

青山学院大学 社会情報学研究科 2019 年度 修士論文

単語ベクトルを用いた大学レコメンドシステム

学籍番号 38118002

氏名 北堀 達也

指導教員 宮治 裕 教授

2020 年 1 月

論文要旨

要旨には，論文の要約を記述します．要約と言っても全ての章や項目を均等に縮めるのではなく，必要な項目に絞って端的に示します．

論文の概要が，要旨に書かれた文章のみで伝わるようにしなければなりません．したがって，少なくとも「ざっくりとした背景」「研究の目的」「他の研究との違い/関わり」「構築したシステム/提案した事項」「結果/得られた結論」が書かれている必要があります．

謝辞

謝辞には，論文を書くにあたりお世話になった方々へ感謝の言葉を記述します．実は論文内で非常に良く見られる項目でもあるため，漏れが無いように気をつける必要があります．

少なくとも，指導をおこなった教員，一緒に学んだり励まし合ったりした同じ研究室のメンバーに対する感謝の気持ちを書くことをおすすめします．

たとえ，あまり感謝していなかったとしても，礼儀として書いておいた方が良いでしょう．論文は何十年も残るモノですから，誰に見られるかわからないということを想定して下さい．また，何年か後には皆さんの気持ちも変化するものですから，あとで後悔しないように慎重に記述して下さい．

宮治の場合には，上記の他に，両親や研究の際に利用したフリーソフト（今でいうオープンソースのソフトウェア）の作者にも感謝の気持ちを述べました．

目次

論文要旨	i
謝辞	ii
第1章 はじめに	1
1.1 背景	1
1.2 研究目的	4
1.3 関連研究	4
1.4 論文構成	4
第2章 本研究で用いた技術	5
2.1 Word2Vec	5
2.1.1 Word2Vec の構造	5
2.1.2 単語ベクトルの次元数	5
2.1.3 反復回数	6
2.1.4 window サイズ	6
2.1.5 本研究での利用	6
2.2 GloVe	6
2.2.1 概要	6
2.2.2 パラメータ	7
2.3 TF-IDF	7
2.3.1 Term Frequency	7
2.3.2 Inverse Document Frequency	7
2.3.3 本研究での利用	8

第3章	システム概要	9
3.1	本論文で構築したシステム	9
3.1.1	単語ベクトルのモデル	9
3.1.2	ストップワード	9
3.1.3	学部・偏差値データベース	9
3.1.4	キャンパス所在地データベース	9
3.2	システムの利用方法	9
3.2.1	大学名でキーワード検索	9
3.3	表の挿入	9
3.3.1	表の作成と変換方法	9
3.3.2	横罫線の設定	12
3.3.3	表内の基本部分の表示位置の変更	13
3.3.4	表の参照	14
3.3.5	表の位置の設定	14
3.3.6	縦罫線の設定と表内見出しなどの位置変更	14
3.4	参考文献と参照	16
第4章	検証実験	18
4.1	実験概要	18
4.2	評価軸	19
4.3	パラメータの詳細	19
4.3.1	単語ベクトルの次元数	19
4.3.2	反復回数	20
4.3.3	window サイズ	20
4.3.4	x-max	20
4.4	Word2Vec のモデル検証結果	20
4.4.1	WV_A	20
4.4.2	WV_B	21
4.4.3	WV_C	22
4.4.4	WV_D	23

4.4.5	WV_E	23
4.4.6	WV_F	24
4.4.7	WV_G	25
4.4.8	WV_H	25
4.5	GloVe のモデル検証結果	26
4.5.1	GV_A	27
4.5.2	GV_B	27
4.5.3	GV_C	28
4.5.4	GV_D	29
4.5.5	GV_E	29
4.5.6	GV_F	30
4.5.7	GV_G	30
4.5.8	GV_H	31
付録 A	プログラムの動作方法	33
A.1	ファイル構成	33
A.2	起動方法	33
A.3	表示の見方	33
参考文献		34

第1章

はじめに

本論文では，大学に関する記事から文書毎のベクトルを生成することで各大学毎の特色をデータ化し，潜在的な関係性を可視化する研究について記述する．

まず，本研究をおこなう背景となった事柄について述べる．次に，研究目的の詳細について述べ，最後に次章以降の本論文の構成について概略を述べる．

1.1 背景

近年，大学進学という選択肢は高校生にとって，一般的な選択肢として受け入れられるようになった．図 1.1 に過去 10 年間の大学・短大進学率と，大学(学部)進学率の推移を文部省公表 [1] の資料から示す．直近 3 年間の推移は比較的横ばい傾向にあるが，高校卒業後の進路に関して 50% 近い学生が大学へ進学している．

更に，2020 年 4 月から高等教育の修学支援制度 [2] が施行される．この新制度では，学生個人に対する要件と，支援対象者の所得に関する要件を満たした場合に授業料などを減免するか，給付型の奨学金を支給する．対象となる学校種は大学，短期大学，高等専門学校，専門学校となる．この制度の施行により，大学・専門学校進学率は今後増加すると考えられる．



図 1.1: 過去 10 年間の大学進学率推移

一方大学の学校数に関して，日本では774校存在している??．これは2019年4月の入学者を募集した大学の数である．それぞれの内訳としては，国立大学が82大学，公立大学が91大学，私立大学が592大学であり，私立大学が全体の約8割を占めている．このデータのうち，2019年度新設大学は13大学で，内訳は公立大学1校，私立大学10校，専門職大学2校に上る．また新設学部は国立，私立専門職大学合わせて61学部，新設学科は計118学科となっている．

