

対話的なインタラクションを行う 面接官エージェントシステムの開発

Development of an Interactive Interviewer Agent System

宮崎健斗¹ 片上大輔²

Kento Miyazaki¹, Daisuke Katagami²

¹東京工芸大学大学院

¹Graduate School of Tokyo Polytechnic University

²東京工芸大学

²Tokyo Polytechnic University

Abstract: 本研究では、学生の就職活動の支援を目的とした面接官エージェントシステムを開発する。面接官エージェントシステムは、ユーザと面接官エージェントが一对一の対話的なコミュニケーションを行い、面接の練習を行うシステムである。本システムでは面接練習の際、ユーザの言語、韻律、動作、視線情報といったマルチモーダルな情報を抽出する、抽出したマルチモーダル情報を基に、ユーザに対しフィードバックを行う、SVMを用いた評価実験の結果、「実行力」の項目においては、言語・動作・頭部特微量セットは0.9の評価推定精度を示し、他8つの項目において約0.7の推定精度を示した。また、各評価項目において優れた推定精度を示した特微量セットの特微量の重みを調べ、推定に寄与した特微量を明らかにした。この結果に基づき、低群に分類されたユーザに対して高群に見られる特徴をフィードバックすることで、適切な面接支援を行うことが可能である。

1. はじめに

仮想的なエージェントとのインタラクションを通して、人の社会的スキルの成長を目的とした研究が注目されている[1]。Aylettら[2]と Sapounaら[3]は、仮想的なエージェントとのコミュニケーションにより、子供のいじめ被害率を減らす効果を示し、Panら[4]は仮想的な女性エージェントとのコミュニケーションにより、若い成人男性の対人不安を軽減する可能性を示唆した。Cholletら[5]は仮想的なオーディエンスに向けたプレゼンテーション練習を提案し、ユーザに安全で効果的な練習の場を提供した。

これらの仮想的なエージェントとのインタラクションの利点は主に三つ上げられ、一つはユーザのニーズに合わせたシミュレーションを提供できること、二つ目は失敗のリスクが無いこと、三つ目に、繰り返してトレーニングを行えることが上げられる。以上の理由から社会的スキルの向上を目的とした研究のアプローチとして仮想的なエージェントがしばしば用いられている。

社会的スキルは、人が社会で暮らす中で欠かすことのできないスキルであり、特に社会的スキルが求

められる場面として面接が挙げられる。就職活動において面接は避けて通ることのできない工程であり、就活を控えた求職者の面接対策は主にキャリアカウンセラーとの模擬面接に依存している。

本研究では、面接官エージェントとの模擬面接を行い、面接対策や社会的スキルの向上を目的とした、面接官エージェントシステムを開発する。

2. 関連研究

自閉症スペクトラムやニートといった就職困難者の支援として、仮想的なエージェントを用いた社会的スキルの向上を目的とした研究[6]や、学生をターゲットとした面接練習の研究[7]が行われており、仮想的なエージェントとの面接練習が、プロのキャリアカウンセラーとの練習を上回る効果が報告された例も存在する[8]。

面接場面を対象として仮想的なエージェントを用いた研究に見られる点として被面接者の非言語情報が重視され、言語情報については考慮されていないことが多い。他者とのインタラクションにおいて、非言語情報は言語情報に比べ相手に与える印象への

影響が大きく、話す内容より話し方が重視されてきた[9][10].

