社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

隠れマルコフモデルを用いた 論文とプレゼンテーションシートの対応付け

羽山　徹彩[†]　　難波　英嗣[††]　　國藤　　進[†]

† 北陸先端科学技術大学院大学 知識科学研究科
石川県能美郡辰口町旭台 1-1
†† 広島市立大学 情報工学部
広島県広島市安佐南区大塚東 3-4-1
E-mail: †{t-hayama,kuni}@jaist.ac.jp, ††nanba@its.hiroshima-cu.ac.jp

あらまし　本論文では，論文からプレゼンテーションシート自動生成を目指し，隠れマルコフモデル（HMM）を用いた論文とプレゼンテーションシート（シート）との対応付けを試みる．対応付けには，Jing の HMM を用いた対応付け手法をもとに，シートごとに論文の章・節を特定することを行った．しかし，論文とシートを対象とした場合では，新聞記事と比べ一テキスト中の文数と一文当りの長さが大きく，また本文の他に図表が含まれるため，単純に Jing の手法を適用しても十分な対応付けができない．そこで，本論文では，論文とシートの対応位置やレイアウトなどの情報を利用した対応付け手法を提案する．評価実験では，人手で作成した正解データをもとに 77%の対応付け精度が得られた．また，提案した改善手法の有効性も確認した．
キーワード　対応付け，プレゼンテーションシート，論文，隠れマルコフモデル

Alignment between a Technical Paper and Presentation Sheets Using Hidden Markov Model

Tessai HAYAMA[†], Hidetsugu NANBA[††], and Susumu KUNIFUJI[†]

† Knowledge of Science, Japan Advanced Institute of Science and Technology
1-1, Asahidai, Tatsunokuchi-machi, Nomi-gun, Ishikawa-ken
†† Faculty of Information Sciences, Hiroshima City Univercity
3-4-1, Ozuka-higashi,Asaminami-ku, Hiroshima-shi, Hiroshima-ken
E-mail: †{t-hayama,kuni}@jaist.ac.jp, ††nanba@its.hiroshima-cu.ac.jp

Abstract　We have been studying towards automatic generation of presentation sheets from a technical paper. In this paper, we propose an alignment method between sections in a technical paper and presentation sheets. Our method is based on the HMM alignment method by Jing. Though her method is effective to align with short sentences in a newspaper, it is difficult to align with a paper including chart and long sentences. So, we proposed and adapted an alignment approach using feature with a paper and sheets, such as information from text appearance. In experiments, our method showed 77% precision and our strategies were efficiency.
Key words　Alignment, Presentation Sheets, Technical Papers, Hidden Markov Model

1. はじめに

学術会議におけるプレゼンテーションでは，限られた時間で効率的に聴衆へ情報を伝えることが重要である．そのような場では一般的に配布資料とともにそれに基づくプレゼンテーションシートが利用されており，その準備には多大な労力と時間的負担を強いられる．そのため，多くの人はプレゼンテーションシートの効率的な作成を望んでいる．プレゼンテーションシートは，多くの場合において論文の内容に即した内容を持つ．そのため，論文をもとにしてプレゼンテーションシートを作成することが効率的と考えられる．

そのような目的のために，文書からのプレゼンテーション

シート自動生成に関する研究がある．柴田らは，文の結束関係をもとにした表示規則を用いている [4]．安村らは，単語頻度に基づき抽出した重要文に対し自ら作成したテンプレートを利用している [5]．既存研究ではヒューリスティックな規則の適用を行っており，それら規則を人手で作成するには多大な労力が必要である．そのため，我々は，それらのプレゼンテーションシート生成規則の自動獲得を目指している．

近年では，WEB 上に論文データだけでなく，PowerPoint 等のプレゼンテーションデータも公開する人が増えおり，これらデータがプレゼンテーションシート生成のための知識の自動獲得に利用できる．そこで，我々は，論文からのプレゼンテーションシート自動生成を目指して，論文データとプレゼンテーションシート（以後,「シート」と略す）データを収集し，それらの対応関係を分析している．論文からのシート自動生成を行うためには，各シートへの論文分割処理を行い，その結果をもとに並び替え，利用する情報選択，およびシート上への表示変換をする必要があり，それら処理を自動的に行う知識を獲得するためには，論文とシートとの対応付けを行って分析する必要がある．

