

技術書の索引から抽出した用語関係データに基づく 技術用語マッピング手法の提案

猪口 真吾[†] 奥 健太[†]

[†] 龍谷大学理工学部 〒520-2194 滋賀県大津市瀬田大江町横谷 1-5

E-mail: ^{††}t150447@mail.ryukoku.ac.jp, ^{†††}okukenta@rins.ryukoku.ac.jp

あらまし 知識を共有するためのさまざまなサービスを利用するにあたり、技術分野の初学者にとっては、どの技術用語で検索をかけるべきかの選定が難しく、適切な情報を得るためには、どの技術用語とどの技術用語が関連するのかが把握する必要がある。このような問題に対し、本研究では、技術用語間の関係性を可視化するインタフェースを検討する。技術用語間の関係性を抽出する基となる情報源としては技術書の索引に着目する。同一技術書において出現ページが近い技術用語は互いに関連性が強いと仮定することで、技術用語間の類似性を定義する。類似性は、技術書の索引から抽出した技術書-索引語（技術用語）-出現ページの関係データを基に算出する。算出した類似性から索引語間距離行列を作成し、多次元尺度構成法（MDS）により関係性を可視化する。索引語間距離行列においては、同一の技術書に出現しない索引語同士の要素は欠損値となるが、MDS では距離行列上に欠損値が存在することは許されない。提案手法では、距離行列を作成するにあたり、このような欠損値に対して定数で補完する。提案手法により、技術用語間の類似性を算出、2次元空間への可視化が可能であるか検討する。また、提示される関連語が妥当であるか考察する。

キーワード 技術用語推薦, 可視化

1. はじめに

現在、Wikipedia^(注1)や Qiita^(注2)、SlideShare^(注3)など、知識を共有するためのさまざまなサービスがある。こうしたサービスには基本的にはキーワード検索機能が提供されている。ある技術について調べたいとき、技術用語でキーワード検索することで、関連する情報を得ることができる。しかしながら、ある技術分野における初学者にとっては、どの技術用語で検索をかけるべきか、その選定が難しい。適切な情報を得るためには、どの技術用語とどの技術用語とが関連するのかが把握しておく必要がある。

このような問題に対し、本研究では、技術用語間の関係性を可視化するインタフェースを検討する。技術用語間の関係性を抽出する基となる情報源としては技術書の索引に着目する。技術書の索引には、その技術書における索引語（技術用語）の出現ページが一覧としてまとめられている。この索引から技術書-索引語（技術用語）-出現ページの関係データを抽出する。ここで、同一技術書において出現ページが近い技術用語は互いに関連性が強いと仮定することで、技術用語間の類似性を定義する。この類似性を基に、技術用語の関係性を2次元空間に可視化したインタフェースを構築する。

なお、推薦システムの基本方式として、大きく次の三つの方式がある；(a) 協調ベース推薦 [1], (b) 内容ベース推薦 [2], (c)

知識ベース推薦 [3] [4]。 (a) および (b) の方式には、大量の履歴データが必要となる。また、(a) では新規ユーザおよび新規アイテムのコールドスタート問題がある。(b) では、新規アイテムのコールドスタート問題については解消されるものの、新規ユーザのコールドスタート問題は依然として残る。これらに対し、(c) は履歴データは必要とせず、新規ユーザに対しても推薦可能である。

知識ベース推薦には、制約ベース推薦方式と事例ベース推薦方式がある。制約ベース推薦は、ユーザが提示した制約を満たすアイテムを推薦する。事例ベース推薦は、ユーザが例示したアイテムや属性に類似するアイテムを推薦する。推薦されたアイテムを参考にしながら、ユーザが条件を修正（批評）していくことで、ユーザのアイテム探索を支援する。本研究では、履歴データをもたない新規ユーザにも推薦可能なシステムを想定しているため、知識ベース推薦方式に位置付けられる。

