

聴解テストから得られたラベルセットによるアニメーション会話の 属性のゼロショット分類の検証

Verifying Anime Scene Dialogues' Attribute Zero-shot Classification via Extracted Label Set in Listening Test

Yangdi NI^{1*} Junjie SHAN² 西原 陽子²
Yangdi NI¹ Junjie SHAN² Yoko NISHIHARA²

¹ 立命館大学大学院情報理工学研究科

¹ Graduate School of Information Science and Engineering, Ritsumeikan University

² 立命館大学情報理工学部

² College of Information Science and Engineering, Ritsumeikan University

Abstract: 従来のテキスト分類手法では各分類タスクにより、分類モデルを訓練する必要がある。そこで本研究では、ゼロショット分類モデルの分類ラベルを調整することにより、分類精度を向上する手法を提案する。日本語の聴解テストに付与した分類ラベルと日本アニメの会話シーンに提案手法を用いて付与されたラベルを用いて属性分類を行い、分類精度を検証した。会話相手の人間関係、会話の場所、会話の形式の三つの属性を設け、聴解テストの頻出パターンに限定し、アニメーションでの分類精度を属性ごとに評価した。結果としては、3種類の会話属性の中に1種類以上が正解である割合は88.5%となった。

1 はじめに

異なるデータセット中のテキストを属性ごとにマッチングするタスクにおいて、機械学習の分類の方法を用いることができる。テキストに属性を示すラベルを付与して学習データを作成し、学習データを用いて分類モデルを作成することで、テキストの自動分類とマッチングが行える。一般的にテキストにラベルを付与する作業は時間がかかる。なお、ラベルは分類タスクが変化すると付け直す必要がある。さらには分類モデルの再作成にも時間がかかる。

本研究では、学習データへのラベリングと分類モデルの再作成の作業負担を減らすため、ゼロショット分類を用いたテキストのマッチング手法を提案する。ゼロショット分類では、入力するテキストが同じであってもラベルセットが変わると分類結果も変化する。そこで本研究では、ゼロショット分類に用いるラベルセットとして分類精度の高いものを抽出し、抽出されたラベルセットを用いてテキストのマッチングを行う。

本論文では、1つのデータセットから得られたラベルセットを用いて、もう1つのデータセットの属性の分類が行えるかを検証し、マッチングの精度評価は今

後の課題とする。

2 関連研究

本研究に関する既存研究を示し、本研究の位置付けを行う。

2.1 属性を用いたテキスト分類

テキスト分類を扱った研究は多く、それぞれの研究において分類の属性が用意されている。Hanazono ら [1] の研究は、テキストに感情を表す属性ラベルを付与し、テキストを分類した。Elberichi [2] らの研究では、単語間の関係を分析し、構文属性と文法属性により、テキストを分類した。また、対話文のテキストを分類するため、会話の形式、会話の場所、会話の活動、話者間の関係性など8種類のメタ属性を取り入れた研究もある [3]。また、Pazzani ら [4] の研究では、ユーザが投稿した商品説明に記載された属性を分類することで、ユーザの興味に合致するアイテムをピンポイントで特定する。テキストの内容からテキストの場所や人物間の関係を自動的に属性で分類する研究もある [5, 6]。

既存研究ではテキストを分類するため、手作業でラベルを付与し、学習データを用意することが多い。本

*連絡先：立命館大学情報理工学研究科
〒567-8570 大阪府茨木市岩倉町 2-150

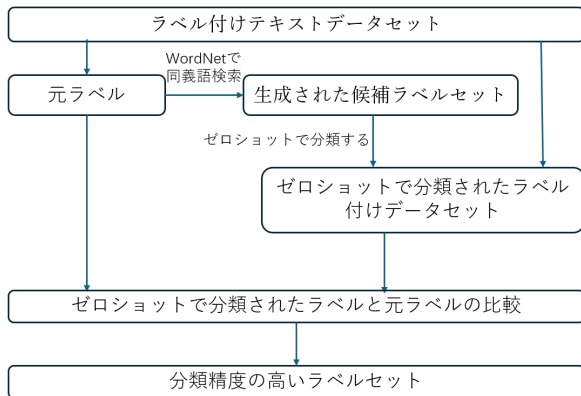


図 1: 分類精度の高いラベルセットの抽出の流れ。

研究では、ゼロショット分類を用いて異なるテキストの属性のマッチングを行う手法を提案する。

2.2 ゼロショット分類

ゼロショット分類を用いた様々な研究が提案されている。例えば、銀行が公表する文書をゼロショットで分類することにより、銀行の政策スタンスを評価する手法 [7] や、声質変換をゼロショットにより行う手法 [8] などがある。