テキスト自動要約分野においては，より自然な要約生成を目指し人手で作成した要約と原文との対応付けに関する研究がある．加藤らは，ニュース原稿からのニュースの字幕自動生成を目的とし DP マッチングを用いた対応付けを行っている [2]．しかし，DP マッチングは対応付ける語の出現順序の異なりを考慮してないため，論文とシートを対象とした場合では対応付けがほぼ不可能である．そこで本研究では，語の出現順序が入れ替わっても対応できる手法を用いる．

Jing は，隠れマルコフモデル（以後,「HMM」と略す）を用いた対応付け手法を提案している [1]．Jing の手法は，対応付ける語の出現順序の異なりや同じ語が繰り返し出現する場合でもある程度柔軟に対応できる手法であり，新聞記事を対象として良い精度が報告されている．しかし，Jing の手法を論文とシートを対象としてそのまま適用した場合では，新聞記事と比べ一テキスト中の文数と一文当りの長さが大きく，また本文の他に図表が含まれるため，十分な対応付けができない．そのため，新聞記事と異なる論文とシートの特性を考慮する必要がある．新聞記事と比べ論文とシートには，レイアウト，キャプション，インデントなどの情報がある．また，論文では本文が章・節によって分割されており，シートでも予め分割され文脈を持った順序がある．そのため，お互いの分割された位置が対応付けをある程度の範囲の絞込みに利用できる．

本研究では，隠れマルコフモデルを用いて論文とプレゼンテーションシートとの対応付けを行う．論文とシートの対応付けでは，Jing の HMM による対応付けをもとに，シートごとに論文中の章・節を特定する．また，論文とシートの特性を考慮した対応付け精度の改善方法を提案し，適用する．

以後，2 章では Jing の HMM による論文とシートの対応付けとその問題点について述べ，3 章では 2 章で述べた対応付けの問題点を改善する手法について説明する．そして，4 章では対応付け精度の改善手法について評価を行い，5 章でまとめと今後の展開について述べる．

2. HMM による論文とシートの対応付け

2.1 Jing の対応付け手法

Jing の手法は，単語の並びを状態とし最もらしい並びをルールとした変異確率からなる HMM をもとに対応箇所を特定する．

HMM の状態集合の作成には，まず対応基となる文書を単語分割し，各単語に対し'(文番号, 単語番号)' となる固有の ID を割り振る．文番号は単語が含まれる文の文書中の出現順番であり，単語番号は含まれる文中の単語の出現順番である．そのため，同じ単語が複数回出現する場合，その単語は出現回数個の ID を持つこととなる．

次に，対応付け対象となる文書も単語分割し，対応基と同じ単語に対しすべての ID を対応させる．その結果，単語の ID が状態とし出現順番に沿って変異する HMM の状態集合ができる．その具体例を図 1 に示す．

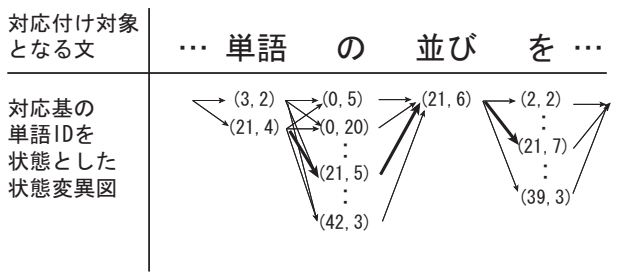

図 1　Jing の手法における状態変異図の具体例

図 1 の状態集合から対応付けを決定するためにはには，最もらしい単語の並びを考慮した状態変異確率を導入する．Jing の手法において用いた状態変異のルールと変異確率を以下に示す．

