

複数人におけるクイズ対話を対象とした主観的難易度の推定

大村 舞[†] 嶋田 和孝[†]

[†]九州工業大学 情報工学部 知能情報工学科

〒 820-8502 福岡県飯塚市川津 680-4

E-mail: ††shimada@pluto.ai.kyutech.ac.jp

あらまし 本論文では、3名によるクイズ対話を対象とし、それぞれが対話で感じた主観的な難易度を推定する手法について提案する。提案手法では、(1) 発話に関する特徴、(2) 沈黙に関する特徴、(3) オーバーラップに関する特徴の3つを利用する。発話特徴としては、発話の時間長や形態素数、前後の発話の分布などを利用する。沈黙特徴としては、沈黙の数や時間の長さなどを利用する。オーバーラップ特徴としては、その頻度や時間の平均、分散などを利用する。これらの特徴について特徴選択を行い、機械学習器である C4.5 と SVM に適用して、主観的難易度を推定する。実験の結果より、特徴選択をした C4.5 が正解率の面でも推定誤差の面でも最も高い精度を得た。

キーワード 主観的難易度, クイズ対話, 発話, 沈黙, オーバーラップ

Estimation of Subjective Impressions of Difficulty in Quiz Dialogue

Mai OMURA[†] and Kazutaka SHIMADA[†]

[†] Department of Artificial Intelligence, Kyushu Institute of Technology

680-4 Kawazu Iizuka Fukuoka, 820-8502 Japan

E-mail: ††shimada@pluto.ai.kyutech.ac.jp

Abstract In this paper, we propose a method for estimating subjective impressions of difficulty in quiz dialogue. The quiz dialogue consists of utterances from three persons. Our method applies three features, utterance, silence and overlap features, to machine learning. The utterance feature set consists of length of each utterance, the number of morphemes and so on. The silence and overlap feature sets also include the similar features to the utterance feature. We select appropriate features from them, and apply the selected features to C4.5 and SVM. In the experiment, C4.5 with the selected features produced the best performance in them.

Key words Subjective impressions of difficulty, Quiz dialogue, Utterance, Silence, Overlap

1. ま え が き

近年、人間と対話を行う対話エージェントの研究が盛んに行なわれている。人間がどのように感じているのか、あるいは対話はどのような状況であるのかを検出・推定することができれば、対話エージェントによる適切な対応が可能になる。例えば、対話エージェントが、状態を踏まえた上で適切な発話を行うことができれば、より表現豊かな対話のできるエージェントになると期待される。そのため、対話の雰囲気や話者の感情といった状態の推定について研究が行われている [7], [12].

また、近年の教育分野への IT 化の導入に伴い、e-Learning が多く利用されている [9]. e-Learning では、情報通信技術を利用することで、時間や場所に制約されないといった利点がある。一般的に対面授業では、「学習者の学習の理解度」、「学習者の主観的な評価」を基に講義内容を適宜調節しながら授業を進

めている。「学習者の主観的な評価」は、教師が学習者の様子などから判断していると考えられる。しかし、e-Learning で利用されるシステムでは一般的な対面授業とは異なり、学習者の様子を基にした評価がないため、学習者が「難しい」と感じているのか、「易しい」と感じているのかを判断することが難しい。そこで、学習者が感じている「難しさ」「易しさ」を、システムが学習者の様子から推定する研究が行われている [6], [11].

中村ら [11] は、擬人化エージェント間の音声対話場面を対象として、対話内容に対して難しいと感じていたか否かを、時間区分ごとの非言語動作の頻出頻度を基に推定している。繁田ら [6] は、英語のリスニング教材を対象として、学習内容に対して主観的に感じた「簡単」、「難しい」の評価を、眼球運動の特徴から推定を行っている。これらの研究では、ユーザの動作や眼球運動といった特徴を用いて推定している。しかし、エージェントや人間同士のやりとりで最も重要な要素の一つである

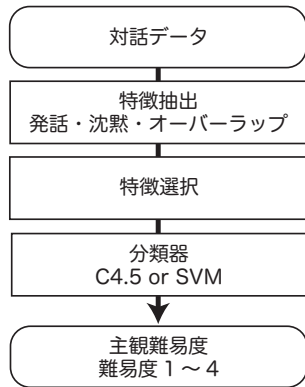


図1 主観的難易度推定の概要.

