卒 業 論 文 概 要 書

Summary of Bachelor's Thesis

Date of submission: 2/7/2020

学科名（専門分野） Department	情報通信	氏名 Name	小林洋生	指 導 教 員 Advisor	渡辺 裕 ㊞
研究指導名 Research guidance	オーディオビジュアル 情報処理研究	学籍番号 Student ID number	1W153052-7		
研究題目 Title	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 A Study on Marking Support System Using Natural Language Processing Technique				

1. はじめに

今日，大学入学試験において選択式の試験を廃止し，記述式問題を導入することが検討されている．しかし，以下の問題により予定されていた 2020 年度の導入は見送りとなった． 50 万枚に上る答案を，限られた時間の中で，公平に採点できるのかという問題である．

大学入学試験に限らず，記述式問題の答案の採点には多大な時間と労力を要し，採点の誤りや採点者による採点結果のばらつきが生じる．また，答案の返却が遅れることで受験者の復習効果も低下する．こうした問題の解決方法として，答案採点の自動化が挙げられる．

そこで本研究では，答案採点の自動化を図るため，形態素解析器の JUMAN++や，自然言語を分散表現化する Doc2Vec，Bidirectional Encoder Representations from Transformers (BERT)といった自然言語処理技術を用いた答案採点支援システムを検討する．

2. 関連研究と問題点

コンピュータによる自動採点の関連研究として，水本らによる“採点項目に基づく国語記述式答案の自動採点”[1]が挙げられる．水本らによる自動採点モデルは，答案中の単語を入力として，それらを Word2Vec によって分散表現 (ベクトル) に変換する．このベクトルを Long short-term memory (LSTM)に渡し, 項目点を予測し sigmoid 関数を用いて項目ごとに得点を出力する．

この自動採点モデルには「人工的に高得点答案を作成可能である」という問題点がある．Word2Vec を用いて単語のみを分散表現に変換し，答案を単語ごとに評価する．そのため，答案に重要単語が含まれていれば，単語の使い方が間違っている場合や，説明不足である場合，文章全体の意味が異なっている場合でも高得点を得られる．

3. 自然言語処理技術

3.1. Doc2Vec

Doc2Vec[2]は，Quoc らによって開発された文章を分散表現に変換するニューラルネットワークである．大量のテキストデータを解析することによって，自然言語で記述された文章を，意味を持った分散表現に変換できる．そのため，解析後のベクトル同士の類似度を測定し，文章分類や似た文章の検出ができる．

3.2. BERT

BERT[3]は，Jacob らによって開発された，汎用言語表現モデルである．自然言語で記述された単語を，文脈に基づいた観点から適切な意味を持った分散表現に変換できる．そのため，文脈の意味を理解し，文脈の情報を加味した単語の適切なベクトルを得られる．この各単語ベクトルは，文脈の情報を付加することで，文章内でのその単語の役割と文章全体の意味を含む．

4. 提案手法

4. 1 提案手法 1

Doc2Vec を用い，文章全体を分散表現に変換して答案の類似度を測定，評価する手法．

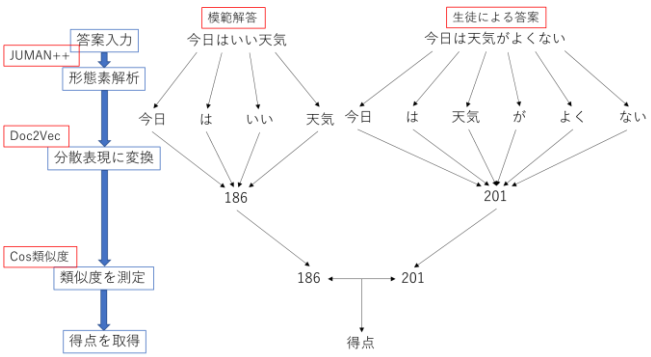

図 1 Doc2Vec を用いた答案採点支援システム

4.2　　提案手法 2

BERT を用い，答案中の単語を，文脈情報を加味した分散表現に変換して答案の類似度を測定，評価する手法.

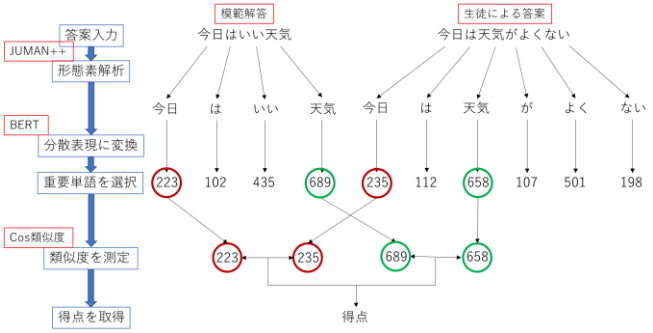

図 2　BERT を用いた答案採点支援システム

5. 提案手法による評価実験

5.1　　精度評価実験の概要

国語の表現自由型短答記述式問題（4 題）の答案を収集し（内容，人手による採点結果が多様なもの，$n = 10$），模範解答と答案の類似度を算出した．人手による採点結果は中学校の国語教師が模範解答をもとに得点を与えた．本研究では，提案手法より得られる答案の類似度と人手による採点から得られる得点の無相関検定を行い，各提案手法の精度を評価する.

5.2　　提案手法 1 の精度評価実験の結果と考察

提案手法 1 の精度評価実験の結果を図 3 と図 4 に示す．図 3 は問題 1，図 4 は問題 4 における答案の類似度と人手による採点結果の相関を，散布図と最小二乗法により求めた近似直線で表したものである.

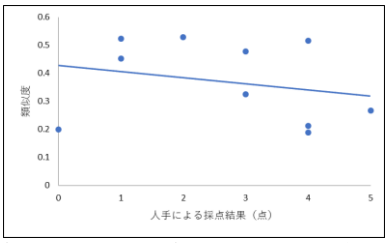

図 3　提案手法 1 の精度評価実験の結果（問題 1）

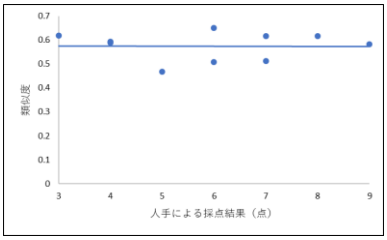

図 4　提案手法 1 の精度評価実験の結果（問題 4）

問題 1 と問題 4 において人手による採点結果と類似度の無相関検定を行った結果，有意な正の相関の傾向が得られなかった（問題 1: $r = -0.244, p > 0.05$, 問題 4: $r = -0.001,\ p > 0.05$）．提案手法 1 の評価実験において，自動採点に対し有効性を示さなかった．改善策として，十分な内容と量の学習データを収集することと複数の模範解答を用いることが挙げられる.

5.3　　提案手法 2 の精度評価実験と結果と考察

提案手法 2 の精度評価実験の結果を図 5 と図 6 に示す.

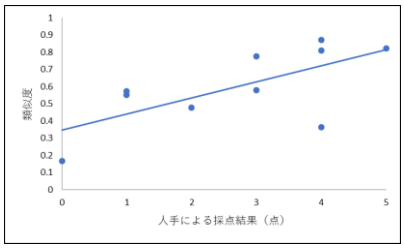

図 5　提案手法 1 の精度評価実験の結果（問題 1）

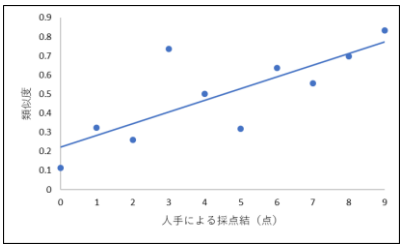

図 6　提案手法 1 の精度評価実験の結果（問題 3）

問題 1 と問題 3 において人手による採点結果と類似度の無相関検定を行った結果，有意な正の相関の傾向が得られた．（問題 1: $r = 0.68$, $p < 0.05$，問題 3: $r = 0.786,\ p < 0.05$）．提案手法 2 の評価実験において，自動採点に対し有効性を示した． より採点精度を上げる方法として，比較する単語数を増やすことや，BERT より得られる単語ベクトル列から文章ベクトルを得ることが挙げられる.

6. まとめ

Doc2Vec や BERT を用い，答案中の個々の単語だけでなく文章全体の意味を捉え評価する方法を提案した． 実験により，提案したシステムの有効性を確認した.

参考文献

[1]　水本智也，磯部順子，関根聡，乾健太郎：“採点項目に基づく国語記述式答案の自動採点”，言語処理学会第 24 回年次大会発表論文集，pp.552-555，2018.

[2]　Quoc V. Le, Tomas Mikolov: “Distributed Representations of Sentences and Documents”, ICML’14, Vol.32, pp.1188-1196, 2014.

[3]　Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, NAACL, pp.4171-4186, 2019.

2019 年度　卒業論文

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

A Study on Marking Support System Using Natural Language Processing Technique

提出日　2020 年 2 月 7 日

指導教員　渡辺 裕 教授

早稲田大学基幹理工学部 情報通信学科

1W153052-7

小林　洋生

目次

第1章　　序論

1.1 研究の背景

今日，文部科学省や大学入試センター等により，大学入学試験の形態が大きく変更されようとしている．選択式の試験を廃止し，記述式問題を導入することが検討されている．グローバル化や技術革新に伴う急速な社会変化が背景にある．

文部科学省は記述式問題の導入理由について以下のように述べている．記述式問題の導入により，思考力・判断力・表現力を評価できる．また，高等学校に対し「主体的・対話的で深い学び」に向けた授業改善を促し，大学においても，質の高い教育が期待される[1]．

しかし，予定されていた 2020 年度大学入学共通テストへの導入は見送りとなった．文部科学大臣は，受験生の不安を払拭し，安心して受験できる体制を早急に整えることが現時点では困難であるため導入を見送ると発表した[2]．記述式問題の導入に関しては，以前から様々な問題が懸念されていた．その中でも，50 万枚に上る答案を，限られた時間の中で，公平に採点できるのかが最大の問題点である．

大学入試センターは，民間事業者に答案の採点を委託することを予定していた．しかし，採点ミスを完全になくすことについて技術的に限界があり，自己採点の不一致を大幅に改善することは困難であると判断した．

大学入学試験に限らず，記述式問題の答案の採点には多大な時間と労力を要し，採点の誤りや採点者による採点結果のばらつきが生じる．また，答案の返却が遅れることで受験者の復習効果も低下する．こうした問題の解決方法として，答案採点の自動化が挙げられる．答案採点を自動化できれば，労力を費やすことなく短時間で，多くの答案を正確かつ公平に採点できる．大学入学試験においても，記述式問題導入のため活用できる．

コンピュータによって答案の正誤を判断することで，答案採点の自動化を図ることができる．そのため，コンピュータが，自然言語を理解し，処理できる必要がある．自然言語は，人間が日常的に使っている言語のことであり，自然言語処理（Natural Language Processing，NLP）は，自然言語をコンピュータに処理させる一連の技術である．自然言語処理技術を用いることで，コンピュータが答案を理解し正誤を判断できると考える．

そこで本研究では，答案採点の自動化を図るため，形態素解析器の JUMAN++や，自然言語を分散表現化する Doc2Vec, Bidirectional Encoder Representations from Transformers (BERT)といった自然言語処理技術を用いた答案採点支援システムを検討する．

1.2 研究目的，及び関連研究と問題点

1.1 節で述べたように，記述式問題の答案の手動採点は多大な時間と労力を要し，採点の誤りや採点者による採点結果のばらつきが生じる．また，答案の返却が遅れることで受験者の復習効果も低下すると考えられる．答案の採点を自動化できれば，採点の効率は上り，労力を費やすことなく短時間で，多くの答案を正確かつ公平に採点できる．また，採点結果を受験者に即時にフィードバックできる．

Copper（1984）によれば，小論文等の記述式問題の得点に関して誤差要因として働くものは六つある．書き手（writer），題目（topic），形式（mode），制限時間（time-limit），テスト状況（examination situation），採点者（marker）である[3]．石岡ら（3003）によれば，これらの大部分はいわゆる「試験」に共通している要因であるが，特に「採点者」の要因は記述式問題においては決定的なものであるという[4]．記述式試験の得点の誤差要因として，「採点者」はとても重大であり，採点者が変わると同じ回答でも得点が変わる．また，同じ採点者でも，その答案を何番目に採点したか，採点の順番・回数によって採点基準が変わり，得点に影響を与えることがある．

このような誤差要因を排除し，公平な採点を行うためにはコンピュータによる自動採点が必要不可欠であり，近年，自動採点の研究が精力的に行われている．

コンピュータによる自動採点の関連研究として，石岡ら（2003）による“コンピュータによる小論文の自動採点システム Jess の試作”[4] と，水本ら（2018）による“採点項目に基づく国語記述式答案の自動採点”[5] が挙げられる．

石岡らによる Jess は，アメリカで実施される適性試験のひとつである Graduate Management Admission Test (GMAT) において，実際に小論文の採点に用いられている e-rater を参考にしたものである．文章の形式的な側面，いわゆる文章作法を評価する「修辞」と，アイディアが理路整然と表現されていることを示す「論理構成」，トピックに関連した語彙が用いられているかを示す「内容」の三つの観点から小論文を評価し採点する[4]．

水本らによる自動採点モデル[5]は，答案中の単語を入力として，それらを Word2Vec によって分散表現（ベクトル）に変換する．このベクトルを Long short-term memory (LSTM)に渡し，項目点を予測し sigmoid 関数を用いて項目ごとに得点を出力する．

これらの関連研究において共通する問題点として，「人工的に高得点答案を作成可能である」ということが挙げられる．単語の使い方が間違っている場合や，説明不足である場合，文章全体の意味が異なっている場合でも，答案に重要単語が含まれ文法的に正しければ高得点を取得できる．

そこで，本研究では，自然言語処理技術 Doc2Vec や BERT を用い，答案中の個々の単語だけでなく文章全体の意味を捉え評価する，短時間で公平な採点を可能とする答案採点支援システムの作成を目的とする．

1.3 本論文の構成

以下に本論文の構成を示す.

第1章　本章であり，本研究の背景，目的，及び関連研究と問題点について述べる.

第2章　水本らによる関連研究について述べる.

