

平成 2018 年度 情報数理工学コース卒業論文概要

平成 2014 年度入学	学籍番号 1411029
指導教員 植野 真臣	氏名 遠藤 立
題目	条件付き確率場を用いた論文構成自動推定による 論文評価システム

概要

アカデミックライティングでは、文章を適切に構造化することで読者に対して効果的に情報を伝達できることが示されている。文章構造の中でも特に論文要素の系列を指す論文構成が重要であることが知られている。今まで論文構成の構築を支援する様々なシステムが提案されてきた。しかし、これらのシステムはどれもユーザー自身が論文構成を入力する必要がある。この問題を解決するためには論文構成の自動推定を行う必要がある。論文構成の自動推定の様々な手法が提案されており、条件付き確率場 (CRF) を用いて高精度に論文構成が推定できることが示されている。そこで、本研究では条件付き確率場を用いて論文構成の自動推定を行い、論文構成のルールに違反しているものがあればルール違反している内容があればルール違反の内容及び違反箇所を指摘するシステムを提案する。論文構成の推定を推定するためには特徴量を定める必要がある。本研究では、先行研究で用いられている特徴量を改善して用い、改善前後の特徴量で推定精度の比較実験を行った。結果、改善後の特徴量の方が推定精度が良いことを示せた。作成した特徴量を用いて、被験者実験を行ったところ本システムを用いることで「学術論文として妥当な形式になる」、「論理的な文章になる」といった利点が示せた。

条件付き確率場を用いた論文構成自動推定による論文評価シ
テム

平成30年1月31日

情報数理工学コース

学籍番号 1411029

遠藤 立

指導教員 植野真臣

1 はじめに

アカデミックライティングでは、文章を適切に構造化することで読者に対して効果的に情報を伝達することが重要である [1]~[3]。アカデミックライティングにおける重要な文章構造として論文構成が知られている [4]。論文構成とは「研究背景」、「関連研究」、「本研究の目的」といった論文の構成要素の系列を指す。

そのため、これまでに論文構成の構築を支援する多数のシステムが開発されてきた。例えば、O'Rourke and Calvo[5] はパラグラフ間の二元性を可視化するシステムを開発している。Fujimoto and Matsuo[6] は、論文を段落単位に自動分割し、その配列の変更や文章の改定を他の利用者との議論を通じて行うシステムを開発している。更に、Kunichika et al.[7] は、利用者による文章の自由なグループ化と、そのグループの階層的な構造化を支援するシステムを開発している。また、西村ら [8]、甲斐ら [9] は、文章の表層表現に基づいて論文構成を解析し、表示するシステムを開発している。これらのシステムは論文構成の画一的な形式化のみを支援しており、多様な利用者の意図を表現することが難しい。これらの問題に対して、Aluisio and Oliveira[4] や岩田 [10] らは形式的なアプローチを用いてこの問題の解決に取り組んでいる。しかし、形式的なアプローチを用いた場合、無数のテンプレートの作成が必要となるため、このアプローチは現実的ではないと考えられる。

この問題を解決するために、宇都・植野 [11] は初心者であっても容易に論文構成を構築できるように確率的なアプローチを用いた論文構成構築支援システムを開発した。このシステムは論文構成が m 重マルコフ情報源に従うと仮定し、過去の優良論文 100 件から学習した m 重マルコフ情報源を用いて次に書くべき内容を利用者に推薦するシステムである。このシステムを利用することで、1) 初心者であっても容易に論文構成を構築することができること、2) 学術論文として妥当な形式を保持しつつ、多様な利用者の意図にあう論文構成を構築できること、が示されている。しかし、このシステムはユーザー自身が論文構成の入力を行う必要がある。

一方で、近年では、論文構成の自動推定技術が提案されている。Guo, Korhonen and Poibeau や, Ronzano and Saggion[12]~[14] はサポートベクターマシン (Support Vector Machine; SVM), 条件付き確率場 (Conditional Random Field; CRF), ナイーブベイズ (Naive Bayes; NB) を用いた英語論文の論文構成自動推定精度の評価実験を行い、それぞれ SVM で 81%, CRF で 71%, NB で 53%の精度で正解の

論文構成要素を各文に割り当てることができることを示している。これらの分類に使う特徴量としては、コーパス内に出現した単語全てのユニグラムやバイグラム、時制や態などを用いている。しかし、これらの特徴量のいくつかは英語に特化しており、英語以外の論文の推定を行うには不適切な特徴量が含まれている。また、特徴量には改善の余地があり、改善により推定精度の向上が期待できる。

