

単語の重要度と難易度を考慮した講義スライド抽出に基づく 学習コンテンツ生成方式の提案

古川 一輝[†] 王 元元[†]

[†] 山口大学工学部知能情報工学科 〒755-8611 山口県宇部市常盤台 2-16-1

E-mail: †{b066ff,y.wang}@yamaguchi-u.ac.jp

あらまし 新型コロナウイルスの影響により、講義スライドを用いたオンデマンド型授業が普及してきた。オンデマンド型授業では、ユーザはいつでもどこでも講義スライドで用いて学習することが可能となった。しかしながら、オンデマンド型授業の講義スライドが大量に存在し、ユーザの関心や興味によって内容の理解度が異なるため、講義スライドに出現している単語に対して、「理解できない、知識が深まらない、興味が広がらない」といった問題が生じてしまう。そこで本研究では、講義スライドの特徴語の重要度を算出し、その特徴語の Wikipedia のリンク情報を用いて特徴語の難易度判定を行い、各特徴語の重要度と難易度の高低に基づき予習・復習用の学習コンテンツを生成する方式を提案する。本論文では、講義スライドから特徴語の抽出を行い、特徴語の難易度判定およびスライドの重要度推定を行うことで、初級・中級・上級の3つの知識レベルに分類された学習コンテンツを生成する。最後に、生成した学習コンテンツを用いて被験者による評価実験を行い、本提案方式の有用性を検証する。

キーワード オンデマンド型授業, 講義スライド, 特徴語抽出, 難易度判定, 学習支援

1 はじめに

近年、新型コロナウイルスの影響によりオンデマンド型オンライン授業が普及してきたことで、ユーザはいつでもどこでも講義スライドを閲覧可能となり、より学習しやすい環境が整ってきた。しかしながら、オンデマンド型授業で使われる講義スライドには、30ページや40ページにも上る多数のスライドで構成されるものもあるため、予習も復習もしにくいといった問題が生じる場合がある。さらに、ユーザの事前知識や授業内容に対する関心や興味によって内容の理解度が異なるため、その講義スライドに出現している単語に対して、「理解できない、知識が深まらない、興味が広がらない」といった問題が生じてしまう。その結果、学習意欲を失ってしまい、授業内容を理解できないまま勉強を放棄してしまう可能性がある。そこで、ユーザにとって、より予習・復習しやすい学習環境が求められる。

そこで本研究では、講義スライドにおける特徴語の難易度とユーザの知識レベルを考慮した学習コンテンツを生成し、より予習・復習しやすい学習環境を構築することを目的としている。具体的に、本研究では、講義スライドから特徴語を抽出し、その特徴語の重要度算出および難易度判定を行い、各特徴語の重要度と難易度の高低に基づき初級・中級・上級の3つのレベルに分類された特徴語が出現するスライドの重要度推定を行うことで、初級者・中級者・上級者向けの3種類の学習コンテンツを生成する。難易度判定には、不特定多数の利用者が投稿、編集可能なオンライン百科事典 Wikipedia¹を用いる。各スライドに出現する特徴語に対して、Wikipediaでの被リンク数に基づき難易度を算出する。

本論文の構成は以下のとおりである。2章では、本研究の概要と関連研究について述べる。3章では、講義スライドにおける特徴語の重要度算出と難易度判定手法について説明する。4章では、各特徴語の重要度と難易度高低に基づき初級・中級・上級の3つのレベルに分類された特徴語が出現するスライドの重要度推定を行い、それらを基に初級者・中級者・上級者向けの3種類の学習コンテンツを生成する手法について説明する。5章では、提案手法で生成した学習コンテンツの有用性を図るための評価実験について述べる。最後に、6章では、まとめと今後の課題について述べる。

2 研究概要と関連研究

2.1 研究概要

本研究では、スライド内から抽出された特徴語の重要度算出および Wikipedia での被リンク数に基づき特徴語の難易度判定を行う。さらに、各特徴語の重要度と難易度の高低に基づき特徴語が出現するスライドの重要度を推定することで、図1のように初級者、中級者、上級者向けの3種類の学習コンテンツを生成する。ここで、重要度が低く被リンク数が多い特徴語は、3つの知識レベルに当てはまらないため、「無し」とする。

この3種類の学習コンテンツの中で、初級者向けのコンテンツには、スライドにおいて内容がやさしく（単語の難易度が低い）、その講義を理解するのに必要となる内容（単語の重要度が高い）が分類される。その講義において基礎的な単語である。中級者向けのコンテンツには、スライドにおいて内容が難しく（単語の難易度が高い）、その講義において重要な内容（単語の重要度が高い）が分類される。その講義で理解しなければならない、覚えなければならない単語が中級者に当たる。上級者向

