

ベイズアプローチを用いた  
アカデミックライティング支援システム

宇都 雅輝

電気通信大学大学院 情報システム学研究科

学位申請論文 博士(工学)

2013年9月

ベイズアプローチを用いた  
アカデミックライティング支援システム

博士論文審査委員会

主査: 植野 真臣 教授  
委員: 大須賀 昭彦 教授  
委員: 長岡 浩司 教授  
委員: 田中 健次 教授  
委員: 田原 康之 准教授

著作権所有者

宇都 雅輝

2013 年

# Academic Writing Support System using Bayes approach

Masaki Uto

## **Abstract**

The purpose of this paper is to develop and evaluate an article structure and argument construction support system for academic writing.

At first, we propose an article structure construction support system. We assume that an article structure, which is defined as a sequence of article component categories, follows  $m$ -th order Markov source. Bayes code is known to estimate the optimal Markov order. However, the code often suffers from over fitting when data length is short. To solve the problem, we propose a bias corrected estimation method based on Bayes code. By using the method, we develop a system which recommends article structures.

Next, we propose an argument construction support system. The traditional systems support construction of an argument as to fit Toulmin model, which is known as a scheme of argument. However, to construct and elaborate an argument, focusing on “Strength of an argument”, which means strength of causal relationship between sentences, is more important than model fitting. However, it is difficult to elaborate all strength of arguments when the argument scale grows up. To solve the problem, we propose a Bayesian network representation of Toulmin model. Using the Bayesian network representation, the proposed system estimates “Strength of an argument”, “Validity of a sentence” and “Influence of a sentence on the claim” and then generates optimal advices to a user.

Moreover, we carried out simulation and subjective experiments to show the effectiveness of the proposed methods and systems.

# ベイズアプローチを用いた アカデミックライティング支援システム

宇都 雅輝

## 要旨

本研究では、アカデミックライティングにおける論文構成と論証の構築を支援するシステムを開発した。

論文構成構築支援システムでは、論文構成が  $m$  重マルコフ情報源に従うと仮定し、過去の優良論文 100 件から推定した予測精度の高い  $m$  重マルコフ情報源に基づき論文構成の構築過程を支援する。更に、予測精度の高い多重度を高精度に推定するために、ベイズ符号語長を用いた多重度の推定補正手法を提案した。

論証構築支援システムでは、文章間の因果関係の強さを意味する「論証の強さ」を重視した論証推敲を支援する。このために、論証の規範モデルとして知られる Toulmin モデルのベイジアンネットワーク表現を用いて、1. 論証の強さ、2. 文章の正当性、3. 主張への影響度、という論証特性指標を提案し、システムに組み込んだ。本システムは、論証構造を可視化するのみでなく、論証の弱い箇所を指摘し、内省を促す。

更に、実験により、各提案手法と提案システムの有効性を確認する。

# 目次

第 1 章	緒言	1
第 2 章	ベイズ符号を用いた論文構成構築支援システム	3
2.1	はじめに	3
2.2	論文要素のカテゴリ化	5
2.3	$m$ 重マルコフ情報源	5
2.4	多重度の推定	9
2.4.1	統計的アプローチ	9
2.4.2	情報論的アプローチ	10
	(1) Normalized Maximum Likelihood	10
	(2) ベイズ符号	11
2.5	ベイズ符号による多重度推定	12
2.5.1	ベイズ符号による多重度推定の分析	12
2.5.2	補正手法	15
2.5.3	補正手法の有効性評価	15
2.6	論文構成構築支援システム	16
2.6.1	$m$ 重マルコフ情報源の推定	16
2.6.2	論文構成構築支援システム	17
2.6.3	好ましくない論文構成の排除	18
2.7	提案システムの有効性評価	21
2.7.1	$m$ 重マルコフ情報源の予測性能の評価	21

2.7.2	システムの有効性評価 . . . . .	22
2.8	むすび . . . . .	25
第3章	Toulmin モデルのベイジアンネットワーク表現を用いた論証構築支援システム	27
3.1	はじめに . . . . .	27
3.2	Toulmin モデル . . . . .	29
3.3	Toulmin モデルのベイジアンネットワーク表現 . . . . .	31
3.3.1	Toulmin モデルのベイジアンネットワーク表現 . . . . .	32
3.3.2	Noisy or モデル . . . . .	35
3.3.3	論証への Noisy or モデル適用の妥当性評価 . . . . .	37
3.4	論証特性指標 . . . . .	38
3.4.1	論証の強さ . . . . .	38
3.4.2	文章の正当性 . . . . .	39
3.4.3	主張への影響度 . . . . .	39
3.4.4	論証特性指標の妥当性評価 . . . . .	40
3.4.5	ユーザによる主観評価の妥当性評価 . . . . .	42
3.5	論証構築支援システム . . . . .	44
3.5.1	論証特性指標に基づく論証改訂方法のフィードバック . . . . .	44
3.5.2	論証構築支援システム . . . . .	46
3.6	評価実験 . . . . .	49
3.6.1	実験方法 . . . . .	49
3.6.2	論証の推敲内容の妥当性評価 . . . . .	50
3.6.3	改訂された論証の質の評価 . . . . .	52
3.7	むすび . . . . .	53
第4章	結言	54
参考文献		56
.1	付録 . . . . .	66

# 目次

2.1	シミュレーションデータにおけるベイズ符号語長 . . . . .	14
2.2	システム構成図 . . . . .	18
2.3	システムインターフェイス . . . . .	19
2.4	好ましい論文構成の割合 . . . . .	21
3.1	Toulmin モデル . . . . .	28
3.2	Toulmin モデルの構成規則 . . . . .	30
3.3	構成規則に従う Toulmin モデルの例 . . . . .	31
3.4	文献 [55] で例示された, 動物実験の禁止を主張する論証の一部 . . . . .	32
3.5	文献 [56] で例示された, 集合論における選択公理を認めるべきであると主張する論証の一部 . . . . .	32
3.6	本博士論文第二章中の論証の一部 . . . . .	32
3.7	Toulmin モデルのベイジアンネットワーク表現 . . . . .	34
3.8	ベイジアンネットワーク表現の例 . . . . .	35
3.9	実験に利用した論証構造 . . . . .	37
3.10	Sensitivity function . . . . .	40
3.11	論証改訂方法のパターン . . . . .	44
3.12	システムインターフェイス . . . . .	47
3.13	自己評価ウィンドウ . . . . .	48
1	論証 1 への適用例 . . . . .	66



---

2	論証 2 への適用例 . . . . .	67
3	学术论文 ( 本論第 2 章 ) への適用例 . . . . .	68

# 表目次

2.1	論文要素カテゴリ	6
2.2	論文構成データの例	7
2.3	論文構成データのデータ長と $m$ 重マルコフ情報源の推定に利用できるデータ数	8
2.4	シミュレーションによる多重度の推定結果	13
2.5	論文構成データにおけるベイズ符号語長とベイズ符号語長の変化率	16
2.6	出現頻度上位 5 位のマルコフ系列	17
2.7	好ましくない論文構成の規則	20
2.8	$m$ 重マルコフ情報源の予測性能の比較結果	23
2.9	被験者実験の結果	25
3.1	Noisy or モデルの妥当性評価結果	38
3.2	実験に利用した論証	39
3.3	論証特性指標の妥当性評価結果	41
3.4	被験者へのフィードバックの妥当性評価	43
3.5	アドバイス文の生成規則	45
3.6	論証改定カテゴリ	49
3.7	実験結果	51

# 第1章

## 緒言

アカデミック・ライティングでは、情報を読者に効果的に伝達できるように文章を執筆することが重要である [1]。このためには、論理構造を意識した文章執筆が不可欠である。特に、アカデミック・ライティングでは、以下の二つの論理構造が重要となる [2][3][4]。

1. 「研究背景」や「先行研究の問題点」、「本研究の目的」といった論文要素の系列で定義される文章全体の大局的な構造。本論では、このような論理構造を「論文構成」と呼ぶ。
2. 「主張」とそれを支持する「データ」や「論拠」などで構成される「論証」。

論文構成を適切に構築することで議論の流れを明確化し、さらに、「主張」が正当化されるように論証を構築することで、文章の論理性向上が期待できる。しかし、これらの作業は、特に初心者にとって難度が高いことが知られている。そこで、本論では、アカデミック・ライティングにおける論文構成と論証の構築を支援するシステムの開発を目的とする。

第2章では、妥当かつ多様な論文構成の構築を支援するシステムを開発する。ここでは、「論文構成」を情報理論における情報源からの出力符号系列とみなしたメタファとしてとらえ、論文構成の構築過程を定式化する。具体的には、論文構成を論文要素カテゴリの系列データとし、それが  $m$  重マルコフ情報源に従うと仮定する。多重度の推定法として、情報論的アプローチでは、ベイズ符号語長最小化による推定法が高精度であると知られている。しかし、本論で扱うようなデータ長の短いデータから学習する場合、多重度の増加に伴いベイズ符号語長が単調減少し、多重度を正しく推定できないことがある。そこで、本研究では、ベイズ符号語長が単調減

少する場合の推定補正法を提案し、過去の優良論文 100 件から予測精度の高い  $m$  重マルコフ情報源を推定する。さらに、推定されたマルコフ情報源に基づき論文構成の構築過程を逐次的にナビゲーションするシステムを開発する。

第 3 章では、論証の構築を支援するシステムを開発する。従来の論証構築支援システムでは、論証の規範モデルとして知られる Toulmin モデルにユーザの論証を当てはめ可視化する支援を行っていることが多い。しかし、論証の主目的である「主張」の正当化のためには、Toulmin モデルへの当てはまりの良さよりも、文章間の因果の強さ、すなわち「論証の強さ」を重視した論証の構築が重要である。しかし、一般の論証では、複数の「論拠」や「データ」、「裏付け」から主張を導くことが多く、論証の構成が複雑化すると、以下の問題が生じると考えられる。1. 「論証の強さ」を全ての文章間について評価することが困難である。2. 論証中の各文章がどの程度正当化できているかの推定が難しい。3. 「主張」の正当化に対して各文章がどのように影響しているかを把握することが困難である。これらの問題を解決するために、本論では、Toulmin モデルのベイジアンネットワーク表現を用いて、1. 論証の強さ、2. 文章の正当性、3. 主張への影響度、という 3 つの指標を算出し、その値に応じて論証改訂のためのアドバイスをフィードバックする論証構築支援システムを開発する。

さらに、本論では、シミュレーション実験と被験者実験により、それぞれのシステムの有効性を示す。

最後に、第 4 章において、本研究で得られた主な研究成果を総括し、本論文をまとめるとともに本研究の課題について述べる。

## 第2章

# ベイズ符号を用いた論文構成構築支援システム

### 2.1 はじめに

アカデミック・ライティングでは、情報を読者に効果的に伝達できるように文章を執筆することが重要であり [1]，このためには文章の適切な構造化が必要であることが指摘されている [2][3]．特に，アカデミック・ライティングに特有かつ重要な文章構造のひとつとして論文構成が知られている [5]．論文構成とは、「研究背景」や「先行研究の問題点」，「本研究の目的」といった論文要素の系列を指す．一般に，論文執筆者は論文構成の一般的な形式を理解し，それに準拠しつつ自身の意図する論文構成を構築することが求められる [6][7]．しかし，特に初心者にとって，このような作業は容易ではないことが指摘されている [8]．

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[9] は，パラグラフ間の関係性を2次元空間に可視化するシステムを開発している．また，Fujimoto and Matsuo[10] 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訂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．さらに，Kunichika *et al* [11] 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 [12]，甲斐ら [13] は，文章の表層表現に基づき利用者の論文構成を解析し，表示するシステムを開発している．Aluisio and Oliveira[5][14]，Aluisio

and Gantenbein[15] は、過去の論文の論文構成を利用者の論文構成と比較し、類似した論文構成を提示するシステムを開発している。また、岩田ら [6]、溝淵ら [7]、山崎ら [16] は、画一的に定義された論文構成の規範と利用者の論文構成とを比較することで、好ましくない構成を指摘・変更するシステムを開発している。しかし、これらのシステムには次の問題点が考えられる。

1. いったん論文を完成させて入力するバッチ処理にのみ対応しており、オンライン処理に対応していないため、論文を構成する過程が支援されない。
2. 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。

ここで、(2) の問題点に対して、Aluisio and Oliveira[5] や岩田ら [6] など同様の形式手法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。

以上の問題を解決するため、本研究では、確率的アプローチを用いた論文構成構築支援システムを開発する。ここでは、「論文構成」を情報理論における情報源からの出力符号系列とみなしたメタファとしてとらえ、論文構成の構築過程を定式化する。具体的には、論文構成を論文要素カテゴリの系列データとし、それが  $m$  重マルコフ情報源に従うと仮定する。多重度の推定法としては、従来から、BIC(Bayesian Information Criteria)[17] や AIC(Akaike Information Criteria)[18] といった統計的アプローチが知られているが [17]、これらの手法では、本論で扱う論文構成データのようなデータ長が短い場合の学習において、多重度を正しく推定できないことが多い。そこで、本研究では、情報論的アプローチにおいて最も高精度なベイズ符号 (Bayes Code)[19][20] により、データ長の短い論文構成データから多重度を推定することを目標とする。しかし、この推定法を用いても、データ長が極端に短い場合、ベイズ符号語長が多重度の増加に従い単調減少し、多重度を正しく推定できない場合がある [21]。そこで、本研究では、ベイズ符号語長が単調減少する場合の推定補正法を提案し、シミュレーション実験により補正手法が多重度の推定精度を向上させることを示す。この補正法を用いて過去の優良論文 100 件から予測精度の高い  $m$  重マルコフ情報源を高精度に推定する。さらに、推定されたマルコフ情報源に基づき論文構成の構築過程を逐次的にナビゲーションするシステムを開発する。本システムの特徴・利点は以下の通りである。

1. 論文構成の構築過程において，利用者が次に選択すべき論文要素を逐次推薦するため，初心者であっても容易に論文構成を構築できる．
2. 論文構成データから推定された  $m$  重マルコフ情報源に従うことで，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の構築を支援できる．

さらに，シミュレーション及び被験者実験により，本システムの有効性を示す．

## 2.2 論文要素のカテゴリ化

本研究では，論文要素をカテゴリ化し，過去の優良論文の論文構成から論文要素の系列データを収集した．対象は，「情報システム」に関する論文とした．本研究において，論文要素とは，「研究背景」や「先行研究の問題点」など，論文中で特定の役割を果たす文章を表す．論文要素は，章や節，パラグラフのように論文執筆者が明示的に構造化した文章と必ずしも対応せず，同じ論文要素を表す連続した文を単位とする．

ここでは，まず，「論文の書き方」など，論文構成の形式を示した文献（科学技術論文の「序論」の形式を示した論文 3 件，情報システム論文の書き方に関する解説論文 3 件，論文構成の一般的な形式を示した著書 6 件）を収集し，これらの文献で定義されている論文要素のカテゴリを抽出した．この際，同義と判断できるカテゴリはグループ化し新たに命名した．次に，抽出したカテゴリを実際の情報システム論文（電子情報通信学会論文誌 D-2009，2010 より 20 件）の論文構成に対応づけながら，新たに必要なカテゴリがあれば追加し，反対に一度も出現しないカテゴリについては削除した．作成した論文要素カテゴリを表 2.1 に示す．

次に，電子情報通信学会論文誌 D(2002～2010) に掲載された「情報システム」に関する優良論文 100 件の論文構成を研究者とともに分析し，表 2.1 の論文要素カテゴリと対応付けることで，論文要素カテゴリの系列データを作成した．収集したデータ例を表 2.2 に示す．以降，このデータを論文構成データと呼ぶ．

## 2.3 $m$ 重マルコフ情報源

本研究では，論文構成データの発生モデルが， $m$  重マルコフ情報源に従うと仮定する．

表 2.1 論文要素カテゴリ

序論	本論	結論
A:研究領域の目標	a:関連研究のレビュー	U:今後の課題
B:研究領域の現状	b:研究領域の詳細	V:まとめ
C:研究領域における問題点	c:システムに用いる理論の概要	W:謝辞
D:先行研究の概要	d:システムに用いる理論の詳細	
E:先行研究の利点	e:理論の評価実験の目的	共通
F:先行研究の問題点	f:理論の評価実験の方法	X:開始
G:本研究の目的	g:理論の評価実験の結果	Y:終了
H:本研究の方法	h:理論の評価実験の考察	
I:本研究の特徴	i:システムの概要	
J:本研究の利点	j:システムの全体設計	
K:本研究の評価	k:システムの詳細設計	
L:用語等定義	l:システムの機能	
M:本論の構成	m:システムの動作例	
	n:システムの運用実績	
	o:システムの評価実験の目的	
	p:システムの評価実験の方法	
	q:システムの評価実験の結果	
	r:システムの評価実験の考察	
	s:システムの評価実験の概要	
	t:本研究の目的	
	u:本研究の利点	
	v:今後の課題	

*m* 重マルコフ情報源は、系列データを扱う最も基礎的なモデルであり、パラメータの推定や実装が容易かつ頑健であるという利点から多様な分野で利用され、その有効性が確認されている[22]。さらに、最先端の時系列データ解析技術であるダイナミック・ベイジアン・ネットワーク



表 2.2 論文構成データの例

序論	本論	結論
XACDFGHJKY	XdijkmopqraY	XVUWY
XAGJMY	XcdijksopqropqraY	XVUWY
XBADFGHMY	XjkopqrY	XVUY
XABDFDFGY	XcdjlkopqropqropqrY	XVY
XBCBDFDFGHKMY	XacdijksopqrY	XVUWY

や隠れマルコフモデルなどの下位モデルであり，拡張性にも優れている．そこで，ここでは， $m$  重マルコフ情報源を予測に利用する．

$m$  重マルコフ情報源は，任意の時点  $t$  における状態  $X_t$  の生起確率が，その直前の  $m$  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  $m$  重マルコフ過程とみなせる，すなわち，等式 (2.1)

$$p(X_t | X_1^{t-1}) = p(X_t | X_{t-m}^{t-1}) \quad (2.1)$$

を満たす情報源である．式 (2.1) において， $m = 0$  とすると，各状態が独立に発生する無記憶情報源と等価となる．また， $p(X_t | X_{t-m}^{t-1})$  は遷移確率と呼ばれる．

ここで，変数  $X_t$  は情報源アルファベット  $A = \{a_1, \dots, a_k\}$  から値を取り，状態系列  $X_{t-m}^{t-1}$  は  $m$  重マルコフ情報源の状態集合  $A^m$  から値を取るとする．また，状態  $s \in A^m$  のもとでシンボル  $a \in A$  が発生する確率  $p(a|s)$  を  $\theta_{sa}$  で表し，このパラメータ集合を  $\Theta = \{\theta_{sa}\}$  と表記する．

ここで，長さ  $n$  の系列データ  $X_1^n$  が得られたとき，これが  $m$  重マルコフ情報源に従うとすると，このデータの尤度は以下の多項分布で与えられる．

$$L(X_1^n | \Theta) \propto \prod_{s \in A^m} \prod_{a \in A} (\theta_{sa})^{n_{sa}} \quad (2.2)$$

ただし， $n_{sa}$  は状態  $s$  のもとでシンボル  $a$  が発生したデータの数を表す．また，パラメータ  $\theta_{sa}$  の最尤推定量は， $\hat{\theta}_{sa} = n_{sa}/n_s$  となる．ただし， $n_s = \sum_{a \in A} n_{sa}$  である．

他方，本論で扱う論文構成データは，100 件の独立した論文構成の集合として定義される．本研究では，論文構成データ中の各論文構成が，同一の  $m$  重マルコフ情報源に従って，独立に

