

# 理工系学生の論理的文書作成支援を目的とした 論文可視化システム

松本 章代<sup>†</sup> 鈴木 雅人<sup>‡</sup> 市村 洋<sup>§</sup>

静岡大学大学院理工学研究科<sup>†</sup> 東京工業高等専門学校<sup>‡</sup> サレジオ工業高等専門学校<sup>§</sup>

## 1 はじめに

本研究は、理工系学生を対象とした、論文執筆のための教育システムの構築を目的とする。論理的な思考が必要なプログラムは書けるのに論理的な文章を書くことは苦手である、という理工系学生は多い。そこで、プログラミングの技術を論文執筆に活かす方法として、自然言語処理において意味表現として用いられる「意味ネットワーク」を仲介させる手法を考案した。意味ネットワークで表現される図は、オブジェクト指向を学んだ学生にとって容易に理解でき受け入れ易い。本システムは、学生が書いた論文を意味ネットワークによって可視化する。客観的に解析されることによって、論理的な文章へのヒントとなり、それを元に学生自身が考え直すことによって文章表現能力の向上を図る仕組みである。

## 2 従来の作文教育システムと本システムの特徴

米国では既に、複数の論文自動採点ソフトが開発され、その中の一つ Criterion(e-rater) [1]は、2001年の時点で全米の中学・高校を中心に200以上の機関で5万人の生徒に利用されている。日本でも、石岡ら[2]による日本語小論文自動評価採点システム「Jess」というソフトが e-rater とほぼ同じ原理で開発されている。これらの自動採点ソフトの仕組みは、上手な文章が持つさまざまなパターンを基準とし、その基準からどれだけ隔たっているかによって文章の上手でない度合いを評価するものである。

また、日本語作文小論文検定協会による日本語の文章解析ソフト「森リン」[3]は、「文章の上手さは、一般に語彙の豊富さと高い相関関係にある」とし、「語彙の量」や、文の長さに基づく「文章の持つリズム」を、数値化し評価する。

これら現在の文章添削システムは、文法の正しさ、語彙の豊富さ、文の長さ、漢字の量といった、表面的な指摘を行うものが主流である。これらは主に小中学生の作文を対象としており、高等教育機関における科学技術論文の執筆指導を目的とし

たものではない。

一方で近年、作文教育システムの分野には「文法」レベルに留まらず「意味」に踏み込もうという動きも存在する。学生の作文自体を解析・添削するのではないが、豊富な語彙・多様な表現を支援しようというものである。作文支援システムと意味ネットワークを結び付けた鄭ら[4]による研究では、コーパスから意味ネットワークを構築し、意味論的な側面から「連想」のサポートを行う。

このような従来の文法的な支援や語彙支援に対し本研究は、「著者本人が自分の意図が適切に読者に伝わる論理的な文章となっているかどうかを確認・修正を行う」ためのヒントを提供するものである。「文章」を理工系の学生にとって抵抗感の少ない「図」の形に可視化し、それを元に学生本人が考え直すことによって学生自身の文章表現能力の向上を見込めると考えられる。さらに、教員側の基本的な文章表現の指導といった余計な労力を減らし、高度な指導に専念できる。科学技術論文の執筆指導において、学生・教員双方にメリットがあるものと期待できる。

## 3 システム構成

### 3.1 入力から出力までの流れ

本システムは、利便性を考慮し現在ウェブアプリケーションとして構築している。システム自体は VineLinux 上で動作しており、開発言語は Ruby、日本語の係り受け解析には (株) CSK が開発したパーザを用いている。

システムの流れは以下ようになる。

- (1) 学生が書いた論文を入力データとし受け取る。
- (2) 論文を構成する各文について形態素解析・構文解析を行う。
- (3) (2)の結果から、重要と思われる語を抽出する。
- (4) (3)で抽出した語とさらに関連の深い語を抽出し、その語間の係り受け関係から、図を作成し出力する。

学生は、出力結果と論文の意図とを照合することにより、ミスマッチな部分を検出して問題点に気づく、という仕組みである。

### 3.2 可視化の手順

意味ネットワークの設計にあたり、できるだけ学生にとって馴染みのある「クラス図」に即した形とするため、「要求仕様からクラス化を行うためのステップ」[5]に沿って、論文を可視化する仕