2. 関連研究

2.1 書籍推薦システム

Amazon.com における推薦システムでは、ユーザの購入履歴に基づき書籍を推薦している [5]。Yang ら [6] の手法は、協調フィルタリングを用い、ユーザのアクセスログを基にユーザの嗜好に合った書籍を推薦する。Givon ら [7] は、書籍の内容を表すソーシャルタグと協調フィルタリングを組み合わせた書籍推薦システムを提案している。Sieg ら [8] は、ドメインオントロジーの形で表現される意味的知識を取り入れたユーザベース協調フィルタリングを用い、ユーザの嗜好を抽出している。BReK12 [9] は、ソーシャルブックマークサイトにおけるブック

(注1) : <https://ja.wikipedia.org/>

(注2) : <https://qiita.com/>

(注3) : <https://www.slideshare.net/>



図 1 システムインターフェース

マークを基に K-12—アメリカにおける幼稚園年長から高校卒業までの 13 年間—向けの書籍を推薦するシステムである。これらの推薦システムでは、ユーザの購入履歴やブックマーク、オントロジー等が必要となる。これらに対し、K3Rec [10] は、新規ユーザを対象とし、履歴データやオントロジー等を必要としない書籍推薦システムである。K3Rec は、Google Books^(注4) や Amazon.com 等から利用可能なメタデータを用いている。K3Rec に対し、我々は、書籍の索引に着目しており、索引語ベースで書籍を特徴化することを想定している。本研究では、これに先立ち、書籍の索引から抽出した索引語間の関連性について分析する。

3. 定義

定義 1：技術書。 技術書を $b_k \in B$ と表す。

定義 2：索引語。 技術書の索引に登録されている語を索引語とよび、 $t_i \in T$ と表す。

定義 3：索引語間距離行列。 索引語間距離行列を $D = [d_{ij}]_{|T| \times |T|}$ と表す。距離行列 D は $|T| \times |T|$ の対称行列である。 d_{ij} は索引語 t_i と索引語 t_j の間の距離を表す。

定義 4：索引語マップ。 索引語マップは、索引語間の類似関係を 2 次元特徴空間に可視化したものである。類似している索引語同士は近く、類似していない索引語同士は離れて配置される。

4. システム

本章では、提案手法である技術用語マッピング手法について説明する。

4.1 システム概要

図 1 は、提案システムのインターフェースである。インターフェースは、検索ビューとマップビューから構成される。

ユーザは検索ビューから任意のキーワードで索引語を検索することができる。マップビューには、検索された索引語周辺の索引語マップが表示される。ユーザは、このマップビューから興味のある索引語に関連する索引語を探索し、調べることができる。

このようなシステムを実現するためには、あらかじめ索引語

マップを用意しておく必要がある。本研究では、技術書から索引語集合を抽出し、技術書内における索引語間のページ差に基づき、索引語間距離行列を作成する。この索引語間距離行列を基に、多次元尺度構成法により、2 次元空間による索引語マップを作成する。

以下、提案システムについて詳細に説明する。

4.2 索引語の登録

技術書集合 B の各索引に登録されている索引語を抽出し、次の関係スキーマをもつ各テーブルに格納する：

```
books(isbn, book_title)
terms(term_id, term_name)
books-terms(isbn, term_id, page)
```

ここで、**books** は技術書テーブルであり、**isbn_id** は技術書の ISBN、**book_title** は技術書のタイトルを表す。**terms** は索引語テーブルであり、**term_id** は索引語 ID、**term_name** は索引語名を表す。**books-terms** は技術書と索引語の関係テーブルであり、**isbn**, **term_id** は、それぞれ **books** テーブル、**terms** テーブルを参照する外部キーである。**page** は、その索引語がその技術書に出現するページ番号を表す。索引語が同一の技術書に複数回出現する場合は、出現回数分のタプルが登録される。例えば、「確率的ブロックモデル」（仮に索引語 ID を 1 とする）は『関係データ学習』（ISBN:9784061529212）の 44 ページと 48 ページに出現する。この場合は、二つのタプル（“9784061529212”, 1, 44）、（“9784061529212”, 1, 48）が **books-terms** テーブルに登録される。