対話から得られる特徴を考慮していない。もし、対話からの特徴を用いてシステム利用者が難しいと感じていたか否かを推定することができれば、実際の間人同士の対話のように、ユーザが難しいと感じていたか否かを推定することができ、例えば、e-Learning など、さまざまな対話的なシステムへ応用ができる。

これらの研究では、1対1の人間対エージェントを想定して研究されている。一方で、人間同士の対話ややりとりを分析し、その結果を対話システムに組み込むことができれば、より自然な対話エージェントが実現できる。本研究では、人が対話中などに、その対話や内容はどの程度難しいと感じたかの評価を「主観的難易度」と定義する。クイズを解くタスク(以降、クイズ対話)を対象とし、複数人対話からの特徴を基にした主観的難易度の推定の有効性について検証する。

2. 主観的難易度の推定法の概要

本節では、複数人によるクイズ対話を対象とした主観的難易度の推定手法の概要を述べる。提案手法の概要を図1に示す。提案手法では、クイズ対話から「発話に関する特徴」、「沈黙に関する特徴」、「オーバーラップに関する特徴」の3つの特徴量を抽出する。そして、抽出した特徴から有効な特徴を選別し、それらを機械学習に適用することで、主観的難易度を推定する。

本研究では、汎用的な主観的難易度を推定することを目的とする。そこで、「すべての人は主観的難易度の程度によって同じような対話の特徴を表出させる」と仮定し、個人用の分類器をそれぞれ作成するのではなく、全体の事例を基に分類器を作成する。例えばAさん固有の特徴、Bさん固有の特徴を探しだし、それぞれのモデルを構成することが目的ではなく、多くの人が「難しい」あるいは「易しい」と感じるときに現われる一般的な特徴とその特徴を基にしたモデルの構成が本研究の目的である。以降の節では、推定手法に適用する特徴とその選択手法について説明する。

3. 特徴抽出

本研究では、対話データから対話に関する特徴を抽出し、抽出した特徴を基に機械学習を行う。主観的難易度推定に用いる特徴は大きく分けて以下の3つとなる。

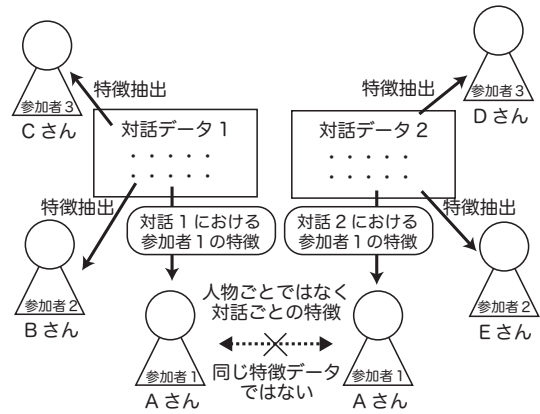


図2 対話から対象者についての特徴を抽出.

- (1) 発話に関する特徴
- (2) 沈黙に関する特徴
- (3) オーバーラップに関する特徴

これらの特徴に加えて、全体の対話時間、対話者全員の発話の合計も特徴として加える。主観的難易度は個人が感じる評価であるため、抽出する特徴は、図2のように主観的難易度を推定したいある対話に参加している人に対して、その参加者ごとの特徴をその対話のみから抽出して用いる。例えば、Aさんが「他の対話」に存在していても、その「他の対話」からAさんの特徴を抽出し、Aさん用の特徴データを作成するのではなく、現在の対象となっている対話からしか特徴を抽出しない。これは前述したように、本研究では個人用のモデルを構成した上での主観的難易度推定ではなく、前述のように一般的な主観的難易度推定を目的としているためである。各特徴の詳しい説明は3.1節から3.3節で行う。

