

文書ベクトル演算に基づく 考察型レポートの自動評価とフィードバック生成

細川 大和[†] 服部 峻^{††}

^{†,††}室蘭工業大学 ウェブ知能時空間研究室 〒050-8585 北海道室蘭市水元町 27-1

E-mail: [†]18024157@mmm.muroran-it.ac.jp, ^{††}hattori@csse.muroran-it.ac.jp

あらまし 近年、情報技術の発展と共に、教育現場にも ICT を活用する動きが活発になっており、例として、e-ラーニングシステム上で教員と学生間での講義情報や提出物の管理などが行われている。教育の場において、教員は学生に対して課題を与え、学生は与えられた課題に取り組み、教員は学生の課題に対して評価点やフィードバックを与えている。これらは学生の習熟度を高めることが期待出来るが、教員と学生間の人数比によって、迅速かつ適切に行うことが困難な場合がある。そこで、本稿ではこの問題を解決するために、学生が提出する考察型レポートを文書ベクトルへ変換し、文書ベクトル演算を用いることで評価点の推定やフィードバックの生成を行うシステムを提案する。

キーワード 自然言語処理, 教育工学, EdTech, 文書ベクトル化, 文書ベクトル演算, 自動採点, フィードバック

1 まえがき

近年、AI を始めとした情報技術の発展の中で、人間が行っていた仕事を機械に置き換えることによって、時間的・経済的に効率化を図ることが様々な分野において行われている。

教育現場の中で、指導者である教員は、自身が開講する講義の中で問題演習や試験、レポートなど、様々な課題を受講者である学生に与えている。そして、学生は教員より与えられた課題に取り組み、教員へ提出し、教員は学生の提出した課題に対して評価の結果やフィードバックを与えている。

ここで教員が学生に与える「評価」や「フィードバック」の例として、「評価」では課題に対する評価点、「フィードバック」では、文章の添削指導、見落としているポイントを気づかせるためのヒントであるとする。このような「評価」や「フィードバック」は、学生に対して復習しやすい環境ときっかけを与え、講義で取り扱った題材について習熟度を高めることが期待出来る。しかし、1つの講義について、教員と学生の人数比は教員1人に対して学生が40人から100人という形式が多く、この場合、全ての学生に対して適切な評価とフィードバックを与えることは、時間と労力のコストが高いといった問題がある [1]。

本稿では、この問題を解決するため、「評価」や「フィードバック」の対象として大学の講義などで行われている、ある課題について問題点を考え解決策を考案するといった「考察型レポート」に着目する。自然言語処理技術の1つである文書ベクトル化を用いて「考察型レポート」をベクトル化し、差ベクトルやベクトル間の類似度の算出といったベクトル演算を用いることによって、評価とフィードバックの自動化を行い、学生に対して迅速かつ適切に評価とフィードバックを与えるシステムを提案する。また、評価とフィードバックの自動化に対して有効な文書ベクトル演算の諸検討を行う。

2 関連研究

本章では、本稿の関連研究として、1章で述べた「評価」と「フィードバック」について述べられている研究を挙げる。

2.1 「評価」

「評価」について、小林は国語の入試問題について、形態素解析された解答文を Doc2Vec 及び BERT を用いてベクトル化し、予め用意された模範解答のテキストとの類似度を算出し、人間が与えた評価点と、類似度との相関を求めている [2]。また、鈴木らは記述式の問題について Doc2Vec を用いてベクトル化及び模範解答との類似度を算出することで、人間が与えた評価点と類似度の相関及び有意差検定を行っている [3]。これらの研究では、Doc2Vec は多義語や言い換え表現の判定が困難であり、多義語や言い換え表現を含む解答文に対して自動評価を行うことは難しいと述べられている。

また、安田らによって発明された特許技術では、参考文献 [4] の代表図面で示されるように、特定のベクトル化手法に依らないシステムが考案されている。ある問題に対する解答テキストについて、それぞれのベクトル化手法によって模範解答との類似度を算出し、採点済みデータから与えられた評価点を目的変数に、類似度を説明変数とした重回帰分析によって求められた重回帰式に代入し評価を行うことで、それぞれのベクトル化手法の弱みを軽減することが期待出来る。

2.2 「フィードバック」

「フィードバック」について、atama+と呼ばれるサービスでは、図1のように、ある単元を学習する学生に対して練習問題を与え、学生の解答が正解であった場合、反復練習を行うための類題や、次に学ぶべき項目に関わる問題を学生に対して与えている。また、学生の解答が不正解であった場合、与えた問

表 1 従来手法に対する提案手法との比較

	参考文献 [4]	atama+ [5]	提案手法
採点対象	様々	答えが明確な問題	考察型レポート
採点手法	複数のベクトル化手法	パターンマッチング	複数のベクトル化手法
用意するデータ	模範解答	模範解答	模範解答または過去の採点済みデータ
フィードバック	推定評価点	評価点, 解答, 解説	推定評価点, ヒント

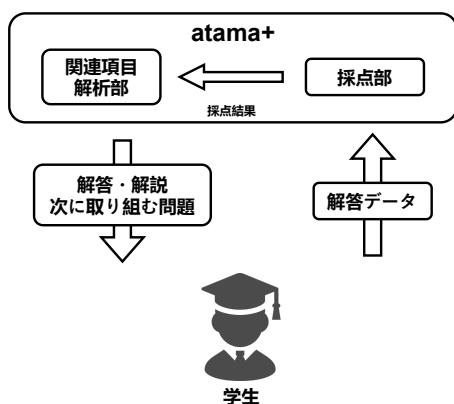


図 1 関連研究「atama+」のシステム概観図

題に関連する既習事項を解析し、学生が復習すべき既習事項を復習問題として与えている [5]。これを反復することによって、学習におけるつまづきを解消し、学習単元に至る基礎知識を十分に理解することが出来、現在学習している単元や、次に学ぶ単元に対してより効率良く学ぶことが期待出来るサービスとなっている。