志願者数の観点から，大学に関する志願者数の推移を日本私立大学振興・共済事業団の資料[3]からまとめた．

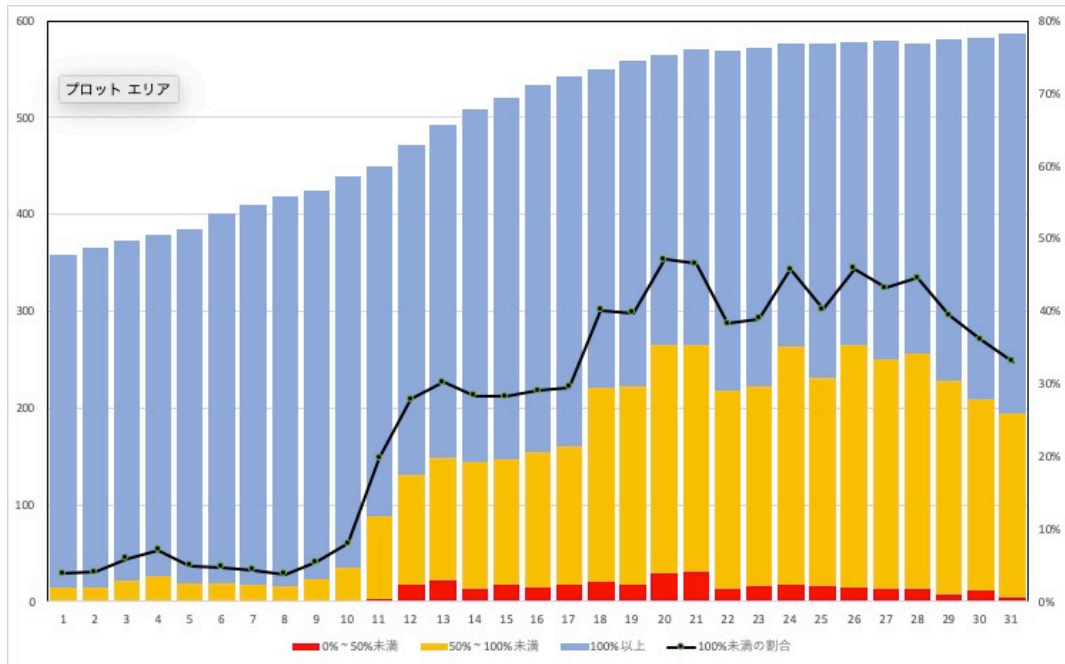


図 1.2: 平成元年～31 年の私立大学志願者数の推移

大学数は増加傾向にあったものの、ここ 10 年間はほぼ横ばいである。志願者数が募集人数を下回った大学の比率は近年減少傾向にある。これは 2018 年から大学に対して入学者の超過率を厳格に制限 [4] したためであると考えられる。入学者の超過率が厳格に制限されたため、入学者数が絞られる形になり、その分の学生が他の大学に流れた結果定員割れの大学数が減ったと考えられる。更に平成 31 年からは入学定員充足率が 0.9 ～ 1.0 倍の場合に入学定員充足率に応じて補助金が増額される。[5] そのため今後、有名大学の定員は減少傾向になると予想される。しかしそれでは志願者の一部が第一志望の学校から第二志望の学校になっただけであり、本質的な解決策としては限界がある。

ここで問題として、大学を選ぶ際の基準が画一化しているため、特定の大学に志願者が集中することとなっていると考えた。大学進学者数は増加傾向にあり、大学数も増加傾向にある。その中で志願者数が偏る原因は、大学選びの基準が画一化していることが原因の一つである。

1.2 研究目的

本研究の目的は、大学に関する記事から各大学の特徴をベクトルで表現し、潜在的な関係性を可視化することを目的とする。対象とするのは大学受験を控えた高校生で、特定の大学から様々な要素を通して大学間の関係性を可視化することで大学を選ぶ際の基準として機能するか考察する。

1.3 関連研究

類似研究（同じような研究）とは、どこが違うのか（ターゲット、手法、想定結果など）を述べる必要がある。また、参考にする先行研究（他組織の研究でも良い）とどのような関連性があるのかを述べる。

場合によっては、関連研究が研究目的より先に書いてあった方が「ながれ」が良い場合もある。また、関連研究を背景の中に入れてしまった方が良いケースもある。これらについては、文章を書きながら、判断するしかない。

1.4 論文構成

2章では本研究で使用した詳細な技術について説明する。3章では変研究で提案するシステムの説明と実際に構築したシステムの使用方法を解説する。4章ではシステムの有効性を検証した結果についてまとめる。最後に5章で有効性の考察を考察し、本研究のまとめと今後の課題について述べる。

第 2 章

本研究で用いた技術

本章では本研究で用いた詳細な技術について述べる。

2.1 Word2Vec

本節では Word2Vec の概要とパラメータについて説明する。Word2Vec[6] は、Tomas Mikolov ら [7] によって提案された、単語をベクトルに変換するためのニューラルネットワークの実装である。単語をベクトルで表現することで、単語同士の関連性を定量的に扱うことができる。またベクトルに変換することで、単語同士でベクトルの距離の足し引きができるようになるため、単語の演算が可能になる。

2.1.1 Word2Vec の構造

Word2Vec の構造は入力層、隠れ層、出力層からなる単純なニューラルネットワークとなっている。入力層と出力層は学習する単語の数だけ存在する。隠れ層はあらかじめ指定した次元数 \times 単語数 (入力層の数) のベクトルからなる。

入力層で受ける入力文章を 1-of-K 形式に変換したものとなり、出力結果が最適になるように隠れ層の単語ベクトルの重みを学習する。最終的に得られるモデルはこの隠れ層で学習した単語ベクトルになる。

2.1.2 単語ベクトルの次元数

word2vec で単語ベクトルを学習する際に指定する size オプションでは、隠れ層の単語ベクトルの次元数を指定する。このオプションで指定したサイズ * 全体

の単語数のサイズのベクトルに全ての単語を圧縮し、分散表現を得る。この次元数が大きすぎると効率的な分散表現を学習できないが、次元数が小さすぎると単語の特徴を十分にとらえきれなくなり、学習に時間を要する。

2.1.3 反復回数

`iter` オプションでは、学習の反復回数を指定する。反復回数が少ないと、最適な分散表現が得られる前に学習が終了してしまう。また、反復回数を増やすと学習に要する時間が増加する。検証実験では、最適な反復回数として、100 1000 回を比較対象とする。

2.1.4 window サイズ

ある単語の単語ベクトルを学習する際に、文書に出現した学習対象の単語から指定した単語数まで離れた単語を対象として学習する。`window` オプションではこの単語数を指定する。

2.1.5 本研究での利用

本研究で構築するシステムでは、大学間の関連を分析するために Word2Vec を用いる。具体的には、大学プレスセンター [8] の記事から、あらかじめリストアップした大学に関する記事をスクレイピングにより取得し、それぞれの記事を学習データとして Word2Vec のモデルを学習した。

2.2 GloVe

本節では GloVe について簡易的に解説する。

2.2.1 概要

GloVe は Jeffrey Pennington ら [9] によって提案された単語ベクトルを取得する実装である。Word2Vec よりも後に提案された手法であり、C 言語で実装されている。GloVe では、コーパス全体の単語間の共起の数を最小二乗法で学習するモデルとなっている。そのため、Word2Vec と比較して学習時間が短縮できる。ま

た，小さなコーパスでも学習ができ，精度の高さも論文の実験から得られている．Word2Vecでは考慮できない，出現回数が極端に少ない単語に重要度が偏るといった問題も回避できる．これはGloVeでは共起頻度が極端に高い単語と，低い単語を重視しないためである．

2.2.2 パラメータ

基本的なパラメータはWord2Vecと共通であるため割愛する．ここでは本研究で適用したx-maxオプションの値について説明する．x-maxオプションはGloVeの学習の際に指定するパラメータで，共起頻度の閾値を表す．(2.1)式で示すように，重み関数 $f(x)$ は2つの単語 i, j の共起頻度がx-maxオプションで指定した値未満の時に $(x/x_{max})^\alpha$ となり，それ以上の場合は1となる． α の値は論文から0.75とする．

$$f(x) = \begin{cases} (x/x_{max})^\alpha & \text{if } x < x_{max} \\ 1 & \text{otherwise} \end{cases} \quad (2.1)$$

これにより，共起頻度が低い単語は重みが低くなり，共起頻度の閾値以上の単語は重みが1になる．

2.3 TF-IDF

本節ではTF-IDFについて説明する．TF-IDFは...