本研究においては、面接官エージェントとの模擬面接の際に抽出された話す内容、話し方の両方を含めたマルチモーダル情報と、評定者による評定を基に機械学習を行い、ユーザに模擬面接内容のフィードバックを行う。仮想的な面接官エージェントとしてLive2D エージェントを用いる。面接官エージェントとの対話シナリオは、ユーザによってシステムに入力された「長所」、「学生時代に力を入れていたこと」といったユーザ情報に基づき生成される。

3. 面接官エージェントシステムの開発

面接官エージェントシステムは、ユーザと面接官エージェントが対一の対話的なコミュニケーションを行い、面接の練習を行うシステムである(図 1)。面接官役のエージェントには、Live2D モデルのハルを用いる。

3. 1. マルチモーダル情報の抽出

本システムは、面接練習の際、ユーザの言語、動作、視線情報といったマルチモーダルな情報を抽出する。抽出する特徴量は以下の通りである。

- ・言語特徴量：

ユーザの発話をマイクの音声入力により取得する。取得した発話に対し、形態素解析器の MeCab[11]を用い、ユーザの発話内容の品詞情報を得る。名詞数、動詞数、感動詞数、フィラー数、形容詞数、副詞数、同一名詞・動詞繰り返し最大数と名詞・動詞出現種類数の計 10 項目を言語特徴量として用いる。

- ・韻律特徴量：

ユーザの声のピッチ、インテンシティをマイクの音声入力により取得し、発話内容の韻律情報を得る。ピッチ平均、最大・最小ピッチ、抑揚、インテンシティの平均、最大・最小インテンシティ、話速の計 8 項目を韻律特徴量として用いる。

- ・動作特徴量：

ユーザの腕部座標を取得し、座標の時系列データから動作量の平均を計算し、動作特徴量として用いる。動作特徴量は LeapMotion を用いて両手の X,Y,Z 座標を取得する。

- ・視線特徴量：

Tobii Pro Fusion を用いて、ユーザの両目の視線座標(X,Y 座標)を取得する。取得した視線座標の時系列データから、視線座標の平均と動作量の平均を計算し、特徴量として用いる。

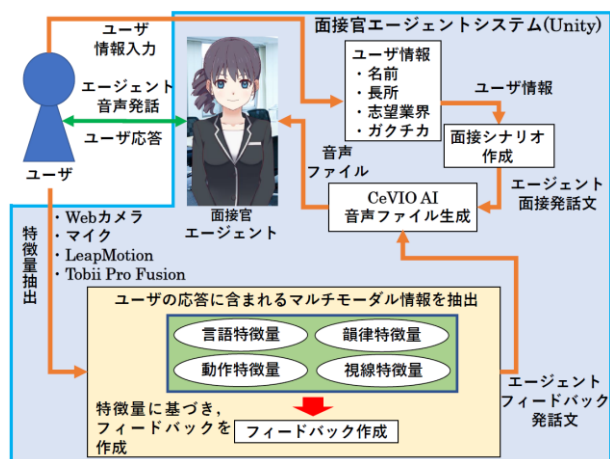


図 1. 面接官エージェントシステム概要図

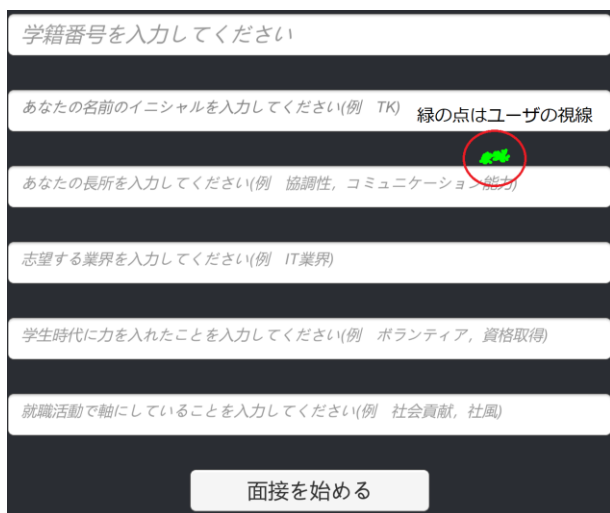


図 2. 面接官エージェントシステムタイトル画面。緑の点は Tobii Pro Fusion により取得したユーザの視線

- ・頭部特徴量

ユーザの顔座標を取得し、座標の時系列データから動作量の平均を計算し、頭部特徴量として用いる。顔座標は OpenCV, Dlib Face Landmark Detector を用い、Web カメラで顔認識を行い、鼻(Nose)、上唇(Upperlip)、下唇(Lowerlip)、顎(Chin)の X,Y 座標を取得する。

