

計量テキスト分析を基にした

テクニカルライティング支援システムの開発

Development of Technical-Writing Support System Based on Quantitative Analysis of Text

吉川 洋希[†] 鈴木 孝幸[‡] 納富 一宏^{†‡}
 Hiroki Yoshikawa Takayuki Suzuki Kazuhiro Notomi

1. はじめに

現在、多くの大学で文章表現技術に関する講義が広く開講されており、実践的な内容による演習を通して基本的な文章作成技術の習得を目指す場合もある。また、キャリア教育の一環としての実施や企業の新人研修においても、報告書や議事録など、ビジネス文書を作成する上で正確な情報伝達を行うための技能を身に付けることが重要な課題となっている。

理工系の大学生にとっての文書作成技術の習得は、実験レポート・卒業論文の執筆や、就職活動における履歴書やエントリーシートあるいはビジネスメールを作成するために必要な重要な課題である。

実験レポートなどの技術文章は簡潔・明解・論理的に書くことが求められるが^[1]、教員が講義において全ての学生に対して知識量に合わせた適切なフィードバックを行うことは、負荷が大きく困難である場合が多い。

以上の背景を踏まえて、本稿では論文の執筆経験がない理工系の学生を対象に、文書作成支援を目的とした個別学習システムの開発について述べる。

2. 提案システム

提案するシステムは、文章の計量分析の結果を元にユーザに対して逐一修正を促し、段階を踏んで論文の序論部分を完成させる過程を体験させるものである。文章作成支援に関する関連研究としては、リスト化した技術文書の規則を適用し、文章中の不適切な表現を自動で検出するものや^[2]、ルールベースに基づく校正と係り受け解析を利用した推敲支援を行うツールの開発^[3]など数多くの研究がなされている。また、文章中の漢字や品詞、読点の使用率などに着目した研究としては性別を偽装して文章を書いた際の文体的特徴の変化を検証したものなどがあるが^[4]、論文の文体的特徴について検討した研究は少ない。

システムによって学生の文書作成時の内省を促し、序論部分の各段落を規律正しく、徐々に内容を膨らませる形で書かせること、それに伴い教員の添削作業の負荷を軽減することを目的とする。

2.1 卒業論文序論における文章構成要素について

大枠として、本学情報学部卒業論文の第 1 章は、「序論」「はじめに」「はしがき」などのタイトルで記述され、節に分割することなく、1300~2000 文字程度の分量となる。構成要素となる段落（パラグラフ）は、①テーマの現状を

説明し問題点を提起する「研究背景」、②研究で明らかにすべき内容や目標とする到達点を表明する「研究目的」、③研究の進め方や方針の骨子となる「研究方法」、④最終的に導かれるべき「到達点」、⑤当該論文の章立てを説明する「論文の構成」の 5 つに大別することができる。

2.2 システム概要

Step1

まず、ユーザである学生がシステムからの入力指示に従って、文章入力画面に文章を入力する。最初の指示内容として「はじめに」を構成する段落のうち、どの部分を書くかを示す。また、空白から文章を作成するのは困難であるため、例文などのヒント提示を行う。

Step2

文章の入力が完了すると、システムが文章の計量分析を行う。文章全体の文字数や読点数、漢字の使用率などの文体的特徴を数値として表現し、評価基準とする標準値データとの比較を行う。その結果から修正アドバイスを作成し、ユーザに提示する。

Step3

提示された修正アドバイスを基にして、ユーザは文章の編集（加筆・修正）を行う。その後、Step2 と同様に文章の計量分析と比較を繰り返し、設定した水準に達するまで文書作成を繰り返す。最終的に「はじめに」を完成させる。提案するシステムの処理の流れを図 1 に示す。

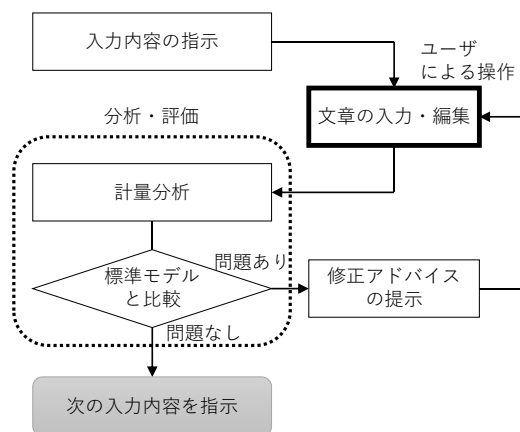


図 1 システムの処理の流れ

2.3 文章修正に関するアドバイスの提示例

2.1 で示した「序論」の 5 つの文章構成要素に対応する形式で、ユーザがパラグラフ単位で作成した文章に対して、2.2 で示した「入力・編集」→「分析・評価」→「修正ア

[†] 神奈川工科大学大学院 Graduate School of Engineering, Kanagawa Institute of Technology

[‡] 神奈川工科大学情報工学科 Dept. of Info. & Comp. Sciences, Kanagawa Institute of Technology

