

授業用教材スライドにおける「話の流れ」の 解析と可視化について

下園 幸一^{†1} 菅沼 明^{†2}

近年の大学等の授業では、PowerPoint等のスライドを用いて行われることが多くなってきている。このため学生が授業の内容を理解するためには、「スライドの質」が重要となっている。我々は、スライドの質を向上させるために「学生がどのようにスライドを理解するか」という点に着目し、これを教員に提示するシステムの構築を目指している。このシステムを利用することにより、教員は自分の意図する内容との相違点を認識でき、スライドの修正支援に役立てることができると考える。本稿では、システムの機能の1つであるスライドの「話の流れ」の可視化を行う。そのために、まずスライドより学生がキーワードと考える語の抽出を行う。その後、そのキーワードのページ間でつながりを「話の流れ」と考え有向グラフで表し、この有向グラフを見やすい形で教員に提示する方法を述べる。

An analysis and a visualization of the story line in the teaching materials slides

KOICHI SHIMOZONO^{†1} and AKIRA SUGANUMA^{†2}

Recently, many approaches of effective teaching at universities and colleges have been performed. One of them is presentation of the teaching material slides such as PowerPoint by using a computer. Therefore, “the quality of a slide” is important in order for a student to understand the contents of the class. We aim to develop a system which presents “how a student understands a slide” to the teacher for improvement of the quality of a slide. By using our system, the teacher can recognize difference between the contents which he means and the contents which students understand and then he can correct a slide.

In this paper, we propose a visualization of “the story line in the teaching materials slides” which is as one of the functions of our system. Firstly, we describe the extraction method of words which students consider as keywords from a slide. Then, some directed graphs which express the story line of the slide based on the pages linked by extracted keywords are presented.

1. はじめに

近年、日本の大学では、授業を効果的に行うためにさまざまな取り組みがなされている。その1つにICTの利活用がある。その中でもWebCTやmoodle等のLMS(Learning Management System)に注目が集まっている。しかしながら、システムや設備に多大なコストがかかることや、教員へ教材作成の負担がかかるため、学内の一部組織でのみ取り組んでいる場合や、教員が個人的に取り組んでいる場合も多い。また、導入コストと比較して、その効果を疑問視する声も存在する。吉田らの調査¹⁾でも、2006年現在で、LMSを利用している教員は20%程度である。現状では、LMSを導入している場合であっても、授業で利用する教材スライド(PowerPointファイル)を授業開始以前にLMS上で公開するのみにとどまっている場合が多い。

しかしながら、同調査によると、PowerPoint等プレゼンテーションソフトを利用して授業を行う教員は72%にのぼっており、実際の授業では多用されていることがわかる。この理由は、第一に“授業準備における教材の蓄積性、編集の容易性、再利用性等”、第二に“授業場面において映像資料を容易に用いることができること、授業のスピードアップなど”が指摘されている。

このような状況の中で、我々は“授業用教材スライドの質”が重要となっていると考える。教員が授業用教材スライドを作成する際、まず考えることは「学生が授業スライドをどのように理解するか」である。実際に授業を行ってみて、学生の反応を見ることにより、授業スライドに変更を加える場合もある。つまり、“授業用教材スライドの質”とは「どれだけ学生がそのスライドを理解できるか」にかかっている。そのため、授業スライドの質を高めるには、まず、学生がどのように授業用教材スライドの内容を理解するかを知ることが重要である。

我々は、計算機を用いて学生がどのように授業用教材スライドを理解するのかを推測し、それを教員に提示することによって、講義以前に教員のスライド改良支援ができるのではないかと考えた。本稿では、このスライド改良支援環境の全体像を示すとともに、その機能の1つである「話の流れ」の解析と可視化について述べる。

