

中国古典文学作品に対する虚詞出現率によるクラスタリングを用いた特徴分析支援システムの開発

油野 紗也加¹ 宮原 哲浩^{2,a)} 黒河 康平¹ 林 哲郎¹ 鈴木 祐介^{2,b)} 内田 智之^{2,c)} 久保山 哲二³
松本 哲志⁴ 東 英寿⁵

概要：中国古典文学作品に対する虚詞出現率によるクラスタリングを用いた特徴分析支援システムについて報告する。虚詞の出現場所に注目し、専門家が指定した虚詞出現率、距離、クラスタリング手法で特徴分析を支援するシステムを開発した。唐宋八大家散文の分析を行った実験結果についても報告する。

キーワード：中国古典文学作品、特徴分析支援、クラスタリング

Development of a Feature Analysis Support System using Clustering based on the Frequency of Function Words for Chinese Classical Literary Works

SAYAKA YUNO¹ TETSUHIRO MIYAHARA^{2,a)} KOUHEI KUROKAWA¹ TETSURO HAYASHI¹
YUSUKE SUZUKI^{2,b)} TOMOYUKI UCHIDA^{2,c)} TETSUJI KUBOYAMA³ SATOSHI MATSUMOTO⁴
HIDETOSHI HIGASHI⁵

Abstract: We report a feature analysis support system using clustering based on the frequency of function words for Chinese classical literary works. Focusing on the place where function words appear, we have developed a feature analysis support system that uses the frequency of function words, distance measures, and clustering methods specified by experts. We also report experimental results analyzing prose works of the Eight Great Prose Masters of the Tang and Song.

Keywords: Chinese classical literary works, feature analysis support, clustering

1. はじめに

文章から著者の文体の計量的な特徴を抽出し、その分析によって問題の解決を試みる研究が行われている [1], [2]. 文章の計量的な分析を行う場合には、どのような特徴情報を用い、どのような方法で処理を行うかが重要である。日本語の助字・助動詞・接続詞等の単独で意味を有さない字を総称して、漢文では虚詞または助字という。本研究では中国古典文学作品を対象として、虚詞の出現場所に注目し、専門家が指定した虚詞出現率、距離、クラスタリング手法で特徴分析を支援するシステムを開発した。唐宋八大家とは、中国の唐代の韓愈、柳宗元、宋代の欧阳脩、蘇洵、曾

¹ 広島市立大学情報科学部
Faculty of Information Sciences, Hiroshima City University
² 広島市立大学情報科学研究科
Graduate School of Information Sciences, Hiroshima City University
³ 学習院大学計算機センター
Computer Centre, Gakushuin University
⁴ 東海大学理学部
Faculty of Science, Tokai University
⁵ 九州大学比較社会文化研究院
Graduate School of Social and Cultural Studies, Kyushu University
a) miyares21@hiroshima-cu.ac.jp
b) y-suzuki@hiroshima-cu.ac.jp
c) uchida@hiroshima-cu.ac.jp

鞏, 王安石, 蘇軾, 蘇轍の8人の作家を指す。句頭または句末に出現する虚詞に注目して, 唐宋八大家の記と序の作品を分析する実験を行い, 特徴分析支援システムの有用性を確認した。

2. 中国古典文学作品に対する虚詞出現率によるクラスタリングを用いた特徴分析支援システム

2.1 特徴分析支援システム

中国古典文学作品に対する虚詞出現率によるクラスタリング [3] を用いた特徴分析支援システムをデータ解析ソフトウェア R で開発した。本特徴分析支援システムの機能は以下の通りである。分析の対象とする1つの作品データが1つのテキストファイルで管理されているとする。分析の対象とする作品名(ファイル名)の一覧をテキストファイルで指定する。出現率を調べる対象とする虚詞の一覧をテキストファイルで指定する。それぞれの虚詞が句頭または句末にあるときだけ出現を数えるのか, または全ての出現を数えるのかを指定することができる。以下で説明する2つの虚詞出現率(全文字数出現率, 全虚詞数出現率)のどちらを使うのかを指定する。作品間の距離として, 5つの距離(標準化ユークリッド距離, キャンベラ距離, マンハッタン距離, 最大距離, バイナリ距離)のどれを使うのかを指定する。最終クラスター数を指定する。階層的クラスタリング手法として, 6つの手法(最長距離法, 最短距離法, ウォード法, 群間平均法, 重心法, メディアン法)で分析が行われる。特徴分析支援システムによるクラスタリングの実行例を図1に示し, 出力される樹形図の例を図2に示す。

2.2 中国古典文学作品を対象とする分析

中国古典の散文の特色は虚詞の用いられ方にあるという劉德清『欧陽脩論稿』の見解がある [4]。虚詞の考察は中国古典散文作品を分析する上で, 有意義な視点になる。本特徴分析支援システムでは, 句頭または句末に出現する虚詞に注目し, 特徴分析を行うことができる。対象とした中国古典散文のテキストデータでは, “,” , “、”, “.” , “。” , “!” , “?” , “:” , “;” の記号で句が終わるため, 句頭とは文の最初とこれらの記号の直後, 句末とはこれらの記号の直前のことである。

2.3 虚詞出現率によるクラスタリング

クラスタリングとは, 作品(個体)の集合をいくつかのクラスター(部分集合)に分割し, それぞれのクラスターの中で作品間の距離が小さくなるようにすることである。作品の特性を示す変数の数を p , 作品 i の k 番目の変数の値を x_{ik} , 作品 i と作品 j 間の距離を d_{ij} で表す。次節で説明する実験結果で使用された, 作品 i と作品 j 間のマンハッタン

> source("jikko.R")
0:全虚詞数出現率, 1:全文字数出現率->0
1:バイナリ距離, 2:キャンベラ距離, 3:ユークリッド距離, 4:マンハッタン距離, 5:最大距離->4