また、索引語には表記揺れの問題がある。例えば、「SVM」の同義語として、「SVM」、「サポート・ベクトル・マシン」、「サポートベクトルマシン」、「support vector machine」などがある。このような関係は、下記のテーブルに登録しておく：

```
synonymous_terms(term_a_id, term_b_id)
```

ここで、**term_a_id** は代表となる索引語の ID であり、**term_b_id** はその代表語と関連付ける同義語となる索引語の ID である。

4.3 索引語間距離行列の作成

books-terms テーブルを基に、索引語間距離行列 $D = [d_{ij}]$ を作成する。索引語 t_i と索引語 t_j の間のページ差を距離 d_{ij} とする。

索引語 t_i と索引語 t_j の間の距離 d_{ij} の算出は、次の 2 通りを考える。

a) 線形方式

索引語 t_i と索引語 t_j の間のページ差と距離 d_{ij} との関係が線形であると仮定する。つまり、同一の技術書内において t_i と t_j が同一ページに出現するときには $d_{ij} = 0$ 、最大限離れて出現するときには $d_{ij} = 1$ とし、その間は線形的に変化するものとする。具体的には、 d_{ij} は次の手順により求める：

(1) **books-terms** テーブルから、二つの索引語 t_i, t_j が共起する技術書集合 $B \subseteq B$ を取得する。

(2) 各技術書 $b_k \in B$ について、それぞれ次の (3)–(4) の処理をする。

(注4) : <http://books.google.com/>

(3) その総ページ数 p_k を取得する.

(4) その技術書 b_k 内における二つの索引語 t_i, t_j の間のページ差 d_{ijk} を取得する. ここで, b_k 内において t_i や t_j が複数回出現する場合には, 最小のページ差を d_{ijk} とする.

(5) 次式により, 距離 d_{ij} を算出する:

$$d_{ij} = \frac{1}{|B|} \sum_k \frac{d_{ijk}}{p_k} \quad (1)$$

b) 非線形方式

索引語 t_i と索引語 t_j の間のページ差と距離 d_{ij} との関係が線形であると仮定する. ここでは, 同一の技術書内において t_i と t_j が同一ページに出現するときには $d_{ij} = 0$, ページ差が一定の大きさ以上になると $d_{ij} = 1$ とし, その間は曲線的に変化するものとする. 具体的には, d_{ij} は次の手順により求める:

(1) **books-terms** テーブルから, 二つの索引語 t_i, t_j が共起する技術書集合 $B \subseteq \mathcal{B}$ を取得する.

(2) 各技術書 $b_k \in B$ について, それぞれ次の (3)–(4) の処理をする.

(3) その総ページ数 p_k を取得する.

(4) その技術書 b_k 内における二つの索引語 t_i, t_j の間のページ差 d_{ijk} を取得する. ここで, b_k 内において t_i や t_j が複数回出現する場合には, 最小のページ差を d_{ijk} とする.

(5) 次式により, 距離 d_{ij} を算出する:

$$d_{ij} = \frac{1}{|B|} \sum_k \frac{1}{1 + e^{-10(\frac{2d_{ijk}}{p_k} - 1)}} \quad (2)$$

4.4 2次元空間への索引語のマッピング

4.3 節で作成した索引語間距離行列を 2 次元空間にマッピングする. 本研究では, 2 次元空間へのマッピングには多次元尺度構成法 (MDS) を用いる.

MDS は, 距離データを用いて, 2 次元空間上に距離が小さいものは近く, 大きいものは遠く配置する. これにより, 索引語間の相対関係の可視化を可能とする. MDS には, 距離データを低次元に配置できる計量多次元尺度法と類似度を低次元に配置できる非計量多次元尺度法がある. 本研究で MDS を適用する索引語間距離行列は距離データであるので, 計量多次元尺度法を適用する.

MDS では距離行列上に欠損値が存在することは許されない. 索引語間距離行列においては, 同一の技術書に出現しない索引語同士の要素は欠損値となる. そのため, このような欠損値に対して何らかの定数で補完することで距離行列を補完する必要がある. 二つの索引語が同一の技術書に出現しないということは, 距離の考え方から, それらの索引語間の距離は十分大きい値を設定する必要がある. 本研究では, 予備実験の結果, 欠損値を 10 で補完することとした.