3. 2. 面接シナリオ

本システムは、面接練習を始める前にユーザにユーザ情報を入力してもらい(図 2)、ユーザに合った面接シナリオが作成される。例えば長所に「協調性」と入力した場合は、「あなたの長所は「協調性」ですが、そのように考える理由を教えてください。」といった面接シナリオが作成される。作成された面接シナリオを基に、CeVIO AI 小春六花を用いて音声



図3. 面接官エージェントを使用している様子

ファイルを作成し、面接官エージェントが発話を行い、面接シナリオが進行する。

4. 面接官エージェントシステムを用いたデータ集め実験

東京工芸大学学生 11 人に面接官エージェントシステムを使用してもらい(図 3), マルチモーダル情報の抽出と、模擬面接を行っている様子の撮影を行った。実験は東京工芸大学倫理審査委員会の承認を得て行っている。

クラウドワークスを通じて人事採用担当経験者 5 名と契約を結び、撮影されたデータ 11 件の学生を評価してもらった。学生を評価する指標として、社会人基礎力、JAVADA の定める就職基礎能力の目安に基づいた評価項目に、印象を評価する項目を加えた計 13 項目を設定した[12,13]。使用した評価項目を以下にまとめる。

- (1) 主体性：物事に進んで取り組む力
- (2) 実行力：目的を設定し確実に行動する力。
- (3) 課題発見力：現状を分析し目的や課題を明らかにする力。
- (4) 計画力：課題の発見に向けたプロセスを明らかにし準備する力。
- (5) 発信力：自分の意見をわかりやすく伝える力。
- (6) 柔軟性：意見の違いや相手の立場を理解する力。
- (7) 意見の主張：発言の道筋が明確で論理的な主張ができる。
- (8) 明確な説明：筋道の通ったわかりやすい表現で自己表現できる。
- (9) 挨拶と話し方：話し方、聞き方のコツをつかみ、相手から好感のもたれる対応ができる。
- (10) 第一印象：良い第一印象を与えたか。
- (11) 姿勢と態度：面接に臨む姿勢、態度は適切だったか。

たか。

(12) 視線：適切なアイコンタクトが行われたか。

以上の 12 項目に加えて、総合的な評価を設定した。

(13) 総合評価：全 12 項目を総合的に判断して評価する。

5 人の人事採用担当経験者に模擬面接動画の評定を Google フォーム上で行ってもらった。各模擬面接動画を見てもらい、その後、評価項目に従い、総合評価以外の項目を 7 件法、総合評価のみ 3 件法で学生(被面接者)評価してもらった。総合評価においては、fleiss' s kappa[14]を用いて、回答の一致度を算出した。その結果、 $\kappa=0.369$ の値が得られ、評価者間で総合評価の値に大きな差は見られないことが確認された。また、評価項目ごとに平均値を基準に本システム使用者を高群・低群に分類した。なお、模擬面接動画には、学生の氏名や年齢などの個人情報は含まれていない。

5. 評価実験

抽出したマルチモーダル情報と人事採用担当経験者による評定結果を用いた評価実験を行う。抽出したマルチモーダル情報の評価推定精度や、各特徴量の寄与を調査し、マルチモーダル情報と評価項目との関係を明らかにする。

5. 1. 比較対象の特徴量セット

各モダリティの推定性能への寄与を検証するため、言語特徴量、韻律特徴量、視線特徴量、動作特徴量、頭部特徴量の計 31 通りの組み合わせの特徴量セットを用いて、推定精度の比較を行う。

