーリーも俳優も良くて、素晴らしい映画だったよ</p> <p>B：そうなんだ、じゃあ観てみよう！</p>
k	<p>A：いつもどんな映画をみるの？</p> <p>B：ん～あまりジャンルは問わないかな</p> <p>A：それなら、この映画を是非観て！</p> <p>B：そうなの??</p> <p>A：この映画の評価は賛否両論あるけど、それも含めて見るべき映画だと思うんだよね</p> <p>B：そうなんだね～！チェックしてみよう！</p>
l	<p>A：この映画知ってる？</p> <p>B：あんまり映画観ないんだよね</p> <p>A：漫画を読んだり、ドラマを観たりはするの？</p>

B：それは、たまにするよ！

A：じゃあ、長い時間映画を見ることが少し面倒なのかな？、

B：ん～そうかも！、

A：この映画は最初から最後までドキドキ感があるし、きっと最後まで飽きずに見ることができるよ！

B：そっかあ。じゃあ思い切って観てみよう！

謝辞

本研究を遂行し、学位論文をまとめるにあたり、多くのご支援とご指導を賜りました。博士論文を上梓するにあたり、お世話になった皆様方に、この場をお借りして感謝の意を申し上げます。

はじめに、指導教員である村尾元教授には、深く感謝しております。六甲道のバーで出会ってから5年、博士課程への進学および、研究全般にわたる多大なご支援を賜りました。時には厳しくご指導いただいたこと、時には親身になって私の研究と学校生活について助言してくださったこと、心より感謝しております。

次に、常に懇切丁寧なご指導をしてくださった情報コミュニケーションコースの教授の皆様方に御礼申し上げます。大月一弘教授、康敏教授、森下淳也教授、清光英成准教授、西田健志准教授に深く感謝いたします。

また、村中先生をはじめとするキャンパスライフ支援センターの皆様は、いつも笑顔で私を励ましてくださいました。皆様からいただく助言や温かい励ましの言葉はいつも胸に響いて、私の努力の糧となりました。私の学校生活や研究活動をいつも見守って、応援してください、心より感謝しております。

そして、研究活動、学校生活において、様々な助言や励ましを与えてくださった研究室の先輩や後輩、図書館で共に勤務した皆様にありがとうございます。みんなと共に過ごした楽しい時間や真剣に研究について話し合った時間、何気ない世間話をした時間は私の学生生活において、かけがえのない時間だったと思います。

最後に、私を支えてくれた友人、これまで温かい目で見守ってくれた家族にも深く感謝いたします。

皆様本当にありがとうございました。

業績一覧

1. 論文

- ① Megumi Kawata, Hajime Murao “Estimating desk work status from video stream using a deep neural network” ICIC Express Letters, Part B: Applications, 2020

2. 口頭発表

・国際

- ① Megumi Kawata, Hajime Murao “Study on Auditory Representation of Blind Spot Warning using Kinect Sensor”, 廈門台港日四地ワークショップ, 中国, 廈門理工学院, March 28, 2017
- ② Megumi Kawata, Hajime Murao “Study on Kinect-based sonication system for blind spot warning”, IEEE International Conference on Information, Communication and Engineering (ICICE2017), Xiamen, Fujian, China, November 17-20, 2017
- ③ Megumi Kawata, Hajime Murao “Estimating desk work status from video stream using a deep neural network”, 14th International Conference on Innovative Computing, Information and Control (ICICIC2019), Seoul, Korea , August 26-29, 2019
- ④ Megumi Kawata, Hajime Murao “Study on the Effect of Appearance of Personified Agents in Persuasion”, The Society of Instrument and Control Engineers Annual Conference (SICE2020), Online, September 23- 26, 2020

