

ヒューマノイドロボットを用いた講義代行システムのための Wikipedia を用いた多肢選択問題生成手法の性能評価

芳尾哲¹ 荒木健治¹

概要: 従来研究の講義代行システムでは、プレゼンテーションスライドと教科書の情報のみを用いており、単調かつ一方的にロボットが話し続けるために受講者にとって退屈な講義であることが課題であった。そこで、本稿では Wikipedia データから自動生成した多肢選択問題に対して Web ページから受講者が回答できるシステムを講義に取り入れることで、受講者が能動的に参加できる講義を行うシステムを開発し、印象評価実験を行った。その結果、提案システムが従来システムを 1 から 7 の 7 段階評価の 15 項目の平均値で 0.19 ポイント上回り、その有効性を確認した。

キーワード: ヒューマノイドロボット, Wikipedia, 多肢選択問題, 文生成

Performance Evaluation for Multi-choice Quiz Generation System using Wikipedia for Lecture Substitution System using Humanoid Robot

SATOSHI YOSHIO^{†1} KENJI ARAKI^{†1}

Abstract: In the past, the lecture substitution system used only presentation slides and textbook information, and the problem was that the lecture was boring for the students because the robot continued to talk monotonously and unilaterally. Therefore, in this research, we developed a system to provide more diverse and interactive lectures that allows students to participate actively by incorporating a system that allows students to respond from a Web page to multiple-choice questions automatically generated from Wikipedia data, and evaluated impressions. An experiment was performed. As a result, the proposed system exceeded the conventional system by 0.19 points in the average of 15 items with 7-level evaluation from 1 to 7. Therefore we confirmed the effectiveness of the proposed method.

Keywords: Humanoid Robot, Wikipedia, Multi-choice Quiz, Sentence Generation

1. はじめに

近年、高性能なヒューマノイドロボットの登場により、ヒューマノイドロボットを用いて講義を行う試み[1]がなされている。実社会でこのような講義を行うことができれば、人間の労力の大きな削減となる。しかし、多くの場合講義時に用いるシナリオは人手で作成されたものであり、労力が大きくかかるのが現状である。そこで、我々の従来研究[2]では、教科書の内容をもとにシナリオを自動生成することで、労力の削減を試みた。その結果、労力の削減という意味では一定の成果が得られたものの、人間による講義には劣る面が多く、特に講義を退屈に感じる受講者が多いという課題が残った。その理由として、生成されたシナリオは教科書から抽出した内容のみを用いたものであること、ロボットが一方的に話すシステムであったことなどが挙げられた。そこで、受講者が退屈せず集中力を保てるように、受講者が能動的に講義に参加できるような要素を持った講義を行うべきであると考えた。そのよ

うな要素として、講義に関連する単語と意味の一致を問う多肢選択問題出題システムの導入を試みた。講義代行システムの導入理由である労力の削減という面を損なわないよう、多肢選択問題は自動で生成し、Web 回答システムを通して問題に回答する形で受講者が能動的に参加できる講義を設計した。ここで、本稿における講義は大学生を対象にしたもので、多人数の受講者に対しヒューマノイドロボットが説明するという形式のものを指す。

本稿では、提案手法を用いて生成した多肢選択問題を組み込んだ講義代行システムによる実験を行い、印象評価と集中力に関する評価の結果とその分析結果について述べる。

2. 関連研究

多人数の受講者に能動的に授業に参加してもらうためには、多肢選択問題を出題することが有効な手段の一つであるとされている[3]。また、多肢選択問題は選択肢を選んで答えるという性質上、自由記述問題や

¹ 北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology Hokkaido
University