5. 2. 評価実験手順

分類学習には、線形カーネルの SVM を使い、模擬面接参加者の特徴量セットを入力 X 、人事採用担当経験者の評定値を Y として、入出力関係を学習する。以下の処理を各評価項目、特徴量セットごとに 1000 回繰り返し、推定精度の平均をまとめる。

11 個の模擬面接データをランダムに 2 個のデータと 9 個のデータに分類する。9 個のデータに対し、3 分割交差検証を行い、 $\{0.001, 0.01, 0.1, 1, 10, 100\}$ の範囲でハイパーパラメータ C のグリッドサーチを行う。得られた最適なパラメータ C を使い、9 個データで学習した後、2 個のテストデータに対する推定精度を確認する。

5. 3. 評価実験結果

学生(被面接者)を評価する 13 個の各評価項目において最も優れた推定精度を示した特徴量セットによ

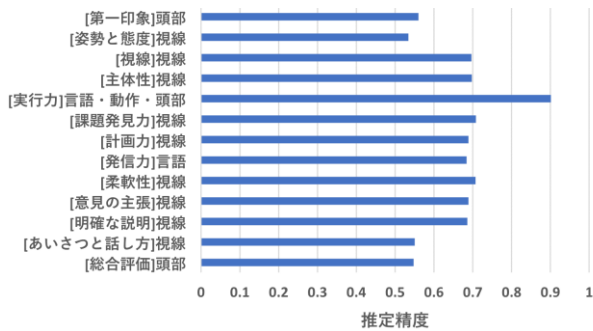


図 4. 最も優れた推定精度の特徴量セットの結果

る推定結果を図 4 に示す。推定精度は小数点第 2 位で四捨五入した。

5. 3. 1. 単一モダリティ特徴量セットのモデルの精度

「視線」の項目においては視線特徴量が 0.7 の精度を得た。「主体性」においては視線特徴量が 0.7 の精度を得た。「課題発見力」においては視線特徴量が 0.71 の精度を得た。「計画力」においては視線特徴量が 0.69 の精度を得た。「発信力」においては言語特徴量が 0.68 の精度を得た。「柔軟性」においては視線特徴量が 0.71 の精度を得た。「意見の主張」においては視線特徴量が 0.69 の精度を得た。「明確な説明」においては視線特徴量が 0.69 の精度を得た。

5. 3. 2. マルチモーダル特徴量セットのモデルの精度

「実行力」においては言語・動作・頭部特徴量セットが 0.9 の精度を得た。

6. 考察

6. 1. SVM を用いた推定精度結果

実行力の項目では、言語・動作・頭部特徴量セットが 0.9 の精度を示した。多くの項目において 0.7 に近い精度を得ており、特徴量と評価との間に相関が見られたが、総合評価、あいさつと話し方、姿勢と態度、第一印象の項目においては、2 クラスの分類タスクのチャンスレベル (0.5) に近い値が示され、特徴量との相関が見られなかった。

6. 2. 特徴量の寄与率

各特徴量の持つ重みを分析し、各評価項目の識別に寄与した特徴量を調査する。学習で構築されたモデル 10 個の寄与率の平均を採る。寄与率が正の値であれば高群の識別に寄与し、負の値であれば低群の識別に寄与する。特徴量の寄与率を表 1、表 2 にまとめる。チャンスレベルに近い評価項目 4 つ(第一印象、姿勢と態度、あいさつと話し方、総合評価)を除

いた計 9 個の評価項目の特徴量の重みについて考察する。

6. 2. 1. 視線、主体性、課題発見力、計画力、柔軟性、意見の主張、明確な説明

これらの評価項目の推定には視線特徴量が有効であった。共通する点として L_Eye_Y, R_Eye_Y が大きく負例の推定に寄与している点が挙げられる。これは下を向いている程評価が悪くなることを示しており、多くの評価項目において同様の結果を示していることから、目線を上げて相手と目を合わせる事が重要だと考えられる。