・国内

- ① 川田恵, 村尾元, 「Kinect を用いた日常生活における死角の可聴化に関する研究」, SCI '17-第 61 回システム制御情報学会研究発表講演会, 京都, 2017 年 5 月 23 日
- ② 川田恵, 村尾元, 「3D イメージセンサを用いたビデオ通話向けの効果音生成に関する研究」, SCI'18-第 62 回システム制御情報学会研究発表講演会, 京都, 2018 年 5 月 18

参考文献

- [1] 後藤学, and 大坊郁夫. "大学生はどんな対人場면을苦手とし, 得意とするのか?: コミュニケーション場面に関する自由記述と社会的スキルとの関連." *対人社会心理学研究* 3, p57-63, 2003.
- [2] 木村昌紀, 大坊郁夫, and 余語真夫. "社会的スキルとしての対人コミュニケーション認知メカニズムの検討." *社会心理学研究* 26 卷 1 号, p13-24, 2010.
- [3] 大坊郁夫. "コミュニケーション・スキルの重要性." *日本労働研究雑誌* 48 卷 1 号, p13-22, 2006.
- [4] 長尾確. "マルチモーダルインタフェースとエージェント." *人工知能学会誌* 11 卷 1 号, p32-40, 1996.
- [5] 中野利彦, et al. "Traveling cafe: 分散型オフィス環境におけるコミュニケーション促進支援システム." *インタラククション*, p227-228, 2006.
- [6] 椎尾一郎, and 美馬のゆり. "Meeting pot: アンビエント表示によるコミュニケーション支援." *インタラククション*, p163-164, 2001.
- [7] 山田誠二, 小野哲雄. "マインドインタラククション AI 学者が考える「ココロ」のエージェント." 近代科学社出版, 2019.
- [8] Walster, E., & Festinger, "The effectiveness of "overheard" persuasive communications." *Journal of Abnormal and Social Psychology*, 65, p395-402, 1962.
- [9] 藤原規行, 小林一樹, and 北村泰彦. "マルチエージェント説得におけるエージェント数の評価." *電子情報通信学会技術研究報告. AI, 人工知能と知識処理* 108 卷 382 号, p21-26, 2009.
- [10] 齋藤ひとみ. "擬人化エージェントによるオーバーハードコミュニケーション: 被説得エージェントの反応の違いについての検討." *日本認知科学大会発表論文集* 29 卷, p2-26, 2012.
- [11] Kantharaju, Reshmashree B., et al. "Is two better than one?: Effects of multiple agents on user persuasion." *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. ACM, 2018
- [12] 間瀬健二. "II 擬人化エージェントの生成." 信学会専門講習会「知的エージェント研究の最前線」, p9-15, 1996.