MDS による索引語間距離行列 D に基づく索引語集合 T の 2 次元空間へのマッピングは以下の手順で行う:

(1) 距離行列 D の欠損値を定数で補完する.

(2) ヤング・ハウスホルダー変換で得られた行列をスペクトル分解し, 固有値と固有ベクトルを求める.

表 1 実験対象の技術書リスト.

ISBN	タイトル	出版社
978-4061529014	機械学習のための確率と統計	講談社
978-4061529021	深層学習	講談社
978-4061529038	オンライン機械学習	講談社
978-4061529045	トピックモデル	講談社
978-4061529052	統計的学習理論	講談社
978-4061529069	サポートベクトルマシン	講談社
978-4061529076	確率的最適化	講談社
978-4061529083	異常検知と変化検知	講談社
978-4061529090	劣モジュー最適化と機械学習	講談社
978-4061529106	スパース性に基づく機械学習	講談社
978-4061529113	生命情報処理における機械学習 多重検定と推定量設計	講談社
978-4061529120	画像認識	講談社
978-4061529137	ヒューマンコンピューテーションとクラウドソーシング	講談社
978-4061529144	変分ベイズ学習	講談社
978-4061529151	ノンパラメトリックベイズ 点過程と統計的機械学習の数理	講談社
978-4061529168	グラフィカルモデル	講談社
978-4061529175	バンディット問題の理論とアルゴリズム	講談社
978-4061529182	ウェブデータの機械学習	講談社
978-4061529199	データ解析におけるプライバシー保護	講談社
978-4061529212	関係データ学習	講談社
978-4061529274	音声認識	講談社
978-4061538214	イラストで学ぶ機械学習 最小二乗法による識別モデル学習を中心に	講談社
978-4061538252	イラストで学ぶディープラーニング	講談社
978-4274131493	わかりやすいパターン認識	オーム社
978-4274215308	続・わかりやすいパターン認識—教師なし学習入門—	オーム社
978-4274218026	進化計算と深層学習 一創発する知能—	オーム社
978-4274218873	機械学習と深層学習 —C 言語によるシミュレーション—	オーム社
978-4274219344	Chainer による実践深層学習	オーム社
978-4274219993	実装ディープラーニング	オーム社
978-4274220333	自然言語処理と深層学習 C 言語によるシミュレーション	オーム社
978-4274221071	Chainer v2 による実践深層学習	オーム社
978-4274221149	強化学習と深層学習 C 言語によるシミュレーション	オーム社
978-4320120365	情報検索アルゴリズム	共立出版
978-4339028515	Python で体験する深層学習	コロナ社
978-4839960889	TensorFlow で学ぶディープラーニング入門	マイナビ

(3) 固有値を降順に出力したとき, 上位 2 個の固有値に対応する固有ベクトル $\mathbf{e}_1, \mathbf{e}_2$ を得る.

(4) 2 個の固有ベクトルを基に基底行列 $P = [\mathbf{e}_1, \mathbf{e}_2]$ を得る.

以上の手順により得られた基底行列 P の i 番目のベクトルがアイテム a_i の 2 次元削減後の特徴ベクトルとなる.