このサービスは、小学校から高校までの学習範囲を取り扱う問題としており、各教科で扱われる単元項目について、関連性を内部のシステムが解析しており、学生が取り組んでいる問題とその解答によって、次に取り組むべき演習問題を推薦している。しかし、残念ながら内部処理に関わる詳しい解析技術や推薦技術は不明である。

3 提案手法

本章では、2章で述べた関連研究との類似点や相違点を挙げた上で、学生の考察型レポートに対して、自動評価とフィードバック生成を行うシステムについて提案する。

3.1 従来手法との類似点と相違点

関連研究と、本稿の提案手法の類似点及び相違点を以下で述べ、その結果を表 1 に示す。まず、「評価」について、参考文献 [4] のシステムでは、提出された解答テキストに対してそれぞれのベクトル化手法について模範解答との類似度を計算し、その値を予め算出された重回帰式へ代入することで評価点の推定を行っている。後述する 3.5 節によって、様々なベクトル化手法を用いて、類似度を算出することや、重回帰分析を行うことで総合的な観点から判断を下すといった点で、本手法と類似している。しかし、このシステムは必ず模範解答を用意する必要があり、レポート課題のような、学生によって様々な書き方で書かれ、取り上げられるトピック数が多くなるほど、数多く

の模範解答を作らなければならないと考えられる。本稿の提案手法では、模範解答または過去の採点済みレポートのどちらかが存在すれば良く、採点済みの解答データが多くなるほど多くのトピックに対して対応出来るようになるため、数多くの模範解答を作成する労力を減らすことが期待出来る。

次に、「フィードバック」について、atama+では、提出された解答データに対してフィードバックを与えることは本手法と類似しているが、atama+では小学校から高等学校までの学習事項と対応している。

一方、本稿で対象としているものは大学で課されるような「考察型レポート」であり、解答が一意に定まる設問に加え、複数解が存在する場合や、決まり切った答えが存在しない設問についても採点対象としており、様々な問題と解答に対応が出来ると期待出来る。

以上から、提案手法は複数のベクトル化手法を用い、模範解答だけでなく過去の採点済みデータを利用することで、答えが1つとは限らない様々な問題にも対応し、教員の負担を減らしつつ、適切な評価とフィードバックが生成されると期待出来る。

3.2 提案手法の概観

本稿の提案システムについて、概観を図 2 に示す。初めに、学生が提出した未採点のレポートを入力としてシステムに与える。次に、システムは与えられたレポートの文書に対して形態素解析を行い単語に分割し、その結果を基に複数のベクトル化手法を用いてベクトル化を行う。このとき、ベクトル化手法によって事前学習が必要な場合が存在する場合も存在するが、解答データベースや語彙データベースから分かち書きされたテキストを入力とし、学習する。さらに、模範解答または過去に同様の問題に取り組み、採点済みとなった学生レポートに対しても、同様にベクトル化手法を適用し任意の次元で文書ベクトル化を行う。以上の手順により、学生が提出した未採点のレポートと、保存されたデータから生成されたベクトルに基づいて、評価点の推定、及びフィードバックの生成を行う。また、参考文献 [4] と同様に、複数のベクトル化手法を用い、それらを総合的に扱うことで、それぞれのベクトル化手法の強みから、多角的で適切な自動評価とフィードバックの生成を行うことが期待出来る。

ここでのフィードバックは1章で述べたように、文章の添削指導や学生が見落としている点を気づかせるヒントといった情報が考えられる。システムが学生に対してフィードバックを与える場合の理想の1つとして、現在の教育現場で行われているように、関連項目を認識させつつ、学生が直すべき箇所やそれに関連するヒントなどを自然言語で出力することが考えられる。

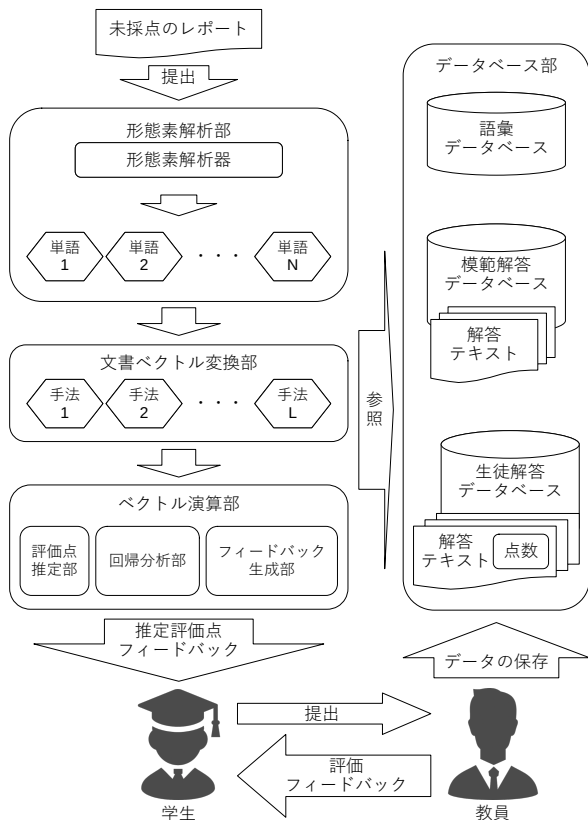


図 2 提案手法の概観

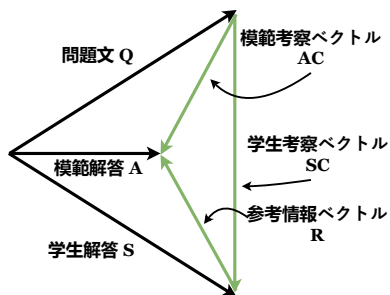


図 3 文書ベクトルの例

本稿では、以上のフィードバックの1つとして、学生によって提出されたレポートに対して、改善するために調べるべき重要単語を挙げ、ベクトル化手法とその演算によって求める手法について、後述する3.6節で述べる。

3.3 文書ベクトルの定義と演算

本研究で用いるベクトル化手法について、問題文、模範解答、学生のレポートの3つに対して得られた特徴ベクトルをそれぞれ Q , A , S とし、図3に示す。

また、この3つのベクトルに対して差ベクトルを求めることで得られるベクトルは、問題文から模範解答へ至る考察、問題文から学生のレポートへ至る考察、学生の解答が模範解答へ近づく参考情報と仮定する。本稿では、この参考情報ベクトルが学生のレポートにある考察のオリジナリティを満たしつつ、学習目標を満たすことが出来ると考え、これを R とおき、フィードバックの生成に用いることとする。

