2ZB-05

# 郷土研究に着目した聞き書きの提示手法の提案

寺嶋　一将†　　植竹　俊文†　　竹野　健夫†
岩県立大学大学院　ソフトウェア情報学研究科†

## 1. はじめに

岩手県花巻市では地域住民の記憶を歴史的な史料として，取材により聞き書きとして収集し，デジタルアーカイブにて公開している．しかし，聞き書きの内容による整理がされていないため，その活用が困難な状態である．

本研究では，聞き書きに記された情報を整理したうえで提示するしくみを提案する．聞き書きの利用者としては近隣の児童・学生・生徒，郷土史家，行政の職員等を想定している．提示に関しては郷土についての学習や研究の側面からの利用を考え，聞き書きに記された内容を地理的・時間的に整理したうえで，地域の特性を用いて提示する．

## 2. 聞き書きの現状と課題

聞き書きとは古老等の話を語り口調を活かしつつ書き起した文章である．聞き書きには話者の半生に合わせて地域の歴史や文化がまとめられている．

### 2.1 聞き書きの現状

聞き書きは 2018 年 1 月現在，約 100 人分が公開されている．取材は継続して実施しているため，今後も公開される聞き書きは増加し続ける．公開は花巻市の郷土史研究団体が運営するデジタルアーカイブにて行われている．

### 2.2 聞き書きの課題

公開される聞き書きの数が増え続けることにより，閲覧者が蓄積された史料の全体を把握することが困難な状態になりつつある．ここから，聞き書きが抱える課題として以下の 3 点が挙げられる．

#### (1) 内容による分類がされていない

内容に踏み込んだ整理がされていないため，閲覧者が蓄積された聞き書きから任意の情報を探し出すことが困難な状態となっている．閲覧者にとって関心のある情報が含まれていたとしても埋没してしまい発見できない恐れがある．

Proposal of presentation method of "Kikigaki" focusing on local studies
†Kazumasa Terashima, Toshifumi Uetake, Takeo Takeno
Iwate Prefectural University

#### (2) 聞き書き間での関連性等が不明瞭

聞き書きの中には同一の話題を取り上げているものもある．それらは地域の特定の出来事について異なる視点から語っているため，地域を理解するうえでの手掛かりになる．しかし，内容の関連性を把握することは難しい状態である．

#### (3) 聞き書きの活用が困難

聞き書きは文章であり，利用する際には十分に読み込んで内容を理解する必要がある．しかし，閲覧者として周辺地域の児童など，低年齢での利用も想定されるため，提示手法を見直し，活用しやすい形で提示する必要がある．

## 3. 提案・検証内容の概要

先行研究[1]にて蓄積された聞き書きの整理と提示に関する手法を提案する．本研究で提案する内容は以下の**図 1** にまとめられる．蓄積された聞き書きを内容に従い，時代・地域の側面から整理し，そこから地域の特性を語として発見する．最終的に，整理された時空間情報及び地域の特徴をもとに，聞き書きを閲覧者に対して提示する．

### 3.1 内容による時代・地域の整理

聞き書きは話者の半生として綴られているため，出来事等の記述を基に地域と時代の側面から，その内容を整理することができる．その際に，地域は花巻市を合併前の 4 つの旧市町に分割し，整理する．

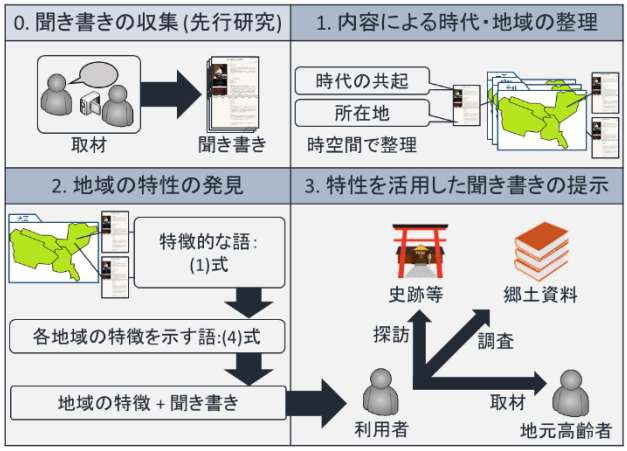


**図 1**　聞き書きの整理と提示の流れ

### (1) 地域による整理

話者が所在地周辺の事柄を中心に語ると仮説立て，文中に出現する地名と話者の所在地が，どの程度一致するかを調査した．結果，文中に出現する地名は大半が話者の所在地内の地名であることが分かった．そこで，本研究では話者の所在地を聞き書きの内容に関連した地域として扱う．

### (2) 時代による整理

時代を整理するために，時代表記として西暦と和暦，及び『時代』が含まれる表記を抽出した．結果，全体の約三分の一の文章にそれらの表記が使われていた．ここから，時代からの整理は文中の時代表記を基に行う．