対話形式の問題などと比べ、ロボットと人との情報のやりとりが行いやすいという利点もある。そこで、多肢選択問題の導入を試みることにした。多肢選択問題生成の研究として、田村らの研究[4]がある。ここでは DBpedia Japanese[5]の抽象情報を変換し、平叙文から問題文を生成することをやっている。この研究では日本の歴史上の人物を対象にして問題を生成しているが、提案手法では主に専門用語を対象にして問題を生成する点が異なっている。また、田村らの研究[6]では、多肢選択問題の選択肢としては DBpedia のリンク情報を用いて共通の属性をもつ人物を選択肢としている。しかし、この方法では属性を持たないものに関しては選択肢を生成できず、専門用語は属性を持たないものが多かったため適さなかった。提案手法では選択肢生成の方法として、代わりに Word2vec[7]により得た単語分散表現を用いて計算した単語間類似度が高いものを選択肢として用いる。また、田村らの研究では講義に対して多肢選択問題を応用してはなかったが、本稿では生成した問題を Web ページ上で問題に答える形式として過去の我々の研究の提案手法による講義代行システムに組み込み、講義形式で実験を行った。

3. 提案手法

3.1 処理過程

提案手法である講義代行システムの処理過程を図 1 に示す。実験を行う際には、ヒューマノイドロボット Pepper[8]を用いて講義シナリオを読み上げ、途中で Web ページを用いた回答システムを併用して多肢選択問題を出題するという形で行う。また、本手法では、このうちプレゼンテーションスライドの内容に合致する講義シナリオと、それに関連する多肢選択問題を自動生成する。

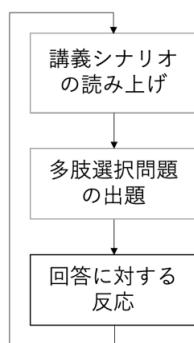


図 1 処理過程

3.2 講義シナリオの生成

はじめに、ロボットが読み上げる講義シナリオを生成する。シナリオの生成手法には、文献[2]で我々の提案したものと同じものを用いる。この手法では、大学

の講義で実際に使用されているスライドの内容に合致する内容を対応する教科書から抽出することでシナリオを生成する。また、抽出した文を話し言葉に変換する処理を行うことで、より人間の発話に近い文体のシナリオを生成する。これらの処理によって、最終的にシナリオはスライドの枚数と同じ数のテキストファイルとして出力される。

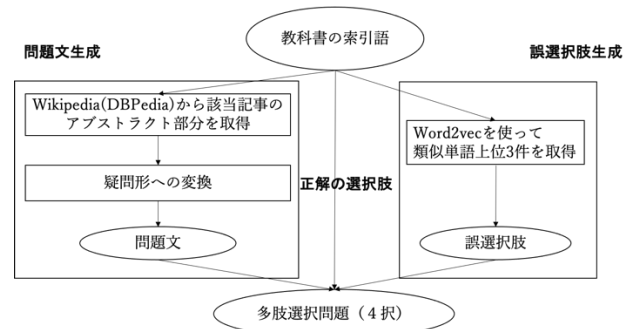


図 2 多肢選択問題生成過程

3.3 多肢選択問題の生成

多肢選択問題の生成について述べる。はじめに、問題文の生成を行う。これには、関連研究で挙げた田村らの手法[4]を参考にした。データセットとして、DBpedia Japanese が提供する Wikipedia の記事タイトルと抽象部分のセットのデータを用いる (long_abstracts_ja.ttl, 487MB, データ数は 934772)。処理の手順としては、はじめに入力となる単語に対し、データセットからその単語に対する Wikipedia の抽象部分を取り出す。次に、抽象部分の 1 文目のみを抽出する。これは 1 文目がタイトルの定義を説明する場合が多く、専門用語の問題を作るという用途に適していると判断したためである。その後、田村らの手法によりその文を問題文の形式に変換する。ここで、本手法では文末表現を田村らの手法とは別のものを用いるよう変更した。この変更点を表 1 に示す。この変更の理由は、田村らの手法では人物についての問題であったため疑問詞を「誰」としていたが、本手法では一般的な講義に対応させるため、また人物に限らない問題文とする必要があったためである。そしてその場合、多肢選択問題としては「どれ」が適切であると考えた。

続いて、選択肢の生成を行う。選択肢の数は田村らの手法[6]と同じく 4 つとし、そのうち 1 つには問題文の正答となる単語、すなわち前述した問題文を生成する際の入力単語を用いる。残り 3 つの誤選択肢については、問題文の答えとなる単語に対して、Word2vec により生成された全ての単語ベクトルの中からコサイン類似度が最も高いものを上から順に 3 つ取得したも