第3章　関連技術について述べる. 本研究に用いる自然言語処理技術とコサイン類似度の概要，及び自動採点における利点と問題点について述べる.

第4章　自然言語処理技術を用いた答案採点支援システムの提案手法及び評価方法について述べる.

第5章　提案手法の実験結果及び考察について述べる.

第6章　本研究の結論と今後の課題について述べる.

第2章　　関連研究

2.1 まえがき

本章では，関連研究である水本らによる“採点項目に基づく国語記述式答案の自動採点”の手法について述べる．

2.2 採点項目を考慮した自動採点モデル

水本らによる自動採点モデルは，既に人間の手によって採点された答案を学習データとし，入力文章に対し採点項目を予測し項目ごとに評価する深層学習モデルである．答案中の単語を入力として分散表現に変換し，各項目で得点を出力する．最後に，各項目の得点を加算し，全体の得点を予測する．水本らによる自動採点モデルのイメージ図を図 2.1 に示す．

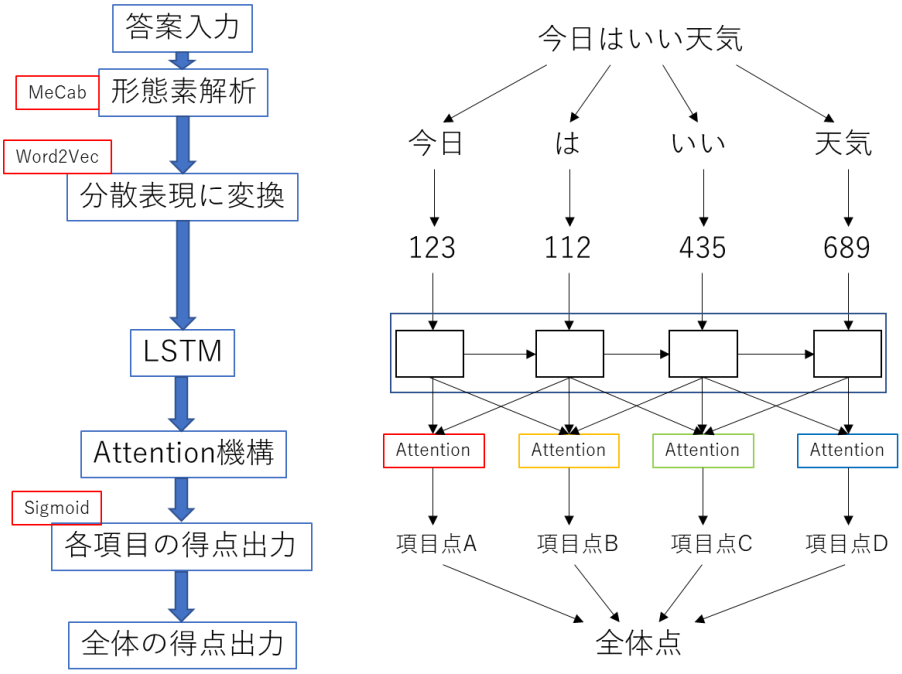

図 2.1　水本らによる，自動採点モデルのイメージ図

答案の文章を日本語形態素解析器 MeCab により文章を単語ごとに分割し，その単語を入力とする．各単語を Word2Vec により分散表現に変換し，このベクトルを LSTM に入力する．各項目で重要箇所に注目させるため，項目ごとに Attention 機構を用意し，LSTM の出力に対して適用する．その後，sigmoid 関数を用いて項目ごとに得点を出力する．各項目の得点を加算し，全体点を予測する．

しかし，この自動採点モデルには「人工的に高得点答案を作成可能である」という問題点がある．Word2Vec を用いて単語のみを分散表現に変換し，答案を単語ごとに評価する．そのため，答案に重要単語が含まれていれば，単語の使い方が間違っている場合や，説明不足である場合，文章全体の意味が異なっている場合でも高得点を得る．

2.3 むすび

本章では，関連研究である水本らによる“採点項目に基づく国語記述式答案の自動採点”の手法について述べた．

第3章　　関連技術

3.1 まえがき

本章では，関連技術について述べる．本研究で用いる自然言語処理技術とコサイン類似度の概要，及び自動採点における利点と問題点について述べる．

3.2 形態素解析器

3.2.1　　形態素解析の概要

形態素解析とは，自然言語からなる文を，意味を持つこれ以上分割できない最小の単位（単語，形態素）に分割する解析である．文法や単語における品詞等の情報（辞書）に基づき分割する．また，それぞれの単語の品詞等を判別する．日本語は英語と異なり，文章中の単語が連なっている．そのため，形態素解析は自然言語処理において主要な解析の一つである．形態素解析の概要図を図 3.1 に示す．

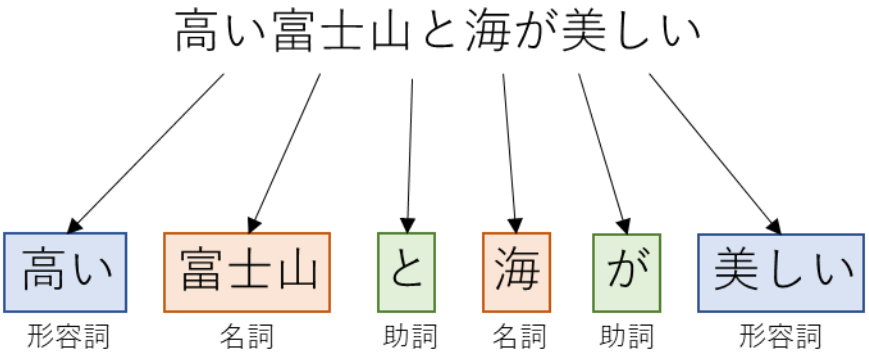

図 3.1　形態素解析の概要図

3.2.2　　MeCab

MeCab[6] は，京都大学情報学研究科と日本電信電話株式会社コミュニケーション科学基礎研究所の共同研究によって開発された形態素解析器である．品詞情報を利用した日本語の形態素解析を行うことができる．

言語，辞書，コーパスに依存しない汎用的な設計を基本方針とし，パラメータの推定に Conditional Random Fields (CRF)を用いている．そのため，ChaSen が採用している隠れマルコフモデルに比べ解析性能が高い．また， ChaSen，JUMAN，KAKASI より高速に動作する[7]．ChaSen，JUMAN，KAKASI は，MeCab と同様に一般的に広く知られている形態素解析器である．

3.2.3 JUMAN++

JUMAN++[6] は，京都大学大学院情報学研究所の黒橋・河原研究室によって開発された高性能な日本語の形態素解析器である．言語モデルとして Recurrent Neural Network Language Model (RNNLM)を用いることにより，単語の並びの意味的な自然さを考慮した解析を行うことができる．そのため，MeCab や JUMAN に比べ大きく解析性能が向上している．文法・辞書・出力フォーマット等は JUMAN から引き継いだものを利用している[8].

JUMAN++のもととなった JUMAN[6] は計算機による日本語解析の研究を目指す多くの研究者に共通に使える形態素解析器を提供するため開発された．国語文法が計算機向きではないという問題を考慮し，使用者によって文法の定義，単語間の接続関係の定義などを容易に変更できる．JUMAN は独自の辞書とともに配布されており，MeCab で一般的に用いられる IPADic に比べ，付加情報が豊富である[9].

3.3 Word2Vec

3.3.1 Word2Vec の概要

Word2Vec[10] は，Tomas らによって開発され，2014 年に Google から発表されたテキスト処理を行うニューラルネットワークである．大量のテキストデータを解析することによって，自然言語である単語を，意味を持った分散表現（ベクトル）に変換できる．形態素解析済みのテキストコーパスを入力とし，コーパスにある単語の特徴量ベクトルを出力する．単語を単語間の関連性や類似度に基づいてベクトルで表現する．そのため，単語同士の類似度を算出することや，意味の演算処理（加減算）を行うことができる．

出力される一つの単語ベクトルは，数百の次元から成る．例えば，単語「海」の分散表示は[0.98, 0.91, 0.11, …]のような形になり，単語「海」におけるすべての要素の重みが一つのベクトルに圧縮される．つまり，単語ベクトルは，対象の単語におけるそれぞれの要素との相関性数値で表現される．

Word2Vec による，単語を分散表現に変換するイメージ図を図 3.2 に示す．

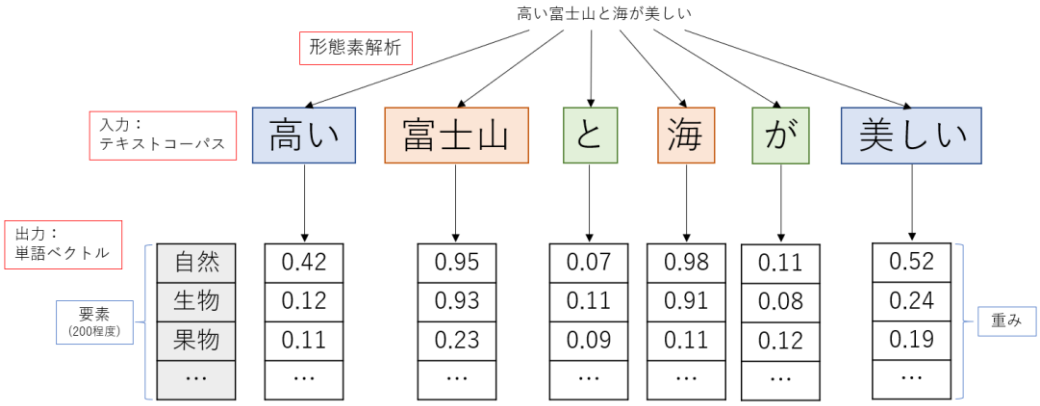

図 3.2　Word2Vec による，単語を分散表現に変換するイメージ図

Word2Vec は，類似語のベクトルをベクトル空間上にグループ化することで，数値に基づいて類似性を検知できる．そのため，十分な学習データが与えられれば過去の出現例をもとに，高い精度で単語の意味を推定できる．また，単語同士の関連性を確立できる．

ベクトル空間上における各単語ベクトルの関係と単語ベクトルの演算処理のイメージ図を図 3.3 に示す．

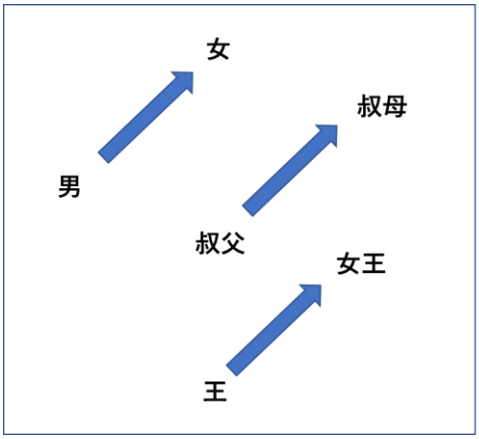

図 3.3　ベクトル空間上における各単語ベクトルの関係と演算処理のイメージ図

図 3.3 では，左下に向かうほど男を意味する単語がグループ化され，右上に向かうほど女を意味する単語がグループ化されている．このように，男女間の関係がそれぞれ近い値のベクトルで表される．このように，類似語のベクトルをベクトル空間上にグループ化することで単語間の関連性を確立している．

Word2Vec は，ベクトルを得るための手法が二つある．前後の単語から対象の単語を予測するニューラルネットワーク Continuous Bag-of-Words (CBoW)[10] と，ある単語から前後の単語を予測するニューラルネットワーク Skip-gram[10] である．学習時間は CBoW の方が Skip-gram より短いが，解析精度は Skip-gram の方が CBoW より高い．

CBoW の概要図を図 3.4，Skip-gram の概要図を図 3.5 に示す．

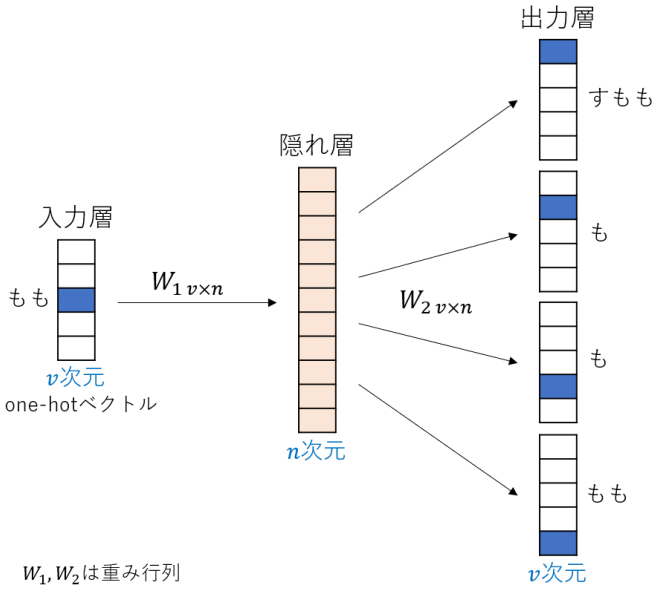

図 3.4　CBoW の概要図

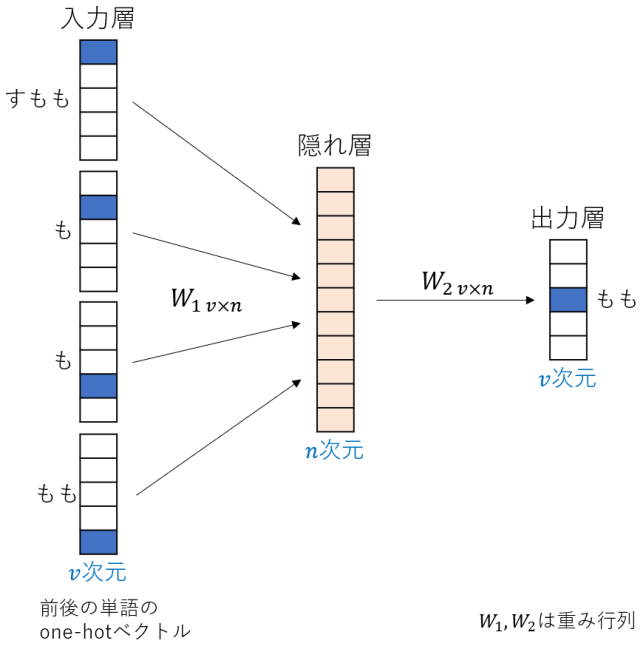

図 3.5　Skip-gram の概要図

3. 3. 2　　自動採点における Word2Vec の利点と問題点

コンピュータによって答案を自動的に採点するために，自然言語で記述された答案をコンピュータが処理可能な数値に変換する必要がある．この問題を解決するには Word2Vec が有効である．Word2Vec は，自然言語で記述された単語を，意味を持った分散表現に変換することで，単語同士の類似度の測定，類似度の高い単語の検出，単語同士の演算等ができる．また，十分な学習データが与えられればそれらの精度は高くなる．