3.1 発話に関する特徴

発話とは、対話中における発言のことを指す。発話に関する特徴として、(1) 発話の時間長・頻度、(2) 発話の形態素および(3) 前後の発話についての分布を用いる。

3.1.1 発話の時間長・頻度

発話に関する特徴として、発話の時間長と頻度を特徴として用いる。発話の時間長とは、発話自体の時間長のことを指す。対話中の対象者の発話について以下の特徴を抽出する。

- 発話時間長の総和
- 発話の頻度
- n 秒以上である発話の頻度 ($n = 0.5, 1.0, 3.0, 5.0$)
- 発話時間長の最大と最小
- 発話時間長の平均と分散
- 全体時間に対する発話時間長の割合

発話時間長の総和とは、対話中における対象者の発話の時間長の合計である。発話の頻度とは、対話中における対象者の発話の回数である。単純な発話の頻度のみではなく、0.5秒以上、1.0秒以上、3.0秒以上、5.0秒以上の時間長である発話の頻度も特徴として加える。発話時間長の平均 ($UtAve$) は、発話時間長の総和を発話の頻度で割った式1で求まる値とする。

$$UtAve = \frac{\text{発話の時間長の合計}}{\text{発話の頻度}} \quad (1)$$

全体時間に対する時間長の割合 ($UtRatio$) は、式 2 で求まる値である。

$$UtRatio = \frac{\text{発話の時間長の合計}}{\text{全体の対話時間}} \quad (2)$$

時間長の分散 ($UtVar$) は、式 1 で求めた値を平均 \bar{u} とし、対象者の発話時間を $u_i (i = 1, 2, \dots, n, n$ は発話の頻度) としたとき、式 3 で求まる値である。

$$UtVar = \frac{\sum_{i=1}^n (u_i - \bar{u})^2}{n} \quad (3)$$

また、対話全体での発話のばらつきを考慮するため、対話の序盤、中盤、終盤での発話の頻度、発話の時間長の総和、発話時間長の平均も特徴として加える。

3.1.2 発話の形態素

発話に関する特徴として、形態素数を特徴として用いる。形態素への分割は MeCab^(注1) を用いる。対象者の発話の形態素数について以下の特徴を抽出する。

- 形態素数の総和と名詞の総和
- 形態素数が n 個以上である発話の数 ($n = 5, 10, 15, 20$)
- 形態素数の最大と最小
- 形態素数の平均

また、事前の対話分析により、主観的難易度が高く評価されるほど、同じ形態素が頻出されやすい傾向が見られた。そこで、発話中の形態素の重複に関して以下の特徴を用いる。

- 他の発話に同じ形態素が含まれていた発話の数
- 全発話の形態素集合の数
- 全発話数に対する「他の発話に同じ形態素が含まれていた発話の数」の割合
- 形態素数の総和に対する「全発話の形態素集合の数」の割合

形態素集合とは、同じ発話から重複したものを取り除いたものである。全発話の形態素のリストを構成し、重複を取り除いたものの数が「全発話の形態素集合の数」である。この数は、形態素数の総和との差が大きいかほど形態素が重複しており、上で示した後半 2 つの特徴は値が小さいほど形態素が重複していることになる。また、全発話の形態素の集合の中でも「重複数が最も大きかった形態素の数」と「重複数が最も小さかった形態素の数」も特徴として用いる。

3.1.3 前後の発話の分布

コールセンターのような 1 対 1 の対話では、自分の発話か相手の発話しかなく、それぞれが順番に交替しながら対話していくモデルを当てはめやすい。しかしながら、本研究で扱うような複数人による対話では、必ずしもそれぞれが順番に交代しながら発話を繰り返すとは限らない。同じ人が続けて発話を繰り返すこともあれば、複数人の中の一部のみが発話する可能性もある。この特徴を考慮するため、「対象者の発話から前後 2 発話が自分の発話であるか否かの分布」を特徴として加えた。各発話について、前後の発話が「自分の発話」か「他の発話」か分布を見る。図 3 の対話の場合、四角で囲まれている発話は前後 2