既存研究でもゼロショット分類の精度の改善が行われているが、ラベルセットの調整を試みたものは少ない。本研究では、分類精度の高いラベルセットを自動的に抽出することでゼロショット分類を行う手法を提案する。

3 提案手法

図 1 と図 2 に提案手法の流れを示す。提案手法は主に 2 つの部分で構成され、1 つ目は属性分類の精度の高いラベルセットを特定する部分であり、2 つ目は分類された属性によりテキストをマッチングする部分である。

図 1 のラベルセットの抽出では、先にラベル付けテキストデータセットを用意する。そして、ラベル付けテキストデータセットの元ラベルを WordNet を利用し、同義語を探し、異なる候補ラベルセットを生成する。生成されたラベルセットでラベル付けテキストデータセットをゼロショットで分類し、分類の結果と元ラベルを比較し、分類精度の高いラベルセットを抽出する。

図 2 の属性マッチングでは、抽出されたラベルセットでラベルなしのデータセットとラベル付けデータセットをゼロショットで分類し、同じ属性ラベルを持つデータ同士ををマッチングする。

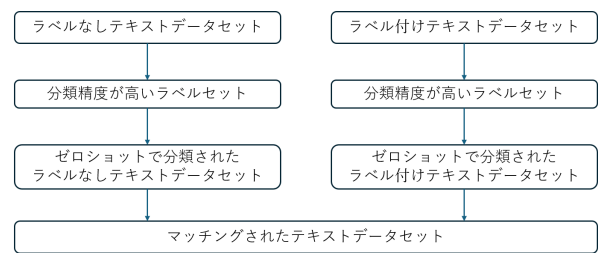


図 2: 属性マッチングの流れ。

3.1 ゼロショット分類のためのラベルセットの抽出

分類精度の高いラベルセットを抽出するため、先に WordNet で大量のラベルセットを生成する。そして、それぞれ生成されたラベルセットでテキスト属性进行分类し、正解ラベルと比べて、分類精度を算出する。

本研究は WordNet を利用し、(1) 同義語置換、(2) 反義語と組合せ、(3) フレーズを追加の 3 つの手法でラベルセットを生成する。図 3 は (1) 同義語置換の手法を示している。ラベルセット「雑談、相談」の中のラベル「雑談」に対し「ゴシップ」が同義語として示され、もう一方のラベル「相談」に対し「協議」が同義語として示されている。これらの同義語を組合せることにより、ラベルセットの候補「ゴシップ、協議」を生成する。図 4 は (2) 反義語と組合せの手法を示している (2 分類の場合)。ラベル「雑談」に対し「ゴシップ」が同義語として示され、「ではない」を追加し、「ゴシップ、ゴシップではない」を生成する。図 5 は (3) フレーズを追加の手法を示している。ラベル「雑談」に対し「ゴシップ」が同義語として示され、もう一方のラベル「相談」に対し「協議」が同義語として示されている。フレーズを追加し、「ゴシップ+フレーズ、協議+フレーズ」を生成する。図 5 ではフレーズの例として「の会話」が使われており、生成されたラベルは「ゴシップの会話、協議の会話」となる。

生成されたラベルセットでラベル付けのデータセットをゼロショットで分類し、分類結果と元付けたラベルを比較し、ラベルセットの分類精度を評価する。そして、分類精度の高いラベルセットを抽出する。

3.2 属性マッチング

抽出されたラベルセットを用いて、2 つの異なるデータセット中のテキストをゼロショットで分類する。分類された属性ラベルをテキストに付与し、属性の相違によりテキスト間でマッチングを行う。