ドバイス提示」という修正サイクルの中で、ユーザに編集を促すための修正アドバイスの例を表 1 に示す。

表 1 修正アドバイスの例

問題点	修正アドバイス
[1] 文の数が少ない	パラグラフを構成する文の数が標準値より少なくなっています。より詳しい説明を記述するために、省略されていると考えられる部分について追加してください。①具体例を示す、②専門用語・術語を補足する、③カタカナ表記には英語も併記する、④参考文献から引用する方法を試してみてください。
[2] 読点が多い	文中の読点数が標準値より多くなっています。過剰な読点は文章を読みにくくする原因になります。文をゆっくり音読して区切りが多いと感じる部分の読点を削除してみてください。
[3] 漢字の使用率が低い	漢字表記が標準値より少なくなっています。名詞、形容詞、動詞などの自立語は漢字表記しましょう。

3. 「はじめに」の計量分析結果

図 1 の分析・評価における「比較」を行うための評価基準とする標準値データを作成するために、本学情報学部の卒業生 12 名分の卒業論文の第 1 章「はじめに」の計量分析を行った。分析対象とした文体的特徴の一覧を表 1 に示す。分析ツールとしては KH Coder を使い、一部については目視で計量を行った。

表 2 計量分析を行う項目

No.	計量項目	No.	計量項目
1	文字数	9	半角英数字
2	文の数	10	漢字率
3	読点の数	11	ひらがな率
4	一文中の平均文字数	12	カタカナ率
5	一文中の平均読点数	13	動詞 (自立)
6	最短文	14	動詞 (非自立)
7	最長文	15	形容詞 (自立)
8	中央値	16	形容詞 (非自立)

分析項目は文献[1]を参考に設定した。

分析結果のうち、例として、文字数、文章中の文の数、文章中の動詞 (自立語) の数、漢字の使用率に対するヒストグラムをそれぞれ図 2~図 5 に示す。

文字数

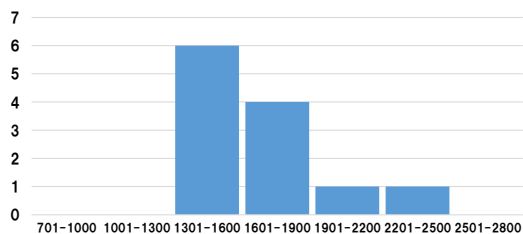


図 2 文字数の度数分布図

分析対象とした卒業論文の「はじめに」はいずれも 1 ページ強の分量で、2000 字を超えるものも見られたが、ほとんどは 1300 字から 1600 字程度の分量だった。

文の数

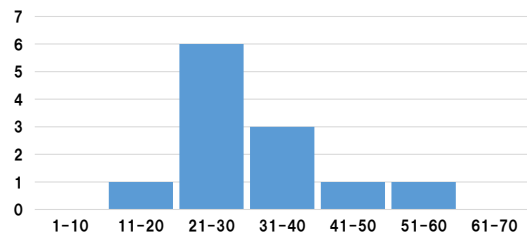


図 3 文の数の度数分布図

文字数の分布と同様に、文の数の分布にも特徴がみられた。

動詞(自立語)の数

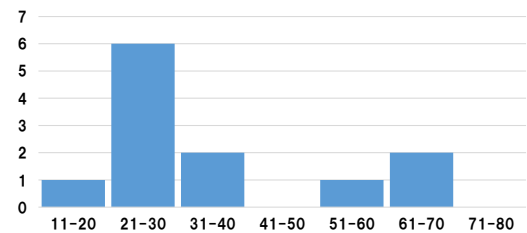


図 4 動詞 (自立語) の度数分布図

動詞 (自立語) の数は 21-30 個に集中した。動詞の増減を指示するアドバイスには工夫が必要だと考えられる。

漢字率

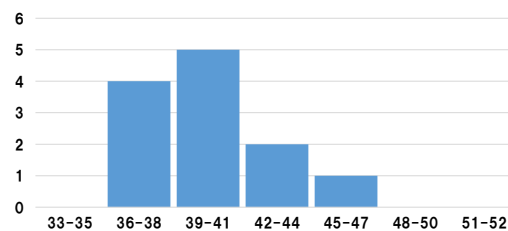


図 5 漢字率の度数分布図

分析したカタカナ率・ひらがな率と同様に、扱うテーマによって大きく影響を受けるため、単純な使用率では指標として扱うのは困難だと考えられる。

4. 結論・今後の展望

本学情報学部の卒業論文の第 1 章「はじめに」の計量分析を実施し、文書修正の際の指標としての利用を検討した。結果からテーマによって特徴が異なり、単体では指標としての利用が難しい文体的特徴を確認できた。今後は分析するデータ量を増やし、修正アドバイスの提示の妥当性の検証および内容の充実化を進める予定である。

参考文献

- [1] 黒木 登志夫, “知的文書とプレゼンテーション-日本語の場合、英語の場合”, 中公新書 (2011).
- [2] 坂本 俊介, 須藤 嵩志, 丸山 広, 中村 太一, “形態素解析を利用した文章校正手法の提案”, 情報処理学会研究報告デジタルドキュメント, Vol.2009-DD-72, No.17 (2009).
- [3] 大野 博之, 稲積 宏誠, “文章作成能力の育成を目指した教育支援ツールの開発”, 電子情報通信学会第 18 回データ工学ワークショップ DEW2007, D9-4 (2007).
- [4] 財津 亘, 金 明哲, “性別を偽装した文章における文体的特徴の変化”, 同志社大学ハリス理化学研究報告, 59(3), 181-188, (2018).