5. 評価

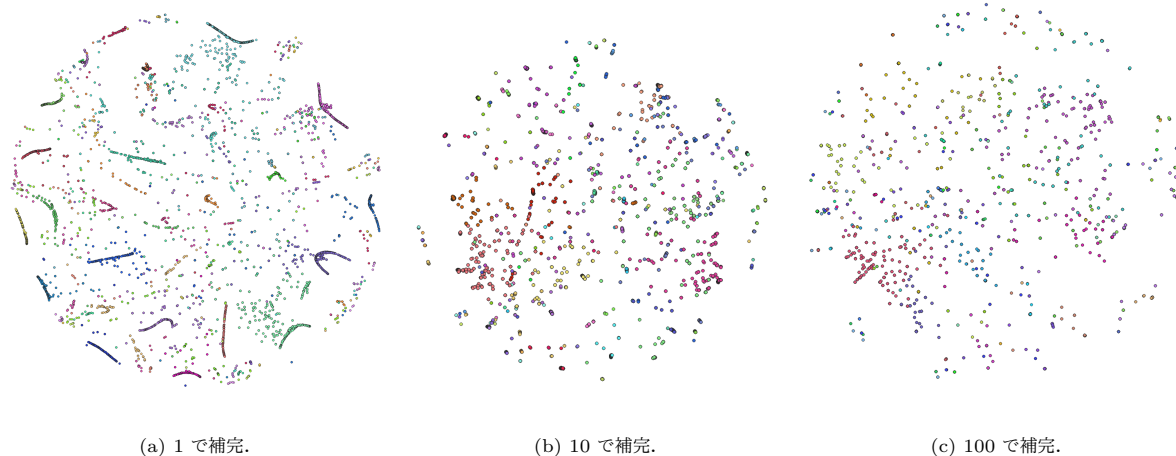
5.1 データセット

本実験では, 表 1 に示す 35 冊の技術書を対象とし, これらの技術書情報を **books** テーブルに登録した. 各技術書の索引から人手により索引語を抽出し, **terms** テーブルおよび **books-terms** テーブルに登録した. ユニークな索引語数は 5,263 であった. また, **books-terms** テーブルに登録されたタプル数は 9,072 であった.

5.2 欠損値の補完に関する予備実験

4.4 節では, 予備実験により欠損値を 10 で補完することであることを述べた. 本節では, その予備実験について説明する.

books-terms テーブルを基に作成した距離行列 D に含まれる欠損値を $\{1, 10, 100\}$ で補完したとき, それぞれで構築された 2 次元空間を検証する. 図 2 は, それぞれ欠損値を 1, 10, 100 で補完したときに構築した 2 次元空間である. 図中の点の一つの索引語を示す.



4.3 節で述べたとおり，同一の技術書内において t_i と t_j が最大限離れて出現するときには $d_{ij} = 1$ となる．欠損値を 1 で補完した場合，同一の技術書に出現しない用語対のページ距離がこの d_{ij} の値と一致する．そのため，同一の技術書内に存在するか否かにかかわらず同等の距離とみなされてしまう．

欠損値を 10 または 100 で補完した場合、このような問題は解消される。100 で補完した場合（図 2(b)）を検証する。補完値を [1, 10, 100] のいずれに設定した場合でも、欠損値の多い用語同士が塊となってマッピングされる問題点がある。しかし、100 で補完した場合は、ページ距離が存在する用語同士と比較し、塊に属する用語同士の距離が顕著に小さく表示される問題が生じる。これは、ページ距離の存在する用語同士に対して、存在しない用語同士のページ距離が極端に異なる値として扱われたことが原因と考えられる。そこで、本研究では、欠損値を 10 で補完した距離行列を用いてマッピングを行う。

5.3 定性分析

4.4 節で作成したマップについて，索引語間距離からそれらの関連を可視化することが可能であるかを検証する．検証は距離行列の算出方法および提示される関連語について行う．

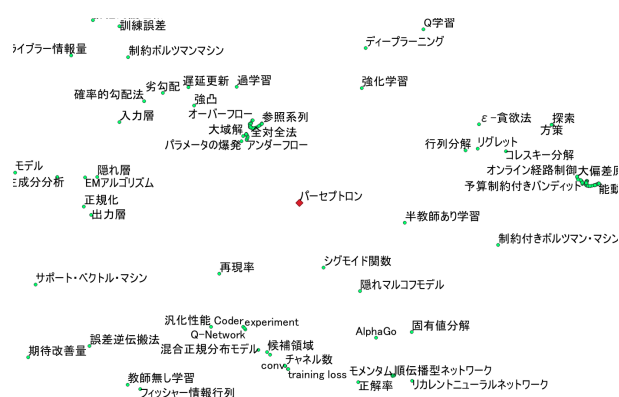
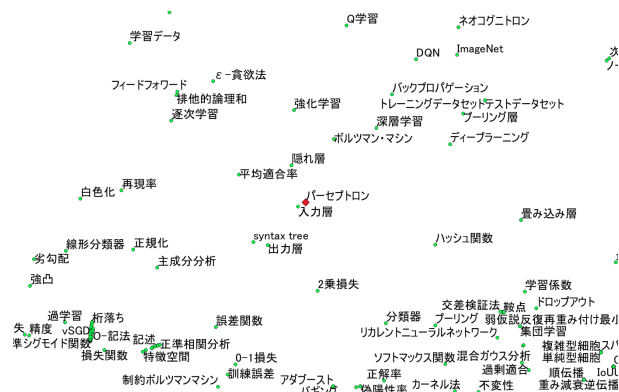
5.3.1 距離行列の算出方法についての検証

