

論文作成支援システム

Paper Creation Support System

砂山 渡 橋 啓八郎
Wataru Sunayama Keihachiro Tachibana

広島市立大学 情報科学部
Faculty of Information Sciences, Hiroshima City University

A paper is one of the things that is published by a researcher to become a medium for the evaluation of their study. Therefore, a system that supports paper creation is suggested. This system is going to support improving the quality and the readability of a paper to assist its publication. The system extracts objective data of a paper and supplies those in visual interfaces. Authors of the paper interpret such data and modifies their paper repeatedly. As a result, papers are going to be published quickly to evolve social activities.

1. 近未来チャレンジ

研究者にとって、論文とは自らの成果を発表し客観的にその価値を認めてもらうための媒体である。そのため、論文の質および可読性を高めることで論文の生産効率を上げる論文作成支援システムの構築を提案する。

1.1 チャレンジの内容

研究論文は、他の研究者による審査によってその採否が決定される。論文の評価項目には、論文の質を問うものとして新規性、有効性、独創性などがあり、より多くの項目において高く評価される必要がある。また、論文の形式を問うものとして、論文のタイトルを初めとする内容の首尾一貫性、未定義語の使用などによる曖昧性、可読性などがあり、形式の不完全さだけで不採録とはされなくとも、査読者の心象に良い影響を与えることはない。そこで、これらの評価基準を論文原稿の客観的なデータから解析し、各評価基準に対する客観的な指標を設定、計算した上で、改善案を著者に示すことで、論文作成および改善の指針を与えるシステムの構築を目指す。

1.2 チャレンジが実現できる根拠

近年、大量のデータの中から有効な知識を取り出すデータマイニング技術の発展が目覚ましい。中でも電子化されたテキストを対象とするテキストマイニング技術として、キーワード抽出や重要文抽出、文書要約などにより文書の内容や特徴を知るための数多くの技術が開発されてきている。例えば、文書に含まれる単語を頻度順に表示しただけでも、その文書がどのような内容なのかを我々人間は推察することが可能である。このように文書から情報を抽出する技術は開発が進んでいるが、抽出された情報を元に文書の構築を支援する研究は少ない。情報の抽出結果を文書の作成者にフィードバックする機構およびインタフェースの構築によって、本チャレンジの内容は5年以内に実現可能であると考えられる。

1.3 社会への貢献が期待できる根拠

本システムにより、論文の質と可読性の向上および論文の生産効率の向上が見込まれるため、世の中の発展速度を高める効果が期待される。すなわち、論文は読者にとって有益な情報を含んで初めて価値が出るため、その有益となり得る情報をどれだけ多くまた具体的に含むかという点が論文の質に関わ

る部分である。また、論文の内容が読み手にとって理解されなければ、いかに優れた研究であってもその意味は乏しいものとなる。加えて、読みやすく質の高い論文の作成を支援することで、論文の審査に要する時間を短縮する効果があると考えられ、論文の生産効率を高めることにも繋がると期待される。

1.4 AIへの期待

情報抽出や知識発見によって取り出される情報は、情報を取り出すだけでなく取り出した情報をどのように応用して用いるかという点について深く考える必要があり、AIシステムを単なる入力から出力を得るための計算機システムではなく、我々の創造活動に実際に使えるシステムとしての位置づけの確立と意識の改革に役立つと考えられる。

2. 論文作成支援システム

研究の成果を論文として公表することを支援する論文作成支援システムを提案する。本システムは、研究そのものの質を高めることは直接の目的ではなく、質の高い研究とその成果が得られた時に、その内容を十分に伝えうる論文の作成を支援する。

2.1 システム概要

提案する論文作成支援システムの構成を図1に示す。ユーザは自らの論文原稿(テキストデータ)を入力としてシステムに与え、システムは入力されたテキストデータを論文評価部において解析することで論文修正の指針を与える客観データを出力する。論文評価部は、論文の可読性を評価する部分と論文の質を評価する部分からなり、その各部においてより具体的な項目を評価するモジュールが用意される。また、システムには客観データを直感的に理解できる視覚化インタフェースを用意する。ユーザはインタフェース上で各モジュールが出力する客観データを視覚的に見て検討を行なうことで、論文修正の指針を得る。ユーザは得た指針にしたがって自らの論文を修正した後、再び原稿をシステムへの入力として与える。この論文修正と客観データの提示を繰り返すことで、読みやすく質の高い論文の作成を支援する。