1: <https://ja.wikipedia.org/wiki/>

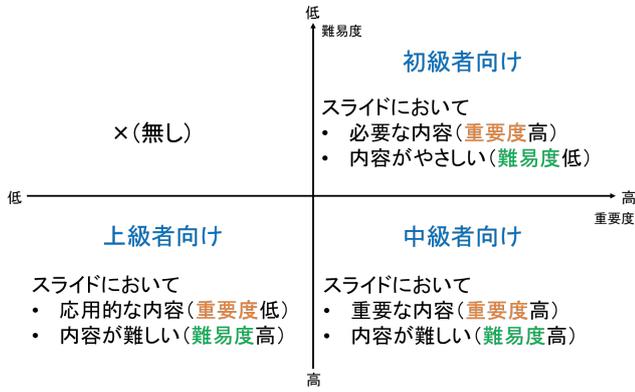


図 1 特徴語の重要度と難易度に基づく 3 種類の学習コンテンツ

けのコンテンツには、そのスライドにおいて内容が難しく（単語の難易度が高い）、その講義において応用的な内容（単語の重要度が低い）が分類される。初級者と中級者に比べて単語の専門性が高くなる。

2.2 関連研究

近年、講義スライドを基にした情報取得やスライドの自動推薦、再構成などの学習支援に関するさまざまな研究やシステムの開発が活発に行われている。特に、学生の批判的思考を促進する質問の生成および質問と関連したスライドの自動推薦する研究 [1]、スライドのインデントの階層構造を利用して質問に関するスライドにおける単語の寄与率を求め、値の高い順で関連スライドを推薦する研究 [2]、スライド全体の専門性を判定し、それに伴いスライドごとの特徴語抽出によりスライドの内容的なグループ化を行い、各スライドグループの内容に関して重要度に基づき新たなスライド生成および既存のスライド削除を行うスライドの再構成を行う研究 [3] などがあげられる。

また、スライドの階層構造に基づき単語を抽出する先行研究も行われている。Wang ら [4] はプレゼンテーションの文脈を考慮した分析手法により講義スライド全体の特徴語を抽出する手法を提案している。この研究では、スライドの階層分析を行い、特徴語にふさわしい単語に重みを付け、重み付けの値が閾値となる単語を特徴語として抽出している。本研究では、こちらの手法を用いて特徴語抽出を行う。

単語の難易度推定においては、教科書コーパス語彙表 [5] を単語難易度の基準データとし、これを学習データとした機械学習によって単語難易度を推定する手法 [6] や単語の出現頻度と視点情報を用いた文書上の主観的高難易度単語の推定 [7] などの研究が行われている。西原ら [8] は展望台システム [9] を用いて特徴語を抽出し、特徴語の出現頻度を用いて難易度を判定する手法を提案している。この研究では、特徴語は出現頻度が低いほど難しいと考え、Web ページにおける特徴語の出現頻度を用いて特徴語の難易度を判定する。本研究では、特徴語の重要度と Wikipedia での被リンク数に着目し、それら 2 つの指標を用いた難易度推定手法を提案する。

スライドの重要度推定においては、スライドの専門性や重要性、階層構造などに基づき重要度を測る研究が行われている。

山田ら [10] は講義スライドにおける重要性に基づきスライドの重要度を推定する手法を提案している。この研究では、教師が講義スライドに付した重要性に対する主観評価値を重要度と定義し、非言語情報から講義スライドの重要度を抽出している。本研究では、特徴語が出現したスライドの出現階層と出現頻度を基にスライドの重要度を推定する。

また、リンク情報を用いた研究も行われている。伊藤ら [11] は Wikipedia のリンク共起性解析によるシソーラス辞書構築手法を提案している。稲子ら [12] は自然言語処理や Wikipedia のリンク情報、カテゴリ構造を利用し、自己学習を振り返るための学習履歴の可視化を行った。田中ら [13] は文書が持つ単語の情報 [14] と、リンク情報を融合した文書分散表現 [15] 取得手法を提案している。小迫ら [16] は閲覧者の動向を考慮した PageRank アルゴリズム [17] を利用し、企業の Web サイト内の重要なページを判別する手法を提案している。