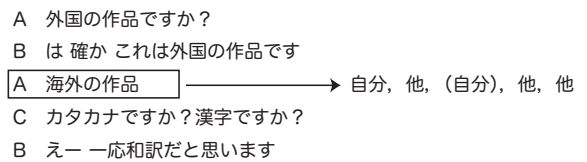


図 3 自分の発話であるか否かの分布の例

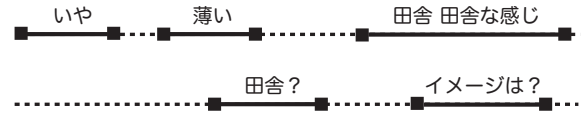


図 4 沈黙の例

発話を見ると「自分, 他, (自分), 他, 他」という分布になっている。各発話についてのこの分布の総和を特徴として用いる。「自分の発話」か「他の発話」の 2 通りであるので、1 つの発話について $2^4 = 16$ 通りのパターンが考えられる。よって、対象者につき 16 個の特徴を持つことになる。

3.2 沈黙に関する特徴

沈黙とは、発話と発話との間の発話していない区間のことを指す。図 4 のような対話で示すと、沈黙は点線の区間である。本研究における沈黙は、対象者が発言していない区間を対象とする^(注2)。この沈黙について以下の特徴を抽出する。

- 沈黙時間の総和
- 沈黙の頻度
- n 秒以上の沈黙の頻度 ($n = 0.5, 3.0, 5.0, 10.0$)
- 沈黙時間の最大と最小
- 沈黙時間の平均と分散
- 全体時間に対する沈黙時間の割合
- 発話時間長に対する沈黙時間の割合

沈黙時間の総和とは、対話中における対象者の沈黙時間の合計である。沈黙の頻度とは、対話中における対象者の沈黙の回数である。単純な沈黙の頻度のみではなく、さらに 0.5 秒以上、3.0 秒以上、5.0 秒以上、10.0 秒以上の時間長である沈黙の頻度も特徴として加える。沈黙時間の平均や分散、全体時間に対する沈黙時間の割合、発話時間長に対する沈黙時間の割合は発話に関する特徴と同様に求める。

また、対話のどのような場面で沈黙が生じているかも難易度の推定に貢献すると考えられる。例えば、序盤の沈黙よりも終盤の沈黙の方が参加者が難易度が高いと感じている状態かもしれない。そこで、対話全体での沈黙のばらつきを考慮するため、対話の序盤、中盤、終盤での沈黙の頻度、沈黙時間の総和、沈黙時間の平均も特徴として加える。

3.3 オーバーラップに関する特徴

オーバーラップとは、発話と発話が重なった区間のことを指す。本研究におけるオーバーラップは、対象者が発話している際に他の参加者との発話が被っている区間のことを指す。このオーバーラップについて以下の特徴を抽出する。

- オーバーラップ時間の総和
- オーバーラップの頻度

(注1) : <http://code.google.com/p/mecab/>

(注2) : 対象者の沈黙中に他の話者が発言しているかどうかは考慮していない。

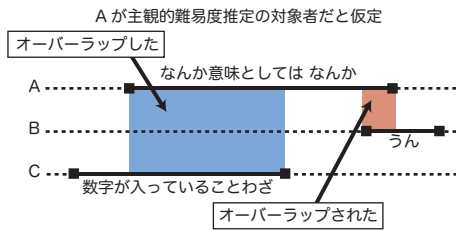


図5 オーバーラップした・オーバーラップされた。

- n 秒以上であるオーバーラップの頻度 ($n = 0.5, 1.0, 3.0, 5.0$)
- オーバーラップ時間の最大と最小
- オーバーラップ時間の平均と分散
- 全体時間に対するオーバーラップ時間の割合
- 発話時間長に対するオーバーラップ時間の割合