^{†1} 鹿児島大学 Kagoshima University

^{†2} 有明高専 Ariake National College of Technology

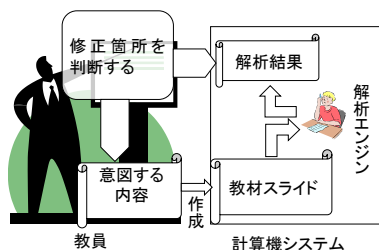


図1 システムの概要

2. 授業用教材スライドの修正支援

先行研究では、学生（受講者）からのフィードバックを利用して、スライドの質の向上および学生の学習支援に役立つものがある²⁾。しかしながら、先行研究の手法では、講義を実際に行わなければフィードバックが得られず、スライドの修正を事前に行うことができない。本研究では受講者からのフィードバックを利用せずに「学生がどのように授業スライドの内容を理解するか」を計算機上で推測し、それを教員に提示することによって、教員のスライド修正支援を行いたい。

システムの概要を図1に示す。支援手順としては以下ようになる。

- (1) 教員はスライドを作成する
 - (2) 解析エンジンが、スライドの内容を解析し、学生がどのようにスライドを理解するかを教員に提示する
 - (3) 提示された内容が教員の意図する内容と異なった場合は、スライドの修正を行う
- この(2)(3)を繰り返すことによってスライドの質を高める。システムは解析エンジンに従って解析結果を提示するのみであり、その結果を見て実際にスライドの修正を行うのは教員である。このようにすることにより、スライド修正のための複雑な処理をシステム側で行わなくてもよいと考えられる。また、(2)(3)を繰り返しながらスライドの改善を行うため、計算機上の解析は、なるべく短時間で終わる方がよい。

3. スライドの「話の流れ」の解析

栗原らがスライドを用いたプレゼンテーションを「スライド提示型プレゼンテーション」と呼び指摘している³⁾ように、この手法は、「発表者の準備通りに進行し、それに合わせて聴衆は理解する」という前提があると思われる。これを授業に特化して言い換えると、教員は「授業の進行は、用意したスライドのページ順通りに進行する」、学生（受講者）は、それに合わせて授業を理解する」という想定を行い、スライドの作成および授業を行うと思われる。

そのため、受講者にとって分かりにくいスライドとは「用意したスライドのページ順通りに授業が進まない」「スライドの順通りだと理解が困難である」ということができる。

亀和田ら⁴⁾はスライド提示型プレゼンテーションにおける認識・理解には2種類あり、1つは1枚のスライドのページに閉じた内容に関する認識・理解、もう1つは複数のページにまたがった関連性に関する認識・理解があると述べ、前者を「ノード理解」、後者を「リンク理解」と呼んでいる。そして、「ノード理解」における発表者と聴衆のズレを認識、修正させることは1枚のページ内に収まっているため容易だが、「リンク理解」におけるズレは、スライド提示型プレゼンテーションでは、ページ単位で情報が断片化され、ページ間の情報を橋渡しすることが難しいため発生しており、ノード理解より確認・修正が難しいと述べ、その確認・修正の支援を行っている。

われわれの研究におけるスライドの「話の流れ」も、上記「リンク理解」の一部と考えることができる。このため、スライドの「話の流れ」を教員に提示することは、スライドの質の向上に役立つと考える。

3.1 解析の方針

学生がスライドの話の流れを認識する際、重要だと思われるものの1つに「各ページに現れるキーワード」があると考えられる。学生は、キーワードを元に「このページは前のページで現れたキーワードに関する説明のページである」や「このキーワードは今回初出なのでこれからのページで説明があるだろう」というような想定をすると考えられる。そのため、学生が認識する話の流れは、各ページに現れるキーワードを手がかりとして、各ページ間を結んでいくことで、スライド全体の話の流れを表現できると考える。

このような想定の上、まず、スライドに現れる語のうち、学生が「キーワードである」と認識する語を計算機上で推定し、その語のページ間でのつながりをもって、学生が認識する話の流れの推定を行う。

3.2 キーワードの抽出

学生が講義の中でキーワードとして認識する語には、専門用語、一般的でない語、頻出語

表 1 2×2 分割表

	b の出現数	b 以外の出現数
a の出現数	O_{11}	O_{12}
a 以外の出現数	O_{21}	O_{22}

表 2 「システム」と A1 との 2×2 分割表

	A1 での出現数	A1 以外のスライドでの出現数	合計
「システム」の出現数	7	81	88
「システム」以外の語の出現数	171	3,983	4,154
合計	178	4,064	4,242