のとした。このようにしたのは、正解の単語と意味が異なる単語であるだけでなく、正解の単語と意味的な類似度が高く消去法で選びにくいような誤選択肢を生成するためである。これにより、多肢選択問題の難易度が易しくなりすぎないようにした。問題文の例を図3に示す。

表1 変更した文末表現

変更前	変更後
「したのは誰ですか？」	「したのは次のうちどれですか？」
「であるのは誰ですか？」	「であるのは次のうちどれですか？」
「のは誰ですか？」	「のは次のうちどれですか？」

<p>(例1) 問題文: 浮動小数点数の一部で有効数字を構成する部分であるのは次のうちどれですか? 選択肢: 仮数, 小数, 最上位ビット, 整数 正解: 仮数</p>
<p>(例2) 問題文: コンピュータなどにおいて中心的な処理装置として働く電子回路のことであるのは次のうちどれですか? 選択肢: CPU, プロセッサ, GPU, ハードウェア 正解: CPU</p>

図3 問題文の例

3.4 Web ページを用いた多肢選択問題の出題

先述の多肢選択問題の出題方法として、多肢選択問題を表示し、回答できる Web ページを作成した。これを図3に示す。Web ページには問題番号、問題文、4つの選択肢、回答の送信ボタンを表示し、受講者が回答できるようにした。なお、Web ページ上の回答の送信ボタンを押すと、次の問題が表示される。また、ロボットが話す内容のテンプレートを図4に示す。多肢選択問題出題の流れとしては、まずロボットが受講者に回答を促し、受講者が1分の回答時間内に答え、回答時間終了後にロボットが問題の正答の単語と正解率を話し、そして再度シナリオに沿って講義を続けるという形式をとった。より受講者が集中して講義に臨めるよう、問題を出すだけでなく、ロボットから正答と正解率を伝えることで受講者全体に対してフィードバックを与えるように設計した。

Web システムを用いた理由としては、受講者全員が参加できるため全員の注意を引くことができることと、

ロボットとの情報のやりとりが音声認識や文字認識を使う場合と比べ行いやすいことが挙げられる。

また、生成された複数の問題のうちどの問題を講義で用いるかという点に関しては、今回人手を使って行なった。今後はこれに関しても自動で行うようにしたいと考えている。

発話1: ここで、問題です。 発話2: ウェブページ上の問題<問題番号>に教えてください。 発話3: <問題文> 発話4: 回答時間は、一分です。 (1分間待機する) 発話5: 回答終わりです。 発話6: 問題<問題番号>の答えは、<正解>でした 発話7: 問題<問題番号>の正解率は、<正解率>パーセントでした。
--

図4 問題出題発話のテンプレート

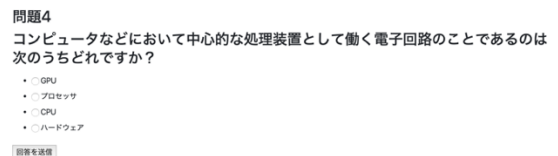


図5 Web システムの例

3.5 生成された問題についての考察

提案手法により多肢選択問題を生成した結果について考察を行う。まず、今回生成された問題数は62個であった。適切な問題が生成された例としては、図3に示す問題が挙げられる。これは正解の選択肢と意味の近い誤選択肢が生成されており、問題文もこれを読んだだけで正解を選択肢の中から一つに特定するために十分な情報を含んでいる。一方で、適切でない問題の例として、一つは選択肢に同義語が含まれる図6のような例が挙げられる。このような例は62個中9個(14.5%)生成された。この問題では、bitとビットという同じ意味をもつ単語が選択肢として含まれてしまうことがあった。同じような例としては、「コンピュータ」と「コンピューター」、「ソフトウェア」と「ソフトウエア」などもあった。本手法の目的は一つの選択肢のみを正解とする多肢選択問題を生成することであるため、このような選択肢は適切ではない。改善策としては、同義語の辞書を用いて同義語を除去することや、表記揺れの統一を行うこと、選択肢の間で単語間類似度を測り、その類似度が非常に高い値のものを誤選択肢から除去することなどが考えられる。もう一つの適切でない例として、多肢選択問題の選択肢には本稿の場合正解の選択肢は一つである一方で、実際に生成された選択肢には正解となる選択肢が複数存在することもあった。その例を図7に示す。このような例は