- [13] 長尾圭一郎, 吉田直人, and 米澤朋子. "生活に寄り添い自発行動を促す親近アンビエントエージェントの設計." *情報処理学会, 紀要論文, C-8*, 2015.
- [14] 黒田拓也, 山田誠二, and 寺田和憲. "オンラインショッピングにおける商品推薦エージェントの外見と振る舞いの関係が購買意欲に与える影響." *人工知能学会論文誌* 31 卷 2 号, G-F78_1, 2016.
- [15] Verhagen, T., Nes, van J., Feldberg, F., Dolen, van W., Virtual customer service agents: Using social presence and personalization to shape online service encounters, *Journal of Computer-Mediated Communication*, Vol. 19, No. 3, pp. 529, 2014.
- [16] Reshmashree B. Kantharaju, Alison Pease., et al. "Integrating Argumentation with Social Conversation between Multiple Virtual Coaches. " In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents (IVA '19)*, p203–205 Association for Computing Machinery, New York, NY, USA, 2019.
- [17] 中野紀子, 比企野純太, and 安田清. "認知症患者のための語りかけエージェント." *人工知能学会全国大会論文集 第 25 回全国大会*, 一般社団法人 人工知能学会, 2011.
- [18] 梁静, 山田誠二, and 寺田和憲. "擬人化エージェント・人間・システムによる商品推薦効果の実験的比較と行動デザイン." *人工知能学会全国大会論文集 第 27 回全国大会*, 一般社団法人 人工知能学会, 2013.
- [19] 松井哲也, and 山田誠二. "ユーザの信頼を誘発する商品推薦エージェントデザイン 感情と知識量の遷移による信頼向上." *人工知能学会論文誌* 32 卷 2 号, C-G92_1, 2017.
- [20] 松井哲也, and 山田誠二. "擬人化エージェントの外見がユーザに与える印象." *人工知能学会全国大会論文集 第 31 回全国大会*. 一般社団法人 人工知能学会, 2017.
- [21] Gray, K., Wegner, D. M., Feeling robots and hu-man zombies: Mind perception and the uncannyvalley, *Cognition*, Vol. 125, No. 1, pp. 125, 2012.
- [22] 鈴木聡, and 山田誠二. "擬人化エージェントによるオーバーハードコミュニケーションのユーザの態度への影響." *情報処理学会論文誌* 46 卷 4 号, p1093-1100, 2005.
- [23] P Wühr, B P Lange, and S Schwarz. "Tears or fears? Comparing gender stereotypes about movie preferences to actual preferences. " *Frontiers in psychology* 8, p428, 2017.
- [24] Gers, Felix A., Jürgen Schmidhuber, and Fred Cummins. "Learning to forget: Continual prediction with LSTM." (1999): 850-855.

- [25] Ji, Shuiwang, et al. "3D convolutional neural networks for human action recognition." *IEEE transactions on pattern analysis and machine intelligence* 35.1, p221-231, 2017.
- [26] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [27] Bradski, Gary, and Adrian Kaehler. "OpenCV." *Dr. Dobb's journal of software tools* 3, 2000.
- [28] Liu, Wei, et al. "Ssd: Single shot multibox detector." *European conference on computer vision*. Springer, Cham, 2016.
- [29] Girshick, Ross. "Fast r-cnn." *Proceedings of the IEEE international conference on computer vision*. 2015.
- [30] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.
- [31] Ji, Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- [32] Lahoud, Jean, and Bernard Ghanem. "2d-driven 3d object detection in rgb-d images." *Proceedings of the IEEE International Conference on Computer Vision*. 2017.
- [33] Martinez, Julieta, et al. "A simple yet effective baseline for 3d human pose estimation." *Proceedings of the IEEE International Conference on Computer Vision*. 2017.
- [34] 大和淳司, 大谷淳, and 石井健一郎. "隠れマルコフモデルを用いた動画像からの人物の行動認識." *電子情報通信学会論文誌 D76.12*, p2556-2563, 1993.
- [35] 小川宏高, 中田秀基, and 工藤知宏. "多数の動画像を対象とするリアルタイム異常値検出の検討." *研究報告ハイパフォーマンスコンピューティング (HPC)*, p1-6, 2015.
- [36] Sigurdsson, Gunnar A., Olga Russakovsky, and Abhinav Gupta. "What actions are needed for understanding human actions in videos?." *Proceedings of the IEEE International Conference on Computer Vision*. 2017.
- [37] Laptev, Ivan, et al. "Learning realistic human actions from movies." 2008.

- [38] Soomro, Khurram, Amir Roshan Zamir, and Mubarak Shah. "UCF101: A dataset of 101 human actions classes from videos in the wild." *arXiv preprint arXiv:1212.0402*, 2012.
- [39] Kuehne, Hildegard, et al. "HMDB: a large video database for human motion recognition." *2011 International Conference on Computer Vision*. IEEE, 2011.
- [40] G.A.Sigurdsson,G.Varol,X.Wang,A.Farhadi,I.Laptev, and A. Gupta. Hollywood in homes: Crowdsourcing data collection for activity understanding. In *European Conference on Computer Vision (ECCV)*, 2016.
- [41] Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." *arXiv preprint arXiv:1611.08050*, 2016.