そこで、本研究では日本語論文の論文構成の自動推定を行い、論文構成のルールに違反しているものがあればルール違反の内容及び違反箇所を指摘するシステムを提案する。このために本研究では、論文構成推定手法と論文構成のルールを開発する。論文構成推定手法としては一般に個々の文の役割ラベルがその前後の文の役割ラベルに依存することを利用して、前後の役割ラベルを考慮した分類が可能なCRFを用いる。また、特徴量としては、Guo, Korhonen and Poibeau[12]の用いたものから日本語論文に適用できるものを抽出したうえで、分類精度の改善に有効と考えられる特徴量を追加したものを用いた。「論文構成のルール」は電子情報通信学会論文誌Dに掲載された「情報システム」関連の論文50件を分析し得られた結果を元に作成した。

また、本研究では、論文要素の推定精度比較実験及び被験者実験により本システムの有用性を示す。論文要素の推定精度比較実験では、本研究で提案した特徴量と先行研究で提案された特徴量の2パターンの特徴量を用いてCRFとSVMで推定精度の比較実験を行う。被験者実験では、システム利用群とシステム未利用群に分けて比較実験を行い、本システムを用いた場合の方がより論文構成が適切になることを有意差検定を用いて示す。

2 論文構成支援システム

宇都・植野[11]は論文構成支援システムを開発している。このシステムは、論文を書く際に使用者が次に書くべき論文構成の要素を逐一推薦するシステムである。各論文構成要素からなる論文構成は論文を書く上で重要であることが示されており、このシステムを利用することで、初心者であっても自分の意図に沿った論文構成を容易に構築できることが示されている。このシステムでは論文構成が m 重マルコフ情報源に従うと仮定しており、過去の優良論文100件に基づいた m 重マルコフ情報源を推定し、推薦に用いる。

2.1 論文構成要素

宇都・植野 [11] は「情報システム」関連の論文を対象として論文構成要素のカテゴリー化を行った。論文構成要素は、一文に一つ与えられる。論文構成は論文構成要素の系列で定義され、一文書に一つ与えられる。Table 1 に宇都・植野 [11] が定義した「情報システム」関連の論文の序論における論文構成要素を示す。このシステムは、論文の執筆過程で次に書くべき要素を Table 1 で示した論文構成要素の中から m 重マルコフ情報源に従って推薦する。

Table 1: 論文構成支援システムにおける論文構成要素

ラベル	内容
A	研究領域の目標
B	研究領域の現状
C	研究領域の問題点
D	先行研究の概要
E	先行研究の利点
F	先行研究の問題点
G	本研究の目的
H	本研究の方法
I	本研究の特徴
J	本研究の利点
K	本研究の評価
L	用語など定義
M	本論の構成

2.2 m 重マルコフ情報源と多重度の推定

m 重マルコフ情報源とは、任意の時点 t における状態 X_t の生起確率が、その直前 m 個の状態系列 $X_{t-m}^{t-1} = X_{t-m}, \dots, X_{t-1}$ のみに依存する m 重マルコフ過程である。すなわち、式 (1)

$$p(X_t | X_1^{t-1}) = p(X_t | X_{t-m}^{t-1}) \quad (1)$$

を満たす情報源である。式 (1) は、遷移確率と呼ばれ、論文構成支援システムは、この遷移確率が大きい順に論文構成 X_t を推薦する。

m 重マルコフ情報源は多重度 m に依存するため、多重度を正確に推定する必要がある。一般的な多重度の推定方法は L を系列データ X_1^n に対する尤度とした時、次式を最小化する問題として与えられる。

$$\hat{m} = \arg \min_m \{-\log L + Kpen(n)\} \quad (2)$$

ただし、 $pen(n)$ はデータ数の関数、 K は自由パラメータであり $Kpen(n)$ はモデルの複雑さに対するペナルティ項である。この推定方はデータの長さが十分でない場合、正しく推定することができない。そこで、宇都・植野 (2010) はベイズ符号を用いた多重度の推定方法を提案している。宇都・植野 (2010) によるとデータ長が短く、多重度に応じてベイズ符号長 $l_{bayes}^m(X_1^n)$ が単調減少する場合、ベイズ符号長の減少量が真の多重度を境に大きく減少することが示されている。これを利用して真の多重度を求める。すなわち、

$$R(m, X_1^n) = \frac{l_{bayes}^{m+1}(X_1^n) - l_{bayes}^m(X_1^n)}{l_{bayes}^m(X_1^n) - l_{bayes}^{m-1}(X_1^n)} \quad (3)$$

