学術論文における結論の構成要素特定手法の提案

高木陽介　　高間康史
首都大学東京
takagi-yousuke@ed.tmu.ac.jp, ytakama@tmu.ac.jp

1 はじめに

本稿では，学術論文の結論を構成する要素を，特徴的な表現および機械学習を組み合わせて特定する手法を提案する．

論文の構成を分析し，序論や結論などを構成する各文をその構成要素として特定することは，論文のサーベイの効率化を図る上で重要な技術である.一般的な特定手法としては，各構成要素に共通する文の特徴素を機械的に学習する手法や，特徴的な表現や記載位置を元に特定する手法が挙げられる．これまでの研究では論文概要やアブストラクトが主な特定の対象範囲とされてきた．しかし結論はこれまで特定の対象にされてこなかった．結論には論文の重要な情報が記載されており，それらの情報を機械的に特定することができれば ,論文サーベイの効率化に貢献することができる．

提案手法では，結論の構成要素として研究内容，研究成果 ,考察 ,今後の課題を対象とする .これらの構成要素に共通する特徴的な表現を分析して作成した手掛かり表現辞書に基づく抽出手法と，機械学習を用いて特定する手法を併用し，両者の長所を生かして結論の構成要素を特定する．評価実験により，提案手法の有効性を示す．

2 関連研究

論文から特定の情報を抽出する研究として問題文や実験情報を対象にした研究が挙げられる .酒井ら[1]は，論文概要に含まれる問題文をSVMを用いて判定している．人手で抽出した手掛かり語，機械的に抽出した特徴語 ,問題文の全ての単語の3通りの特徴語集合についてSVMで学習させた結果，すべての単語を用いた場合のF値が0.93と最も高い値を示したとしている．平井ら[2]は，学術論文から実験情報を抽出する手法を提案している．論文からルールベースで論文構成要素を抽出し，その論文構成要素を元に機械的に実験情報を抽出している .論文の構成要素として ,表 ,図 ,脚注 ,参考文献領域などを対象としている．

論文から各構成要素に該当する文を抽出する研究として，論文概要を対象にした研究やアブストラクトを対象にした研究が挙げられる．廣川ら[3]は論文概要の各文がどの観点に該当するのか複数人で判定して作成した学習用データに対して，SVMを用いて各観点のモデルを構築している．各文のベクトル化において，全ての単語を使う方法とSVMスコアの絶対値上位の単語を使う方法の二通りで評価を行った結果，後者の判定性能の方が良いという結果を示している．徳永ら[4]は学術論文のアブストラクトを動機 ,手法 ,結果の3つに分けてそれぞれに該当する文を手掛かり語や記載位置を参考に抽出した．実験の結果，動機の抽出率79%,手法の抽出率76%,結果の抽出率93%が得られ，全体で53%の抽出率が得られたことを示した．

重要文抽出に基づく要約生成のために，構成要素に着目した研究も存在する .SHINら[5]は論文全体を構成する「序論」「関連研究」「提案手法」「評価実験」から重要文のみを抽出し統合する手法を提案している．「序論」ではアブストラクトを正解データとし，SVMによって重要文か否かを判定している .「関連研究」では文中の単語のTF-IDFの和と手掛かり語を参考に重要文を抽出している .「提案手法」「評価実験」については，記載位置を手掛かりにする構想を示している．

3 提案手法

本研究では論理的な文章を構成する各文を論文の「構成要素」と定義する．章・節などの役割により，出現する構成要素は異なるが，論文の最後に位置する結論の場合，研究内容，研究成果，考察，今後の課題が主な構成要素となる．提案手法では，人手で作成する手掛かり表現辞書に基づく手法と，機械学習を用いた手法を併用する．前者の手法では，各構成要素に書かれる頻度が高い表現(特徴的表現)が所定の位置に書かれているかどうかを基準に構成要素を特定する．後者の手法では，前者の手法を用いて抽出した各文を学習データとして，SVMを用いて各構成要素の識別器を構築する．前者の手法では適合率は高いが再現率は低いことが想定される．一方，後者の手法では再現率が前の手法より高いことが想定されるため，両者を組み合わせることで再現率の向上が期待できる．そこで，両者を統合したハイブリッド型の特定手法も提案する．

