対話システムが共有する第三者情報に対するユーザの興味度推定モデルの構築

金山凜吾 [1]　三野星弥 [1]　石黒浩 [1]　吉川雄一郎 [1]
[1] 大阪大学大学院 基礎工学研究科
{kanayama.ringo, mitsuno.seiya, ishiguro, yoshikawa}@irl.sys.es.osaka-u.ac.jp

概要

継続的にユーザと関わる非タスク指向型対話システム実現のため，対話中にシステムが第三者情報をうわさとして共有する手法が注目されている．この手法によりシステムへの満足度を向上させるためには，第三者情報を共有する際に，ユーザにとってより興味深い内容を選択することが重要であると考えられる．本研究では，ユーザにとって興味度の高い情報を共有する対話システム実現の第一歩として，様々な第三者情報に対するユーザの興味度を推定するモデルの構築を目指す．我々は，独自に作成した第三者情報の興味度データセット[1)]を用いて，GPT-4o をファインチューニングし，また，推定精度を高めるために多段階推定手法，ユーザ属性を考慮する手法を組み合わせることで，第三者情報についての興味度の推定精度をベースモデルよりも 45.4% 向上させたモデルを構築した．

1　はじめに

近年，娯楽 [1]，教育 [2]，福祉 [3] といった様々な領域で，人の代わりの雑談相手となる非タスク指向型対話システムが注目されている．このような非タスク指向型対話システムは，継続的なユーザとの関わりを必要とするため [4]，ユーザの対話満足度を高め，長期的にユーザと対話できるようになることが求められている [5]．

非タスク指向型対話システムによる長期対話を実現するために，第三者の情報をシステムがうわさとして共有する手法が提案されている（図 1）．先行研究では，あるユーザ（A さん）の趣味や過去の行動について，システムが別のユーザ（B さん）に第三者情報として共有する（e.g.,“A さんは海を眺めるのが好きらしいよ”）ことで，ユーザのシステムに対する対話意欲の減退を抑えられること [6] や，システムの印象（e.g., 社会性，擬人性）を向上させられること [7][8] が報告されている．しかし，これらの研究では，数多ある第三者情報の候補の中からシステムがユーザに共有する情報は，ランダムに選択されていたり [6]，実験者によって事前に設定されていたりする [7][8] など，単純な方法で決定されており，十分な検討がなされていない．そのため，システムがどのように共有する第三者情報を選択するかは課題として挙げられている [6]．

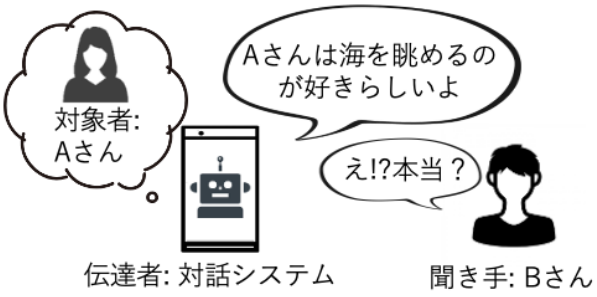

図 1　対話システムによる第三者情報の共有

先行研究では，システムがユーザの趣味・嗜好をもとに，ユーザにとって興味深いと考えられる情報を推定しながら対話することで，その満足度を向上させられることが報告されている [9][10]．これを受け，本研究では，ユーザの興味度の高い第三者情報を選択して共有することで，ユーザの対話満足度を高め，継続的にユーザと関わる対話システムを実現できるのではないかと考えた．第三者情報に対する聞き手の興味には，個人の趣味・嗜好だけでなく，人間社会において，その情報がどの程度珍しく [11][12]，人に知られていない（プライベートな）ものなのか [13][14]，また，ポジティブ/ネガティブなものであるか [10][15] といった社会的な側面も影響すると考えられる．よって，第三者情報に対するユーザの興味度を推定する手法について新たに検討