Document Visualization System Aiming at Supporting that  
Science and Engineering Students Write Logical Papers

<sup>†</sup> Akiyo MATSUMOTO, Shizuoka University

<sup>‡</sup> Akiyo MATSUMOTO, Masato SUZUKI,  
Tokyo National Collage of Technology

<sup>§</sup> Hiroshi ICHIMURA, Salesian Polytechnic

組みを考案した。

- **ステップ1：論文から名詞と動詞を抽出する。**
  - (1) 文書全体を形態素解析し、文単位で TF-IDF を求め上位 50 語を抽出する（ひらがなのみの語と記号は除去）。
  - (2) (1)に該当する語を含む文について構文解析を行い、その結果から「(1)に係っている語」と「(1)に係っている語」を抽出する。
- **ステップ2：主語を中心に目的語と動詞を登録する。**
  - (1) 抽出した名詞ごとに動詞をまとめる。
  - (2) 主格として使われた名詞とそれ以外とで分ける。
  - (3) ある主語に特有に含まれる目的語を、その主語と結び付け“has-a”関係として登録する。
  - (4) 主語に含まれる目的語が決まれば、それを関係付けている動詞を主語に登録する。
- **ステップ3：代表的な主語を選び、階層化する。**
  - (1) 複数の主語の中から、共通の動詞を有する主語をグループ化する。
  - (2) グループ化された類似する主語について概念階層辞書と照らし合わせ、一般的なものから特殊なものへと順番をつけて階層化する。（“is-a”関係）
  - (3) 共通する語を1つにまとめ、全体を統合したグラフにする（図1）。

なお、ステップ2及び3において「する」「なる」「ある」といった非常に一般的な動詞については、同一文のみノードをまとめるものとする。

#### 4 動作実験

3節で示したアルゴリズムに則りシステムを構築し、実際に学生が書いた論文<sup>1</sup>をサンプルデータとして動作実験を行った。すると、主語と目的語、それが係る動詞のみを抽出することになるため、各文に共通する語が少なく語同士が思うように連結しないという問題点が生じた（図1）。そこで、主語・目的語に限定せずキーワードの係り受け関係を表現するように変更した（図2）。なおその際、係り方の違い（主語、目的語、並列、といった関係）によって、エッジを色分けする。

#### 5 おわりに

本稿では、理工系の学生を対象とした論文執筆のための教育システムとして、「学生が書いた論文を可視化して客観的に解析することにより、書いた本人に自分の論文の問題点を気づかせるシステム」を提案した。

今回実装したシステムについては、現在のところ



図1 ステップ3



図2 実行例

る以下の点の対応を検討している。

- 接続助詞や助動詞などの細かいニュアンスをどう表現するか。
  - 章（節）の遷移をどう表現するか。
- また今後は、実際に学生が論文を書く場面において、このシステムを使用してもらい、
- (1) 出力結果は入力した論文を適切に図示したものとなっているか。
  - (2) 出力結果はその論文の修正に役に立ったか。
  - (3) それ以降の論文の書き方に何か影響はあるか。
- について検証を行う予定である。

#### 謝辞

本研究は文部科学省科学研究費補助金（若手B, 課題番号18700664）の交付を受けている。

#### 参考文献

- [1] <http://criterion1.ets.org/cwe/>
- [2] 石岡 恒憲, 亀田 雅之: コンピュータによる小論文の自動採点システム Jess の試作, 計算機統計学, Vol. 16, No. 1, pp. 3-18, (2003).
- [3] <http://www.mori7.info/moririn/index.php>
- [4] 鄭在玲, 三宅真紀, 畑中伸幸, 赤間啓之: 反復クラスタリングによる意味ネットワークに基づく作文支援システムの開発, 情報処理学会研究報告, Vol. 2005, No123, pp. 99-105.
- [5] 山野修・大山恭弘: C++道工具箱, 日経 BP 社, (1993).

<sup>1</sup>今回は、ある高専専攻科2年生の書いた学修成果レポートの1章をデータとして利用した。