3.1 手掛かり表現辞書に基づく特定手法

この特定手法では最初に既存の論文で頻出している文頭表現を機械的に抽出する．次に抽出した表現の中から特定の構成要素と関連性の高い文頭表現を抽出し，手掛かり表現辞書に登録する．また，それらと文中での共起率が高い文末表現を求め，特徴的表現として手掛かり表現辞書に登録する．表現X, Yの共起率CO(X,Y)は式(1)に示すSimpson係数によって求める．

$$\mathrm{CO}(\mathrm{X},\mathrm{Y}) = \frac{|X \cap Y|}{min(|X|,|Y|)} \quad (1)$$

|X|と|Y|は対象の文章の中でキーワードX，Yが含まれている文の数を表す．分子はXとYが同時に含まれている文の数を表す．

2009~16年度の言語処理学会年次大会の予稿と2013~2015年度の人工知能学会全国大会の予稿の結論約2 万文を対象として手掛かり表現辞書を構築した．表1に，研究内容に頻出する文末表現と文頭表現との共起率を示す．また，表2に研究成果，今後の課題における共起率を示す．研究内容における文頭表現は「本稿では」「本研究では」「本論文では」の3表現である．研究成果における文頭表現は「結果」，今後の課題の場合は「今後」である．これらの基準に基づき，共起率0.3以上を基準値として手掛かり表現を決定した．表内の赤字の表現は基準値を上回る表現である．青字の表現は基準値よりも下回っているが，特徴的な表現と判断して手掛かり表現辞書に登録した．考察は頻出する文頭表現が存在しなかったため，機械的に特徴的表現を抽出できなかった．そのため，既存の論文から手作業で，「実験結果から」「この結果から」「考えられる」「思われる」を抽出し，手掛かり表現辞書に登録した．最終的に，研究内容で32語，研究成果で16語，考察で4語，今後の課題で15語を手掛かり表現辞書に登録した．

抽出したこれらの特徴的表現が対象文に含まれているかどうかを基準に，結論内の各文が各構成要素に該当するかを判定する．文頭表現の場合は文頭10文字以内，文末表現の場合は文末15文字以内に含まれていることを条件とする．またできるだけ多くの構成要素を特定するため，どの構成要素にも該当しない文については，対等接続詞や指示語が文頭表現である場合，前の文と同じ構成要素と判定する．対等接続詞は「そして」「また」「さらに」などを指し，指示語は「この」「これ」「それ」などを指す．このようにした理由は，これらの文頭表現が書かれている文は前の文と同じ内容を述べているケースが多いためである．

表 1：研究内容における共起率

文末表現	全体	対象文	共起率	文末表現	全体	対象文	共起率
提案した	1392	912	0.655	抽出した	45	19	0.422
行った	1059	398	0.376	開発した	45	11	0.244
述べた	329	201	0.611	達成した	41	4	0.098
検討した	135	72	0.533	比較した	40	13	0.325
構築した	121	50	0.413	説明した	37	19	0.514
報告した	90	56	0.622	調べた	36	13	0.361
評価した	90	24	0.267	成功した	35	2	0.057
検証した	84	29	0.345	取り組んだ	31	24	0.774
考察した	82	38	0.463	おこなった	30	18	0.600
明らかにした	80	18	0.225	定義した	28	7	0.250
用いた	73	8	0.110	利用した	27	4	0.148
紹介した	71	38	0.535	論じた	24	9	0.375
実装した	63	19	0.302	提案する	23	10	0.435
分析した	61	32	0.525	目指した	21	12	0.571
調査した	60	20	0.333	実施した	21	5	0.238
作成した	56	15	0.268	議論した	21	12	0.571
実現した	52	10	0.192	分類した	20	4	0.200
行なった	51	20	0.392				