1) 本研究で作成したデータセットは GitHub（https://github.com/IshiguroLab/Thirdperson-info-Dataset）にて公開している．

する必要がある．

本研究では，ユーザにとって興味度の高い第三者情報を共有する対話システムを実現するための第一歩として，第三者の趣味や行動についての様々な文章に対しての興味度を推定するモデルを構築することを目指す．

2 提案手法

本研究では，第三者情報の興味度推定モデルの構築を目指し，独自に作成した興味度データセットで大規模言語モデル（LLM）をファインチューニングする（2.1). また，推定精度を高めるために多段階推定（2.2）およびユーザ属性の考慮（2.3）といった手法を組み合わせ，提案モデルを構築する（図 2）．

2.1 ファインチューニング

近年，医療分野での患者の症状診断推定 [16]，テキストからの話者の感情推定 [17] など，様々な分野で LLM が活用されている．また，LLM をそれぞれのタスクに特化するようにファインチューニングすることにより，ベースモデルに比べ推定精度が向上することが知られている [14][17]．そこで本研究では，第三者の趣味や行動に関する文章と興味度得点のデータセットを作成し，OpenAI の GPT-4o-2024-08-06（以下 GPT-4o）をファインチューニングして興味度を推定できるモデルを構築する．

2.2 多段階推定

推定モデルの構築において，段階的な推定手法は推定精度向上に効果的であることが知られている．例えば，感情推定において，感情の生起要因を推定してから感情の推定を行うことで，直接推定を行う場合より精度が向上することが知られている [18]．LLM においても，思考過程を出力させた後に，最終的な答えを出力させる多段階推論手法（e.g., Chain of Thought [19]）を用いることで，その精度を向上させられることが報告されている．本研究では，Chain of Thought の考え方を参考にし，まず，第一段階として文章の興味度に寄与すると考えられる特徴因子を先に推定し，その後，その特徴因子を基に興味度の推定を行う多段階推定モデルを構築する（図 2）．本研究では特徴因子として，先行研究により第三者情報に対するユーザの興味深さに関連すると考えられる文章の情報量 [11][12]，プライバシー性 [13][14]，極性 [10][15] の 3 つを選択した．

2.3 ユーザ属性を考慮した推定

第三者情報の共有において，共有される内容が同じでも，その対象者や聞き手の性別や情報が共有されるユーザ間の関係性によって，情報への興味度は異なることが知られている [20][21][22]．そこで本研究では，情報の対象者（A さん）と聞き手（B さん）の性別や関係性といったユーザ属性に基づき，興味度を推定するモデルを構築する．具体的には，性別については 2 人が同性か異性か，関係性については知り合ったばかりであるか，仲の良い友人であるかの，2 つの要因を組み合わせた 2 × 2 の 4 パターンのユーザ属性を考慮した（図 2）．

3 モデル

提案モデルを実現するために，興味度データセットを作成し，GPT-4o をファインチューニングすることで，ユーザ属性を考慮しながら多段階で興味度を推定するモデルを構築する．

3.1 データセット

我々は，第三者の趣味や行動についての文章リストを用意し，各文章の興味度，情報量，プライバシー性，極性それぞれのカテゴリについて，人間のアノテータがラベルを付与したものを，ファインチューニングに使用するデータセットとした．

3.1.1 第三者情報についての文章リスト

まず，人手で初期文章リストを作成した．具体的には，大学生 6 名に「A さんは〜」に続く形で，趣味や行動について，162 文を作成させた．初期文章リストは，情報量（ありふれた⇔珍しい），プライバシー性（パブリック⇔プライベート），極性（ネガティブ⇔ポジティブ）のカテゴリについて“どちらともいえない”を含めた 3 段階に分け，全体で 3 × 3 × 3 の 27 のブロックについて，各ブロック 6 文ずつ文章を作成させた．