3 単語の重要度算出と難易度判定

本章では、講義スライド内から特徴語を抽出し重要度を算出する手法、および Wikipedia での被リンク数を用いた特徴語の難易度判定手法について述べる。

3.1 講義スライドにおける特徴語抽出と重要度算出

本研究では、既存の分析手法 [3] を用いて特徴語抽出を行う。具体的には、スライドの階層構造の分析を行い、特徴語に適した単語に重みを付け、重み付けの値が閾値以上となる単語を特徴語として抽出する。スライドのプレゼンテーションには説明の流れがあり、文脈を表すような単語は、あるスライドではタイトルに出現し、他のスライドでは本文で出現するといったようにスライドによって出現する階層が異なる場合がある。プレゼンテーションに出現する特徴語も同じようにスライドによって出現階層が異なる場合があるため、単語の重み付けの前処理として、説明の流れにおけるスライドごとの出現階層の違いを、ある単語 k の出現階層が変わるスライド s_i とスライド s_j というペア T として抽出する。 T は以下の式を満たす単語 k に関するスライド s_i とスライド s_j のペアの集合とする。

$$V = \{(k, s_i, s_j) | l_{max}(k, s_i) \neq l_{max}(k, s_j)\} \quad (1)$$

$l_{max}(k, s_i)$ はスライド s_i 内で単語 k が出現する最も上位の階層である。スライドのタイトルを階層構造の最上位である 1 とする。インデントの始まりを 2 とし、階層が深くなるたびに 3, 4 と 1 ずつ増えていくものとする。

次に、単語 k の出現階層が変化するスライドのペアの集合 T に含まれるスライドを用いて、単語の k の重み付けの値 $I(k)$ を以下の式により算出する。

$$I(k) = \frac{1}{l_{max}(k, s_i)} + \sum_{k \in T} \left(\Delta * \frac{1}{dt(s_i, s_j)} \right) \quad (2)$$

$$\Delta = \left| \frac{1}{l_{max}(k, s_i)} - \frac{1}{l_{max}(k, s_j)} \right|$$

Δ は異なるスライドでの出現階層の違いを表し、単語 k のス

ライド s_i と s_j における出現階層の変化量を返す関数である。 $dt(s_i, s_j)$ は T に含まれるスライド s_i と s_j のスライド間の距離を関数であり、隣接するスライドであれば 1 を返す。式 (2) において、スライド内で高い階層に出現し、距離が近いスライドに集中して出現する単語の値を高くする。

3.2 Wikipedia リンク情報に基づく難易度判定

本研究では、難易度の高い単語は講義内であり説明されておらず、一般性が低く、多くの人から周知されていないという仮説を立て、その仮説に基づき研究を行う。そこで、3.1 節で抽出した特徴語の被リンク数を用いて特徴語の難易度判定を行う。被リンク数の取得には、世界最大規模の百科事典である Wikipedia を用いる。Wikipedia は誰でも編集できるという特徴があり、多数の編集者に編集された結果、情報の正確性が近年向上しているため、本研究で活用する。Wikipedia は、より多くの人に周知されている項目については、その項目へのリンク数が多く、周知されていない項目については、その項目へのリンク数が少ないことから、多くの人から認知されているか否かを判定できる。つまり、ある特徴語に対するリンク数、すなわち特徴語の被リンク数を用いることで特徴語の難易度を判定する。ある特徴語について被リンク数が多い場合、難易度が低いと判定される。ある特徴語について被リンク数が少ない場合、難易度が高いと判定される。

4 重要度と難易度を用いた学習コンテンツの生成

本章では、3 章で求めた特徴語の重要度と難易度に基づきスライドの重要度推定および学習コンテンツの生成について述べる。

特徴語の重要度と Wikipedia での被リンク数にそれぞれ閾値を設けて、初級者、中級者、上級者向けの 3 つの知識レベルに振りわけると、特徴語の重要度の閾値 α は、全ての特徴語の重要度の平均値とし、特徴語の被リンク数の閾値 β は、全ての特徴語の被リンク数の平均値とする。特徴語の重要度が高いほど、よく説明されているため、その講義を理解するために必要な単語であり、特徴語の重要度が低いほど、あまり説明されていないため、その講義において応用的な単語である。また、特徴語の被リンク数が多いほど一般性が高く、よく知られている単語であるため難易度が低く、特徴語の被リンク数が少ないほど専門性が高く、あまり知られていない単語であるため難易度が高くなる。そして、各特徴語の重要度と難易度の高低に基づき「初級・中級・上級」の 3 つの知識レベルに振りわけると、