表 2：研究成果と今後の課題における共起率

文末表現	全体	対象文	共起率	文末表現	全体	対象文	共起率
確認した	456	156	0.342	予定である	693	461	0.665
わかった	294	126	0.429	必要がある	560	128	0.229
分かった	290	98	0.338	挙げられる	468	309	0.660
確認できた	184	81	0.440	課題である	267	23	0.086
得られた	117	46	0.393	必要である	185	34	0.184
確認された	94	39	0.415	目指す	115	64	0.557
示唆された	84	33	0.393	検討する	95	43	0.453
明らかになった	71	32	0.451	期待できる	94	13	0.138
見られた	70	20	0.286	期待される	79	19	0.241
向上した	39	26	0.667	検討している	70	26	0.371
出来た	38	21	0.553	検討したい	68	38	0.559
判明した	36	20	0.556	行いたい	50	21	0.420
				課題となる	48	10	0.208
				課題としたい	34	2	0.059
				目指したい	29	16	0.552
				予定している	29	15	0.517
				検討していく	28	13	0.464
				課題とする	20	2	0.100

3.2 機械学習を用いた特定手法

機械学習を用いた特定手法は構成要素特定の再現率を向上させることを目的として導入する．訓練データは手掛かり表現辞書構築に用いた結論約2万文から，3.1節で述べた特定手法で各構成要素に該当すると判定された文を用いる．SVMの学習にはLIBSVM[1]を用いる．各構成要素の特徴的な表現は文中に書かれる傾向にないため，訓練データの各文から文頭・文末の10文字を対象とし，単語2-gramにより素性ベクトルに変換する．各構成要素で頻出している素性ベクトルのみを学習に用いる．予備実験として，各構成要素で出現頻度が上位50,70,80,100個の素性ベクトルを元にそれぞれ学習を行い，10-分割交差検証によってそれぞれの分類性能を評価した．その結果，各構成要素の上位70個(計218個)を学習に用いた場合の判定性能が最も高かったため，これを採用する．

3.3 ハイブリッド型特定手法

ハイブリッド型特定手法では，3.1節，3.2節でそれぞれ述べた特定手法の判定結果両方を元に最終的な判定を行う．一般論として，人手で求めた手掛かり表現に基づく特定手法の方が適合率は高く再現率は低い傾向にあり，機械的な手法の方が再現率は高く適合率が低い傾向にあるとされている．このことから，本特定手法では基本的に適合率が高い手掛かり表現に基づく特定手法の判定結果を反映させる．もし対象文がどの構成要素にも該当しないと判定された場合，機械学習を用いた特定手法の判定結果を反映させる．このようにすることで，手掛かり表現に基づく特定手法に影響を与えることなく補完することができる．

4 評価実験

3節で提案した3種類の特定手法の性能を評価する実験を行った．実験データとして用いたのは2017年度の言語処理学会の予稿集から無作為に抽出した100件の論文，合計739文である．事前に手作業で各文を構成要素に分類した結果，全739文中619文がいずれかの構成要素に該当しており，内訳は研究内容が243文で，研究成果が130文，考察が40文，今後の課題が206文であった．評価指標は再現率，適合率， F値を採用し，構成要素ごとに算出する．

4.1 手掛かり表現に基づく特定手法の実験結果

3.1節で述べた手掛かり表現に基づく特定手法の再現率・適合率・F値を表3に示す．いずれの構成要素も適合率の方が再現率よりも高い傾向にあり，全体でも0.18程度，適合率の方が高い結果が得られた．研究内容と今後の課題のF値は0.8を超えているが，研究成果は0.7，考察は0.6程度と低い値となった．また考察の適合率は他の構成要素よりも0.2以上低い値をとなっている．以上より，本手法は特定の精度は高いが網羅性は低いという特徴があることがわかる．

4.2 機械学習を用いた特定手法の実験結果

3.2節で述べた機械学習を用いた特定手法の実験結果を表4に示す．当初の想定通り，手掛かり表現に基づく手法と比べて全体的に再現率が向上し，適合率が低下している傾向があることがわかる．また，考察に関しては表3と同様，他の構成要素よりも低い値となっている．この理由として，収集した手掛かり表現が少ないため，学習データも同じような表現しか書かれておらず学習の効果が薄かったことが考えられる．