- If$((S_1 == S_2) and (W_1 = W_2 - 1))$,then 0.9
- If$((S_1 == S_2) and (W_1 < W_2 - 1))$,then 0.8
- If$((S_1 == S2) and (W_1 > W_2))$,then 0.7
- If$(S_2 - CONST < S_1 < S_2)$,then 0.6
- If$(S_2 < S_1 < S_2 + CONST)$,then 0.5
- If$(|S_2 - S_1| > CONST)$,then 0.4

S_1，S_2 は状態変異の前後の単語が含む文，W_1，W_2 は状態変異の前後の単語，$CONST$ は文が影響を及ぼす範囲を表す設定変数である．

そして，Viterbi アルゴリズムにより最短経路を求めることで，単語ごとに対応箇所が特定される．また，隣接する単語が同じ文に対応する場合は，連結し一つの対応語として纏める．

2.2 Jing の手法の問題点

Jing の手法は，対応基に含まれる単語が対応付け対象となる文書中に出現する場合，すべての出現単語に対し対応付けを行う．しかし，論文中の文末表現がシートにおいて削除されたり，「そこで,」などの接続表現がシートでは矢印などの記号や配置のような表現に置き換わったりすることが多い．そのように論文とシートの対応付けにおいては，内容語以外の表現の対応が取れない場合があるため，不要語だけの不自然な対応付け結果を削除する必要がある．

また，シートには，論文と比べ冗長的な内容を挿入したり，内容を省略したりすることがある．わかり易く説明するための具体例など論文に含まれない冗長的な内容がシートに付け加えられることがあり，そのような場合は十分な対応付けが取れない．逆に，論文では詳細に述べているにも関わらず，シートでは簡単な内容として取り除かれることも多い．そのような場合では，対応付ける情報が少ないため重複する別の箇所などへ対応付けられ易くなる．そのため，対応付け範囲を限定したり，冗長的内容の影響を軽減したりするような対応付けの考慮が必要である．

3. 対応付け精度の改善手法

本章では，2 章で述べた Jing の手法の問題点を改善する手法を提案する．また，改善するために用いた論文とシートの特有の情報についても述べる．

3.1 対応付け範囲の限定

対応付け範囲の限定では，HMM による語句の対応付けにおいてシートと論文との相対位置によって論文の対応付け範囲を定める．論文中の章・節とシートの順番が相関関係があると仮定すると，シートの位置によって論文中の対応する章・節の範囲がある程度決まる．対象を絞ることで対応付け精度が向上すると考えられる．

3.2 タイトルによる対応付け

タイトルによる対応付けでは，各シートのタイトルと論文中の章・節のタイトルをもとに対応するシートを特定する．

タイトルによる対応付けでは，以下の手順で行う．

1. 以下の手掛かり語集合をもとにして，同じ手掛かり語集合に属するシートと論文中の章・節を対応付ける．
   - 集合 A …「はじめに」「序論」「背景」
   - 集合 B …「今後の課題」「今後の展望」「今後の発展」
   - 集合 C …「おわりに（終わりに）」「結論」「さいごに（最後に）」「むすび（結び）」
2. 1. において対応付けされなかった場合では以下のルールを適用する．
   - 2 枚目のシートである場合では最初の章に対応付ける
   - 1. の集合 B，C に属すか最後のシートである場合では最後の章に対応付ける

タイトルによる対応付けは，HMM の対応付け結果よりも優先的に適用する．

3.3 タイトルに対する重み付け

タイトルに対する重み付けでは，章・節との対応付け情報の作成においてシートのタイトル箇所に重みを付ける．章・節との対応付けでは，全ての対応語句を一様に考え頻度計算を行っている．しかし，シートのタイトル箇所はシートの内容を最も反映しているため，タイトルに対応する語句に対して他の対応語句よりも頻度の値を大きくする．