6 2 個中 4 個 (6.5%) 生成された。こういった問題文が曖昧な表現であった場合、正解の選択肢のみから誤選択肢を生成する手法では対応できないため、問題文の情報も用いて生成することが必要であり、問題文の情報を用いない本手法ではこのような例を防げなかった。今回は問題文を Wikipedia のアブストラクトの 1 文目をもとに生成したが、2 文目以降も用いることで、一意に正解の選択肢を決定できるような問題文を生成できる可能性があると考えられる。

問題文: ほとんどのデジタルコンピュータが扱うデータの最小単位であるのは次のうちどれですか?
 選択肢: ビット, bit, キロバイト, メモリ
 正解: ビット

図 6 選択肢に同義語が含まれる例

問題文: 哲学および精神分析学における概念であるのは次のうちどれですか?
 選択肢: 自我, 理性, 人格, 肉体
 正解: 自我

図 7 正解の選択肢が複数考えられる例

4. 性能評価実験

4.1 実験設定

性能評価実験として、ヒューマノイドロボット Pepper と多肢選択問題の出題 Web システムを用いた講義を行った。講義は 4 種類を行い、理系大学生及び大学院生計 6 人を 3 人ずつに分け、それぞれに 2 種類ずつを振り分けた。講義と被験者グループの組み合わせを表 3 に示す。この表 3 のように、各被験者には 1 回目と 2 回目で同じ講義内容を受けないような組み合わせにし、問題を出題する場合としない場合で 2 回受けてもらうことにした。これらの理由としては、一つは同じ内容の講義を 2 回受ける場合、2 回目の講義では問題の出題を除きほとんど同じ内容となり、生徒の興味や集中力が著しく低下すると考えられるためである。2 つ目の理由としては、講義の内容に関係なく問題の出題が有効かどうか確かめるためである。

多肢選択問題生成システムの入力単語には教科書の索引語を用い、これにより生成された 65 個の問題から人手で 2 種類の講義内容に応じて 4 つずつを選び、それを講義中に使う問題とした。具体的には、プログラム内蔵方式、補数、文字コード、CPU、アドレスバス、主記憶装置、仮数、ビットを正解とする 8 の問題である。講義ごとに 4 つをランダムに入れ替え、その

単語がシナリオ中に出現する直前と直後に 2 つずつ出題するようにした。

性能評価としては、実験後にアンケート調査を実施した。アンケート項目は 15 項目の形容詞対を用いたものと 1 項目の質問にわけ、合計 16 個の設問に 1 から 7 の 7 段階で評価をするよう指示した。15 項目の形容詞対は、ロボットの性能を評価する際に用いられる GodSpeed[9]を参考に設定した。最後の項目については、GodSpeed の「安全性」の代わりに独自に追加したものであり、「講義としての有効性」のカテゴリとしている。

表 2 講義と被験者グループの組み合わせ表

被験者グループ	講義内容	順番	問題の出題
A	データ表現	前	提案手法
A	コンピュータの基本構成	後	ベースライン
B	データ表現	前	ベースライン
B	コンピュータの基本構成	後	提案手法

4.2 実験用データ

実験用データとして、講義シナリオの生成には実際に大学での講義に使用された教科書[10]と講義スライドを用いた。この講義スライドは教科書の著者が作成したものである。内容としては、データ表現、コンピュータの基本構成に関する 2 種類の講義に該当する部分を使用した。この実験データを用いて、我々の開発した手法[2]により講義シナリオを生成した。そして、問題の生成対象となる単語群には、今回使用した教科書の索引語を用いた。これは、索引語が講義において重要な意味を持つ単語であると考えられるからである。また、選択肢生成の際に利用した Word2vec の学習コーパスは Wikipedia のダンプデータ (7GB) [11]を用いた。本来学習コーパスには講義データが考えられるが、十分な量の講義データを集めることができなかったためである。

4.3 実験結果

実験結果を表 3、表 4、そして図 8 に示す。表の各数値は、被験者がつけた各項目の評価値の平均値である。表 4 に示すように、提案手法を用いて行なった講義が従来手法を用いた講義を全 15 項目の平均値で 0.19 ポイント上回った。また、5 カテゴリのうち 4 カテゴリの平均値が上回っており、下回ったのは「知性認識」のカテゴリのみであった。加えて、表 3 に示すように、「集中力を保つことができたか」については 1.

