

LMS 向け日本語テキストマイニング支援ツール TeMP の開発と運用

梅村 信夫¹⁾, 山本 恵²⁾

1) 名古屋学芸大学 ICT 活用教育推進室

2) 名古屋外国語大学 現代国際学部

d_chaser@nuas.ac.jp

概要：著者らは、LMS（学習管理システム）に蓄積された日本語文書から有用な知識を抽出するため、簡便性を重視した解析支援ツール TeMP の開発を継続的に進めている。TeMP は、LMS に配置するデータ入力部と、クラウド型アプリケーションとして機能するデータ解析・出力部から成る。本稿では、後者が有する諸機能と要素技術を概説するとともに、授業での運用事例を考慮しつつ TeMP の有用性について検討した。

1 はじめに

LMS（学習管理システム）には、大量の学習履歴が日々刻々と蓄積される。中には日本語テキスト（例えば学生が提出したレポートや、形成的授業評価アンケートの自由記述文）が存在するが、これらを解析するための環境が十分に整備されているとは言い難い。

そこで著者らは、LMS におけるテキストマイニング（日本語テキストからの有用な知識抽出）を支援する、簡便性を重視した解析支援ツールの開発を進めてきた[1][2]。

本研究では、開発の最中にあるツールの諸機能・要素技術について概説し、進行中の授業科目における利活用実態に照らしつつ、e ラーニングへの適用性について検討する。以下の文中においては、開発ツールを TeMP (Text Mining Plugin) と称する。

2 TeMP の諸機能

2.1 構成要素

TeMP は、①データ入力部（LMS のプラグイン）と②データ解析・出力部（クラウド型アプリケーション）から成る。本研究では、代表的な LMS の Moodle に①を実装した。また②は、LAMP (Linux, Apache, MySQL and PHP)、形態素解析器 MeCab [3]、そして統計計算ソフト R [4]とそのパッケージ RMeCab [5]を使い構築

した。

①と②は、クッキーによってセッションを共有しながら、「データの取得・検査・整形」と「データ解析・可視化」の役割をそれぞれが担う。

2.2 解析の要素技術

Moodle に実装したプラグインは、データ入力用のブロック（プラグイン）として機能する（図 1）。

The screenshot shows the 'TEMP: 日本語テキストマイニング' Moodle block interface. It is divided into three main sections: 1. '【外部ファイルの参照】' (External File Reference) with instructions to upload a file and buttons for 'アップロード' (Upload) and 'リセット' (Reset). 2. '【データの直接入力】' (Direct Data Input) with instructions to click a button to open an input window and a button labeled '入力ウィンドウを開く' (Open input window). 3. '【Moodle DB から取得】' (Retrieve from Moodle DB) with instructions to select a module type and a button labeled 'DB から一覧を取得' (Get list from DB). The module types listed are: 'オンラインテキスト' (selected), 'データベース', 'Wiki', and 'Feedback'.

図 1 データ入力用 GUI (Moodle ブロック)

Moodle ユーザーがブロックを操作すると、プラグインはデータベース等から日本語テキストを取得し、検査・整形の過程を経てクラウド

型アプリケーションへ送信する。アプリケーションはデータを受け取ると、形態素解析によって文書から単語（索引語）を抽出する。抽出時には、日本語 Wikipedia・はてなキーワードを登録した MeCab のユーザー辞書によって、複合語を認識する。また、Zipf の法則によるフィルタリング（不要語の自動削除）を適宜施す。

索引語の抽出が終わると、「索引語-文書行列」などの各種文書ベクトルを生成する。一部のベクトル生成においては、LSA (潜在的意味分析) によって、同義語を考慮しつつ次元縮退（索引語の集約）を行う場合もある。

処理の最終段階では、多変量解析の手法を用いてベクトルの持つ多次元情報を 2 次元に縮約し、表組み（テーブル）あるいはグラフィックスによって解析結果を可視化する。

クラウドアプリケーションが出力するテーブルとグラフィックスを、操作メニューと共に図 2 に例示した。

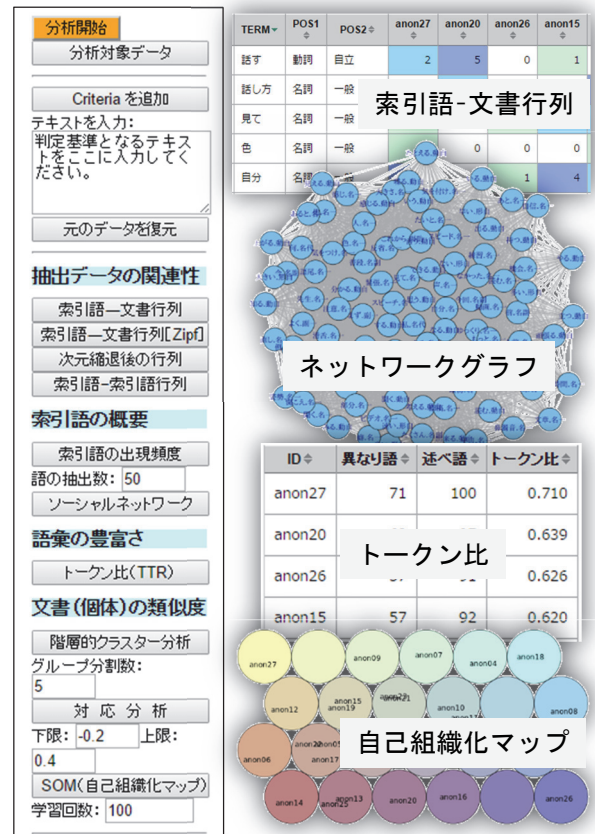


図 2 データ解析・出力用 GUI と出力例 (クラウド型アプリケーション)