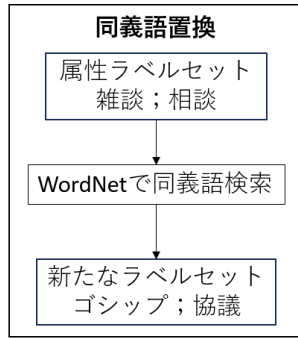


図 3: WordNet でラベルセットの生成（同義語置換）。

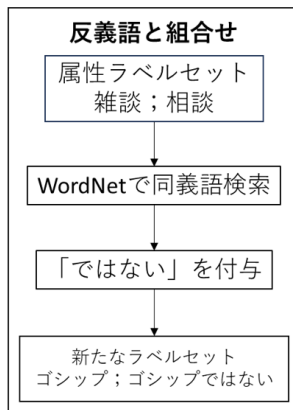


図 4: WordNet でラベルセットの生成（反義語と組み合わせ）。

4 予備実験

属性分類に用いるラベルセットを抽出するための予備実験を行った。予備実験と評価実験では、データセットとして日本語能力検定試験の聴解テストの会話テキスト集合と、日本のアニメ会話シーンのテキスト集合を用いた。これは、既存研究において聴解テストと属性が一致するアニメ会話シーンを提供する妥当性が確認されているためである [10]。

4.1 実験手順

実験は以下の手順に従って行った。

1. 聴解テストの属性ラベルを手作業で付与する。
2. WordNet を利用し、候補ラベルセットを生成する。
3. 候補ラベルセットを用いて聴解テストをゼロショットで分類する。
4. 得られた分類結果と手作業の結果を比較し、分類精度の高いラベルセットを抽出する。

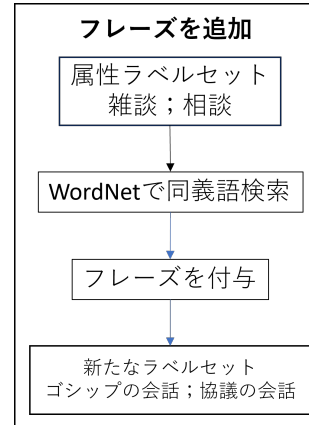


図 5: WordNet でラベルセットの生成（フレーズを追加）。

会話テキストに付与する属性は、既存研究 [3] を参照し、「場所」「形式」「関係」の 3 種類とした。それぞれに初期ラベルセットとして、

- **場所（6 分類）**：自宅、職場、学校、施設、屋外、電話
- **形式（2 分類）**：雑談、用談相談
- **関係（4 分類）**：初対面、知り合い、友達、家族

を手作業で付与した。

本研究は 2 つのゼロショット分類方法を利用した。1 つ目はテキストを Embedding してベクトルを作成し、ベクトルのコサイン類似度を評価することで分類する手法である（図 6）。言語モデルを用いてテキストとラベルを Embedding してベクトルを作成し、ベクトルのコサイン類似度を評価して、類似度が高いラベルをテキストの分類結果とする。実験は Open-AI の text-embedding-ada-002 モデル¹を用いた。

2 つ目はテキストとラベルセットを与えて、事前訓練済み言語モデルを用いてゼロショット分類を行う手法である（図 7）。テキストとラベルセットを事前訓練済み言語モデルに与え、各ラベルの尤度を算出し、尤度が最も高いラベルをテキストの分類結果とする。実験は MoritzLaurer モデル²を用いた。

2 つのゼロショット分類方法ごとに分類精度の高いラベルセットを評価した。ラベルセットの分類精度を評価するため、既存研究 [11] を参照し、分類結果の平均二乗平方誤差（RMSE）を用いて評価を行った。

予備実験に用いた聴解テストの内容を説明する。実験には N1 から N5 まで各レベル 50 件合計 250 件聴解テストを利用し、「場所」、「形式」、「関係」3 種類の属性