次に，作成した初期文章リストを以下の手順で 1,399 文にまで拡張した．まず，LLM を用いたデータセット拡張手法である Evol-instruct [23][24] を参考に，gpt-4-turbo に文章の幅（種類）を増加させる指示を与え，文章リストの各文を 2 倍に拡張した（付録 A 参照）．また，拡張した文章リストの各文を OpenAI の text-embedding-ada-002 でベクトル化し，類似度の高い（コサイン類似度が 0.95 以上となる）

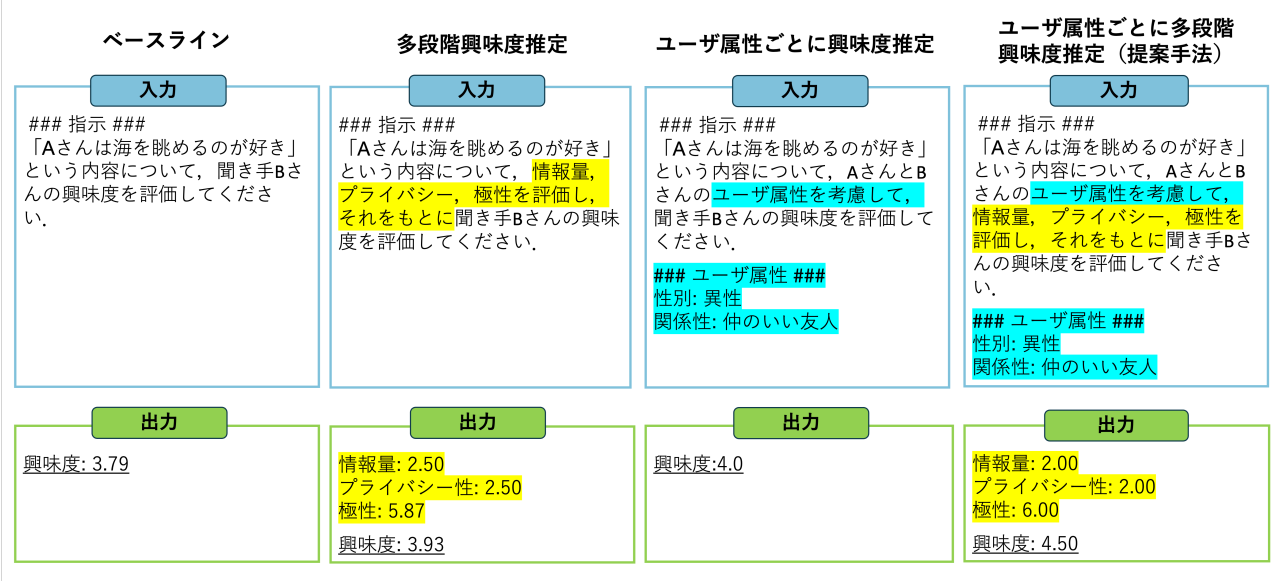

図 2　興味度推定の際に LLM に与えた入力および出力の例（入力プロンプトは一部分のみ表示）．提案手法は正解ラベル (興味度 4.41 点) に最も近い値を推定できていることが分かる．

文を削除した．この処理を 6 回繰り返し，最終的に 1,399 文を作成した．

3.1.2　ラベル付与

用意した 1,399 文の文章リストに対し，人手で得点ラベルを付与した（表 1）．具体的には，興味度，情報量，プライバシー性，極性それぞれのカテゴリについて, クラウドソーシングサービスを用いて，各文に対し 48 名ずつに 7 件法で得点を付与させた（付録 B 参照）．ラベルを付与させる際，各文ごとに A さんの属性を性別（同性，異性），関係性（仲のいい友人，知り合ったばかり）の組み合わせにより 4 種類に分類し，それぞれ 12 名ずつに得点を付与させ，その代表値 (平均値) を最終的なラベルとした．

さらに，評価実験の際に，ユーザ属性を考慮しないベースラインモデルを構築するため，A さんの属性 4 種類についてそれぞれ 12 名ずつに付与させた得点をまとめて，48 名の代表値 (平均値) を最終的なラベルとしたものを別で用意した．