3.4 背景情報を用いた対応精度の改善

背景情報を用いた対応結果の改善について図 2 をもとに述べる．

背景情報を用いた対応精度の改善では，各シートに対する論文の章・節の対応付けを行った後，シートの対応付け結果の順序を考慮して不適切な対応結果を変更する．

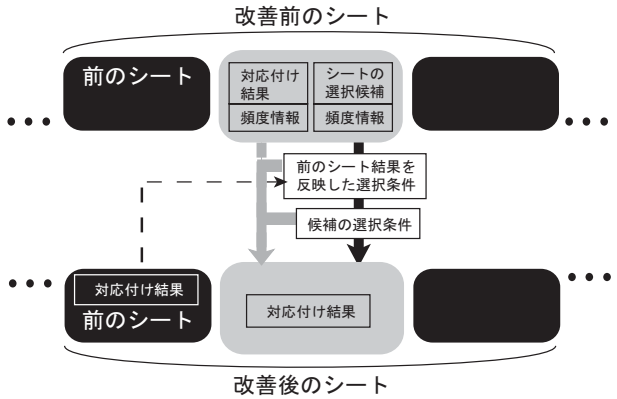

図 2 背景情報を用いた対応結果の改善の概要

対応結果の改善では，シートの選択候補から前のシート結果を反映した選択条件と候補の選択条件を満たす場合，その結果を改善結果として対応付け結果を変更する．また，条件を満たす候補が存在しない場合は，対応付け結果の変更をしない．

背景情報を用いて精度を改善するための設定条件を以下に示す．

前のシート結果を反映した選択条件

前のシート結果を反映した選択条件では，前のシートの対応結果をもとに対応付けする章・節の範囲を設定する．範囲の設定には，正解データをもとに前後のシート間の距離を求め，その値を用いる．シート間の距離とは，論文から章・節の出現する順番を位置として前シートの位置から後シートの位置を差し引いた値である．

シートの選択候補

シートの選択候補では，適切な対応付け結果として可能性のある章・節を選び出す．選択候補としては，対応付け結果と同じ章に含む節を候補とする方法と対応付け結果の高い順位の章・節を候補とする方法が考えられる．

候補の選択条件

候補の選択条件では，選択候補から適切に選択するために候補に制約を設ける．制約の設定には，対応付けに用いた対応頻度をもとに対応順位 1 位との対応頻度差と下限値の 2 つのパラメータを用いる．それらパラメータによって，設定した対応頻度差内かつ対応頻度の下限値を超える章・節を各シートの選択候補とする．

4. 評価実験

4.1 実験の概要

評価実験では，正解データをもとして対応付け精度の改善手法ごとに行う．対応付けに使用するデータは WEB から収集した 90 組の日本語の論文であり，平均シート枚数は 23.6 枚，論文ページ数は 5.6 ページである．正解データは，それらに対し人手でシートごとの対応付けを行った．

正解データにおける対応付けでは，以下の対応付けのタイプ分けを行っている．

type1　論文中と対応が取れるシート
type2　論文中に内容を含まないが内容的に対応個所が特定できるシート
type3　論文中に含まれない内容をもつシート

これらタイプをもとに，各評価実験では，全てのデータ（type1〜type3）に対する対応精度と特定可能であるデータ（type1）に対する対応精度について求める．

HMM の対応付けに用いた論文データとシートデータは，本文に対し章・節のタイトルと文の境界にタグを付与したテキスト情報とシート内のテキストに対し各シートの境界とタイトル箇所にタグを付与したテキスト情報である．

論文データの作成には，PDF 形式と PS 形式のファイルに対し ps2pdf(注1)と pdftohtml(注2)を用いて XML ファイルへ変換し，XML ファイルに含まれるフォント情報と位置情報，および本文中の手掛り語をもとに行う．また，図表のキャプションを説明がある章・節の最後に再挿入する．シートデータの作成には，PPT 形式(注3)のファイルに対し独自に作成したフィルタプログラムによりテキスト情報とその位置情報の抽出した結果をもとに行う．また，手掛り語(注4)を用いて概要を示すシートと実際に発表で使用しない補助シートを特定し，補助シートに対しては対応付けの対象から外すことを行っている．

論文とシートの対応付け手順は，論文データとシートデータをもとに HMM による対応付け手法を適用，不要語による対応付け除去，シートに対応する章・節の特定を行う．図 3 に HMM による対応付け結果の具体例を示す．