[1] https://www.csie.ntu.edu.tw/~cjlin/libsvm/

4.3 ハイブリッド型特定手法の実験結果

3.3節で述べたハイブリッド型特定手法の実験結果を表5に示す．提案手法では2種類の特定手法のどちらかで判定できた結果を出力するため，再現率は表3と表4よりも向上していることがわかる．一方，適合率はどの構成要素においても表3と表4の中間の値となっている．F値は向上していることから，2つの手法を併用することは有効であるといえる．

2つの手法の判定結果が異なっていた文は24文あったが，その中で手掛かり表現による特定手法が正解であった場合は19文であった．機械学習を用いた特定手法のみ判定できた文は113文あり，その中で77文が正解していた．このことからも，ハイブリッド型特定手法の有効性がわかる．

表 3 手掛かり表現に基づく特定手法の実験結果

	研究内容	研究成果	考察	今後の課題	全体
再現率	0.795	0.603	0.600	0.738	0.723
適合率	0.924	0.851	0.625	0.973	0.900
F値	0.855	0.706	0.612	0.839	0.802

表 4 機械学習を用いた特定手法の実験結果

	研究内容	研究成果	考察	今後の課題	全体
再現率	0.836	0.656	0.550	0.869	0.805
適合率	0.855	0.891	0.742	0.854	0.855
F値	0.845	0.756	0.632	0.861	0.829

表 5 ハイブリッド型特定手法の実験結果

	研究内容	研究成果	考察	今後の課題	全体
再現率	0.906	0.725	0.650	0.883	0.844
適合率	0.860	0.890	0.641	0.951	0.882
F値	0.882	0.799	0.645	0.916	0.863

5 おわりに

本研究では学術論文の結論を構成する要素として研究内容，研究成果，考察，今後の課題の4種類に着目し，これらに該当する文を特定する手法を提案した．手掛かり表現に基づく特定手法と機械学習を用いた特定手法を提案し，それら2つを組み合わせたハイブリッド型特定手法も提案した． 評価実験の結果，手掛かり表現に基づく手法は適合率が高く再現率が低く，機械学習を用いた手法は適合率が低く再現率が高いことがわかった．ハイブリッド型特定手法は両者の長所を組みあわせることで，適合率を維持しつつ再現率を向上させることが可能であることを示した．

今後の課題として，全ての指標において低い結果となった考察文の判定を改善する必要がある．また，前後の文の構成要素や，判定対象文の記載位置など，他の情報も判定の参考にすることで，より特定の性能が向上すると考えられる．さらに，構成要素によっては機械学習を用いた手法の方が適合率が高い場合も観測されたため，構成要素ごとに2つの特定手法の組み合わせ方を変更するアプローチも考えられる．今回の評価実験では正解データの作成を1名で行ったが，複数人による判定を行い，正解データの信頼性を高めることも必要と考える．

参考文献

[1] 酒井敦彦,廣川 佐千男.手掛り語に着目した論文概要からの課題抽出,火の国情報シンポジウム 2012, B-4-1, 2012.

[2] 平井久貴,新妻弘崇,太田学,高須淳宏.学術論文からの実験情報抽出の一手法, DEIM Forum2015, F3-1, 2015.

[3] 廣川 佐千男,酒井敦彦.学術論文概要中の各文の観点推定, 第 44 回「ディジタル図書館」ワークショップ, pp.20-24 , 2014.

[4] 徳永康次,延澤志保,太原育夫.テキスト構造に着目した学術論文の要旨自動生成のための重要文抽出, 第 6 回情報科学フォーラム, E-032, pp. 215-216, 2007.

[5] SHIN Wonha, 白井清昭.セグメント構造を考慮した学術論文の包括的要約の自動生成の提案,言語処理学会 第 23 回年次大会, pp.230-233,2017.