- 初級者向けの特徴語
特徴語の重要度 $\geq \alpha$ (重要度高)
特徴語の被リンク数 $\geq \beta$ (難易度低)
- 中級者向けの特徴語
特徴語の重要度 $\geq \alpha$ (重要度高)
特徴語の被リンク数 $< \beta$ (難易度高)
- 上級者向けの特徴語
特徴語の重要度 $< \alpha$ (重要度低)

特徴語の被リンク数 $< \beta$ (難易度高)

4.1 スライドの重要度推定

知識レベル「初級・中級・上級」に振り分けられた特徴語のみを対象とし、講義スライド全体において、それらの特徴語が出現する全てのスライドの重要度を 1 枚づつ推定する。以下の式を用いて特徴語が出現したスライドの重要度を 1 枚づつ算出する。

$$Weight(k, s_i) = \sum \left(\frac{1}{n(k, s_i)} * c(k, s_i) \right) \quad (3)$$

$n(k, s_i)$ はスライド s_i において特徴語 k が出現した階層番号、 $c(k, s_i)$ はスライド s_i 内の階層番号 $n(k, s_i)$ において特徴語 k が出現した回数を表す。式 (3) では、スライドにおいて特徴語の出現階層が高く、集中してよく出現している場合はスライドの重要度が高くなる。

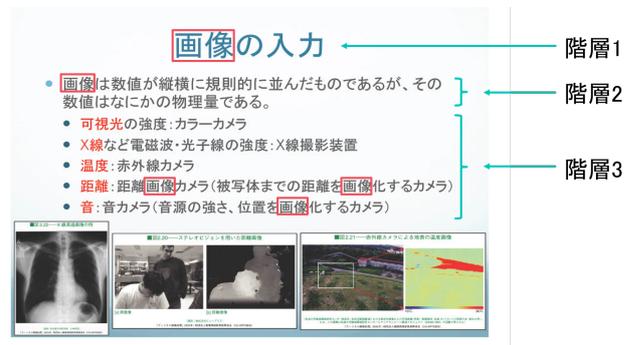


図 2 スライドの重要度算出例

図 2 にスライドの重要度算出の例を示す。この例を用いてスライドの重要度推定手法について説明する。この例では、特徴語「画像」が出現するスライドの重要度を測る。「画像」は階層 1 に 1 回、階層 2 に 1 回、階層 3 に 3 回出現している。これらの出現階層番号と出現回数を式 (3) に当てはめると、以下の式となる。

$$\text{「画像」の重要度} = \frac{1}{1} * 1 + \frac{1}{2} * 1 + \frac{1}{3} * 3 = 2.5$$

このように算出した重要度を基に特徴語ごとのスライドの重要度の高さ順でランキングを付ける。

4.2 重要度に基づくスライド抽出

図 3 に学習コンテンツ生成の手順を示す。4.1 節で求めたスライドの重要度を基に学習コンテンツを作成するためのスライドを抽出する。具体的には、ランキング付けしたスライドに閾値を設け、その閾値以上の重要度を持つスライドを抽出する。ここで、閾値は特徴語ごとのスライドの重要度の平均値とする。抽出したスライドを学習コンテンツとする。

5 評価実験

本章では、生成した学習コンテンツの有用性に関する評価を行う。本研究では、講義スライドの特徴語の重要度を算出し、

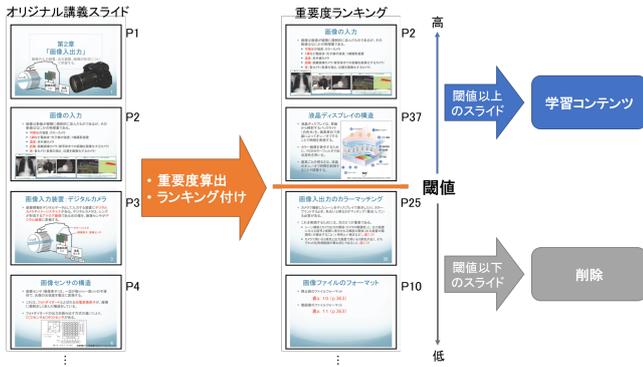


図3 学習コンテンツ生成の手順

表1 抽出した初級者・中級者・上級者向けの特徴語

講義スライド	初級	中級	上級
第1回コンピュータ基礎	情報	コンピュータ	世界
第2回画像処理	構造	画像	光, 電圧, カメラ
第2回認知科学	心身二元論, デカルト	哲学	神
第3回画像処理	収差, 系	色, 人間, カメラ	—