興味度，情報量，プライバシー性，極性それぞれのカテゴリについてラベルを付与させる際，意味が分かりにくい文章（e.g.,“A さんは映画鑑賞を「楽しむ」にすると心地いい”）を LLM が作成した可能性があるため，該当文を除外するため，“文章の意味がわからない”という項目を同時に用意した．そして，文ごとにこの項目を選択したアノテータの人数を算出し，カテゴリごとに平均値と標準偏差を算出した．そして，いずれかのカテゴリにおいて，一定 (mean+2SD) 以上の人数のアノテータがこの項目を選択した場合，その文は除外した（35 文が該当).

また，Directed Questions Scale（DQS）[25] を参考に，質問中にチェック項目（e.g.,“この項目は 1 を選択してください”）を複数設け，これらの項目に一つでも違反したアノテータのデータはラベルの得点（平均値）の算出時に除外した（8 人が該当).

3.2　ファインチューニング

作成した興味度データセットを用いて，OpenAI の gpt-4o をファインチューニングし，第三者情報の興味度推定モデルを構築した. 具体的には，興味度データセットを訓練用，テスト用に 9:1 に分割し，OpenAI の API を用いてファインチューニングを実施し，ユーザ属性を考慮しながら特徴因子（情報量，プライバシー性，極性）を基に興味度を多段階で推定するよう学習させた gpt-4o（以下，ユーザ属性ごとに多段階興味度推定）を作成した．また，ファインチューニング時のハイパーパラメータは，OpenAI が事前に用意した，訓練データに合わせて自動で設定される値のセットを用いた (batch_size: 9, learning_rate_multiplier: 2, n_epochs: 3).

4　評価実験

4.1　ベースライン

GPT-4o をファインチューニングして作成した提案モデルを評価するための実験を行った．ベースラ

表 1　第三者情報の内容ごとの興味度の得点と特徴因子の得点の例

第三者情報の内容	興味度	情報量	プライバシー性	極性
A さんは足の爪を切った	1.500	1.604	3.312	4.542
A さんはコーヒーを飲んだ	2.596	1.170	1.810	5.063
A さんはスキーをしたことがない	3.500	3.167	3.250	3.333
A さんはバスに乗ったことがない	4.190	3.900	4.420	3.830
A さんは海を眺めるのが好き	4.062	2.041	2.208	5.958
A さんは絵画展で最優秀賞を受賞した	5.417	6.060	2.917	6.833
A さんは人を殺めた	6.021	6.938	6.771	1.000
A さんは宝くじで 1 億当たった	6.532	6.745	6.660	6.766

表 2　ファインチューン前後の興味度推定精度（MAE）

手法	FT 前	FT 後
直接興味度推定	0.718	0.445
多段階興味度推定	0.767	0.439
ユーザ属性ごとに興味度推定	0.904	0.395
ユーザ属性ごとに多段階興味度推定	**0.646**	**0.392**

インとして，直接興味度を推定するよう学習させた gpt-4o（以下，直接興味度推定），特徴因子の推定結果から興味度を推定するよう学習させた gpt-4o（以下，多段階興味度推定），ユーザの属性を考慮して興味度を推定するよう学習させた gpt-4o（以下，ユーザ属性ごとに興味度推定）の 3 種類のモデルを用意した．

さらに，提案モデルと 3 種類のベースラインモデルについて，ファインチューニング前のモデルも比較対象としてそれぞれ用意した．

4.2　実験結果

本研究で提案した，ユーザ属性ごとに多段階興味度推定を行うモデルと各ベースラインモデルそれぞれについて平均絶対誤差 (MAE) を算出し，モデルの性能を比較した．実験の結果，提案モデルが他のファインチューニングモデルと比べて最も高い推定精度を示していたことが確認された（表 2）．また，提案モデルの推定精度 (MAE= 0.392) は，ファインチューニング前の直接興味度推定モデルの精度 (MAE= 0.718) と比較して約 45.4%性能が向上していたことが確認された．