3.4 使用するベクトル化手法について

本稿では評価点の推定とフィードバックの生成について、Bag of Words (以下 BoW と表記する), Term Frequency (以下 TF と表記する), TF-IDF, Word2Vec, Doc2Vec といった5種類のベクトル化手法を使用する。なお、TF-IDF, Word2Vec, Doc2Vec の3つの手法は Python のオープンソースライブラリである gensim を用いて実装した [6-8]。

BoW は、単語ごとに分割された文書に対して、使用された単語の値に1を与え、使用していない単語の値に0を与えるベクトル化手法である。2文書間の類似度を算出するとき、BoW の性質から同じ単語をどれだけ使用しているか判別することが出来る。しかし、BoW は2つの値しか持たないため、ベクトルとしての表現能力に乏しいといった問題がある。これにより、文書中の主題となる重要語が抽出しづらいと考えられる。

TF は、単語ごとに分割された文書に対して、使用された単語について、文書中での出現頻度を値としたベクトル化手法である。BoW の問題に対し、単語の出現頻度を用いて表現能力を拡張することで、BoW より高精度に類似度が算出され、文書中の主題となる重要語が抽出されると期待出来る。しかし、助詞や記号など、多くの文章で高頻度に出現する単語も高い値を示すため、重要語を抽出する場合にノイズになると考えられる。

TF-IDF は、TF と同様に、使用されている単語について文書中での出現頻度を考慮したベクトル化手法である。TF の問題に対し、自身の文書に使用されている単語が他の文書でどれだけ使用されているのかといった、文書間での単語の出現頻度である IDF (Inverse Document Frequency) を用いている。これにより、多くの文書に用いられ、ノイズとなりやすい共通単語の値を下げることで、文書中の主題を捉えた類似度の算出と重要語の抽出が期待出来る。単語 i を含む文書 j における $t_{i,j}$ の値 $weight_{i,j}$ について、様々な定義が提案されているが、本稿では式(1)で定義する。なお、 $frequency_{i,j}$ は単語 i が文書 j 中に出現した頻度を表し、 $document_freq_i$ は単語 i が全文書 D 中に何件出現しているのかといった、文書間での単語の出現頻度を表している。本稿での全文書集合とは、図2で示される学生解答データベース、模範解答データベース、語彙データベースに含まれる文書数の合計である。

$$weight_{i,j} = frequency_{i,j} \cdot \log_2 \frac{|D|}{document_freq_i} \quad (1)$$

Word2Vec は、Mikolov らによって開発されたベクトル化手法で、単語に分割された文書群をテキストコーパスとして学習し、単語を任意の次元のベクトルへ埋め込むベクトル化手法である [9,10]。Word2Vec の特徴として、単語間の類似度を算出することが出来、単語ベクトルによるベクトル演算として加算や減算、類似度計算などが出来ることが知られている。

一般に、Word2Vec は単語間でのベクトル演算を目的に使用されており、単語の列である文書でのベクトル演算では用いられていない。本稿では、Word2Vec によって文書ベクトルを生成する場合、文書 j における次元 i の値 $weight_{i,j}$ について、式(2)を用いて生成した。なお、 $w_{i,j,k}$ は、文書 j を単語ごとに分

割したときの先頭から k 番目の単語から得られるベクトルの i 次元目の値を表している。

$$\text{weight}_{i,j} = \sum_k w_{i,j,k} \quad (2)$$

Doc2Vec は、Le らによって Word2Vec を基に開発されたベクトル化手法で、タグ付けされた文書を単語ごとに分割し、タグと文書を対にした文書群をテキストコーパスとして学習し、文書を任意の次元のベクトルへ埋め込むベクトル化手法である [11]。Word2Vec は単語間の類似度を算出することが出来るが、一方、Doc2Vec は文書間の類似度を算出することが出来る、文書ベクトルによるベクトル演算が出来ることが知られている。

本稿では Word2Vec と Doc2Vec について、学習手法では Word2Vec に Skip-gram 法、Doc2Vec に PV-DM 法を使用した。また、gensim のモデルに与える共通のパラメータとして、ベクトルの次元数 M (vector_size) を 200, epochs を 20, window, negative, min_count をそれぞれ 10, workers を 8, hs を 0 とした [7, 8]。

3.5 評価点の推定

評価点の推定を行うとき、まず、提出されたレポートに対してベクトル化手法を適用する。得られたベクトルに対して、模範解答や過去のデータを参照し、類似度が高いと算出されたレポートは、実際的评价点も類似しており、差の絶対値が小さく、負の相関があると考えられる。以上の仮説から、評価点の推定を行う手法を提案する。

提出された未採点のレポートをベクトル化した文書ベクトル \mathbf{S} に対して、模範解答または過去に評価された採点済みレポートから類似度の高い文書のうち上位 n 件を抽出する。取り出された j 件目の文書 D_j について、実際に与えられた評価点と、提出された未採点のレポート \mathbf{S} とのコサイン類似度をそれぞれ $\text{point}(D_j)$, $\text{sim}(\mathbf{S}, \mathbf{D}_j)$ とする。以上の評価点と類似度を用いて、提出されたレポート \mathbf{S} の評価点を推定する関数を $\text{score}(\mathbf{S})$ とし、以下の式 (3) で表す。

$$\text{score}(\mathbf{S}) = \frac{\sum_{j=1}^n \text{point}(D_j) \cdot \text{sim}(\mathbf{S}, \mathbf{D}_j)}{\sum_{j=1}^n \text{sim}(\mathbf{S}, \mathbf{D}_j)} \quad (3)$$

このとき、それぞれのベクトル化手法において、類似度が高いとはどの程度を指すのか、抽出する件数が何件である方がそのベクトル化手法による点数推定として精度が良くなるのかといった最適化の問題が考えられる。このような調整をシステム上で行う場合、取得する件数 n を調整することや、類似度が設定された閾値以上であれば推定するためのデータとして取り入れるといった閾値 t などを調整することで可能であると考えられる。