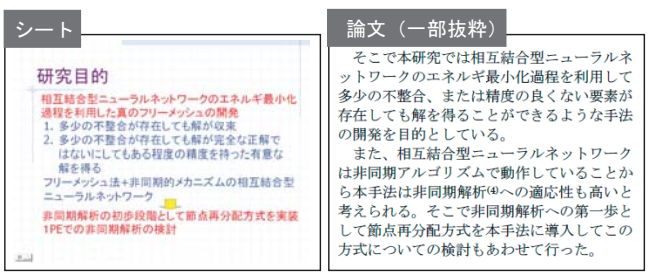

図 3　HMM による対応付け結果の具体例

図 3 のタグに含まれている数字は論文中の対応する文の番号を示す．また，’⟨ SP ⟩’，’⟨SW⟩’，’⟨Non⟩’ は，シートの語句の境界，不要語による対応付け除去を適用した箇所，および論文中に含まれない箇所を示す．

HMM の対応付け結果をもとに章・節の対応付けでは，シートごとに論文の章・節に含まれる対応箇所の頻度を求め，最も多くの対応付け箇所を含む章・節を対応付け結果とする．

4.2　実験方法

評価実験では，３章で述べた対応付け範囲の限定，タイトルによる対応付けとタイトルに対する重み付け，および背景情報を用いた対応結果の改善の順に行う．

対応付け範囲の限定では，まず設定する対応付け範囲について調べ，その結果をもとに評価する．対応付け範囲では，一枚目のタイトルシートを除く正解データをもとにシートと論文との相対位置のずれ割合を調べ，その範囲内で評価を行う．タイトルの重み付けの評価では，タイトル箇所に対する重みを他の箇所の 2 倍とした．背景情報を用いた対応精度の改善では，前のシートを反映した選択条件の設定パラメータを求め，利用するシートの選択候補の検証を行い，それら設定条件をもとに評価を行う．

また，各評価実験では，前の評価実験で有効であった手法を適用して評価する．

4.3　対応付け範囲の限定

4.3.1　設定する対応付け範囲

シートと論文と対応する相対位置のずれ割合の結果を表 1 に示す．

表 1　正解データから求めたシートと論文のずれ割合

ずれ割合	全体に対する割合	ずれ割合	全体に対する割合
0.0 〜	0.55	0.5 〜	0.01
0.1 〜	0.26	0.6 〜	0.00
0.2 〜	0.11	0.7 〜	0.00
0.3 〜	0.05	0.8 〜	0.00
0.4 〜	0.01	0.9 〜	0.01

表 1 の結果では，ずれの割合 0.5 までを考慮することでほぼ全ての対応付けが可能である．そのため，対応付け範囲の限定では，シートの位置においてずれ割合を 0.1〜0.5 の 0.1 刻みおよび 0.3〜0.5 の 0.01 刻みでの範囲で対応付けの精度評価を行う．

4.3.2　評価結果

対応付け範囲の限定による対応付け精度結果を表 2 に示す．

表 2 において，Jing の HMM による対応付けでは対応付け精度 67.5%であったが，対応付け範囲の限定を適用した場合では，ずれ割合 0.3 以上において優位な結果であった．また，最良の評価結果は，ずれ割合 0.36 において対応付け精度 72.2%であった．以上から，対応付け範囲の限定を導入することは有効である．しかし，約 27.8%の誤った対応付けがなされている．その原因の一つとしては，論文に含まれない内容がシートへ挿入されているため，その影響によって十分な対応付けができないことが考えられる．

4.4　タイトルによる対応付けとタイトルに対する重み付け

タイトルによる対応付けとタイトルに対する重み付けを行っ

(注1)：http://www.cs.wisc.edu/~ghost/
(注2)：http://pdftohtml.sourceforge.net/
(注3)：MicroSoft 社の PowerPoint ツールのフォーマット形式
(注4)：概要を示すシートに対しては，「発表の流れ」，「目次」などを用い，補助シートに対しては「まとめ」「今後の課題」などを用いた