表 2.3 論文構成データのデータ長と  $m$  重マルコフ情報源の推定に利用できるデータ数

データ長 $l_e$	$m$ 重マルコフ情報源の推定に利用できるデータ数 $n_{em}$					
	m=0	m=1	m=2	m=3	m=4	m=5
序論 (819)	819	719	619	519	419	321
本論 (1240)	1240	1140	1040	940	840	740
結論 (460)	460	360	260	160	62	0

生起すると仮定する．ただし，論文構成の「序論」「本論」「結論」は，異なる  $m$  重マルコフ情報源に従うとする．

ここで，論文構成データ中の論文構成の個数を  $J$ ，{序論，本論，結論}の要素を  $e, j$  番目の論文構成の  $e$  に関する系列データを  $X_{ej}$ ， $\mathbf{X}_e = \{X_{ej}\}(j = 1, \dots, J)$ ， $e$  に関する論文要素の集合を  $A_e$  とすると，データ  $\mathbf{X}_e$  の尤度は次式で与えられる．

$$L(\mathbf{X}_e|\theta) \propto \prod_{s \in A_e^m} \prod_{a \in A_e} \prod_{j \in J} (\theta_{sa})^{n_{sa}^j} \quad (2.3)$$

ただし， $n_{sa}^j$  は， $j$  番目の論文構成において，状態  $s$  のもとでシンボル  $a$  が発生したデータの数を表す．また，式 2.3 におけるパラメータ  $\theta_{sa}$  の最尤推定量  $\hat{\theta}_{sa}$  は次式で与えられる．

$$\hat{\theta}_{sa} = \frac{\sum_{j \in J} n_{sa}^j}{\sum_{a \in A_e} \sum_{j \in J} n_{sa}^j} = \frac{n_{sa}}{n_s} \quad (2.4)$$

このとき， $n_{sa} = \sum_{j \in J} n_{sa}^j$ ， $n_s = \sum_{a \in A_e} \sum_{j \in J} n_{sa}^j$  とする．

本研究では，系列データ  $X_{ej}$  のデータ長を  $l_{ej}$  で表し， $l_e = \sum_{j \in J} l_{ej}$  を論文構成データ  $\mathbf{X}_e$  のデータ長と呼ぶ．また，データ長  $l_e$  の論文構成データ  $\mathbf{X}_e$  を用いて， $m$  重マルコフ情報源を推定する場合，推定に利用できるデータ数  $n_{em}$  は， $n_{em} = \sum_{s \in A_e^m} \sum_{a \in A_e} \sum_{j \in J} n_{sa}^j$  となる．

ここで，本論で扱う論文構成データのデータ長と， $m$  重マルコフ情報源の推定に利用できるデータ数を表 2.3 に示す．表 2.3 から，本論で扱う論文構成データはデータ長が短いため，多重度の増加に伴い  $m$  重マルコフ情報源の推定に利用できるデータ数が極端に少なくなることがわかる．本研究では，このデータから  $m$  重マルコフ情報源の多重度  $m$  の正確な推定を目指す．一般に，このようなデータから多重度を正確に推定することは一般に容易ではないが，近年，従来

のモデル選択基準より予測精度を高めた手法が提案されている．本研究では，これをさらに改良し，より正確に論文構成データから  $m$  重マルコフ情報源の多重度  $m$  を推定する．

## 2.4 多重度の推定

$m$  重マルコフ情報源の予測精度は多重度  $m$  に依存するため，多重度を正確に推定することは重要な課題である．多重度の推定法として，以下に紹介する統計的アプローチと情報論的アプローチが知られている [20] ．

### 2.4.1 統計的アプローチ

統計的アプローチでは，データ  $X_e$  が得られたとき，データ数の関数である  $pen(n_{em})$  を所与として，多重度の推定値  $\hat{m}_e$  を次式で与える．

$$\hat{m}_e = \operatorname{argmin}_m \left\{ -\log L(X_e | \hat{\theta}) + K pen(n_{em}) \right\} \quad (2.5)$$

ここで， $K$  はモデルの自由パラメータ数であり， $K = (|A_e| - 1)|A_e|^m$  である．式 (2.5) の右辺第一項はモデルのデータへの当てはまりの良さと解釈でき，多重度の増加に伴うモデルの複雑化に従い単調減少するのに対し，第二項は単調増加する．すなわち，第二項はモデルの複雑さに対するペナルティ項と解釈できる．このため，式 (2.5) の最小化により，過学習を避け，データのモデルへの当てはまりの良さとモデルの複雑さとのトレードオフを解消した多重度の推定を実現できる．統計的アプローチでは， $pen(n_{em}) = \frac{1}{2} \log n_{em}$  とした BIC (Bayesian Information Criterion) や  $pen(n_{em}) = 1$  とした AIC (Akaike Information Criterion) [18] が知られている [17] ．

しかし，統計的アプローチによる従来の情報量基準では，本論で扱う論文構成データのようなデータ長が不十分な場合の学習において，正しく多重度を推定できないことが多い．そこで，本研究では，以下に示す情報論的アプローチを用いて，データ長の短い論文構成データから高精度に多重度を推定することを試みる．

## 2.4.2 情報論的アプローチ

情報論的アプローチでは、データとモデルを一意復号可能な条件のもとで符号化したとき、その総符号語長を最小化するモデルが、データを最もよく表現するモデルであるとする。すなわち、モデルと系列データ  $X_e$  の総符号語長を  $l^m(X_e)$  とすると、多重度の推定値  $\hat{m}_e$  は、

$$\hat{m}_e = \underset{m}{\operatorname{argmin}} \{l^m(X_e)\} \quad (2.6)$$

で与えられる。このようなモデル選択原理は MDL 原理 (minimum description length principle)[23] として知られている。MDL 原理は BIC としばしば混同されるが、Rissanen は BIC と同形の符号長を対数予測分布の漸近的な近似値として導いており、BIC は対数予測分布の漸近的な近似の一つに過ぎないと述べている [24]。実際、BIC は、対数予測分布の  $O(1)$  の近似であるが、以下で紹介する符号語長を用いた最先端のモデル選択手法では、この誤差をより小さくでき、より少数データから高精度の学習が実現できる。

情報論的アプローチにおける符号語長として、ユニバーサル符号化理論より導かれ、ミニマックス冗長度を達成することが知られる NML(Normalized Maximum Likelihood)[24] とバイズ符号 [19] による符号語長が知られている。

### (1) Normalized Maximum Likelihood

NML は次式のように尤度を正規化した形式で定義される。

$$p_{nml}^m(X_e) = \frac{L(X_e|\theta)}{\sum_{y \in A_e^{l_e}} L(y|\hat{\theta}(y))} \quad (2.7)$$

ただし、 $\hat{\theta}(y)$  はデータ系列  $y$  を所与として最尤推定されたパラメータ集合を表す。

NML の符号語長  $l_{nml}^m(X_e)$  は、式 (2.7) の対数を取り、 $l_{nml}^m(X_e) = -\log p_{nml}^m(X_e)$  で与えられ、データ  $X_e$  の最小符号語長となることが示されている [24]。しかし、式 (2.7) の分母は、長さ  $l_e$  の全パターンのデータ系列に対する尤度の和となっており、計算量が膨大である。近年、Kontkanen and Myllymaki[25] によって、効率的な計算手法が提案されているが、実用レベルには至っていない。

## (2) ベイズ符号

ベイズ符号は、情報源の確率分布のクラスのみが既知であり、そのパラメータについては未知である場合を扱うユニバーサル符号において、以下の冗長度

$$\int_{\Theta} \sum_{X_e} L(X_e|\theta) \log \frac{L(X_e|\theta)}{Q(X_e)} p(\theta) d\theta \quad (2.8)$$

を最小にする符号である [26] . ここで、 $Q(X_e)$  は、データ  $X_e$  に対する符号化確率を表し、 $p(\theta)$  は、パラメータ  $\theta$  の事前分布を表す . また、式 2.8 において、 $\Theta$  はパラメータ  $\theta_{sa}$  のベイズ推定値  $\hat{\theta}_{sa}^{bayes}$  の集合を表す . ベイズ推定値は次式で定義される .

$$\hat{\theta}_{sa}^{bayes} = \frac{\sum_{j \in J} n_{sa}^j + n'_{sa}}{\sum_{a \in A_e} (\sum_{j \in J} n_{sa}^j + n'_{sa})} = \frac{n_{sa} + n'_{sa}}{\sum_{a \in A_e} (n_{sa} + n'_{sa})} \quad (2.9)$$

ここで、 $n'_{sa}$  は事前分布のハイパーパラメータを表す .

ベイズ符号は、式 (2.8) を最小にする符号化確率  $Q(X_e)$  として、式 (2.10) で与えられる .

$$Q(X_e) = \int_{\Theta} L(X_e|\theta) p(\theta) d\theta \quad (2.10)$$

$m$  重マルコフ情報源における標準的なベイズ符号として、以下の Krichevsky-Trofimov(KT) 符号 [27] が知られている .

$$Q_{KT}(X_e) = \prod_{\substack{s \in A_e^m: \\ n_{sa} \geq 1}} \frac{\prod_{a \in A_e: n_s \geq 1} \frac{1}{2} \cdot \frac{3}{2} \cdots (\frac{1}{2} + n_{sa} - 1)}{\frac{|A_e|}{2} \cdot (\frac{|A_e|}{2} + 1) \cdots (\frac{|A_e|}{2} + n_s - 1)} \quad (2.11)$$

KT 符号は、個別冗長度の一樣な上下界が可能にし、無記憶情報源を対象にする場合、ミニマックス冗長度の定数項を最適にするという利点を持つ [28] . また、KT 符号は、ベイズ予測分布の事前分布  $p(\theta)$  に、Jeffreys の事前分布、すなわち、ハイパーパラメータを  $1/2$  としたディレクレ分布を仮定した場合と等価となる .

これより、ベイズ符号語長  $l_{bayes}^m(X_e)$  は次式で与えられる .

$$l_{bayes}^m(X_e) = -\ln Q_{KT}(X_e) \quad (2.12)$$

また、ベイズ符号語長は、理論的には、現在最も高精度である NML と漸近的に同等の推定結果を与えることが知られており [29]、NML より容易に計算できるという利点をもつ . そこで、

本研究では、ベイズ符号語長を用いて、データ長の短い論文構成データから、より正確に多重度を推定することを試みる。

## 2.5 ベイズ符号による多重度推定

### 2.5.1 ベイズ符号による多重度推定の分析

ここでは、まず、ベイズ符号に基づく多重度推定法の有効性を示すために、データ長の短いシミュレーション・データを用いた実験を行う。実験手順は以下の通りである。

1. アルファベット数  $|A| = 10$  の  $0, \dots, 4$  重マルコフ情報源を構成し、真のモデルとする。
2. それぞれのモデルから長さ  $n = 100, 200, 500, 1000, 2000$  のデータをランダムにサンプリングする。
3. これらのデータからベイズ符号を用いて式 (2.6) によりモデルを推定する。また、比較のために、BIC と AIC による推定も行う。
4. (1)~(3) の手順を 1000 回繰り返し、それぞれの手法で各モデルが推定された回数を記録する。

表 2.4 に実験結果を示す。表中では、多重度の推定値が真値と一致した回数を“ ”、真値より小さくなった回数を“-”、真値より大きくなった回数を“+”、単調減少し多重度の上限 5 以上となった回数を“ $\hat{m} \geq 5$ ”に示している。

表より、 $m \leq 1$  のとき、ベイズ符号語長を用いたモデル推定が最も精度が高いことがわかる。しかし、この推定法でも、パラメータ数に対してデータ長が不十分な場合、ベイズ符号語長が単調減少し、多重度を決定できないことがある [21]。実験結果からも、データ長が短いときやパラメータ数が増大したときに、このような現象が多数確認できる。2.3 で述べたように、本研究で扱う論文構成データは、「序論」、「本論」、「結論」の状態集合サイズが  $|A| = 15, 24, 5$  であるのに対し、データ長が  $n = 819, 1240, 460$  と短いため、同様の現象が予測される。

この問題の原因を明らかにするために、以下、より詳細な分析を行う。

多重度の増加に伴うベイズ符号語長の単調減少は、真値より大きい多重度で本来増加すべきベイズ符号語長が、何らかの要因により減少するために起こる。Silander et.al [30] は、ベイズ

表 2.4 シミュレーションによる多重度の推定結果

真の多重度  $m = 0$  パラメータ数  $|A| = 10$ 

n	ベイズ符号語長				BIC				AIC			
	-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$	
100	922	0	0	78	1000	0	0	0	1000	0	0	0
200	975	0	0	25	1000	0	0	0	999	0	1	0
500	995	0	0	5	1000	0	0	0	1000	0	0	0
1000	994	0	1	5	1000	0	0	0	1000	0	0	0
2000	999	0	0	1	1000	0	0	0	1000	0	0	0

真の多重度  $m = 1$  パラメータ数  $|A^2| = 100$ 

n	ベイズ符号語長 + 補正手法				ベイズ符号語長				BIC				AIC			
	-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$	
100	929	0	71	0	650	0	6	344	0	1000	0	0	996	4	0	0
200	998	1	1	0	985	1	1	13	0	1000	0	0	999	1	0	0
500	1000	0	0	0	1000	0	0	0	805	195	0	0	1000	0	0	0
1000	1000	0	0	0	1000	0	0	0	1000	0	0	0	1000	0	0	0
2000	1000	0	0	0	1000	0	0	0	1000	0	0	0	1000	0	0	0

真の多重度  $m = 2$  パラメータ数  $|A^3| = 1000$ 

n	ベイズ符号語長 + 補正手法				ベイズ符号語長				BIC				AIC			
	-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$	
100	902	51	76	0	40	0	3	957	0	1000	0	0	0	1000	0	0
200	884	1	97	0	412	0	34	554	0	1000	0	0	0	1000	0	0
500	997	0	3	0	988	0	3	9	0	1000	0	0	0	1000	0	0
1000	1000	0	0	0	1000	0	0	0	0	1000	0	0	999	1	0	0
2000	1000	0	0	0	1000	0	0	0	0	1000	0	0	1000	0	0	0

真の多重度  $m = 3$  パラメータ数  $|A^4| = 10000$ 

n	ベイズ符号語長 + 補正手法				ベイズ符号語長				BIC				AIC			
	-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$	
100	432	568	0	0	4	0	0	996	0	1000	0	0	0	1000	0	0
200	867	126	7	0	21	0	7	972	0	1000	0	0	0	1000	0	0
500	945	0	55	0	391	0	55	554	0	1000	0	0	0	1000	0	0
1000	943	0	57	0	830	0	57	113	0	1000	0	0	0	1000	0	0
2000	1000	0	0	0	998	0	0	2	0	1000	0	0	0	1000	0	0

真の多重度  $m = 4$  パラメータ数  $|A^5| = 100000$ 

n	ベイズ符号語長 + 補正手法				ベイズ符号語長				BIC				AIC			
	-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$		-	+	$\hat{m} \geq 5$	
100	0	1000	0	0	0	0	0	1000	0	1000	0	0	0	1000	0	0
200	2	998	0	0	2	0	0	998	0	1000	0	0	0	1000	0	0
500	89	911	0	0	89	1	0	910	0	1000	0	0	0	1000	0	0
1000	328	672	0	0	328	0	0	672	0	1000	0	0	0	1000	0	0
2000	684	316	0	0	684	0	0	316	0	1000	0	0	0	1000	0	0

符号語長においてディリクレ分布のハイパーパラメータを 1 未満とすると、過疎なデータから学習した際、より複雑なモデルが選択されてしまうと報告している。これは、多重度の増加に伴うベイズ符号語長が単調減少することを意味している。したがって、ベイズ符号語長の単調減少は、多重度の増加に伴い、パラメータ推定のために割り振られるデータが過疎化するために起こ

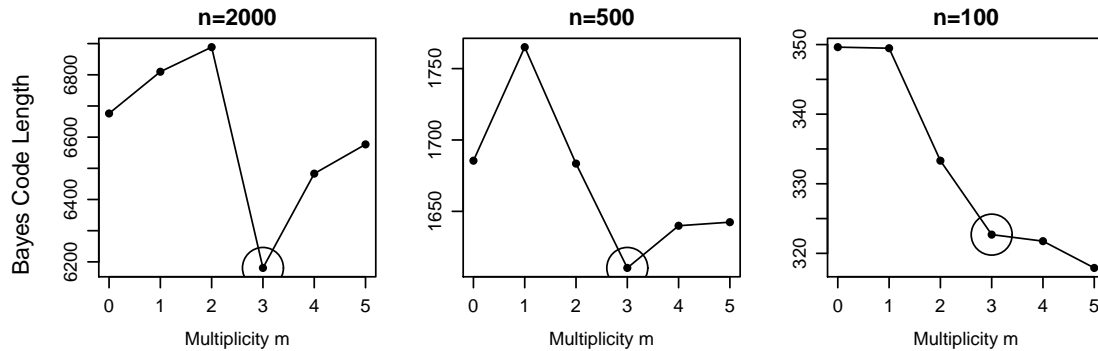


図 2.1 シミュレーションデータにおけるベイズ符号語長

と考えられる。

例として、図 2.1 に、3 重マルコフ情報源から発生させたシミュレーションデータ ( $n = 100, 500, 2000$ ) に対する、多重度  $m$  を所与としたときのベイズ符号語長を示す。図 2.1 では、データ長が十分にある場合 ( $n = 2000, 500$ )、真の多重度でベイズ符号語長が最小となっているが、データ長が  $n = 100$  と短い場合、多重度の増加に従いベイズ符号語長が単調減少している。

また、図 2.1 から、ベイズ符号語長が単調減少する場合、真値より大きい多重度においてベイズ符号語長の減少量が小さくなっていることがわかる。多重度が真値より小さい場合、ベイズ符号語長が真値で最小になるように減少していくとともに、パラメータの増加に伴うデータの過疎化が、ベイズ符号語長の減少を促進する。一方、多重度が真値より大きくなると、データの過疎化によるベイズ符号語長の減少と本来のデータによるベイズ符号語長の増加が打ち消しあうため、結果としてベイズ符号語長の減少量が小さくなると解釈できる。また、Silander et al[30] は、データの過疎化によるベイズ符号語長の減少量がパラメータ数の増加に伴い減少していくことを報告しており、本論の解釈に一致する。

本研究では、データ長が短く、ベイズ符号語長が単調減少する場合に、ベイズ符号語長の減少量が真の多重度を境に大きく減少することを利用して、真の多重度を推定する手法を提案する。



### 2.5.2 補正手法

ベイズ符号語長の減少量が大きく減少する点 (以下, 変化点と呼ぶ) を見つけるには, 次式で定義する変化率を最小化する点を見つければよい.

$$R(m, X_e) = \frac{l_{bayes}^{m+1}(X_e) - l_{bayes}^m(X_e)}{l_{bayes}^m(X_e) - l_{bayes}^{m-1}(X_e)} (m = 1, \dots, M-1) \quad (2.13)$$

ここで,  $M$  は多重度の上限を表し, 問題に応じて分析者により決定される [20].

式 (2.13) の変化率は,  $m$  から  $(m+1)$  の間でのベイズ符号語長の減少量を  $(m-1)$  から  $m$  の間の減少量で除して求めるため, ベイズ符号語長が単調減少する場合, ベイズ符号語長の変化量が大きく減少する変化点において, この値は最小となる. したがって, ベイズ符号語長が単調減少する場合, 多重度の推定値は次式で求められる.

$$\hat{m}_e = \underset{m}{\operatorname{argmin}} \{R(m, X_e)\} \quad (2.14)$$

この補正手法は, ベイズ符号語長の最小化の一般形となっているため, ベイズ符号語長が単調減少する場合だけでなく, 真値で最小値をとる場合にも真の多重度が得られる.