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

Console Terminal Jobs
//mac/Home/Desktop/実数YUNQ/

[1] "cluster = 2"
call:
hclust(d = typeDist, method = m[i])
cluster method : ward.D2
distance : manhattan
Number of objects: 491
[1] "joi-manhattan"
[1] "ward.D2"

[1] "clust.number = 1"
[13] "鶴堂-記-01.txt" "鶴堂-記-03.txt" "鶴堂-記-04.txt" "鶴堂-記-05.txt" "鶴堂-記-08.txt"
[6] "鶴堂-序-11.txt" "鶴堂-序-15.txt" "鶴堂-序-19.txt" "鶴堂-序-22.txt" "鶴堂-序-24.txt"
[11] "鶴堂-序-26.txt" "柳宗元-記-03.txt" "柳宗元-記-05.txt" "柳宗元-記-10.txt" "柳宗元-記-11.txt"
[16] "柳宗元-記-12.txt" "柳宗元-記-24.txt" "柳宗元-記-26.txt" "柳宗元-記-38.txt" "柳宗元-序-01.txt"
[21] "柳宗元-序-29.txt" "柳宗元-序-33.txt" "柳宗元-序-38.txt" "柳宗元-序-41.txt" "柳宗元-序-46.txt"
[26] "柳宗元-序-04.txt" "柳宗元-序-08.txt" "柳宗元-序-10.txt" "柳宗元-序-20.txt" "柳宗元-序-21.txt"
[31] "柳宗元-序-22.txt" "柳宗元-序-23.txt" "柳宗元-序-24.txt" "柳宗元-序-27.txt" "柳宗元-序-38.txt"
[36] "柳宗元-序-39.txt" "柳宗元-序-41.txt" "柳宗元-序-43.txt" "柳宗元-序-44.txt" "柳宗元-序-46.txt"
[41] "柳宗元-序-49.txt" "柳宗元-序-55.txt" "柳宗元-序-56.txt" "歐陽修-記-01.txt" "歐陽修-記-02.txt"
[46] "歐陽修-記-03.txt" "歐陽修-序-29.txt" "歐陽修-序-30.txt" "歐陽修-序-46.txt" "歐陽修-序-47.txt"
[51] "歐陽修-記-08.txt" "歐陽修-記-09.txt" "歐陽修-記-10.txt" "歐陽修-記-13.txt" "歐陽修-記-14.txt"
[56] "歐陽修-記-15.txt" "歐陽修-記-16.txt" "歐陽修-記-17.txt" "歐陽修-記-18.txt" "歐陽修-記-19.txt"
[61] "歐陽修-記-22.txt" "王安石-記-04.txt" "歐陽修-序-31.txt" "歐陽修-序-32.txt" "歐陽修-序-33.txt"
[66] "歐陽修-記-29.txt" "歐陽修-記-30.txt" "歐陽修-記-32.txt" "歐陽修-記-36.txt" "歐陽修-序-01.txt"
[71] "歐陽修-記-40.txt" "歐陽修-記-41.txt" "歐陽修-記-43.txt" "歐陽修-序-02.txt" "歐陽修-序-03.txt"
[76] "歐陽修-序-06.txt" "歐陽修-序-07.txt" "歐陽修-序-18.txt" "歐陽修-序-21.txt" "歐陽修-序-28.txt"
[81] "歐陽修-序-23.txt" "歐陽修-序-24.txt" "歐陽修-序-25.txt" "歐陽修-序-27.txt" "歐陽修-序-28.txt"
[86] "歐陽修-序-34.txt" "歐陽修-序-35.txt" "歐陽修-序-36.txt" "歐陽修-序-37.txt" "歐陽修-序-38.txt"
[91] "歐陽修-序-39.txt" "歐陽修-序-40.txt" "歐陽修-序-41.txt" "歐陽修-序-43.txt" "歐陽修-序-44.txt"
[96] "歐陽修-序-46.txt" "蘇海-記-01.txt" "蘇海-序-03.txt" "蘇海-序-04.txt" "蘇海-序-05.txt"
[101] "蘇海-記-04.txt" "蘇海-記-05.txt" "蘇海-序-03.txt" "蘇海-序-04.txt" "蘇海-序-05.txt"
[106] "蘇海-記-06.txt" "蘇海-記-07.txt" "蘇海-序-03.txt" "蘇海-序-04.txt" "蘇海-序-05.txt"
[111] "蘇海-序-06.txt" "蘇海-序-07.txt" "蘇海-序-08.txt" "蘇海-序-10.txt" "蘇海-序-12.txt"
[116] "蘇海-序-11.txt" "蘇海-序-12.txt" "蘇海-序-13.txt" "蘇海-序-14.txt" "蘇海-序-15.txt"
[121] "蘇海-序-16.txt" "蘇海-序-17.txt" "蘇海-序-18.txt" "蘇海-序-19.txt" "蘇海-序-20.txt"
[126] "蘇海-序-21.txt" "蘇海-序-22.txt" "蘇海-序-23.txt" "蘇海-序-24.txt" "蘇海-序-25.txt"
[131] "蘇海-序-26.txt" "蘇海-序-27.txt" "蘇海-序-28.txt" "蘇海-序-29.txt" "蘇海-序-30.txt"
[136] "蘇海-序-31.txt" "蘇海-序-32.txt" "蘇海-序-33.txt" "蘇海-序-34.txt" "蘇海-序-35.txt"
[141] "蘇海-序-36.txt" "蘇海-序-37.txt" "蘇海-序-38.txt" "蘇海-序-39.txt" "蘇海-序-40.txt"
[146] "蘇海-序-41.txt" "蘇海-序-42.txt" "蘇海-序-43.txt" "蘇海-序-44.txt" "蘇海-序-45.txt"
[151] "蘇海-序-46.txt" "蘇海-序-47.txt" "蘇海-序-48.txt" "蘇海-序-49.txt" "蘇海-序-50.txt"
[156] "蘇海-序-51.txt" "蘇海-序-52.txt" "蘇海-序-53.txt" "蘇海-序-54.txt" "蘇海-序-55.txt"