そのため，模範解答と生徒によって作成された答案を比較し，評価する自動採点において，Word2Vec は有効である．

しかし，問題点も挙げられる．答案は，複数の単語の組み合わせである文章から成るが，Word2Vec は，全体の文章ではなく一部の単語のみを分散表現に変換する．そのため，「人工的に高得点答案を作成可能である」という自動採点における決定的な問題が生じる．単語の使い方が間違っている場合や，説明不足である場合，文章全体の意味が異なっている場合でも，答案に重要単語が含まれていれば高得点を取得できる．例として，模範解答が「今日は天気がいい」，生徒による答案が「今日の天気は良くない」の場合を挙げる．この場合，文章全体の意味は正反対であるが，どちらの答案にも全く同じ単語「今日」と「天気」が含まれ，それぞれの答案に意味の似た単語「いい」と「良く」が含まれている．そのため，自動採点の結果，文章全体の類似度は高くなり高得点を得られる可能性がある．このように，Word2Vec を用いて，単語のみを分散表現に変換し単語ごとに比較することは，大きな採点ミスにつながる．

3.4 Doc2Vec

3.4.1 Doc2Vec の概要

Doc2Vec[11] は, Quoc ら(2014)によって開発された文章を分散表現に変換するニューラルネットワークである．大量のテキストデータを解析することによって，自然言語で記述された文章を，意味を持った分散表現に変換できる．そのため，解析後のベクトル同士の類似度を測定し，文章分類や似た文章の検出ができる．

Doc2Vec は，Word2Vec のアルゴリズムを応用して開発された．形態素解析済みのテキストコーパスを入力とし，文章の特徴量ベクトルを出力する．任意の長さの文章を分散表現に変換することができ，特定のタスクに依存することがない．そのため，コンテンツベースのレコメンドや，感情分類，文章分類，スパムフィルタリング等に活用される．さらに，機械学習のモデルにおける入力には固定長のベクトルが用いられるため，Doc2Vec による解析結果を入力ベクトルとして使用できる．

Doc2Vec による，文章を分散表現に変換するイメージ図を図 3.6 に示す．

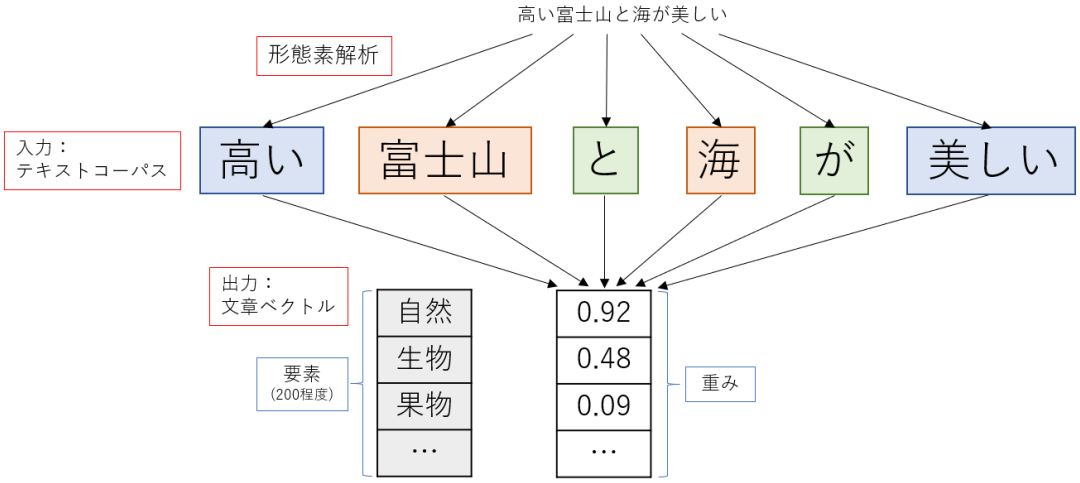

図 3.6　Doc2Vec による，文章を分散表現に変換するイメージ図

Doc2Vec は，ベクトルを得る手法が二つある．Distributed Memory (dmpv)[11] と Distributed Bag-of-Words (DBoW)[11] である．dmpv は，Word2Vec の前後の単語から対象の単語を予測する CBoW に似た構造を持つ．入力ベクトルは，単語列だけでなく文章 ID が付加される．文章 ID を入力に加えることで，文脈を意識した学習が可能になる．一方，DBoW は，Word2Vec のある単語から前後の単語を予測する Skip-gram に似た構造を持つ．入力ベクトルは，単語ではなく文章 ID である．該当文章内に含まれる単語を予測するように学習する．学習時間は DBoW の方が dmpv より短いが，解析精度は dmpv の方が DBoW より高い．

dmpv の概要図を図 3.7，DBoW の概要図を図 3.8 に示す．

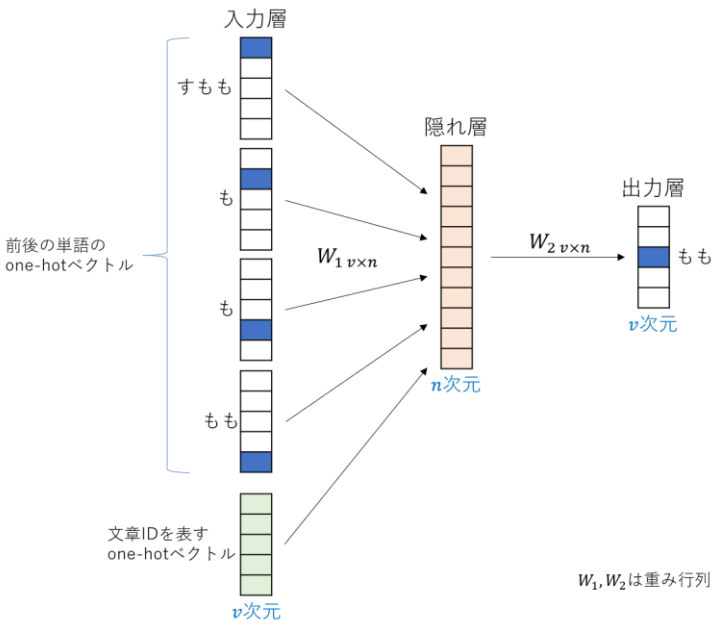

図 3.7　dmpv の概要図

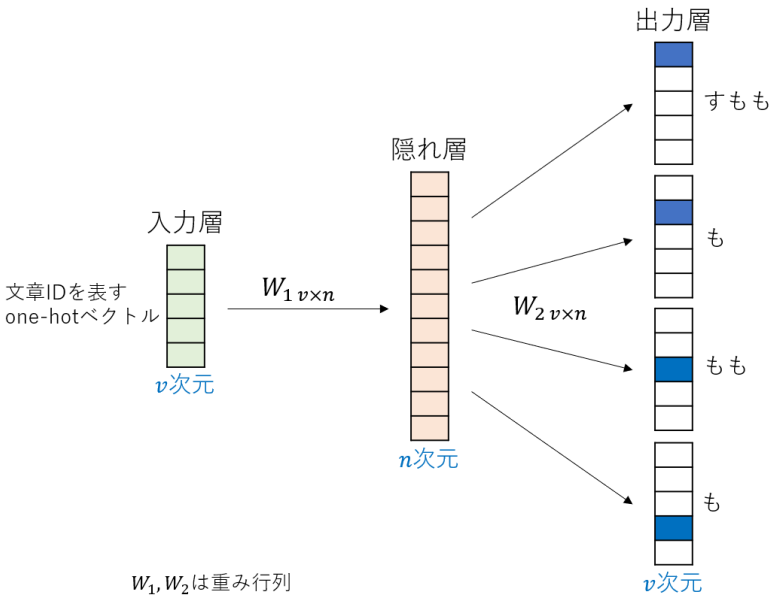

図 3.8　DBoW の概要図

3. 4. 2　　Doc2Vec と Word2Vec の違い

Word2Vec は，自然言語で記述された単語を，意味を持った分散表現に変換できる．一方で，Doc2Vec は，単語のみでなく自然言語で記述された文章を，意味を持った分散表現に変換できる．

3. 4. 3　　自動採点における Doc2Vec の利点と課題

Doc2Vec は Word2Vec とは異なり，一部の単語だけでなく文章全体を分散表現に変換できる．そのため，自動採点においても文章全体の意味を捉え，模範解答と生徒による答案の類似度を測定，評価できる．したがって，答案を分散表現化する手法として Doc2Vec を用いることで，Word2Vec を用いる際に生じる「人工的に高得点答案を作成可能である」といった問題を解決できると考えられる．つまり，生徒による答案に重要単語が含まれていたとしても，文章全体の意味として間違っている場合得点は低くなり，採点の精度は Doc2Vec を用いることで Wod2Vec に比べ高くなると考えられる．

しかし，Doc2Vec による文章を分散表現に変換する解析精度は，学習させるデータの内容と量に依存する．十分な学習データが与えられれば過去の出現例をもとに，単語及び文章の意味を高い精度で推定できる．一方，学習させるデータが不十分で，解析する文章内に学習データにはない未知の単語（以下，未知語）が含まれていた場合，その単語のベクトルに関して，DBoW では無視され，dmpv では乱数で処理される．DBoW は，文章 ID のみを入力とするため，未知語は無視する．dmpv は文章 ID だけでなく文章内の単語ベクトルも入力とするため，未知語を乱数で処理する．つまり，文章内に未知語が含まれていた場合，同一文章であっても実行ごとに得られるベクトルは異なる．した

がって，未知語を含んだ文章の意味推定の精度は低く，出力された文章ベクトルの信頼性も低い．模範解答と生徒による答案の類似度を評価する自動採点において，これは重大な問題点であり，学習にどれだけ多くのデータを与えられるかが課題である．

3.5 BERT

3.5.1 BERT の概要

BERT[12] は，Jacob ら（2018）によって開発された，汎用言語表現モデルである．双方向 Transformer[13] で言語モデルを事前学習することで汎用性を獲得し，自然言語処理の多くのタスクで State Of The Art (SOTA)が達成されている．転移学習をすることが可能であり一つのモデルをファインチューニングすることで，自然言語に関わる様々な問題に用いることができる．

BERT は，様々な自然言語処理に対して汎用的に使用可能なベクトルを算出できる．つまり，BERT を用いることで単語を，文脈を含んだ正確な意味を表す分散表現に変換できる．

BERT の特徴として文脈を理解できることが挙げられる．BERT は，Attention 機構を採用した Transformer を活用している．Attention 機構は，文字列における単語の間に存在する文脈的な関係を学習する．そのため，文脈に沿って単語を理解できる．これまでの自然言語処理技術は，解析をする文章内に複数の意味を持つ単語（多義語）が含まれているとき，その単語がどの意味で使われているのか理解できなかった．しかし BERT は，文章全体を参照することで，様々な意味の中から適切な意味を選ぶことが可能である．

BERT は，Transformer の Encoder ブロックから構成され，ネットワーク側ではなく学習データ側にマスクをかけることで双方向 Transformer を実現した．入力は，単語ベクトルの列で，出力も単語ベクトルの列になる．入力された単語ベクトルの列に文章の情報を付与して出力する．入力される単語ベクトルは，以下の三つのベクトルの和である．対象単語に対応した個別のトークンベクトル，入力された複数の文の内，対象単語が何文目に含まれているのかを表すセグメントベクトル，その文の中で，対象単語は何単語目に現れるのかを表すポジションベクトルである．出力は，入力された単語ベクトルと同じ数の単語ベクトルの列であり，各単語ベクトルは各単語が文章中での働きを示す情報が付与されている．BERT の入出力のイメージ図を図 3.9 に示す．

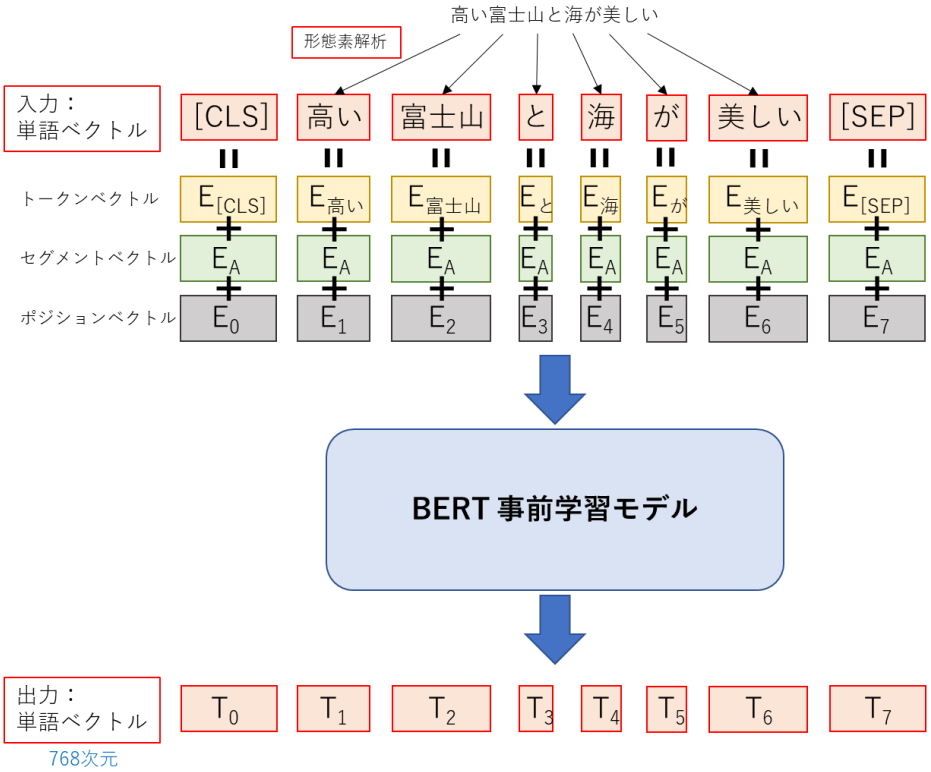

図 3.9　BERT の入出力のイメージ図

BERT は，2 種類の事前学習を行い，言語表現モデルを学習させる．Masked Language Model (MLM)と Next Sentence Prediction (NSP)である．MLM は，一つの文の内，15%の単語をランダムに隠して単語を予測することで文脈を学習する．NSP は，二つの文を与え，二つの文が後続で隣り合っているのか判定することで文脈を学習する．