## 3.2 地域の特性の発見

聞き書きを時代と地域で整理した後，特定の時代，地域による文章の集合内から名詞を抽出し，$TFIDF$により文章としての特徴量を付与する．

### (1) 名詞に対する特徴量の付与

文書$T_j$における名詞$W_i$の$TFIDF$は次式で表される．

$$TFIDF_{i,j} = TF_{i,j} * IDF_i \quad (1)$$

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (2)$$

$$IDF_i = \log \frac{D}{d_i} \quad (3)$$

$n_{i,j}$は文書$T_j$における単語$W_i$の出現回数であり，$\sum_k n_{k,j}$は文書$T_j$における全単語の出現回数の総和である．$D$は全文書数であり，$d_i$は単語$W_i$を含む文書数である．

付与された値は文章内における名詞の特徴を示したものである．この数値に合わせ，名詞がどの程度地域の特性を示しているかを意味する特徴量$FV_{i,j}$を，本研究で提案する以下の(4)式で求める．

$$FV_{i,j} = TFIDF_{i,j} * \frac{l_{i,m}}{L_i} \quad (4)$$

$l_{i,m}$は任意の地域$R_m$における単語$W_i$の出現する文書であり，$L_i$は全地域における単語$W_i$の出現する文書数の総和である．

### (2) 地域の個性を示す語に対する特徴量の付与

各地域の個性として定義できる名詞[2]をまとめ，それぞれに(1)，(4)，(5)式を用いて特徴量を付与した(**表 1**)．その後，付与された特徴量から，名詞に対して地域ごとに順位付けを行い，各手法を比較した(**表 2**)．(5)式は残差 IDF で特徴量を求める式である．$E_i$は名詞，$n$は全ての文書数，$n_i$は名詞$E_i$を含む文書数，$F_i$は名詞$E_i$の出現頻度である．

$$RIDF(E_i) = \log \frac{n}{n_i} + \log\left(1 - e^{-\frac{F_i}{n}}\right) \quad (5)$$

**表 1**　各地域で定義した特徴語への特徴量付与

| | | TF-IDF (1)式 | | | | 案手法 (4)式 | |
|---|---|---|---|---|---|---|---|
| 地域 | 個性 | 花巻 | 石鳥谷 | 大迫 | 東和 | 大迫 | 東和 |
| 花巻 | (宮沢)賢治 | 0.13 | 0.09 | 0.19 | - | 0.05 | - |
| | (高村)光太郎 | 0.11 | - | - | - | - | - |
| 石鳥谷 | 南部杜氏 | - | 0.13 | 0.1 | - | 0.02 | - |
| 大迫 | 早池峰 | - | - | 0.06 | - | 0.06 | - |
| | 神楽 | 0.04 | 0.1 | 0.22 | 0.1 | 0.14 | 0.02 |
| | ワイン | - | - | 0.04 | - | 0.04 | - |
| 東和 | (萬)鉄五郎 | - | - | - | 0.04 | - | 0.04 |
| | 田瀬(湖) | 0.02 | - | - | 0.07 | - | 0.07 |

**表 2**　各地域で定義した特徴語の出現順位

| 地域 | 花巻 | | 石鳥谷 | 大迫 | | | 東和 | |
|---|---|---|---|---|---|---|---|---|
| 特徴語 | (宮沢)賢治 | (高村)光太郎 | (南部)杜氏 | 早池峰 | 神楽 | ワイン | (萬)鉄五郎 | 田瀬(ダム) |
| TF-IDF | 173 | 267 | 124 | 595 | 22 | 1248 | 783 | 293 |
| 残差 IDF | 225 | 219 | 164 | 438 | 61 | 181 | 978 | 508 |
| 提案手法 | 310 | 148 | 102 | 281 | 48 | 602 | 351 | 164 |

**表 2** の灰色で示した箇所は 3 手法の中で最も順位の高いものである．手法を比較すると，提案手法によって該当する地域において，特性を示す語を際立たせることができる傾向が見られた．提案手法を用いることで，名詞がもつ地域の特性を明確化させることができた．

## 3.3 特性を活用した聞き書きの提示

語に対して付与された特徴量をもとに各文章に対して，地域の特性をどの程度示しているかをまとめる．それに合わせて文章間の関係について，特徴量を付与された語をもとに整理する．ここから，聞き書きが示す地域の特徴量をまとめ，閲覧者に対して提示する．

## 4. おわりに

聞き書きを時代・地域で分類し，地域の特徴を発見するために特徴量の付与を実施した．特徴量の値と語の出現順を 3 つの手法で比較した結果，提案手法で最も各地域の個性を際立たせることができた．今後はこの手法で得られる地域の特徴を活用し，聞き書きを閲覧者に対して提示する仕組みの構築に取り組む．

## 参考文献

[1] 小田島瑞希，竹野健夫，植竹俊文，菅原光政：地域コミュニティにおけるデジタルアーカイブシステム，情報文化学会 第 20 回全国大会講演予稿集，pp. 38-41(2012)．

[2] 花巻市：花巻市消防本部管内の概況，平成 27 年版花巻市消防年報，p. 1(2015)．