[156] "魯寧-序-25.txt" "魯寧-序-26.txt" "魯寧-序-27.txt" "魯寧-序-28.txt" "魯寧-序-29.txt"
[161] "魯寧-序-30.txt" "魯寧-序-31.txt" "魯寧-序-32.txt" "魯寧-序-33.txt" "魯寧-序-34.txt"
[166] "魯寧-序-35.txt" "魯寧-序-36.txt" "魯寧-序-37.txt" "魯寧-序-38.txt" "魯寧-序-39.txt"
[171] "王安石-記-02.txt" "王安石-記-03.txt" "王安石-記-04.txt" "王安石-記-05.txt" "王安石-記-06.txt"
[176] "王安石-記-07.txt" "王安石-記-08.txt" "王安石-記-09.txt" "王安石-記-10.txt" "王安石-記-11.txt"
[181] "王安石-記-12.txt" "王安石-記-13.txt" "王安石-記-14.txt" "王安石-記-15.txt" "王安石-記-16.txt"
[186] "王安石-記-17.txt" "王安石-記-18.txt" "王安石-記-19.txt" "王安石-記-20.txt" "王安石-記-21.txt"
[191] "王安石-序-01.txt" "王安石-序-02.txt" "王安石-序-03.txt" "王安石-序-04.txt" "王安石-序-05.txt"
[196] "王安石-序-06.txt" "王安石-序-07.txt" "王安石-序-08.txt" "王安石-序-10.txt" "王安石-序-12.txt"
[201] "蘇軾-記-01.txt" "蘇軾-記-02.txt" "蘇軾-記-03.txt" "蘇軾-記-04.txt" "蘇軾-記-06.txt"
[206] "蘇軾-記-07.txt" "蘇軾-記-08.txt" "蘇軾-記-09.txt" "蘇軾-記-10.txt" "蘇軾-記-11.txt"
[211] "蘇軾-記-12.txt" "蘇軾-記-13.txt" "蘇軾-記-14.txt" "蘇軾-記-15.txt" "蘇軾-記-16.txt"
[216] "蘇軾-記-17.txt" "蘇軾-記-18.txt" "蘇軾-記-19.txt" "蘇軾-記-20.txt" "蘇軾-記-21.txt"
[221] "蘇軾-記-22.txt" "蘇軾-記-23.txt" "蘇軾-記-24.txt" "蘇軾-記-25.txt" "蘇軾-記-26.txt"
[226] "蘇軾-記-27.txt" "蘇軾-記-28.txt" "蘇軾-記-29.txt" "蘇軾-記-30.txt" "蘇軾-記-31.txt"
[231] "蘇軾-記-32.txt" "蘇軾-記-33.txt" "蘇軾-記-34.txt" "蘇軾-記-35.txt" "蘇軾-記-36.txt"
[236] "蘇軾-記-37.txt" "蘇軾-記-38.txt" "蘇軾-記-39.txt" "蘇軾-記-40.txt" "蘇軾-記-41.txt"
[241] "蘇軾-記-42.txt" "蘇軾-記-43.txt" "蘇軾-記-44.txt" "蘇軾-記-45.txt" "蘇軾-記-46.txt"
[246] "蘇軾-序-01.txt" "蘇軾-序-02.txt" "蘇軾-序-03.txt" "蘇軾-序-04.txt" "蘇軾-序-05.txt"
[251] "蘇軾-序-06.txt" "蘇軾-序-07.txt" "蘇軾-序-08.txt" "蘇軾-序-09.txt" "蘇軾-序-10.txt"
[256] "蘇軾-序-11.txt" "蘇軾-序-12.txt" "蘇軾-序-13.txt" "蘇軾-序-14.txt" "蘇軾-序-15.txt"
[261] "蘇軾-序-16.txt" "蘇軾-序-17.txt" "蘇軾-序-18.txt" "蘇軾-序-19.txt" "蘇軾-序-20.txt"
[266] "蘇軾-序-21.txt" "蘇軾-序-22.txt" "蘇軾-序-23.txt" "蘇軾-序-24.txt" "蘇軾-序-25.txt"
[271] "蘇軾-序-26.txt" "蘇軾-序-27.txt" "蘇軾-序-28.txt" "蘇軾-序-29.txt" "蘇軾-序-30.txt"
[276] "蘇軾-序-31.txt" "蘇軾-序-32.txt" "蘇軾-序-33.txt" "蘇軾-序-34.txt" "蘇軾-序-35.txt"
[281] "蘇軾-序-36.txt" "蘇軾-序-37.txt" "蘇軾-序-38.txt" "蘇軾-序-39.txt" "蘇軾-序-40.txt"
[286] "蘇軾-序-41.txt" "蘇軾-序-42.txt" "蘇軾-序-43.txt" "蘇軾-序-44.txt" "蘇軾-序-45.txt"
[291] "蘇軾-序-46.txt" "蘇軾-序-47.txt" "蘇軾-序-48.txt" "蘇軾-序-49.txt" "蘇軾-序-50.txt"
[296] "蘇軾-序-51.txt" "蘇軾-序-52.txt" "蘇軾-序-53.txt" "蘇軾-序-54.txt" "蘇軾-序-55.txt"