2.2 論文の可読性の向上

まず、論文が客観的に理解されるものであるかを表す指標を導入することで、論文の可読性の向上を支援する。これには、内容の首尾一貫性の確保、未定義語の除去、曖昧な表現の具体化や説明不足箇所における説明の追加などを目的として、これ

連絡先: 砂山渡, 広島市立大学情報科学部,
731-3194 広島市安佐南区大塚東 3-4-1, 082-830-1705,
sunayama@sys.im.hiroshima-cu.ac.jp

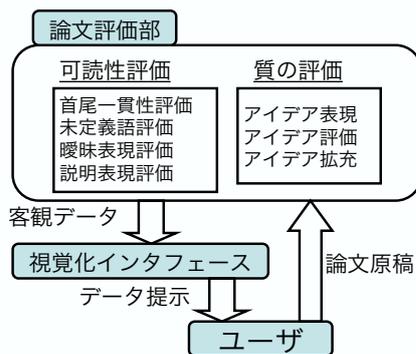


図 1: 論文作成支援システム

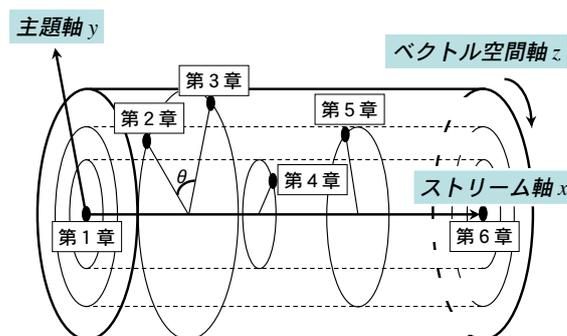


図 2: 論文構成表示インタフェースの例

らの程度を自動的に評価して修正を促したり、著者であるユーザが客観的に判断できる指標を導入する。

2.21 論文構成視覚化モジュール

論文の首尾一貫性および各章の位置づけを明確にするモジュールの構築例を以下に示す。まず、文章の首尾一貫性を評価するために、次の2つの尺度を導入する。

1. 文章の流れを表すストリーム軸
2. 文章の主題からの距離を表す主題軸

すなわちストリーム軸を、文章の第1章（第1段落）から最終章（最終段落）に至る道筋において、中間の章（段落）がその流れのどのあたりに位置するのかを示す尺度として定義し、主題軸を、文章の第1章（第1段落）から最終章（最終段落）までを貫く主題に関して、中間の各章（段落）が主題との程度関連するかを示す尺度として定義する。また対象とする文章の形式として、1章で全体を見通して話を展開した上で2章以降で展開された各部分について論理を展開し、最終章でそれらを再びまとめあげる構成のものを仮定する。このとき、文章中の各章（段落）がもつストリーム軸 x と主題軸 y の座標の値を、以下のように与える。

まず、 n 章構成の文章の第 i 章に含まれる単語（名詞、動詞、形容詞）の集合を X_i として、主題を表す単語の集合 T を、式 (1) で与える。

$$T = X_1 \cap X_n \quad (1)$$

すなわち、最初と最後の章に共通に現れる単語を文章の主題および主題と密接に関連する単語の集合として捉える。また、最初と最後の章の単語集合から集合 T に含まれる単語を除いた集合 S と G を式 (2) と式 (3) のように与える。ただし記号「-」は集合差を表す。

$$S = X_1 - T \quad (2)$$

$$G = X_n - T \quad (3)$$

すなわち、主題部分を除いた文章の背景や前提知識を表す単語が集合 S となり、主題に関連して得られた知見や導かれる結論を表す単語が集合 G となる。これらを用いて、各章 X_i のストリーム軸と主題軸の座標の値 (x_i, y_i) を式 (4) と式 (5) とで与える。ただし、 $n(A)$ は集合 A の要素数を表す。

$$x_i = \frac{n(S \cap X_i)}{n(S \cap X_i) + n(G \cap X_i)} \quad (4)$$

$$y_i = \frac{n(T \cap X_i)}{n(T)} \quad (5)$$

すなわち式 (4) によって、第1章の背景と最終章の結論のどちら側の単語をより多く含むかを測ることで、各章が論文全体のストーリーにおいて背景と結論のどちら側に近い内容であるのかを表す指標となる。また、式 (5) によって、主題を表す単語を各章が含む割合をはかることで、主題とどれだけ強く関連している内容であるのかを表す指標として用いられる。