距離行列の算出方法についての検証を以下の 2 パターンで行う.

- (1) 4.3 節 線形方式で算出
- (2) 4.3 節 非線形方式で算出

1 の場合 (図 3), 関連語は近くに表示されているが, 関連語同士のほうが関連性が高い場合に, その距離が正確に表れていないという問題点がある. 原因として, 距離行列は索引語間距離をそのまま格納しているため, 複数の書籍に索引語間距離が存在する場合において外れ値があった場合, 顕著にそれが表れることが考えられる.

2 の場合 (図 4), 関連語が近くに表示されるだけでなく, 関連語同士の距離も比較的正確に表れている. これは, 1 で指摘した外れ値がある場合でも, マップ全体で見ればそれらは近い



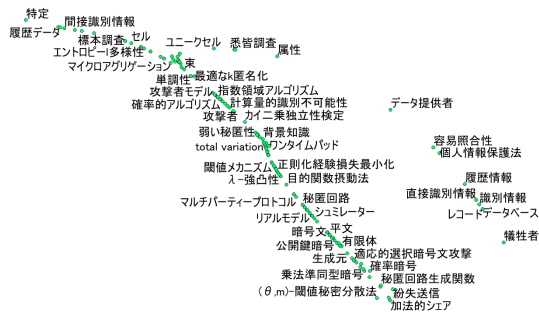


図 5 索引語の塊部分の拡大図

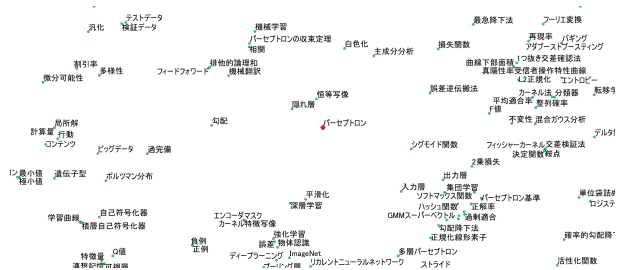


図 6 2 冊以上の技術書に掲載されている索引語のみを対象とし、2 で作成した距離行列に MDS を適用

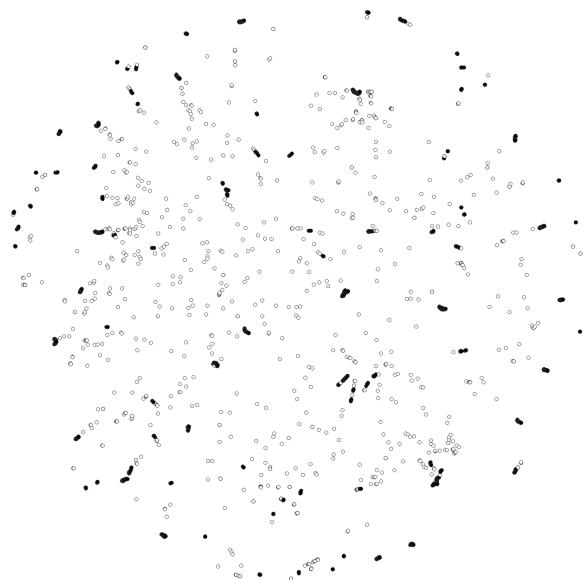


図 7 索引語マップを DBSCAN でクラスタリング
 白丸: どのクラスにも属さない独立した索引語
 黒丸: いずれかのクラスに属す索引語

ので，4.3 節 非線形方式により近似され，周囲と比較し極端な距離があっても表れにくくなった結果と考えられる．さらに，1 と比較して差は小さくなるが，他の関連語からの距離も考慮されることで比較的正確なマッピングが可能となる．