オーバーラップ時間の総和とは、対話中における対象者のオーバーラップ時間の合計である。オーバーラップの頻度とは、対話中における対象者のオーバーラップの回数である。単純なオーバーラップの頻度のみではなく、さらに0.5秒以上、1.0秒以上、3.0秒以上、5.0秒以上の時間長であるオーバーラップの頻度も特徴として加える。ここでも、平均や割合は発話や沈黙に関する特徴の場合と同じように算出する。

また、発話や沈黙の場合と同様に、対話全体でのオーバーラップのばらつきを考慮するため、対話の序盤、中盤、終盤でのオーバーラップの頻度、オーバーラップ時間の総和、オーバーラップ時間の平均も特徴として加える。

また、オーバーラップが発生した状態に関する特徴を用いる。オーバーラップの状態として「自分が発話をしたことによりオーバーラップが起きた」場合と「自分が発話している最中に他の人が発話したことによりオーバーラップが起きた」場合が考えられる。そこで、前者を「オーバーラップした」、後者を「オーバーラップされた」と呼ぶ。図5の対話例で説明する。Aを対象者としたとき、「なんか意味としてはなんか」という発言に対して、最初のオーバーラップはAが発言したためにオーバーラップが起きたので「オーバーラップした」となり、後のオーバーラップはBが後から発言したためにオーバーラップが起きたので「オーバーラップされた」となる。この2つの頻度を特徴として用いる。

4. 特徴の選択方法

前節で述べた特徴の総数は108個になる^(注3)。しかし、すべての特徴が主観的難易度の推定に有効であるとは限らない。そこで、適切な特徴を選択したのち分類器を作成することにする。最も有効であると評価された特徴のサブセットを求め、このサブセットから分類器を学習する。特徴のサブセットを評価方法として Correlation-based Feature Selection(CFS) [1] を用いる。この評価方法は、各特徴間の冗長性具合からそれぞれの特徴の予測能力を踏まえることで、特徴のサブセットを評価している。

(注3)：前節で説明した項目の合計ではなく、すべての特徴を累計した数であることを注意。

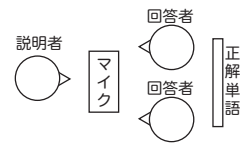


図6 頭上方向からのクイズ対話の配置。



図7 クイズ対話の様子。

各クラスと相関が高く、特徴のサブセット間では相関が低くなるようなサブセットは良い特徴のサブセットであると仮定し、そのようなサブセットを高く評価する。CFSは式4のように定式化される。

$$CFS = \max_{S_k} \left[\frac{r_{cf_1} + r_{cf_2} + \dots + r_{cf_k}}{\sqrt{k + 2(r_{f_1 f_2} + \dots + r_{f_i f_j} + \dots + r_{f_k f_1})}} \right] \quad (4)$$

ここで、 S_k は特徴のサブセットであり、 r_{cf_i} はクラス c と特徴 f_i との相関、 $r_{f_i f_j}$ は特徴 f_i と特徴 f_j との相関を表わす ($f_i, f_j \in S_k$)。

5. 実験

主観的難易度推定の手法について有効性を検証するため、クイズ対話を収録し、実験を行なった。本節では、まず、本研究にて収録したクイズ対話について述べ、実験結果と考察を述べる。

5.1 対話データの作成

実験データは、複数人によるクイズ対話である。クイズ対話では、考える、悩む、相談するといった様子が比較的現われやすく、主観的難易度を個人が評価しやすいと考えられる。

本実験で行ったクイズ対話の内容は「指定された単語を対話を通じて当てる」といった形式のものである。クイズ対話は3人1グループで行う。この3人を図6のように配置する。実際の図7にクイズ対話の様子を示す。一方の側にいる1人(以下説明者)のみに見えるホワイトボードに正解単語を提示し、質問・対話のやり取りを通じながら、残りの2人(以下回答者)に当ててもらおうという流れである。説明者は、対話の最初に提示したヒントを最初に言い、以降は回答者からの質問にのみ解答するという制限を加えている。回答者から正解単語が言われた、あるいは制限時間である210秒を越えたら終了とする。1グループにつき3問取り組んでもらった。この対話を中央に配置されたマイクで収録した。

クイズ対話終了後、参加者全員に各問題の主観的難易度について以下の4段階で評価をしてもらった。

難易度1 「簡単だった」

難易度2 「やや簡単だった」

表 1 主観的難易度の評価結果.