3.5.2　BERT と Word2Vec の違い

BERT は Word2Vec と同様に，自然言語で記述された単語を，意味を持った分散表現に変換できる．しかし，BERT は，文脈を理解することができ，出力する単語ベクトルに文脈の情報も付加する．

Word2Vec は, 単語の内部状態を多次元であるベクトル空間の 1 点に収束させるため，多義語の解釈ができない．つまり，解析する文章が異なるものでも，未知語でなければ同じ単語に対し，文脈に関わらず常に同じベクトルを出力する．一方 BERT は，文脈によって単語の内部状態が異なるため多義語の場合, 別々の語義はベクトル空間上の別の場所に位置できる．また，Word2Vec は，固定長のウィンドウサイズの範囲でのみ共起性を考慮しているのに対し，Attention 機構を用いた BERT は文章全体を考慮している．そのため，BERT は，文脈に基づいた観点から単語の表現を定義・獲得する．つまり，BERT は Word2Vec と異なり，同じ単語でも文脈によって出力するベクトルが異なる．このように，同じ表層系を持つ単語の多義性を正しく認知できるため，Word2Vec より

も適切に単語を分散表現に変換できる．

BERT と Word2Vec による，分散表現化の違いのイメージ図を図 3.10 に示す．

図 3.10　BERT と Word2Vec による，分散表現化の違いのイメージ図

3. 5. 3　　自動採点における BERT の利点と課題

BERT は，自然言語で記述された単語を，文脈に基づいた観点から適切な意味を持った分散表現に変換する．そのため，Word2Vec と異なり，文脈の意味を理解し，文脈の情報を加味した単語の適切なベクトルを得る．得られる各単語ベクトルは，文脈の情報を付加することで，文章内でのその単語の役割と文章全体の意味を含む．そのため，自動採点においても，模範解答と生徒による答案の類似度を測定し，評価する際，答案中のいくつかの重要単語を比較することで文章全体の類似度を比較できると推測される．したがって，BERT を用いることで，　Word2Vec を自動採点に用いる際に生じる「人工的に高得点答案を作成可能である」といった問題を解決できると推測できる．

BERT は，得られた各単語ベクトルの和を求めることで，文章ベクトルを得ることができる．しかし，各単語ベクトルは 768 次元であり，その和を求めた文章ベクトルは，768 × トークン数（単語ベクトルの数）〔次元〕になる．そのため，自動採点において，模範解答と生徒による答案の単語数が異なると次元数も異なり，類似度を比較できない．次元数を揃えるため，畳み込みニューラルネットワークにより一つのベクトルに集約す

る技術は存在するが，平均ベクトルと最大値からなるベクトルを結合したものであり，もともとのベクトルから大きく変化され，ベクトルの持つ単語及び文章の意味の精度が下がると推測される．そのため，本研究では，答案中のいくつかの重要単語を比較することで文章全体の類似度を比較する．

3.6 コサイン類似度

コサイン類似度は，二つのベクトルの内積を用いて類似度を算出する．-1 から 1 までの値をとり，数値が大きくなるほど類似度が高くなる．コサイン類似度は以下の式(3.1)で計算できる．

$$Similarity = \cos\theta = \frac{\vec{x}\cdot\vec{y}}{|\vec{x}||\vec{y}|} = \frac{\vec{x}^T\vec{y}}{|\vec{x}||\vec{y}|} \tag{3.1}$$

コサイン類似度を用いて，Doc2Vec や BERT によって得たベクトルの類似度を測定できる．

3.7 むすび

本章では，関連技術について述べた．本研究で用いる自然言語処理技術とコサイン類似度の概要，及び自動採点における利点と問題点について述べた．

第4章　　提案手法

4.1 まえがき

本章では，自然言語処理技術を用いた答案採点支援システムの提案手法及び評価方法について述べる．提案手法として，Doc2Vec を用いた文章全体を分散表現に変換し比較する手法と，BERT を用いた単語を，文脈の情報を加味した分散表現に変換し比較する手法を検討する．

4.2 提案手法 1

4.3.1　　Doc2Vec を用いた文章全体を分散表現に変換する手法

自動採点において Doc2Vec を用い，文章全体を分散表現に変換して答案の類似度を測定，評価する手法を検討する．

模範解答と生徒による答案からそれぞれの文章ベクトルを取得し，類似度を測定し生徒による答案を評価する．まず初めに，JUMAN++を用いて模範解答と答案の文章に対し，形態素解析を行う．次に，Doc2Vec を用いてそれぞれの文章を，意味を持った分散表現に変換する．最後に，コサイン類似度によって，Doc2Vec より得た二つの文章ベクトルの類似度を測定し，類似度から得点を得る．

Doc2Vec を用いた答案採点支援システムのイメージ図を図 4.1 に示す．

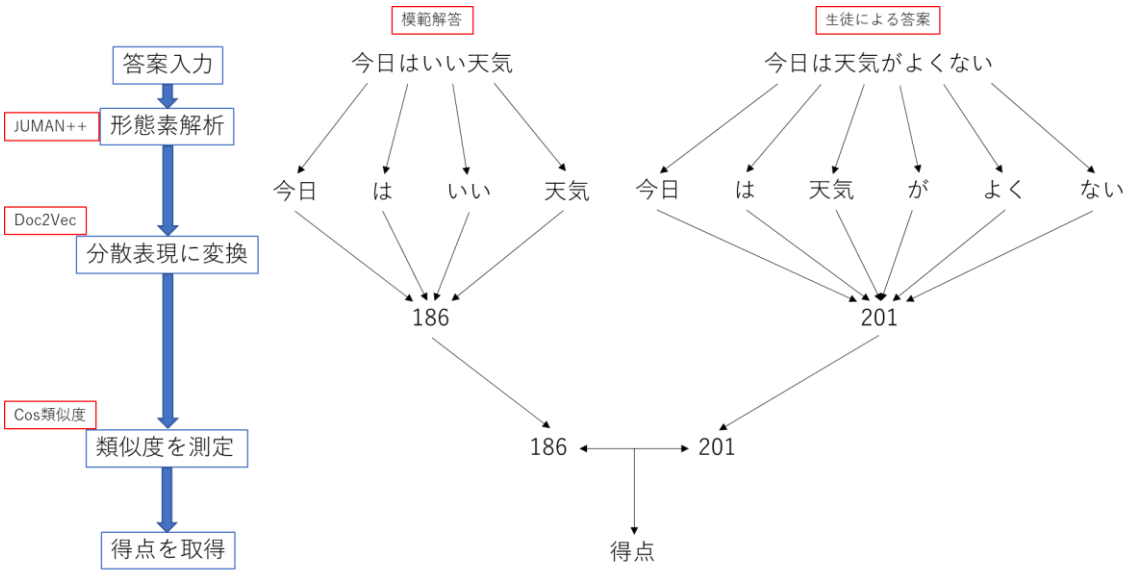

図 4.1　Doc2Vec を用いた答案採点支援システムのイメージ図

4.3.2 Doc2Vec の学習モデル

Doc2Vec の学習モデルを作成するにあたり，学習データとして複数の短編小説を青空文庫[14] よりダウンロードし用いる．青空文庫とは，著作権の保護期間が切れた作品や著者が許諾した作品のテキストを公開しているインターネット上の電子図書館である．

学習に用いる総データ数は, 12,651 テキストである. 学習データの例を図 4.2 に示す.

オシャベリ姫
夢野久作

ある国に王様がありまして，夫婦の間にたった一人，オシャベリ姫というお姫さまがありました.
このお姫様は大層美しいお姫様でしたが，どうしたものか生れ付きおしゃべりで，朝から晩まで何かしらシャベッていないと気もちがわるいので，おまけにそれを又きいてやる人がいないと大層御機嫌がわるいのです.

（中略）

そのときに二人の女中は王様から沢山の御褒美をいただきました.
そうして死ぬまで忠義にムクチ王子とオトナシ姫に仕えました.

底本：「夢野久作全集 1」ちくま文庫，筑摩書房
1992（平成 4）年 5 月 22 日第 1 刷発行
※底本の解題によれば，初出時の署名は「かぐつちみどり」です.
入力：柴田卓治
校正：江村秀之
2000 年 5 月 17 日公開
2006 年 5 月 3 日修正
青空文庫作成ファイル：
このファイルは，インターネットの図書館，青空文庫（http://www.aozora.gr.jp/）で作られました.
入力，校正，制作にあたったのは，ボランティアの皆さんです.

図 4.2　学習データの例

学習するにあたり，テキストデータを事前に処理する．まずそれぞれのテキストデータから不要な部分を取り除く. 本文の前後にある説明部分は学習に不要なため削除する. その後, 文章を単語に分割し学習に用いる品詞を選択する. 本研究では, 文章中の動詞, 形容詞，名詞を用いる．処理したテキストデータを学習データとして，学習モデル作成に用いる．学習時のパラメータを表 4.1 に示す．パラメータのそれぞれの説明を以下に箇条書きで示す.

表 4.1　学習時のパラメータ

パラメータ	設定値
size	400
alpha	0.0015
sample	$1e^{-4}$
min_count	1
workers	4

・size: 分散表現の次元数である．
・alpha: 学習率である．値が大きいほど収束するのが早いが，大きすぎると発散する．また，値が小さいほど解析精度は高いが，収束するのが遅い．
・sample: 単語を無視する際の頻度の閾値である．あまりに高い頻度で出現する単語は意味のない単語である可能性が高いため無視する．
・min_count: 学習に用いる単語の最低出現回数である．sample とは逆に，出現頻度が低すぎる単語もその文章に適切でない可能性があるため無視する．しかし，本研究では，すべての単語を対象とする．
・workers: 学習時のスレッド数である．

表 4.1 のパラメータで，12,651 テキストの短編小説を用いた際の学習時間は約 3 時間 30 分であった．

4.3 提案手法 2

4.3.1　BERT を用いて単語を，文脈情報を加味した分散表現に変換する手法

自動採点において BERT を用い，答案中の単語を，文脈情報を加味した分散表現に変換して答案の類似度を測定，評価する手法を検討する．

模範解答と生徒による答案に含まれる重要単語のベクトルを取得し，類似度を測定し生徒による答案を評価する．まず初めに，JUMAN++を用いて模範解答と答案の文章に対し，形態素解析を行う．次に，BERT を用いて，各文章中の単語を，文脈の情報も加味した分散表現に変換する．得られた単語ベクトル列の中から模範解答と答案に共通して含まれる重要単語のベクトルを複数選択する．コサイン類似度によって，対応する重要単語同士でそれぞれ類似度を測定する．平均類似度をもとに得点を得る．

BERT を用いた答案採点支援システムのイメージ図を図 4.3 に示す．

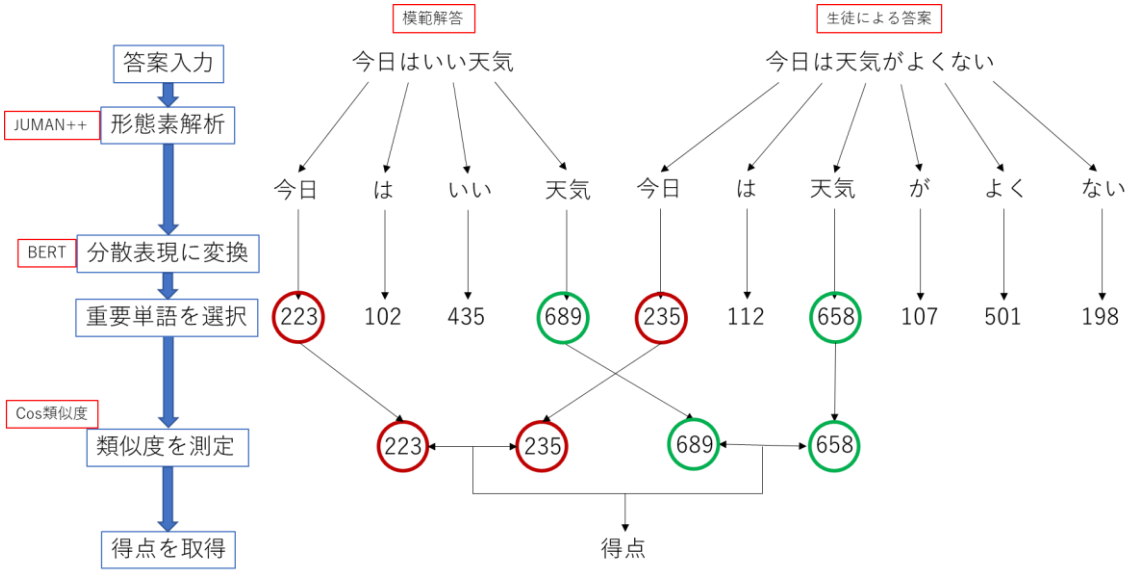

図 4.3　BERT を用いた答案採点支援システムのイメージ図

4. 3. 2　　BERT の学習済みモデル

本研究では，京都大学の黒橋・河原研究室で公開されている学習済みモデル[15] を用いる．BERT の公式サイトでは英語 pretrained モデルや多言語 pretrained モデルが公開されている．そのモデルを使って対象タスクで finetuning することによりそのタスクを高精度に解くことができる．多言語 pretrained モデルには日本語も含まれているため，日本語のタスクに多言語 pretrained モデルを利用できる．しかし，黒橋・河原研究室では，基本単位が文字になっていることは適切でないと考え，日本語 pretrained モデルを作成した．入力テキストを形態素解析し，形態素を subword に分割したものを基本単位として，日本語テキストのみで pretraining したものである．

京都大学黒橋・河原研究室による日本語 pretrained モデルの詳細を以下に箇条書きで示す[15].