2.3.1 Term Frequency

Term Frequencyは文書内における単語の出現頻度を表す．文書内である単語が出現する頻度が高ければ，その単語は重要であると考えられる．単語 t のTF値の計算方法は，文書 d 内の単語 t の出現回数を文書 d 内の全ての単語の出現回数の総和で割ることで求められる．

$$\text{tf}(t, d) = \frac{n_{t,d}}{\sum_{s \in d} n_{s,d}}$$

2.3.2 Inverse Document Frequency

Inverse Document Frequencyは逆文書頻度と呼ばれるもので，ある文書内の単語の，全体の文書における出現頻度の対数になっている．IDFが低いほど，他の文

書で出現しないため、重要な単語であると考えられる。単語 t の IDF 値の計算方法は、全ての文書数を単語 t が出現する文書の数で割った自然対数に 1 を足すことで求められる。1 が足されているのは、全ての文書に出現する単語の IDF 値が 0 にならないためである。

$$\text{idf}(t) = \log \frac{N}{df(t)} + 1$$

2.3.3 本研究での利用

本研究で構築したシステムでは、大学間で共通している近い意味の単語を学習した単語ベクトルから取得する。その際に普遍的な単語や、TF-IDF 値から重要度が低いと判断できる単語を削除するために TF-IDF を用いた。また、トピックモデルを推測する際に TF-IDF を利用して大学ごとの頻出単語をリストした。

第 3 章

システム概要

3.1 本論文で構築したシステム

概要図載せる

3.1.1 単語ベクトルのモデル

3.1.2 ストップワード

3.1.3 学部・偏差値データベース

3.1.4 キャンパス所在地データベース

3.2 システムの利用方法

本研究で構築したシステムの利用方法について述べる．

3.2.1 大学名でキーワード検索

3.3 表の挿入

本節では，表の挿入に関して記述する．また，表の作成方法や書式の変更方法についても実例を元に解説する．

3.3.1 表の作成と変換方法

馴れてくれば $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ の命令を直接記述して表を作るのは容易だが，最初の内は手こずるかもしれない．ここでは論文に用いる一般的な表を元に，一番簡単と思われる作成と貼り付けの方法を記載する．

表の作成

図 3.1 の様な表を仕上がりのイメージと仮定する .

データセット番号	抽出率	抽出数	データ数(全)	正答率	正答数	データ数(対象)
A	0.856	202	236	0.874	180	206
B	0.717	86	120	0.816	80	98
C	0.798	150	188	0.822	162	197
平均	0.790	N/A	N/A	0.837	N/A	N/A

図 3.1: 貼り付けたい表の仕上がりのイメージ

まず , Excel 等で表を作成しておく .

	A	B	C	D	E	F	G
1	データセット番号	抽出率	抽出数	データ数(全)	正答率	正答数	データ数(対象)
2	A	0.856	202	236	0.874	180	206
3	B	0.717	86	120	0.816	80	98
4	C	0.798	150	188	0.822	162	197
5	平均	0.790	N/A	N/A	0.837	N/A	N/A
6							

図 3.2: Excel で作成した表

表の変換

CSV2TeX(<http://naisodewafurenu.web.fc2.com/csv2tex.html>) に接続する .

CSV2TeX

CSV形式からTeXのtabular環境を自動的に作成します (タブ区切り形式からの変換にも対応)
ダウンロード不要、ブラウザ上で動きます。

↓ Excelからコピー＆ペーストした場合は、タブ区切り形式になっているはずなので[tab-delimited]を選択してください。

Input: Output:

図 3.3: CSV2TeX のサイト

Input のプルダウンメニューから「tab-delimited」を選択する .

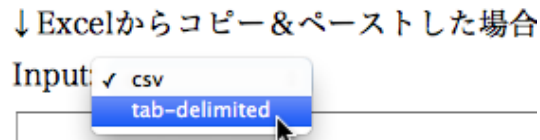


図 3.4: 入力方式の変更

その直下のボックスに Excel の表をペーストし , 「 CONVERT 」 ボタンを押下する . Output のボックス内に \LaTeX の表組みの命令が埋め込まれたデータが出力される .

Input:

tab-delimited

データセット番号	抽出率	抽出数	データ数 (全)	正答率	正答数	データ数 (対象)
A	0.856	202	236	0.874	180	206
B	0.717	86	120	0.816	80	98
C	0.798	150	188	0.822	162	197
平均	0.790	N/A	N/A	0.837	N/A	N/A

Output:

```
\begin{table}[htbp]
\begin{center}
\begin{tabular}{lllllll}
データセット番号 & 抽出率 & 抽出数 & データ数 (全) & 正答率 & 正答数 & データ数
(対象) \YY
A & 0.856 & 202 & 236 & 0.874 & 180 & 206\YY
B & 0.717 & 86 & 120 & 0.816 & 80 & 98\YY
C & 0.798 & 150 & 188 & 0.822 & 162 & 197\YY
平均 & 0.790 & N/A & N/A & 0.837 & N/A & N/A\YY
\end{tabular}
\end{center}
\end{table}
```

図 3.5: CSV2TeX の操作例

これを tex ファイルに貼り付ける .

```

\begin{table}[htbp]
\begin{center}
\begin{tabular}{lllllll}
データセット番号 & 抽出率 & 抽出数 & データ数 (全) & 正答率 & 正答数 & データ数 (対象) \\
A & 0.856 & 202 & 236 & 0.874 & 180 & 206 \\
B & 0.717 & 86 & 120 & 0.816 & 80 & 98 \\
C & 0.798 & 150 & 188 & 0.822 & 162 & 197 \\
平均 & 0.790 & N/A & N/A & 0.837 & N/A & N/A
\end{tabular}
\end{center}
\end{table}

```

表の修正

このままでは , キャプションとラベルが不足しているので , それらの情報を足すことにする . また , 現在の \LaTeX では , 図表内において $\text{\begin{center}}$ と $\text{\end{center}}$ の代わりに \centering 命令を使うことが推奨されている為 , それも

変更する．

まず，命令の上部

```
\begin{table}[htbp]
\begin{center}
\begin{tabular}{llllllll}
```

に対して，以下の様に修正する．

```
\begin{table}[htbp]
\caption{実験1の結果}
\centering
\begin{tabular}{llllllll}
```

また，命令下部

```
\end{tabular}
\end{center}
\end{table}
```

に対して，以下の様に修正する．

```
\end{tabular}
\label{table:resultEx1}
\end{table}
```

その出力を，表 3.1 に示す．

表 3.1: 実験 1 の結果

データセット番号	抽出率	抽出数	データ数（全）	正答率	正答数	データ数（対象）
A	0.856	202	236	0.874	180	206
B	0.717	86	120	0.816	80	98
C	0.798	150	188	0.822	162	197
平均	0.790	N/A	N/A	0.837	N/A	N/A