また、それぞれのベクトル化手法について式 (3) を適用することで、1つのレポートに対して複数の手法による推定評価点を得ることが出来る。このことから、模範解答や過去に提出された採点済みレポートに対して評価点の推定を行い、実際に与えられた評価点を目的変数に、推定評価点を説明変数と置くことで、重回帰分析が可能となる。これによって、それぞれの手

法による推定評価点に対して、偏回帰係数との積によって線形結合することが出来、3.2 節で述べたような、それぞれのベクトル化手法の強みを利用した評価点の推定が期待出来る。

この場合の評価点は、使用する L 個のベクトル化手法のうち ℓ 番目の手法によって推定された推定評価点 $\text{score}_\ell(\mathbf{S})$ と、求められた偏回帰係数 $\hat{\alpha}_\ell$ の積の総和に切片 c を足した値で表すことが出来る。これを以下の式 (4) で表す。

$$\text{scoreEx}(\mathbf{S}) = c + \sum_{\ell=1}^L \hat{\alpha}_\ell \cdot \text{score}_\ell(\mathbf{S}) \quad (4)$$

3.6 フィードバックの生成

TF-IDF は重要語抽出や、トピック分析などに用いられている手法であるため、この手法は重要語抽出として適していると考えられる。また、Doc2Vec は文書や単語 (タグ) を入力に与えることで、類似する文書を出力することが出来ることから、トピック分析などに用いることが出来ると考えられる。少なくとも、TF-IDF と Doc2Vec の2つの手法について、提出されたレポートに対して模範解答との差ベクトルを求めることで、学生のレポートに対して足りない点として単語を出力出来るのではないかと考えた。以上の仮説から、学生の提出したレポートに対して足りない点を出力するような手法を提案する。

図 3 から、学生のレポートベクトル \mathbf{S} と模範解答ベクトル \mathbf{A} との差ベクトル \mathbf{R} を利用して、重要度の高い単語を上位 m 件抽出し学生に与える。

差ベクトル \mathbf{R} からフィードバックとして単語を出力するとき、任意の単語 w について与えられる重要度を $\text{importance}(w, \mathbf{R})$ とし、式 (5) 及び式 (6) で表す。ただし、式 (5) は BoW, TF, TF-IDF での重要語抽出に用い、式 (6) は Word2Vec, Doc2Vec で重要語抽出に用いる。このとき、 $\mathbf{R}(w)$ はベクトル \mathbf{R} における単語 w に対応する値である、また、 $\text{sim}(\mathbf{W}(w), \mathbf{R})$ は単語 w に対応する埋め込みベクトル $\mathbf{W}(w)$ と、ベクトル \mathbf{R} との類似度とする。また、各ベクトル化手法から式 (5) 及び式 (6) によって求められた重要語について、重要度の相加重平均をレポート全体の重要語として抽出し、学生に与える。

$$\text{importance}(w, \mathbf{R}) = \begin{cases} \text{sim}(\mathbf{S}, \mathbf{A}) \cdot \mathbf{R}(w) & (5) \\ \text{sim}(\mathbf{S}, \mathbf{A}) \cdot \text{sim}(\mathbf{W}(w), \mathbf{R}) & (6) \end{cases}$$

さらに、3.5 節で求めた重回帰分析の結果を用いるといった手法や、総和が 1 となる係数列による最適化、ユーザが調整した値による代入などによって、ベクトル化手法ごとに求めた単語の重要度を組み合わせることも考えられる。そのような係数を適用する場合、式 (4) と同様に、使用する L 個のベクトル化手法のうち ℓ 番目の手法によって抽出された、単語 w の重要度を $\text{importance}_\ell(w, \mathbf{R})$ と、それぞれのベクトル化手法に設定された標準偏回帰係数 α_ℓ を掛けた値の総和によって表すことが出来る。これを式 (7) で表す。

$$\text{importanceEx}(w, \mathbf{R}) = \sum_{\ell=1}^L \alpha_\ell \cdot \text{importance}_\ell(w, \mathbf{R}) \quad (7)$$

表 2 考察毎の解答例に対する評価点の推定例

問題 ID	考察 2.3.1	考察 2.3.2	考察 2.3.3
Truth	70.000	63.000	50.000
BoW	57.993	36.500	0.000
TF	65.628	41.350	90.000
TF-IDF	62.094	29.510	0.000
Word2Vec	63.600	41.700	70.005
Doc2Vec	66.671	0.000	0.000
scoreEx	62.920	44.547	65.311

4 評価実験

本章では、本稿の提案手法である評価点の推定及びフィードバック生成の有効性を検証するため、3つの実験を行う。

4.1 データセット

データセットとして、室蘭工業大学で開講されている情報学応用演習 A という講義の中で、学生が提出したレポート及びその模範解答を使用する [12]。本稿では、レポート内にある設問から、3つの設問それぞれに対して、問題文 1 件、模範解答 1 件、学生の解答文 99 件、著者が作成した採点済みの解答文 2 件の合計 309 件を用いた。表 3 に文章例を示す。

また、学習などの事前処理が必要なベクトル化手法について、語彙データ及び文書データとなるテキストコーパスとして、Cirrus Search が提供している日本語版 Wikipedia のダンプデータを使用した。このとき、前処理として形態素解析に MeCab を使用し、辞書は mecab-ipadic-NEologd を使用した。

4.2 評価点推定手法の精度評価

評価点の推定手法による精度を評価するため、式 (4) によって推定される評価点と、教員によって実際に与えられた評価点との差の平均を求めた。また、提出された未採点のレポートに対して評価点を推定したとき、類似度の閾値 t と、取得する文書数の上限となる件数 n の 2 つの変数がどのように影響するのか調べた。その結果を図 4 に示す。また、著者が作成した考察毎の解答例に対する評価点の推定結果を表 2 に示す。