が挙げられる。本稿では、スライドにある情報のみからキーワードの抽出を行いたいため、シソーラス等外部知識は用いない。そのため、頻出語の中からキーワードの特定を行う。

しかし、単に頻出語を抽出しただけでは次のような誤りを犯すと考えた。

誤り 1 全体を通して出現頻度は少ないが、ある特定の回で多く出現しているキーワードを抽出できない

誤り 2 各回に頻繁に出現する一般的な語をキーワードとして抽出してしまう
そのため、統計的手法を用いてキーワードの特定を行った。

大規模コーパスからある単語の共起関係を求める統計的手法として、尤度比検定の 1 つである G 検定 (G スコア : 以下 G 値)⁵⁾ が用いられている。G 値を求める一般的な式は

$$G = 2 \sum_{i=1}^n O_i \ln \left(\frac{O_i}{E_i} \right) \quad (1)$$

で表すことができる。 O_i が観測値、 E_i が期待値である。 n 語のコーパスから表 1 のような 2×2 分割表を用いて単語 a と単語 b の連関性検定を行う場合は以下の計算になる。

$$A = O_{11} \ln O_{11} + O_{12} \ln O_{12} \\ + O_{21} \ln O_{21} + O_{22} \ln O_{22}$$

$$B = (O_{11} + O_{12}) \ln(O_{11} + O_{12}) \\ + (O_{21} + O_{22}) \ln(O_{21} + O_{22}) \\ + (O_{11} + O_{21}) \ln(O_{11} + O_{21}) \\ + (O_{12} + O_{22}) \ln(O_{12} + O_{22})$$

$$C = (O_{11} + O_{12} + O_{21} + O_{22}) \\ \times \ln(O_{11} + O_{12} + O_{21} + O_{22})$$

$$G = 2(A - B + C)$$

O_{11} は単語 a と単語 b が共起している回数、 $O_{11} + O_{12}$ はコーパス内の単語 a の出現回数、 $O_{11} + O_{21}$ はコーパス内の単語 b の出現回数、4 つのセルの合計が n (コーパス内の単語総数) である。G 値が大きければ、単語 a と単語 b の共起関係は強いといえる。

この G 値の計算手法は、語と語の共起関係を数値化するものである。しかしながら、今

回は、各講義で使用するスライドとそのスライドに含まれる語の共起関係から G 値を求めることで、“ある語 a は、ある講義とどの程度関係が深いか”を数値で表すことができるのではないかと考えた。

以下のような極端な例を想定して説明する。

- ある科目は 10 回の講義からなり、それぞれの講義に使用するスライドは 500 語からなる
- 語 c は、各講義でまんべんなく 20 回使用されている
- 語 d は、ある特定の講義 e で 10 回使用されており、他の講義では毎回 1 回のみ使用されている

この場合、語 c の G 値は各講義で全て 0 となり、各講義においてキーワードではないと判断できる。従って、“誤り 2”を犯さない。語 d の G 値は、講義 e において 21.8 となり、他の講義では 0.56 となる。講義 e において大きな値を示しているため、“誤り 1”を犯さずに、講義 e におけるキーワード候補と判断できる。

実際にスライドから語を抽出するには形態素解析を行わなければならない。形態素解析エンジンには京都大学で開発された MeCab⁶⁾ を、解析辞書には IPA で作成された辞書を用いて語の抽出を行っている。なお、MeCab による形態素解析では、名詞が接続している場合に名詞毎に形態素として抽出してしまうため、学生が考えるキーワードとそぐわない場合がある。例えば、「公開鍵暗号」という語の解析結果は、「公開」「鍵」「暗号」の 3 つの語になるが、学生は「公開鍵暗号」を 1 つの語と認識すると考えられる。そのため、形態素解析結果で名詞が接続している場合は 1 つの語としている。

12 回の講義で構成されたある科目 A の第 1 回講義 A1 のスライドでの例を表 2 に挙げる。この表により、A1 のスライドに出現する「システム」という名詞との G 値の計算を行うと 2.537 となる。