表 2　対応付け範囲の限定の適用結果（ずれ割合：0.1〜0.5）

ずれ割合	ALL	IAD	ずれ割合	ALL	IAD
no	0.627	0.675			
0.10	0.473	0.509	0.40	0.664	0.715
0.20	0.602	0.647	0.41	0.664	0.715
0.30	0.656	0.706	0.42	0.659	0.710
0.31	0.663	0.713	0.43	0.657	0.707
0.32	0.667	0.718	0.44	0.658	0.707
0.33	0.666	0.717	0.45	0.656	0.706
0.34	0.670	0.721	0.46	0.657	0.707
0.35	0.670	0.721	0.47	0.657	0.707
0.36	0.671	0.722	0.48	0.659	0.710
0.37	0.667	0.718	0.49	0.657	0.707
0.38	0.665	0.715	0.50	0.658	0.708
0.39	0.665	0.716			

ALL：全ての正解データを用いた場合，
IAD：特定可能である正解データを用いた場合

た結果を表 3 と表 4 に示す．本評価実験では，対応付け範囲の限定において有効であったずれ割合 0.30 から 0.49 までの値で行う．

表 3　タイトルによる対応付けによる精度結果

ずれ割合	ALL	IAD	ずれ割合	ALL	IAD
0.30	0.672	0.723	0.4	0.682	0.734
0.31	0.678	0.729	0.41	0.682	0.734
0.32	0.682	0.734	0.42	0.678	0.730
0.33	0.682	0.734	0.43	0.675	0.726
0.34	0.686	0.738	0.44	0.675	0.727
0.35	0.686	0.738	0.45	0.674	0.725
0.36	0.687	0.739	0.46	0.675	0.726
0.37	0.684	0.736	0.47	0.675	0.726
0.38	0.682	0.733	0.48	0.677	0.729
0.39	0.683	0.734	0.49	0.675	0.727

表 4　タイトルに対する重み付けによる精度結果

ずれ割合	ALL	IAD	ずれ割合	ALL	IAD
0.30	0.698	0.751	0.40	0.704	0.757
0.31	0.700	0.754	0.41	0.704	0.757
0.32	0.703	0.756	0.42	0.698	0.752
0.33	0.705	0.758	0.43	0.698	0.752
0.34	0.709	0.763	0.44	0.697	0.750
0.35	0.709	0.763	0.45	0.696	0.748
0.36	0.710	0.764	0.46	0.695	0.748
0.37	0.706	0.760	0.47	0.695	0.747
0.38	0.706	0.759	0.48	0.695	0.747
0.39	0.706	0.760	0.49	0.693	0.745

表 3 では，表 2 に比べ全体的に優位な結果が得られた．また，表 4 では，表 3 よりも全体的に優位な結果が得られた．そのため，タイトルによる対応付けとタイトルに対する重み付けは，ともに有効な方法であり，ずれ割合 0.36 において最良の対応付け結果が得られた．

このようにシートのタイトルに着目した対応付けは，タイトルが最も重要な内容であるため論文に含まれる可能性が高いことを示している．特に，タイトルに対する重み付けは，論文に含まれない挿入内容の影響を抑えることができたと考える．しかし，23.6%のシートに対して誤った対応付けがなされており，タイトル箇所における誤対応付けの可能性もあることが考えらる．

4.5　背景情報を用いた対応結果の改善

本評価実験では，前節で有効であった対応付け範囲の限定（ずれ割合 0.36），タイトルによる対応付け，およびタイトルに対する重み付けを適用し評価する．

4.5.1　前のシート結果を反映した選択条件

正解データから求めたシート間の章・節の距離を表 5 に示す．

表 5　正解データから求めたシート間の距離

シート間距離	全体に対する割合	シート間距離	全体に対する割合
-4	0.01	1	0.22
-3	0.01	2	0.07
-2	0.01	3	0.04
-1	0.03	4	0.01
0	0.51		

表 5 から，シート間の距離が −1〜3 の間において全体の 86.9 %のシートが占めていることがわかる．そのため，次のシートの対応する章・節を選択条件においては，設定する範囲を −1〜3 とする．また，前のシートの誤対応結果であった場合，連鎖的な誤対応付けが発生することから，二つ前のシートの対応結果も選択条件として扱う．