[1] "clust.number = 2"
[6] "鶴堂-記-02.txt" "鶴堂-記-03.txt" "鶴堂-序-04.txt" "鶴堂-序-05.txt" "鶴堂-序-06.txt"
[11] "鶴堂-序-07.txt" "鶴堂-序-08.txt" "鶴堂-序-09.txt" "鶴堂-序-10.txt" "鶴堂-序-12.txt"
[16] "鶴堂-序-13.txt" "鶴堂-序-14.txt" "鶴堂-序-15.txt" "鶴堂-序-16.txt" "鶴堂-序-18.txt"
[21] "鶴堂-序-20.txt" "鶴堂-序-21.txt" "鶴堂-序-23.txt" "鶴堂-序-25.txt" "鶴堂-序-27.txt"
[26] "鶴堂-序-28.txt" "鶴堂-序-29.txt" "鶴堂-序-30.txt" "鶴堂-序-31.txt" "鶴堂-序-32.txt"
[31] "鶴堂-序-33.txt" "鶴堂-序-34.txt" "鶴堂-序-35.txt" "鶴堂-序-37.txt" "鶴堂-序-38.txt"
[36] "鶴堂-序-39.txt" "鶴堂-序-40.txt" "鶴堂-序-41.txt" "鶴堂-序-42.txt" "鶴堂-序-43.txt"
[41] "柳宗元-記-04.txt" "柳宗元-記-06.txt" "柳宗元-記-07.txt" "柳宗元-記-15.txt" "柳宗元-記-17.txt"
[46] "柳宗元-記-18.txt" "柳宗元-記-19.txt" "柳宗元-記-20.txt" "柳宗元-記-22.txt" "柳宗元-記-24.txt"
[51] "柳宗元-記-25.txt" "柳宗元-記-27.txt" "柳宗元-記-30.txt" "柳宗元-記-31.txt" "柳宗元-記-32.txt"
[56] "柳宗元-記-33.txt" "柳宗元-記-34.txt" "柳宗元-記-35.txt" "柳宗元-記-36.txt" "柳宗元-記-37.txt"
[61] "柳宗元-序-05.txt" "柳宗元-序-06.txt" "柳宗元-序-07.txt" "柳宗元-序-09.txt" "柳宗元-序-11.txt"
[66] "柳宗元-序-12.txt" "柳宗元-序-13.txt" "柳宗元-序-14.txt" "柳宗元-序-15.txt" "柳宗元-序-16.txt"
[71] "柳宗元-序-17.txt" "柳宗元-序-18.txt" "柳宗元-序-19.txt" "柳宗元-序-25.txt" "柳宗元-序-26.txt"
[76] "柳宗元-序-28.txt" "柳宗元-序-29.txt" "柳宗元-序-30.txt" "柳宗元-序-31.txt" "柳宗元-序-32.txt"
[81] "柳宗元-序-33.txt" "柳宗元-序-34.txt" "柳宗元-序-35.txt" "柳宗元-序-36.txt" "柳宗元-序-37.txt"
[86] "柳宗元-序-40.txt" "柳宗元-序-42.txt" "柳宗元-序-45.txt" "柳宗元-序-47.txt" "柳宗元-序-48.txt"
[91] "柳宗元-序-50.txt" "柳宗元-序-51.txt" "柳宗元-序-52.txt" "柳宗元-序-53.txt" "柳宗元-序-54.txt"