その特徴語の Wikipedia のリンク情報を用いて特徴語の難易度判定を行い、各特徴語の重要度と難易度の高低に基づき3種類の学習コンテンツを生成した。

5.1 実験概要

本実験は、実際に講義で使用されている4つのオリジナル講義スライド、「第1回コンピュータ基礎」「第2回画像処理」「第2回認知科学」「第3回画像処理」を対象とし、提案手法により、1つのオリジナル講義スライドにつき、初級者・中級者・上級者向けの3種類の学習コンテンツを生成した。表1に抽出した初級者・中級者・上級者向けの特徴語を示す。なお、「第3回画像処理」の知識レベル「上級」については、特徴語が振りわけられなかったため、生成した学習コンテンツは「第3回画像処理」の「上級者」を除いた計17個となった。評価実験では、生成した学習コンテンツとオリジナル講義スライドを比較するために、学習コンテンツ17個とオリジナル講義スライド4個の計21個のコンテンツを用いて評価を行う。

4つのオリジナル講義スライドと生成した初級者・中級者・上級者向けの3種類の学習コンテンツに対して、15名の大学生にアンケート調査を実施した。なお、実験を行う際に、まず被験者に講義スライドのタイトルのみを確認していただき、その講義に対して自身の知識レベルを初級・中級・上級で記入してもらった。アンケート調査の設問項目は以下の通りである。

- Q1: この講義スライドでのあなたの知識レベルをお教えてください
- Q2: 講義スライド全体の内容が理解できた
- Q3: 講義スライドの内容に関する知識が深まった
- Q4: 講義スライドの内容に対する興味・関心が広がると感じた
- Q5: 今回の実験で感じたこと、ご意見などがございましたらご記入をお願いします

これらのアンケートでは、Q2~Q4は5段階（1:全くそう

思わない, 2:あまり思わない, 3:どちらともいえない, 4:ややそう思う, 5:とてもそう思う)で評価してもらった。

5.2 理解度・知識・興味に関する評価結果

5段階評価によるQ2~Q4の評価平均値を図4~6に示す。

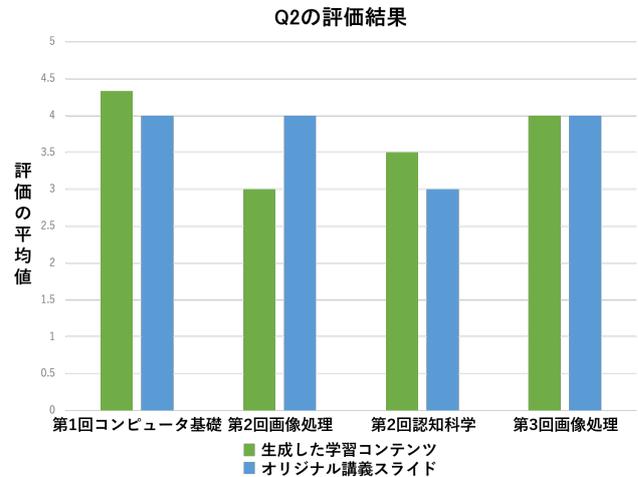


図4 Q2の「講義スライド全体の内容が理解できた」の評価結果

Q2の「講義スライド全体の内容が理解できた」において、講義スライド「第1回コンピュータ基礎」と「第2回認知科学」では、提案手法により生成した学習コンテンツの方がオリジナルスライドより高い評価を得ることができた。「第2回画像処理」の評価が低くなった原因として、全体的に抽出されたスライドの枚数が少ないことにより、内容を理解しにくかったことがあげられる。

学習コンテンツの評価を行った被験者17名のうち、12名が「ややそう思う」もしくは「とてもそう思う」と回答したことから、7割の被験者が少なくとも内容を理解できたと感じていることがわかった。