訓練	難易度 1	難易度 2	難易度 3	難易度 4	合計
	27	16	27	38	108
テスト	難易度 1	難易度 2	難易度 3	難易度 4	合計
	8	6	11	10	36

表 2 1 グループ分の対話データの一部.

話者	開始時間	終了時間	発話時間	発話内容
C	1.340	3.410	2.070	最初のヒントは スポーツ
A	2.410	3.020	0.610	うん
B	2.430	3.080	0.650	はい
B	4.440	6.720	2.280	はい それは きゅ 球技ですか
A	5.880	6.860	0.980	スポーツかあ
C	7.510	8.660	1.150	それは球技じゃありません
A	8.695	9.445	0.750	球技じゃない
A	10.710	11.42	0.710	球技じゃない
A	12.725	13.655	0.930	球技じゃないスポーツ
A	14.230	14.890	0.660	格闘技?
C	15.580	16.540	0.960	格闘技に近いです
A	16.560	17.510	0.950	格闘技に近い
B	16.955	17.675	0.720	近いです
...

難易度 3 「まあまあ難しかった」

難易度 4 「難しかった」

評価は、上から順に主観的難易度が低く、下の方の評価であるほど主観的難易度は高いものとする。本手法では前述のように「すべての人は主観的難易度の程度によって同じような対話の特徴を表出させる」と仮定している。そのため、1 グループにつき、3 問×3 人=9 の主観的難易度の事例を得ることになる。機械学習の訓練用のデータとして 12 グループ、テストデータとして 4 グループ分の収録を行った。各参加者の判断した主観的難易度の分布は表 1 のようになった。

収録した対話をテキストデータとして書き起した。作成した書き起しデータの一部を表 2 に示す。発話は 0.3 秒以上の沈黙があれば、2 つの発話として分割した。また、対話データの範囲は、説明者がヒントに関する発話をしてから、回答者から正解のワードが出るあるいは制限時間を経過する直後の発話までとした。

5.2 実験設定

対話データから特徴量を抽出し、機械学習に適用した。分類器としては C4.5 [4] と SVM [5] を採用し、特徴選択も含め実装には Weka [2] を利用した。実験で比較するのは以下の 3 種類である。

C4.5_{no_select} 機械学習アルゴリズムに C4.5 を用い、特徴選択をせずに学習

C4.5_{do_select} 機械学習アルゴリズムに C4.5 を用い、特徴選択してから学習

SVM_{do_select} 機械学習アルゴリズムに SVM を用い、特徴選択してから学習

ここで、「特徴選択をせずに学習」というのは 4. 節で評価したサブセットを用いるのではなく、3. 節で取り挙げた全特徴 108 個を用いて学習することを指す。作成した対話データのうち、

表 3 評価結果.

利用した特徴	正解率	推定誤差
C4.5 _{no_select}	38.89	1.13
C4.5 _{do_select}	50.00	0.96
SVM _{do_select}	50.00	1.35

訓練用のデータを機械学習で学習し、分類器を作成する。そして、テストデータの主観的難易度の評価を分類器によって推定し、実際の主観的難易度の評価と比較する。

主観的難易度の推定結果の評価は正解率と推定誤差の 2 点で行う。正解率とは、推定結果と実際の主観的難易度の一致度である。一方で、推定誤差とは以下の式で求められる値である。

$$R = \sqrt{\frac{\sum_{i=1}^n (a_i - p_i)^2}{n}} \quad (5)$$

ここで a_i は実際の主観的難易度の評価値、 p_i は提案手法における主観的難易度の推定値である (いずれも $n = 1, 2, \dots, n$ であり、 n はテストデータ数である)。