[96] "柳宗元-序-50.txt" "柳宗元-序-51.txt" "柳宗元-序-52.txt" "柳宗元-序-53.txt" "柳宗元-序-54.txt"
[101] "柳宗元-序-55.txt" "柳宗元-序-56.txt" "柳宗元-序-57.txt" "柳宗元-序-58.txt" "柳宗元-序-59.txt"
[106] "歐陽修-記-21.txt" "歐陽修-記-22.txt" "歐陽修-記-23.txt" "歐陽修-記-25.txt" "歐陽修-記-26.txt"
[111] "歐陽修-記-35.txt" "歐陽修-記-37.txt" "歐陽修-記-38.txt" "歐陽修-記-39.txt" "歐陽修-記-42.txt"
[116] "歐陽修-序-01.txt" "歐陽修-序-04.txt" "歐陽修-序-05.txt" "歐陽修-序-08.txt" "歐陽修-序-10.txt"
[121] "歐陽修-序-11.txt" "歐陽修-序-13.txt" "歐陽修-序-16.txt" "歐陽修-序-17.txt" "歐陽修-序-22.txt"
[126] "歐陽修-序-26.txt" "歐陽修-序-42.txt" "歐陽修-序-49.txt" "歐陽修-序-50.txt" "歐陽修-序-51.txt"
[131] "歐陽修-序-52.txt" "歐陽修-序-53.txt" "歐陽修-序-54.txt" "蘇海-序-01.txt" "蘇海-序-02.txt"
[136] "魯寧-記-01.txt" "魯寧-記-02.txt" "魯寧-記-03.txt" "魯寧-記-10.txt" "魯寧-記-12.txt"
[141] "魯寧-記-13.txt" "魯寧-記-14.txt" "魯寧-記-17.txt" "魯寧-記-24.txt" "魯寧-記-28.txt"
[146] "魯寧-記-35.txt" "魯寧-記-37.txt" "魯寧-序-01.txt" "魯寧-序-03.txt" "魯寧-序-08.txt"
[151] "魯寧-序-05.txt" "魯寧-序-07.txt" "魯寧-序-08.txt" "魯寧-序-09.txt" "魯寧-序-10.txt"
[156] "魯寧-序-11.txt" "魯寧-序-12.txt" "魯寧-序-13.txt" "魯寧-序-15.txt" "魯寧-序-18.txt"
[161] "魯寧-序-20.txt" "魯寧-序-23.txt" "魯寧-序-32.txt" "魯寧-序-35.txt" "魯寧-序-40.txt"
[166] "魯寧-序-42.txt" "王安石-記-08.txt" "王安石-記-10.txt" "王安石-記-12.txt" "王安石-記-15.txt"
[171] "王安石-記-18.txt" "王安石-記-22.txt" "王安石-記-23.txt" "王安石-記-24.txt" "王安石-記-26.txt"
[176] "王安石-記-27.txt" "王安石-序-02.txt" "王安石-序-09.txt" "王安石-序-11.txt" "王安石-序-13.txt"
[181] "蘇軾-記-05.txt" "蘇軾-記-16.txt" "蘇軾-記-17.txt" "蘇軾-記-23.txt" "蘇軾-記-25.txt"
[186] "蘇軾-記-37.txt" "蘇軾-記-38.txt" "蘇軾-記-40.txt" "蘇軾-記-46.txt" "蘇軾-記-47.txt"
[191] "蘇軾-記-49.txt" "蘇軾-記-50.txt" "蘇軾-記-51.txt" "蘇軾-記-53.txt" "蘇軾-記-55.txt"
[196] "蘇軾-記-56.txt" "蘇軾-記-58.txt" "蘇軾-序-02.txt" "蘇軾-序-03.txt" "蘇軾-序-04.txt"
[201] "蘇軾-序-05.txt" "蘇軾-序-06.txt" "蘇軾-序-07.txt" "蘇軾-序-16.txt" "蘇軾-序-20.txt"
[206] "蘇軾-序-22.txt" "蘇軾-序-23.txt" "蘇軾-序-24.txt" "蘇軾-序-26.txt" "蘇軾-序-28.txt"
[211] "蘇軾-序-29.txt" "蘇軾-序-30.txt" "蘇軾-序-31.txt" "蘇軾-序-32.txt" "蘇軾-序-33.txt"
[216] "蘇軾-序-34.txt" "蘇軾-序-35.txt" "蘇軾-序-36.txt" "蘇軾-序-37.txt" "蘇軾-序-38.txt"
[221] "蘇軾-序-39.txt" "蘇軾-序-40.txt" "蘇軾-序-41.txt" "蘇軾-序-42.txt" "蘇軾-序-43.txt"
[226] "蘇軾-序-44.txt" "蘇軾-序-45.txt" "蘇軾-序-46.txt" "蘇軾-序-47.txt" "蘇軾-序-48.txt"
[231] "蘇軾-序-49.txt" "蘇軾-序-50.txt" "蘇軾-序-51.txt" "蘇軾-序-52.txt" "蘇軾-序-53.txt"
[236] "蘇軾-序-54.txt" "蘇軾-序-55.txt" "蘇軾-序-56.txt" "蘇軾-序-57.txt" "蘇軾-序-58.txt"
[241] "蘇軾-序-59.txt" "蘇軾-序-60.txt" "蘇軾-序-61.txt" "蘇軾-序-62.txt" "蘇軾-序-63.txt"
[246] "蘇軾-序-64.txt" "蘇軾-序-65.txt" "蘇軾-序-66.txt" "蘇軾-序-67.txt" "蘇軾-序-68.txt"
[251] "蘇軾-序-69.txt" "蘇軾-序-70.txt" "蘇軾-序-71.txt" "蘇軾-序-72.txt" "蘇軾-序-73.txt"
[256] "蘇軾-序-74.txt" "蘇軾-序-75.txt" "蘇軾-序-76.txt" "蘇軾-序-77.txt" "蘇軾-序-78.txt"
[261] "蘇軾-序-79.txt" "蘇軾-序-80.txt" "蘇軾-序-81.txt" "蘇軾-序-82.txt" "蘇軾-序-83.txt"
[266] "蘇軾-序-84.txt" "蘇軾-序-85.txt" "蘇軾-序-86.txt" "蘇軾-序-87.txt" "蘇軾-序-88.txt"