・入力テキスト：日本語 Wikipedia 全て (約 1,800 万文，半角を全角に正規化)
・入力テキストに Juman++で形態素解析を行い，さらに BPE を適用し subword に分割
・BERT_{BASE}と同じ設定 (12-layer, 768-hidden, 12-heads)
・30 epoch (1GPU で 1epoch に約 1 日かかるため pretraining に約 30 日)
・語彙数: 32,000 (形態素，subword を含む)
・max_seq_length: 128

4. 4 評価方法

提案する二つの答案採点支援システムの精度を，人手による採点結果と比較することで評価する．本研究では，システムより得られる答案の類似度と人手による採点から得られる得点の無相関検定を行い，提案手法の精度を評価する．

無相関検定とは，標本から算出した相関係数を用いて，母集団の相関係数（母相関係数）が 0 かどうか検定することである．「2 変数間に相関がない（母相関係数が 0 である)」という帰無仮説のもと，2 変数間の相関の有無を調べることができる．有意であると判定された場合この帰無仮説が棄却され，「2 変数間に相関がある」ということを統計的に確認できる．

直線的な相関関係の強さを表す指標の一つに「相関係数」がある．二つの要素xとyからなるn個のデータ(x_i, y_i)が得られたとき，相関係数r_{xy}は以下の式(4.1)で算出される．

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \times \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4.1}$$

相関係数r_{xy}の範囲は$-1 \leq r_{xy} \leq 1$である．値が正のとき正の相関関係，値が負のとき負の相関関係，値が 0 に近いとき$(|r_{xy}| < 0.2)$，無相関であることを示す．

無相関検定は，t分布を用いる．統計量tは自由度$(n-2)$のt分布に従い，以下の式(4.2)で算出される．

$$t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} \tag{4.2}$$

母相関係数pは，t分布において$|r|$以上の値が発生する確率である．この母相関係数pの値で，帰無仮説をもとに 2 変数間の相関の有無を調べることができる．$p < 0.05$（有意水準）のとき，帰無仮説が棄却され，「2 変数間に相関がある」と確認できる．

4.5 むすび

本章では，自然言語処理技術を用いた答案採点支援システムの提案手法及び評価方法について述べた．提案手法として，Doc2Vec を用いた文章全体を分散表現に変換し比較する手法と，BERT を用いた単語を，文脈の情報を加味した分散表現に変換し比較する手法を検討した．

第5章　　実験及び結果・考察

5.1 まえがき

本章では，提案手法の実験結果及び考察について述べる．

5.2 形態素解析器の精度比較

5.2.1　　実験概要

MeCab と JUMAN++による形態素解析器の解析精度を比較する．五つの文章をそれぞれ，MeCab と JUMAN++を用いて形態素解析し，その結果を比較する．解析に用いる五つの文章を以下に示す．

(1) すももももももものうち
(2) 今日はいい天気
(3) 今日は天気が良くない
(4) 高い富士山と海が美しい
(5) ある日の暮方のことである．一人の下人が，羅生門の下で雨やみを待っていた．

5.2.2　　結果と考察

MeCab による形態素解析の結果例を図 5.1，JUMAN++による形態素解析の結果例を図 5.2 に示す．

```
kobayashi@kobayashi:~$ mecab
すももももももものうち
すもも　名詞,一般,*,*,*,*,すもも,スモモ,スモモ
も　助詞,係助詞,*,*,*,*,も,モ,モ
もも　名詞,一般,*,*,*,*,もも,モモ,モモ
も　助詞,係助詞,*,*,*,*,も,モ,モ
もも　名詞,一般,*,*,*,*,もも,モモ,モモ
の　助詞,連体化,*,*,*,*,の,ノ,ノ
うち　名詞,非自立,副詞可能,*,*,*,うち,ウチ,ウチ
EOS
```

図 5.1　MeCab による形態素解析の結果例

```
kobayashi@kobayashi:~$ jumanpp
すももももももものうち
すもも すもも すもも 名詞 6 普通名詞 1 * 0 * 0 "自動獲得:テキスト"
も も も 助詞 9 副助詞 2 * 0 * 0 NIL
もも もも もも 名詞 6 普通名詞 1 * 0 * 0 "代表表記:股/もも カテゴリ:動物-部位"
@ もも もも もも 名詞 6 普通名詞 1 * 0 * 0 "代表表記:桃/もも 漢字読み:訓 カテゴリ:植物;人工物-食べ物 ドメイン:料理・食事"
も も も 助詞 9 副助詞 2 * 0 * 0 NIL
もも もも もも 名詞 6 普通名詞 1 * 0 * 0 "代表表記:股/もも カテゴリ:動物-部位"
@ もも もも もも 名詞 6 普通名詞 1 * 0 * 0 "代表表記:桃/もも 漢字読み:訓 カテゴリ:植物;人工物-食べ物 ドメイン:料理・食事"
の の の 助詞 9 接続助詞 3 * 0 * 0 NIL
うち うち うち 名詞 6 副詞的名詞 9 * 0 * 0 "代表表記:うち/うち"
EOS
```

図 5.2　JUMAN++による形態素解析の結果例

文章(1)から文章(5)の五つの文章の，MeCab と JUMAN++を用いた形態素解析の結果を表 5.1 に示す.

表 5.1　MeCab と JUMAN++による形態素解析の結果

入力文章	MeCab による分割数（単語）	JUMAN++による分割数（単語）
(1)	7	7
(2)	4	4
(3)	6	6
(4)	6	7
(5)	27	24

文章(4)と文章(5)において解析結果に差が生じた．文章(4)の MeCab による解析結果を図 5.3，JUMAN++による解析結果を図 5.4，文章(5)の MeCab による解析結果を図 5.5，JUMAN++による解析結果を図 5.6 に示す.

```
kobayashi@kobayashi:~$ mecab
高い富士山と海が美しい
高い	形容詞,自立,*,*,形容詞・アウオ段,基本形,高い,タカイ,タカイ
富士山	名詞,固有名詞,一般,*,*,*,富士山,フジサン,フジサン
と	助詞,並立助詞,*,*,*,*,と,ト,ト
海	名詞,一般,*,*,*,*,海,ウミ,ウミ
が	助詞,格助詞,一般,*,*,*,が,ガ,ガ
美しい	形容詞,自立,*,*,形容詞・イ段,基本形,美しい,ウツクシイ,ウツクシイ
EOS
```

図 5.3　MeCab による文章(4)の形態素解析結果

```
kobayashi@kobayashi:~$ jumanpp
高い富士山と海が美しい
高い たかい 高い 形容詞 3 * 0 イ形容詞アウオ段 18 基本形 2 "代表表記:高い/たかい 反義:形容詞:安い/やすい;形容詞:低い/ひくい"
富士 ふじ 富士 名詞 6 地名 4 * 0 * 0 "代表表記:富士/ふじ 地名:日本:静岡県:市"
@ 富士 ふじ 富士 名詞 6 地名 4 * 0 * 0 "代表表記:富士/ふじ 地名:日本:静岡県:郡"
山 やま 山 名詞 6 普通名詞 1 * 0 * 0 "代表表記:山/やま 漢字読み:訓 地名末尾 カテゴリ:場所-自然"
@ 山 さん 山 名詞 6 普通名詞 1 * 0 * 0 "代表表記:山/さん 漢字読み:音 地名末尾 カテゴリ:場所-自然"
と と と 助詞 9 格助詞 1 * 0 * 0 NIL
海 うみ 海 名詞 6 普通名詞 1 * 0 * 0 "代表表記:海/うみ 漢字読み:訓 地名末尾 カテゴリ:場所-自然"
@ 海 かい 海 名詞 6 普通名詞 1 * 0 * 0 "代表表記:海/かい 漢字読み:音 地名末尾 カテゴリ:自然物"
が が が 助詞 9 格助詞 1 * 0 * 0 NIL
美しい うつくしい 美しい 形容詞 3 * 0 イ形容詞イ段 19 基本形 2 "代表表記:美しい/うつくしい 反義:形容詞:醜い/みにくい"
EOS
```

図 5.4　JUMAN++による文章(4)の形態素解析結果

```
kobayashi@kobayashi:~$ mecab
ある日の暮方のことである。一人の下人が、羅生門の下で雨やみを待っていた。
ある    連体詞,*,*,*,*,ある,アル,アル
日      名詞,非自立,副詞可能,*,*,*,日,ヒ,ヒ
の      助詞,連体化,*,*,*,*,の,ノ,ノ
暮方    名詞,副詞可能,*,*,*,*,暮方,クレガタ,クレガタ
の      助詞,連体化,*,*,*,*,の,ノ,ノ
こと    名詞,非自立,一般,*,*,*,こと,コト,コト
で      助動詞,*,*,*,特殊・ダ,連用形,だ,デ,デ
ある    助動詞,*,*,*,五段・ラ行アル,基本形,ある,アル,アル
。      記号,句点,*,*,*,*,。,。,。
一      名詞,数,*,*,*,*,一,イチ,イチ
人      名詞,接尾,助数詞,*,*,*,人,ニン,ニン
の      助詞,連体化,*,*,*,*,の,ノ,ノ
下人    名詞,一般,*,*,*,*,下人,ゲニン,ゲニン
が      助詞,格助詞,一般,*,*,*,が,ガ,ガ
、      記号,読点,*,*,*,*,、,、,、
羅生門  名詞,固有名詞,一般,*,*,*,羅生門,ラショウモン,ラショーモン
の      助詞,連体化,*,*,*,*,の,ノ,ノ
下      名詞,一般,*,*,*,*,下,シタ,シタ
で      助詞,格助詞,一般,*,*,*,で,デ,デ
雨      名詞,一般,*,*,*,*,雨,アメ,アメ
やみ    名詞,一般,*,*,*,*,やみ,ヤミ,ヤミ
を      助詞,格助詞,一般,*,*,*,を,ヲ,ヲ
待っ    動詞,自立,*,*,五段・タ行,連用タ接続,待つ,マッ,マッ
て      助詞,接続助詞,*,*,*,*,て,テ,テ
い      動詞,非自立,*,*,一段,連用形,いる,イ,イ
た      助動詞,*,*,*,特殊・タ,基本形,た,タ,タ
。      記号,句点,*,*,*,*,。,。,。
EOS
```

図 5.5　JUMAN++による文章(5)形態素解析結果

```
kobayashi@kobayashi:~$ jumanpp
ある日の暮方のことである。一人の下人が、羅生門の下で雨やみを待っていた。
ある ある ある 連体詞 11 * 0 * 0 * 0 "代表表記:或る/ある"
日 にち 日 名詞 6 時相名詞 10 * 0 * 0 "代表表記:日/にち 漢字読み:音 カテゴリ:時間"
@ 日 ひ 日 名詞 6 時相名詞 10 * 0 * 0 "代表表記:日/ひ 漢字読み:訓 弱時相名詞 カテゴリ:時間"
の の の 助詞 9 接続助詞 3 * 0 * 0 NIL
暮方 暮方 暮方 名詞 6 普通名詞 1 * 0 * 0 "自動獲得:Wikipedia Wikipediaページ内一覧:伊東深水 読み不明"
の の の 助詞 9 接続助詞 3 * 0 * 0 NIL
こと こと こと 名詞 6 形式名詞 8 * 0 * 0 NIL
である である だ 判定詞 4 * 0 判定詞 25 デアル列基本形 15 NIL
。 。 。 特殊 1 句点 1 * 0 * 0 NIL
一 一 一 名詞 6 数詞 7 * 0 * 0 "カテゴリ:数量"
人 にん 人 接尾辞 14 名詞性名詞助数辞 3 * 0 * 0 "代表表記:人/にん 準内容語 カテゴリ:人"
の の の 助詞 9 接続助詞 3 * 0 * 0 NIL
下人 げにん 下人 名詞 6 普通名詞 1 * 0 * 0 "自動獲得:Wikipedia"
が が が 助詞 9 格助詞 1 * 0 * 0 NIL
、 、 、 特殊 1 読点 2 * 0 * 0 NIL
羅生門 らしょうもん 羅生門 名詞 6 普通名詞 1 * 0 * 0 "自動獲得:Wikipedia Wikipedia多義"
の の の 助詞 9 接続助詞 3 * 0 * 0 NIL
下 した 下 名詞 6 普通名詞 1 * 0 * 0 "代表表記:下/した 漢字読み:訓 カテゴリ:場所-機能"
で で で 助詞 9 格助詞 1 * 0 * 0 NIL
雨 う 雨 名詞 6 普通名詞 1 * 0 * 0 "代表表記:雨/う 漢字読み:音 カテゴリ:抽象物"
@ 雨 あめ 雨 名詞 6 普通名詞 1 * 0 * 0 "代表表記:雨/あめ 漢字読み:訓 カテゴリ:抽象物"
やみ やみ やみ 名詞 6 普通名詞 1 * 0 * 0 "代表表記:闇/やみ カテゴリ:抽象物"
を を を 助詞 9 格助詞 1 * 0 * 0 NIL
待って まって 待つ 動詞 2 * 0 子音動詞タ行 6 タ系連用テ形 14 "代表表記:待つ/まつ"
いた いた いる 接尾辞 14 動詞性接尾辞 7 母音動詞 1 タ形 10 "代表表記:いる/いる"
。 。 。 特殊 1 句点 1 * 0 * 0 NIL
EOS
```

図 5.6　JUMAN++による文章(5)形態素解析結果

表 5.1 より，文章(4)，文章(5)の解析結果で多少の差は生じたが，全体的に大きな差はないことが確認できる．図 5.3 と図 5.4 より，「富士山」を MeCab は一つの固有名詞ととらえ，JUMAN++は地名の「富士」と「山」に分割していることがわかる．図 5.5 と図 5.6 より，「である」を MeCab は二つの助動詞に分割し，JUMAN++は一つの判定詞とらえていることがわかる．また，「待っている」を MeCab は二つの動詞と助詞，助動詞の三つに分割し，JUMAN++は一つの動詞と接尾語の二つに分割していることがわかる．