この G 値をスライドに出現する全ての語に対して計算し、スライド毎に語の順位付けを行った。ある科目 B(13 回の講義からなる)の第 8 講義 B8 での例を表 3 に示す。この例で、「データ」という語は、出現頻度で見ると「パレート図」「特性要因図」「管理図」より多い

表 3 順位づけされた語の例

順位	G 値	出現数	語	順位	G 値	出現数	語
1	16.63	3	パレート図	8	11.07	2	散布図
2	16.63	3	特性要因図	9	11.07	2	関係
3	16.63	3	管理図	10	11.07	2	管理限界線
4	11.07	2	チェックシート	11	11.07	2	中心線
5	11.07	2	ヒストグラム	12	11.07	2	相関
6	11.07	2	折れ線グラフ	13	10.15	3	仕事
7	11.07	2	近く	14	6.20	4	データ

が、G 値はこれらより小さい。これは「データ」という語が他のスライドでも多用されているからである。このように G 値を用いることでスライドに出現する一般的な語の順位を下げることができる。

実際に学生がキーワードと認識する語と G 値に順位づけられた語の間に相関関係があるか、いくつかのスライドで予備調査を行った。調査結果の分析はまだ不完全ではあるが、少なくとも「スライドタイトルで語が出現しているかどうか」を G 値での順位に考慮したほうがよいことがわかっている。そのため、修正 G 値 = G 値 × (1 + タイトルでの出現数) という修正値を用いて順位付けを行った。

4. スライドの「話の流れ」の可視化

第 3 節で得られた結果を元に可視化を行う。可視化の機能として、以下を考えた。

- (1) スライド中のどのページとどのページに関連があるのか
 - (2) なぜ、そのページ同士が関連があると判断されたのか
 - (3) 他と関連がないページが存在するか
 - (4) 複数のページで 1 つのまとまりを形成しているものがあるかどうか
 - (5) 1 つのスライドにあるページ全ての関連性やまとまりを一望できる
- (1), (2) を提示することで、教員は、自分の考えた「話の流れ」と比較でき、スライド順に学習することで内容の理解が進むかどうかの吟味を行うことができる。(3) を提示することで、そのページの存在意義や他ページとの関連性を再考できる。(4) を提示することで、まとまりの連続性や関連性の吟味ができる。(5) を提示することで、話題のバランス (あるまとまりでスライドが集中していないかどうか) を吟味できる。

4.1 有向グラフの提示方法

提示方法として、まず、抽出したキーワードをそれぞれのノードとし、ページ毎にグルー

プ化を行い、ノードを出現順に有向矢で結ぶ有向グラフで可視化を行う。有向矢で結んだ理由は、スライド提示型プレゼンテーションの場合、話の流れが基本的にページの出現順になっていると考えるからである。

有向グラフ作成には Common Public License の元に配布されているグラフ描画ソフトウェア Graphviz⁷⁾ を用いた。

第 3 節で抽出したキーワード列は、キーワード候補と思われる順に語句に順番がついている。そのためグラフ作成の際、どの程度の順位までをノードとして提示するかを決めなければならない。キーワード候補が少なすぎると関連性のないページが多くなってしまふ。多すぎると、本来関係のないページまで関連性があると提示してしまう。本稿では、キーワード語数はスライドのページ数に依存すると考え、スライドページ数の倍の順位までをノードとした。

前述の講義 B8 で使用したスライド (ページ数 17 ページ) をグラフ化した例を図 2 に示す。それぞれの矩形領域が各ページを示しており、左から右方向へページ順に並んでいる。またそれぞれの矩形領域の上部にページ番号とページタイトルを表示している。矩形領域の中に存在するノードが抽出したキーワードであり、キーワードがタイトルに含まれる場合は、ノードに網掛けを行っている。

このグラフからわかることを以下に列挙する。

- 前半の p.1~p.7 までに強い結びつきがない。
- 後半の p.8~p.17 まででは、p.8 でその後説明する「QC 七つ道具」を列挙し、その後のページで「QC 七つ道具」の説明を行っている。