ここでこれら2つの指標と、昨今よく用いられているベクトル空間モデル [Salton 75] との融合を測る。ベクトル空間モデルにおいては、論文の中で用いられる単語 k を各章が含むか否かを1,0の整数値で表し、論文の第 i 章はベクトル $X_i = \{x_{ik}\}$ (単語 k を含むとき $x_{ik} = 1$, 含まないとき $x_{ik} = 0$) と表現される。このとき、各章間の関連度を表す値として、第 i 章と第 j 章の角度の余弦は式 (6) のように表される。

$$\cos\theta(i, j) = \frac{\sum_k x_{ik} x_{jk}}{\sqrt{\sum_k x_{ik}^2} \sqrt{\sum_k x_{jk}^2}} = \frac{n(X_i \cap X_j)}{\sqrt{n(X_i)n(X_j)}} \quad (6)$$

視覚化インタフェースを用いた支援を考えた場合、人間が視覚的に理解可能なのは3次元までであるため、先のストリーム軸と主題軸に加えて、ベクトル空間モデルによる情報を1次元に圧縮して表現する必要がある。そこで、連続する2つの章のなす角度をベクトル空間軸 z の値として式 (7) で与える。すなわち、論文の話がどれだけ連続して展開しているかを表す指標として用いられる。

$$z_i = \cos\theta(i, i-1) \quad (7)$$

これら3つのストリーム軸、主題軸、ベクトル空間軸を視覚化したインタフェースの例（論文は6章構成を仮定）を図2に示す。これにより、各章の位置づけを視覚的に確認することが可能になり、著者が意図する位置づけと異なる場合には、足りない単語を用いた話を追加することで形で論文修正が可能となる。すなわち主題軸は主題との関係が明示的に表現されてい

るか否かを表す指標であるため、論文全体を通して値が小さいことが望まれ、この指標が大きな値をもつ章においては主題を表す単語 T に含まれる単語を用いて説明を追加する等の修正案が考えられる。ストリーム軸に関しては、一概に述べることは難しいが、論文全体の流れを作る意味において章間での値が異なるように内容を漸次変化させるために、前の章とは異なる新しい言葉の使用を積極的に促すことや、章を追うごと最終章に近くなるように使用する言葉を変化させていく修正案が考えられる。ベクトル空間軸においては連続する章間で値の差が大きい場合に、話の連続性が保てておらずストーリーの理解が困難な可能性があるとして、前後の章の関係を明確にするために連続する章で同じ言葉を使った文を加える修正が考えられる。

ここまでで述べてきた各軸の値の与え方も式 (4)、式 (5) と式 (7) に限定されるものではない。例えば主題軸の基準として、ユーザが与える単語（たとえば著者が論文に付与するキーワード）の集合をもとに、既存手法 [砂山 02] により単語を補充して主題となる単語集合を作成した上で、主題単語集合 T と作成することも考えられる。

2.22 未定義語除去モジュール

本モジュールは論文中の未定義の専門用語に適切な説明を与えることを促す。文章中の専門用語を抽出する研究は古くからあり [吉村 86]、専門用語の説明を作成する研究 [藤井 02] も現れている。そこでこれらの研究を応用することで論文中の用語の説明文をユーザである著者に提示し、定義そのものがなされているか、また定義を読者が理解することが可能であるかを著者が客観的に判断できる環境を構築する。すなわち、適切な言葉の定義が抽出されていない専門用語について、定義の有無の確認やその表現方法についての改善を促す。

2.23 曖昧表現評価モジュール

本モジュールは首尾一貫性モジュールとも関連して、論文の主題や主旨が明確に現れているかを評価し、主題の曖昧さを除去することを目的とする。人間ではなく機械が客観的な評価にもとづいて抽出するキーワードや重要文が、筆者が意図するキーワードや重要文と一致するのであれば、人間が理解することも容易であると考えられる。そこで、従来のキーワード抽出法や自動要約システム的应用を目指す [砂山 99, 相良 04]。システムが提示する論文のキーワードや重要文がユーザの意図とは異なる場合、適切な単語や文を補うことによる論文の改善を促す。

2.24 説明表現評価モジュール

本モジュールは論文中の説明が不十分な箇所を示唆して説明の追加を促すと共に、よりわかりやすい説明表現への改善を促す。まず、説明不足箇所の示唆には、語彙的連鎖によるテキストセグメンテーションの研究 [本田 94] を応用する。すなわち、論文中で単語の関係にもとづいて得られるセグメントを提示し、ユーザの意図とは異なるセグメント分割が行なわれている場合、そのセグメント間における説明が不足しているとしてセグメントをつなぐ説明の追加を促す。またセグメントの大きさが小さい場合にも、そのセグメントを提示することで、そこで述べられている事柄に対する十分な説明がなされているかの確認を促す。