¹<https://platform.openai.com/docs/guides/embeddings>

²<https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7>

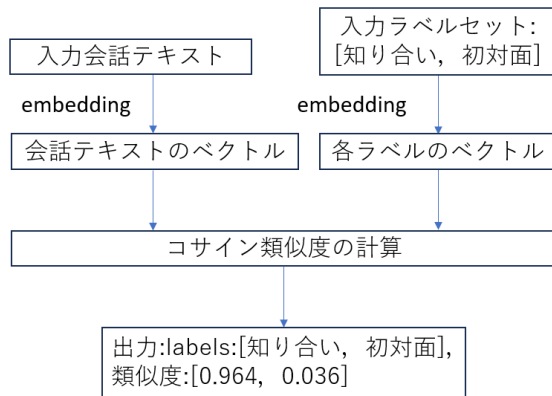


図 6: テキストと分類ラベルの Embedding の類似度によりゼロショット分類を行う例。

表 1: 各ゼロショット分類の属性ごとの RMSE が一番低いラベルセット

ゼロショット分類方法	属性	分類精度の高いラベルセット	RMSE
MoritzLaurer	場所 6 分類	[自宅で話す, 職場で話す, 学校で話す, 施設で話す, 街並みで話す, 電話で話す]	1.75
	形式 2 分類	[世間話, 相談]	3.43
	関係 4 分類	[初対面間の会話, 御近付き間の会話, 身方間の会話, 親類間の会話]	2.39
Open-AI の embedding	場所 6 分類	[自宅で話す, 職場で話す, 学校で話す, 施設で話す, 街並みで話す, 電話で話す]	2.54
	形式 2 分類	[雑談, 相談]	3.17
	関係 4 分類	[初対面, 知辺, 馴じみ, 親類]	2.74

のラベルを手作業で付けた。そして、WordNet を利用し、162 個の候補ラベルセットを生成した。

4.2 実験結果

RMSE が最も低いラベルセットを表 1 に示す。2 つのゼロショット分類方法で異なるラベルセットが得られた。MoritzLaurer モデルで抽出したラベルセットの RMSE の平均値は 2.54, Open-AI の embedding モデルでの場合の平均値は 2.81 であった。平均値より, MoritzLaurer モデルの方が分類精度の高いラベルセットが抽出されたことがわかった。

一方で、「形式」の分類においては、Open-AI の embedding モデルを用いた方が RMSE の値が低いラベルセットが抽出された。そこで、次章の評価実験においては表 1 のラベルセットのうち、(1) MoritzLaurer で得られたラベルセットを用いた分類と (2) MoritzLaurer で得られた「場所」「関係」のラベルセットと Open-AI の embedding モデルで得られた「形式」のラベルセットを用いた分類の 2 種類の評価を行う (表 2)。

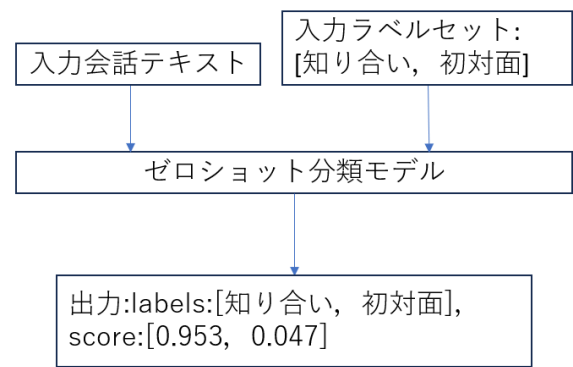


図 7: 事前訓練モデルによりゼロショット分類を行う例。

表 2: 評価実験で用いるゼロショット分類方法とラベルセット

ゼロショット分類方法	属性	分類精度の高いラベルセット
MoritzLaurer	場所 6 分類 (MoritzLaurer)	[自宅で話す, 職場で話す, 学校で話す, 施設で話す, 街並みで話す, 電話で話す]
	形式 2 分類 (MoritzLaurer)	[世間話, 相談]
	関係 4 分類 (MoritzLaurer)	[初対面間の会話, 御近付き間の会話, 身方間の会話, 親類間の会話]
Open-AI の embedding と MoritzLaurer	場所 6 分類 (MoritzLaurer)	[自宅で話す, 職場で話す, 学校で話す, 施設で話す, 街並みで話す, 電話で話す]
	形式 2 分類 (Open-AI の embedding)	[雑談, 相談]
	関係 4 分類 (MoritzLaurer)	[初対面間の会話, 御近付き間の会話, 身方間の会話, 親類間の会話]

5 評価実験：アニメーションの会話の属性分類の評価

評価実験では、聴解テストで抽出したラベルセットがアニメ会話シーンの属性分類に有効であるかを評価した。

5.1 実験手順

実験は以下の手順に従って行った。

1. 聴解テストとアニメーションの会話のテキストデータを用意する。
2. 表 2 に示すラベルセットを用いてゼロショットで分類する。
3. 聴解テストに手作業で付与された 3 種類の属性の組合せの数をカウントし、上位 10 個を頻出の組合せとする (表 3)。
4. 表 3 の属性の組合せごとに分類されたアニメーションの会話テキストを 10 個ずつサンプリングする。
5. サンプリングされたアニメ会話シーンの分類が正解か否かを手作業で確認する。

表 3: 聴解テストの各属性の組合せ.