3.3.2 横罫線の設定

横罫線には「\hline」命令を利用する．横罫線を引きたい場所で，\hlineを入力する．二重線を引きたい場合には，\hline \hlineと記述すれば良い．

例えば，以下の命令を出力すると表 3.2 の様に出力される．

```
\begin{tabular}{llllllll}
\hline
```

```

データセット番号 & 抽出率 & 抽出数 & データ数 ( 全 ) & 正答率 & 正答数 & データ数 ( 対象 )
\\ \hline \hline
A & 0.856 & 202 & 236 & 0.874 & 180 & 206\\ \hline
B & 0.717 & 86 & 120 & 0.816 & 80 & 98\\ \hline
C & 0.798 & 150 & 188 & 0.822 & 162 & 197\\ \hline \hline
平均 & 0.790 & N/A & N/A & 0.837 & N/A & N/A\\ \hline
\end{tabular}

```

表 3.2: 実験 1 の結果

データセット番号	抽出率	抽出数	データ数 (全)	正答率	正答数	データ数 (対象)
A	0.856	202	236	0.874	180	206
B	0.717	86	120	0.816	80	98
C	0.798	150	188	0.822	162	197
平均	0.790	N/A	N/A	0.837	N/A	N/A

「`\begin{tabular}`」の次の行の「`\hline`」だけが記載されている行は、一番上の横罫線を示す。また表の一行目（見出し）の最後の部分をみると、改行「`\\`」をし、その後に 2 回罫線を引く「`\hline \hline`」命令が書かれている為、二重罫線が表示されている。その他の行は、必要に応じて 1 回または 2 回の「`\\ \hline`」が記載されている。

3.3.3 表内の基本部分の表示位置の変更

表内の文字の表示位置を制御している命令は、`\begin{tablar}{l1l1l1l1l}` の部分である。この`{l1l1l1l1l}`は、表が 7 列（1 が 7 個ある）でできており、それらを「1: 左寄せ」で表示することを意味している。

表内の表示位置の修正としては、表の大部分を占める部分（2 行目から 4 行目のデータが入っている部分）を元に、位置の指定をしていく。まず、1 列目「A と B と C」とかかれた部分の部分は文字情報なので、センタリング「`c`」にする。また、2 列目から 7 列目は、数値データが入っていることから、右寄せ「`r`」にする。したがって、位置の指定は、`\begin{tablar}{crrrrrrr}` とすれば良い。以上の修正をした表を表 3.3 に示す。

表 3.3: 実験 1 の結果

データセット番号	抽出率	抽出数	データ数 (全)	正答率	正答数	データ数 (対象)
A	0.856	202	236	0.874	180	206
B	0.717	86	120	0.816	80	98
C	0.798	150	188	0.822	162	197
平均	0.790	N/A	N/A	0.837	N/A	N/A

欧文の論文の場合には、表の縦線を入れないことが多い為、以上で表が完成したと考えて良い。

和文の論文に見られる様に、表の縦線を入れこともできるが、縦線を入れたが為に見出しの位置などが気になり、さらなる調整が必要になることが多い。通常は、ここまでで良いだろう。

3.3.4 表の参照

表の環境でもラベルが設定でき、図と同様の手法で参照することができる。例えば、表 3.1 の命令内では、「`\label{table:resultEx1a}`」と設定されているため、「表 `\ref{table:resultEx1a}`」の様に指定すれば表 3.1 と参照できる。

3.3.5 表の位置の設定

表の位置の設定は、図の位置の設定と同様「`[htbp]`」の様な指定ができる。である。また、`float` パッケージを読み込んでいる為、強制的にその場所に出力する「`[H]`」も利用できる。

3.3.6 縦罫線の設定と表内見出しなどの位置変更

縦方向罫線の入れ方と、それに伴う見出しなどの表示位置変更方法について記載する。

縦罫線の設定

列の要素のどこに縦罫線を引くのかを「`|`」を使って指示する。また、2重罫線は「`||`」を利用する。例えば、今回の仕上がりのイメージでは、1列目と4列目の右が2重罫線、残りと外側が通常の縦罫線だったので、

「`\begin{tabular}{|c||r|r|r||r|r|r|}`」と書く。

表 3.4: 実験 1 の結果

データセット番号	抽出率	抽出数	データ数 (全)	正答率	正答数	データ数 (対象)
A	0.856	202	236	0.874	180	206
B	0.717	86	120	0.816	80	98
C	0.798	150	188	0.822	162	197
平均	0.790	N/A	N/A	0.837	N/A	N/A

表内の見出し行などの部分の表示位置の変更

表の 1 行目 (ラベル部分) は, 2 行目から 4 行目と異なり, それぞれの列を説明する言葉が書かれている。これは, 表内右寄せでは無く, センタリングとしたい。

このように書かれている部分を (表示の関係から行を折り返している)

```
\begin{tabular}{|c||r|r|r||r|r|r|}
\hline
データセット番号 & 抽出率
& 抽出数 & データ数 (全)
& 正答率 & 正答数
& データ数 (対象) \\ \hline \hline
```

以下の様に変更する。

```
\begin{tabular}{|c||r|r|r||r|r|r|}
\hline
\multicolumn{1}{|c|}{データセット番号} & \multicolumn{1}{c|}{抽出率}
& \multicolumn{1}{c|}{抽出数} & \multicolumn{1}{c|}{データ数 (全)}
& \multicolumn{1}{c|}{正答率} & \multicolumn{1}{c|}{正答数}
& \multicolumn{1}{c|}{データ数 (対象)} \\ \hline \hline
```

なお, 命令を見て分かるように `multicolmun` 命令を利用する際には, 罫線情報は再度指定する必要がある。

また, 最後の行は「N/A」の文字は表内でセンタリングとしたい。

同様にこのように書かれている部分を (表示の関係から行を折り返している)

```
平均 & 0.790
& N/A & N/A
& 0.837 & N/A
```

```
& N/A \\ \hline
\end{tabular}
```

以下の様に修正する .

```
平均 & 0.790
& \multicolumn{1}{c|}{N/A} & \multicolumn{1}{c|}{N/A}
& 0.837 & \multicolumn{1}{c|}{N/A}
& \multicolumn{1}{c|}{N/A} \\ \hline
\end{tabular}
```

以上の修正を施すと , 表 3.5 の様に出力される .

表 3.5: 実験 1 の結果

データセット番号	抽出率	抽出数	データ数 (全)	正答率	正答数	データ数 (対象)
A	0.856	202	236	0.874	180	206
B	0.717	86	120	0.816	80	98
C	0.798	150	188	0.822	162	197
平均	0.790	N/A	N/A	0.837	N/A	N/A

3.4 参考文献と参照

参照命令に応じて拡張子が bib であるファイルから情報が読み取られ , 適切な番号が割り振られ , その番号順に参考文献が作成される . 例えば , 「湖上ら [10] は , 人間共生型ロボットの感性出力に関する研究を行った . 」という文章において , 参照の命令は「\cite{Kogami2009}」の様に記載されている . これで参考文献の出力順に応じた番号が自動的に割り振られ , 参考文献のページに適切なフォーマットにて出力がなされる .