表 1 解析結果可視化の手法

種別 [表示形式]	可視化の手法	概要
索引語の関連性 [テーブル]	索引語-文書行列 (Zipf 適用可)	文書から抽出した索引語の出現頻度行列表を提示。Zipf の法則を任意に適用することで、不要語の自動削除が可能。
	索引語-文書行列 (LSA 適用)	
	索引語-索引語行列	
	索引語-索引語行列 (Zipf 適用可)	
索引語の要約 [グラフィックス]	バーチャート (Zipf 適用可)	索引語の出現頻度を可視化。
	ワードクラウド (Zipf 適用可)	
	ネットワークグラフ (Zipf 適用可)	
語彙の豊富さ [テーブル]	トークン比	索引語の異なり数と延べ数から、語彙力の指標を算出。
	ユールの K 特性値	
文書および索引語の類似度 [グラフィックス]	デンドログラム	階層的クラスター分析から得た文書間・索引語間の距離を樹状図で表現。
	対応分析布置図	対応分析 (コレスポンデンス分析) から得た成分値に従い、類似の文書・索引語を座標平面の近傍に布置。
	自己組織化マップ	ニューラルネットワークモデルにもとづき、類似の文書・索引語をグリッドの近傍に布置。

※Zipf: Zipf の法則にもとづく索引語のフィルタリング LSA: 潜在的意味分析

3 授業での利活用

2015 年度第 2 セメスター開講の情報リテラシー科目（2 クラス編成，各クラス受講者数 19 名・23 名のブレンド型授業）において，TeMP の運用を開始した．適用対象は，毎時実施する形成的授業評価の自由記述文である．

具体的な活用方策として，以下 2 点を実施している：①TeMP が出力したワードクラウドをスライド資料にまとめ，毎時の冒頭で学生に提示しながら前回の授業内容を振り返る，②トークン比を手がかりに，記述文の内容改善を求めるフィードバックコメントを該当学生に送信する．

①については，授業評価の内容を教員が毎時精査している証しとして，また授業に対する他学生全体の意見・提案をブラウズする手段として，学生は好意的に受け止めているようである．

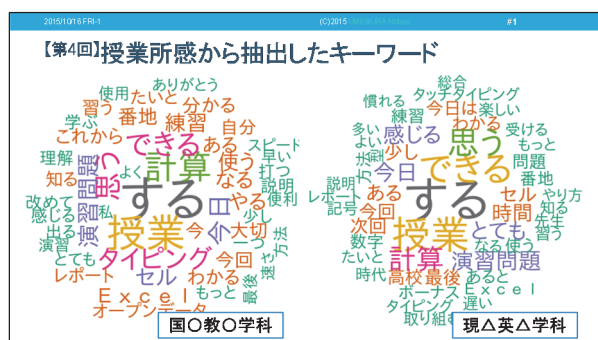


図 3 授業用資料内でのワードクラウド活用例

「語彙の豊富さ」を示すトークン比・ユールの K 特性値については，情報リテラシー以外の科目（例えばゼミナール）において，採点済みレポートの解析に活用している．

以上，限られた授業科目の範囲内でも，TeMP の活用場面を見いだせる．表 1 に掲げた諸機能を組み合わせることで，利活用の機会はさらに高まるだろう．以上より，機能に関するアンケートの実施など，客観的評価の蓄積は必要であるが，取り上げた事例から TeMP の有用性は高いと推測する．

4 おわりに

TeMP は，簡易な支援ツールとして，一次スクリーニングを目的に設計している．つまり，大まかな知見を得た後，必要に応じ詳細な分析を高機能な専用ツールに委ねる，という考え方である．そのため，サーバースペックに対する敷居も低く，文献[2]で実施したサーバー負荷テストを含む性能評価試験から，50 名の教員が同時接続して使用する場面を想定しても，遅滞などの問題が発生しない点を確認している．

上述した性能評価と授業での運用事例を通じて，細部に課題を抱えるものの，TeMP は機能的に実用水準に達したと考える．

ところで，レポート評価に関する諸機能の充実化に対しては，教員からの要望が特に多い．このようなニーズに照らし，機械学習にもとづく自動採点支援システムの実装を計画中である．

参考文献

- [1] 梅村信夫、山本恵、学習履歴テキストからの知識抽出を支援する Moodle プラグインの開発、2013 年度学術講演会発表論文集、2013.
- [2] 梅村信夫、山本恵、教師用テキストマイニング・プラグインの開発と評価、日本ムードル協会全国大会発表論文集、vol.3、pp.59-64、2015.
- [3] 工藤拓、山本薫、松本裕治、Conditional Random Fields を用いた日本語形態素解析（解析）、情報処理学会研究報告、自然言語処理研究会報告、vol.2004、no.47、pp.89-96、2004.
- [4] R Core Team, R: A Language and Environment for Statistical Computing. Vienna, Austria, 2015.
- [5] M. Ishida, RMeCab: interface to MeCab. 2014.