### 2.5.3 補正手法の有効性評価

次に, 補正手法により多重度の推定精度がどの程度改善されるかを評価するために, 2.5.1 のシミュレーション・データに対して補正手法を適用した. ただし, 式 (2.13) より, 補正手法では  $\hat{m} = 0$  を推定できない. しかし, 表 2.4 から確認できるように, 真の多重度が  $m = 0$  である場合, パラメータ数が少なくデータの過疎化が起こらないため, 式 (2.6) のベイズ符号語長の最小化で高精度に真の多重度を推定できる. そこで, 実際には, まず, ベイズ符号語長の最小化により多重度を推定し, ベイズ符号語長が単調減少となった場合に補正手法を用いた. 結果を表 2.4 に示す. 表 2.4 から, パラメータ推定に割り当てられるデータ数が極端に少ない場合 (真の多重度  $m = 4$  の場合) には, 補正手法でも真の多重度が得られていないが, そうでない場合には補正手法により多重度の推定精度を大幅に向上できたことがわかる. 以上より, 補正手法は, パラメータ数の増加により短いデータから正しく多重度を推定できない場合の推定補正法として有

表 2.5 論文構成データにおけるベイズ符号語長とベイズ符号語長の変化率

	ベイズ符号語長						ベイズ符号語長の変化率			
	m=0	m=1	m=2	m=3	m=4	m=5	m=1	m=2	m=3	m=4
序論	2617	2292	2211	2205	2077	1832	0.25	0.07	21.33	1.91
本論	3840	3622	3435	3344	3309	3224	0.85	0.49	0.38	2.43
結論	734	719	606	556	440	287	7.53	0.44	2.32	1.32

効といえる。

## 2.6 論文構成構築支援システム

### 2.6.1 $m$ 重マルコフ情報源の推定

ここでは、論文構成データから、「序論」、「本論」、「結論」の多重度を推定する。まず、それぞれのベイズ符号語長を表 2.5 に示す。なお、多重度の上限は、 $M = 5$  とした。表 2.5 より、ベイズ符号語長が多重度の増加に従い単調減少していることがわかる。そこで、次に補正手法を用いて多重度の推定を行った。表 2.5 にベイズ符号語長の変化率を示す。表 2.5 より、多重度の推定値は「序論」 $\hat{m} = 2$ 、「本論」 $\hat{m} = 3$ 、「結論」 $\hat{m} = 2$  となった。一方、統計的アプローチによる多重度の推定値は全て  $\hat{m} = 1$  となっている。これは、ペナルティ項の値が多重度の増加に従い指数的に増大するのに対し、尤度の増加量がデータの過疎化のために小さくなることが原因であると考えられ、2.5.1 のシミュレーション実験より妥当な推定値ではないといえる。

推定された多重度の大きさから、「論文構成」は画一的な形式となるのではなく、直前の論文要素の系列パターンに応じて多様な形式を取ることがわかる。ここで、表 2.6 に、「序論」、「本論」、「結論」における、出現頻度が上位 5 位の長さ ( $\hat{m} + 1$ ) の系列パターンを示す。これらの結果から、以下の考察が可能である。

1. 「序論」は、比較的近い過去の論文要素に依存して論文構成が構築されており、基本的な議論の流れは、「研究領域を概括し、複数の関連研究を議論した後、本研究の目的を述べる」となっている。

表 2.6 出現頻度上位 5 位のマルコフ系列

	序論	本論	結論
1 位	D, F, G (37)	o, p, q, r (120)	X, V, U (76)
2 位	X, B, C (36)	p, q, r, Y (81)	U, W, Y (48)
3 位	D, F, D (23)	p, q, r, o (30)	V, U, W (47)
4 位	B, C, D (21)	q, r, o, p (30)	V, U, Y (29)
5 位	X, B, A (20)	r, o, p, q (30)	X, V, Y (2)

2. 「本論」は，論文構成が比較的長距離の論文要素に依存しており，ほとんどの情報システム論文で，複数の評価実験が「目的，方法，結果，考察」の反復で記述されている．
3. 「結論」は，論文要素数が少ないため，表 2.6 の 5 つのパターンによりほぼ全ての構成が表現されている．

### 2.6.2 論文構成構築支援システム

本節では，推定された  $m$  重マルコフ情報源に基づく，論文構成構築支援システムについて詳述する．本システムは，構築過程の利用者の論文構成を，推定された  $m$  重マルコフ情報源に当てはめることで，利用者が次に選択すべき論文要素を逐次予測し，遷移確率の大きい順に利用者に推薦する．なお，この推薦は「序論」，「本論」，「結論」ごとに独立に行う．

本システムの構成図を図 2.2 に示す．本システムは，論文構成の構築・編集を管理する論文構成構築システムとマルコフ情報源と論文構成データを保持するデータ部により構成されている．

本システムでは，図 2.3 の画面で論文構成を構築することができる．図 2.3 では，画面左に構築過程の利用者の論文構成が表示され，画面右上に次に選択すべき論文要素とその遷移確率が，遷移確率の大きい順にリスト表示されている．また，リスト内の論文要素をクリックすると，画面右下のウィンドウに，その論文要素の解説と例文が表示される．これらを参考に，利用者が自身の意図に合った論文要素を選択すると，利用者の論文構成の末尾にその論文要素が追加

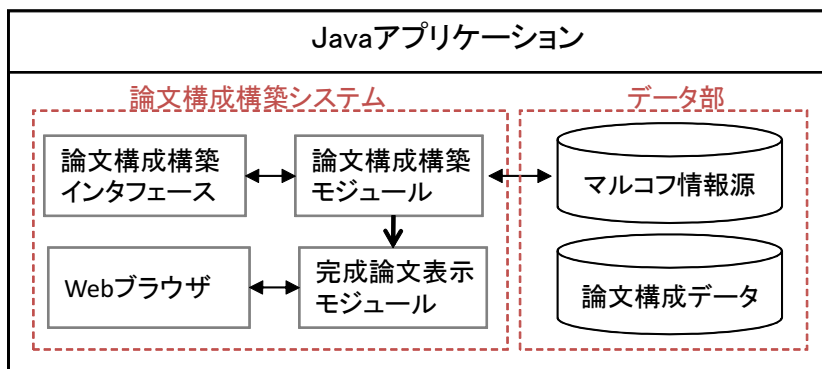


図 2.2 システム構成図

され、同時に推薦リストも更新される。この反復により、利用者は自身の意図する論文構成を、マルコフ情報源にしたがって構築できる。なお、本システムでは、構築過程の論文構成の長さ  $n$  が多重度  $m$  より短い間は、全論文要素系列  $X_1^n$  を所与として  $p(X_{n+1}|X_1^n)$  に従い論文要素を推薦する。さらに、画面左の各論文要素の下にテキストボックスが設置されるため、利用者は論文構成を逐次編集しながら、論文を執筆することができる。また、執筆した論文は HTML 形式に変換して出力できる。なお、本システムは、Java(Swing) を用いて実装した。

### 2.6.3 好ましくない論文構成の排除

本システムでは、論文要素の推薦の際、既に選択された論文要素も再度推薦されるため、論文要素の選択の仕方（遷移確率の低い論文要素を連続して選択するなど）によって好ましくない論文構成が構築される可能性がある。そこで、ここでは、本システムにより好ましい論文構成が構築される割合を以下の手順で評価する。

1. 論文構成データから推定した  $m$  重マルコフ情報源 ( $m = 0, \dots, 5$ ) に従い、論文構成をランダムに 1000 件ずつ発生させる。
2. これらが好ましい論文構成であるか否かを、研究者と作成した「好ましくない論文構成の規則」(表 2.7) を用いて判定する。具体的には、表 2.7 の規則がひとつでも当てはまれば、その論文構成を好ましくないものと判定する。
3. 上記の (1), (2) を「序論」、「本論」、「結論」ごとに行い、好ましい論文構成が構築される

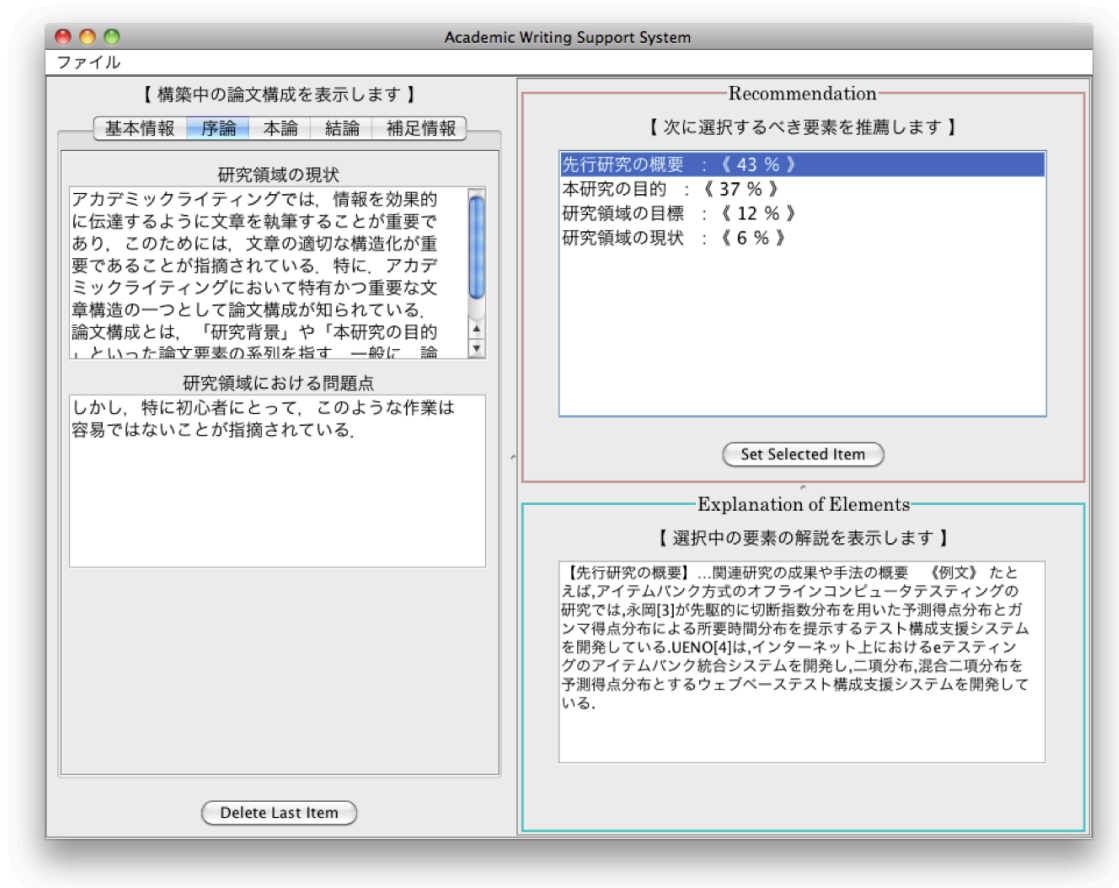


図 2.3 システムインターフェイス

割合を式 (2.15) で算出する .

$$F_{me} = \frac{1}{|C_{me}|} \sum_{i=1}^{|C_{me}|} z_{mei} \quad (2.15)$$

ただし,  $C_{me}$  は  $m$  重マルコフ情報源に従いランダムに構築した論文構成の  $e$  の集合 ( $|C_{me}| = 1000$ ),  $z_{mei}$  は  $C_{me}$  の  $i$  番目のデータに好ましくない構成が含まれるとき 0, それ以外で 1 となるダミー変数とする .

結果を図 2.4 に示す . 図 2.4 より, 多重度の増加に伴い好ましい論文構成の出現割合が増加していることがわかる . これは, 多重度の増加に伴い, 予測時に利用される論文要素系列が長く

表 2.7 好ましくない論文構成の規則

序論	<ul style="list-style-type: none"> <li>・「研究領域の目標，現状，問題点」より前に「本研究の目的，方法，特徴，利点，評価」が存在</li> <li>・「研究領域の目標・現状・問題点」と「先行研究の概要・利点・問題点」が混在</li> <li>・「本研究の目的」より前に「本研究の方法，特徴，利点，評価」が存在</li> <li>・「本研究の目的」がない</li> <li>・「本研究の目的」が複数存在</li> <li>・「本論の構成」が最後以外に存在</li> </ul>
本論	<ul style="list-style-type: none"> <li>・「システム」に関する論文要素がないにも関わらず「システムの評価実験」が存在</li> <li>・「理論」に関する論文要素がないにも関わらず「理論の評価実験」が存在</li> <li>・「システムの概要」より前に「システムの設計，機能，動作例，評価」が存在</li> <li>・「システムの全体設計」が複数存在</li> <li>・「システム」も「理論」もない</li> <li>・「理論の評価実験」より前に「システムの評価実験」が存在</li> <li>・「評価実験の目的，方法」がないにも関わらず「評価実験の結果，考察」が存在</li> <li>・「今後の課題」が最後以外に存在</li> </ul>
結論	<ul style="list-style-type: none"> <li>・「まとめ」がない</li> <li>・「謝辞」が最後以外に存在</li> </ul>

なり，推薦される論文要素の多様性が減少するため，論文構成データと同様の論文構成が出現しやすくなるためと考えられる．また，図 2.4 より，「結論」は，論文要素数が少ないため，多重度の推定値  $m = 2$  で好ましくない論文構成の出現割合が 0 となったが，「序論」「本論」では，各々の多重度の推定値  $m = 2, 3$  において，2 割程度好ましくない論文構成が出現してしまうことがわかる．そこで，システムの最終的な実装では，表 2.7 の規則に従い，好ましくない論文構成の出現を排除している．

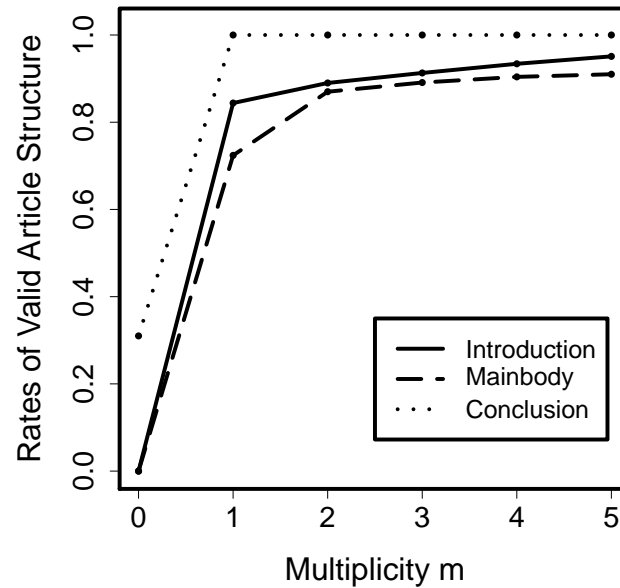


図 2.4 好ましい論文構成の割合

## 2.7 提案システムの有効性評価

### 2.7.1 $m$ 重マルコフ情報源の予測性能の評価

本節では，推定された  $m$  重マルコフ情報源が，予測に優れたモデルであることを実験により示す．予測性能の指標には「 $F$  値 ( $F$ -measure)」を用いる． $F$  値は，予測の正確さの指標と解釈できる「適合率 ( $Precision$ )」と予測の網羅性・多様性の指標と解釈できる「再現性 ( $Recall$ )」との調和平均であり，情報検索・推薦システムにおける予測性能の評価指標として一般に用いられる [31][32]．ここでは，電子情報通信学会論文誌 D(2008～2011) から評価用の論文構成データ (以下，評価データ) を新たに 20 件収集し，次式で  $F$  値を算出した．

$$F\text{-measure}_{me} = \frac{2Precision_{me} \cdot Recall_{me}}{Precision_{me} + Recall_{me}} \quad (2.16)$$

$$Precision_{me} = \frac{1}{\sum_{t=1}^T l_t} \sum_{t=1}^T \sum_{l=1}^{l_t} \frac{R_{tl}}{N_{tl}} \quad (2.17)$$

$$Recall_{me} = \frac{1}{\sum_{t=1}^T l_t} \sum_{t=1}^T \sum_{l=1}^{l_t} \frac{R_{tl}}{C_{tl}} \quad (2.18)$$

ここで、 $T$  は評価データの総数 ( $T = 20$ )、 $l_t$  は  $t$  番目の評価データのデータ長、 $N_{tl}$  は  $t$  番目の評価データにおける系列データ  $X_{l-m}^{l-1}$  を所与として  $m$  重マルコフ情報源から推薦される論文要素数、 $R_{tl}$  は  $t$  番目の評価データにおける  $l$  番目の論文要素が、 $m$  重マルコフ情報源から推薦された論文要素集合の要素である時に 1 そうでないときに 0 を取るダミー変数、 $C_{tl}$  は  $t$  番目の評価データの  $l$  番目の論文要素が論文要素集合の要素である時に 1 そうでないときに 0 を取るダミー変数を表す。

$F$  値の算出結果を表 2.8 に示す。表より、 $F$  値は「序論」 $m = 3$ 、「本論」 $m = 3$ 、「結論」 $m = 2$  で最大となり、「序論」以外は補正手法による多重度の推定値と一致していることがわかる。「序論」では一致していないが、補正法による推定値  $m = 2$  と  $F$  値最大の多重度  $m = 3$  の  $F$  値の差は 0.01 未満と微少であり、予測性能は概ね同等であると解釈できる。以上の結果より、推定された  $m$  重マルコフ情報源により、多様なパターンの論文構成を高精度に予測できることが示せた。

また、多重度の推定法として、本実験のような  $F$  値最大化による手法も考えられる。本実験では、新たに評価用のデータを追加して得られた多重度の推定値が提案手法と同等の結果を示したが、提案手法では、より少数のデータから予測の良い多重度を推定でき、予測効率が高いことがわかる。

## 2.7.2 システムの有効性評価

本節では被験者実験により、提案システムの利点である (1) 初心者でも容易に論文構成を構築できる、(2) 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るの 2 点について評価する。

本実験では、被験者を異なる条件の 3 つのグループに分けて論文を執筆させ、論文執筆後のアンケートと完成した論文の分析により、提案システムの有効性を評価した。グループは、提



表 2.8  $m$  重マルコフ情報源の予測性能の比較結果

適合率						
	m=0	m=1	m=2	m=3	m=4	m=5
序論	0.067	0.239	0.269	0.284	0.257	0.196
本論	0.042	0.322	0.399	0.415	0.428	0.409
結論	0.200	0.379	0.429	0.415	0.415	0.415

再現率						
	m=0	m=1	m=2	m=3	m=4	m=5
序論	1.000	0.973	0.908	0.827	0.714	0.611
本論	1.000	0.951	0.882	0.829	0.766	0.701
結論	1.000	0.973	0.959	0.946	0.946	0.946

F 値						
	m=0	m=1	m=2	m=3	m=4	m=5
序論	0.125	0.384	0.415	0.423	0.378	0.297
本論	0.080	0.486	0.549	0.553	0.549	0.516
結論	0.333	0.546	0.593	0.577	0.577	0.577

案システムを利用するグループ A と画一的なテンプレートを利用するグループ B, Microsoft Office Word により自由に論文執筆を行うグループ C とした。各グループの被験者は、情報システムを学ぶ大学 4 年生 5 名, 大学院修士生 2 名の計 7 名ずつとし、いずれも論文執筆の経験のない者とした。被験者には、「RFID を用いた観光支援システムの開発と評価」という論文の「序論」を、筆者らが用意した「システムの仕様書」、「関連論文」、「評価実験の計画と結果の資料」を参照して執筆させた。なお、グループ B に与えたテンプレートの構成は、「研究領域の現状」「研究領域の問題点」「先行研究の概要」「先行研究の問題点」「本研究の目的」「本研究の方法」「本研究の特徴」「本論の構成」とした。

論文執筆後に回答させたアンケートは、以下の 3 つの質問で構成され、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) で回答させた。