この引用のラベルは , myrefs.bib ファイルにて , 以下の様に記載されている .

```
@article{Kogami2009,
author = "湖上 潤 and 宮治 裕 and 富山 健",
title = "人間共生ロボットにおける擬似感性システムの構築と評価",
journal = "日本感性工学会論文誌",
pages = "601-609",
month = sep,
year = 2009,
}
```

この BibTeX の書式は , 全て各自で記述しても構わないが , 一般的な論文をダ

ウンロードするサイトにおいて出力することができるようになっており、それを利用して良い。

上記例は論文に関する情報であるが、書籍(の一部)[11]、書籍[12]、予稿集[13]、その他(Webサイトなど)[14]で参考文献欄に載せる情報は異なる。それぞれの書式を記載しておいたので、各自で myrefs.bib ファイルを参照して欲しい。

第4章

検証実験

モデルの精度がどの程度妥当かを検証するために、本章ではパラメータを微調整したモデルの出力結果をまとめる。またそれぞれの出力結果に関して評価する。

4.1 実験概要

検証実験では、2通りの方法で生成したモデルを用いて出力を得た。1つはWord2Vecで生成したモデルで、もう一方はGloVeで生成したモデルである。また、それぞれの方法でモデルを生成する際に、異なるパラメータを適用していくつかのモデルを学習した。適用したパラメータの詳細は次節で説明する。

出力した内容は、モデルに対して大学名を入力とし、入力された大学に近い大学名を出力とした。また、入力と出力の2つの大学間において近さを定義する共通の単語について、それぞれの大学からの距離を示した。例として、A大学に近い大学としてB大学が得られた場合、A大学とB大学で共通して近い単語を探す。得られた単語からのそれぞれの大学間の距離が近ければ、それぞれの大学が近い要因としての単語を得ることができる。

モデルを用いて出力する内容は、青山学院大学に近い大学名、青山学院大学ーキリスト教に該当する大学名、青山学院大学と明治学院大学それぞれで共通する意味が近い単語の3項目とした。明治学院大学と比較した理由は、同じミッション系の大学で神学部の統合を通した日本神学校の創立などの関係性を持っているためである。

4.2 評価軸

モデルを評価する際に、それぞれの出力に対して評価軸を定めた。

青山学院大学に近い大学に関する評価項目は、まず偏差値が近いこと、ミッション系の大学であること、キャンパスの立地の近さである。本研究で参考にした偏差値とキャンパスの立地はパスナビ[15]を参考とした。モデルに学部、学科の情報は考慮されていないため、キャンパスの近さは学部を考慮しない。

次に、青山学院大学－キリスト教の評価項目は、非ミッション系の大学であること、偏差値が近いこと、キャンパスの立地が近いこととした。この場合、大学の要素からはキリスト教が消えていることが期待されるため、非ミッション系の大学であることをもっとも重要な評価項目とする。

最後に、青山学院大学と明治学院大学で共通の近い単語の評価項目は、直接的な関係性を示す単語とした。これは、両校がミッション系であることから、キリスト教に関する単語などがあげられる。また、歴史的な背景から日本神学校(現 東京神学大学)の創設に両校の神学部が統合したことから、神学部に関する単語も考慮する。

表 4.1: 各出力結果の評価軸

青山学院大学に近い大学	青山学院大学－キリスト教	青山学院大学と明治学院大学で共通の近い単語
偏差値が近い ミッション系 立地	非ミッション系 偏差値が近い 立地	直接的な関係性を示す単語

4.3 パラメータの詳細

各パラメータの詳細について解説する。

4.3.1 単語ベクトルの次元数

Word2Vec と GloVe で指定する次元数は、小さすぎると単語の特徴を効率的に学習できず、大きすぎると適切な分散表現が学習できない。一般的に、50 300 次元を指定する。本研究で使用したデータセットは比較的サイズが小さいた

め、単語ベクトルの次元数は50次元とした。

4.3.2 反復回数

Word2Vec と GloVe のトレーニングの反復回数を指定する。この数字の大きさに比例して学習に要する時間も大きくなる。また、反復回数が少なすぎると十分に単語の特徴を学習できないため、検証実験では10, 100, 1000のパラメータで比較する。

4.3.3 window サイズ

window サイズは10と1000で比較した。一般的にはwindow サイズは10 20で学習するが、記事の中に出現する大学名の単語ベクトルを学習する際、対象となる単語は記事全体に出現すると考えられるためwindow サイズに1000を適用して比較する。

4.3.4 x-max

GloVe の学習を行う際に、共起頻度の閾値を指定する必要がある。Jeffrey Pennington らの実験では、100,000,000 600,000,000 個のトークンが含まれたコーパスを用いて、x-max オプションに100を指定した。一方本論文で学習したデータは約2,400,000個のトークンが含まれたデータを用いたため、x-max オプションに指定する値は10と5で比較する。

4.4 Word2Vec のモデル検証結果

Word2Vec のモデル検証結果を示す。今回検証したモデル名と、対応するパラメータの一覧を表 4.2 に示す。

4.4.1 WV_A

単語ベクトルの次元数が50次元、反復回数とwindow サイズがそれぞれ10で生成したモデルの結果を表 4.3 に示す。

青山学院大学に2番目に近い大学として明治学院大学が出現したが、それ以

表 4.2: Word2Vec パラメータの詳細

モデル名	単語ベクトルの次元数	反復回数	window サイズ
WV_ A	50	10	10
WV_ B	100	10	10
WV_ C	50	100	10
WV_ D	100	100	10
WV_ E	50	10	1000
WV_ F	100	10	1000
WV_ G	50	100	1000
WV_ H	100	100	1000

外の結果は関連性が不透明である．この結果からは有効なモデルだと判断できない．

表 4.3: WV_ A の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
聖心女子大学	0.74	目白大学	0.62	女学院	1.094
明治学院大学	0.722	帝京大学	0.583	英和	0.968
昭和女子大学	0.704	芝浦工業大学	0.552	芸術	0.742
帝京大学	0.703	女子美術大学	0.546	ライン	0.634
清泉女子大学	0.691	東京薬科大学	0.545	山手	0.579
実践女子大学	0.649	東京家政大学	0.544	校友	0.439
白百合女子大学	0.649	東京情報大学	0.542		
津田塾大学	0.638	東京電機大学	0.529		
東京女子体育大学	0.637	東京理科大学	0.525		
女子美術大学	0.636	多摩美術大学	0.524		

4.4.2 WV_ B

単語ベクトルの次元数が100次元，反復回数とwindowサイズがそれぞれ10で生成したモデルの結果を4.4に示す．

青山学院大学に近い大学として，ミッション系の大学で偏差値も比較的近いと考えられる上智大学が得られた．また青山学院大学－キリスト教の結果から上位10校にはミッション系の大学が含まれなかった．さらに，青山学院大学と明治学院大学で共通の近い意味の単語から，キリスト教などの単語が得られた．

しかし青山学院大学に近い大学の順位を全体的に見ると、関連性があると断言できない。

表 4.4: WV_ B の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
上智大学	0.372	成蹊大学	0.317	定期	0.531
東京外国語大学	0.308	駒澤大学	0.293	神学	0.45
駒澤大学	0.308	獨協大学	0.262	芸術	0.415
東京大学	0.295	東京理科大学	0.253	監督	0.37
筑波大学	0.294	中央大学	0.252	キリスト	0.364
早稲田大学	0.289	名古屋大学	0.221	イギリス	0.363
中央大学	0.276	筑波大学	0.204	キリスト教	0.323
オックスフォード大学	0.271	明治大学	0.19	合同	0.304
東京農業大学	0.256	大阪大学	0.167	前期	0.291
成蹊大学	0.256	関西学院大学	0.167	チャペル	0.272