6. 2. 2. 実行力

「実行力」の推定には言語・動作・頭部特徴量セットが有効であった。Lowerlip_Y_M や Chin_Y_M が正例の推定に寄与していることから、口を大きく開け、ハキハキと話すことが重要と考えられる。反対に、L_Hand, R_Hand が負例の推定に寄与しており、ジェスチャの多用が落ち着きの無さを与え悪印象に繋がった可能性が考えられる。また、フィラー数も負例の推定に寄与しており、フィラーが多いと悪印象に繋がると考えられる。

6. 2. 3. 発信力

「発信力」の推定には言語特徴量が有効であった。正例の推定には感動詞数、副詞数、名詞出現種類数が大きく寄与していた。感動詞には「はい」などの返事が含まれており、適切な返事が良い評価に繋がったと考えられる。動詞の意味を強調させる副詞も、被面接者のエピソードを話す際に効果的に作用した可能性が考えられる。反対に、名詞の意味を強調させる形容詞と、フィラーは負例の推定に大きく寄与した。形容詞は話の冗長さに繋がった可能性が考えられる。フィラー数は「実行力」においても負例の推定に寄与しており、相手に悪い印象を与えないためにはなるべく減らすよう心がける必要があると考えられる。

7. おわりに

マルチモーダル情報を用いた面接官エージェントシステムの開発を行い、システムで抽出したマルチモーダル情報と、評定者による評定との関係を機械学習によりモデル化し、評価の自動推定を行った。

「実行力」の項目においては、言語・動作・頭部特徴量セットが 0.9 の精度を示し、他 8 つの項目において約 0.7 の推定精度を示した。また、特徴量の重みを調べ、推定に寄与した特徴量を明らかにした。この結果に基づき、低群に分類されたユーザに対して高群に見られる特徴をフィードバックすることで、適切な面接支援を行うことが可能である。

表 1. (1/2)特徴量の寄与率. 赤字は負の値を表す. L, R はそれぞれ左目, 右目を表す. X, Y はそれぞれ X 座標, Y 座標を表し, M は動作量の平均を表す.

特徴量	視線	主体性	課題発見力	計画力	柔軟性	意見の主張	明確な説明
L_Eye_X	0.205	0.481	0.443	0.482	0.377	0.383	0.401
L_Eye_Y	-0.818	-0.660	-0.627	-0.625	-0.531	-0.627	-0.623
R_Eye_X	-0.219	0.650	0.634	0.672	0.597	0.634	0.684
R_Eye_Y	-0.900	-0.443	-0.476	-0.432	-0.367	-0.458	-0.434
L_Eye_X_M	-0.414	-0.290	-0.250	-0.277	-0.230	-0.321	-0.298
L_Eye_Y_M	-0.533	-0.063	-0.101	-0.059	-0.085	-0.079	-0.038
R_Eye_X_M	0.127	0.132	0.113	0.124	0.072	0.144	0.139
R_Eye_Y_M	0.119	-0.177	-0.169	-0.167	-0.252	-0.134	-0.190

表 2. (2/2)特徴量の寄与率. 赤字は負の値を表す. L, R はそれぞれ左手, 右手を表す. X, Y, Z はそれぞれ X 座標, Y 座標, Z 座標を表し, M は動作量の平均を表す

特徴量	実行力	発信力
名詞数	0.056	0.124
動詞数	0.064	0.143
感動詞数	0.075	0.276
フィラー数	-0.126	-0.904
形容詞数	-0.083	-0.347
副詞数	-0.034	0.235
同一名詞繰り返し最大数	0.074	-0.022
同一動詞繰り返し最大数	-0.052	0.062
名詞出現種類数	0.104	0.194
動詞出現種類数	0.091	0.189
L_Hand_X_M	-0.130	-
L_hand_Y_M	0.004	-
L_Hand_Z_M	-0.125	-
R_Hand_X_M	-0.164	-
R_Hand_Y_M	-0.150	-
R_Hand_Z_M	-0.156	-
Nose_X_M	0.108	-
Nose_Y_M	0.117	-
Upperlip_X_M	0.109	-
Upperlip_Y_M	0.094	-
Lowerlip_X_M	0.118	-
Lowerlip_Y_M	0.140	-
Chin_X_M	0.099	-
Chin_Y_M	0.128	-