1. 初心者でも容易に論文構成を構築できたか。
2. 自身の意図に合った論文構成であったか。
3. 論文構成を意識して執筆できたか。

ただし、グループ B に対しては (1) の質問を除外した。

さらに、完成した論文を研究者に読んでもらい、以下の 2 点についてアンケート同様の 6 段階で評価してもらった。

1. 論文構成は学術論文として妥当な形式か。
2. 執筆された論文は論理的であるか。

ただし、グループ A と B の質問 (1) では、論文要素の系列のみを提示して論文構成を評価させた。また、論文構成のパターン数と平均論文作成時間、平均総文章量も計測した。以上の結果を表 2.9 に示す。なお、表 2.9 のアンケートと専門家評価では、評価値の平均と分散 (カッコ内) を示している。また、\*\*は分散分析の結果、有意水準 1% で有意差があったことを、\*は 5% で有意差があったことを表している。分散分析には Kruskal-Wallis 検定を用いた。ただし、アンケート (1) には、Wilcoxon-Mann-Whitney 検定を用いた。

さらに、分散分析の結果、有意差の認められた項目について、Scheffe の方法で多重比較を行ったところ、アンケートでは、質問 (2) のグループ A-B 間と A-C 間、質問 (3) のグループ A-C 間に有意水準 5% で有意差がみられた。また、専門家評価では、質問 (1) のグループ A-C 間と B-C 間、質問 (2) のグループ A-C 間に有意差がみられた。平均文章量は、グループ A-C 間と B-C 間に有意差がみられた。以上の結果から、提案システムの以下の利点が確認された。

1. 初心者でも容易に論文構成を構築できる。
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る。

また、表 2.9 より、論文構成自体の妥当性はシステムよりテンプレートが高得点だが、論文全体の論理性はシステムの方が高得点であるとわかる。専門家へのインタビューから、テンプ

表 2.9 被験者実験の結果

	グループ A	グループ B	グループ C
アンケート (1)**	5.14(0.83)		2.86(1.13)
アンケート (2)**	5.14(0.64)	3.57(1.18)	3.57(0.90)
アンケート (3)*	5.14(0.83)	4.00(0.93)	3.00(1.20)
専門家評価 (1)**	5.43(0.25)	6.00(0.00)	2.43(1.29)
専門家評価 (2)*	5.14(0.64)	3.71(1.39)	2.43(1.50)
平均作成時間 (分)	38(11)	40(23)	35(24)
平均総文章量 (文字)**	682(143)	754(227)	313(145)
パターン数	6	1	7

レート利用者の論文では同じ記述を繰り返すような冗長な文章が多いという意見を得ており、画一的なテンプレートを利用した場合、利用者は意図に合わない論文要素にも強制的に文章を挿入しなければならなかったり、前後の文脈を考慮せず各論文要素に独立に文章を執筆しがちになるため、冗長な文章が多くなったと考えられる。一方、本システムでは、利用者の意図に合わせて論文を構成でき、文章執筆と論文構成を同時に行う過程で文章や論文構成を逐次推敲できるため、冗長性の少ない論理的な文章を執筆できたと解釈できる。

また、表 2.9 より、本システムにより論文構成への意識が有意に向上し、初心者でも十分な量の文章を執筆できることもわかった。さらに、グループ A の被験者へのインタビューから、提案システムにより、論文の書き方を学習・理解できたといったポジティブな意見が多数得られた。

以上から、提案システムにより、利用者の意図する論文構成を、妥当な形式を保持しつつ容易に構築できることを示せた。

## 2.8 むすび

本章では、「論文構成」を情報理論における情報源からの出力符号系列とみなしたメタファとしてとらえ、論文構成の構築過程を定式化した。具体的には、論文構成を論文要素カテゴリーの系列データとし、それが  $m$  重マルコフ情報源に従うと仮定した。また、データ長の短い論文構

成データからより正確に多重度を推定するために、多重度の推定補正法を提案し、結果として「序論」 $\hat{m} = 2$ 、「本論」 $\hat{m} = 3$ 、「結論」 $\hat{m} = 2$ という推定値を得た。このことから、論文構成は画一的な形式に従うのではなく、直前の論文要素の系列に応じて多様な形式を取りうるという知見を得た。また、推定された  $m$  重マルコフ情報源により、多様なパターンの論文構成を高精度で予測できることを実験により確認した。さらに、 $m$  重マルコフ情報源に基づく論文構成構築支援システムを開発し、被験者実験により、提案システムが妥当かつ多様なパターンの論文構成の構築を支援でき、論文構成への意識と文章の論理性の向上にも寄与することを示した。なお、本研究では、提案システムの支援対象を情報システム論文としたが、本研究で提案したアプローチは分野が異なっても同様に利用できると思われる。

なお、今回は、最適なモデルを推定するために、ベイズ周辺尤度を最大化した。しかし、確率推論の効率を向上させるためには、周辺尤度ではなく、モデル平均予測分布を最大化させる必要がある。このためには、すべてのあり得るモデルについて、周辺尤度を計算しなければならず、計算アルゴリズムの改良が必要となる。これらを今後の課題としたい。

また、本研究では、全ての論文要素間の遷移関係をデータから推定するアプローチを採用した。一方で、特定の論文要素間について、「遷移すべきでない」といった事前知識がある場合には、この知識を制約とすることで、推定すべきパラメータ数を減少させることが可能といえる。このアプローチについても、今後の課題とする。

## 第3章

# Toulmin モデルのベイジアンネット ワーク表現を用いた論証構築支援シ ステム

### 3.1 はじめに

情報の効果的な伝達を目指すアカデミックライティングでは [1], 説得的な「論証」の構築が重要課題となる [4]. 論証とは「自らの主張を説得的に行うための言語行為 [33]」と定義される. 説得的な論証を構築するためには, 全ての文章間の因果関係を把握し, 客観的かつ反復的に論証を推敲する必要がある. しかし, 特に初心者にとって, このような作業は容易ではないことが知られている [34].

そのため, アカデミックライティングにおける論証の構築を支援するシステムがこれまでに多数開発されてきた [35]. このようなシステムの多くでは, Toulmin モデル [36] と呼ばれる論証スキーマに論証を当てはめることを支援する (例えば, [34, 37, 38, 39, 40]). Toulmin モデルとは, 論証を「主張」, 「データ」, 「論拠」, 「裏付け」, 「限定」, 「論駁」の 6 つの小単位に分類し, それぞれの関係を有向グラフとして形式化したものであり [41], 論証の基準として多様な分野で利用され, その妥当性が示されている (例えば [42, 43, 44, 45]).

Toulmin モデルに基づく論証構築支援システムとして, 館野ら [37] は, 文章中の Toulmin

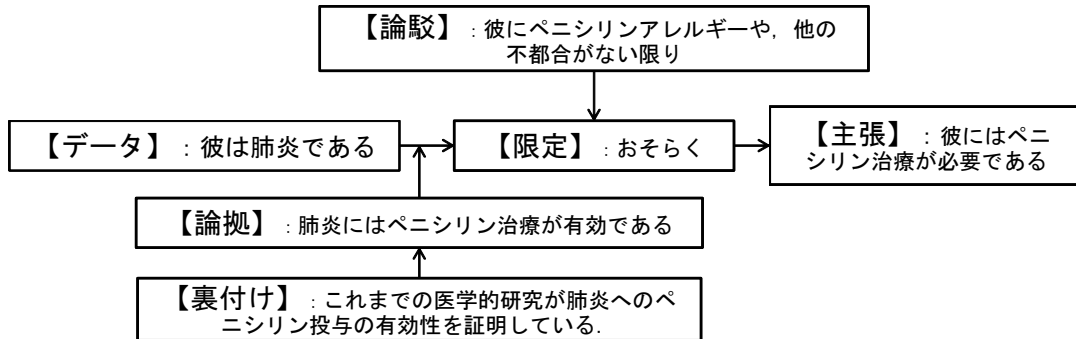


図 3.1 Toulmin モデル

モデルの構成要素に対応する箇所をハイライトでき、更にそれを他のユーザと共有して相互評価できるシステム「カラコメ！」を開発している。Brudvikら [38]、Azilawatiら [39] は、文章に Toulmin モデルの各構成要素に対応するタグ付けを行うことで、それらを階層化して表示するシステム「Voice of Reason」を開発している。また、Chrisら [40] は、論証を Toulmin モデルに当てはめ、ダイアグラムとして可視化できるシステム「Araucaria」を開発している。

これらのシステムでは、論証を Toulmin モデルに表層的に当てはめ可視化する支援に留まっている。しかし、論証の主目的である「主張」の正当化のためには、論証の Toulmin モデルへの当てはまりの良さよりも、文章間の因果の強さ、すなわち「論証の強さ」を重視した論証の構築が重要である。一般の論証では、複数の複数の「論拠」や「データ」、「裏付け」から主張を導くことが多く、論証の構成が複雑化すると、以下の問題が生じると考えられる。1. 「論証の強さ」を全ての文章間について評価することが困難である。2. 論証中の各文章がどの程度正当化できているかの推定が難しい。3. 「主張」の正当化に対して各文章がどのように影響しているかを把握することが困難である。

以上の問題を解決するために、本論では、Toulmin モデルのベイジアンネットワーク表現を用いた論証構築支援システムを開発する。ここでは、論証中の文章を確率変数、文章間の因果を条件付確率で表すことで、Toulmin モデルに当てはめて構築された論証をベイジアンネットワークとして定式化する。確率は主観確率とし、確率値に対応付けられた 6 段階のカテゴリを用いてユーザが評価する。本システムでは、このベイジアンネットワーク表現を用いて、1. 論証の強さ、2. 文章の正当性、3. 主張への影響度、という論証の特性を表す 3 つの指標（論証特性指標

と呼ぶ)を算出し, それらの推定結果に応じて論証改訂のためのアドバイスをフィードバックする. 本システムの特徴は以下の通りである.

1. ベイジアンネットワークの特殊形のひとつである Noisy or モデル [46] を利用することで, ユーザが評価する主観確率の個数を減少させた上で, 全ての論証の強さを推定できる.
2. 論証中の各文章がどの程度正当化されているかを数値化でき, 更に, この値が低い場合, その文章を正当化するための論証改訂方法をアドバイスできる.
3. Sensitivity Analysis[47],[48] を利用した論証の分析により, 「主張」の正当化に対する各文章の影響を定式化できる.

更に, 本論では, 被験者実験により本システムの有効性を示す.

## 3.2 Toulmin モデル

本論では, 論証の規範モデルとして Toulmin モデルを採用する.

Toulmin モデルとは, 論証を「主張 (Claim)」、「データ (Data)」、「論拠 (Warrant)」、「裏付け (Backing)」、「限定 (Qualifier)」、「論駁 (Rebuttal)」の 6 つの小単位に分類し, それぞれの関係を有向グラフとして形式化したものであり [41], 一般に図 3.1 のようなダイアグラムで表現される. ここで, 「主張」とは, 正当化されるべき命題であり, 「データ」とは, 結論を基礎づけるために参照される事実である. 「論拠」とは, データと結論を架橋し, その移行を権威づける一般的な仮言的弁明であり, 「裏付け」は, 論拠に権威と信用を与える言明である. 「限定」は, データから結論への移行を正当化する強さを表し, 「論駁」は, 論拠の一般的正当性が退けられなければならない状況を意味する [49].

Toulmin モデルの構成要素には論証中の文章が対応付けられ, 例えば, 図 3.1 の Toulmin モデルは以下の論証 [50] を表現している.

彼にはペニシリン治療が必要である (主張). なぜなら, 彼は肺炎であり (データ), 肺炎にはペニシリン治療が有効である (論拠) からだ. これまでの医学的研究が肺炎へのペニシリン治療の有効性を証明している (裏付け). 従って, 彼がペニシリンアレルギーであったり, 他の不都合な状況にない限り (論駁), おそらく (限定) 彼にはペニシリン治療

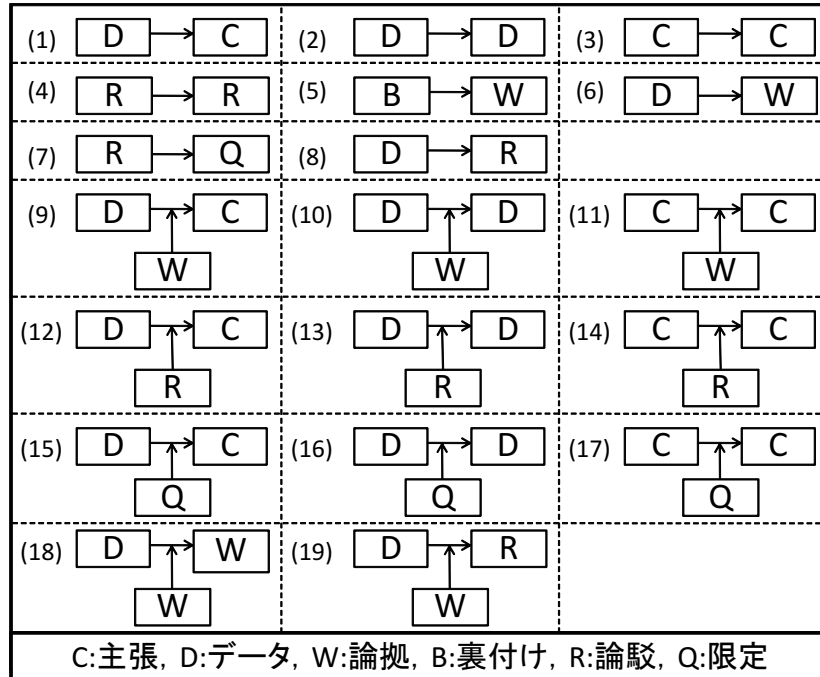


図 3.2 Toulmin モデルの構成規則

を施すべきであろう。

Toulmin モデルは、日常的な文脈の論理に幅広く適用可能であり、法廷での論証過程の分析 [42] やクリティカルリーディング [43]、合意形成 [44]、意思決定 [45] など多様な分野で利用され、その妥当性が示されている。アカデミックライティングにおける論証支援においても、論証構築のための具体的な基準として利用されている [51]。そこで、本論でも Toulmin モデルを論証の規範として利用し、Toulmin モデルへの論証の当てはめを支援する。このために、ここでは Toulmin モデルの構造を定義する。

Toulmin モデル  $T$  は、図 1 に示した 6 つの構成要素を元とする集合  $N$  と構成要素間の支持関係を表す有向辺集合  $E$  により  $T = \langle N, E \rangle$  で定義される。ここで構成要素間への有向辺の付与は任意には行えず、一定の規則（以下、構成規則）に従う必要がある。本論では、この構成規則を定義するために、Toulmin モデルの構成規則に言及した先行研究 [40, 52, 53, 54] から共通する規則を全て抽出した。



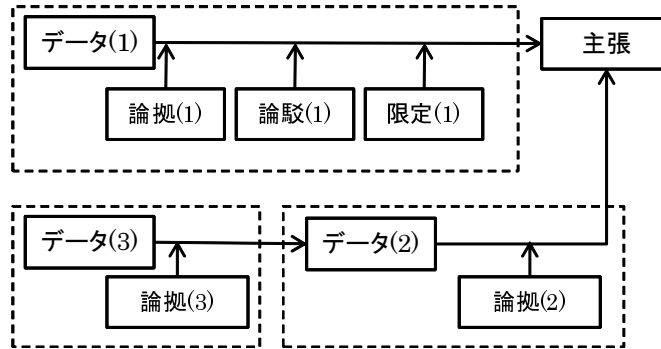


図 3.3 構成規則に従う Toulmin モデルの例

抽出した構成規則を図 3.2 に示す．図 3.2 の構成規則の組み合わせにより，例えば，図 3.3 のような Toulmin モデルを表現できる．図 3.3 の論証は，「主張」に対して，二つの「データ」を提示し，それらの間の論証を支持するために「論拠」や「限定」を追加している．更に，「データ(2)」を補足するために「データ(3)」と「論拠(3)」が付与されている．この論証は，図 3.2 の構成規則 (1),(2),(9),(10),(12),(15) により表現できる．

この例のように，Toulmin モデルでは，「主張」に対して「データ」や「論拠」などを複数提示することで複雑な構造の論証も表現できる．本論では，図 3.3 のように「データ」が複数存在するとき，図中の点線で囲んだような局所的な論証を「局所論証」と呼び，以下のように定義する．ある「主張」(または「データ」) $S_i$  に対し  $R_i$  個の「データ」 $S_d^i = \{S_{d_1}^i, \dots, S_{d_r}^i, \dots, S_{d_{R_i}}^i\}$  が提示されるとき， $S_{d_r}^i$  から  $S_i$  への有向辺を  $S_{d_r}^i \rightarrow S_i$  と表す．更に， $S_{d_r}^i \rightarrow S_i$  に接続する有向辺を持つ構成要素集合を  $\Psi_{id_r}$  と表記し， $S_{d_r}^i \cup \Psi_{id_r}$  を  $A_{ir}$  と表す．「局所論証」は構成要素集合  $A_{ir}$  から  $S_i$  への論証を表す．

### 3.3 Toulmin モデルのベイジアンネットワーク表現

本論では，Toulmin モデルに当てはめて構築された論証をベイジアンネットワークとして表現する．

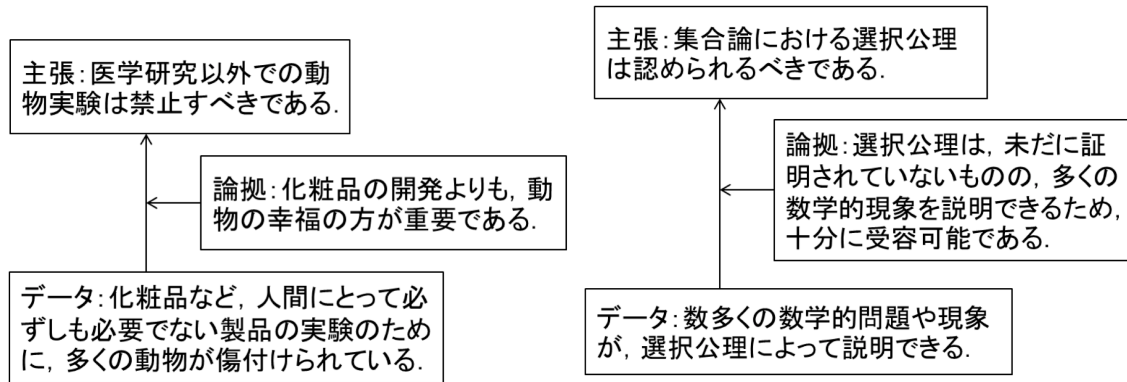


図 3.4 文献 [55] で例示された, 動物実験の禁止を主張する論証の一部

図 3.5 文献 [56] で例示された, 集合論における選択公理を認めるべきであると主張する論証の一部

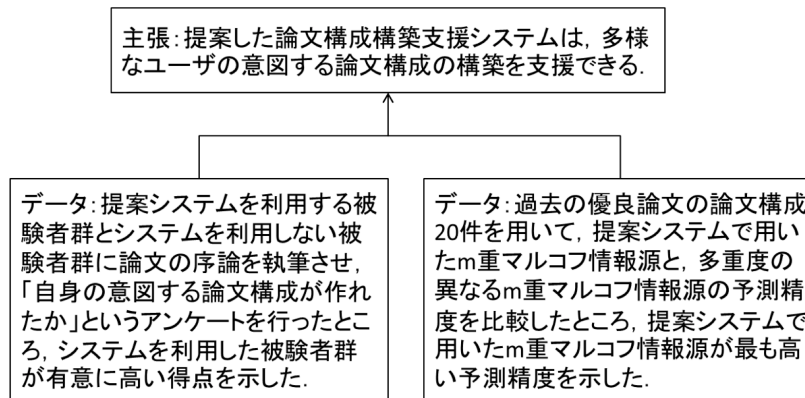


図 3.6 本博士論文第二章中の論証の一部

### 3.3.1 Toulmin モデルのベイジアンネットワーク表現

一般に, 論証内の文章の正しさは真偽の 2 値で断定できない場合が多く, 「文章の正当性が高いかどうか」といった蓋然的な基準で評価される [57]. 蓋然的な文章を含む論証の例を図 3.4, 図 3.5, 図 3.6 に示す. これらの例では, 「データ」や「論拠」の内容, あるいは, 「データ」から「主張」に至る推論過程に蓋然性が含まれており, 結果として導出された「主張」も蓋然的な文

