

レポート作成指導のための統計データ言語表現のデータベース化の試み

An Approach to Constructing a Database of Example Sentences that Explains Statistical Data for Report Writing Instruction

吉根 勝美^{*1}

Katsumi YOSHINE^{*1}

^{*1} 南山大学経済学部

^{*1} Faculty of Economics, Nanzan University

Email: kyoshine@nanzan-u.ac.jp

あらまし：数値データを根拠とするレポート作成では、データをグラフで表現したり、適切に要約した文章を書いたりする必要がある。公的な統計データを利用してレポートを作成する課題を大学生に課すとき、グラフ作成のための表計算ソフトの操作指導はよく行われるが、データの文章表現の指導方法が確立されているとは言いがたい。本研究では、統計データの言語表現をデータベース化して、これをレポート作成指導に役立てる方法を検討する。

キーワード：レポート作成、統計データ、データの言語化、穴埋め問題

1. はじめに

大学の授業では、理解度の確認や成績評価のために学生にレポートを課すことが多い。このとき、一般的に行われる指導内容は、テーマの見つけ方やレポートの構成の仕方、さらに、剽窃は許されないというような倫理的な注意事項であることが多い。

筆者が所属する経済学部では、統計データを根拠としたレポートを課すことが多いので、データを適切に要約した文章を書くという指導も必要となる。この指導の目標は、経済紙誌などに掲載されている統計表やグラフから読み取ったことを、文章として適切に表現できるようにすることである。

ただし、データからテキストを自動生成する研究は従来から行われている。例えば、小林ら⁽¹⁾は、日経平均株価の動向を説明するテキストを、動的に生成するシステムを開発している。このシステムでは、同期間のニュースを要約したり、チャートの形状を認識して言語化したりする機能を有している。また、加藤ら⁽²⁾は、テキスト中に現れる時系列情報の値やその変化に関する言語表現コーパスを構築している。さらに、日本経済新聞社が2017年1月25日に配信開始した『決算サマリー』の記事は、上場企業の開示資料や決算データから人工知能が文章を作成し、配信まで完全自動化されている⁽³⁾。

このように、データからテキストの自動生成は一部実用化も進んでいるが、大学卒業時に求められる学士力の一部をなす汎用的技能を向上させるために、統計データを活用して根拠としたレポートを書くという作業は相変わらず必要であろう。しかし、この指導方法が確立されているとは言いがたい。そこで、本稿では、実際の授業で筆者が行っている指導例を

もとにして、統計データに関わる文章をデータベース化して、これをレポート作成指導に役立てる方法を検討する。

2. 統計データを含むレポートの指導

経済統計データを根拠とするレポート作成指導では、データに簡単な分析を施して、グラフを作成するという表計算ソフトの操作手順の指導が主となり、統計データを適切に要約した文章を書くという指導は、どうしても後回しになる。本章では、統計データを文章で表現する事前練習として作成した穴埋め問題を紹介する。

農林水産省刊行『平成27年度 食料・農業・農村白書』の99ページに記載されているグラフと文章を題材にして作成した穴埋め問題を図1に示す。実際には、WebClassのアンケート機能を使って出題した。

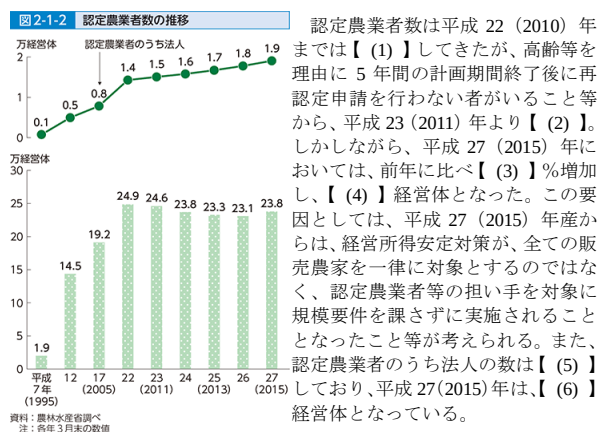


図1 穴埋め問題の例

この問題の(3),(4),(6)には、グラフから読み取った数値、あるいは増加率の計算結果を回答する。

(1),(2),(5)には、白書の原文通りなら「一貫して増加」あるいは「減少に転じた」が正解となるが、単に「増加」「減少した」と書いても、統計グラフの説明として間違いではない。ただし、データを適切に要約するのなら、単なる増加・減少ではなく、データの動向をより適切に表現するような語彙を使いこなせるような指導が必要である。

3. レポート指導支援のシステム化

2章で示した穴埋め問題の題材は、教員が実際に目を通して探しているのが現状である。このままでは、作成できる問題量が圧倒的に不足するため、統計データに関わる文章をデータベース化することを検討する。