よりわかりやすい説明表現への改善には、豊かな文章表現を支援する研究 [北田 01] や、文章の推敲を支援する研究 [?] があり、同義語や関連語を提供するシソーラスや推敲支援システム [Jungle] も販売されている。また、最近の多くのワードプロセッサに見られる、いわゆる「てにをは」の不備の示唆やスペルチェックがあり、用字用語の統一や文末等の表現スタイルの統一などの文章の仕上げ作業を支援する研究 [笠原 01] も

ある。これらのシステムの応用により、ユーザへの記述の改善を促す。

2.3 論文の質の向上

次に、論文が価値のあるものと認められるための、論文の質の向上を支援する。読み手にとっての論文の価値は、述べられている内容や結果の「意味づけ」および「解釈」によって、それがどれだけ世の中に貢献するか、また自らの研究活動に生かすことができるかによって判断される。

よくある論文の例として「すごいシステムを作りました。結果も実験評価によって優れていることが確認できます。」という書き方が見受けられる。確かに、結果が優れていることは論文が評価されるための必要条件かもしれないが、読者はすごいシステムそのものに価値を見いだすのではなく、そのシステムの動作原理や結果に意味を見いだして評価するため、論文の随所で明確な意味づけとともに説明がなされていなければ、十分にその価値が伝わることはない。そこで論文の価値を次の3つに大別して、それぞれを明確に表現することを支援する。

1. アイデア表現：研究の意図、アイデアの他の研究との違いや良さ（新規性、独創性）
2. アイデア評価：アイデアを生かしたシステムを用いて得られた有効な結果（有効性）
3. アイデア拡充：アイデア実現に伴って得られた関連情報と新たな知見

2.31 アイデア表現モジュール

1つ目は、論文が主張する新しいアイデアの位置づけを明確にすることで、論文の新規性と独創性の表現を支援する。そのためにはまず、論文間の参照情報をもとに関連研究間の関係を明確にする研究 [難波 01]、また文献を作成する研究者間の関係を見いだす研究 [松尾 03] などを元に、文献間の関係を表すネットワークを構築する。原稿である論文の参考文献情報を主な入力として、構築されたネットワーク上での研究の位置づけを視覚的に明示にする。

また、関連研究との比較においては、比較するポイント（関連研究とユーザの研究の共通点）と、ポイントに関する両者の特徴と違いを明確にする必要がある。この比較のポイントとはすなわち、関連文献のネットワークをどのような視点から見るかという観点に相当するため、さまざまな観点によって表示が切り替えられる視覚化インタフェースの構築が必要となる。その上でシステムは、ユーザが論文が意図した位置づけにあるかどうかを確かめ、不足文献を補い、他の文献との違いを明確に記述することを促す。

2.32 アイデア評価モジュール

2つ目は、論文に含まれるアイデアを検証するために行なった実験結果の記述を支援する。すなわち図3の上側の流れのように、実験結果の記述においてはまず、結果を数値などにより提示する。次にその数値にどのような意味があるのか解釈を与え、それらの結果と解釈が得られた原因は、アイデアのどの部分の貢献によるものなのかをフィードバックを伴う考察を与えることで初めてアイデアが検証される。

そこで、実験結果を述べる1つのセグメントから、実験結果の「提示」、「解釈」、「アイデアへのフィードバック」の3つのパートのそれぞれに相当する記述を抽出した上でユーザに提示する。これにより、3つの内の足りない項目についてユーザに追加記述を促す。すなわち、語彙的連鎖や形式的な段落情報を利用して1つの結果を述べている部分に相当するセグメント

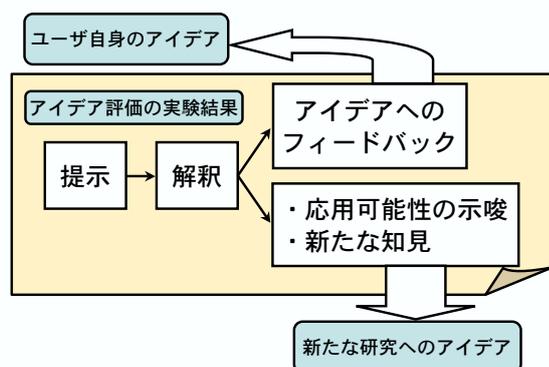


図 3: アイデア評価結果の記述の流れ

を切り出した上で、3つのパートそれぞれの記述の特徴に対して、文章から特定の知識を自動的に抽出する技術 [藤本 01] などに応用する。

さらに、アイデアを構成する要素を、論文中でアイデアを述べている記述部分をセグメント化することで自動抽出し、各要素に対応する「アイデアへのフィードバック」の記述の有無を明示することで、論文に足りない考察や実験をユーザに示唆して修正を促していく。