章となっている。

このような蓋然的な文章を含む論証を定式化するために、本論では、論証中の文章を確率変数  $S_i (i = 1, \dots, N)$  とみなし、文章の正当性を主観確率  $p(S_i = k)$  として表す。ここで、 $k \in \{0, \dots, r\}$  (本論では  $r = 1$ ) は確率変数  $S_i$  の状態値を表し、真偽 (それぞれ 1, 0) の 2 値を取るものとする。更に、文章集合  $\Pi_{S_i}$  が与えられたときの文章  $S_i$  の正当性は条件付確率  $p(S_i = k | \Pi_{S_i} = j)$  で表す。ここで、 $\Pi_{S_i} = j (j = 1, \dots, q_i)$  は  $\Pi_{S_i}$  が  $j$  番目のパターンを取ることを表す。主観確率は、不確実性を含む主観的な信念や信頼の測度として知られており [58]、正当性のような蓋然かつ主観的な事象の数量化に適した測度であると考えられる。なお、論証の定式化のアプローチとして、ファジー論理の利用も考えられる。しかし、ファジー論理は加法性を満たさないため、どのようなデータに対しても当てはまり、計算も容易という利点があるが、推論精度が悪いという欠点があるため、本論では確率的アプローチを採用する。

一方で、このような確率変数間の依存関係を簡潔に表現するモデルとして、ベイジアンネットワークが知られている。ベイジアンネットワークとは、次の定義を満たすモデルである。

$N$  個の変数集合  $X = \{x_1, x_2, \dots, x_N\}$  を持つベイジアンネットワークは、以下の  $\langle \mathcal{G}, p \rangle$  で表現される。

- $\mathcal{G} = \langle V, E \rangle$  は、 $X$  に対応する頂点集合  $V$  と、その変数間の確率的依存関係を表現する有向辺集合  $E$  で構成される非循環有向グラフであり、ネットワーク構造と呼ばれる。
- $p$  は、 $X$  についての同時確率分布を表し、各変数  $x_i (i = 1, \dots, N)$  に対応する条件付確率分布  $p(x_i | \Pi_i, \mathcal{G})$  の積として、 $p(x_1, x_2, \dots, x_N) = \prod_{i=1}^N p(x_i | \Pi_i, \mathcal{G})$  で表される。ただし、 $\Pi_i$  は変数  $x_i$  の親変数集合を示す。

ベイズ統計学では、確率の解釈として、ユーザの知識と客観データを統合的に取り扱える主観確率を用いる。近年のベイジアンネットワークの研究では、データからネットワーク構造  $\mathcal{G}$  やパラメータ  $p$  を推定する統計的学習アプローチが盛んであるが (例えば [59])、そもそもは専門家の因果知識と主観確率を計測し、エキスパートシステムを構築することから始まっている [60],[61]。

ベイジアンネットワークは、1. 確率変数間の因果関係を表すモデルと解釈でき、2. 主観確率を扱うことができ、3. 任意の変数の確率を求める確率推論を効率よく実行できるといった特徴を

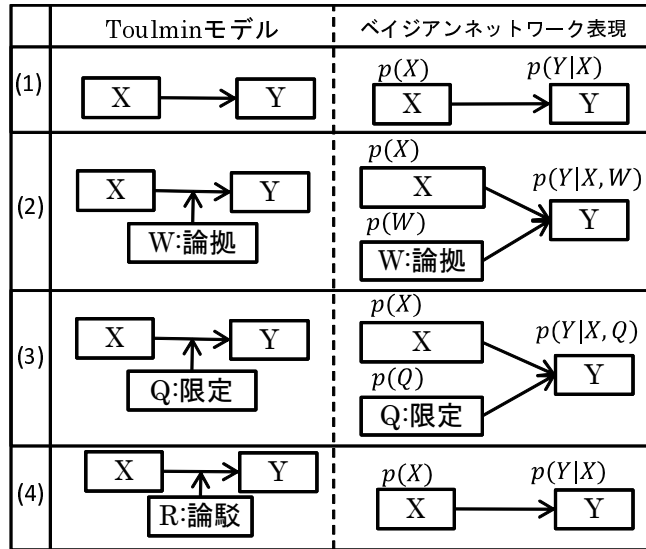


図 3.7 Toulmin モデルのベイジアンネットワーク表現

持っており [62]，本論の目的に適した手法であると考えられる。

そこで，本論では，Toulmin モデルに当てはめて構築された論証をベイジアンネットワークにより定式化する．具体的には，ユーザが構築した Toulmin モデルをもとにベイジアンネットワークの構造  $\mathcal{G}$  を決定し，条件付確率は，主観確率値を対応付けた言葉を用いてユーザに評価させる．言葉を主観確率に変換する手法としては，[63] や [64] が知られている．増田ら [64] は，「十分～である」，「おそらく～である」，「どちらかという～である」，「どちらかという～でない」，「あまり～でない」，「全く～でない」などの言葉を主観確率に変換する手法を提案している．これは各ユーザの主観確率がベータ分布として求められる手法である．ここではこの結果から，言葉のベータ分布の期待値がほぼ等分されることを用いて，次のように主観確率を求めた．6. 十分妥当である (0.95)，5. おそらく妥当だ (0.77)，4. どちらかという～妥当だ (0.59)，3. どちらかという～妥当ではない (0.41)，2. あまり妥当ではない (0.23)，1. 全く妥当でない (0.05)．以降では，これらの言葉を評価カテゴリと呼ぶ．

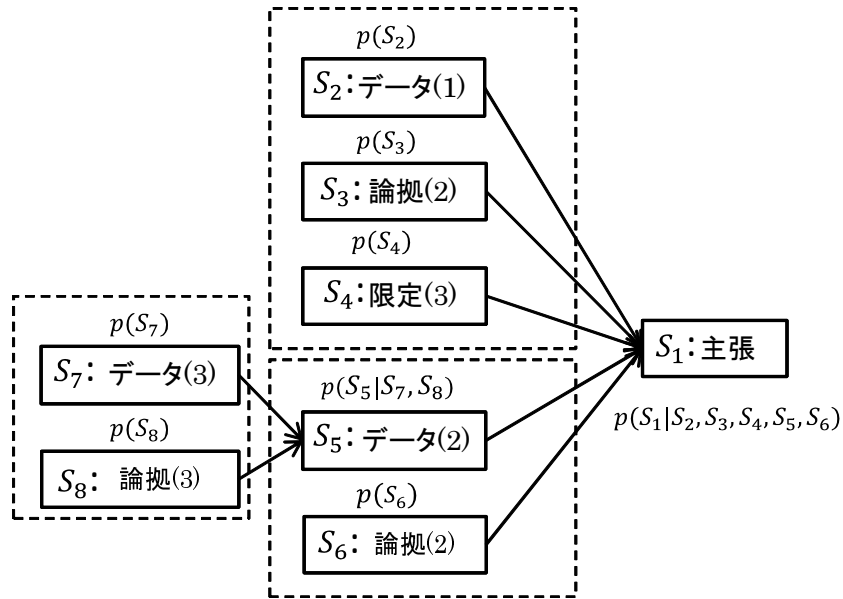


図 3.8 ベイジアンネットワーク表現の例

### 3.3.2 Noisy or モデル

本論では、ベイジアンネットワークの構造を、Toulmin モデルにおける各構成要素間の論理的关系を反映するように、ユーザが構築した Toulmin モデルから決定する。

まず、「論拠」 $W$  が任意の局所論証  $X \rightarrow Y$  に付与された、図 3.2 中 (9),(10),(11),(18),(19) のような Toulmin モデルを考える。これらの Toulmin モデルは、文章  $Y$  の正当性が、文章  $X$  と「論拠」 $W$  に依存して決まることを意味する [53]。更に、文章  $X$  と「論拠」 $W$  は、一方が他方の正当性を高める関係になく独立である。ここで、例として、図 1 の Toulmin モデルを考える。図 1 では、「主張：彼にはペニシリン治療が必要である」が  $Y$ 、「データ：彼は肺炎である」が  $X$ 、「論拠：肺炎にはペニシリン治療が有効である」が  $W$  に対応する。この例から、「主張」の正当性が「データ」と「論拠」の両方に依存して決まること、「データ」と「論拠」が独立な関係であることがわかる。以上より、 $X, Y, W$  を確率変数とみなすと同時確率分布は  $p(X, Y, W) = p(Y|X, W)p(X)p(W)$  と書け、図 3.7(2) 右図のようなベイジアンネットワークで表現できる。

また，図 3.2 中 (15),(16),(17) のように，局所論証  $X \rightarrow Y$  に「限定」 $Q$  が付与された Toulmin モデルについても，同様の説明が可能である．

一方，図 3.2 中 (12),(13),(14) のように，局所論証  $X \rightarrow Y$  に「論駁」 $R$  が付与された Toulmin モデルにおいて，「論駁」 $R$  は文章  $Y$  の正当性向上に直接には寄与しないため [52]，ベイジアンネットワークには反映しない．

以上より，Toulmin モデルからベイジアンネットワークを構築するための規則を図 3.7 にまとめた．例として，図 3.3 の Toulmin モデルのベイジアンネットワーク表現を図 3.8 に示す．

この変換では，論証が複雑化し，ある文章を支持する「データ」「論拠」「限定」が増加すると，ベイジアンネットワークのパラメータ数が指数的に増大し，ユーザが全ての条件付確率を評価することが困難になる．しかし，説明変数のどれか一つが成立したときに目的変数が成立するような場合には，ベイジアンネットワークの特殊形であり，計算量を線形オーダーに抑えることができる Noisy or モデルが利用できる．Noisy or モデルでは，各説明変数の不成立が変数間で独立に生じると仮定できるとき [46]，説明変数集合  $\Pi_{S_i}$  の任意のパターン  $j$  に対する条件付確率を次式で計算できる．

$$p(S_i = 1 | \Pi_{S_i} = j) = 1 - \prod_{S_l \in I_j} p(S_l = 0 | S_l = 1) \quad (3.1)$$

ここで， $I_j$  は  $\Pi_{S_i}$  の  $j$  番目のパターンにおいて 1 を取る変数の集合を表す．

他方，図 3.3 のように，ある文章に対して複数の局所論証が提示される場合，一つでも局所論証が成立すればその文章は正当化されると考えられる．したがって，その文章を目的変数，その文章を支持する各局所論証を説明変数とみなすと，Noisy or モデルの仮定が満たされる．そこで，本論では，Noisy or モデルを以下のように用いる．

任意の「主張」(または「データ」) $S_i$  に対し  $R_i$  個の局所論証  $A_i = \{A_{i1}, \dots, A_{ir}, \dots, A_{iR_i}\}$  があるとき， $A_{ir}$  を所与とする条件付確率  $p(S_i = 1 | A_{ir})$  を  $r = 1, \dots, R_i$  について，それぞれ独立にユーザに評価させ，最終的に条件付確率  $P(S_i = 1 | \Pi_{S_i})$  を式 (3.1) で決定する．これにより，ユーザに評価させる条件付確率の数を， $2^{\sum_{r=1}^{R_i} |A_{ir}|}$  から  $\sum_{r=1}^{R_i} 2^{|A_{ir}|}$  に軽減できる．

例えば，図 3.8 の場合，「 $S_1$ :主張」を支持する 2 つの局所論証を所与とした条件付確率  $p(S_1 = 1 | S_2, S_3, S_4)$  と  $p(S_1 = 1 | S_5, S_6)$  をそれぞれユーザに評価させ，条件付確率  $p(S_1 = 1 | \Pi_{S_1})$  を式 (3.1) により決定する．これにより， $p(S_1 = 1 | \Pi_{S_1})$  に関し，ユーザが評価する確率

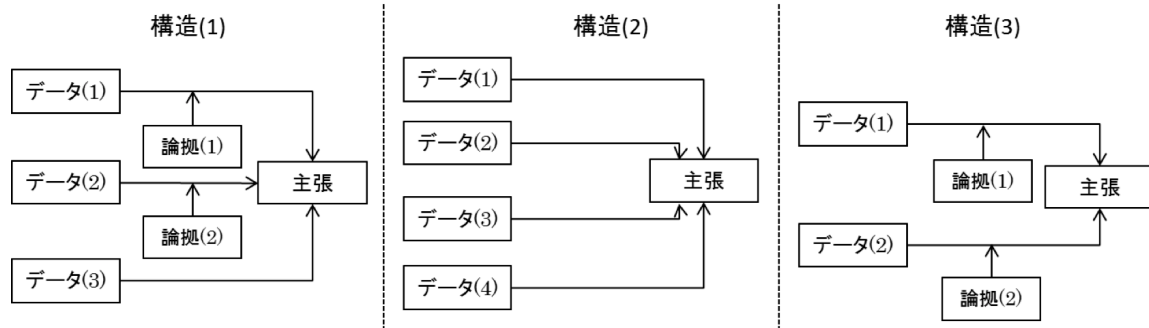


図 3.9 実験に利用した論証構造

値の数は、 $2^5 = 32$  から  $2^3 + 2^2 = 12$  に軽減される。

### 3.3.3 論証への Noisy or モデル適用の妥当性評価

本節では、論証への Noisy or モデルの適用が妥当であるかを被験者実験により評価する。ここでは、まず、図 3.9 の Toulmin モデル構造を持つ 3 種類の論証を筆者らが用意し、理系大学院生 5 名に配布した。構造 (1) と (2) は文献 [33](pp.171-174)、構造 (3) は [65](pp.197-205) で悪文として例示された論証の一部を使用した。被験者には、これらの論証の Toulmin モデルと原文、さらに、その論証のベイジアンネットワーク表現に必要な条件付き確率表を紙面に印刷して配布した。条件付き確率表は、全ての条件付き確率の入力欄がある表と、Noisy or モデルを適用して条件付き確率の入力数を削減した表の二種類を配布した。被験者には、これらの条件付き確率値を、3.3.1 で挙げた 6 段階の評価カテゴリで入力させた。ここでは、この実験で収集したデータを用いて、被験者が入力した評価値と、Noisy or モデルを適用して推定した値とのピアソンの累積相関係数を求めた。

実験結果を表 3.2 に示す。表 3.2 より、全ての構造の論証で、被験者による評価値と Noisy or モデルを仮定して推定した値との間に優位な正の相関があったことがわかる。このことから、Noisy or モデルを利用した論証の近似は、被験者が評価値を入力した場合と同傾向となることが示せた。

表 3.1 Noisy or モデルの妥当性評価結果

	構造 (1)	構造 (2)	構造 (3)
被験者 1	.94 ( $p < .01$ )	.93 ( $p < .01$ )	.95 ( $p < .01$ )
被験者 2	.99 ( $p < .01$ )	.98 ( $p < .01$ )	.91 ( $p < .01$ )
被験者 3	.75 ( $p < .01$ )	.85 ( $p < .01$ )	.81 ( $p < .01$ )
被験者 4	.87 ( $p < .01$ )	.56 ( $p < .05$ )	.82 ( $p < .01$ )
被験者 5	.80 ( $p < .01$ )	.72 ( $p < .01$ )	.80 ( $p < .01$ )

### 3.4 論証特性指標

ここでは, Toulmin モデルのベイジアンネットワーク表現を用いて, 1. 論証の強さ, 2. 文章の正当性, 3. 主張への影響度, という 3 つの論証特性指標を定義する.

#### 3.4.1 論証の強さ

「論証の強さ」は文章間の因果の強さを表すため, 変数間の依存関係の強さを表す条件付相互情報量を用いて定式化する. ここで, 2 変数  $S_i, S_j$  間の依存関係を  $S_i \rightarrow S_j$  とすると, 論証の強さは次式で定義される.

$$I(S_i, S_j | \Psi_{S_{ij}}) = \sum_{\Psi_{S_{ij}}} p(\Psi_{S_{ij}}) \sum_{k=0}^r \sum_{k'=0}^r p(S_j = k | S_i = k', \Psi_{S_{ij}}) p(S_i = k') \log \frac{p(S_j = k | S_i = k', \Psi_{S_{ij}})}{p(S_j = k | \Psi_{S_{ij}})} \quad (3.2)$$

式 (3.2) の計算には, 変数  $S_i$  の親ノード集合  $\Pi_{S_i}$  における任意の部分集合  $\beta \subseteq \Pi_{S_i}$  を所与とする条件付確率  $P(S_i = k | \beta)$  が必要である. これは条件付確率  $P(S_i = k | \Pi_{S_i})$  から  $\beta$  以外の全ての変数  $\Pi_{S_i} \setminus \beta$  を周辺化し,  $\sum_{\Pi_{S_i} \setminus \beta} p(S_i = k | \Pi_{S_i}) p(\Pi_{S_i})$  により得られる.

条件付相互情報量は, 0 を最小値とし変数間の依存関係が強いほど大きい値を取るため,  $I(S_i, S_j | \Psi_{S_{ij}}) < \epsilon_0$  (本論では  $\epsilon_0 = 0.05$ ) のとき論証が弱いと判断する.