4.4.3 WV_ C

単語ベクトルの次元数が50次元、反復回数が100、windowサイズを10で生成したモデルの結果を4.5に示す。

青山学院大学に近い大学で上智大学や明治学院大学などが出現したが、単語ベクトルの類似度の高さでは清泉女子大学と聖心女子大学よりも低い。どちらもミッション系の大学であるが、偏差値などの観点から上智大学や明治学院大学の方が近いと考えられる。

青山学院大学と明治学院大学で共通の近い単語では、神学という単語が出現した。しかし、それぞれの大学との類似度の合計は他の単語に対して高くないため、学習不足であると考えられる。

表 4.5: WV_C の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
清泉女子大学	0.677	東京薬科大学	0.557	女学院	0.981
聖心女子大学	0.674	帝京大学	0.539	英和	0.893
上智大学	0.633	中央大学	0.525	芸術	0.819
明治学院大学	0.616	目白大学	0.516	ライン	0.625
帝京大学	0.599	東京情報大学	0.498	学位	0.538
大東文化大学	0.585	女子美術大学	0.472	山手	0.538
立教大学	0.58	桜美林大学	0.469	校友	0.515
実践女子大学	0.578	東京電機大学	0.46	神学	0.463
桜美林大学	0.563	成蹊大学	0.458		
昭和女子大学	0.561	工学院大学	0.45		

4.4.4 WV_D

単語ベクトルの次元数が100次元，反復回数が100，windowサイズを10で生成したモデルの結果を4.6に示す．

青山学院大学に近い大学として，上智大学や中央大学，学習院大学等が得られた．しかし類似度が高い大学で仏教系大学である駒澤大学などが出現した．青山学院大学と明治学院大学で共通の近い単語はキリストや礼拝，教会などミッション系の大学を連想させるような単語が新しく出現した．

表 4.6: WV_D の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
上智大学	0.393	駒澤大学	0.337	キリスト	0.493
駒澤大学	0.364	中央大学	0.314	定期	0.491
東京外国語大学	0.304	獨協大学	0.261	キリスト教	0.441
中央大学	0.3	関西学院大学	0.244	神学	0.433
東京農業大学	0.288	東洋大学	0.224	礼拝	0.424
関西学院大学	0.283	成蹊大学	0.208	宗教	0.416
早稲田大学	0.254	名古屋大学	0.178	芸術	0.39
筑波大学	0.253	明治大学	0.177	教会	0.358
獨協大学	0.253	筑波大学	0.166	基本	0.34
学習院大学	0.249	法政大学	0.165	チャペル	0.311

4.4.5 WV_E

単語ベクトルの次元数が50次元，反復回数が10，windowサイズを1000で生成したモデルの結果を4.7に示す．

このモデルでは、青山学院大学に近い大学として中央大学、法政大学、立教大学、明治大学が類似度の高い大学としてあげられた。また、青山学院大学－キリスト教はミッション系の大学が出現せずに法政大学や明治大学を得ることができた。

しかし、青山学院大学と明治学院大学で共通の近い単語から、神学やミッション系の大学に関する単語が出現しなくなった。

表 4.7: WV_E の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
東洋大学	0.69	法政大学	0.558	芸術	0.787
日本女子大学	0.651	明治大学	0.527	音楽	0.741
聖心女子大学	0.636	立命館大学	0.49	イギリス	0.732
立命館大学	0.623	東洋大学	0.482	女学院	0.714
昭和女子大学	0.623	昭和女子大学	0.465	バス	0.711
中央大学	0.607	千葉工業大学	0.463	コミュニティ	0.643
法政大学	0.601	横浜市立大学	0.459	相談	0.601
立教大学	0.595	中央大学	0.455		
明治大学	0.584	金沢工業大学	0.455		
東京電機大学	0.569	神奈川大学	0.433		

4.4.6 WV_F

単語ベクトルの次元数が100次元、反復回数が10、windowサイズを1000で生成したモデルの結果を4.8に示す。

青山学院大学に近い大学は近い偏差値の大学やミッション系の大学が多く得られた。青山学院大学と明治学院大学で共通の近い単語から、商業という単語が新しく得られた。これは1944年に専門部を閉鎖し、明治学院に合同した際の高等商業学部から関連性があると考えられる。

表 4.8: WV_F の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
明治大学	0.729	明治大学	0.426	心理	0.857
北里大学	0.713	法政大学	0.423	併設	0.85
法政大学	0.707	北里大学	0.385	統合	0.843
明治学院大学	0.66	九州大学	0.362	商業	0.817
短期大学	0.658	立命館大学	0.341	前期	0.794
上智大学	0.654	名古屋大学	0.337	キリスト教	0.791
立教大学	0.65	駒澤大学	0.331	イギリス	0.72
中央大学	0.6	大阪大学	0.321	教会	0.679
早稲田大学	0.567	早稲田大学	0.317	神学	0.612
日本女子大学	0.562	中央大学	0.317	キリスト	0.606

4.4.7 WV_G

単語ベクトルの次元数が50次元，反復回数が100，windowサイズを1000で生成したモデルの結果を4.9に示す．

青山学院大学に近い単語は九州大学や東北大学，名古屋大学など立地的に遠い大学が出現した．また青山学院大学－キリスト教では，ミッション系の大学である同志社大学が得られたため，モデルの有効性は低いと考えられる．

表 4.9: WV_G の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
九州大学	0.708	九州大学	0.556	イギリス	1.07
東北大学	0.52	名古屋大学	0.424	前期	1.005
明治学院大学	0.503	東北大学	0.392	キリスト教	0.902
上智大学	0.451	東京大学	0.315	基本	0.883
名古屋大学	0.448	大阪大学	0.299	芸術	0.834
短期大学	0.442	同志社大学	0.289	神学	0.824
東京大学	0.44	立命館大学	0.256	併設	0.818
明治大学	0.42	中央大学	0.246	キリスト	0.808
立教大学	0.41	明治大学	0.232	教会	0.774
東京農業大学	0.41	日本女子大学	0.231	山手	0.688

4.4.8 WV_H

単語ベクトルの次元数が100次元，反復回数が100，windowサイズを1000で生成したモデルの結果を4.10に示す．

青山学院大学に近い大学で，偏差値の近い明治大学や中央大学が得られた．

また，同じミッション系の大学として，明治学院大学や立教大学，同志社大学，上智大学も得られた．一方で，青山学院大学－キリスト教でミッション系の大学の同志社大学や，オックスフォード大学などの大学が得られた．

青山学院大学と明治学院大学で共通の近い単語では，ミッション系の大学に関する単語や神学という単語とともに，新しく統合という単語が得られた．この単語は神学部や専門部の統合という歴史的な背景から得られたと考えられる．