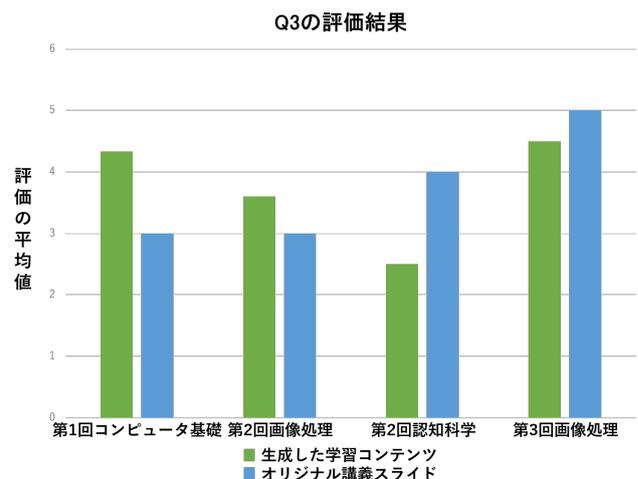


図5 Q3の「講義スライドの内容に関する知識が深まった」の評価結果

Q3の「講義スライドの内容に関する知識が深まった」にお

いて、「第1回コンピュータ基礎」と「第2回画像処理」はオリジナルスライドより高い評価を得たが、そのほかの2つはオリジナルスライドより低い評価となった。学習コンテンツの評価が低くなった原因として、スライドの繋がりがなくなることで内容があやふやになってしまうことや、内容を理解するのに必要な図や表が含まれていないことがあげられる。

学習コンテンツの評価を行った被験者17名のうち、11名が「ややそう思う」もしくは「とてもそう思う」と感じたことから6割以上の被験者が少なくとも知識が深まったと感じていることがわかった。

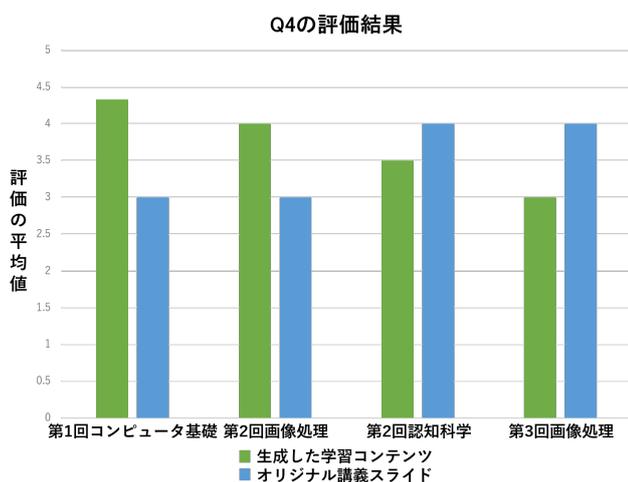


図6 Q4の「講義スライドの内容に対する興味・関心が広がると感じた」の評価結果

Q4の「講義スライドの内容に対する興味・関心が広がると感じた」において、「第1回コンピュータ基礎」と「第2回画像処理」はオリジナルスライドより高い評価を得たが、そのほかの2つはオリジナルスライドより低い評価となった。

学習コンテンツの評価を行った被験者17名のうち、12名が「ややそう思う」もしくは「とてもそう思う」と回答したことから、7割の被験者が少なくとも内容に対する興味・関心が広がると感じていることがわかった。

以上より、Q2とQ4では、7割の被験者が高い評価を付け、Q3では、6割以上の被験者が高い評価を付けたことから全体として有用性は高いと評価できた。しかし、オリジナル講義スライドと比較して相対的に評価が低くなる学習コンテンツが複数存在したため、内容を理解しやすいように改善する必要がある。具体的には、学習コンテンツに適切なスライドの枚数の調整、スライドの繋がりをわかりやすくする、必要な図や表の追加などの改善を行うことでオリジナルスライドよりも評価が高くなり、有用性が高くなると考えられる。

Q5において、改善すべき点は以下の通りである。

- 文字の羅列で見にくく感じた。
- 図がないのでイメージしづらかった。
- スライドの繋がりがよくわからなかった。
- 哲学は内容がなかなか厳しく、スライドを見るだけでは全体理解は難しいと感じた。

5.3 ユーザの知識レベルとアンケート調査結果

事前に調査したユーザの知識レベルは、「第1回コンピュータ基礎」において、学習コンテンツの初級者は1名、中級者は2名、オリジナルスライドは中級者が1名であった。「第2回画像処理」において、学習コンテンツの初級者は3名、中級者は2名、オリジナルスライドは初級者が1名であった。「第2回認知科学」において、学習コンテンツの初級者は3名、中級者は1名、オリジナルスライドは初級者が1名であった。「第3回画像処理」において、学習コンテンツの初級者は4名、中級者は1名、オリジナルスライドは初級者が1名であった。

Q2の「講義スライド全体の内容が理解できた」では、知識レベルが高いユーザが多い学習コンテンツは、知識レベルが低いユーザが多いコンテンツに比べて評価値の平均値が高くなった。理由として、事前知識があると内容を理解しやすいため、評価値が高くなったと考えられる。