[271] "蘇軾-序-89.txt" "蘇軾-序-90.txt" "蘇軾-序-91.txt" "蘇軾-序-92.txt" "蘇軾-序-93.txt"
[276] "蘇軾-序-94.txt" "蘇軾-序-95.txt" "蘇軾-序-96.txt" "蘇軾-序-97.txt" "蘇軾-序-98.txt"
[281] "蘇軾-序-99.txt" "蘇軾-序-100.txt" "蘇軾-序-101.txt" "蘇軾-序-102.txt" "蘇軾-序-103.txt"
[286] "蘇軾-序-104.txt" "蘇軾-序-105.txt" "蘇軾-序-106.txt" "蘇軾-序-107.txt" "蘇軾-序-108.txt"
[291] "蘇軾-序-109.txt" "蘇軾-序-110.txt" "蘇軾-序-111.txt" "蘇軾-序-112.txt" "蘇軾-序-113.txt"
[296] "蘇軾-序-114.txt" "蘇軾-序-115.txt" "蘇軾-序-116.txt" "蘇軾-序-117.txt" "蘇軾-序-118.txt"
[301] "蘇軾-序-119.txt" "蘇軾-序-120.txt" "蘇軾-序-121.txt" "蘇軾-序-122.txt" "蘇軾-序-123.txt"
[306] "蘇軾-序-124.txt" "蘇軾-序-125.txt" "蘇軾-序-126.txt" "蘇軾-序-127.txt" "蘇軾-序-128.txt"
[311] "蘇軾-序-129.txt" "蘇軾-序-130.txt" "蘇軾-序-131.txt" "蘇軾-序-132.txt" "蘇軾-序-133.txt"
[316] "蘇軾-序-134.txt" "蘇軾-序-135.txt" "蘇軾-序-136.txt" "蘇軾-序-137.txt" "蘇軾-序-138.txt"
[321] "蘇軾-序-139.txt" "蘇軾-序-140.txt" "蘇軾-序-141.txt" "蘇軾-序-142.txt" "蘇軾-序-143.txt"
[326] "蘇軾-序-144.txt" "蘇軾-序-145.txt" "蘇軾-序-146.txt" "蘇軾-序-147.txt" "蘇軾-序-148.txt"
[331] "蘇軾-序-149.txt" "蘇軾-序-150.txt" "蘇軾-序-151.txt" "蘇軾-序-152.txt" "蘇軾-序-153.txt"
[336] "蘇軾-序-154.txt" "蘇軾-序-155.txt" "蘇軾-序-156.txt" "蘇軾-序-157.txt" "蘇軾-序-158.txt"
[341] "蘇軾-序-159.txt" "蘇軾-序-160.txt" "蘇軾-序-161.txt" "蘇軾-序-162.txt" "蘇軾-序-163.txt"
[346] "蘇軾-序-164.txt" "蘇軾-序-165.txt" "蘇軾-序-166.txt" "蘇軾-序-167.txt" "蘇軾-序-168.txt"
[351] "蘇軾-序-169.txt" "蘇軾-序-170.txt" "蘇軾-序-171.txt" "蘇軾-序-172.txt" "蘇軾-序-173.txt"
[356] "蘇軾-序-174.txt" "蘇軾-序-175.txt" "蘇軾-序-176.txt" "蘇軾-序-177.txt" "蘇軾-序-178.txt"
[361] "蘇軾-序-179.txt" "蘇軾-序-180.txt" "蘇軾-序-181.txt" "蘇軾-序-182.txt" "蘇軾-序-183.txt"
[366] "蘇軾-序-184.txt" "蘇軾-序-185.txt" "蘇軾-序-186.txt" "蘇軾-序-187.txt" "蘇軾-序-188.txt"
[371] "蘇軾-序-189.txt" "蘇軾-序-190.txt" "蘇軾-序-191.txt" "蘇軾-序-192.txt" "蘇軾-序-193.txt"
[376] "蘇軾-序-194.txt" "蘇軾-序-195.txt" "蘇軾-序-196.txt" "蘇軾-序-197.txt" "蘇軾-序-198.txt"
[381] "蘇軾-序-199.txt" "蘇軾-序-200.txt" "蘇軾-序-201.txt" "蘇軾-序-202.txt" "蘇軾-序-203.txt"
[386] "蘇軾-序-204.txt" "蘇軾-序-205.txt" "蘇軾-序-206.txt" "蘇軾-序-207.txt" "蘇軾-序-208.txt"
[391] "蘇軾-序-209.txt" "蘇軾-序-210.txt" "蘇軾-序-211.txt" "蘇軾-序-212.txt" "蘇軾-序-213.txt"
[396] "蘇軾-序-214.txt" "蘇軾-序-215.txt" "蘇軾-序-216.txt" "蘇軾-序-217.txt" "蘇軾-序-218.txt"
[401] "蘇軾-序-219.txt" "蘇軾-序-220.txt" "蘇軾-序-221.txt" "蘇軾-序-222.txt" "蘇軾-序-223.txt"
[406] "蘇軾-序-224.txt" "蘇軾-序-225.txt" "蘇軾-序-226.txt" "蘇軾-序-227.txt" "蘇軾-序-228.txt"
[411] "蘇軾-序-229.txt" "蘇軾-序-230.txt" "蘇軾-序-231.txt" "蘇軾-序-232.txt" "蘇軾-序-233.txt"
[416] "蘇軾-序-234.txt" "蘇軾-序-235.txt" "蘇軾-序-236.txt" "蘇軾-序-237.txt" "蘇軾-序-238.txt"
[421] "蘇軾-序-239.txt" "蘇軾-序-240.txt" "蘇軾-序-241.txt" "蘇軾-序-242.txt" "蘇軾-序-243.txt"
[426] "蘇軾-序-244.txt" "蘇軾-序-245.txt" "蘇軾-序-246.txt" "蘇軾-序-247.txt" "蘇軾-序-248.txt"
[431] "蘇軾-序-249.txt" "蘇軾-序-250.txt" "蘇軾-序-251.txt" "蘇軾-序-252.txt" "蘇軾-序-253.txt"