表 3.2 実験に利用した論証

	内容	構成要素数	総文字数	ネットワークの平均深さ	ネットワークの平均幅
論証 1	現行の中学校の歴史教科書が教室での利用に不適切であることを主張する論証文	20	1238	3.16(0.75)	2.67(1.92)
論証 2	高等動物を用いた現行の動物実験が倫理的に不当であることを主張する論証文	16	1687	3.80(1.10)	2.80(1.64)

### 3.4.2 文章の正当性

「文章の正当性」は、ある文章が、論証によりどの程度正当化されているかを表す指標である。ここでは、これを文章  $S_i$  が真である周辺確率として式 (3.3) で定義する。

$$p(S_i = 1) = \sum_{j=1}^{q_i} p(S_i = 1 | \mathbf{\Pi}_{S_i} = j) p(\mathbf{\Pi}_{S_i} = j) \quad (3.3)$$

この計算は Variable Elimination アルゴリズム [66] により、根ノードから順に変数消去することで厳密かつ容易に実行できる。

この値が低い場合、すなわち、 $p(S_i) < \epsilon_1$  (本論では  $\epsilon_1 = 0.7$ ) のとき、その文章は十分に正当化されていないと解釈する。

### 3.4.3 主張への影響度

「主張への影響度」は、ある文章  $S_i$  が「主張」 $S_c$  の正当化にどの程度影響を与えるかを表す指標である。より具体的には、文章  $S_i$  の正当性が向上するように論証を改訂したとき、どの程度「主張」の正当性が向上するかを表す。ベイジアンネットワークにおいてこのような分析を行う手法として Sensitivity Analysis [47][48] が有用である。ここで、ある変数  $S_i$  が真である確率  $p(S_i = 1)$  が変化したとき、「主張」 $S_c$  が真である確率  $p(S_c = 1)$  がどのように変化するかを調べたいとする。Sensitivity Analysis では、 $p(S_c = 1)$  を  $p(S_i = 1)$  の関数 (Sensitivity function と呼ばれる) として以下のように表現する。

$$p(S_c = 1) = \mu_i p(S_i = 1) + \nu_i \begin{cases} \mu_i = \frac{p^0(S_c=1, S_i=1)}{p^0(S_i=1)} - \frac{p^0(S_c=1, S_i=0)}{p^0(S_i=0)} \\ \nu_i = \frac{p^0(S_c=1, S_i=0)}{p^0(S_i=0)} \end{cases} \quad (3.4)$$

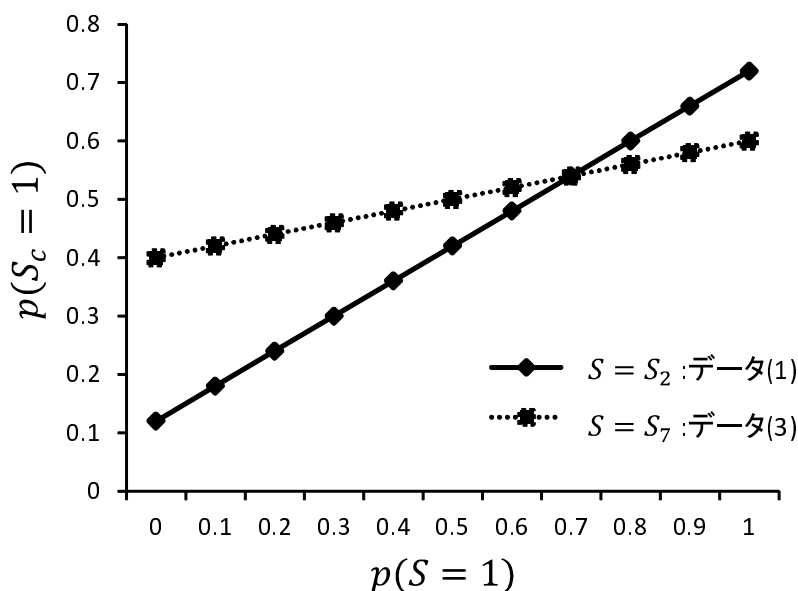


図 3.10 Sensitivity function

ここで、 $p^0$  は  $p(S_i = 1)$  の値を変化させる前の確率分布を表す。

例として、図 3.8 のベイジアンネットワークにおける、「 $S_2$ :データ (1)」及び「 $S_5$ :データ (3)」が真である確率  $p(S_2 = 1)$  と  $p(S_5 = 1)$  を変化させた場合の Sensitivity function を図 3.10 に示す。図 3.10 より、「 $S_2$ :データ (1)」の正当性  $p(S_2 = 1)$  が向上するように論証を改訂した方が、より効率よく「主張」の正当性  $p(S_1 = 1)$  を向上できることがわかる。

このように Sensitivity function を比較することで、論証中の各文章の正当性向上が「主張」の正当化にどの程度影響するかを判断できる。

#### 3.4.4 論証特性指標の妥当性評価

本節では、本論で提案した 3 つの論証特性指標の妥当性を評価する。ここでは、専門家による「論証の強さ」「文章の正当性」「主張への影響度」の評点と、本論で提案した各論証特性指標の推定値との相関を実験により算出する。

実験手順は以下の通りである。はじめに、筆者らが用意した 2 つの論証を専門家に配布した。ここでは、文献 [33](pp.171-174)、[65](pp.197-205) で悪文として例示された論証を用いた。

表 3.3 論証特性指標の妥当性評価結果

	論証の強さ	文章の正当性	主張への影響度
相関係数	.87	.79	.67
p	< .01	< .01	< .01

これらの論証の内容と、論証を Toulmin モデルに当てはめたときの構成要素数、総文字数、ネットワークの平均深さと平均幅を表 3.2 にまとめた。ここで、ネットワークの平均幅は、Toulmin モデルのベイジアンネットワーク表現における、各ノードの親ノード数の平均を表し、ネットワークの平均深さは、根ノードから各葉ノードまでの各パスに含まれるノード数の平均を表す。それぞれ括弧内は分散を表している。なお、本実験で用いた論証には、「弱い論証」、「正当性の低い文章」が各一つ以上出現していた。専門家には、原文とその論証の Toulmin モデルを記載した紙資料を配布した。

専門家には、まず、2つの論証中の全ての「論証の強さ」と「文章の正当性」、「主張への影響度」を3段階（3：高，2：中，1：低）で評価させ、Toulmin モデルが記載された紙面上に評点を記入させた。次に、これらの Toulmin モデルのベイジアンネットワーク表現に必要な条件付確率値を、筆者らが作成した条件付確率表に、3.3.1 で挙げた6段階の評価カテゴリで入力させた。

ここでは、入力された評価カテゴリを条件付確率値に変換し、これを用いて算出した各論証特性指標の推定値と、専門家による評点とのスピアマンの順位相関係数を計算した。ただし、「主張への影響度」は関数で定義されるため、専門家による評点との相関係数を直接求めることができない。「主張」への影響は、式 (3.4) で表される Sensitivity function の傾きが大きいほど強いと解釈できるため、本実験ではこの値を用いて相関係数を算出した。結果を表 3.3 に示す。

表 3.3 より、全ての論証特性において、有意水準 1% で専門家の評点と論証特性指標の推定値に正の相関が認められたことがわかる。このことから、提案した各論証特性指標の推定値は専門家による評点と類似した傾向となることが示せた。

### 3.4.5 ユーザによる主観評価の妥当性評価

本システムでは、論証特性指標の推定値やフィードバックの内容が、ユーザによる条件付確率の主観評価の結果に依存して決まるため、これらの主観評価が適切に行われていることが重要となる。本論では、ユーザは、あらかじめ論証と文章の妥当性評価に関する指導や説明を受け、適切に主観評価を行うことができると仮定する。

ここでは、まず、これらの指導を受けたユーザが、主観評価を適切に行うことができるかを評価する。被験者として論文執筆経験のない理系大学生 10 名ずつを集めた。被験者には、実験内容を説明した後、Toulmin モデルと各論証特性指標に関する説明、および、文章の正当性に対応する条件付確率の評価例を例文を用いて 20 分程度で解説した。更に理解を深めさせるために簡単な例題を用いた演習問題と解説を 10 分程度実施した。その後、被験者には、原文とその論証の Toulmin モデルを記載した紙資料、更に、Toulmin モデルのベイジアンネットワーク表現に必要なすべての条件付確率表について、確率値の箇所を空欄にしたものを記載した紙資料を配布した。被験者には、これらを用いて、各条件付き確率表の空欄に 6 段階の評価カテゴリを記入させた。

これらのデータを用いて、ここではまず、3.4.4 の実験で専門家が入力した条件付確率に対応する評価カテゴリと、被験者が入力した評価カテゴリについて、スピアマンの順位相関係数を算出した。結果として、論証 1 では  $r = 0.82 (< .01)$ 、論証 2 では  $r = 0.71 (< .01)$  となり、有意水準 1% で、被験者と専門家の主観評価結果に正の相関が認められた。このことから、本実験の被験者が、条件付確率を適切に評価できたことがわかる。

次に、被験者による条件付確率の主観評価値を用いて算出される各論証特性指標値の妥当性を確認する。ここでは、被験者と専門家の主観評価結果をそれぞれ用いて各論証特性指標の値を算出し、これらの相関を求めた。ここではピアソンの累積相関係数を用いた。また、3.4.4 の実験同様、「主張への影響度」に関する相関を求めるために Sensitivity function の傾きを利用した。結果を表 3.4 に示す。

表 3.4 より、全ての論証特性指標について、専門家の主観評価結果を用いたときの推定値と、被験者の主観評価結果を用いたときの推定値に、有意な正の相関があったことが確認できる。このことから、本実験で被験者にフィードバックされた各論証特性指標値は、専門家の主観評価結

表 3.4 被験者へのフィードバックの妥当性評価

		論証の強さ	文章の正当性	主張への影響度
論証 1	相関係数	.43	.65	.89
	p	< .01	< .01	< .01
論証 2	相関係数	.67	.76	.27
	p	< .01	< .01	< .05

果から推定した値と同傾向であったことが示せた。

ただし、表 3.4 より、論証 2 の「主張への影響度」は、他の指標と比べ、相関係数の値が小さいことがわかる。このことは、論証 2 が弱い論証を繰り返して主張を導いている悪文であり、ネットワークが深く、主観評価値の設定が難しいことに原因していると考えられる。

「主張への影響度」は、ネットワークが深い場合、個別の局所論証に対する主観評価値の誤差が累積し、推定値の信頼性が低下することがある。論証 1 に比べ、論証 2 はネットワークが深く、かつ、専門家と被験者間の主観評価値の誤差が若干大きかったことから、論証 2 の「主張への影響度」では相関係数が小さく算出されたと考えられる。

更に、一般的な論証でどのようなネットワークが構築されるかを調べるために、鈴木ら [51] の実験で用いられたテーマ「最寄駅から大学までの最も良い経路について論じよ」で、7 人の被験者にレポートを執筆させた。完成した論証から Toulmin モデルを構築したところ、平均ノード数は 6.29(1.89)、ネットワークの平均深さは 2.22(0.37)、平均幅は 2.61(1.01) であった。更に、2012 年の電子情報通信学会論文誌からランダムに選択した 5 件の情報システム論文の序論の論証を分析し、Toulmin モデルを構築したところ、平均ノード数は 8.80(2.48)、ネットワークの平均深さは 2.33(0.34)、平均幅は 3.40(0.55) であった。

以上より、論証 2 のように弱い論証が繰り返される悪文の場合には、初心者の構成する「主張への影響度」の信頼性は低くなる傾向があるが、それでも有意水準 5% で専門家の判断とも相関があり有用である。文章を推敲して修正することにより、「主張への影響度」も専門家の判断に近づいていくと考えられる。

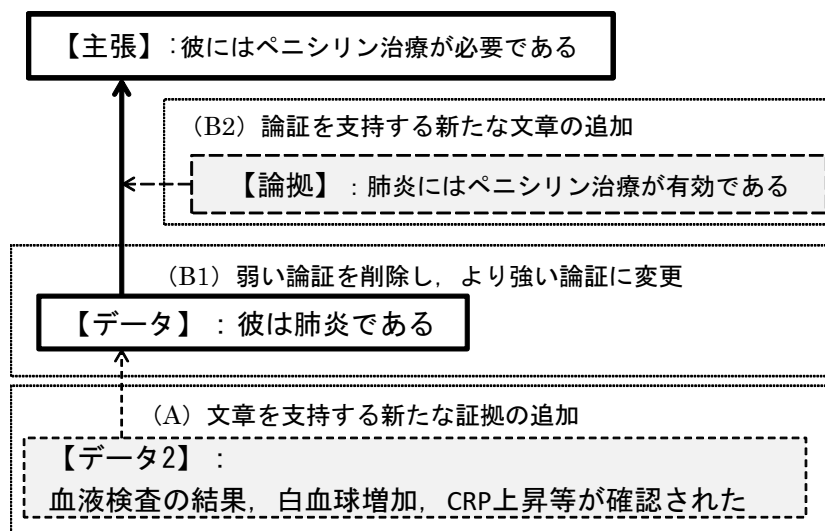


図 3.11 論証改訂方法のパターン

## 3.5 論証構築支援システム

### 3.5.1 論証特性指標に基づく論証改訂方法のフィードバック

ここでは、論証特性指標の推定結果に応じて論証改訂のためのアドバイスをフィードバックする手法を提案する。

論証中の任意の文章を正当化するためには、以下の論証改訂方法のパターンが考えられる。

- (A) その文章を支持する証拠を新たに追加する。
- (B) その文章を支持する既存の局所論証を次のいずれかの方法で改訂する。
  - (B1) その論証を削除し、より強い論証に変更する。
  - (B2) その論証を支持する文章を新たに追加する。

これらについての具体例を図 3.11 に示す。図 3.11 において、文章「彼は肺炎である (データ)」の正当性を高めたい場合には、(A) この文章を支持する証拠 (この例では「血液検査の結果、白血球増加および CRP 上昇が確認された (データ)」) を追加すればよい。一方、文章「彼にはペニシリン治療が必要である (主張)」の正当性を高めたい場合には、以下の改訂方法が考えられる。(A) この文章を支持する証拠を新たに追加する。(B) 既存の論証「彼は肺炎である

表 3.5 アドバイス文の生成規則

論証特性指標の推定結果	アドバイス文
文章の正当性が低い $p(S_i = 1) < \epsilon_1$	この文章は正当性が低いようです。 この文章を正当化すると、 $\Delta p(S_c = 1)$ 程度の「主張」の正当性向上が見込めます。 この文章を正当化するためには、この文章の正当性を $\epsilon_1 - p(S_i = 1)$ 以上高める必要があります。 以下のような方法で、この文章の正当性を高めましょう。 ・この文章を支持する以下のような文章を追加しましょう。 *この文章を支持する「データ」を追加しましょう *この文章に対する「裏付け」を追加しましょう ・この文章を支持している既存の局所論証を改訂しましょう。**
論証が弱い $I(S_i, S_{d_r}^i   \Psi_{S_{id_r}}) < \epsilon_0$	この論証は弱いようです。以下のような方法で論証を改訂しましょう。 ・この論証を削除し、より強い論証に変更しましょう。 ・この論証を支持する以下のような文章を追加し、論証を強化しましょう。 *この論証に対する「論拠」を追加しましょう *この論証に対する「限定」を追加しましょう

\* : Toulmin モデルの構成規則に従い、その文章/論証に追加可能な構成要素のみ提示

\*\* : その文章を支持する局所論証が、少なくとも一つ存在するときに提示

(データ) . よって、彼にはペニシリン治療が必要である (主張)」を次のいずれかの方法で改訂する。(B1)「彼は肺炎である (データ)」を削除し、より強い「データ」を提示する。(B2) この論証を支持する文章 (例では「なぜなら、肺炎にはペニシリン治療が有効であるからだ (論拠) 」) を追加する。

本システムでは、これらの論証改訂方法を用いて、論証特性指標の推定結果に応じた次のようなアドバイスのフィードバックを行う。

まず、任意の文章  $S_i$  の正当性が低い、すなわち、 $p(S_i = 1) < \epsilon_1$  である場合、その文章の正当性を高めるために、上記の (A) または (B) の方法で改訂を行うようにアドバイスを与える。ただし、改訂方法 (B) に対応したアドバイス文「この文章を支持する既存の局所論証を改訂しましょう」は、その文章を支持する局所論証が少なくとも一つ存在するときのみ提示される。

更に, その文章  $S_i$  を正当化したとき (具体的には  $p(S_i = 1) = \epsilon_1$  となったとき), どの程度「主張」の正当性向上が見込めるかを, 「主張への影響度」の式 (3.4) を用いて次式で計算し, フィードバックする.

$$\Delta p(S_c = 1) = (\mu_i \epsilon_1 + \nu_i) - p(S_c = 1) \quad (3.5)$$

また, その文章を正当化するためにはどの程度の正当性向上が必要か, すなわち,  $\epsilon_1 - p(S_i = 1)$  もフィードバックする.

一方, 任意の論証について, 論証が弱い, すなわち  $I(S_i, S_{d_r}^i | \Psi_{S_{i,d_r}}) < \epsilon_0$  である場合には, (B1) または (B2) の方法で論証を改訂するようにアドバイス文を提示する.

なお, 改訂方法 (A) と (B2) に対応するアドバイスを提示する際には, Toulmin モデルの構成規則 (図 3.2) を用いて, 着目している文章/論証に追加可能な Toulmin モデルの構成要素を決定し, 追加を促すアドバイス文 (例えば「この論証に対する「論拠」を追加しましょう」) を提示する.

表 3.5 に, 各論証特性指標の推定結果に応じて生成されるアドバイス文をまとめる. 表 3.5 では, 任意の文章/論証に関する各論証特性指標の推定結果が, 各行の 1 列目の条件を満たすとき, その行の 2 列目に列挙されたアドバイス文が提示されることを表す. 例えば, ある文章  $S_i$  に着目したとき, 「文章の正当性が低い  $p(S_i = 1) < \epsilon_1$ 」ならば, 表 3.5 の 1 行目に記述されたアドバイス文が全て提示される.

### 3.5.2 論証構築支援システム

本節では, ここまでに述べた手法を実装した論証構築支援システムについて詳述する.

本システムは図 3.12 右のようなインターフェイスを持つ. ユーザは, 任意のワードプロセッサ (図 3.12 左) を用いて文章執筆を行いながら, 本システムを用いて以下の手順で Toulmin モデルを構築し, 論証の推敲を行う.

1. システムを起動すると, システム下部にアドバイス文「主張を入力してください」が表示される. ユーザはワードプロセッサに文章を執筆し, 「主張」に対応する文章をシステム左下のテキストボックスにコピーしてアドバイス文をクリックする.
2. 入力された文章が Toulmin モデルのノードとしてシステム上部に可視化され, 同時に,



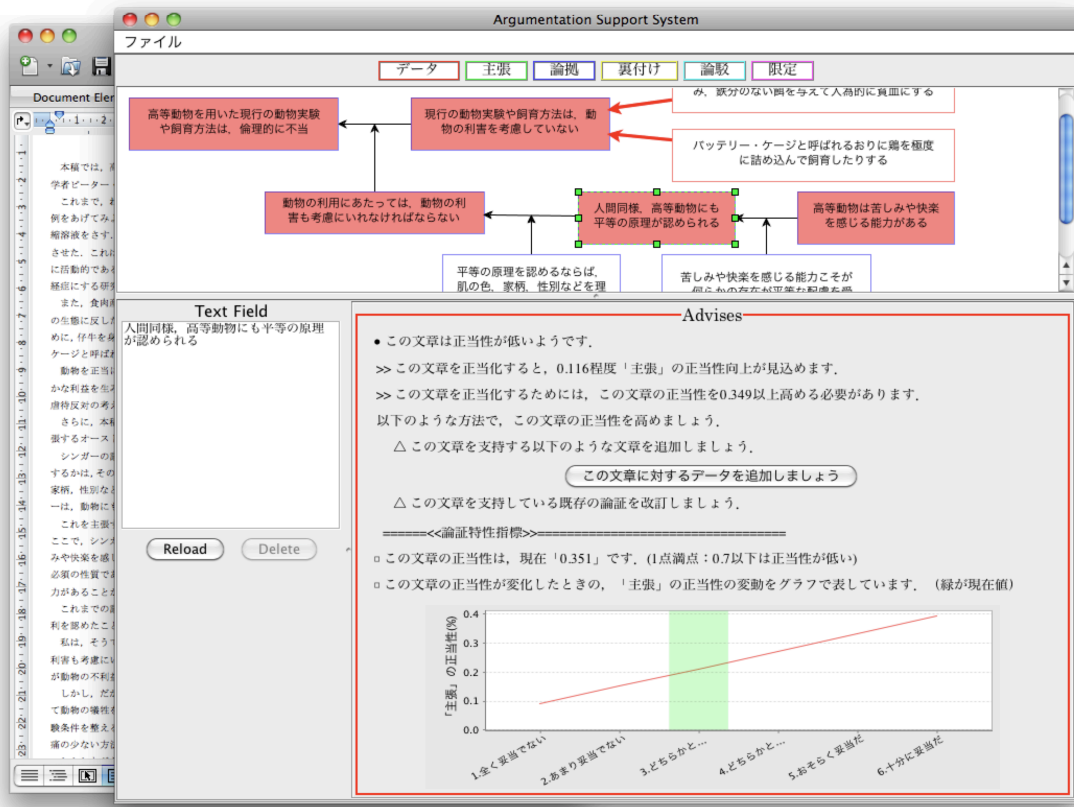


図 3.12 システムインターフェイス

その文章の正当性を自己評価するように促す図 3.13 のようなウィンドウが表示される。

3. ユーザは 3.3.1 で挙げた 6 段階の評価カテゴリを用いて自己評価を行う。システムは評価結果を主観確率に変換した後、論証中の全ての文章と局所論証について論証特性指標を計算する。
4. システム上に可視化されている Toulmin モデルのノードまたは有向辺をクリックすると、システム下部に各論証特性指標の算出結果が表示される。更に、これらの値が指標ごとに定めた閾値  $\epsilon$  未満である場合、表 3.5 の規則に従い、論証改訂のためのアドバイス文を表示する。
5. これらの論証特性指標及びアドバイスをもとに、ユーザはワードプロセッサ上の文章を改