属性の組合せ
用談相談, 職場, 知り合い
用談相談, 施設, 初対面
用談相談, 学校, 知り合い
雑談, 職場, 友達
用談相談, 学校, 友達
雑談, 学校, 友達
用談相談, 職場, 友達
雑談, 自宅, 友達
用談相談, 自宅, 家族
用談相談, 電話, 初対面

初めに聴解テストとアニメーションの会話のテキストデータを用意した. 聴解テストは予備実験で使ったものと同じであった. アニメーションの会話は, 複数のアニメから既存研究の手法 [9] を用いてシーンごとにカットされたものを用意し, 合計 314,930 個であった.

実験では, 4 章で説明したように, (1) MoritzLaurer と (2) Open-AI の embedding と MoritzLaurer の 2 つの手法でアニメーションをゼロショット分類した.

5.2 実験結果

表 4 に, MoritzLaurer モデルを用いて分類された結果を示す. 10 種類の属性の組合せに対し, 形式, 場所, 関係のいずれかが一致した数, 3 種類の属性, 2 種類の属性, 1 種類の属性が一致した数, および一致した属性がなかった数を示す. 表 4 では, 100 個のサンプリングされたアニメーションの中, 形式が一致したのは 81 個, 場所が一致したのは 53 個, 関係が一致したのは 58 個であった. 3 種類の属性が全部一致したアニメーションは 31 個, 2 種類が一致したのは 36 個, 1 種類が一致したのは 27 個, 全部一致しなかったのは 6 個であった.

表 5 は Open-AI の embedding と MoritzLaurer を用いて分類された結果を示す. 100 個のサンプリングのアニメーションの中, 形式が一致したのは 75 個, 場所が一致したのは 54 個, 関係が一致したのは 51 個であった. 3 種類の属性が全部一致したのは 33 個, 2 種類が一致したのは 29 個, 1 種類が一致したのは 21 個, 全部一致しなかったのは 17 個であった.

6 考察

100 個のサンプリングされたアニメーションの分類結果として, 1 種類以上の属性で正解であったものに注目する. MoritzLaurer モデルを用いた場合に 94 個 (31+36+27)

一致し, Open-AI の embedding と MoritzLaurer モデルを用いた場合に 83 個 (33+29+21) 一致した. 両者の平均は 88.5 となった. この結果から, データセットから得られたラベルセットを用いて, 異なるデータセットの分類が行えることが確認された.

属性ごとの分類の結果に注目する. いずれのモデルにおいても「形式」の一致数が最も多く, それぞれ 81 個と 75 個であった. このことから「形式」の分類精度の高いことが確認できた.

7 おわりに

本研究では, ゼロショット分類を利用し, テキストを分類属性によりマッチングする手法を提案した. 提案手法は, 良い分類結果が得られるラベルセットを抽出し, ラベルセットを用いてテキストのゼロショット分類を行う. 同一ラベルが付与されたテキスト同士をマッチングする. 評価実験では提案手法で得られた 1 つのデータセットから得られたラベルセットを用いて, 別のデータセットの属性の分類が行えるかを評価した.

実験では, 日本語能力試験の聴解テストの会話テキストから得られたラベルセットを用いて, アニメ会話シーンのテキストを分類する実験を行った. ゼロショットの分類モデルとして, 2 種類のゼロショット分類方法を用い, (1) MoritzLaurer モデル 3 種類属性のラベルセットと, (2) MoritzLaurer モデルの「場所」, 「関係」のラベルセットと Open-AI の embedding の「形式」のラベルセットの組合せを用いた. 100 個のアニメーションに対し, 分類された聴解テストの正解数を評価した. 実験の結果, 1 種類以上の属性が一致した分類結果は 100 個中 94 個 (MoritzLaurer) と 87 個 (Open-AI の embedding と MoritzLaurer) で, 平均が 88.5 個となり, 1 つのデータセットから得られたラベルセットが他のデータセットの属性分類に適用できることを確認した.

今後, 提案手法のマッチング精度を評価する実験を行う.

参考文献

- [1] Masaya Hanazono, Takashi Nishiyama, Keiichi Tokuda: An optimization method of constructing corpora for hmm-based emotional speech synthesis, *Proceedings of the Annual Conference of the Institute of Systems, Control and Information Engineers*, Vol. SCIO9, No. 0, pp. 303-303(2009)

表 4: MoritzLaurer モデルを用いてアニメーションを分類した結果. 各属性の組合せに一致した属性・属性種類数の統計.