[436] "蘇軾-序-254.txt" "蘇軾-序-255.txt" "蘇軾-序-256.txt" "蘇軾-序-257.txt" "蘇軾-序-258.txt"
[441] "蘇軾-序-259.txt" "蘇軾-序-260.txt" "蘇軾-序-261.txt" "蘇軾-序-262.txt" "蘇軾-序-263.txt"
[446] "蘇軾-序-264.txt" "蘇軾-序-265.txt" "蘇軾-序-266.txt" "蘇軾-序-267.txt" "蘇軾-序-268.txt"
[451] "蘇軾-序-269.txt" "蘇軾-序-270.txt" "蘇軾-序-271.txt" "蘇軾-序-272.txt" "蘇軾-序-273.txt"
[456] "蘇軾-序-274.txt" "蘇軾-序-275.txt" "蘇軾-序-276.txt" "蘇軾-序-277.txt" "蘇軾-序-278.txt"
[461] "蘇軾-序-279.txt" "蘇軾-序-280.txt" "蘇軾-序-281.txt" "蘇軾-序-282.txt" "蘇軾-序-283.txt"
[466] "蘇軾-序-284.txt" "蘇軾-序-285.txt" "蘇軾-序-286.txt" "蘇軾-序-287.txt" "蘇軾-序-288.txt"
[471] "蘇軾-序-289.txt" "蘇軾-序-290.txt" "蘇軾-序-291.txt" "蘇軾-序-292.txt" "蘇軾-序-293.txt"
[476] "蘇軾-序-294.txt" "蘇軾-序-295.txt" "蘇軾-序-296.txt" "蘇軾-序-297.txt" "蘇軾-序-298.txt"
[481] "蘇軾-序-299.txt" "蘇軾-序-300.txt" "蘇軾-序-301.txt" "蘇軾-序-302.txt" "蘇軾-序-303.txt"
[486] "蘇軾-序-304.txt" "蘇軾-序-305.txt" "蘇軾-序-306.txt" "蘇軾-序-307.txt" "蘇軾-序-308.txt"
[491] "蘇軾-序-309.txt" "蘇軾-序-310.txt" "蘇軾-序-311.txt" "蘇軾-序-312.txt" "蘇軾-序-313.txt"
[496] "蘇軾-序-314.txt" "蘇軾-序-315.txt" "蘇軾-序-316.txt" "蘇軾-序-317.txt" "蘇軾-序-318.txt"
[501] "蘇軾-序-319.txt" "蘇軾-序-320.txt" "蘇軾-序-321.txt" "蘇軾-序-322.txt" "蘇軾-序-323.txt"
[506] "蘇軾-序-324.txt" "蘇軾-序-325.txt" "蘇軾-序-326.txt" "蘇軾-序-327.txt" "蘇軾-序-328.txt"
[511] "蘇軾-序-329.txt" "蘇軾-序-330.txt" "蘇軾-序-331.txt" "蘇軾-序-332.txt" "蘇軾-序-333.txt"
[516] "蘇軾-序-334.txt" "蘇軾-序-335.txt" "蘇軾-序-336.txt" "蘇軾-序-337.txt" "蘇軾-序-338.txt"
[521] "蘇軾-序-339.txt" "蘇軾-序-340.txt" "蘇軾-序-341.txt" "蘇軾-序-342.txt" "蘇軾-序-343.txt"
[526] "蘇軾-序-344.txt" "蘇軾-序-345.txt" "蘇軾-序-346.txt" "蘇軾-序-347.txt" "蘇軾-序-348.txt"
[531] "蘇軾-序-349.txt" "蘇軾-序-350.txt" "蘇軾-序-351.txt" "蘇軾-序-352.txt" "蘇軾-序-353.txt"
[536] "蘇軾-序-354.txt" "蘇軾-序-355.txt" "蘇軾-序-356.txt" "蘇軾-序-357.txt" "蘇軾-序-358.txt"
[541] "蘇軾-序-359.txt" "蘇軾-序-360.txt" "蘇軾-序-361.txt" "蘇軾-序-362.txt" "蘇軾-序-363.txt"
[546] "蘇軾-序-364.txt" "蘇軾-序-365.txt" "蘇軾-序-366.txt" "蘇軾-序-367.txt" "蘇軾-序-368.txt"
[551] "蘇軾-序-369.txt" "蘇軾-序-370.txt" "蘇軾-序-371.txt" "蘇軾-序-372.txt" "蘇軾-序-373.txt"
[556] "蘇軾-序-374.txt" "蘇軾-序-375.txt" "蘇軾-序-376.txt" "蘇軾-序-377.txt" "蘇軾-序-378.txt"
[561] "蘇軾-序-379.txt" "蘇軾-序-380.txt" "蘇軾-序-381.txt" "蘇軾-序-382.txt" "蘇軾-序-383.txt"
[566] "蘇軾-序-384.txt" "蘇軾-序-385.txt" "蘇軾-序-386.txt" "蘇軾-序-387.txt" "蘇軾-序-388.txt"
[571] "蘇軾-序-389.txt" "蘇軾-序-390.txt" "蘇軾-序-391.txt" "蘇軾-序-392.txt" "蘇軾-序-393.txt"
[576] "蘇軾-序-394.txt" "蘇軾-序-395.txt" "蘇軾-序-396.txt" "蘇軾-序-397.txt" "蘇軾-序-398.txt"
[581] "蘇軾-序-399.txt" "蘇軾-序-400.txt" "蘇軾-序-401.txt" "蘇軾-序-402.txt" "蘇軾-序-403.txt"
[586] "蘇軾-序-404.txt" "蘇軾-序-405.txt" "蘇軾-序-406.txt" "蘇軾-序-407.txt" "蘇軾-序-408.txt"
[591] "蘇軾-序-409.txt" "蘇軾-序-410.txt" "蘇軾-序-411.txt" "蘇軾-序-412.txt" "蘇軾-序-413.txt"
[596] "蘇軾-序-414.txt" "蘇軾-序-415.txt" "蘇軾-序-416.txt" "蘇軾-序-417.txt" "蘇軾-序-418.txt"