としたとき、 $R(m, X_1^n)$ を最小にする m を求めることで多重度の推定を行うことができる (式 (4))。

$$\hat{m} = \arg \min_m \{R(m, X_1^n)\} \quad (4)$$

宇都・植野 [11] はこのように推定された m 重マルコフ情報源を用いて、次に書くべき要素を逐次推薦することで (1) 初心者でも容易に論文構成を構築でき、(2) 学術論文として妥当な形式を保持しつつ多様な利用者の意図に合う論文構成を構築できことを示した。しかし、このシステムはユーザー自身が論文構成の入力を行う必要がある。この問題を解決するためには論文構成の自動推定を行う必要がある。

3 論文構成の自動推定技術

近年、英語論文の構成を自動推定できる手法が提案されている。Guo, Korhonen and Poibeau [12] は、サポートベクターマシン (Support Vector Machine; SVM)、条件付き確率場 (Conditional Random Field; CRF)、ナイーブベイズ (Naive Bayes; NB) を用いた評価実験を行い、それぞれ SVM で 81%、CRF で 71%、NB で 53% で正解のラベルを割り当てることができると示している。また、Saggion らはロジスティック回帰モデルと SVM を用いて評価実験を行い、同様に高精度で

Table 2: Guo,Korhonen and Poibeau[12] が提案した特徴量

特徴量	詳細
名詞ユニグラム	コーパス内で出現した名詞全て
動詞ユニグラム	コーパス内に出現した動詞全て
バイグラム	コーパス内に出現した単語全て
動詞クラス	動詞ごとのクラス番号
時制	文の時制
主語	
直接目的語	
関節目的語	
第二目的語	
態	文の態
文章中の位置	文章全体を 10 分割したときの文の位置
単語数	文に含まれている単語数
前の文章の予測ラベル	

の推定ができることを示している。Guo,Korhonen and Poibeau[12] は特徴量として Table 2 で示したものを採用している。

Table 2 で示した特徴量を本研究でそのまま用いるのは適切でない。理由としては、第一に先行研究では英語の論文を対象にしているが本研究で対象としているのは日本語の論文である点である。直接目的語や関節目的語は日本語には存在しないためこれらを直接特徴量化するのは不可能である。また、受け身や時制は日本語にも存在するものの英語ほど正確に処理するのは困難である。第二にこれらの特徴量は改良の余地があると考えられる。以上より、本研究で使用する特徴量は Table 2 の特徴量をベースにアレンジしたものを採用する。

4 提案手法

本研究では、日本語の論文の序論に対して「論文構成のルール」を設け、このルールを違反していた場合に、ルール違反の内容及び違反箇所を指摘するシステムを提案する。論文構成のルールは電子情報通信学会論文誌 D に掲載された「情報システム」関連の論文 50 件を分析し得られた結果を元に作成した。論文構成

の推定を行うための訓練データとして、本研究で示す論文構成要素を役割ラベルとして各文にラベリングを行ったコーパスを作成した。本システムはこのコーパスを利用して CRF 及び本研究で作成した特徴量による論文構成要素の推定を行い、推定により得られた構成が「論文構成のルール」を違反していた場合に指摘を行う。

4.1 本研究で用いる論文構成のルール

本研究では、論文構成のルールを作成するにあたって、電子情報通信学会論文誌 D に掲載された「情報システム」関連の論文 50 件の分析を行ったところ、「不適切な論文構成のルール」として以下が抽出された。

1. 「研究領域の社会的意義や課題について述べている文」, 「本研究に関連する先行研究のレビューを述べている文」, 「本研究の目的を明確に述べている文」が存在しない。
2. 「本研究の手法, 評価, 利点」のいずれかが存在しない。
3. 「2 章以降で述べる内容及び章番号を述べる文」が章末以外に存在する。
4. 「本研究の目的を明確に述べている文」は分散して複数存在することがない。

この結果を基に本研究で用いる論文構成のルールを作成した。(Table 3)

Table 3: 本研究で用いる論文構成のルール

ルール	内容
1	研究領域の背景（研究領域の社会的意義や課題）に関する記述が存在する
2	関連研究（本研究に関連する先行研究のレビュー）が存在する
3	本研究の目的が存在する
4	本研究の手法・評価・利点のいずれかが存在する
5	本論の構成（2 章以降で述べる内容及び章番号を述べる文）が最後にのみ存在する
6	本研究の目的が分散して複数存在しない