4.5.2　シートの選択候補

利用する選択候補を検証するために，章だけを対象とした対応付けと対応順位 2 位までの結果と対応順位 3 位までの結果による対応付けの精度を求める．その結果を表 6 に示す．

表 6 では，章だけを対象とした結果よりも対応順位をもとにした結果の方が全体的に優位であった．そのため，対応付けに用いる選択候補としては，対応順位の結果を用いる．

4.5.3　評 価 結 果

背景情報を用いた対応結果の改善による精度結果を表 7，表 8，表 9 に示す．表 7 では選択候補として順位付け 2 位までの対応結果を用いた結果である．表 8，表 9 では選択候補として対応順位 3 位までの対応結果を用いた結果であり，表 9 では，対応順位 3 位の結果を選択され易くするために，パラメータを 2 位よりも緩やかに設定した．

実験結果においては，表 8，表 9 よりも表 7 の方が全体的に優位な結果であった．また，順位付け 3 位の結果がより含まれている表 9 よりも表 8 の方が優位な結果であった．そのため，順位付け 2 位までの対応付け結果を選択候補として扱うことが有効である．また，最も高い対応結果は，対応頻度差 3 と対応頻度の下限値 1 から 4 までのパラメータを用いた結果であり，対応付け精度 77%の結果が得られた．

しかし，対応付け 2 位の結果では 86.7%の対応付け精度が得られるにも関わらず，実際の結果は，約 9.3%ほど低い精度であった．その原因の一つとしては，内容語自体の言い換えにより対応付けが不十分であった可能性がある．

表 7　背景情報を用いた対応結果の改善による精度結果（選択候補：1 位〜2 位）

DIS ＼ DL	0		1		2		3		4		5		6	
	ALL	IAD	ALL	IAD	ALL	IAD	ALL	IAD	ALL	IAD	ALL	IAD	ALL	IAD
1	0.712	0.766	0.714	0.768	0.714	0.768	0.712	0.766	0.711	0.765	0.711	0.765	0.710	0.764
2	0.714	0.768	0.715	0.769	0.715	0.769	0.714	0.769	0.714	0.768	0.713	0.767	0.711	0.765
3	0.714	0.769	0.716	0.770	0.716	0.770	0.716	0.770	0.716	0.770	0.714	0.768	0.712	0.766
4	0.709	0.763	0.711	0.765	0.711	0.765	0.713	0.767	0.713	0.767	0.712	0.766	0.710	0.764
5	0.706	0.760	0.708	0.762	0.709	0.763	0.711	0.765	0.711	0.765	0.709	0.763	0.708	0.762

ALL:すべての正解データを用いた場合，IAD：特定可能である正解データを用いた場合，DIS：順位 1 位の結果との対応頻度差，DL：対応頻度の下限値

表 8　背景情報を用いた対応結果の改善による精度結果（選択候補：1 位〜3 位）

DIS ＼ DL	0		1		2		3		4		5		6	
	ALL	IAD	ALL	IAD	ALL	IAD	ALL	IAD	ALL	IAD	ALL	IAD	ALL	IAD
1	0.711	0.765	0.713	0.767	0.713	0.767	0.712	0.766	0.710	0.764	0.710	0.764	0.709	0.763
2	0.707	0.761	0.710	0.764	0.711	0.765	0.711	0.765	0.711	0.765	0.711	0.765	0.710	0.764
3	0.703	0.757	0.706	0.759	0.706	0.760	0.709	0.763	0.711	0.765	0.711	0.765	0.710	0.764
4	0.694	0.747	0.697	0.749	0.698	0.751	0.703	0.756	0.707	0.761	0.708	0.762	0.707	0.761
5	0.683	0.734	0.685	0.737	0.689	0.741	0.698	0.751	0.703	0.757	0.705	0.758	0.706	0.760