[601] "蘇軾-序-419.txt" "蘇軾-序-420.txt" "蘇軾-序-421.txt" "蘇軾-序-422.txt" "蘇軾-序-423.txt"
[606] "蘇軾-序-424.txt"
```

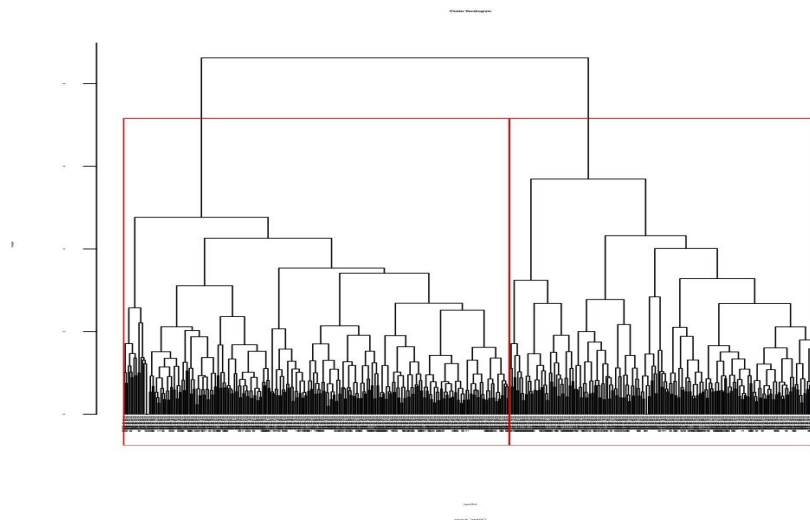


図 2 特徴分析支援システムによる樹形図の出力例

距離（市街地距離ともいう） d_{ij} は $d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}|$ と定義される [1]。本特徴分析支援システムでは作品の特性を示す変数として、以下のように定義される作品 i に出現する k 番目の虚詞の出現率 x_{ik} を使う。すなわち、作品 i に出現する、虚詞一覧で k 番目に指定した虚詞の出現率として、以下の 2 つの出現率（全文字数出現率、全虚詞数出現率）を用いて作品間の距離を求める。作品の特性を示す変数の数 p は、指定した虚詞の数のことである。

$$\text{全文字数出現率 } x_{ik} = \frac{\text{作品 } i \text{ での } k \text{ 番目の虚詞の出現回数}}{\text{作品 } i \text{ の全文字数}}$$

$$\text{全虚詞数出現率 } x_{ik} = \frac{\text{作品 } i \text{ での } k \text{ 番目の虚詞の出現回数}}{\text{作品 } i \text{ での全虚詞の出現回数の合計}}$$

3. 実験

虚詞出現率によるクラスタリングを用いた特徴分析支援システムを用いて、中国古典文学作品のうち唐宋八大家在作成した、記と序の作品データの分析を行った実験結果 [5] について説明する。

この実験の設定は以下の通りである。唐宋八大家在作成した散文のうち、今回は記と序を分析の対象とする。すなわち、記 244 作品（韓愈 9 作品、柳宗元 38 作品、歐陽脩 43 作品、蘇洵 5 作品、蘇軾 58 作品、蘇轍 25 作品、曾鞏 39 作品、王安石 27 作品）と、序 247 作品（韓愈 40 作品、柳宗元 58 作品、歐陽脩 57 作品、蘇洵 4 作品、蘇軾 26 作品、蘇轍 7 作品、曾鞏 42 作品、王安石 13 作品）の計 491 作品を分析の対象とする。入力作品データのファイル名は、作家、記か序か、通し番号を表す。例えば、ファイル“王安石-記-16.txt”は、王安石的記の 16 番の作品データであることを表す。本実験で使用する虚詞は、句頭に出現する「夫」、「凡」、「且」、「蓋」、「又」、「亦」、「故」、「因」、「然」、「而」、「則」、「便」、「乃」、「輒」と句末に出現する「矣」、「焉」、「也」、「哉」、「乎」、

“歟”、“邪”、“爾”、“耳”、“已”の 24 字の虚詞である。ただし、それぞれの字が句頭または句末に出現とするとだけ数える。これらの虚詞 24 字を用いて 2 種類の出現率（全文字数出現率、全虚詞数出現率）を用いて距離を求め、クラスタリングを行う。用いる距離は、標準化ユークリッド距離、キャンベラ距離、マンハッタン距離、最大距離、バイナリ距離 (5 通り) である。用いるクラスタリング手法は、階層的な手法である最長距離法、最短距離法、ウォード法、群間平均法、重心法、メディアン法 (6 通り) である。最終クラスター数は、2, 3, 4 (3 通り) で行った。

表 1 クラスタリング結果の例

	クラスター 1	クラスター 2
韓愈	11 (22%)	38 (78%)
柳宗元	32 (33%)	64 (67%)
歐陽脩	69 (69%)	31 (31%)
蘇洵	7 (78%)	2 (22%)
曾鞏	50 (62%)	31 (38%)
王安石	26 (65%)	14 (35%)
蘇軾	55 (65%)	29 (35%)
蘇轍	23 (72%)	9 (28%)

表 2 唐と宋の作家の作品数と割合

	クラスター 1	クラスター 2
唐の作家	43 (30%)	102 (70%)
宋の作家	230 (66%)	116 (34%)

例として、全虚詞数出現率、マンハッタン距離、ウォード法を用いて行った最終クラスター数 2 の設定でのクラスタリング結果を示す。この設定での特徴分析支援システムによるクラスタリングの実行の様子を図 1 に示し、出力される樹形図を図 2 に示す。この設定でのクラスタリング結果を表 1 に示す。この結果を唐と宋の作家の作品で分類

した結果を表 2 に示す。各行が各クラスターに含まれる作家の作品数と作家の全作品における割合を表している。表 1 では、例えば韓愈について 11 作品（割合は 22%）がクラスター 1 に、38 作品（割合は 78%）がクラスター 2 に属することを示している。表 2 では、例えば唐の作家について 43 作品（割合は 30%）がクラスター 1 に、102 作品（割合は 70%）がクラスター 2 に属することを示している。

4. おわりに

本稿では、中国古典文学作品に対する虚詞出現率によるクラスターリングを用いた特徴分析支援システムについて報告した。虚詞の出現場所に注目し、専門家が指定した虚詞出現率、距離、クラスターリング手法で特徴分析を支援するシステムを開発した。中国古典文学作品の中で、今回唐宋八大家の作成した記と序の作品データの分析を行った実験により本特徴分析支援システムの有用性が確認できた。今後の課題として、分析手法の追加などが挙げられる。

謝辞 本研究は JSPS 科研費 JP18H00651 の助成を受けたものです。

参考文献

- [1] 村上 征勝: 真贋の科学 -計算文献学入門-, 朝倉書店 (1994).
- [2] 東 英寿: 欧陽脩『六一詩話』文体の特色, 中国古典論集, Vol. 34, pp. 27-39 (2005).
- [3] 新納 浩幸: R で学ぶクラスタ解析, オーム社 (2007).
- [4] 劉 徳清: 『欧陽脩論稿』, pp.273-274, 北京師範大学出版社 (1991).
- [5] 東英寿, 宮原哲浩: 唐宋八大家の文章におけるクラスター分析試論, 『唐宋八大家の探究』所収, 花書院 (2021 年 3 月刊行予定).