

シラバスを用いた講義スライドの重要ページ抽出に関する研究

1240372 宮本 健太 【教育情報工学研究室】

1 はじめに

学生は、講義の内容を限られた時間の中で理解するため、効率的に学習を進めたいという傾向にある。そのため、学生自身が理解できていない内容を講義資料全体から効率的に探す必要がある。これに対して、学生が効率よく学習するために講義スライドの自動要約の研究がある [1]。これは、教師が講義スライドのページ毎に学習にかかる時間を入力し、テキスト、画像処理を組み合わせることで要約するというものである。結果として、要約したスライドは閲覧にかかる時間が短くなったにもかかわらず、講義前の小テストではより良い点を取ったと報告されている。しかし、この研究は一つの講義スライドに対して要約するものであり、講義全体のスライドに対しては適用できない。

そこで本研究では、シラバスを利用して全講義スライドのそれぞれのページ単位において、各キーワードに対する重要度を計算し重要ページを抽出することで、教師に負担をかけず学生の学習支援に役立てる。

2 提案手法

前提として教師が講義で使用するスライド資料をPDFに限定し、シラバスはテキスト形式で提供されるものとする。また、PDFで提供されるスライド資料から抽出できるテキスト情報のみを対象とし、PDFに埋め込まれている図は対象外とする。

まず、シラバスのテキストに対して「授業の目的」「授業の概要」「到達目標」「授業計画」「キーワード」の内容を抽出し、形態素解析を行う。解析結果に対して、それぞれの名詞句を対象キーワードとする。また、これらの対象キーワードの出現回数もカウントする。

次に、対象キーワードそれぞれを全講義スライドのPDFに対して検索をかけ、ヒットしたキーワードの個数、文字のサイズと色を抽出する。このサイズや色によって重みをつける。スライドの各ページに対してこれらのサイズと色で対象キーワードの重み付けを行う。対象キーワードの重み付けは以下の計算式で行う。

$$W(k, n, p) = \sum_{0 \leq i < c} SC(k) * SW_i(k, n, p) * CW_i(k, n, p)$$

$W(k, n, p)$: n 回目の講義スライドのページ p におけるキーワード k の重要度

c : ページに出現したキーワードの個数

$SC(k)$: キーワード k がシラバスで出現した回数

$SW(k, n, p)$: 対応するサイズの重み $\{0.5, 1, 2\}$

$CW(k, n, p)$: 対応する色の重み $\{1, 1.5, 2\}$

3 実験

実装にはプログラミング言語として Python を用いた。また、PDF 解析ツールである PyMuPDF、形態素解析ツールの MeCab、形態素解析で抽出する単語の判別には UniDic 辞書を使用した。

対象科目は本学の「人工知能基礎」とした。講義資料数は 10 回分のスライドで全 360 ページである。今回、シラバスより抜き出したキーワードは 97 個あり、その内、不適切なものとして抽象的なキーワードが 48 個、シラバスと講義スライドで表記ゆれが生じているものが 10 個あり除外した。残った 39 個のキーワードに対して結果が妥当か評価した。

評価項目としては、重要度の上位に対象キーワードが説明されている講義スライドが適切に抜き出せているかどうかを「◎」「○」「△」「×」の 4 段階で判定した。「◎」は必要かつ十分なスライドが抽出できた場合、「○」は重要度の上位に必要なスライドは抽出しているが他のふさわしくないスライドが抽出された場合、「△」は重要度の上位に必要なスライドは抽出しているが十分ではない場合、「×」はふさわしくないスライドが抽出された場合を意味する。

表 1 実験結果

	◎	○	△	×
個数	30 個	4 個	5 個	0 個
割合	76.9%	10.3%	12.8%	0%

表 1 に実験結果を示す。結果から、「○」や「△」が出てきた主な原因としては重みづけの対象となるキーワード (例: 推論における「帰納」) が他のキーワード (例: 帰納学習) を説明する講義スライドに多く出現しているため重要度が高くなっていた。これは、重みづけを含めて計算式には改善の余地があると考えられる。

4 まとめ

本研究では、シラバスを利用して講義のキーワードに対するスライドの重要度を計算する手法を提案した。今後は、情報検索の評価尺度である再現率や適合率を考慮して結果の上位から何番目までを採用するかを決め、また他の科目でも適用できるかを検討したい。

参考文献

- [1] A. Shimada, F. Okubo, C. Yin and H. Ogata, "Automatic Summarization of Lecture Slides for Enhanced Student Preview—Technical Report and User Study—," in IEEE Trans. on Learning Tech., vol. 11, no. 2, pp. 165-178, 1 April-June 2018.