Q3の「講義スライドの内容に関する知識が深まった」でも同様に、ユーザの知識レベルが高いユーザが多い学習コンテンツは、オリジナルスライドよりも評価値の平均値が高くなった。理由としては、ユーザの知識レベルが高ければコンテンツの内容が難しくても知識に直結できるため、知識が深まったと感じやすいからだと考えられる。知識レベルが低いユーザが多い学習コンテンツの評価値の平均値がオリジナルスライドよりも低くなった理由としては、ユーザの知識レベルが低ければ、難しい内容や新しい知識をすぐに身につけるのが難しいからだと考えられる。

Q4の「講義スライドの内容に対する興味・関心が広がると感じた」では、知識レベルが高いユーザが多い学習コンテンツほどオリジナルスライドよりも評価値の平均値が高くなった。理由としては、事前知識があればコンテンツの内容を理解しやすく、講義全体の内容を想像しやすいため、評価値が高くなったと考えられる。

5.4 ユーザの知識レベルと学習コンテンツの関係

Q1で調査した被験者の知識レベルとQ2~Q4の結果の関係について考察を行った。今回は初級者と中級者しかいなかったため、上級者の考察はできなかった。自身の知識レベルを「初級者」と回答した被験者は11名であった。初級者が回答したQ2~Q4のアンケート結果および初級者が使用した学習コンテンツのレベルの内訳を示す。

- Q2「とてもそう思う：1名（初級者）、ややそう思う：6名（初級者2名、中級者2名、上級者2名）、どちらでもない：1名（上級者）、あまりそう思わない：3名（初級者1名、中級者2名）、全くそう思わない：0名」
- Q3「とてもそう思う：1名（初級者）、ややそう思う：5名（初級者2名、中級者1名、上級者2名）、どちらでもない：1名（中級者）、あまりそう思わない：4名（初級者1名、中級者2名、上級者1名）、全くそう思わない：0名」
- Q4「とてもそう思う：0名、ややそう思う：6名（初級者2名、中級者1名、上級者3名）、どちらでもない：3名（初級者2名、中級者1名）、あまりそう思わない：1名（中級者）、

全くそう思わない：1名（中級者）」

アンケート結果を基に、「とてもそう思う」を5点、「全くそう思わない」を1点としてスコア化し、評価を行う。Q2の結果では、初級者は15点、中級者は12点、上級者は11点であった。Q3の結果では、初級者は15点、中級者は11点、上級者は10点であった。Q4の結果では、初級者は14点、中級者は10点、上級者は12点となり、初級者レベルのコンテンツの方が中級者、上級者レベルよりも高い評価値となった。よって、知識レベルが初級者のユーザに対しては初級者レベルのコンテンツが有効であることがわかった。

次に、自身のレベルを「中級者」と回答した被験者は6名であった。中級者が回答したQ2~Q4のアンケート結果および中級者が使用した学習コンテンツのレベルの内訳を示す。

- Q2「とてもそう思う：2名（中級者1名、上級者1名）、ややそう思う：3名（初級者2名、上級者1名）、どちらでもない：0名、あまりそう思わない：1名（中級者）、全くそう思わない：0名」

- Q3「とてもそう思う：1名（上級者）、ややそう思う：4名（初級者2名、中級者1名、上級者1名）、どちらでもない：0名、あまりそう思わない：1名（中級者）、全くそう思わない：0名」

- Q4「とてもそう思う：2名（初級者1名、上級者1名）、ややそう思う：4名（初級者1名、中級者2名、上級者1名）、どちらでもない：0名、あまりそう思わない：0名、全くそう思わない：0名」

Q2の結果では、初級者は8点、中級者は7点、上級者は9点であった。Q3の結果では、初級者は8点、中級者は6点、上級者は9点であった。Q4の結果では、初級者は9点、中級者は8点、上級者は9点となり、全体的に上級者レベルのコンテンツの方が、初級者、中級者レベルのコンテンツより高い評価値となった。理解しやすい点と知識を深める点に関しては上級者レベルのコンテンツが優れているが、興味が広がる点に関しては初級者レベルのコンテンツの方が優れている。よって、知識レベルが中級者のユーザに対しては上級者レベルのコンテンツが有効であることがわかった。

6 おわりに

本研究では、講義スライドの特徴語の重要度を算出し、その特徴語のWikipediaのリンク情報を用いて特徴語の難易度判定を行い、各特徴語の重要度と難易度に基づきスライドの重要度を推定した。最後に、スライドの重要度に基づいて生成した学習コンテンツを用いて評価実験を行い、学習コンテンツの有用性を確認した。