図 4 から、取得件数 n に関わらず、閾値 t が高くなることで、推定評価点と実際に与えられた点数との差の平均が大きくなるのが分かる。これは、閾値 t が高くなることで式 (3) に取り入れられるレポートの数が減り、推定評価点との誤差が広がったと考えられる。しかし、図 4 及び表 2 から、式 (4) による総合評価で、取得件数 n や閾値 t に関わらず、実際に与えられる点数に近い推定評価点を求めることが可能であると分かった。

4.3 フィードバック生成の有効性評価

式 (5) 及び式 (6) を用いて、表 3 中の著者が作成した解答例を入力として、得られたフィードバックを表 4 に示す。

表 4 から、BoW, TF, TF-IDF による重要語抽出では、問題文と模範解答に関連が見られる単語を出力することが出来た。例えば、考察 2.3.1 ではランダム性がキーワードであり、ランダム性が問題なのだ気づかせるような単語が出力されていると考えられる。また、考察 2.3.2 の結果では、TF-IDF では

wing という単語が出力されており、X-wing 法と呼ばれる数独を解くための考え方に繋がる可能性が考えられる。この場合、形態素解析による前処理の段階で複合語の処理を高度化させ「X-wing 法」と単一の名詞として出力させることで、より具体的なヒントとして与えることが出来るのではないかと考えられる。さらに、考察 2.3.3 では、模範解答が一意に決まるような問題ではなかったことから、数独に関わる全般的な話題に関わるような単語が多く出力されていた。本稿では模範解答を使用した、過去データから対象の問題について、高得点者のグループを抽出して差ベクトルを取るといった手法によって、改善が出来るのではないかと考えられる。

これに対して、本稿で提案した Word2Vec 及び Doc2Vec を用いたフィードバック生成は期待する出力を得ることが出来なかった。これは、本稿で提案した式 (2) による文書ベクトル化手法について再検討すべきであると考えられる。

4.4 重回帰分析及び相関分析

3.5 節で述べた仮説が正しいのか検証をするため、図 5 から 2 文書間の評価点の差と類似度との関係を考察する。また、式 (4) 及び式 (7) について、各ベクトル化手法が取得件数 n と類似度の閾値 t といったパラメータに依って、各手法による評価点の推定がどのようになるのか検討するため、図 6 から標準偏重回帰係数にどのような影響を及ぼしているのか考察する。

図 5 から、それぞれの文書ベクトル化手法で点数差に関わらず類似度が正で求められやすいことが分かる。これは BoW, TF, TF-IDF の 3 つの手法では、定義から 0 以上の類似度が算出されており、Word2Vec は、式 (2) によって作られたベクトルは 0.9 以上の類似度が出やすく、出現する単語のベクトルの和では類似性と差の平均での相関が取れないことが分かる。また、Doc2Vec は同じ問題に対する解答テキストのみを比較しているため、負の値が出にくい結果となったと考えられる。

5 まとめと今後の課題

本稿では、様々なベクトル化手法を用いて、学生が提出する考察型レポートを文書ベクトルへ変換し、文書ベクトル演算を用いることで評価点の推定やフィードバックの生成を行う手法及びシステムを提案した。

4.2 節の結果から、重回帰分析を行うことで単一のベクトル化手法を用いた点数の推定に比べ、精度が高い結果となった。さらに、4.3 節の結果から、式 (7) によって期待したフィードバックを得ることが出来た。しかし、ベクトル化手法に基づいた生成過程によって、誤ったフィードバックを与えてしまう恐れがあると分かった。また、4.4 節の結果から、文書ベクトルが類似しているならば与えられた評価点が類似しているという仮説に誤りがあると考えられる。そのため、本稿で用いたベクトル化手法及び評価点の推定、フィードバックの生成手法について、再検討が必要であると考えられる。

今後の課題として、先述の課題に加え、提案システムを改良する要素として 4 点ほど考えられる。

1つ目は「剽窃の検出機能」である。何故ならば、学生自身の力で課題に取り組ませるために必要であると考えられるからである。

2つ目は「評価基準をシステムに反映する機能」である。本稿のシステムでは、教員による評価データを用いているが、それ以外にシステムに干渉出来る余地がないという問題がある。システムに対して自由度を持たせることで、教員の価値基準や社会の要請に柔軟に対応出来るシステムになると考えられる。

3つ目は「適切な言葉を使用しているか判別する機能」である。本稿で使用したベクトル化手法では、文脈依存でないベクトル化手法を用いており、これによる悪影響を防止するために必要であると考えられる。また、適切な言葉を用いて文書を書いているか考え指摘することは、実際の教育現場で行われていることであり、必要な要素であると考えられる。

4つ目は「フィードバックの対象を判別する機能」である。本稿のシステムでは、入力されたテキストに対して評価点と重要語を与える設計となっている。しかし、実際に与えられる評価点の高い学生に対してフィードバックを多く与える必要はない可能性がある。また、ただ提出することで、出力結果を基にレポートを作成することを防ぎ、真にフィードバックを必要としている学生に向けて与えるために必要であると考えられる。

文 献

- [1] Shun Hattori, Hiroyuki Kameda, Knowledge-based Compiler with e-TA for Software Engineering Education, Proceedings of the 9th Joint Conference on Knowledge-Based Software Engineering (JCKBSE'10), pp.265-278 (2010).
- [2] 小林 洋生, 自然言語処理技術を用いた答案採点支援システムの