4.2 使用したラベル

本研究では論文構成のルール違反を調べるのに必要最低限なラベルとして Table 4 のように定めた。

Table 4: 本研究で使用するラベル

ラベル	内容	詳細
BKG	研究領域の背景	研究領域の社会的意義や課題について述べている文
REL	関連研究	本研究に関連する先行研究のレビューを述べている
OBJ	本研究の目的	本研究の目的を明確に述べている
CON	本研究の手法・評価・利点	本研究の手法, 評価, 利点のいずれかについて述べている
STR	本論の構成	2章以降で述べる内容及び章番号を述べている

Table 4 のラベルを使用して, 訓練データとして使用するコーパスを作成する。

4.3 コーパス作成

本研究の目的に利用できる公開コーパスは存在しないため, 本研究では独自にコーパスを作成した。論文構成要素は論文の分野によって大きく異なるため, 異なる分野の論文を混ぜることは望ましくない。本研究では電子情報通信学会論文誌 D「情報システム」関連の日本語の優良論文 50 件に対してラベリングを行った。作成したコーパスの例として論文 [15] へのラベリングの一部を Table 5 に示す。本コーパスは Table 5 のように一つの文に対して一つのラベルを付与したデータとして定義した。本システムはこのコーパスを訓練データとして, CRF 及び本研究で作成した特徴量を用いて論文構成要素の推定を行う。

4.4 ラベル推定手法と特徴量

一般に, 個々の文の役割ラベルは前後の文のラベルに強く依存すると考えられる。そのため本研究では, 前後の順序情報を考慮した上で分類することができる CRF を用いてラベルの推定を行う。また, 本研究で使用する特徴量は Guo, Korhonen and Poibeau[12] の用いていた特徴量をベースに本研究で使用するのに適した特

Table 5: 本研究で作成したコーパス中の論文 [15] のラベリング例

文番号	ラベル	文
1	BKG	機械翻訳においても、また、翻訳者による翻訳においても、高い訳質を保証するためには大規模で正確な対訳辞書が必要不可欠である。
2	BKG	しかし、人手によって対訳辞書を作成し、継続的に収録語数を増やし辞書を維持・管理していく作業は膨大な時間と労力を要する。
3	REL	そこで、自然言語処理分野においては、多様なテキストデータを情報源として、対訳辞書を自動若しくは半自動的に作成する技術に関する研究が行われてきた。
4	REL	例えば、複合語専門用語の構成要素の訳語を連結して訳語の候補を生成し（要素合成法）、ウェブから収集した目的言語の専門分野コーパスを用いて、生成された訳語候補を検証する手法 [3]、検索エンジンから訳語が併記された文書を収集し、訳語対を獲得する手法 [4] 等がある。
(中略)		
9	OBJ	以上の背景のもとで、本論文では、句に基づく統計的機械翻訳モデル [5] により学習されるフレーズテーブル、及び、既存の対訳辞書を用いる要素合成法 [3] という、性質の異なる 2 種類の知識源（大規模対訳文対から統計的に学習された訳語知識、及び、既存の対訳辞書）を用いる手法を併用して、対訳文から専門用語対訳辞書を獲得する手法を提案する。
10	CON	本論文の手法では、まず、専門用語対訳辞書獲得の情報源として用いる日英対訳文対に対して、句に基づく統計的機械翻訳モデルを適用することにより、フレーズテーブルを学習する。
(中略)		
30	STR	以下に本論文の構成を述べる。
31	STR	2. で、本論文で用いた日英対訳特許文について説明し、 3. で、本論文で用いた訳語推定手法について述べる。
(中略)		

微量にアレンジしたものをを用いる。また、本研究でのラベル推定の精度の結果は5章で述べる。

4.4.1 CRF

CRFは多値分類問題を解くのに使われる手法であり、主に系列ラベリング問題を解くときに用いられる。系列ラベリング問題とは、系列 $\mathbf{x} = x_1, \dots, x_M$ に対してラベル列 $\mathbf{y} = y_1, \dots, y_M$ を割りあてる問題として定式化される。例えば、「This is a apple.」という英文に対して「This(代名詞) is(動詞) a(冠詞) apple(名詞).」というように品詞を割りあてる問題である。本研究で行われている各文に対して表6のようなラベルを割りあてる問題も系列ラベリング問題として考えることができる。

CRFは入力系列 \mathbf{x} に対するラベル列 \mathbf{y} の予測スコア $s(\mathbf{x}, \mathbf{y})$ を素性ベクトル $\mathbf{f}(x, y)$ と重みベクトル $\boldsymbol{\lambda}$ との内積で計算する。

$$s(\mathbf{x}, \mathbf{y}) = \boldsymbol{\lambda}^T \mathbf{f}(x, y) = \sum_{i=1}^k \lambda_i f_i(x, y) \quad (5)$$