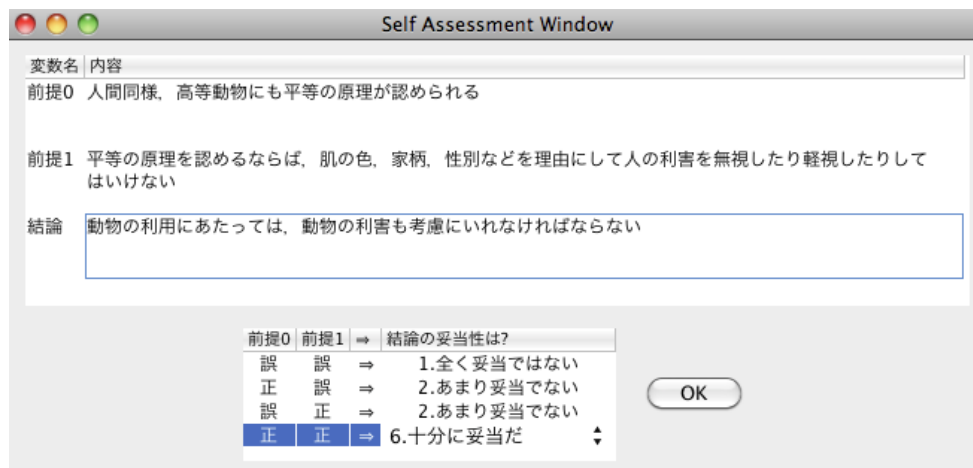


図 3.13 自己評価ウィンドウ

訂または加筆する。このとき、改訂方法 (A) または (B2) のアドバイスに従い、Toulmin モデルの構成要素に対応する文章を執筆した場合には、システム左下のテキストボックスにその文章をコピーしアドバイスをクリックする。

- Toulmin モデルにノードと有向辺が追加され、システム内部では Toulmin モデルがベイジアンネットワーク表現に変換される。同時に、追加された文章に関する自己評価を促すウィンドウが表示される。例えば、「主張」 $S_C$  に対して「データ」 $S_D$  を追加した場合、 $p(S_D = 1)$  及び  $p(S_C = 1|S_D = 1)$ ,  $p(S_C = 1|S_D = 0)$  の主観確率を問う質問文「 $D$  が真である確率」、「 $D$  が真のとき  $C$  が真となる確率」、「 $D$  が偽のとき  $C$  が真となる確率」が図 3.13 のように提示される。
- 以降、(3) ~ (7) を繰り返すことで、論証を Toulmin モデルに当てはめて可視化しながら、「主張」が十分に正当化されるまで文章と論証の推敲を繰り返す。

なお、システム上に可視化される Toulmin モデルのノードは、枠線が構成要素ごとに異なる色でハイライトされている。更に、文章の正当性が低い箇所と論証の弱い箇所は、対応するノード背景と有向辺が赤色でハイライトされる。

本システムは Java (Swing) を用いて実装した。

本システムが支援対象とする文章は、ある問題に対する自らの「主張」を、それを支える

表 3.6 論証改定カテゴリ

指摘理由カテゴリ	出現頻度		改訂方法カテゴリ	出現頻度	
	A 群	B 群		A 群	B 群
論証が弱い	4.65(2.84)	3.65(2.09)	その論証を削除する	3.80(2.34)	2.25(1.44)
			その論証を支持する文章を追加する	0.55(1.35)	1.35(1.63)
			その論証を支持する文章を変更する	0.30(0.45)	0.25(0.56)
文章の正当性が低い	1.30(0.81)	0.60(1.01)	その文章を削除する	0.05(0.23)	0.05(0.23)
			より正当性の高いもの書き換える	0.15(0.32)	0.07(0.27)
			その文章を支持する「データ」を追加する	1.10(0.88)	0.35(0.68)

「データ」や「論拠」を用いて論理的に展開する論証である。論文のように段落単位の構成が求められる文章であっても、Toulmin モデルの各構成要素として、段落全体あるいは各段落のトピックセンテンスを対応付けることで、Toulmin モデルを構築できるため、本システムを利用できる。また、論証が複雑な場合でも、画面右にスクロールバーが出現し、ネットワークとアドバイスの提示が可能となる。

## 3.6 評価実験

ここでは、被験者実験により、本システムの有効性を評価する。具体的には、論証が複雑なとき、1. 論証の強さ、2. 文章の正当性、3. 主張への影響度、を考慮した論証の構築を本システムが支援できるかを評価する。

### 3.6.1 実験方法

上記の評価を行うために、本実験では、筆者らが用意した比較的複雑な論証を、本システム及び比較用システムを用いて被験者に推敲させ、被験者が行った改訂の内容と改訂後の論証の質を評価する。具体的な実験手順は以下の通りである。

1. 被験者を、本システムを利用する A 群と、論証を Toulmin モデル [36] に当てはめ可視化する機能のみを持つシステムを利用する B 群にランダムに振り分けた。各グループの被

験者は論文執筆経験のない理系大学生 10 名ずつとした。

2. 実験内容を説明した後、Toulmin モデルと各論証特性指標、システムの利用方法に関する説明を 20 分程度行った。また、文章の正当性に対応する条件付確率の評価例を例文を用いて解説した。更に、実験内容とシステムの操作について理解を深めさせるために簡単な例題を用いた演習と解説を 10 分程度実施した。
3. 3.4.4 の実験で利用した 2 種類の論証をユーザに配布した。被験者には、原文とその論証の Toulmin モデルを記載した紙資料を配布した。
4. 被験者には、まず、資料と同じ構造の Toulmin モデルをシステム上に構築させた後、システムを自由に操作させながら、次の作業を行わせた。
  - (a) 論証の改訂が必要だと思う箇所（以下、指摘箇所）を Toulmin モデルに記載された紙面上に全て指摘する。
  - (b) 指摘した箇所について、改訂が必要だと思った理由（以下、指摘理由）と、どのように改訂すべきだと思うか（以下、改訂方法）を自由記述する。
  - (c) 論証中の全ての文章について、その文章が「主張」の正当化のためにどの程度重要だと思うか（以下、影響度）を 3 段階（1. 重要、2. どちらとも言えない、3. 重要でない）で評価する。
  - (d) 実際に文章と論証（Toulmin モデル）を改訂する。

### 3.6.2 論証の推敲内容の妥当性評価

次に、3.6.1 の実験手順 (4a),(4b),(4c) で被験者が行った「指摘理由」「改訂方法」「影響度」の評価が、専門家（筆者らの一人）が B 群と同じ条件で同様の活動を行った結果とどの程度一致するかを評価する。

そのために、まず、上記の実験手順 (4b) で被験者が自由記述した「指摘理由」と「改訂方法」を表 3.6 のカテゴリに分類した。表 3.6 には、各群における各カテゴリの出現頻度の平均と標準偏差（括弧内）も示した。表より、どちらの群の被験者も「論証の強さ」「文章の正当性」の観点で論証を推敲していることがわかる。

表 3.6 のカテゴリを用いると、実験手順 (4a),(4b),(4c) から得られたデータは、被験者

表 3.7 実験結果

		$\kappa$ 係数			専門家評価	
		指摘理由	改訂方法	影響度	論証の強さ	文章の正当性
論証 1	A 群	.39(.15)	.36(.18)	.38(.22)	3.70(.79)	3.70(.79)
	B 群	.21(.18)	.18(.17)	.15(.13)	2.10(1.10)	2.10(1.10)
	$p$	< .05	< .05	< .01	< .01	< .01
論証 2	A 群	.31(.11)	.31(.12)	.41(.21)	3.00(1.15)	3.10(1.19)
	B 群	.17(.18)	.17(.18)	.11(.12)	1.90(1.10)	1.80(1.03)
	$p$	< .05	< .05	< .01	< .05	< .05
		追加要素数	削除要素数	総文章量	所要時間	
論証 1	A 群	2.50(2.14)	5.70(2.06)	936(294)	44.00(6.96)	
	B 群	1.40(1.77)	2.60(2.01)	1088(140)	39.30(4.14)	
	$p$	< .10	< .01	< .10	< .05	
論証 2	A 群	1.30(1.33)	3.00(3.71)	1434(377)	32.60(8.86)	
	B 群	1.20(1.69)	1.30(1.06)	1641(96)	26.90(4.32)	
	$p$		< .10	< .10	< .05	

$e \in \{1, \dots, n\}$  が各 Toulmin モデル中のノードと有向辺に付与した { 指摘理由カテゴリ, 改訂方法カテゴリ, 影響度 } の集合として表される.

ここでは, このデータを用いて, 複数の評価者による評定の一致率の指標として知られる  $\kappa$  係数 [67] を, 専門家と各被験者間について算出した.  $\kappa$  係数は, 被験者  $e$  がカテゴリ  $i$  と評価し, 専門家がカテゴリ  $j$  と評価した頻度を  $O_{ij}(e)$  とすると, 以下で定義される.

$$\kappa(e) = \frac{\sum_{i=1}^K O_{ii}(e)/N - \sum_{i=1}^K E_{ii}(e)/N}{1 - \sum_{i=1}^K E_{ii}(e)/N} \quad (3.6)$$

ここで,  $E_{ii}(e) = \sum_{i=1}^K O_{ij}(e) \sum_{j=1}^K O_{ij}(e)/N$ ,  $N = \sum_{i=1}^K \sum_{j=1}^K O_{ij}(e)$  を表す.

表 3.7 に, 専門家と各群の被験者間について算出した  $\kappa$  係数の平均と標準偏差 (括弧内) を示す. 表中の,  $p < .05(.01)$  は対応のない  $t$  検定の結果, 有意水準 5% (1%) で有意差があったことを表す. 表より, 全ての項目において, A 群の  $\kappa$  係数が B 群より有意に大きいことがわかる. このことから, 本システムを利用することで, 被験者は「論証の強さ」「文章の正当性」「主張への影響度」の観点で, より専門家に近い論証の推敲を行うことができたといえる.

### 3.6.3 改訂された論証の質の評価

次に、被験者が改訂した論証が、「論証の強さ」「文章の正当性」の観点で改善されたかを評価する。そのために、被験者が改訂した文章と論証を、以下の2項目について4段階(1. 非常に多い, 2. やや多い, 3. あまりない, 4. ほとんどない)で専門家に採点させた。

1. 論証が弱く、改訂が必要な個所がどの程度みられるか。
2. 文章の正当性が低く、改訂が必要な個所がどの程度みられるか。

表 3.7 に、各項目における評価値の平均と標準偏差(括弧内)を示す。表 3.7 中の専門家評価の列に記載した  $p < .05(.01)$  は、Wilcoxon-Mann-Whitney 検定の結果、有意水準 5%(1%)で有意差があったことを表す。表より、全ての項目において、本システムを用いた A 群が有意に高い評点を得ていることがわかる。このことから、本システムの利用により、「論証の強さ」「文章の正当性」の観点で論証の改訂が促進されることが示せた。

ここで、被験者による論証の改訂と専門家による評点との関係を分析するために、改訂時に追加・削除された Toulmin モデルの構成要素数と改訂後の総文章量、所要時間を計測した。結果を表 3.7 に示す。表より、被験者による改訂では、追加要素数に比べ削除要素数が多いことが確認できる。これは、本実験で利用した論証が、冗長な文章を多く含むこと、領域知識を要するテーマであり証拠を新たに創出することが困難であったことに起因するといえる。

一方、専門家による評点が高かった A 群の特徴として、追加・削除要素数ともに B 群に比べて多いことが挙げられる。ここで、追加・削除要素数と、専門家が採点した「論証の強さ」「文章の正当性」の平均点との相関係数を算出したところ、それぞれ  $r = .38(p < .05)$ 、 $.54(p < .01)$  で有意な正の相関が確認された。このことから、A 群の被験者は、文章の追加・削除が必要な箇所を B 群の被験者より多く同定し効果的に改訂できたため、高い評点を得られたといえる。

更に、表 3.7 より、A 群の被験者の方が所要時間が有意に長かったことが確認できる。このことから、本システムが、被験者による注意深い論証の推敲を促進したと解釈できる。

以上の実験より、本システムが、1. 論証の強さ、2. 文章の正当性、3. 主張への影響度を考慮した論証の構築を支援できることを示せた。

### 3.7 むすび

本章では、アカデミックライティングにおける論証構築支援として、論証が複雑なときにも、「論証の強さ」を重視した論証の構築を支援できるシステムを開発した。このために、本章では、Toulmin モデルに当てはめて構築された論証をベイジアンネットワークとして定式化する手法を提案した。更に、Toulmin モデルのベイジアンネットワーク表現を用いて、1. 論証の強さ、2. 文章の正当性、3. 主張への影響度という 3 つの論証特性指標を定義し、これらの指標を用いて論証改訂方法をアドバイスする手法を提案した。

また、被験者実験により、本システムが 1. 論証の強さ、2. 文章の正当性、3. 主張への影響度、を考慮した論証の構築を支援できることを示した。

なお、本論では、論証のスキーマとして Toulmin モデルを採用した。しかし、Toulmin モデルが、演繹法、帰納法、背理法などの全ての推論パターンを表現できるかは検証していない。今後は、この検証を実施し、アカデミックライティングにおける論証スキーマとしての Toulmin モデルの適性や課題を明らかにしたい。また、本論では、本システムを利用することで論証の推敲能力が向上することまでは主張できていない。今後はこの能力を伸ばすための利用方法についても研究する。

## 第4章

# 結言

本論では、アカデミック・ライティングにおける論文構成と論証の構築を支援するシステムの開発と評価について述べた。

第2章では、論文構成構築支援システムの開発を目的とし、「論文構成」を情報理論における情報源からの出力符号系列とみなしたメタファとしてとらえ、論文構成の構築過程を定式化した。具体的には、論文構成を論文要素カテゴリの系列データとし、それが  $m$  重マルコフ情報源に従うと仮定した。また、データ長の短い論文構成データからより正確に多重度を推定するために、多重度の推定補正法を提案し、結果として「序論」 $\hat{m} = 2$ 、「本論」 $\hat{m} = 3$ 、「結論」 $\hat{m} = 2$  という推定値を得た。このことから、論文構成は画一的な形式に従うのではなく、直前の論文要素の系列に応じて多様な形式を取りうるという知見を得た。また、推定された  $m$  重マルコフ情報源により、多様なパターンの論文構成を高精度で予測できることを実験により確認した。さらに、 $m$  重マルコフ情報源に基づく論文構成構築支援システムを開発し、被験者実験により、提案システムが妥当かつ多様なパターンの論文構成の構築を支援でき、論文構成への意識と文章の論理性の向上にも寄与することを示した。

第3章では、アカデミックライティングにおける論証の構築支援として、論証が複雑なときにも、「論証の強さ」を重視した論証の構築を支援できるシステムを開発した。このために、Toulmin モデルに当てはめて構築された論証をベイジアンネットワークとして定式化する手法を提案した。更に、Toulmin モデルのベイジアンネットワーク表現を用いて、1. 論証の強さ、2. 文章の正当性、3. 主張への影響度という3つの論証特性指標を定義し、これらの指標を用いて論



証改訂方法をアドバイスする手法を提案した。また，被験者実験により，本システムが 1. 論証の強さ，2. 文章の正当性，3. 主張への影響度，を考慮した論証の構築を支援できることを示した。

以上より，本論で開発したシステムを利用することで，初心者であっても論文構成と論証を意識した文章執筆を行うことができ，結果として，文章の論理性が向上したことが示された。ただし，本論では，本システムを利用することでアカデミックライティングの能力が向上することまでは主張できていない。この能力を伸ばすための利用方法については今後の課題としたい。

## 参考文献

- [1] Noriko Kando. Text-level structure of research papers: Implications for text-based information processing systems. In *In Proceedings of the British Computer Society Annual Colloquium of Information Retrieval Research*, pp. 68–81, 1997.
- [2] Christine Neuwirth, David Kaufer, Richard Chimera, and Terilyn Gillespie. The notes program: A hypertext application for writing from source texts. In *Hypertext*, pp. 121–141, 1987.
- [3] H Shibata and K Hori. A framework to support writing as design using multiple representations. In *Proceedings of 5th Asia pacific Conference on Computer Human Interaction*, 2002.
- [4] 大井恭子, 石川直美. 「知識伝達モデル」から「知識変形モデル」への発展を志向するアカデミック・ライティング指導: 「問いかけ」とピア・レビューの重要性に着目して. 千葉大学教育学部研究紀要, Vol. 54, pp. 105–117, 2006.
- [5] Sandra M. Aluisio and Osvaldo Novaes de Oliveira Jr. A detailed schematic structure of research paper introduction: An application in support-writing tools. In *Procesamiento del Lenguaje Natural*, 1996.
- [6] 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. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, Vol. 97, No. 593, pp. 41–48, 1998.
- [7] 溝淵全人, 松居辰則, 岡本敏雄. 論理的構造を考慮した論文作成支援. 電子情報通信学会技術研究報告. ET, 教育工学, Vol. 100, No. 682, pp. 139–146, 2001.
- [8] Valeria Delisandra Feltrim, Jorge Marques Pelizzoni, Simone Teufel, Maria das Gracias Volpe Nunes, and Sandra M. Aluisio. Applying argumentative zoning in an automatic

- 
- critiquer of academic writing. In Ana L. C. Bazzan and Sofiane Labidi, editors, *SBIA*, Vol. 3171 of *Lecture Notes in Computer Science*, pp. 214–223. Springer, 2004.
- [9] Stephen T. O'Rourke and Rafael A. Calvo. Analysing semantic flow in academic writing. In *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*, pp. 173–180, Amsterdam, The Netherlands, The Netherlands, 2009.
- [10] Takayuki Fujimoto and Tokuro Matsuo. An approach to cooperative research paper writing / edit system based on the interactive-edit method. In *IJCSNS International Journal of Computer Science and Network Security*, Vol. 7, pp. 15–20, 2007.
- [11] Hidenobu Kunichika, Chiaki Miyazaki, Yuto Matsuki, Robert Chartrand, and Akira Takeuchi. An intelligent partner for organizing a paragraph. In *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*, pp. 549–556, Amsterdam, The Netherlands, The Netherlands, 2009. IOS Press.
- [12] 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. 情報処理学会研究報告, 情報基礎研究報告会, 第 93 巻, pp. 59–62, 1993.
- [13] 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訂支援システムの試作と評価. 情報処理学会研究報告. 自然言語処理研究会報告, Vol. 95, No. 27, pp. 79–84, 1995.
- [14] Sandra M. Aluisio, Valeria Tomas de Aquino, Rafael Pizzirani, and Osvaldo Novaes de Oliveira Jr. Assessing high-order skills with partial knowledge evaluation: Lessons learned from using a computer-based proficiency test of english for academic purposes. *JITE*, Vol. 2, pp. 185–202, 2003.
- [15] Sandra M. Aluisio and R.E Gantenbein. Toward the application of systemic functional linguistics in writing tools. In *Published in the Proceedings of the ISCA 12th International Conference*, pp. 181–185, 1997.
- [16] 山崎通弘, 山村毅, 大西昇. スタイル選択可能な文章作成支援システム. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, Vol. 98, No. 660, pp. 1–8, 1999-03-12.