JUMAN++は RNNLM を用いることにより意味的な自然さを考慮した解析ができ，MeCab に比べ性能が大きく向上している．本実験の結果からも，MeCab がより細かく分割している一方で，JUMAN++は意味的に自然な境界で分割していることが確認できる．そのため，本研究では意味的な自然さを考慮した解析ができる JUMAN++を用いることが適切であると考えられる．

5.3 Doc2Vec を用いた自動採点

5.3.1 Doc2Vec による文章分散表現化の精度

5.3.1.1 実験概要

Doc2Vec による文章の分散表現化の精度を確認する．三つの文章を，Doc2Vec を用いて 2 度解析し，それぞれの文書で二つずつ文章ベクトルを得る．1 回目の解析より得られる文章ベクトルと 2 回目の解析より得られる文章ベクトルのコサイン類似度を算出し，精度を確認する．解析に用いる文章を以下に示す．

(1) 今日はいい天気
(2) 今日は天気が良くない
(3) ある日の暮方のことである．一人の下人が，羅生門の下で雨やみを待っていた．

また，文章(1)と文章(2)の文章ベクトルの類似度を算出し，正確な意味情報を含んだ文章ベクトルを得られているか，Doc2Vec による分散表現化の意味的精度を確認する．

5.3.1.2 結果と考察

Doc2Vec を用い，1 回目と 2 回目の解析より得られた文章ベクトルのコサイン類似度，及び文章(1)と(2)の文章ベクトルのコサイン類似度を表 5.2 に示す．

表 5.2　Doc2Vec による文章分散表現化の精度

解析文章	解析 1 回目と 2 回目のコサイン類似度	文章(1)と(2)のコサイン類似度
(1)	0.787	-
(2)	0.837	-
(3)	0.881	-
(1)と(2)	-	0.462

表 5.2 より，1 回目と 2 回目の解析から得られた文章ベクトルのコサイン類似度はすべての文章において，理想値である 1.0 を得られなかったことがわかる．全く同じ文章であるため，何度解析しても同じ値のベクトルを得られることが理想である．しかし，Doc2Vec は dmpv と DBoW の性質上未知語に弱い．そのため，解析文章に未知語が含まれていると本実験のような結果になったと考えられる．このことから，本研究で用いる

Doc2Vec の学習モデルは学習データが不十分であると考えられる．しかし，解析 1 回目と 2 回目のコサイン類似度は理想値ではないもののどれも 1.0 に近い値を得られている．そのため，ベクトル空間上で大まかな文章の意味分類ができているといえる．また，動作確認のため「私」の 1 語で実験を行ったところ，解析 1 回目と 2 回目のコサイン類似度は 1.0 を得られた．

文章(1)と文章(2)は全く反対の意味を持つ文章である．しかし，文章(1)と文章(2)のコサイン類似度は 0.46 となり，負の値を得られなかった．Doc2Vec を用いることで文章全体の意味を捉え，全く意味の異なる二つの文章ベクトルを，ベクトル空間上の離れたところに位置できると推測した．しかし，本実験からはそのことを確認できなかった．Doc2Vec は Word2Vec のアルゴリズムを応用したものである．そのため，文章中に似た単語が多く含まれていた場合，得られる文章ベクトルの類似度は比較的高くなるのだと考えられる．

5.3.2 Doc2Vec を用いた答案採点支援システムの有効性の評価

5.3.2.1 実験概要

本実験では，中学校の国語試験の採点を想定し Doc2Vec を用いた答案採点支援システム（以下，提案手法 1）の有効性を評価する．高校入試問題の中から表現工夫型の短答記述式問題を選択し，中学 3 年生が実際に解いた 40 枚の答案を収集する．表現工夫型とは，文章中の表現を抽出するだけでなく，自分なりの表現で言い換える必要がある問題である．

人手による採点結果として，現役の中学国語教師が模範解答をもとに採点する．問題は 4 題あり，1 題のみ 0 から 5 点までのいずれかの得点を与え，残りの 4 題は 0 から 9 点までのいずれかの得点を与える．人手によって採点された 40 枚の答案の中から，答案の内容と得点共に多様なものを各問題で 10 種ずつ選択する．これらの答案を提案手法 1 のシステムに入力することで，選択した答案と模範解答の類似度を算出する．人手による得点とシステムから得る答案類似度の無相関検定を行い，提案手法 1 の有効性を評価する．

問題 1 と問題 2，問題 3 は 20 字程度，問題 4 は 30 字程度の答案である．実験に用いる四つの問題における，模範解答と収集した 10 種の答案を図 5.7 に示す．

問題 1↵
［模範解答］↵
音をたてずに飛ぶフクロウの羽根↵
［生徒による答案］↵
1. 音をたてずに飛ぶフクロウ↵
2. フクロウの羽根前方についたくし状の細い毛↵
3. 生態系サービス↵
4. くし状の細い毛があるフクロウ↵
5. フクロウの音のたたない羽↵
6. 音をたてずに飛ぶ↵
7. フクロウのくし状の細い毛↵
8. 獲物に気づかれることのなく音を立てずに飛ぶ↵
9. 音を立てずに飛ぶ鳥↵
10. 獲物に気づかれず音を立てず飛ぶフクロウ↵
↵
問題 3↵
［模範解答］↵
草の上の露に、しきりに風が吹きつける↵
［生徒による答案］↵
1. 露がある場所に風が吹いている↵
2. 白露に吹いてくる風↵
3. 水をはじく葉の表面構造↵
4. 葉上の水玉が風に吹かれる↵
5. 露が風に吹かれてこぼれ落ちる↵
6. 秋の野原で葉が露や水をはじいている↵
7. 白露に暖かい風が吹きついている↵
8. 草におく露や、葉上の水玉などの↵
9. 草におく露に向かって風が吹きつけられる↵
10. 草や葉上の水玉に風が吹く↵

問題 2↵
［模範解答］↵
騒音の防止に役立つパンタグラフ↵
［生徒による答案］↵
1. 新幹線の車両の防止↵
2. 騒音の防止ができる新幹線の車両↵
3. 騒音防止のパンタグラフ↵
4. 騒音防止の車↵
5. 騒音の防止に役立つ集電装置↵
6. 騒音の防止↵
7. パンタグラフの騒音防止↵
8. パンタグラフ↵
9. 生物が人類に与えてくれる恩恵↵
10. フクロウを真似したパンタグラフ↵
↵
問題 4↵
［模範解答］↵
生物のもつ様々な情報に心を動かし、それを読み取る感性と知性↵
［生徒による答案］↵
1. 心を動かし、読み取る感性と知性↵
2. 保全にも活用にも、それに心を動かして読み取る感性と知性↵
3. 生物がもっているものに心を動かし、読み取る感性と知性↵
4. あらゆる戦略に関する情報に心を動かし読み取る感性と知性↵
5. 生物多様性の保全に心を動かし、読み取る感性と知性↵
6. 生物多様性の保全が必要で、それに心を動かし読み取る感性と知性↵
7. 生物多様性の保全、活用に心を動かし、読み取る感性と知性↵
8. それに心を動かし、読み取る感性と知性↵
9. 生物の貴重な情報は、心を動かし、読み取る感性と知性↵
10. 生物多様性や、その活用にも心を動かし、読み取る感性と知性↵

図 5.7　提案手法の評価実験に用いる答案

5.3.2.2　　結果と考察

問題 1（20 字程度の文章，$n = 10$）における，人手による採点結果と提案手法 1 による答案の類似度を表 5.3 に示す．また，散布図と最小二乗法により求めた近似直線で表したものを図 5.8 に示す．

表 5.3　問題 1 における人手による採点結果と答案の類似度（提案手法 1）

生徒による問題 1 の答案	人手による採点結果	答案の類似度
1	4	0.516
2	3	0.325
3	0	0.199
4	4	0.189
5	5	0.266
6	1	0.523
7	4	0.212
8	1	0.452
9	2	0.529
10	3	0.478

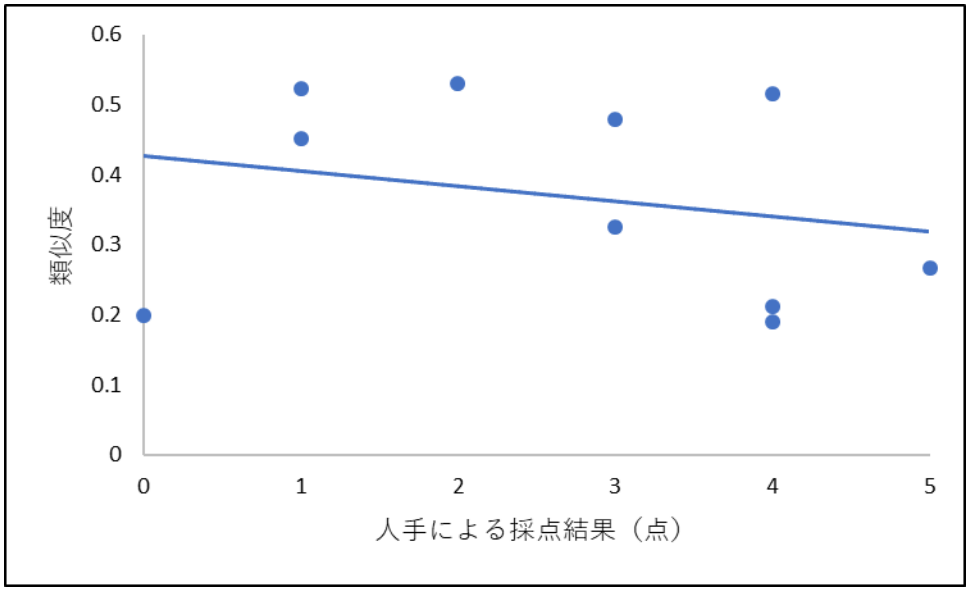

図 5.8　問題 1 における人手による採点結果と答案の類似度（提案手法 1）

問題 2（20 字程度の文章，$n = 10$）における，人手による採点結果と提案手法 1 による答案の類似度を表 5.4 に示す．また，散布図と最小二乗法により求めた近似直線で表したものを図 5.9 に示す．

表 5.4　問題 2 における人手による採点結果と答案の類似度（提案手法 1）

生徒による問題 2 の答案	人手による採点結果	答案の類似度
1	1	0.229
2	5	0.222
3	9	0.393
4	2	0.306
5	9	0.550
6	4	0.396
7	6	0.445
8	7	0.171
9	0	0.354
10	8	0.294

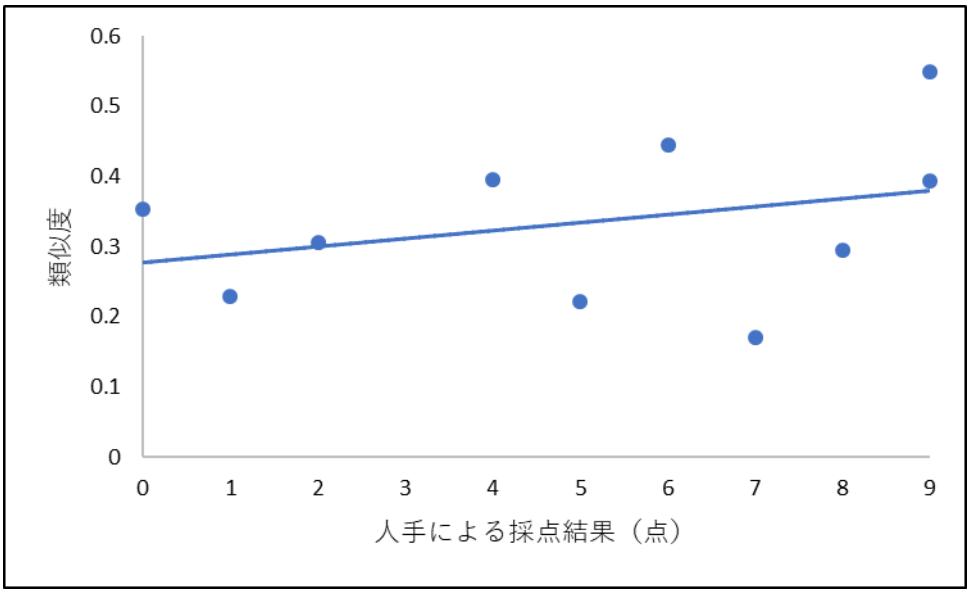

図 5.9　問題 2 における人手による採点結果と答案の類似度（提案手法 1）

問題 3（20 字程度の文章，$n = 10$）における，人手による採点結果と提案手法 1 による答案の類似度を表 5.5 に示す．また，散布図と最小二乗法により求めた近似直線で表したものを図 5.10 に示す．

表 5.5　問題 3 における人手による採点結果と答案の類似度（提案手法 1）

生徒による問題 3 の答案	人手による採点結果	答案の類似度
1	8	0.517
2	5	0.330
3	0	0.172
4	4	0.580
5	3	0.494
6	2	0.334
7	6	0.409
8	1	0.313
9	9	0.504
10	7	0.434