各省庁が刊行する白書、年次報告等には、統計表やグラフが豊富に掲載されており、それを説明する文章が書かれている。これらの白書類のほとんどはウェブで公開されているので、これをデータベース化し、レポート指導支援のシステム化を図りたい。

データベースの構成を図2に示す。白書に含まれる文章は文章データベースに、図表はグラフデータベースにそれぞれ格納する。また、グラフと文章の対応表を別途作成する。さらに、統計データの説明文に特有な表現集を随時蓄積していく。例えば、前章の穴埋め問題の指導例で示したように、「一貫して増加」、「減少に転じた」のような表現を、表現集データベースに格納しておく。

中小企業庁刊行『2016年版中小企業白書』を例に試作しているデータベースで、「転じ」をキーワードとして検索すると、図3の検索結果が得られる。

この検索結果を利用した指導例として、以下が考えられる。

- ・検索結果を学習者に提示して、「転じ」の用例を読ませる。
- ・「転じ」た時期を空欄にした穴埋め問題を作成する。
(例)【 】より再び減少に転じた。
- 【 】から全業種で不足に転じて
- ・「転じ」と合わせて使う用語を穴埋め問題にする。
(例)2012年より再び【 】に転じた。

2013年第4四半期から全業種で【 】に転じて

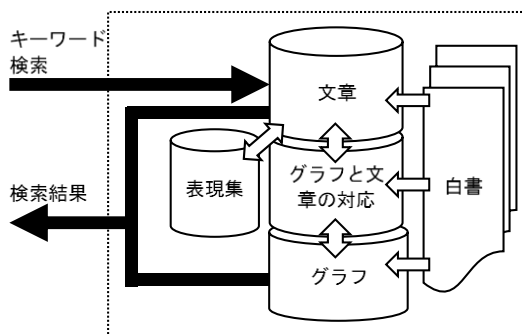
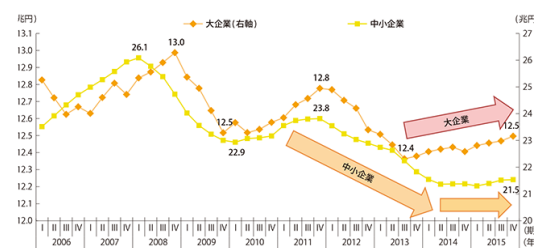


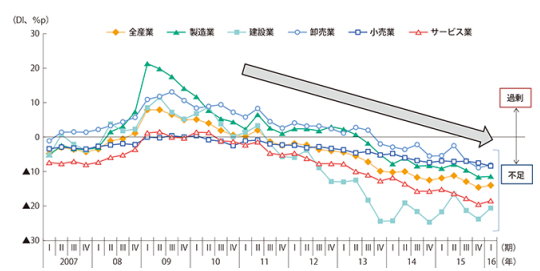
図2 データベースの構成の概要

第1-2-18図 人件費の推移（規模別）



人件費の推移を規模別に見ていくと、大企業・中小企業共に2008年後半から2010年前半にかけて低下し、翌年に上昇したものの、2012年より再び減少に転じた（第1-2-18図）。

第1-2-20図 中小企業の従業員数過不足DIの推移（業種別）



中小企業の従業員過不足DIの推移を確認すると、2013年第4四半期から全業種で不足に転じており、特に建設業、サービス業では不足感が強い（第1-2-20図）。

図3 検索結果の例

4. おわりに

本稿では、大学生に対して、統計データを根拠としたレポートを課すときに、データを適切に要約する指導方法を確立することを目的として、統計データに関わる文章を多く含む白書をデータベース化して、これをレポート作成指導に役立てる方法を検討した。

昨今は、人工知能が作成した新聞記事が見られるようになっているものの、統計データをよく観察して、特徴を見つけ出し、レポートの文章にまとめる能力自体は、人工知能の進展に関わらず、大学生・社会人に必要であろう。本研究では、大量の統計データとその文章表現をデータベース化し、大学生にその能力を育成するための指導を支援することを目標としている。

今後は、このデータベースを活用して、穴埋め問題の自動生成や、学生が書いた文章の評価への応用について検討したい。

参考文献

- (1) 小林一郎, 渡邊千明, 奥村奈穂子: “グラフとテキストの協調による知的な情報提示手法—日経平均株価テキストとグラフの提示を例にして”, 情報処理学会論文誌, vol.48, no.3, pp.1058-1070 (2007)
- (2) 加藤恒昭, 松下光範, 神門典子: “時系列情報の値と変化に関する言語表現コーパスの構築”, 人工知能学会論文誌, vol.25, no.5, pp.637-650 (2010)
- (3) 日本経済新聞社, 言語理解研究所, 東京大学: “完全自動「決算サマリー」ベータ版(試用版)”, <http://pr.nikkei.com/qreports-ai/> (2017年5月29日確認)