式 5 は実際に得られた主観的難易度の評価値 a_i ($i = 1, 2, \dots, n$) と本手法における主観的難易度の推定値 p_i ($i = 1, 2, \dots, n$) との差の二乗平均平方根である (n はテストデータ数)。この推定誤差 R は、推定の誤りの程度を示すものである。本実験では、「難易度 1」から「難易度 4」までの 4 クラスの分類について評価する。このとき、「難易度 1」と評価していたものを「難易度 4」と推定するものよりも、「難易度 1」と評価していたものを「難易度 2」と推定するような分類器のほうが本研究の目的に適していると判断したため、評価尺度 R を用いることにした。

5.3 実験結果と考察

実験結果を表 3 に示す。正解率については、特徴選択をした C4.5 および SVM がしていない場合と比較して高い値を示した。この結果は特徴選択の有効性を示している^(注4)。

ここで、特徴選択をした C4.5 と SVM についてその内訳をみる。C4.5_{do_select} の推定結果が表 4 であり、SVM_{do_select} の推定結果が表 5 である。この 2 つの表から分かるように、C4.5 と比較して、SVM は大きな間違いが多いことが分かる。例えば、難易度 4 を難易度 1 と 3 つの事例で誤っている。これは、特徴選択をしていない C4.5 と比較しても同様の傾向があった。そのため、C4.5 と比較して、SVM は推定誤差が大きくなっている (表 3 を参照)。

次に、特徴選択について述べる。最も評価の良かった分類器 C4.5_{do_select} について、選択された特徴は以下のようなものであった。

- 全体のクイズ時間
- 全体の発話数
- 最小の発話の時間長
- 沈黙の総時間
- 前後 1 つの発話が他の人の発話である発話の数

(注4) : 特徴選択をしていない SVM についても評価したが同様の結果が得られている。

表 4 主観的難易度の推定結果 C4.5_{do_select}.

	推定			
	難易度 1	難易度 2	難易度 3	難易度 4
難易度 1	5	2	2	0
難易度 2	0	2	2	2
難易度 3	0	4	4	3
難易度 4	0	1	2	7

表 5 主観的難易度の推定結果 SVM_{do_select}.

	推定			
	難易度 1	難易度 2	難易度 3	難易度 4
難易度 1	7	0	2	0
難易度 2	1	1	2	2
難易度 3	1	4	3	4
難易度 4	3	0	0	7

● 全発話数に対する「同じ形態素が含まれていた発話の数」の割合

● 形態素数の総和に対する「全発話の形態素集合の数」の割合

● 対話の始め、中頃、終盤での沈黙時間

選択された特徴は、主に「発話に関する特徴」「沈黙に関する特徴」であり、「オーバーラップに関する特徴」は選択されていなかった。実際のクイズ対話中에서도、主観的難易度を高く評価していた人は低く評価していた人よりも沈黙が長い傾向や、発話の頻度が多い傾向が見られたため、発話や沈黙の特徴が主に選択されたと考えられる。オーバーラップについて、主観的難易度の推定に有効であると判断されるためには、さらにオーバーラップが起きたときの発話内容や前の沈黙に着目すべきであると考えられる。主観的難易度の評価を高く評価していた人の傾向として、独り言のようにつぶやいている最中にオーバーラップが起きたというものがあつた。したがって、「発話中の形態素が少ない発話」で起きたオーバーラップ、「長い沈黙が発生したのち」のオーバーラップなど詳細に、かつ「発話」や「沈黙」と組み合わせ主観的難易度の分析を行う必要がある。

選択された特徴を見ると、「発話に関する特徴」の中でも単純な発話の時間長・頻度についての特徴よりも、前後の発話に関する特徴や形態素の重複に関する特徴が主に選択されていた。単純な数のみに着目するよりも、対話の流れがどのようなものなのか、何を発言していたのかに着目した特徴でないと主観的難易度の推定に有効に働かないと考えられる。

6. まとめ

本稿では、クイズ対話を対象とした主観的難易度の推定手法について述べた。対話からの特徴として、発話、沈黙、オーバーラップに関する特徴から選出されたものを分類器に適用し、主観的難易度を推定した。特徴選択は有効に機能したが、正解率は50%で、十分なレベルとは言いがたい。実験の結果から、主観的難易度の推定には、単純な時間長や頻度に関する特徴よりも、形態素の重複や発話の前後に着目した特徴の方が有効に働くことが分かった。推定精度向上のためには、これらについ