より詳細な分析として，2.1〜2.3 で提案した手法の効果をそれぞれ確認する．はじめに，ファインチューニングの効果を確認するために，提案モデルおよびベースラインモデルについて，ファインチューニング前後で MAE を比較した結果，全ての手法で，ファインチューニングにより興味度の推定精度が向上していたことが確認された．これは第三者情報の興味度推定においても，LLM をファインチューニングすることにより推定精度を向上させられることを示している．

次に，多段階興味度推定モデルと直接興味度推定モデルを比較すると，ファインチューニング前の多段階興味度推定モデルは，直接興味度推定モデルに比べて精度が低下していたが，ファインチューニング後は精度が向上し，直接興味度推定モデルを上回る結果となっていた．以上より，適切な学習を行うことにより，多段階興味度推定の手法が推定精度の向上に寄与すると考えられる．

最後に，ユーザ属性ごとに興味度推定モデルと直接興味度推定モデルを比較した．ユーザ属性ごとに興味度推定モデルは，ファインチューニング前は精度が低下していたが，ファインチューニング後には精度が向上し，直接興味度推定モデルを上回る結果となっていた．以上より，ユーザの属性の違いによる興味度の差を適切に学習することで，推定精度の向上に寄与すると考えられる．

5　おわりに

本研究では，独自に作成した興味度データセットを用いて，GPT-4o をファインチューニングし，提案手法による第三者情報の興味度推定モデルを構築した．評価実験から，ファインチューニング，多段階推定，ユーザ属性ごとに推定の各手法が，興味度推定の精度向上に有効である可能性が示された．特に，これらの手法を組み合わせた提案モデルが最も高い精度を示しており，総合的なアプローチの有効性が示唆された．今後は，提案モデルを用いて，ユーザの興味度の高い第三者情報を共有する非タスク指向型対話システムの実現を目指す予定である．

謝辞

本研究は JSPS 科研費 JP24H00165 の助成を受けた.

参考文献

[1] 吉田裕介, 萩原将文. 複数の言語資源を用いたユーモアを含む対話システム. 知能と情報, Vol. 26, No. 2, pp. 627–636, 2014.

[2] アイエドゥンエマヌエル, 林佑樹, 瀬田和久. 会話エージェントと学習支援. 教育システム情報学会誌, Vol. 36, No. 4, pp. 221–232, 2019.

[3] 大津耕陽, 西田勇樹, 木内敬太, 林勇吾. チャットボットによる個人適応型ヘルスケアの実現に向けた対話型課題の導入：解決志向アプローチを題材として. ヒューマンインタフェース学会論文誌, Vol. 24, No. 4, pp. 285–294, 2022.

[4] Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. Challenges in building intelligent open-domain dialog systems. **arXiv preprint arXiv:1905.05709**, 2019.

[5] 角森唯子, 東中竜一郎, 吉村健, 磯田佳徳. ユーザ情報を記憶する雑談対話システムの構築とその複数日にまたがる評価. 人工知能学会論文誌, Vol. 35, No. 1, pp. DSI–B_1–10, 2020.

[6] 三野星弥, 吉川雄一郎, 伴碧, 石黒浩. 友人グループ内での長期間利用による他者情報のやり取りを行う日常対話チャットボットの評価˜対話体験とプライバシー意識の調査˜. 人工知能学会論文誌, Vol. 37, No. 3, pp. IDS–I_1, 2022.

[7] C. Fu, Y. Yoshikawa, and H. Ishiguro. Sharing experiences to help a robot present its mind and sociability. **International Journal of Social Robotics**, 2020.

[8] H. Mahzoon, K. Ogawa, Y. Yoshikawa, M. Tanaka, K. Ogawa, R. Miyazaki, Y. Ota, and H. Ishiguro. Effect of self-representation of interaction history by the robot on perceptions of mind and positive relationship: a case study on a home-use robot. **Advances in Robotics**, Vol. 33, No. 21, p. 1112–1128, 2019.