- [17] Imre Csiszar and Paul C. Shields. The consistency of the bic markov order estimator. In *Annals of Statistics*, Vol. 28, pp. 1601–1619, 2000.
- [18] H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, Vol. 19, No. 6, pp. 716–723, 1974.
- [19] Bertrand Clarke, Andrew, and R. Barron. Information-theoretic asymptotics of bayes methods. *IEEE Transactions on Information Theory*, Vol. 36, pp. 453–471, 1990.
- [20] Zsolt Talata. Divergence of information-criterion based markov order estimators for infinite memory processes. In *ISIT*, pp. 1378–1382. IEEE, 2010.
- [21] I. Csiszár. Large-scale typicality of Markov sample path and consistency of MDL order estimators. *IEEE Trans. Inform. Theory*, Vol. 48, No. 6, pp. 1616–1628, 2002.
- [22] Tie Liu. Application of markov chains to analyze and predict the time series. *Modern Applied Science*, Vol. 4, No. 5, pp. 162–166, 2010.
- [23] J. Rissanen. Modeling by shortest data description. *Automatica*, Vol. vol. 14, pp. 465–471, 1978.
- [24] J. Rissanen. Complexity and information in modeling. In *Complexity and Constructivity in Economic Analysis*. Blackwell Publishings, 2005.
- [25] Petri Kontkanen and Petri Myllymäki. A linear-time algorithm for computing the multinomial stochastic complexity. *Inf. Process. Lett.*, Vol. 103, No. 6, pp. 227–233, 2007.
- [26] Toshiyasu Matsushima, Hiroshige Inazumi, and Shigeichi Hirasawa. A class of distortionless codes designed by bayes decision theory. *IEEE Transactions on Information Theory*, Vol. 37, No. 5, pp. 1288–1293, 1991.
- [27] R. Krichevsky and V. Trofimov. The performance of universal encoding. *Information Theory, IEEE Transactions on*, Vol. 27, No. 2, pp. 199–207, 1981.
- [28] Bertrand S. Clarke and Andrew R. Barron. Jeffreys’ prior is asymptotically least favorable under entropy risk. *Journal of Statistical Planning and Inference*, Vol. 41, No. 1, pp. 37–60, 1994.
- [29] Z. Talata. Model selection via information criteria. In *Periodica Mathematica Hungar-*

- 
- ica*, Vol. 5, pp. 99–117, 2005.
- [30] Tomi Silander, Petri Kontkanen, and Petri Myllymaki. On sensitivity of the map bayesian network structure to the equivalent sample size parameter. In *Proceedings of the Twenty-Third Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-07)*, pp. 360–367, Corvallis, Oregon, 2007.
- [31] Haroon Malik and Elhadi Shakshuki. Predicting function changes by mining revision history. In *Proceedings of the 2010 Seventh International Conference on Information Technology: New Generations*, ITNG '10, pp. 950–955, 2010.
- [32] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, Vol. 22, No. 1, pp. 5–53, 2004.
- [33] 戸田山和久. 論文の教室. 日本放送出版協会, 2002.
- [34] 館野泰一, 大浦弘樹, 望月俊男, 西森年寿, 山内祐平, 中原淳. アカデミックライティングにおける協同推敲活動環境の構築と評価. 日本教育工学会研究報告集, Vol. 8, No. 3, pp. 59–62, 2008.
- [35] Oliver Scheuer, Frank Loll, Niels Pinkwart, and Bruce M. McLaren. Computer-supported argumentation: A review of the state of the art. *I. J. Computer-Supported Collaborative Learning*, Vol. 5, No. 1, pp. 43–102, 2010.
- [36] Stephen Edelston Toulmin. *The Use of Argument*. Cambridge University Press, 1958.
- [37] 館野泰一, 大浦弘樹, 望月俊男, 西森年寿, 山内祐平, 中原淳. アカデミック・ライティングを支援する ict を活用した協同推敲の実践と評価. 日本教育工学会論文誌, Vol. 34, No. 4, pp. 417–428, 2011.
- [38] Ole C. Brudvik, Khai Seng Hong, Yam San Chee, and Libo Guo. Assessing the impact of a structured argumentation board on the quality of students' argumentative writing skills. In *Proceeding of the 2006 conference on Learning by Effective Utilization of Technologies: Facilitating Intercultural Understanding*, pp. 141–148, 2006.
- [39] Jamaludin Azilawati, Caroline Mei Lin Ho, and Yam San Chee. The impact of structured argumentation and enactive role play on students' argumentative writing skills.

- In *ICT : providing choices for learners and learning, 2007*, pp. 430–439, 2007.
- [40] Chris Reed and Glenn Rowe. Araucaria: Software for argument analysis, diagramming and representation. *International Journal of AI Tools*, Vol. 14, pp. 961–980, 2004.
- [41] 堀田昌英, 榎戸輝揚, 岩橋伸卓. 多元的議論構造の可視化手法 : 社会技術としての政策論議支援. *社会技術研究論文集*, Vol. 1, pp. 67–76, 2003.
- [42] 新田克己, 柴崎真人, 安村禎明, 長谷川隆三, 藤田博, 越村三幸, 井上克巳, 白井康之, 小松弘. ダイアグラムに基づく法的論争支援システム. *人工知能学会論文誌*, Vol. 17, pp. 32–43, 2002.
- [43] 平柳行雄. ツールミン・モデルを利用したクリティカル・リーディング力向上. *日本英語コミュニケーション学会紀要*, Vol. 16, No. 1, pp. 82–92, 2007.
- [44] Caiquan Xiong, Ying Pan, and Dehua Li. A discussion information-structuring model based on the toulmin formalism. In *Proceedings of the 1st international conference on Forensic applications and techniques in telecommunications, information, and multimedia and workshop, e-Forensics '08*, pp. 441–446, 2008.
- [45] 水山光春. ツールミンモデルを用いた意志決定過程を組み込んだ環境教育の授業設計. *京都教育大学環境教育研究年報*, Vol. 4, pp. 27–40, 1996.
- [46] Finn V. Jensen and Thomas D. Nielsen. *Bayesian Networks and Decision Graphs*. Springer Publishing Company, Incorporated, 2nd edition, 2007.
- [47] Hei Chan and Adnan Darwiche. Sensitivity analysis in bayesian networks: from single to multiple parameters. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence, UAI '04*, pp. 67–75, 2004.
- [48] Linda C. van der Gaag and Silja Renooij. Analysing sensitivity data from probabilistic networks. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence, UAI'01*, pp. 530–537, 2001.
- [49] 氏川雅典. ツールミンの議論モデルの変容—批判から寛容へ. *ソシオロゴス*, Vol. 31, pp. 1–19, 2007.
- [50] 富田英司, 丸野俊一. 思考としてのアーギュメント研究の現在. *心理学評論*, Vol. 47, No. 2, pp. 187–209, 2004.

- 
- [51] 鈴木宏昭, 舘野泰一, 杉谷祐美子, 長田尚子, 小田光宏. Toulmin モデルに準拠したレポートライティングのための協調学習環境. 京都大学高等教育研究, Vol. 13, pp. 13–24, 2007.
- [52] Bart Verheij. Evaluating arguments based on toulmin’s scheme. *Argumentation*, Vol. 19, No. 3, pp. 347–371, 2005.
- [53] Chris Reed and Glenn Rowe. Translating toulmin diagrams: Theory neutrality in argument representation. *Argumentation*, Vol. 19, pp. 267–286, 2005.
- [54] F. V. James. Toulmin ’ s model and the solving of ill-structured problems. In David L Hitchcock and Bart Verheij, editors, *Arguing on the Toulmin Model. New Essays in Argument Analysis and Evaluation*, pp. 303–311. Springer, 2006.
- [55] J.T. Gage. *The Shape Of Reason: Argumentative Writing In College*. Longman Publishing Group, 2005.
- [56] Andrew Aberdein. The uses of argument in mathematics. *Argumentation*, Vol. 19, No. 3, pp. 287–301, 2005.
- [57] 野矢茂樹. 論理トレーニング. 産業図書株式会社, 2006.
- [58] 松原望. ベイズ統計学概説-フィッシャーからベイズへ. 培風館, 2010.
- [59] 植野真臣. ベイジアンネットワークの統計的学習. 人工知能学会誌, Vol. 25, No. 6, pp. 803–810, 2010.
- [60] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1 edition, September 1988.
- [61] David Earl Heckerman. *Probabilistic similarity networks*. Stanford University, Stanford, CA, USA, 1990.
- [62] 繁枘算男, 植野真臣, 本村陽一. ベイジアンネットワーク概説. 培風館, 2006.
- [63] Ruth Beyth-Marom. How probable is probable? a numerical translation of verbal probability expressions. *Journal of Forecasting*, Vol. 1, No. 3, pp. 257–269, 1982.
- [64] 増田規彦, 繁枘算男. 言葉を尺度とした主観確率の測定. 日本ファジィ学会誌, Vol. 8, No. 6, pp. 84–90, 1996.
- [65] 福澤一吉. 議論のレッスン. 生活人新書, 2002.
- [66] Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge Uni-

versity Press, New York, NY, USA, 1st edition, 2009.

- [67] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37–46, 1960.



# 謝辞

本研究を進めるにあたり、終始懇切なる御指導を賜った、電気通信大学大学院准教授の植野真臣先生に、心より感謝を申し上げます。本論文の審査過程において、数々の貴重な御助言と御指導を賜りました大須賀昭彦教授、長岡浩司教授、田中健次教授、田原康之准教授に、深謝申し上げます。また、本研究における議論・検討に当たって、ご教示とご激励を頂いた電気通信大学大学院植野真臣研究室の皆様にご礼申し上げます。

## 関連論文の印刷公表の方法及び時期

### 査読付き論文（本学位申請論文関連論文）

宇都雅輝，植野真臣，「ベイズ符号を用いた論文構成構築支援システム」，電子情報通信学会論文誌 D Vol.J94-D No.12 pp.2069-2081，2011 年

宇都雅輝，鈴木宏昭，植野真臣，「Toulmin モデルのベイジアンネットワーク表現を用いた論証推敲支援システム」，電子情報通信学会論文誌 D Vol.J96-D No.4, pp. 998-1011, 2013 年

### 査読付き論文（その他）

植野真臣，宇都雅輝「他者からの学びを誘発する e ポートフォリオ」日本教育工学会論文誌，35 巻，3 号，pp13-26，2011 年

### 国際会議

Maomi Ueno, Masaki Uto,「LERNING COMMUNITY USING SOCIAL NETWORK SERVICE」,IADIS Web Based Communities and Social Media 2011 Conference,July,2011

### その他（研究会等）

宇都雅輝，植野真臣，「形式手法に基づくアカデミック・ライティング支援システム」，教育システム情報学会学生研究発表会，2010 年 3 月

宇都雅輝，宮澤芳光，鈴木宏明，植野真臣，「確率的アプローチに基づく論文構成構築支

援システム」, 日本教育工学会研究会, 日本教育工学会研究報告集 10(3), 2010 年 7 月

宇都雅輝, 宮澤芳光, 鈴木宏明, 植野真臣, 「確率的アプローチに基づく論文構成構築支援システム」, 教育システム情報学会 第 35 回全国大会, 2010 年 8 月

宇都雅輝, 植野真臣, 「情報論的アプローチに基づく論文構成構築支援システム」, 日本教育工学会 第 26 回全国大会, 2010 年 9 月

宇都雅輝, 鈴木宏昭, 植野真臣, 「ベイジアンネットワークを用いた論証支援システム」, 人工知能学会 第 83 回 人工知能基本問題研究会 (SIG-FPAI), 2011 年 11 月 27 日

宇都雅輝, 鈴木宏昭, 植野真臣 「Toulmin モデルとベイジアンネットワークを用いた論証構築支援システムの開発」, 電子情報通信学会教育工学研究会, 2011 年 12 月 9 日

## 1.1 付録

第3章で提案した論証構築支援システムの適用例として、第3章の実験で利用した論証と、情報システムの学術論文における論証を本システムで分析した例を示す。学術論文への適用の例としては、本論文の第2章の一部を用いた。

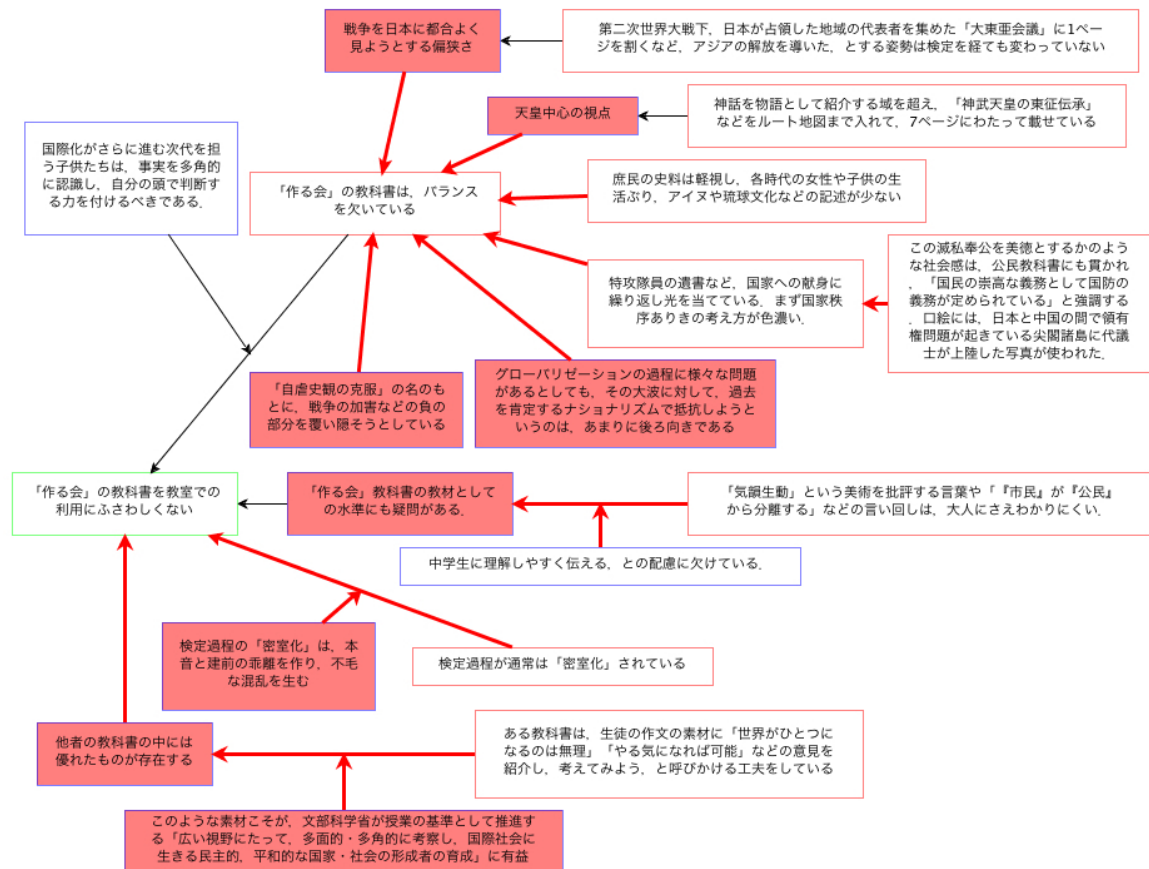


図1 論証1への適用例

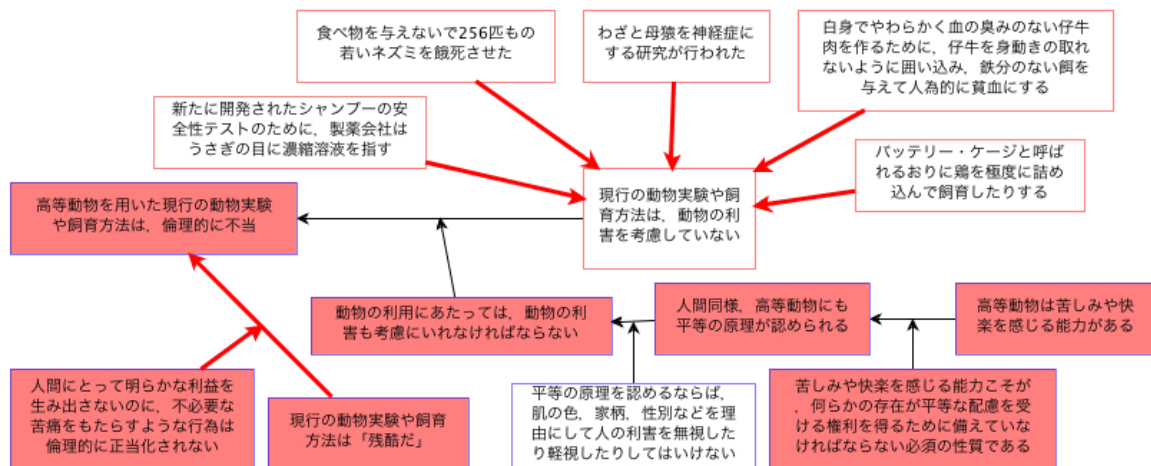


図 2 論証 2 への適用例

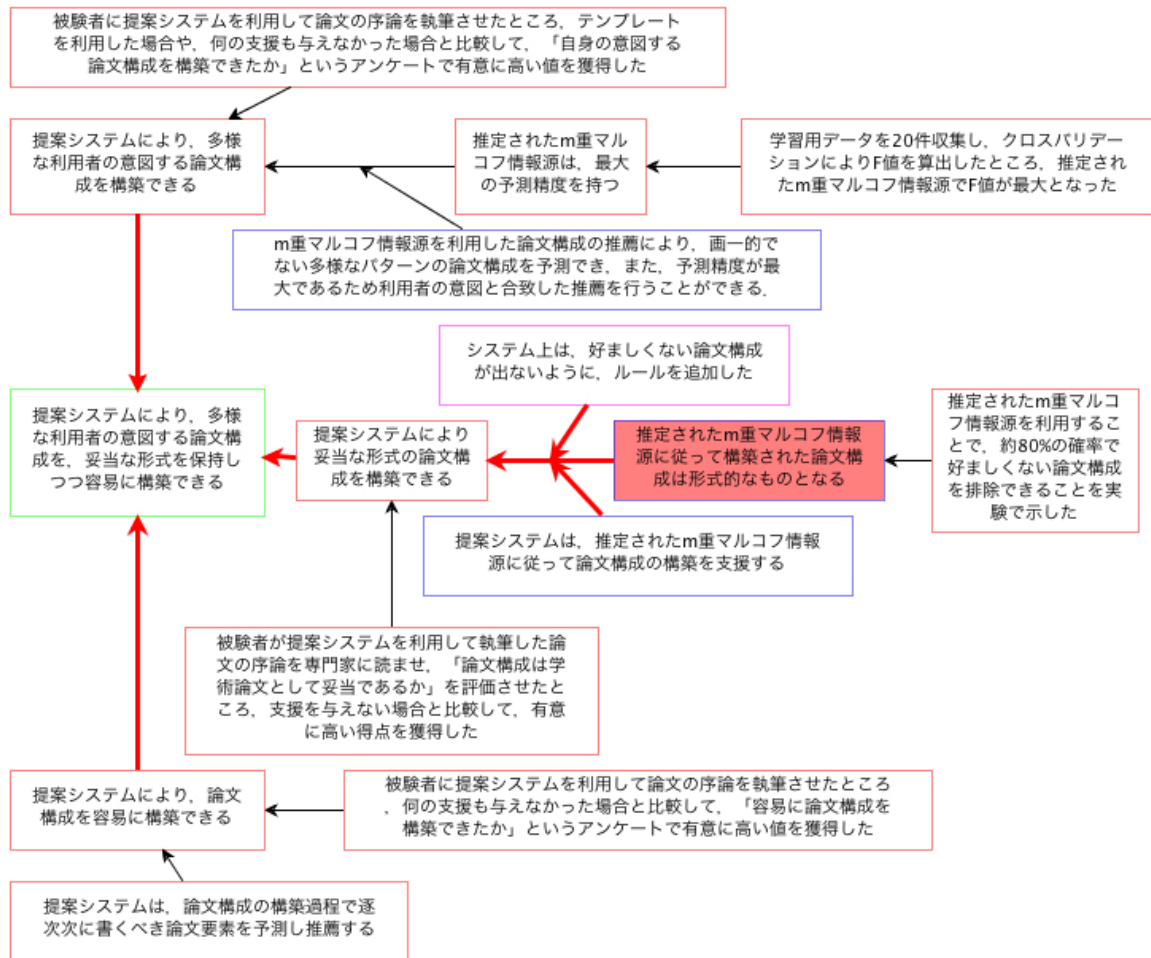


図3 学術論文（本論第2章）への適用例