表 4.10: WV_H の検証結果

青山学院大学に近い大学		青山学院大学－キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
明治大学	0.494	明治大学	0.263	キリスト教	0.84
明治学院大学	0.457	北里大学	0.255	イギリス	0.764
立教大学	0.435	東京外国語大学	0.23	キリスト	0.703
同志社大学	0.429	同志社大学	0.225	教会	0.671
北里大学	0.428	学習院大学	0.219	前期	0.661
東京外国語大学	0.424	オックスフォード大学	0.197	神学	0.655
関西学院大学	0.377	東北大学	0.189	統合	0.613
短期大学	0.375	名古屋大学	0.185	併設	0.608
上智大学	0.344	九州大学	0.175	山手	0.606
中央大学	0.337	関西学院大学	0.156	チャペル	0.605

4.5 GloVe のモデル検証結果

本節では GloVe のモデル検証結果をまとめた．GloVe で生成したモデル名と，対応するパラメータの一覧を表 4.11 に示す．

表 4.11: GloVe パラメータの詳細

モデル名	単語ベクトルの次元数	反復回数	window サイズ
GV_A	50	10	10
GV_B	100	10	10
GV_C	50	100	10
GV_D	100	100	10
GV_E	50	10	1000
GV_F	100	10	1000
GV_G	50	100	1000
GV_H	100	100	1000

4.5.1 GV_A

単語ベクトルの次元数が50次元，反復回数が10，windowサイズを10で生成したモデルの結果を4.12に示す．

このモデルからは，青山学院大学に近い大学として中央大学や立教大学，上智大学，明治大学，法政大学など，偏差値が近い大学とミッション系の大学が取得できた．また，青山学院大学－キリスト教では，上位10校からはミッション系の大学が出現しなかった．しかし東京工業大学や大阪体育大学などは関連性が低いと考えられる．国際連合大学は大学ではないが，大学院は存在し青山学院大学と立地は非常に近いと出現したと考えられる．

青山学院大学と明治学院大学で共通の近い単語に関しては，学習不足であった．

表 4.12: GV_A の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
中央大学	0.764	中央大学	0.559	学位	0.922
立教大学	0.747	東京工芸大学	0.557		
上智大学	0.71	立教大学	0.537		
駒澤大学	0.71	法政大学	0.506		
明治大学	0.683	国際連合大学	0.499		
学習院大学	0.672	関西大学	0.482		
法政大学	0.671	明治大学	0.482		
実践女子大学	0.623	大阪芸術大学	0.472		
東京理科大学	0.616	大阪体育大学	0.471		
成蹊大学	0.609	金沢医科大学	0.461		

4.5.2 GV_B

単語ベクトルの次元数が100次元，反復回数が10，windowサイズを10で生成したモデルの結果を4.13に示す．

このモデルでは学習不足により青山学院大学と明治学院大学で共通の近い単語が得られなかった．青山学院大学－キリスト教の結果から，中央大学を除いて関連性の低い大学が出現した．

表 4.13: GV_ B の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
中央大学	0.639	東京工芸大学	0.52		
立教大学	0.588	中央大学	0.471		
学習院大学	0.562	大阪体育大学	0.46		
大東文化大学	0.551	大東文化大学	0.449		
明治大学	0.545	大阪芸術大学	0.435		
上智大学	0.537	神戸市外国語大学	0.415		
愛知淑徳大学	0.532	金沢医科大学	0.412		
大阪体育大学	0.53	京都女子大学	0.408		
芝浦工業大学	0.513	名城大学	0.407		
駒澤大学	0.496	会津大学	0.396		

4.5.3 GV_ C

単語ベクトルの次元数が50次元, 反復回数が100, window サイズを10で生成したモデルの結果を4.14に示す.

青山学院大学に近い大学として得られた大学は偏差値, 立地の観点から妥当なものと考えられる. しかし青山学院大学-キリスト教から, 上智大学と立教大学が得られた. これらの大学はミッション系であるため, モデルの妥当性は低いと考えられる. また, このモデルでは学習不足により青山学院大学と明治学院大学で共通の近い単語が得られなかった.

表 4.14: GV_ C の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
立教大学	0.691	中央大学	0.476		
中央大学	0.667	明治大学	0.466		
上智大学	0.632	成蹊大学	0.437		
明治大学	0.616	法政大学	0.426		
駒澤大学	0.612	武蔵大学	0.407		
法政大学	0.601	上智大学	0.404		
学習院大学	0.563	駒澤大学	0.391		
早稲田大学	0.497	立教大学	0.366		
成蹊大学	0.479	福岡大学	0.354		
東京理科大学	0.476	成城大学	0.347		

4.5.4 GV_D

単語ベクトルの次元数が100次元，反復回数が100，windowサイズを10で生成したモデルの結果を4.15に示す．

青山学院大学に近い大学として新しく大東文化大学が得られた．しかし評価項目を考慮すると，法政大学や上智大学よりも類似度が高い結果は妥当性が低いと考えられる．また，このモデルでは学習不足により青山学院大学と明治学院大学で共通の近い単語が得られなかった．

表 4.15: GV_D の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
中央大学	0.642	中央大学	0.55		
明治大学	0.549	法政大学	0.412		
立教大学	0.542	明治大学	0.377		
学習院大学	0.539	関西大学	0.375		
大東文化大学	0.52	淑徳大学	0.37		
上智大学	0.52	武蔵大学	0.359		
成城大学	0.509	東洋大学	0.355		
駒澤大学	0.486	東京理科大学	0.349		
法政大学	0.474	駒澤大学	0.348		
芝浦工業大学	0.455	東京工科大学	0.348		

4.5.5 GV_E

単語ベクトルの次元数が50次元，反復回数が10，windowサイズを1000で生成したモデルの結果を4.16に示す．

青山学院大学に近い大学からは，女子美術大学がもっとも類似度が高い大学として得られた．青山学院大学との関連性はキャンパス所在地が相模原市という点があげられる．青山学院大学－キリスト教の結果からは相模女子大学が得られた．キャンパスの立地が近いための結果であると考えられる．

青山学院大学と明治学院大学で共通の近い単語は，ミッション系の大学を連想する単語が出現したものの，類似度の合計値が高い単語は関連性が不明確なものであった．

表 4.16: GV_E の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
女子美術大学	0.647	相模女子大学	0.442	芸術	1.024
立教大学	0.631	神奈川工科大学	0.396	コミュニティ	0.93
聖心女子大学	0.622	芝浦工業大学	0.391	音楽	0.925
東京工業大学	0.615	北海道教育大学	0.389	学位	0.874
津田塾大学	0.613	女子美術大学	0.373	礼拝	0.854
上智大学	0.611	東京農業大学	0.372	心理	0.854
関西学院大学	0.61	国際連合大学	0.362	合同	0.825
明治学院大学	0.605	目白大学	0.362		
帝京大学	0.602	創価大学	0.359		
東京農業大学	0.599	帝京大学	0.351		