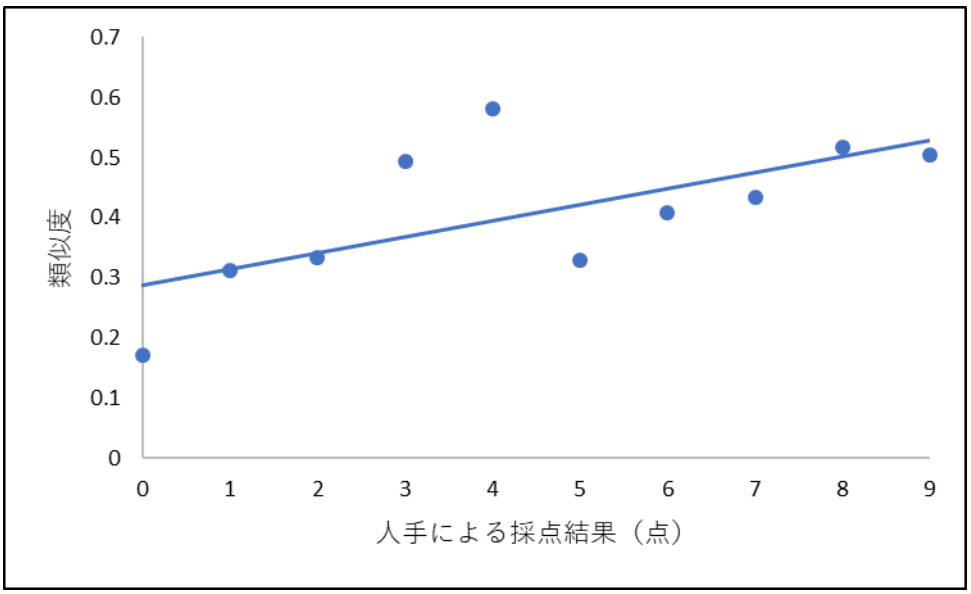

図 5.10　問題 3 における人手による採点結果と答案の類似度（提案手法 1）

問題 4（30 字程度の文章，$n = 10$）における，人手による採点結果と提案手法 1 による答案の類似度を表 5.6 に示す．また，散布図と最小二乗法により求めた近似直線で表したものを図 5.11 に示す．

表 5.6　問題 4 における人手による採点結果と答案の類似度（提案手法 1）

生徒による問題 3 の答案	人手による採点結果	答案の類似度
1	8	0.517
2	5	0.330
3	0	0.172
4	4	0.580
5	3	0.494
6	2	0.334
7	6	0.409
8	1	0.313
9	9	0.504
10	7	0.434

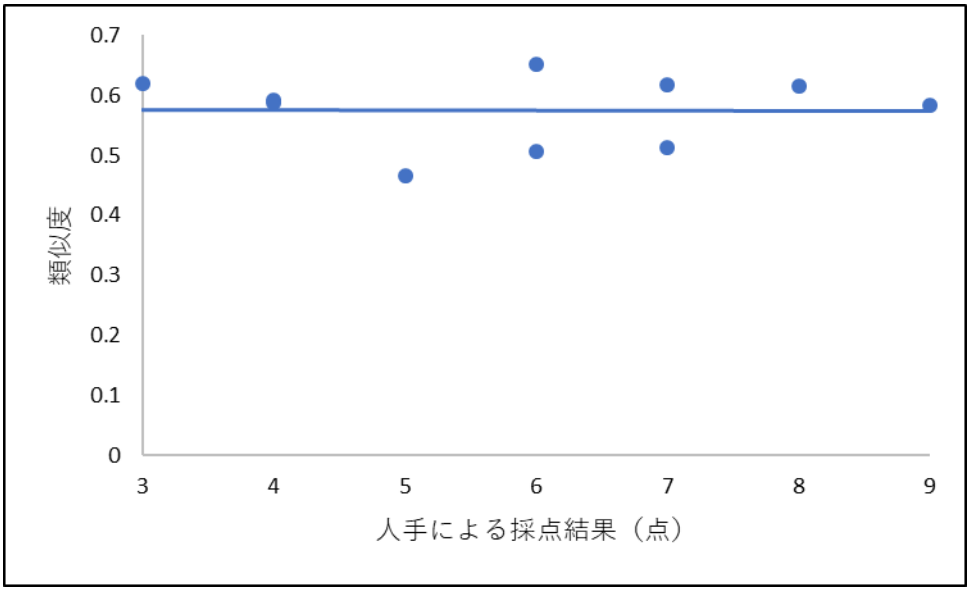

図 5.11　問題 4 における人手による採点結果と答案の類似度（提案手法 1）

表 5.3，表 5.4，表 5.5，表 5.6 より算出した，四つの問題における人手による採点結果と提案手法 1 による答案類似度の相関係数，及び無相関検定の結果を表 5.7 に示す．

表 5.7　人手による採点結果と答案類似度の相関関係（提案手法 1）

	問題 1	問題 2	問題 3	問題 4
標本数	10	10	10	10
自由度	8	8	8	8
相関係数	-0.244	0.325	0.661	-0.001
t値	0.713	0.973	2.490	0.002
p値	0.496	0.359	0.038	0.999

表 5.7 より，問題 2 と問題 3 において正の相関を得られたが，問題 1 と問題 4 は正の相関関係を得られなかったことがわかる．また，無相関検定の結果，問題 3 のみ$p \fallingdotseq 0.038 < 0.05$で帰無仮説が棄却され，人手による採点結果とシステムによる答案類似度に相関があることが示された．しかし，他の 3 題については，相関があることを示さなかった．

このような結果になった原因として大きく分けて二つ考えられる．

一つ目は，Doc2Vec の学習モデルを作成した際に用いた学習データが不十分であったことである，5.3.1 節で述べた実験の結果からもわかるように，Doc2Vec の解析精度は学習データに依存し未知語に弱いという特徴を持っている．本実験では，用いた学習データの量及び内容が不十分だったため，答案に未知語が含まれ，適切な分散ベクトルを得られなかったと考えられる．これは，学習データとして，国語の教科書や試験に出題

される文献を用い, かつより多くの資料を収集することで多少改善できると推測される.

二つ目は, 採点の対象が表現自由型の答案であったことである. Doc2Vec は, Word2Vec のアルゴリズムを応用したものであるため，5.3.1 節で述べた実験と同様に，模範解答と同じ単語が答案に多く含まれていると比較的高い類似度を得た．一方で，模範解答の言葉を他の表現に言い換えた答案は低い類似度を得た．そのため，単語の使い方が間違っている場合や，説明不足である場合，文章全体の意味が異なっている場合でも，模範解答と同じ単語が多く含まれた答案がより高い類似度を得る結果になったと考えられる．これは，一つの模範解答だけでなく，複数の模範解答を用いることで改善できる可能性があると推測される.

5.4 BERT を用いた自動採点

5.4.1 BERT による単語分散表現化の精度

5.4.1.1 実験概要

BERT による単語の分散表現化の精度を確認する．三つの文章を，BERT を用いて二度解析し，それぞれの文章で二つずつ単語ベクトル列を得る．1 回目の解析より得た単語ベクトル列と 2 回目の解析より得た単語ベクトル列のコサイン類似度を算出し, 精度を確認する．解析に用いる文章を以下に示す.

(1) 今日はいい天気
(2) 今日は天気が良くない
(3) ある日の暮方のことである．一人の下人が，羅生門の下で雨やみを待っていた.

また，文章(1)と文章(2)の単語ベクトル列から，二つの文に共通して含まれる単語「今日」と「天気」の単語ベクトルを取り出しそれぞれの類似度を計る．二つの類似度から平均値を求め，二つの文章の類似度を推定する．正確な意味情報を含んだ単語ベクトルを得られているか，BERT による分散表現化の意味的精度を確認する.

5.4.1.2 結果と考察

BERT を用い, 1 回目と 2 回目の解析より得られた単語ベクトル列のコサイン類似度, 及び文章(1)と文章(2)に共通して含まれる 2 単語の平均コサイン類似度を表 5.8 に示す.

表 5.8　BERT による単語分散表現化の精度

解析文章	解析 1 回目と 2 回目の コサイン類似度	文章(1)と(2)中の 2 単語の 平均コサイン類似度
(1)	1.0	-
(2)	1.0	-
(3)	1.0	-
(1)と(2)	-	0.674

表 5.8 より，1 回目と 2 回目の解析から得られた単語ベクトル列のコサイン類似度はすべての文章において，理想値である 1.0 を得られたことがわかる．このことから，BERT は何度解析しても，毎回同じ値のベクトルを得られることがわかった．

文章(1)と文章(2)に共通して含まれる 2 単語の平均コサイン類似度は 0.674 と比較的高い値を得た．文章(1)と(2)は全く反対の意味を持つ文章だが，類似度を計った単語「今日」と「天気」は，どちらの文章でも全く同じ意味を持ち似た働きをする．そのため，高い値が得られたと考えられる．しかし，全く同じ単語同士の類似度を計ったにも関わらず類似度は 1.0 を示さなかった．そのため，BERT を用いて得られる単語ベクトルは単語の意味だけでなく，文章中の単語の役割など文脈の情報が付加されていることが本実験から確認できた．

5.4.2 BERT を用いた自動採点の評価

5.4.2.1 実験概要

BERT を用いた答案採点支援システム（以下，提案手法 2）の有効性を評価する．5.3.2 節で述べた実験と同様に，中学校の国語試験の採点を想定する．高校入試問題の中から表現工夫型の短答記述式問題を選択し，中学 3 年生が実際に解いた 40 枚の答案を収集する．

人手による採点結果として，現役の中学校国語教師が模範解答をもとに採点する．問題は 4 題あり，1 題のみ 0 から 5 点までのいずれかの得点を与え，残りの 4 題は 0 から 9 点までのいずれかの得点を与える．人手によって採点された 40 枚の答案の中から，答案の内容と得点ともに多様なものを各問題で 10 種ずつ選択する．システムにより，選択した答案と模範解答の類似度を算出する．

本実験では，答案と模範解答の類似度を算出するため，答案中の重要単語を用いる．BERT より得られる単語ベクトル列の中から，模範解答と答案に共通して含まれる重要単語の単語ベクトルを二つずつ選択する．コサイン類似度によって，対応する重要単語同士でそれぞれ類似度を算出する．二つの類似度から平均値を求め，模範解答と答案の類似度を推定する．人手による得点とシステムより推定される答案類似度の無相関検定を行い，提案手法 2 の有効性を評価する．

5.4.2.2 結果と考察

問題 1（20 字程度の文章，$n = 10$）における，人手による採点結果と手安手法 2 により推定された答案の類似度を表 5.9 に示す．また，散布図と最小二乗法により求めた近似直線で表したものを図 5.12 に示す．

表 5.9　問題 1 における人手による採点結果と答案の類似度（提案手法 2）

生徒による問題 1 の答案	人手による採点結果	重要単語 1 の類似度	重要単語 2 の類似度	平均類似度（答案類似度）
1	4	0.986	0.754	0.870
2	3	0.850	0.306	0.578
3	0	0.336	0	0.168
4	4	0.542	0.189	0.366
5	5	0.852	0.794	0.823
6	1	0.946	0.200	0.573
7	4	0.843	0.778	0.811
8	1	0.879	0.223	0.551
9	2	0.953	0	0.476
10	3	0.884	0.669	0.777

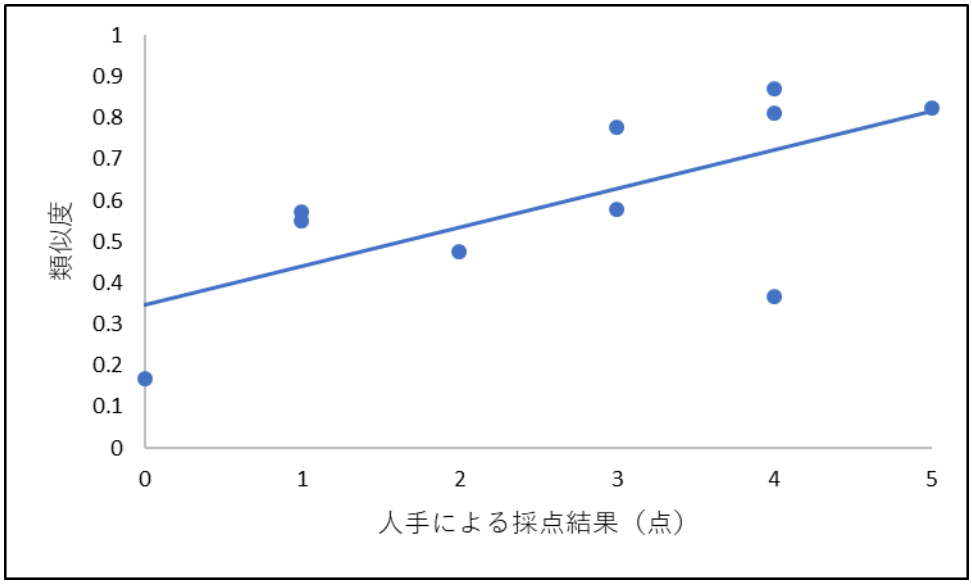

図 5.12　問題 1 における人手による採点結果と答案の類似度（提案手法 2）

問題 2（20 字程度の文章，$n = 10$）における，人手による採点結果と提案手法 2 により推定された答案の類似度を表 5.10 に示す．また，散布図と最小二乗法により求めた近似直線で表したものを図 5.13 に示す．

表 5.10　問題 2 における人手による採点結果と答案の類似度（提案手法 2）

生徒による問題 2 の答案	人手による採点結果	重要単語 1 の類似度	重要単語 2 の類似度	平均類似度（答案類似度）
1	1	0.238	0.104	0.171
2	5	0.921	0	0.461
3	9	0.900	0.993	0.947
4	2	0.877	0	0.439
5	9	0.947	0.773	0.860
6	4	0.815	0.141	0.478
7	6	0.183	0.698	0.441
8	7	0.999	0	0.500
9	0	0.176	0	0.088
10	8	0.999	0	0.500

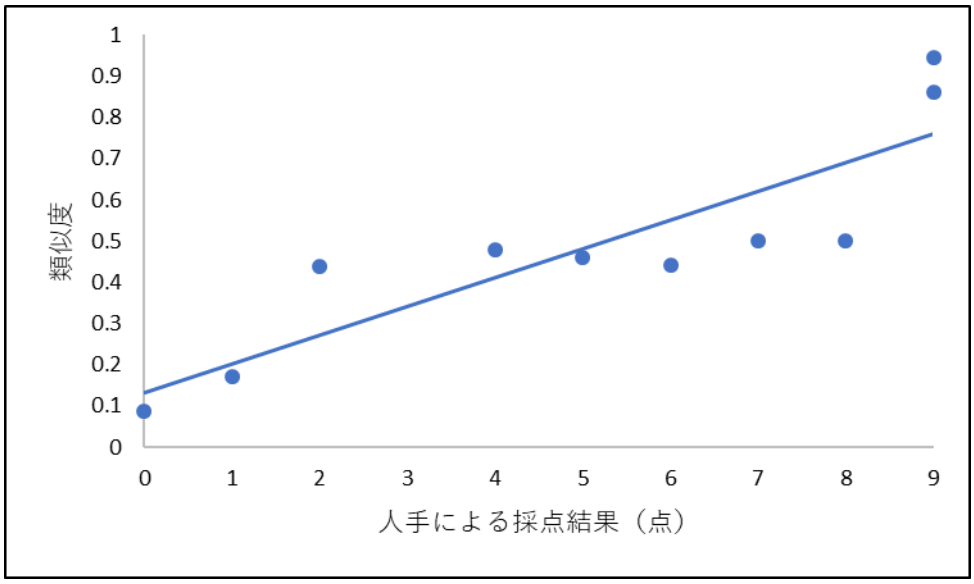

図 5.13　問題 2 における人手による採点結果と答案の類似度（提案手法 2）

問題 3（20 字程度の文章，$n = 10$）における，人手による採点結果と提案手法 2 により推定された答案の類似度を表 5.11 に示す．また，散布図と最小二乗法により求めた近似直線で表したものを図 5.14 に示す．