ここで、素性ベクトルとは素性関数 $f_i(x, y)$ を要素とするベクトルであり、素性関数は系列の特徴を反映したものである必要がある。例えば、品詞割り当ての系列ラベリング問題を考えた時、 $f_1(x, y)$ を直前の x に対する y が前置詞であるか否かという素性関数として定義するとする。この時、 $f_1(x, y)$ は以下のように表現される。

$$f_1(x, y) = \begin{cases} 1 & (\text{前のラベルが前置詞}) \\ 0 & (\text{それ以外}) \end{cases}$$

一般に前置詞は名詞の直前にのみ使用されるという特徴があるため、素性関数 $f_1(x_n, y_n)$ が1をとった(発火した)場合、 x_n のラベル y_n が名詞である可能性が高いと言える。このように、素性関数が系列の特徴を捉えたものとなるような素性空間を構築する必要がある。

重みベクトルは対応する素性関数の有用性を示すものであり、有用な素性関数に対応する重みベクトルであれば値は大きくなるため発火した時にスコアに与える影響は大きくなり、逆に有用でない素性関数であるならば重みベクトルの値は0に近づく。重みベクトルの値は訓練データから求める。

スコアからラベル列を求めるためには、スコア $s(\mathbf{x}, \mathbf{y})$ を最大とする \mathbf{y} を求め

ればよい。従ってラベル列 \hat{y} は次式で求められる。

$$\hat{y} = \arg \max_{y \in Y} s(\mathbf{x}, \mathbf{y}) \quad (6)$$

ただし、 Y はラベルの集合である。

次に求めたスコアを確率に変換する。入力系列 \mathbf{x} に対してラベル列 \mathbf{y} が予測される条件付き確率をソフトマックス関数

$$y_i = \frac{\exp(x_i)}{S(\mathbf{x})} \quad (7)$$

ただし、

$$S(\mathbf{x}) = \sum_{i=1}^k \exp(x_i) \quad (8)$$

に適用する。ソフトマックス関数は $\sum_{i=1}^k y_i = 1$ を満たしているため、確率分布の表現に用いられる。式 (7) より求める条件付き確率 $p(y|x)$ は

$$p(\mathbf{y}|\mathbf{x}) = \frac{\exp s(\mathbf{x}, \mathbf{y})}{Z(\mathbf{x})} \quad (9)$$

ただし、

$$Z(\mathbf{x}) = \sum_{\mathbf{y}' \in Y} \exp s(\mathbf{x}, \mathbf{y}') \quad (10)$$

により求めることができる。

前述した通り、CRF では訓練データ $D = (x_i, y_i)_{i=1}^N$ (本研究の場合は表 4 で示したようなコーパスデータ) から重みベクトル λ を求める必要がある。これを学習と呼ぶ。CRF では学習の手法として最尤法を用いる。最尤法は全ての学習事例に対して条件付き確率 $p(y_i|x_i)$ を計算し、それぞれの積をとる。

$$l = \prod_{i=1}^N p(y_i|x_i) \quad (11)$$

式 (7) の l は学習データ D の尤もらしさを表しており、尤度と呼ばれる。最尤法はこの尤度が最大になるような重み λ を求めることで学習を行う手法である。式 (7) は非常に小さい値の積を計算する必要がある。そのためアンダーフローなどの問題を引き起こしてしまう。そのため、一般的には式 (11) の対数をとった対数尤度 L を最大化する問題に置き換えて考えられる。

$$L = \log l = \sum_{i=1}^N \log p(y_i|x_i) \quad (12)$$

L を最大化するためには主に確率勾配降下法を用いる。

Table 6: 本研究で使用する特徴量

特徴量	詳細
文番号	文の先頭から数えて何文目か
段落番号	何段落目に含まれるか
箇条書きか否か	—
箇条書きの先頭か否か	—
段落の開始	—
段落終了	—
文の相対位置	文番号/文の総数
段落の相対位置	段落内の文番号/段落内の文の総数
箇条書きの相対位置	箇条書き内の文番号/箇条書き内の文の総数
被引用参考文献数	—
単語数	文に含まれている文の総数
単語ユニグラム	各ラベルの文内に含まれている単語の頻出
単語バイグラム	全文の頻出