[9] Takahisa Uchida, Takashi Minato, Yutaka Nakamura, Yuichiro Yoshikawa, and Hiroshi Ishiguro. Female-type android’s drive to quickly understand a user’s concept of preferences stimulates dialogue satisfaction: Dialogue strategies for modeling user’s concept of preferences. **International Journal of Social Robotics**, Vol. 13, pp. 1499–1516, 2021.

[10] 小林峻也, 萩原将文. ユーザの嗜好や人間関係を考慮する非タスク指向型対話システム. 人工知能学会論文誌, Vol. 31, No. 1, pp. DSF–A_1, 2016.

[11] 呉健朗, 富永詩音, 武藤佑太, 宮田章裕. 複数対話型エージェントの役割分担によるユーモア生成システム. 情報処理学会論文誌, Vol. 61, No. 8, pp. 1353–1362, 8 2020.

[12] 古志野瑛元, 内田貴久, 吉川雄一郎, 伴碧, 三野星弥, 酒井和紀, 石黒浩. 情報量に基づき共通選好に言及する対話ロボットが人同士の対話意欲に及ぼす影響の評価. 人工知能学会論文誌, Vol. 39, No. 3, pp. IDS6–D_1–11, 2024.

[13] 西川一二, 雨宮俊彦, 楠見孝. 対人的好奇心尺度の開発——人の感情・秘密・属性に関する好奇心探索——. 心理学研究, Vol. 93, No. 5, pp. 436–446, 12 2022.

[14] 三野星弥, 伴碧, 吉川雄一郎, 石黒浩. 初対面ロボットにユーザは何を話すのか—対話場面における話題の深さの検討. 研究報告ヒューマンコンピュータインタラクション, Vol. 2024-HCI-206, No. 32, pp. 1–4, 2024.

[15] Richard H. Smith, Terence J. Turner, Ron Garonzik, Colin W. Leach, Vanessa Urch-Druskat, and Christine M. Weston. Envy and schadenfreude. **Personality and Social Psychology Bulletin**, Vol. 22, No. 2, pp. 158–168, 1996.

[16] Aram Bahrini, Mohammadsadra Khamoshifar, Hossein Abbasimehr, Robert J Riggs, Maryam Esmaeili, Rastin Mastali Majdabadkohne, and Morteza Pasehvar. Chatgpt: Applications, opportunities, and threats. **arXiv preprint arXiv:2304.09103**, 2023.

[17] Zhiwei Liu, Kailai Yang, Qianqian Xie, Tianlin Zhang, and Sophia Ananiadou. Emollms: A series of emotional large language models and annotation tools for comprehensive affective analysis. **arXiv preprint arXiv:2401.08508**, 2024.

[18] 徳久良子, 乾健太郎, 松本裕治. Web から獲得した感情生起要因コーパスに基づく感情推定. 情報処理学会論文誌, Vol. 50, No. 4, pp. 1365–1374, 2009.

[19] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. **arXiv preprint arXiv:2201.11903**, 2022.

[20] 竹中一平. 類型別にみたうわさの伝達に関連する要因 —内容属性と機能の評価からのアプローチ—. 武庫川女子大学紀要. 人文・社会科学編, Vol. 61, pp. 43–52, 2013.

[21] M. W. H. Weenig, A. C. Groenenboom, and H. A. M. Wilke. Bad news transmission as a function of the definitiveness of consequences and the relationship between communicator and recipient. **Journal of Personality and Social Psychology**, Vol. 80, No. 3, pp. 449–461, 2001.

[22] David C. Watson. Gender differences in gossip and friendship. **Sex Roles**, Vol. 67, No. 9–10, pp. 494–502, 2012.