今後の課題として、単語の難易度判定手法の改善、スライドを再構成した際にスライドの繋がりがなくなり、理解しづらくなる問題の改善を行う予定である。

- [1] Saki Inoue, Yuanyuan Wang, Yukiko Kawai, and Kazutoshi Sumiya: Question Support Method to Promote Critical Thinking Using Lecture Slide Structure in On-Demand Courses, Proceedings of the 24th International Conference on Asia-Pacific Digital Libraries (ICADL 2022), pp. 509-515, December 2022.
- [2] 桐原牧紀, 王元元, 河合由起子, 角谷和俊: e-Learning における講義コンテンツの階層構造に基づくスライド推薦方式の提案, 第13回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2021), H14-4, 2021年3月.
- [3] Yuanyuan Wang and Yukiko Kawai: A Lecture Slide Reconstruction System based on Expertise Extraction for e-Learning, Proceedings of the 18th International Conference on Asia-Pacific Digital Libraries (ICADL 2016), LNCS 10075, pp. 167-179, December 2016.
- [4] Yuanyuan Wang, Yukiko Kawai, and Kazutoshi Sumiya: Interactive poster generation based on topic structure and slide presentation, Transactions of the Japanese Society for Artificial Intelligence, Vol. 30, No. 1, pp. 112-123, 2015.
- [5] 田中牧郎, 相澤正夫, 斎藤達哉, 棚橋尚子, 近藤明日子, 河内昭浩, 鈴木一史, 平山允子: 言語政策に役立つ, コーパスを用いた語彙表・漢字表等の作成と活用, 国立国語研究所内部報告書, LR-CCG-10-07, pp. 1-254, 2011年2月.
- [6] 水谷勇介, 河原大輔, 黒橋禎夫: 日本語単語の難易度推定の試み, 言語処理学会第24回年次大会発表論文集, pp. 670-673, 2018年3月.
- [7] 大社綾乃, オリビエ オージュロー, 黄瀬浩一: 視点情報と単語の出現頻度を用いた主観的高難易度単語の推定, 電子情報通信学会技術研究報告, Vol. 116, No. 461, PRMU2016-190, pp. 187-192, 2017年2月.
- [8] 西原陽子, 砂山渡, 谷内田正彦: Web ページの難易度と学習順序に基づく情報理解支援システム, 電子情報通信学会論文誌 D, Vol. J89-D, No. 9, pp. 1963-1975, 2006年9月.
- [9] 砂山渡, 谷内田正彦: 観点に基づいて重要文を抽出する展望台システムとそのサーチエンジンへの実装, 人工知能学会論文誌, Vol. 17, No. 1, pp. 14-22, 2002年.
- [10] 山田博文, 松田和彦, 田口亮, 桂田浩一, 小林聡, 新田恒雄: 講義再現システムにおけるスライド重要度抽出, 人工知能学会論文誌, Vol. 17, No. 4, pp. 481-489, 2002年.
- [11] 伊藤雅弘, 中山浩太郎, 原隆浩, 西尾章治郎: Wikipedia のリンク共起性解析によるシソーラス辞書構築, 情報処理学会論文誌データベース (TOD), Vol. 48, No. SIG20(TOD36), pp. 39-49, 2007年12月.
- [12] 稲子翔太, 中野裕司, 杉谷賢一, 久保田真一郎: 自己学習を振り返るための共起グラフによる学習履歴の可視化, 情報処理学会研究報告コンピュータと教育 (CE), Vol. 2020-CE-154, No. 9, pp. 1-8, 2020年3月.
- [13] 田中義規, 竹川高志: Graph Autoencoders による単語情報とリンク情報を融合した文書分散表現, 2022年度人工知能学会全国大会 (第36回), Vol. 36, 4E3-GS-2-01, p. 4E3GS201, 2022年.
- [14] Lucky Agarwal, Kartik Thakral, Gaurav Bhatt, and Ankush Mittal: Authorship Clustering using TF-IDF weighted Word-Embeddings, Proceedings of the 11th Annual Meeting of the Forum for Information Retrieval Evaluation (FIRE '19), pp. 24-29, 2019.
- [15] Thomas N. Kipf and Max Welling: Variational Graph Auto-Encoders, arXiv preprint, arXiv:1611.07308, 2016.
- [16] 小迫良輔, 森山真光: 企業の Web サイトにおけるアクセスログデータを用いた閲覧者の重視するページの解析に関する研究, 情報システム学会第16回全国大会・研究発表大会論文集, Vol. 16, S1-4, p. S1-4, 2020年.
- [17] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd: The PageRank Citation Ranking: Bringing Order to the Web, Stanford Digital Library Technologies Project, 1998.