- 検討, https://www.ams.giti.waseda.ac.jp/data/pdf-files/2020_kobayashi_bt.pdf (2021-12-16 参照).
- [3] 鈴木 千尋, 佐藤 直行, Doc2Vec を用いた国語記述式答案の自動評価, 情報処理学会第 81 回全国大会講演論文集, pp.135-136 (2019).
- [4] J-PlatPat 特開 2017-188039 複数の類似度算出によってテキストの点数を推定するプログラム, 装置及び方法, <https://www.j-platpat.inpit.go.jp/c1800/PU/JP-2017-188039/OF7797C29FBA9391826891398632F04813D60EB6D5855F34EAE770613829685/11/ja> (2021-12-16 参照).
- [5] atama+. <https://corp.atama.plus/> (2021-12-16 参照).
- [6] models.tfidfmodel - TF-IDF model - gensim - Radim Rehurek. <https://radimrehurek.com/gensim/models/tfidfmodel.html> (2021-12-16 参照).
- [7] models.word2vec - Gensim - Radim Rehurek, <https://radimrehurek.com/gensim/models/word2vec.html#gensim.models.word2vec.Word2Vec> (2022-01-09 参照).
- [8] models.doc2vec - Doc2vec paragraph embeddings - gensim, <https://radimrehurek.com/gensim/models/doc2vec.html#gensim.models.doc2vec.Doc2Vec> (2022-01-09 参照).
- [9] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean, Distributed Representations of Words and Phrases and their Compositionality, 2013. <https://arxiv.org/pdf/1310.4546.pdf> (2022-01-09 参照).
- [10] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, Efficient Estimation of Word Representations in Vector Space, 2013. <https://arxiv.org/pdf/1301.3781.pdf> (2022-01-09 参照).
- [11] Quoc Le, Tomas Mikolov, Distributed Representations of Sentences and Documents, <https://arxiv.org/pdf/1405.4053v2.pdf> (2022-01-09 参照).
- [12] 室蘭工業大学, 情報学応用演習 A - シラバス参照, <https://u.muroran-it.ac.jp/kyomu/syllabustop2021/2603/J4135.html> (2022-01-09 参照).

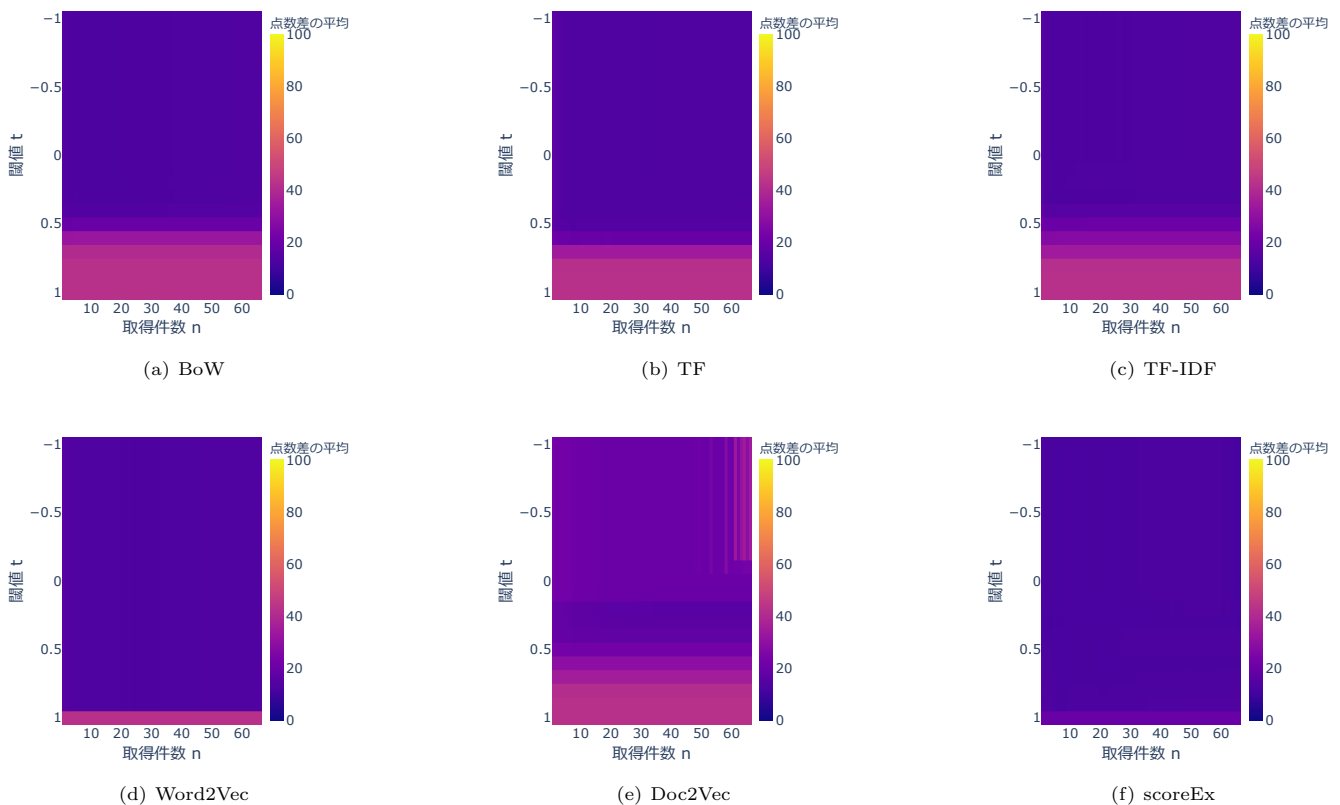


図 4 各手法における推定評価点の誤差 (問題の区別なし)