属性の組合せ	形式	場所	関係	3種類の属性	2種類の属性	1種類の属性	なし
用談相談, 職場, 知り合い	6	7	3	2	4	2	2
用談相談, 施設, 初対面	8	3	3	1	2	7	0
用談相談, 学校, 知り合い	10	9	5	4	6	0	0
雑談, 職場, 友達	8	1	8	1	6	2	1
用談相談, 学校, 友達	8	9	8	7	1	2	0
雑談, 学校, 友達	10	7	9	7	2	1	0
用談相談, 職場, 友達	8	6	7	4	3	3	0
雑談, 自宅, 友達	8	4	7	3	4	2	1
用談相談, 自宅, 家族	8	6	3	2	4	3	1
用談相談, 電話, 初対面	7	1	5	0	4	5	1
合計	81	53	58	31	36	27	6

表 5: Open-AI の embedding と MoritzLaurer モデルを用いてアニメーションを分類した結果. 各属性の組合せに一致した属性・属性種類数の統計.

属性の組合せ	形式	場所	関係	3種類の属性	2種類の属性	1種類の属性	なし
用談相談, 職場, 知り合い	10	5	3	3	3	3	1
用談相談, 電話, 初対面	7	1	5	1	2	5	2
用談相談, 学校, 知り合い	8	8	7	4	5	1	0
雑談, 職場, 友達	7	2	5	2	3	2	3
用談相談, 学校, 友達	7	8	4	4	2	3	1
雑談, 学校, 友達	8	7	9	7	1	1	1
用談相談, 職場, 友達	8	6	7	5	2	2	1
雑談, 自宅, 友達	6	4	5	3	3	0	4
用談相談, 自宅, 家族	7	5	3	1	5	2	2
用談相談, 施設, 初対面	6	8	3	3	3	2	2
合計	75	54	51	33	29	21	17

- [2] Zakaria Elberrichi, Abdelattif Rahmoun, Mohamed Amine Bentaalah: Using wordnet for text categorization, *International Arab Journal of Information Technology*, Vol. 5, No. 1(2008)
- [3] 小磯花絵, 天谷晴香, 居關友里子, 臼田泰如, 柏野和佳子, 川端良子, 田中弥生, 伝康晴, 西川賢哉: 『日本語日常会話コーパス』モニター版の設計・評価・予備的分析, 国立国語研究所論集, pp. 17-33(2020)
- [4] Michael J Pazzani, Daniel Billsus: Content-based recommendation systems, *The adaptive web: methods and strategies of web personalization*, pp. 325-341(2007)
- [5] 篠原宜道, 渡部広一, 河岡司: 常識判断システムを用いた会話意味理解方式~会話文からの場所推定~, 情報処理学会研究報告自然言語処理 (NL), Vol.2003, No.4(2002-NL-153), pp. 89-96(2003)
- [6] 小井出慎, 古宮嘉那子, 佐々木稔, 新納浩幸: 物語における登場人物の親しさ推定, 言語処理学会第22回年次大会発表論文集, pp.985-988(2016)
- [7] 澤木智史, 仲山泰: ゼロショットテキスト分類を活用した含意判定モデルによるFRB金融政策コミュニケーションの読解, 人工知能学会第二種研究会資料, Vol. 2023, No. FIN-030, pp. 72-77(2023)
- [8] 鈴木大志, 鷹合大輔, 中沢実: AutoVC を用いたゼロショットリアルタイム声質変換手法の実装と評価, 情報処理学会論文誌, Vol. 65, No. 2, pp. 529-237(2024)
- [9] Junjie Shan, 西原陽子, 山西良典, 福本淳一: 日本語の単語と文法のレベルを用いたアニメのセリフ

の難易度の推定手法, 人工知能学会全国大会論文
集, Vol. JSAI2017, pp. 4F14 (2017)

- [10] NI Yangdi, SHAN Junjie, 韓毅弘, 西原陽子: 会
話属性により分類されたアニメ会話をを用いた日本
語聴解テスト練習支援での妥当性考察, 第 9 回コ
ミック工学研究会, pp. 60-64(2023)
- [11] Yangdi Ni, Junjie Shan, Yoko Nishihara: Match-
ing of Japanese Listening Test Dialogues and
Anime Scene Dialogues based on Zero-shot At-
tribute Classification *Procedia Computer Sci-*
ence, Vol. 246, pp. 3820-3829(2024)