Intelligent User Interfaces, pp. 17-27 (2015)

- [2] Aylett, R., Vala, M., Sequeira, P., Paiva, A.: Fearnot!—an emergent narrative approach to virtual dramas for anti-bullying education. In International Conference on Virtual Storytelling, pp. 202-205 (2007)
- [3] Sapouna, M., Wolke, D., Vannini, N., Watson, S., Woods, S., Schneider, W., Enz, S., Hall, L., Paiva, A., Andre, E., Dautenhahn, K., Aylett, R.: Virtual learning intervention to reduce bullying victimization in primary school: a controlled trial. Journal of Child Psychology and Psychiatry, Vol. 51, No. 1, pp. 104-112 (2010)
- [4] Pan, X., Gillies, M., Barker, C., Clark, D.M., Slater, M.: Socially Anxious and Confident Men Interact with a Forward Virtual Woman: An Experimental Study, Plos ONE Vol.7 No.4 (2012)
- [5] Chollet, M., Wörtwein, T., Morency, L. P., Shapiro, A., Scherer, S.: Exploring Feedback Strategies to Improve Public Speaking: An Interactive Virtual Audience Framework, in Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 1143-1154 (2015)
- [6] Anderson, K., André, E., Baur, T., Bernardini, S., Chollet, M., Chryssafidou, E., Damian, I., Ennis, C., Egges, A., Gebhard, P., Jones, H., Ochs, M., Pelachaud, C., Porayska-Pomsta, K., Rizzo, P., Sabouret, N.: The TARDIS framework: intelligent virtual agents for social coaching in job interviews. In International Conference on Advances in Computer Entertainment Technology, pp. 476-491 (2013)
- [7] Hoque, M., Courgeon, M., Martin, J. C., Mutlu, B., Picard, R. W.: Mach: My automated conversation coach, In Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, pp. 697-706 (2013)
- [8] Damian, I., Baur, T., Lugrin, B., Gebhard, P., Mehlmann, G., André, E.: Games are better than books: In-situ comparison of an interactive job interview game with

参考文献

- [1] Tanaka, H., Sakti, S., Neubig, G., Toda, T., Negoro, H., Iwasaka, H., Nakamura, S.: Automated social skills trainer. In Proceedings of the 20th International Conference on

conventional training. In International Conference on Artificial Intelligence in Education, pp. 84-94 (2015)

- [9] Langer, M., König, C. J., Gebhard, P., André, E.: Dear computer, teach me manners: Testing virtual employment interview training. *International Journal of Selection and Assessment*, Vol.24, No.4, pp. 312-323 (2016)
- [1 0] Takeuchi, N., Koda, T.: Initial Assessment of Job Interview Training System using Multimodal Behavior Analysis, *Proceedings of the 9th International Conference on Human-Agent Interaction*, November, pp. 407-411 (2021)
- [1 1] Kudo, T.: *McCab: Yet Another Part-of-Speech and Morphological Analyzer* (2006)
- [1 2] 宮崎健斗, 片上大輔: マルチモーダル情報に基づく就職面接場面における被面接者の評価モデルの提案, 2020 年度人工知能学会全国大会(第 34 回), 2F6-GS-13-04 (2020)
- [1 3] 宮崎健斗, 片上大輔: マルチモーダル情報に基づく就職面接場面における被面接者の評価モデルの提案, *人工知能学会論文誌*, Vol. 36, No. 5, A-L23_1-9 (2021)
- [1 4] Fleiss, J. L.: Measuring nominal scale agreement among many raters, *Psychological Bulletin*, Vol. 76, No. 5 pp. 378-382 (1971)