いずれの場合においても、1冊の技術書にしか掲載されていない索引語は他の技術書との距離データが存在せず、それらが塊となってマッピングされる問題がある。

そこで、欠損値削減のため、2冊以上の技術書に掲載されている索引語のみを抽出し、距離行列を作成、検証する。図6は、2冊以上の技術書に掲載されている索引語のみを抽出し、それらの索引語間距離を4.3節 非線形方式で算出した索引語間距離行列にMDSを適用したものである。この場合、大幅に欠損値が削減されるため、一部塊はみられるものの、全索引語を用いた場合と比較して大幅に解消可能である。反面、全索引語数(5,263語)と比較し、約82%(4,312語)が削減されるというデータ数の大幅な減少もみられた。

5.3.2 提示される関連語についての検証

5.3.1 小節の結果より、4.3 節 非線形方式で算出した距離行列に MDS を適用し作成した索引語マップを検証に用いる。

5.3.1 小節で述べたとおり、欠損値の多い索引語同士が塊となつてマッピングされる問題がある。索引語マップにおいて同一縮尺で塊となっている索引語とそうでない索引語を表示すると、塊となっている索引語が視認しづらくなる。そこで、DBSCAN を用いて索引語マップをクラスタリングすることにより、索引語同士の塊を判別する。

DBSCAN は、マップ上の点（索引語）を

- コア点: 半径 ε 以内に minPts 個以上の隣接点を持つ点
- 到達可能点: 半径 ε 以内に隣接点が存在するが minPts 個より少なく, 半径 ε 以内にコア点を持つ点
- 外れ値: 半径 ε 以内に隣接点がない点

に分類し、コア点から到達可能なすべての点とクラスタを形成するクラスタリング手法である。検証に用いるマップの領域は第一軸：-7.86 ~ 7.85, 第二軸：-7.96 ~ 7.68 である。検証では、半径 $\varepsilon = 0.05$, $\text{minPts} = 5$ と設定しクラスタリングを行う。

クラスタリングの結果(図7), 全索引語(5,263語)の内, 約84%(4,416語)がいずれかのクラスタに分類された. 1冊の技術書にしか掲載されていない索引語数(4,312語)より多くの索引語がいずれかのクラスタに分類された点については, 欠損値は比較的少ないものの類似性が特に高い索引語が同一クラスタに分類されたことが原因であると考えられる.

検索語に対して提示される関連語についての検証を以下の 2 パターンで行う.

- (1) どのクラスタにも属さない独立した索引語（外れ値）の関連語
但し、いずれかのクラスタに属す索引語は関連語から排除
- (2) 1冊の技術書にしか掲載されていない索引語と同一クラスタの関連語