表 4 形容詞対によるアンケート結果

指標名	項目	ベースライン	提案手法
擬人観	機械的な-人間的な	2.33	2.5
	人工的-生物的	2.33	2.5
	偽物のような-自然な	2.83	3
知性認識	愚かな-賢明な	4	4.5
	知的でない-知的な	4.67	4.17
	無知な-物知りな	5.67	4.67
好ましさ	嫌い-好き	4	4.17
	親しみにくい-親しみやすい	3.33	4
	不愉快な-愉快的な	3.5	4.33
有生性	熱心でない-熱心な	2.83	3.83
	非活発な-活発な	3	3.17
	活気のない-生き生きとした	3	3
講義の有効性	わかりにくい-わかりやすい	3.17	3
	興味深くない-興味深い	4.83	4.83
	飽きのくる-飽きのこない	2	2.67
	全項目の平均値	3.43	3.62

表 3 個別項目のアンケート結果

項目	ベースライン	提案手法
集中力を保つことができたか	2.33	3.67

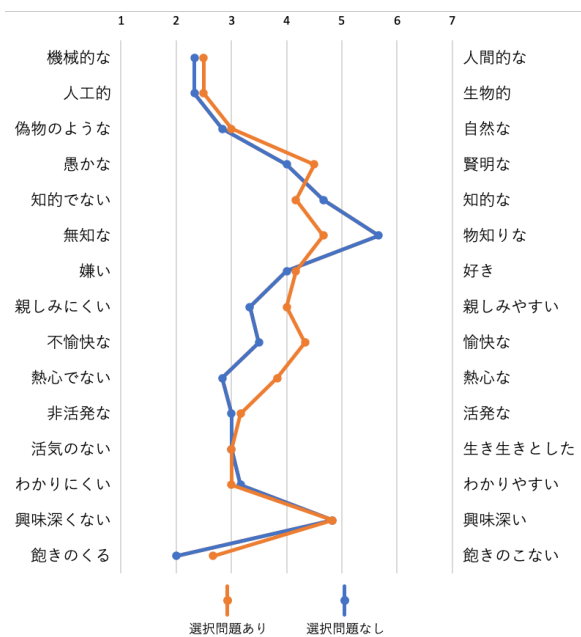


図 8 形容詞対によるアンケート結果のグラフ

34 ポイント向上し、提案手法が受講者の講義への集中力

向上においても有効であることが確認された。

4.4 考察

提案システムは従来システムを全 15 項目の平均値で 0.19 ポイント上回った。これは総合的により評価の高い講義となったということを示している。しかし、著しい改善とは言えない結果となった。

提案手法によって向上した部分について考察する。まず、「好ましさ」のカテゴリが向上していた。個別の項目としては、特に「不愉快な-愉快的な」、「親しみにくい-親しみやすい」が向上しており、これらの理由としては、ロボットが一方向的に話すのではなく、受講者に話しかける形で回答を促し、受講者が回答するというロボットとのインタラクティブな要素を付け加えたことが親しみやすさを向上させた結果につながったと考えられる。また、「有生性」のカテゴリも向上し、特に「熱心でない-熱心な」が向上していた。この理由としては、スライドに沿った内容を淡々と読み上げるだけの従来手法よりも、問題の出題があった分、教え方に工夫をされていて熱心であるという印象を与えたのではないかと考える。また、「集中力を保つことができたか」の項目に関しては提案手法が 3.67 と 7 段階中 1 以上の向上が見られたことから、十分な有効性があったと言える。その要因としては、多肢選択問題への回答を通して集中力を保ちながら講義に参加することができていたためであると考えられる。加えて、アンケートの自由記述欄には、「問題をとくことがないと集中できなかった」との記述