[23] Can Xu, Qingfeng Sun, Kai Zheng, Xiubo Geng, Pu Zhao, Jiazhan Feng, Chongyang Tao, and Daxin Jiang. Wizardlm: Empowering large language models to follow complex instructions. 2023.

[24] Haipeng Luo, Qingfeng Sun, Can Xu, Pu Zhao, Jianguang Lou, Chongyang Tao, Xiubo Geng, Qingwei Lin, Shifeng Chen, and Dongmei Zhang. Wizardmath: Empowering mathematical reasoning for large language models via reinforced evol-instruct. 2023.

[25] Michael R. Maniaci and Ronald D. Rogge. Caring about carelessness: Participant inattention and its effects on research. **Journal of Research in Personality**, Vol. 48, pp. 61–83, February 2014.

A　GPT4 による文章リストの拡張

GPT-4-turbo を用いて，文章リストを 2 倍に拡張した（表 3）.

表 3　GPT4 による文章リストの拡張の例

元の文章	拡張した文章
昨日ゴミ捨てを忘れた	今朝，朝食を作り忘れた
料理が得意	日本料理作るのが得意
ピーマンが嫌い	トマトが好き
帰国子女だ	海外で生活している
インド人の友達がいる	インド料理が好き

文章リスト拡張の際，以下のプロンプトを GPT-4-turbo に与え，文章リストの話題の幅 (種類) を増加させた．なお，{given_text} には，第三者についての文章が挿入される.

 # 指示 #

- **以下の指示に従って，対話における文章の作成をしてください.**
- **具体的には，元となる文章をベースにして，異なる話題で，類似した形式で新しい文章を 1 つ作成してください.**
- **新しい文章を作成する際には，以下の制約条件を守ってください.**

 # 制約条件 #

- **新しい文章は自然な日本語になるようにしてください.**
- **新しい文章は 30 文字以内にしてください.**
- **新しい文章は 1 文で構成してください.**
- **新しい文章は，趣味嗜好や，行動について述べた文章としてください.**
- **趣味嗜好についての文章とは，得意なものや好きなもの，苦手なものや嫌いなものについての文章となります.**
- **行動についての文章とは，これからの予定や，過去の行動，現在行っていることについての文章となります.**
- **新しい文章のフォーマットは「A さんは が得意です」,「A さんは，が好き」,「A さんは が苦手だ」,「A さんは が嫌いだ」,「A さんはこれから するつもりだ」「A さんは したよ」,「A さんは している」のような形を考え，それぞれ形を「 が得意」,「 が好き」,「〜が苦手」,「 が嫌い」,「 するつもり」「〜した」,「 している」という形に修正して書いてください.**
- **新しい文章のフォーマットは，「A さんは」という主語を省略して書いてください.**
- **新しい文章のフォーマットを修正する際には，語尾と句点は省略して書いてください.**

 # 元となる文章 #

{**given_text**}

 # 新しい文章 #

B　アノテータへの教示文

第三者情報の文章リストに得点ラベルを付与する際に，アノテータに与えた教示文を図 3 に示す.

「A さんは海を眺めるのが好き」という文章について評価してください.
・興味度ラベル（1: とても興味度の低い内容〜7: とても興味度の高い内容）
この文章が，興味を感じる内容 (興味度の高い内容) か，あるいは興味を感じない内容 (興味度の低い内容) かを評価してください.
・情報量ラベル（1: とてもありふれた内容〜7: とても珍しい内容）
この文章が，一般的にありふれた内容か，珍しい内容かを評価してください.
・プライバシー性ラベル（1: とてもパブリックな内容〜7: とてもプライベートな内容）
この文章が，一般的にパブリックな（誰にでも抵抗なく話すことができる）内容か，あるいはプライベートな（他の人に話すことに抵抗がある）内容かを評価してください.
・極性ラベル（1: とてもネガティブな内容〜7: とてもポジティブな内容）
この文章が，一般的にネガティブな内容か，あるいはポジティブな内容かを評価してください.

図 3　アノテータに与えた教示内容