2.33 アイデア拡充モジュール

3つ目は、論文の本筋とは直接関係のない部分であるが、アイデアを検証するために行なった実験結果に関連して得られる新たな情報や新たな知見の記述を支援する。これは、図3の下側の流れに相当し、「アイデアへのフィードバック」の代わりに、「応用可能性の示唆や新たな知見」に相当する記述が存在する場合、これを自動的に抽出する。実験結果とその解釈に対して、アイデアへのフィードバックか応用可能性・知見の記述がない場合、そのいずれかの記述の追加をユーザに促す。

また、本モジュールが抽出できた記述すなわち知見の数を明示することで、その数が少ない場合にはより多くの結果と考察を示してアイデアを拡充し、論文の価値を高めることを促していく。

3. 結論

研究者の論文執筆活動を支援する論文作成支援システムの構築を提案した。本システムの構築により論文の質の改善が行なわれ、論文の公表のペース改善を図ると共に、論文を修正するユーザの中でも研究の位置づけや方向性に関する新たな発見、また考察を充実させることによる新たな知見が得られる。そのため本システムの実現によって、論文作成に関する支援効果以外にもユーザの研究の進展への効果が期待される。

参考文献

- [藤井 02] 藤井敦・石川徹也: World Wide Web を用いた事典知識情報の抽出と組織化, 電子情報通信学会誌, Vol.J85-D-II, No.2, pp.300-307, (2002).
- [藤本 01] 藤本和則, 賀沢秀人, 佐藤浩史, 島津光伸, 北寿郎: ネット情報を使った意思決定支援 DSIU における知識獲得技

術, 人工知能学会論文誌, Vol.16, No.1, pp.120 - 129, (2001).

[本田 94] 本田岳夫・奥村学: 語彙の結束性に基づいたテキストセグメンテーション, 情報処理学会自然言語処理研究会資料, No.102, pp.25 - 32, (1994).

[Jungle] 推敲支援システム: <http://www.junglejapan.com/products/ruigo/ver3/index.html>

[北田 01] 北田純弥・萩原将文: 電子辞書を用いた比喻による文章作成支援システム, 情報処理学会論文誌, Vol.42, No.5, pp.1232 - 1241, (2001),

[笠原 01] 笠原健成・小林栄一・荒井真人・絹川博之: マニュアルの校閲作業における文書推敲支援ツールの実適用評価, 情報処理学会論文誌, Vol.42, No.5, pp.1242 - 1253, (2001),

[松尾 03] 松尾豊・友部博教・橋田浩一・石塚満: Web から人間関係ネットワークの抽出と情報支援, 第 17 回人工知能学会全国大会, 1F1-02, (2003).

[難波 01] 難波英嗣, 神門典子, 奥村学: 論文間の参照情報を考慮した関連論文の組織化, 情報処理学会論文誌, Vol.42, No.11, pp.2640 - 2649, (2001).

[相良 04] 相良直樹・砂山渡・谷内田正彦: 重要文抽出による文章からのストーリー抽出とシナリオ創発支援, 第 53 回人工知能基礎論研究会資料, pp.61 - 66, (2004).

[Salton 75] G. Salton, A. Wong, and C. Yang: A Vector Space Model for Automatic Indexing, Communications of the ACM, Vol.18, No.11, pp.613 - 620, (1975).

[砂山 99] 砂山渡・谷内田正彦: キーワード抽出に基づく文章作成支援システム, 第 37 回人工知能学会 SIG-FAI 研究会資料, pp.85 - 89, (1999).

[砂山 02] 砂山渡・谷内田正彦: 観点に基づいて重要文を抽出する展望台システムとそのサーチエンジンへの実装, 人工知能学会論文誌, Vol.17, No.1, pp.14 - 22, (2002).

[上田 94] 上田良寛・水梨豪・竹田幸史: 言語部品表現を用いた推敲支援, 情報処理学会自然言語処理研究会資料, No.103, pp.103 - 113, (1994).

[吉村 86] 吉村賢治・日高達・吉田将: 日本語科学技術文における専門用語の自動抽出システム, 情報処理学会論文誌, Vol.27, No.1, p.33 - 40, (1986).