4.4.2 特徴量

前項でも述べたように適した素性関数を用いるためには適切な素性空間を構築する必要がある。素性空間を作成する上で文章から抽出すべき文章の特徴を特徴量と呼ぶ。本研究では論文構成に適した特徴量の開発にも焦点を当てている。先行研究で述べられている特徴量を Table 2 にまとめた。

2 節でも述べたように本研究にこれらの特徴量を用いるのは不適切である。第一に Guo, Korhonen and Poibeau [12] が示した特徴量は英語に特化しており、日本語論文を対象とする本研究にはそのまま利用できない。第二にこれらの特徴量はコーパス内に出現した名詞・動詞ユニグラム、バイグラム全てを特徴量としているが絞ることができること、文の位置に関する特徴量の細分化、文件数の情報の追加をすることで推定精度の向上が見込めると考えられる。以上の考察を基に本研究では特徴量として Table 6 を作成した。

文番号から箇条書きに対する相対位置までは主に先行研究の文章中の位置を再現したものである。先行研究では文章全体を 10 分割した時に対象の文がどの位置にくるのかを特徴量としていたが、本研究ではより細かく求めた。研究背景は文頭にくる場合が多く、本論の構成は末尾にくるといった位置から求められる

ラベルの情報は非常に大きい。特に相対位置は明確に今見ている文が文章全体の先頭の方なのか末尾の方なのかを調べることができるため、ラベル推定に大きく貢献できると期待できる。引用されている参考文献の数は本研究で新たに提案されたものである。この特徴量は主に REL のラベルであるか否かを推定するのに大きく貢献してくれると期待できる。単語数は先行研究と同様に用いた。特に本論の目的について述べる文は簡素で短い文で書かれることが多い。そのため単語数は OBJ のラベルであるか否かを推定するのに大きく貢献してくれると期待できる。単語ユニグラム、バイグラムに関しては先行研究から大きく数を絞ったものを用いた。前述した通り、先行研究で用いている特徴量としての単語はあまりに多く、推定に時間がかかる上にノイズも大きく含んでしまっている。本研究では使用する単語を最低限に絞ることでノイズを減らし、推定速度及び精度の向上を試みた。使用した単語ユニグラムは各ラベルごとに出現した単語をまとめ、ノイズを取り除いた後に、出現回数が 30 未満のもの 30 件を特徴量とした。出現回数に上限を設けた理由としては、出現回数が多すぎる単語はほぼ全てのラベルに出現するような文章単位での頻出語の可能性が高いためである。従ってどのラベルにも多く出現する単語はラベル推定においては不要な語であるため排除した。またラベルごとに頻出の単語を特徴量として用いれば、ラベルの推定に貢献できると期待できる。バイグラムに関してはノイズを取り除いた後の上位 100 を用いた。

4.5 論文構成のルールの指摘

本研究では、前述のラベル推定手法に基づいて Table 3 で示した「論文構成のルール」を違反していた場合に違反箇所及びその内容の指摘を行う。例えば「REL,REL,OBJ,REL,REL,OBJ,STR」と推定された論文に対しては Table 7 のような指摘を行う。

5 論文要素の推定精度比較実験

本研究では、本システムの有用性を示すために論文構成要素推定の比較実験を行った。具体的には、本研究で提案した特徴量と Guo,Korhonen and Poibeau[12]によって提案された特徴量の 2 パターンの特徴量を用いて CRF と SVM で推定精度の比較実験を行った。評価方法としては本コーパスデータに対して 10 分割の

Table 7: 指摘例

1 REL
2 REL
3 OBJ
4 REL
5 REL
6 OBJ
3,6 文目 本研究の目的が分散して複数存在します
7 STR
研究領域の背景が存在しません
本研究の手法・評価・利点がいずれも存在しません

クロスバリデーションを用いた。

結果を Table 8 に示す。Table 8 より、本研究で提案した特徴量を用いてかつ

Table 8: 特徴量による精度の比較

特徴量	CRF	SVM
本研究で提案した特徴量	71.325%	65.750%
先行研究 [12] の特徴量	63.212%	53.594%

CRF を用いた場合が最も高精度に推定できることが示せた。

6 被験者実験

本節では、被験者実験により提案システムの有用性を示す。被験者をシステム利用群とシステム未利用群に分け、ある研究の資料を元に序論を執筆させた。その後各群に対して異なったフィードバックを与え、フィードバックを踏まえた上で再度執筆させた。これら計 24 本の論文の分析結果及び、執筆後のアンケートの結果を基に本システムの有効性を評価した。各グループの被験者は 6 名とし、いずれも論文執筆経験のないものとした。フィードバックの内容として実験群には「論文構成を意識せよ」のみを与え、対象群にはこれに加えて本システムが指摘したルール違反の内容及び指摘箇所を与えた。