1 の場合 (図 8), 検索語: パーセプトロン に対し,

- 誤差逆伝搬法
- 入力層

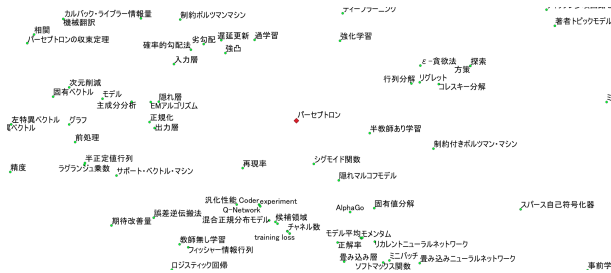


図 8 どのクラスタにも属さない独立した索引語およびその関連語
検索語：パーセプトロン

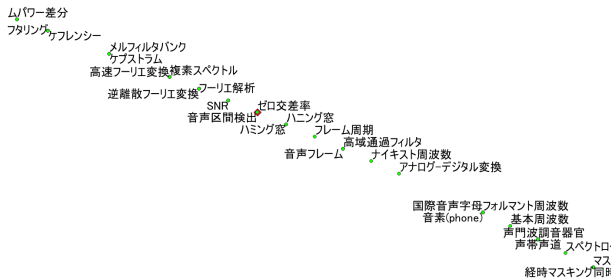


図 9 1冊の技術書にしか掲載されていない索引語と同一クラスタの関連語
検索語：音声区間検出

出力層

などの関連語が提示され、また、関連語同士の距離も正しく表示している。例示すると、パーセプトロンと入力層・出力層はそれぞれ関連するが、パーセプトロンと入力層・出力層との関連度と比較すると、入力層と出力層の関連度の方が高い。また、提示された関連語に関して、極端に関係しない用語は確認できない。これは、用いたデータセットが特定の分野に集中していたためである可能性がある。

2 の場合（図 9），検索語：音声区間検出 に対し、

- SNR（信号対雑音比）
- 音声パワー
- ゼロ交差率

などの関連語を提示できている。しかし、1 冊の技術書にしか掲載されていない索引語は、索引語間距離行列において欠損値が多いため、他の技術書の索引語との位置関係の算出が困難である。また、そのような索引語の場合、同一クラスタ内での位置関係は単純に索引語のページ順となるため、マップとして表示する利点は乏しいと考えられる。

6. おわりに

本研究では、技術用語の関係性を可視化するインタフェースの実現に向けて、技術書の索引から抽出した用語関係データに基づく技術用語マッピング手法を提案した。本稿における用語関係データとは、技術書における索引語の出現ページの差を基に算出したものである。提案手法では、用語関係データから技

術用語間の類似性を算出し、多次元尺度構成法を用いて技術用語を 2 次元空間にマッピングした。索引語間距離を算出可能な技術用語間において提示される関連語の妥当性を確認した。また、提案手法では索引語間距離を算出不可能な技術用語が索引語マップ全体に対して大きな影響を与えていることを確認した。

今後、提案手法で算出不可能な索引語間距離の扱いについて検討する。

文 献

- [1] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, Vol. 22, No. 1, pp. 5–53, 2004.
- [2] Iván Cantador, Alejandro Bellogín, and David Vallet. Content-based recommendation in social tagging systems. In *[RecSys2010]Proceedings of the 4th ACM conference on Recommender systems*, p. 237, 2010.
- [3] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor. *Recommender Systems Handbook*. Springer, 2011.
- [4] Marius Kaminskas, Ignacio Fernández-Tobías, Francesco Ricci, and Iván Cantador. Knowledge-based music retrieval for places of interest. In *Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies - MIRUM '12*, pp. 19–24, New York, New York, USA, 2012. ACM Press.
- [5] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, Vol. 7, No. 1, pp. 76–80, jan 2003.
- [6] Chenxing Yang, Baogang Wei, Jiangqin Wu, Y Zhang, and Liang Zhang. CARES: a ranking-oriented CADAL recommender system. In *[JCDL2009]Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*. ACM, pp. 203–212, 2009.
- [7] Sharon Givon and Victor Lavrenko. Predicting social-tags for cold start book recommendations. In *[RecSys2009]Proceedings of the 3rd ACM conference on Recommender systems*, p. 333, 2009.
- [8] Ahu Sieg, Bamshad Mobasher, and Robin Burke. Improving the Effectiveness of Collaborative Recommendation with Ontology-Based User Profile. In *RecSys HetRec 2010: Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, pp. 39–46, 2010.
- [9] Maria Soledad Pera. Automating Readers' Advisory to Make Book Recommendations for K-12 Readers. In *[RecSys2014]Proceedings of the 8th ACM conference on Recommender systems*, pp. 9–16, 2014.
- [10] Maria Soledad Pera and Yiu-kai Ng. Analyzing Book-Related Features to Recommend Books for Emergent Readers. In *[HT2015]Proceedings of the 26th ACM Conference on Hypertext and Social Media*, pp. 221–230, 2015.