て、さらに詳細な特徴量の分析が必要である。

今回は発話や沈黙に関する長さや時間などの非言語的な特徴のみで推定をしたが、どのような発話がなされたか、というのも重要な特徴である。特定の手がかり語に代表されるような言語特徴の導入は重要な課題である。また、対話には発話の情報だけでなく、さまざまなマルチモーダルな情報が含まれている。我々は、これまでに、対話参加者の動きや姿勢に基づく対話分析手法 [13], [14] や参加者の立ち位置や距離感などからお互いの協力度を推定する手法 [8], 発話に含まれる笑いの情報を考慮した場の盛り上がり推定 [10] の研究を行っている。それ以外にも視線 [3] など、対話参加者の状態を推定するための情報源は多くある。このような情報の統合的利用も今後の課題である。

文 献

- [1] M.A. Hall. Correlation-based feature subset selection for machine learning. Master's thesis, Ph.D thesis, University of Waikato, 1998.
- [2] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: An update: Sigkdd explorations. *ACM SIGKDD Explorations Newsletter*, Vol. 11, No. 1, pp. 10-18, 2009.
- [3] Dineshbabu Jayagopi, Dairazalia Sanchez-Cortes, Kazuhiro Otsuka, Junji Yamato, and Daniel Gatica-Perez. Linking speaking and looking behavior patterns with group composition, perception, and performance. In *Proceedings of the 14th ACM international conference on Multimodal interaction, ICMI '12*, pp. 433-440, 2012.
- [4] Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [5] V. N. Vapnik. *Statistical Learning Theory*. Wiley, 1999.
- [6] 繁田亜友子, 濱本和彦, 野須潔. 英語リスニング電子教材を対象とした眼球運動分析による学習者の主観的難易度の推定. 東海大学紀要. 開発工学部, Vol. 20, pp. 117-125, 2011.
- [7] 豊田薫, 宮越喜浩, 山西良典, 加藤昇平. 発話状態時間長に着目した対話雰囲気推定. 人工知能学会論文誌, Vol. 27, No. 2, pp. 16-21, 2012.
- [8] 坂口勝哉, 嶋田和孝. 二人組共同作業における頭上カメラを用いた協力度推定. Technical report, 人工知能学会先進的学習科学と工学研究会, SIG-ALST, 2013.
- [9] 日本イーラーニングコンソシアム. eラーニング白書 2008/2009年版. 東京電機大学出版局, 2008.
- [10] 嶋田和孝, 楠本章裕, 横山貴彦, 遠藤勉. 複数人談話における笑いの情報を考慮した盛り上がり判定. 電子情報通信学会, 言語理解とコミュニケーション研究会 (NLC), NLC2012-7, pp. 25-30, 2012.
- [11] 中村和晃, 角所考, 正司哲朗, 美濃導彦, 澤木美奈子, 南泰浩, 前田英作. 擬人化エージェントとの音声対話時におけるユーザの非言語動作からの難/易及び興味/退屈の推定. 電子情報通信学会論文誌 A, 基礎・境界, Vol. 95, No. 1, pp. 85-96, 2012.
- [12] 多田和彦, 矢野良和, 道木慎二, 大熊繁. 感情遷移における急激な韻律特徴変化の検出による感情遷移判別. 知能と情報: 日本知能情報ファジィ学会誌: Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 22, No. 1, pp. 90-101, 2010.
- [13] 小松和朗, 嶋田和孝, 遠藤勉. 多人数インタラクション評価のための姿勢推定. 電子情報通信学会, ヒューマンコミュニケーション基礎研究会 (HCS), pp. 25-30, 2012.
- [14] 小松和朗, 嶋田和孝, 遠藤勉. 話者の頭部及び姿勢変化に着目した複数人対話分析. 火の国情報シンポジウム 2013, 2013.