6.1 実験内容

本実験ではまず、本システムが下す指摘が不適ではないことを示すために提出された論文全 24 本に対しての論文構成推定実験を行った。結果は正解率が 66.029% という結果となり、Table 9 で示した結果と比べるとやや精度が落ちてしまっているものの、本実験で指摘の妥当性を示すには十分な正答率であると考えた。次に改定前後の論文の分析結果及びアンケートの結果をそれぞれ得点化し、Mann-Whitney 検定を用いて評価した。分析結果に関しては今回は

1. 改訂前における各群の得点の平均値の差の検定
2. 改訂前後における各群の得点の増加量の平均値の検定

の二つの検定を行った。1. の検定でより群間における得点の有意差がないことを示し、本実験を行う上で群間の能力に有意な差がないことを示し、2. の検定により対象群が実験群より論文の精度を上げられたことを有意差があることで示す。また、アンケート結果に関しては各群の得点の平均値で検定を行った。各 Table では、得点の平均値と分散（括弧内）及びテスト統計量（U 値）と P 値（括弧内）を示しており、*は 5%で有意差があったことを示している。

分析の観点として、

1. 推定によるルール違反数
2. 研究者による論文の評価点
 - (a) 「執筆された論文は学術論文として妥当な形式か」（評価観点 1）
 - (b) 「執筆された論文は論理的か」（評価観点 2）

を用いた。それぞれ、1. はルール違反数、2. については研究者に完成した論文を読んでもらい、6 段階（6:とても当てはまる、5:当てはまる、4:どちらかといえば当てはまる、3:どちらかといえば当てはまらない、2:当てはまらない、1:全く当てはまらない）で評価してもらった。

また、論文執筆後のアンケートとして以下の質問を被験者に与えた。

1. フィードバックの後、論理構成を意識して執筆できたか。
2. フィードバックの内容が妥当であったか。
3. フィードバックの内容が有益であったか。

Table 9: 推定によるルール違反数

	利用群	未利用群	テスト統計量 (p 値)
改訂前	2.00(0.33)	2.00(0.67)	18.00(1.00)
改定前後の増減	-1.33(1.22)	-0.50(0.58)	10.0(0.16)

Table 10: 研究者による評価点

	評価点	利用群	未利用群	テスト統計量 (p 値)
改訂前	評価点 1	2.83(1.77)	2.67(1.70)	17.0(0.87)
	評価点 2	2.83(1.57)	2.50(1.61)	15.5(0.68)
改訂前後の増減	評価点 1	1.50(1.25)	-0.17(0.14)	5.00(0.02)*
	評価点 2	1.83(1.81)	0.17(1.60)	6.50(0.047)*

これらの質問を5段階(5.とても意識できた, 4.意識できた, 3.どちらとも言えない, 2.あまり意識できなかった, 1.意識できなかった)で評価してもらい, その数値を得点とした。

6.2 実験結果

結果をそれぞれ Table 9～Table 11 に示す。

研究者による評価点については, Table10 より改訂前の時点では群間に有意差がないことがわかる。これはどちらの群の被験者も, 初期状態では同程度の能力であったことを示している。一方で, 改定前後の増減では, 2つの評価観点のどちらにおいても, システム利用群の得点が有意に増加した。このことから, 本システムを利用することで, 1) 学術論文として妥当な形式となること, 2) 論理的な文章となること, の二点の利点が示せた。

次に「ルールの違反数」に着目すると, 先ほどと同様に, 改訂前の時点では群間に有意差がないことがわかる。改定前後での違反数の増減については, システム利用群では減少する傾向が確認できるが, 有意差は得られなかった。本実験では, 各群の被験者数が6名と少なかったことが要因の一つと考えられ, 被験者数を増やすことで有意差が得られる可能性は高いと考えられる。

最後に, 「アンケート結果」についてシステム利用群と未利用群で比較すると, Table 11 より, 設問2「フィードバックの内容が妥当であるか」と, 設問3「フィードバックの内容が有益であるか」の二点では, 有意差は認められなかったものの

Table 11: アンケート結果

	利用群	未利用群	テスト統計量 (p 値)
設問 1	3.83(0.90)	4.33(0.47)	16.0(0.72)
設問 2	3.00(1.00)	4.17(0.69)	7.5(0.07)
設問 3	3.00(1.00)	4.00(0.82)	9.00(0.13)