表 3 データセットの文章例

問題 ID	文書 ID	入力文
考察 2.3.1	問題文	(考察) 同じ難易度「Easy」の同じ問題に対して、同じ基本手法を適用しているにも関わらず、「solve」実行の度にアルゴリズム処理時間や空白セルアクセス回数が異なること⇒何故ならば、(理由の分析を各自の言葉で！)
	模範解答	何故ならば、基本手法のアルゴリズムの Step(2) において、未だ空白である注目セルへの入力候補として、generateRandomNumbers() 関数を用いてランダムな順の数字を生成して入力している為である。運良く最良の場合、巻き戻しは発生せず、空白セルの数だけ入力を試せば解に辿り着くが、最悪の場合、全ての空白セルで各々の真の解を入力する順番が最後になり、膨大な巻き戻しが発生し、空白セルアクセス回数やアルゴリズム処理時間が掛かる。このように、アルゴリズムに「ランダム性」を含む為、「solve」実行の度にアルゴリズム処理時間や空白セルアクセス回数が大きく異なってしまう。特に、sudoku の問題の難易度が上がる程、この傾向は一般的に強くなると考えられる。
	解答例	何故ならば、空白セルに入力する候補となる数字の順番をランダムに生成している為である。
考察 2.3.2	問題文	(問題点) 問題の難易度が難しくなる程、巻き戻し (バックトラック) が多く発生し、空白セルに一度入力した数字のリセットも頻発してしまい、効率が悪く、アルゴリズム処理時間も膨大になっていく⇒この問題を解決する為には、(解決アプローチや方法を各自の言葉で！)
	模範解答	この問題を解決する為には、巻き戻し (バックトラック) を減らす、出来れば無くす必要がある。具体的には、単純に左上の (0,0) から右下へと入力を試みるのではなく、注目セルと同じ行、同じ列、同じ 3x3 ボックス内を探索するだけでも解が唯一に確定する空白セルや、入力し得る候補のより少ない空白セルから優先的に入力を試みたり、空白セル毎に入力を試したが巻き戻されてしまった数字を記憶しておき、それと同じ数字の入力を試みないように制限したり (但し、解が唯一の空白セルは先に埋めた後、左上から右下へ、あるいは、入力し得る候補のより少ない順など、入力を試す空白セルの順序は固定にする必要がある)、さらには、人間が sudoku を解く際の X-wing 法 [1] などを模倣して論理的に絞ったりする方法などが考えられる。参考文献 [1] 数独 数学の面白いこと・役に立つことをまとめたサイト、 https://analytics-notty.tech/category/sudoku/
	解答例	この問題を解決する為には、注目する空白セルに対して、解の候補を限定し、一意に定まるセルを優先的に処理する手法によって、巻き戻しやランダム性を減らす事が考えられる。
考察 2.3.3	問題文	(以上の他にも、各自で考察、問題分析する！)
	模範解答	(模範解答を唯一には書き難い) 基本手法の長所は、総当たり法である為、どんな難易度の問題に対しても解が求まる事である。また、アルゴリズムがランダム性を有しているが、運が良いと最短の空白セルアクセス回数で解に辿り着く可能性もある。一方で、短所は、一般的に難易度が上がる程、巻き戻し (バックトラック) が大量に発生し、効率が悪く、計算時間が膨大になってしまい、そういう意味で解を求める事が非現実的になってしまう場合もある。
	解答例	Step(3) の setCellValue 関数による空白セルの入力及び、Step(5) での入力済みセルへの初期化の時 GUI 処理が行われている。これによって巻き戻しの回数が多くなるにつれ、処理時間が長くなる事が考えられる。この問題を解決する為に、問題を解き終わった時のみ GUI 処理を行う事が考えられる。

表 4 考察毎の解答例に対するフィードバック例

問題 ID	手法名	フィードバック (上位 10 件の重要語)
考察 2.3.1	BoW	(, (,), 2, Step, generateRandomNumbers, solve, sudoku, アクセス, アルゴリズム
	TF	候補, 数字, 生成, 順番, は, ランダム, 為, (, (,)
	TF-IDF	アルゴリズム, sudoku, generateRandomNumbers, 解, solve, 回数, アクセス, 処理, (,) Step
	Word2Vec	レエブン, キメラクインケ,))》, ラズファード, overeating, VPCF, WDARL, マーガレット・コール, ザラリオ, ガンダベディア
	Doc2Vec	christo, ビジネスグラウンドワークス, 54965532, kawasak, 赤倉_(列車), yamanashikotsu, コン・ヨンスク, ヤーギン, パケーキ
	importanceEx	sudoku, generateRandomNumbers, アルゴリズム, solve, (,) Step, 解, 回数, アクセス, 処理
考察 2.3.2	BoW	ランダム, 一意, 事, 処理, 性, 手法, 限定, セル, 候補, 優先
	TF	入力, セル, 空白, /, 的, -, 0, 1, 3, [
	TF-IDF	入力, sudoku, 左上, notty, analytics, バックトラック, 数字, 試み, wing, tech
	Word2Vec	レエブン, キメラクインケ,))》, ラズファード, ザラリオ, overeating, WDARL, VPCF, スーパーキュウレンオー, マーガレット・コール
	Doc2Vec	三浦清美, コン・ヨンスク, パケーキ, 54965532, yamanashikotsu, アニソクイズ, 赤倉_(列車), christo, ビジネスグラウンドワークス
	importanceEx	入力, セル, 空白, /, sudoku, 的, 左上, 数字, 試み, -
考察 2.3.3	BoW	(,), 3, 5, GUI, Step, setCellValue, これ, 入力, 処理
	TF	解, 度, 性, 的, 難易, アクセス, アルゴリズム, バックトラック, ランダム, 一般
	TF-IDF	解, 難易, バックトラック, 短所, アルゴリズム, 長所, 模範, 解答, 最短, ランダム
	Word2Vec	キメラクインケ, レエブン, overeating,))》, WDARL, マーガレット・コール, 蕭珪, 》, 翅鏡, ザリユース
	Doc2Vec	アメリカ独立戦争の情報戦略, 坂齋小一郎, 貞山堀, JOD, コラブ, 甲斐智陽, 大友康匠, christo, Boiz, :en:The_Molly_Maguire_(film)
	importanceEx	解, 難易, 度, 性, 的, バックトラック, 短所, アルゴリズム, 長所, 模範