もあった。この被験者は1回目の講義に多肢選択問題の出題があったため、2回目に多肢選択問題がないことに講義の単調さや退屈さを感じ、集中力を失ってしまったと考えられる。

次に、従来システムよりも平均値が低くなった点や改善が必要な点について考察する。「知性認識」のカテゴリは、カテゴリの平均値が唯一低下した。その原因として、多肢選択問題出題システムについては作成された問題を出しているだけで、ロボットが持つ知性として感じなかったということが考えられる。あるいは多肢選択問題の一つに選択肢の意味が同じものが含まれていたことも影響があると考えられる。また、提案手法は15の形容詞対の評価値の平均値が依然として中央値である4を下回ったことから、さらなる改善が求められる結果であった。全体的に評価を上げるためには、講義の大部分を占めるスライドの内容に沿って話すシナリオ生成と発話の部分の改良が必要だと考える。

最後に、多肢選択問題の各問題の正答率を表5に示す。この結果から、全ての問題が正解率100%や0%というわけではなく、問題の難易度としては簡単すぎることも難しすぎることもない適度な難易度で誤選択肢が生成できていたと考える。

表5 問題と正答率

講義内容	問題番号	正解率
データ表現	問題1	33%
データ表現	問題2	33%
データ表現	問題3	33%
データ表現	問題4	100%
コンピュータの基本構成	問題1	67%
コンピュータの基本構成	問題2	67%
コンピュータの基本構成	問題3	100%
コンピュータの基本構成	問題4	100%

5. まとめ

本稿では、自動生成した講義シナリオ及び多肢選択問題出題システムを用いたヒューマノイドロボットによる講義形式の実験を行い、提案手法が従来手法の性能を多くの指標で上回ったことを示した。特に受講者の集中力の向上に一定の成果が見られ、多肢選択問題出題システムがヒューマノイドロボットを用いた長時間の講義でも十分有効に働くことがわかった。今後は、生成された問題の中から出題する問題を人手で選ぶ部分を自動化することで、さらなる労力の削減を行いたい。また、選択肢の生成方法についても同義語の排除や類似度の非常に高い語を排除することで一意な正解選択肢を生成できるよう改善を行う予定である。

参考文献

- [1] Takuya Hashimoto, Igor M. Verner Hiroshi-Kobayashi, Human-Like Robot as Teacher’s Representative in a Science Lesson: An Elementary School Experiment, Robot Intelligence Technology and Applications, pp. 775-786, 2012.
- [2] 芳尾哲, 荒木健治, “ヒューマノイドロボットを用いた講義代行システムのための講義シナリオ文自動生成手法”, 電気・情報関係学会北海道支部連合大会, pp218-219, 2017.
- [3] 佐藤浩章, シリーズ大学の教授法2, 講義法, 2017
- [4] 田村 吉宏, 山内 崇資, 林 佑樹, 中野 有紀子, Wikipedia 記事情報に基づく歴史学習問題の自動生成手法, The 29th Annual Conference of the Japanese Society for Artificial Intelligence, 2014
- [5] DBpedia Japanese, <http://ja.DBpedia.org/>
- [6] 田村 吉宏, 鶴崎泰斗, 高瀬 裕, 中野 有紀子, Wikipedia を用いた質問応答と多肢選択問題による歴史学習 The 30th Annual Conference of the Japanese Society for Artificial Intelligence, 2015
- [7] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado. Jeff Dean, Distributed representations of words and phrases and their compositionality, In Advances in neural information processing systems, pp. 3111-3119, 2013.
- [8] Pepper, <https://www.softbank.jp/robot/pepper/>
- [9] Christoph Bartneck, Dana Kulic and Elizabeth Croft. “Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots.” In Proceedings of the Metrics for Human-Robot Interaction Workshop in Affiliation with the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 71–81, 2008.
- [10] 荒木健治, コンピュータ工学概論-コンピュータはなぜ計算ができるのか?-, オーム社, 東京, 2013.
- [11] Wikimedia Downloads, <https://dumps.wikimedia.org/jawiki/latest>