表 9　背景情報を用いた対応結果の改善による精度結果
（選択候補：1 位〜3 位，3 位の下限値：DL の-1，3 位の対応頻度差：DIS の+1）

DIS ＼ DL	0		1		2		3		4		5		6	
	ALL	IAD	ALL	IAD	ALL	IAD	ALL	IAD	ALL	IAD	ALL	IAD	ALL	IAD
1	0.711	0.765	0.713	0.767	0.713	0.767	0.712	0.766	0.710	0.764	0.710	0.764	0.709	0.763
2	0.707	0.761	0.710	0.764	0.711	0.765	0.711	0.765	0.711	0.765	0.711	0.765	0.710	0.764
3	0.703	0.756	0.705	0.759	0.706	0.760	0.708	0.762	0.710	0.764	0.711	0.765	0.710	0.764
4	0.693	0.746	0.696	0.748	0.697	0.750	0.702	0.756	0.706	0.759	0.708	0.762	0.707	0.761
5	0.683	0.734	0.685	0.737	0.688	0.740	0.697	0.750	0.702	0.756	0.704	0.758	0.706	0.760

表 6　章だけを対象とした対応結果と対応順位による対応結果

	対応順位 2 位		対応順位 3 位		章だけを対象とした場合	
ずれ割合	ALL	IAD	ALL	IAD	ALL	IAD
0.30	0.797	0.852	0.827	0.884	0.765	0.823
0.31	0.800	0.855	0.831	0.888	0.766	0.824
0.32	0.803	0.857	0.834	0.891	0.766	0.824
0.33	0.805	0.860	0.839	0.896	0.767	0.825
0.34	0.807	0.862	0.842	0.899	0.770	0.828
0.35	0.809	0.865	0.843	0.900	0.770	0.828
0.36	0.812	0.867	0.844	0.902	0.771	0.829
0.37	0.812	0.868	0.846	0.903	0.768	0.826
0.38	0.812	0.868	0.846	0.904	0.764	0.822
0.39	0.813	0.869	0.846	0.904	0.766	0.824
0.40	0.811	0.866	0.846	0.903	0.765	0.823
0.41	0.812	0.868	0.847	0.905	0.763	0.821
0.42	0.811	0.866	0.845	0.903	0.762	0.819
0.43	0.807	0.862	0.842	0.899	0.761	0.818
0.44	0.806	0.861	0.840	0.898	0.761	0.818
0.45	0.805	0.860	0.839	0.897	0.757	0.814
0.46	0.803	0.858	0.840	0.898	0.755	0.813
0.47	0.803	0.858	0.843	0.900	0.753	0.810
0.48	0.806	0.862	0.844	0.902	0.751	0.808
0.49	0.804	0.859	0.843	0.900	0.749	0.806

5. ま と め

本研究では，論文からのプレゼンテーションシート自動生成を目指して，隠れマルコフモデルを用いた論文とプレゼンテーションシートの対応付けを行った．本対応付け手法では，Jing の HMM による対応付け手法をもとに，対応付け精度の改善手法を適用した．改善手法では，論文とシートの固有の情報を考慮した，対応付け範囲の限定，タイトルによる対応付け，タイトルに対する重み付け，および背景情報による対応結果の改善を適用した．評価実験では，改善手法を適用前において対応付け精度 67.5%に対し，適用後において 77%の結果が得られた．

今後の展開としては，対応付け結果を利用し，論文とシートとのより詳細な分析を行っていきたい．具体的には，シートに利用する章・節の分析，文書中の引用箇所の分析，引用箇所の位置的表現の分析を行う．

文　献

[1] Hongyan Jing, "Using hidden markov modeling to decompose human-written summaries", Computational Linguistics, 28(4):pp.527 - 544, 2002

[2] 加藤直人，浦谷則好，"局所的要約知識の自動獲得手法 "，自然言語処理，Vol. 6，No. 7，pp.73-92，1999

[3] 松本祐治 他，"形態素解析「茶筌」version2.3.1 仕様報告書"，http://chasen.aist-nara.ac.jp/chasen/，2003

[4] 柴田知秀，河原大輔，黒橋禎夫，"主題と文章構造の解析に基づくスライドの自動生成 "，言語処理学会 第 9 回年次大会，pp.597-600，2003

[5] 安村禎明，武市雅司，新田克己，"論文からのプレゼンテーション資料の作成支援 "，人工知能学会論文誌，Vol.18, No.4, pp.212-220，2003