システム利用群で高い評価が得られていた。このことから、提案システムによるフィードバックが利用者にとって妥当であり、有益なものであったことが読み取れる。ただし、設問 1「フィードバックの後、論理構成を意識して執筆できたか」については改善が見られなかった。このことは、違反箇所をフィードバックするだけでは論理構成を意識させるのに不十分であることを示唆していると解釈できる。より論理構成を意識させるためには、具体的な修正方法やサンプルを提示するなど、支援の方法に工夫を行う必要があると考えられる。

7 まとめ

本研究では、CRF 及び、本研究で作成した特徴量を利用して論文構成の推定を行い、得られた結果が「論文構成のルール」を違反していた場合に違反の内容及び違反箇所を指摘するシステムを提案した。本システムの有用性を示す実験として、本研究では論文構成の推定精度実験を行い、本研究で提案する特徴量及び CRF が実際に有効であることを示した。続いて被験者実験として本システムを利用した被験者と利用していない利用者にそれぞれ論文を執筆してもらい、完成した論文に対しての分析実験及び、被験者を対象とするアンケートの得点から有意差検定を行った。

結果としては、研究者評価による「執筆された論文は学術論文として妥当な形式であるか」と「執筆された論文は論理的であるか」の項目で有意差があった。この結果から、初心者が本システムを利用することで、学術論文として妥当な形式となること、論理的な文章となることの二点の利点が見せた。他の項目では有意差は見られなかったが、システムを利用した方が良い論文になる傾向があることがわかった。従って、被験者の人数を増やすことで有意差があることを示せると期待できる。

References

- [1] N.Kando,"Text-level structure of research papers: Implications for text based information processing systems,"19th Annual BCS-IRSG Colloquium on IR,Aberdeen,UK,April 1997
- [2] C.M. Neuwirth,D.S. Kaufer, R. Chimera, and T.Gillespie,"The notes program: A hypertext application for writing from source texts,"Proc. Hypertext'87,pp121-142,1987
- [3] H. Shibata and K.Hori,"A framework to support writing as design using multiple representations,"Proc. 5th Asia Pacific Conference on Computer Human Interaction,2002
- [4] S.M. Alusio and O.N. Oliveira, Jr., "A detailed schematic structure of research paper introduction: samiento del Lenguaje Natural,1996."
- [5] S.T. O'Rourke and R.A. Calvo, "Analyzing semantic flow in academic writing," Proc. 2009 Conference on Artificial Intelligence in Education, 2009.
- [6] T. Fujimoto and T. Matsuo, "An approach to cooperative research paper writing/edit method," IJCSNS International Journal of Computer Science and Network Security, vol.7, no.3, pp.15-20, March 2007.
- [7] H. Kunichika, C. Miyazaki, Y. Matsuki, R. Chartrand and A. Takeuchi, "An intelligent partner for organizing a paragraph," Proc. 2009 Conference on Artificial Intelligence in Education,2009
- [8] 西村健士, 島津秀雄, "特定表現の重点解析による科学技術論文構造化手法," 情報学研報, 情報基礎研究報告会 " vol.93.no39,pp59-62,1993.
- [9] 甲斐郷子, 中村順一, 吉田 將, "表層表現に基づく文章構造解析を利用した論文改定支援システムの試作と評価," 情処学研報, 自然言語処理研究会報告, vol.95,no.27,pp.79-84,1995.
- [10] 岩田芳明, 山村 毅, 大西 昇, "マークアップ方式による文章作成システム," 信学技報, NLC97-60,1998.

- [11] 宇都雅輝・植野真臣 (2011) ベイズ符号を用いた論文構成支援システム. 電子情報通信学会論文誌 D , Vol. J94-D, No. 12, pp.2069-2081.
- [12] Yufan Guo et. al. (2011) Comparison and user-based evaluation of models of textual information structure in the context of cancer risk assessment. BMC Bioinformatics, Vol.12, No. 69, pp.1-18
- [13] Yufan Guo et.al. (2011) A Weakly-supervised Approach to Argumentative Zoning of Scientific Documents. Conference on Empirical Methods in Natural Language Processing, pp. 273 - 283.
- [14] Ronzano, F., Saggion, H. (2016) Knowledge Extraction and Modeling from Scientific Publications. International World Wide Web Conference, Montreal.
- [15] 渋谷英潔 荒木健治 桃内佳雄 栃内香次. 2006. 単語概念の深層格選好に基づく深層格推測手法. 電子情報通信学会論文誌 D. Vol. J89-D No.6 pp1413-1428