この情報を元に、教員は、前半部分の結びつきを強めるために、例えば、p.3 に p.4 から p.6 で説明する「デシジョンテーブル」「決定表」「QC 七つ道具」に関する記述を加えるという修正を行う。この修正を行った結果を図 3 に示す。p.3 に書き加えた項目が道しるべになり、その後説明する内容の関連性ははっきりするスライドとなる。また、最初にこれから学ぶ内容のキーワードを挙げることで学生も心構えができるようになる。このような「話の流れ」の修正に役立つと考える。

4.2 ページ配置

スライドのページ枚数が多い場合、ページ順に配置を行うとどうしても全体のグラフが横長になってしまう。また、あるページでこれから説明する項目を提示し、その後のスライドで項目を順に説明するようなスライド構成の場合、提示スライドと説明スライドの距離が遠くなってしまい分かりにくい可能性がある。このような状態は図 4 に示す「提示方式 A」に

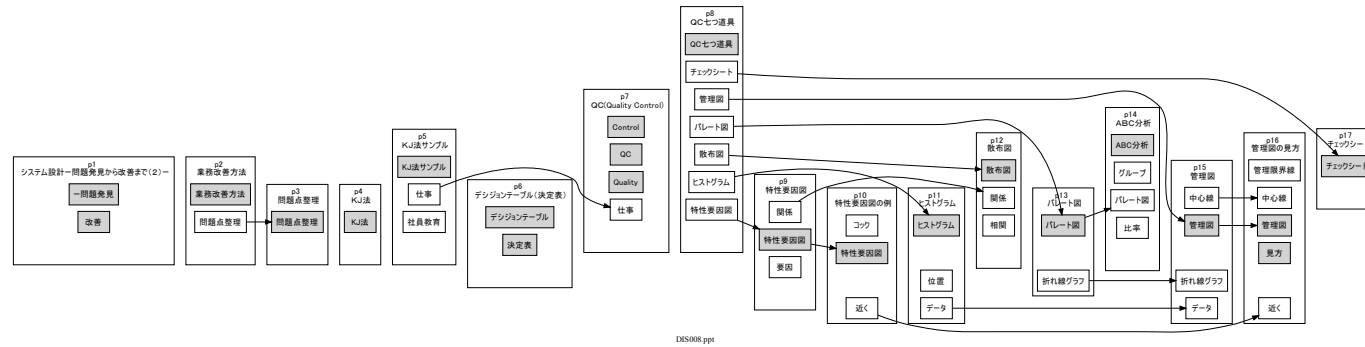


図 2 グラフ化の例

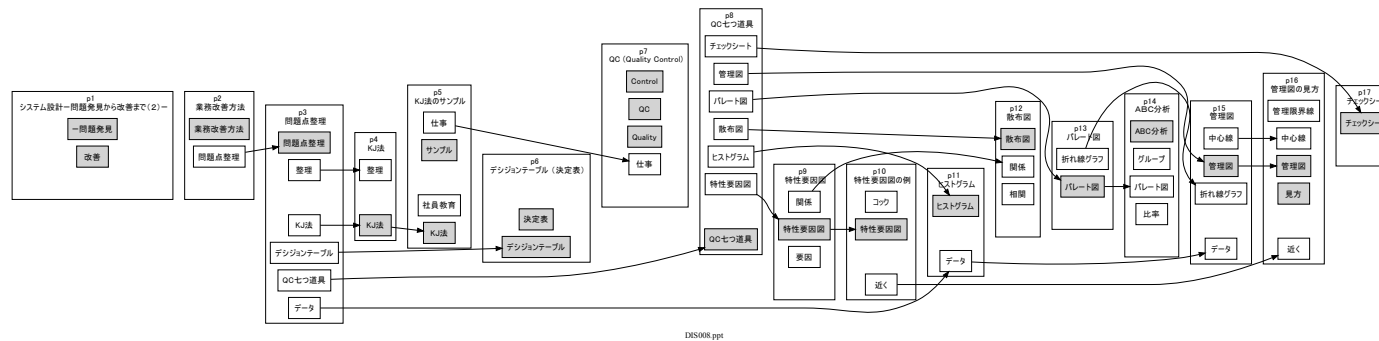


図 3 修正後のグラフ化の例

該当する。提示方法 A では、ページを順に左から並べているため項目 C と項目 C に関連があるページ A+3 との距離が長くなってしまふ。このような場合は、ページ A+1,A+2,A+3 を並列であるとみなし縦に並べる「提示方式 B」の方がわかりやすいと思われる。