4.5.6 GV_F

単語ベクトルの次元数が100次元，反復回数が10，windowサイズを1000で生成したモデルの結果を4.17に示す．

青山学院大学に近い大学からは，関西学院大学や聖心女子大学，明治学院大学などのミッション系の大学が得られた．関西学院大学は立地を除いて青山学院大学に近い大学であると考えられる．また，青山学院大学と明治学院大学で共通の近い単語からは，関係性のある単語が取得できなかった．

表 4.17: GV_F の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
関西学院大学	0.541	神奈川工科大学	0.344	女学院	0.753
聖心女子大学	0.525	相模女子大学	0.334	英和	0.742
早稲田大学	0.514	北海道教育大学	0.333	芸術	0.731
明治学院大学	0.513	目白大学	0.315	前期	0.705
実践女子大学	0.503	国際連合大学	0.315	音楽	0.676
東京工業大学	0.495	関西学院大学	0.299	コミュニティ	0.674
女子美術大学	0.492	関東学院大学	0.293	貢献	0.631
津田塾大学	0.489	会津大学	0.285		
帝京大学	0.485	大妻女子大学	0.283		
東京大学	0.474	聖心女子大学	0.279		

4.5.7 GV_G

単語ベクトルの次元数が50次元，反復回数が100，windowサイズを1000で生成したモデルの結果を4.18に示す．

このモデルでも青山学院大学に近い大学として関西学院大学が得られた。また、青山学院大学と明治学院大学で共通の近い単語からは、関係性のある単語が取得できなかった。

表 4.18: GV_G の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
関西学院大学	0.505	神奈川工科大学	0.41	女学院	0.63
女子美術大学	0.497	東京農業大学	0.38	芸術	0.612
聖心女子大学	0.47	明治大学	0.377	英和	0.572
立教大学	0.459	相模女子大学	0.372	学位	0.485
国際大学	0.456	東京情報大学	0.369		
相模女子大学	0.445	帝京大学	0.355		
同志社大学	0.445	神奈川大学	0.35		
津田塾大学	0.425	帝京平成大学	0.347		
愛知淑徳大学	0.421	東京農工大学	0.333		
東京情報大学	0.411	福島県立医科大学	0.328		

4.5.8 GV_H

単語ベクトルの次元数が100次元、反復回数が100、windowサイズを1000で生成したモデルの結果を4.19に示す。

このモデルでは、青山学院大学にもっとも近い大学として、関西学院大学が得られた。青山学院大学－キリスト教の結果からは、もっとも近い大学として関西学院大学が得られたが、ミッション系の大学であるため、モデルの精度は妥当ではないと考えられる。青山学院大学と明治学院大学で共通の近い単語からは、ミッション系の大学を連想する単語と、統合という歴史的な背景を連想する単語が得られた。また定期という単語からは、1970年から1987年まで行われていた総合定期戦からの結果であると考えられる。

表 4.19: GV_H の検証結果

青山学院大学に近い大学		青山学院大学 - キリスト教		青山学院大学と明治学院大学で共通の近い単語	
大学名	類似度	大学名	類似度	単語	類似度の合計
関西学院大学	0.353	関西学院大学	0.315	聖書	0.471
東京外国語大学	0.322	東京農業大学	0.272	定期	0.458
東京農業大学	0.271	神戸大学	0.214	統合	0.455
静岡大学	0.27	東京理科大学	0.213	女学院	0.451
青山学院女子短期大学	0.257	横浜市立大学	0.213	英和	0.426
オックスフォード大学	0.242	静岡大学	0.18	チャペル	0.369
神戸大学	0.235	昭和女子大学	0.179	前期	0.365
東京工業大学	0.229	北海道大学	0.179	キリスト	0.359
九州大学	0.229	明治大学	0.175	キリスト教	0.357
筑波大学	0.221	九州大学	0.175	申し出	0.351

付録 A

プログラムの動作方法

本研究にて用いたプログラムについて解説する．

A.1 ファイル構成

プログラムのフォルダ内は，主に4つのファイルから構成される．

ああああいいいい

ううううええええ

これらを に設置し，以下の手順にそって起動する．

A.2 起動方法

まず，ウェブサーバを動かした状態にし，外部クライアント（Web ブラウザから），以下の URL にアクセスする．

A.3 表示の見方

実験に利用するための，実行結果は test.log ファイルに出力されている．

このファイルは4つのカラムからなる CSV 形式のファイルである．第1列には，...

参考文献

- [1] 文部科学省. 平成 30 年度画稿基本調査(確定値)の好評について. https://www.mext.go.jp/component/b_menu/other/_icsFiles/afieldfile/2018/12/25/1407449_1.pdf.
- [2] 文部科学省. 高等教育の修学支援新制度. https://www.mext.go.jp/a_menu/koutou/hutankeigen/index.htm.
- [3] 日本私立学校振興・共済事業団. 平成 31(2019) 年度私立大学・短期大学等入学志願動向. <https://www.shigaku.go.jp/files/shigandoukouH31.pdf>.
- [4] 文部科学省. 平成 28 年度以降の定員管理に係る私立大学等経常補助金の取り扱いについて(通知). https://www.mext.go.jp/a_menu/koutou/shinkou/07021403/002/002/_icsFiles/afieldfile/2015/07/13/1360007_2.pdf.
- [5] 文部科学省. 平成 31 年度以降の定員管理に関わる私立大学等経常補助金の取り扱いについて(通知). https://www.mext.go.jp/a_menu/koutou/shinkou/07021403/002/002/_icsFiles/afieldfile/2018/09/19/1409177.pdf.
- [6] 西尾泰和. word2vec による自然言語処理. 株式会社オライリー・ジャパン, 2014.
- [7] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. Technical report, Google Inc. Mountain View, CA, 9 2013.
- [8] 大学プレスセンター. <https://www.u-presscenter.jp/>.
- [9] Jeffrey Pennington, Richard Socher, and Christopher D.Manning. Glove: Global vectors for word representation. Technical report, Computer Science Department, Stanford University, Stanford, CA 94305, 2014.
- [10] 湖上潤, 宮治裕, 富山健. 人間共生ロボットにおける擬似感性システムの構築と評価. 日本感性工学会論文誌, pp. 601–609, September 2009.

- [11] 富山 健, 宮治 裕. 介護者支援ロボットシステムの提案. 小原 信, 神長 勲 (編), 日本福祉, 第 13 章, pp. 301–329. 以文社, July 2001. (ISBN4-7531-0217-3).
- [12] 中田 亨. 理系のための「即効！」論文術. 講談社, 2010.
- [13] Yutaka Miyaji and Ken Tomiyama. Construction of virtual kansei by petri-net with ga and method of constructing personality. In *Proceedings ROMAN2003, 12th IEEE Workshop Robot and Human Interactive Communication*, pp. 6B4(CD-ROM), November 2003.
- [14] 電気通信大学佐藤研究室. Tex マニュアル. [http://hs.hc.uec.ac.jp/index.php?plugin=attach&refer=Tex&openfile=TeX マニュアル.pdf](http://hs.hc.uec.ac.jp/index.php?plugin=attach&refer=Tex&openfile=TeX%20マニュアル.pdf) (参照 2013-10-31).
- [15] 大学受験パスナビ. <https://passnavi.evidus.com/>.