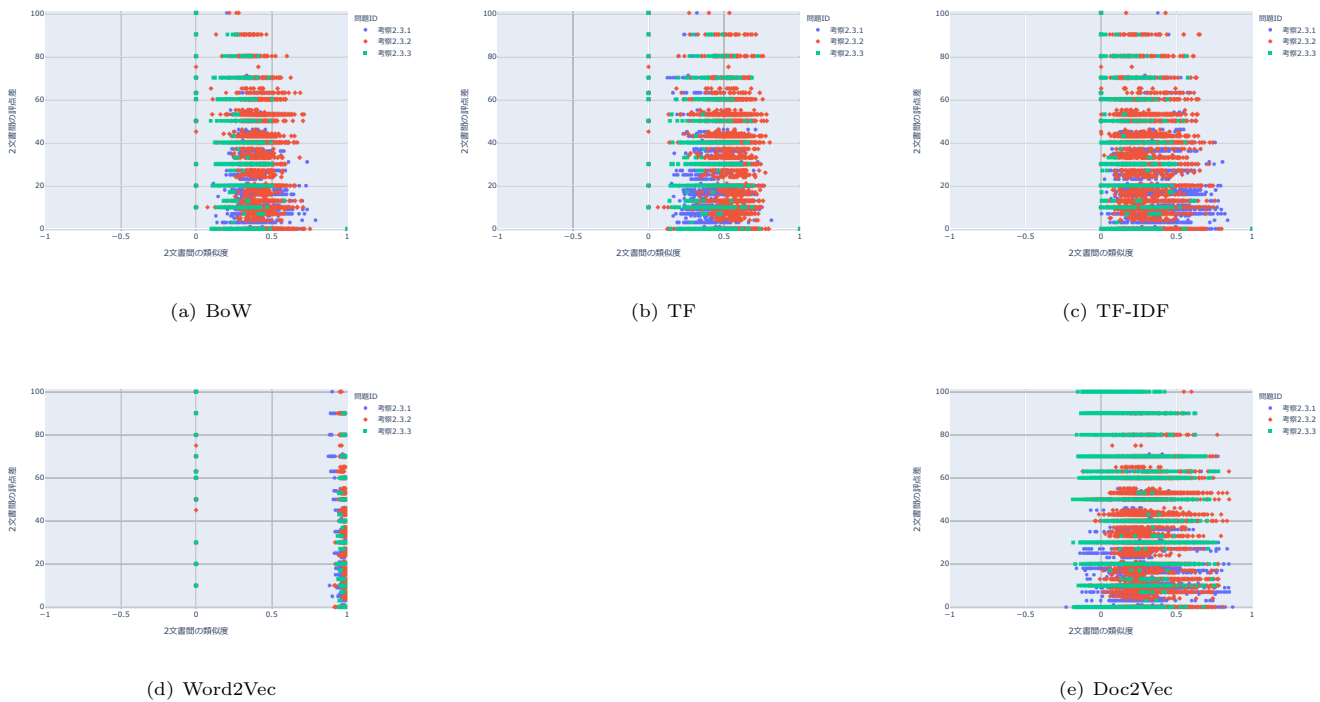


図5 2文書間の評価点の差と類似度の関係

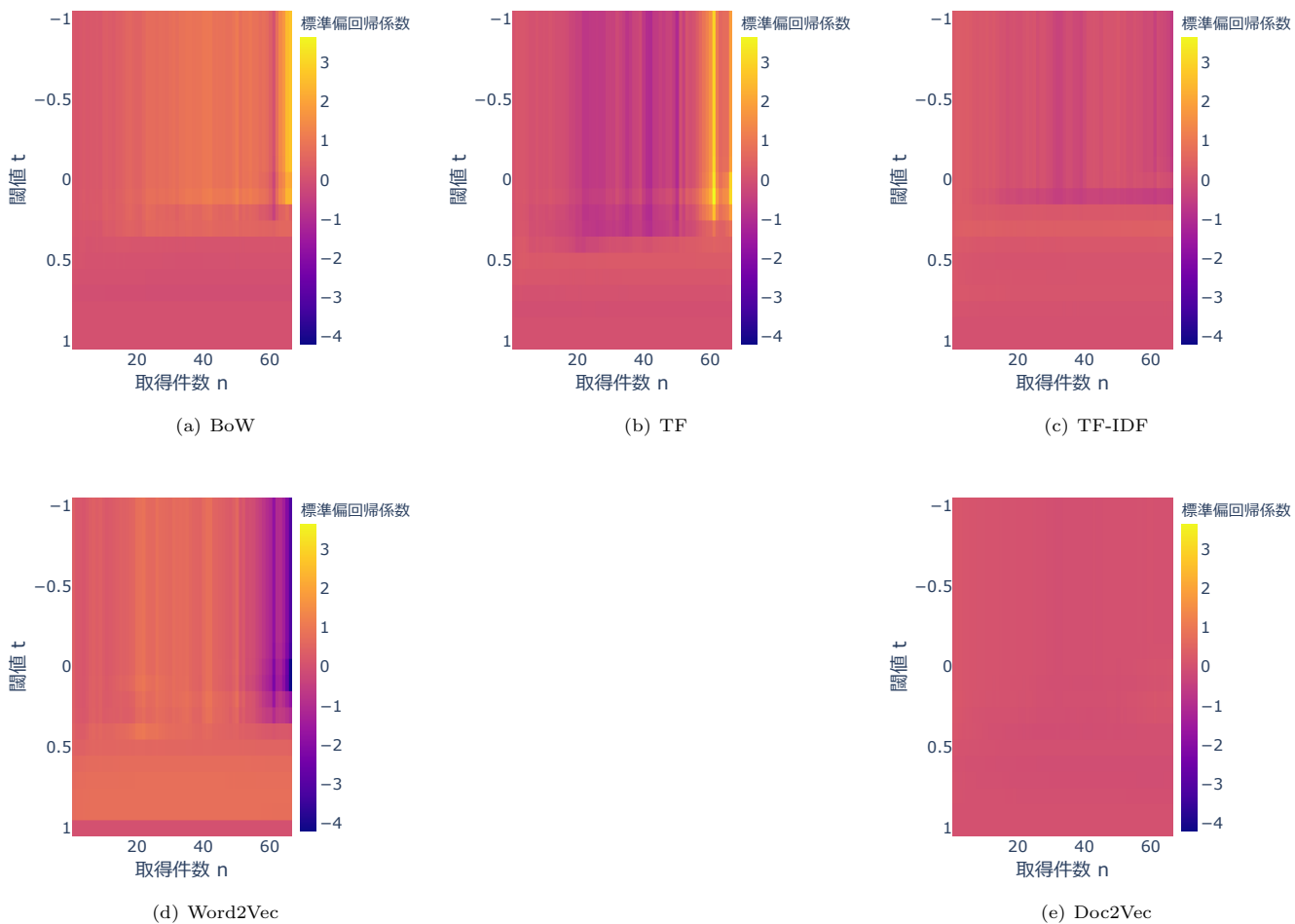


図6 重回帰分析における各手法の標準偏回帰係数（問題の区別なし）