前述の講義 B8 で使用したスライドに対して、提示方式 B で作成したグラフの例を図 5 に示す。この例では、p.3 で全体説明を行い、そこで説明した 3 つの項目について、1 つ目は p.4 から p.5、2 つ目は p.6、3 つ目は p.8 から p.17 で説明を行っていることがわかる。しかしながら、スライドの全体の流れが提示方法 A と比較してわかりにくくなっている。そ

のため、提示方法 A と提示方法 B を相互に切り替えながらスライド全体の流れが把握できるようにすることで、可視化の機能として挙げた (4)、(5) を満たすと考える。

4.3 段階的詳細提示

つながりの強いページ群を提示し、それらをつながりの強さに応じて段階的に提示することによって、可視化の機能として挙げた (4) の機能を強化でき、スライドにあるまとまりを把握しやすくなると考えた。つながりの強いページ群は、同じようなキーワードが含まれていると考えられる。そこで、文書の類似度計算に用いられるベクトル空間法を用いてページ

考える「話の流れ」の有向グラフを作成する。この有向グラフと、システムにより提示された有向グラフの近似度により再現性の評価は可能であると考えている。

5.2 有効性の評価手法

第3節で述べたように、受講者にとってわかりにくいスライドとは「用意したスライドの順番通りに授業が進まない」「スライドの順番だと理解が困難である」である。亀和田ら⁴⁾や栗原ら³⁾は実際のプレゼンテーションにおけるスライド提示順と発表者と聴衆が見ているページの差を提示することによりスライドの質の向上を支援している。特に栗原ら³⁾は、前者の指標を Linearity, 後者の指標を Synchrony と定義し数値化している。我々の手法の有効性の評価でも、この2つの指標は利用できる。つまり、本手法を用いた可視化を行う前と後で Linearity と Synchrony を比較することによって、スライドの質が向上したかどうかを計測できると考えている。

6. まとめ

本稿では、実際のスライドから尤度比検定の1つである G 検定の値を用いてキーワードを抽出し、そのキーワードのつながりを有向グラフで表すことによって、スライドの話の流れの可視化手法を提案した。今後、この可視化手法の評価を行い、システムの実装を行う予定である。

参考文献

- 1) 吉田 文, 田口真奈: 大学教員の IT 利用実態調査, NIME 研究報告第 38 号, 独立行政法人 メディア教育開発センター (2008).
- 2) 仁木啓司, 松浦健二, 後藤田中, 金西計英, 矢野米雄: スライド教材のカスタマイズ環境を用いた教員・学習者支援研究: パワーポイントのレポートを用いた支援例, 電子情報通信学会研究会報告 (教育工学), Vol.106, No.583, pp.139-144 (2007).
- 3) 栗原一貴, 望月俊男, 大浦弘樹, 椿本弥生, 西森年寿, 中原 淳, 山内祐平, 長尾 確: スライド提示型プレゼンテーション方法論の拡張手法を定量的に評価する研究, 情報処理学会論文誌, Vol.51, No.2, pp.391-403 (2010-02).
- 4) 亀和田慧太, 西本一志: 聴衆の注意遷移状況を提示することによるプレゼンテーション構築支援の試み, 情報処理学会論文誌, Vol.48, No.12, pp.3859-3872 (2007-12-15).
- 5) Sokal, R.R. and Rohlf, F.J.: *Introduction to Biostatistics*, W.H.Freeman and Company, San Francisco and London (1973). (藤井宏一訳: 生物統計学, 共立出版株式会社 (1983)).
- 6) Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, *Proceedings of the 2004 Conference on*

Empirical Methods in Natural Language Processing (EMNLP-2004), Association for Computational Linguistics, pp.230-237 (2004).

- 7) Bilgin, A., Ellson, J., Gansner, E., Hu, Y. and North, S.: Graphviz - Graph Visualization Software, 入手先 <<http://www.graphviz.org>> (参照 2010-08-01).