表 5.11　問題 3 における人手による採点結果と答案の類似度（提案手法 2）

生徒による問題 3 の答案	人手による採点結果	重要単語 1 の類似度	重要単語 2 の類似度	平均類似度（答案類似度）
1	8	0.581	0.814	0.698
2	5	0.487	0.152	0.320
3	0	0.227	0	0.114
4	4	0.264	0.736	0.500
5	3	0.739	0.736	0.738
6	2	0.520	0	0.260
7	6	0.483	0.789	0.636
8	1	0.650	0	0.325
9	9	0.758	0.908	0.833
10	7	0.271	0.843	0.557

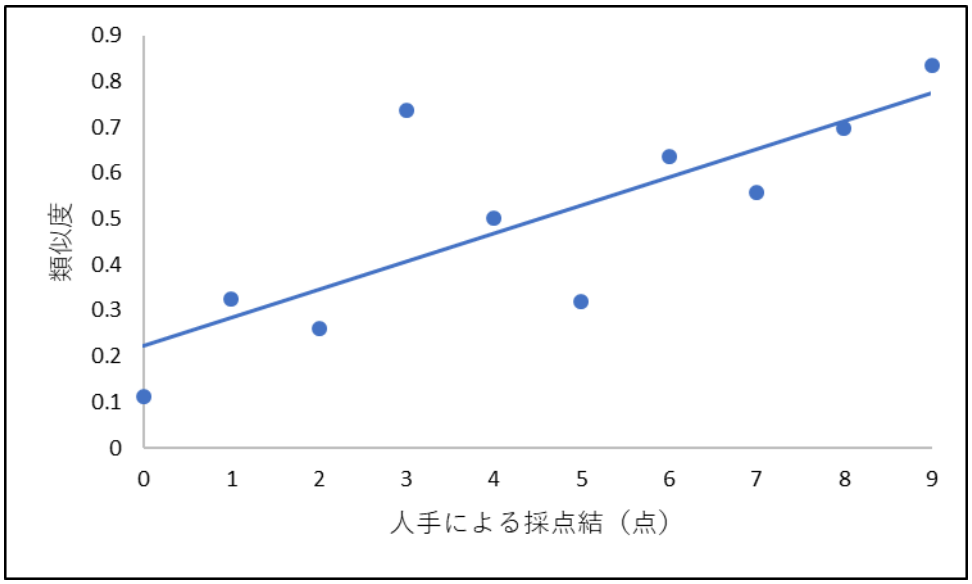

図 5.14　問題 3 における人手による採点結果と答案の類似度（提案手法 2）

問題 4（30 字程度の文章，$n = 10$）における，人手による採点結果と提案手法 2 により推定された答案の類似度を表 5.12 に示す．また，散布図と最小二乗法により求めた近似直線で表したものを図 5.15 に示す．

表 5.12　問題 4 における人手による採点結果と答案の類似度（提案手法 2）

生徒による問題 4 の答案	人手による採点結果	重要単語 1 の類似度	重要単語 2 の類似度	平均類似度（答案類似度）
1	7	0.732	0.917	0.825
2	5	0.858	0.890	0.874
3	9	0.948	0.958	0.953
4	8	0.915	0.936	0.926
5	4	0.838	0.920	0.879
6	6	0.857	0.901	0.879
7	4	0.832	0.905	0.869
8	7	0.888	0.929	0.909
9	6	0.800	0.939	0.870
10	3	0.848	0.910	0.879

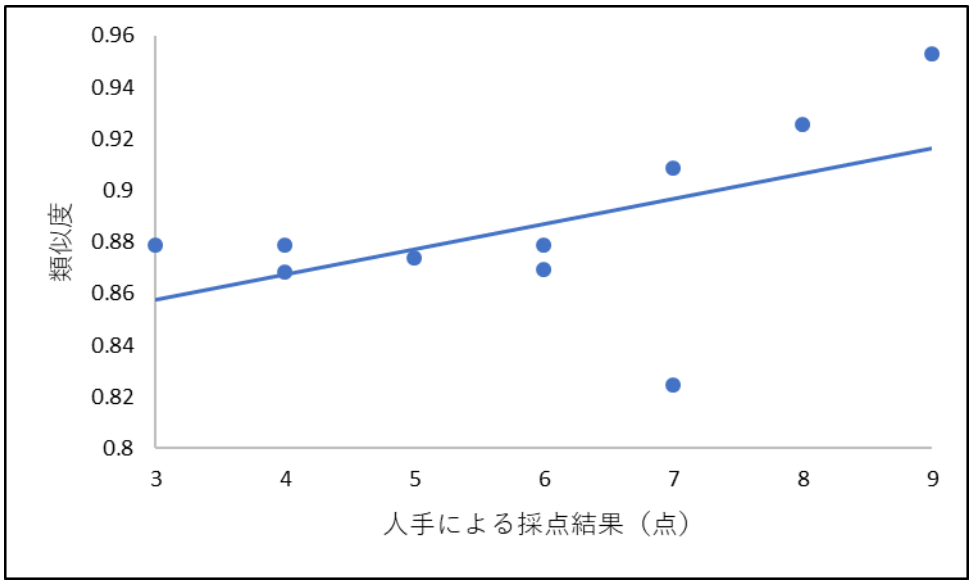

図 5.15　問題 4 における人手による採点結果と答案の類似度（提案手法 2）

表 5.9，表 5.10，表 5.11，表 5.12 より算出した，四つの問題における人手による採点結果と BERT を用いたシステムにより推定された答案類似度の相関係数, 及び無相関検定の結果を表 5.13 に示す.

表 5.13　人手による採点結果と答案類似度の相関関係（提案手法 2）

	問題 1	問題 2	問題 3	問題 4
標本数	10	10	10	10
自由度	8	8	8	8
相関係数	0.680	0.876	0.786	0.532
t値	2.624	5.125	3.601	1.777
p値	0.030	0.001	0.007	0.113

表 5.13 より，全ての問題において正の相関を得られたことがわかる．また，無相関検定の結果，問題 1, 2, 3 は$p < 0.05$で帰無仮説が棄却され，人手による採点結果と提案手法 2 による答案類似度に相関があることが示された. しかし, 問題 4 は$p \fallingdotseq 113 > 0.05$となり, 人手による採点結果とシステムによる答案類似度に相関があることを示さなかった．この原因は，二つの単語のみで答案類似度を推測したためだと考えられる．本実験に用いた問題 4 の答案はいずれも似た文章であった. すべての答案に模範解答と同じように重要単語が含まれていた．人手による採点結果は，説明不足の答案に対し低い得点を与えている．一方，表 5.12 からもわかるように，模範解答と答案の類似度は，どの答案においても高い値を得ている. 本実験では二つの単語のみで答案類似度を推定したため，説明不足の答案においても比較的高い類似度が得られたと考えられる.

本実験の結果より，提案手法 2 では，表現自由型短答記述式問題の採点に対し有効性が確認できた．しかし，重要単語を適切に含んだ似た答案に対し，区別がつかず適切な類似度を得ることができなかった．比較する単語数を増やす，または BERT より得られる単語ベクトル列を一定の次元に圧縮した文章ベクトルに変換し答案の類似度を測定することで改善できると考えられる.

5.5 まとめと考察

本実験より，Doc2Vec 及び BERT の性質，並びに提案したシステムの有効性を調べることができた.

Doc2Vec は未知語に弱く，解析文章内に未知語が含まれると適切な意味を持った文章ベクトルが得られない．そのため，実験においても同じ文章に対し，解析する度に異なるベクトルを得た．また，全く反対の意味を持つ二つの文章に対し，文章ベクトルの類似度は高い値を得た．文章中に似た単語が多く含まれていると，得られる文章ベクトルの類似度は高い値をとることが確認できた.

提案手法 1 の評価実験において，自動採点に対し有効性を示さなかった．改善策として，十分な内容と量の学習データを収集することと複数の模範解答を用いることが挙げられる．

BERT は，単語の意味だけでなく文章中の単語の役割など文脈情報を付加した単語ベクトルを得られることが実験より確認できた．しかし，全く反対の意味を持つ二つの文章に対し，文章の類似度は高い値を得た．比較した共通単語が，全く同じ意味を持ち似た働きをしていたことが原因だと考えられる．

提案手法 2 は，表現自由型短答記述式問題の採点に対して有効性を示した．しかし，模範解答と同じように重要単語を含んだ似た答案に対し，区別がつかず適切な評価を与えることができなかった．改善策として，以下の二つが挙げられる．比較する単語数を増やすこと，BERT より得られる単語ベクトル列を一定の次元に圧縮した文章ベクトルに変換し答案の類似度を算出することである．

5.6 むすび

本章では，提案手法の実験結果及び考察について述べた．MeCab と JUMAN++による形態素解析の精度比較，Doc2Vec と BERT の分散表現化の精度及び自動採点の精度評価の実験を行った．実験結果を踏まえ，考察を述べた．

第6章　結論

6.1 結論

本研究では，答案採点の自動化を図るため，日本語形態素解析器の JUMAN++，自然言語を分散表現化する Doc2Vec 及び BERT といった自然言語処理技術を用いた答案採点支援システムを検討した．

関連研究における「人工的に高得点答案を作成可能である」といった問題点を解決すべく，Doc2Vec や BERT を用い，答案中の個々の単語だけでなく文章全体の意味を捉え評価する手法を提案した．

実験により，提案したシステムの有効性を確認したが，同時に多くの問題点と課題が見つかった．Doc2Vec は自動採点において Word2Vec と同じ問題が生じた．未知語に弱く，答案中に模範解答と同じ単語が多く含まれている場合高い類似度を得た．一方 BERT は Word2Vec の問題に有効であることが分かった．しかし，重要単語を適切に含んだ似た答案に対し，区別がつかず適切な評価を与えることができないといった新たな改善点が見つかった．

6.2 今後の課題

提案した二つのシステムにおいて，多くの問題点と改善点が見つかった．Doc2Vec を用いた答案採点支援システムには二つの問題点がある．未知語に対し適切なベクトルを得ることができないことと，模範解答の言葉を他の表現に言い換えられた答案に対し低い類似度を得ることである．そのため，学習データを見直し Doc2Vec の学習モデルを改良することと複数の模範解答を用いた評価方法を検討する必要がある．一方，BERT を用いた答案採点支援システムには一つの問題点がある．重要単語を適切に含んだ似た答案に対し，区別がつかず適切な評価を与えることができないことである．そのため，比較する単語数を増やすことや，BERT より得られる単語ベクトル列から文章ベクトルを得る方法を検討する必要がある．

謝辞

本研究に際して，丁寧かつ熱心なご指導をしてくださり，実験環境および快適な研究環境を与えてくださった渡辺裕教授に心より感謝いたします.

貴重なご意見や様々なご提案をくださいました笠井裕之教授に感謝いたします.

日頃からアドバイスをくださり，研究室における温かい環境を提供してくださった渡辺研究室の皆様に感謝いたします.

最後に，私をここまで育ててくださり，支えてくださっている家族に感謝いたします.

参考文献

[1] 文部科学省，文部科学省ホームページ：“大学入学者選抜改革について”，https://www.mext.go.jp/a_menu/koutou/koudai/detail/1397731.htm，参照 2020.1.5.

[2] 山本廣基，大学入試センターホームページ：“大学入学共通テストの記述式問題導入見送りについて”，https://www.dnc.ac.jp/，参照 2020.1.5.

[3] Copper, P. L.: “The assessment of writing ability: a review of research”, GREB, No.82-15R, June.1984.

[4] 石岡恒憲，亀田雅之：“コンピュータによる小論文の自動採点システム Jess の試作”，計算機統計学，16 巻，1 号，pp.3-19，2003.

[5] 水本智也，磯部順子，関根聡，乾健太郎：“採点項目に基づく国語記述式答案の自動採点”，言語処理学会第 24 回年次大会発表論文集，pp.552-555，2018.

[6] 工藤拓，山本薫，松本裕司：“Conditional Random Fields を用いた日本語形態素解析”，情報処理学会研究報告．自然言語処理研究会報告 161，pp.89-96，2004.

[7] 京都大学情報学研究所，京都大学情報学研究所ホームページ：“MeCab: Yet Another Part-of-Speech and Morphological Analyzer”, http://taku910.github.io/mecab/, 参照 2020.1.10.

[8] 京都大学大学院情報学研究所，黒橋・河原研究室ホームページ：“日本語形態素解析システム JUMAN++”，http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++，参照 2020.1.15.

[9] 京都大学大学院情報学研究所，黒橋・河原研究室ホームページ：“日本語形態素解析システム JUMAN”，http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN，参照 2020.1.15.

[10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean: “Distributed Representations of Words and Phrases and their Compositionality”, NIPS'13, Vol.2, pp.3111-3119, 2013.

[11] Quoc V. Le, Tomas Mikolov: “Distributed Representations of Sentences and Documents”, ICML’14, Vol.32, pp.1188-1196, 2014.

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, NAACL, pp.4171-4186, 2019.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: “Attention Is All You Need”, NIPS, 2017.

[14] 青空文庫，青空文庫ホームページ：“インターネットの電子図書館青空文庫”，http://www.aozora.gr.jp/，参照 2019.9.20.

[15] 京都大学大学院情報学研究所, 黒橋・河原研究室ホームページ:”BERT 日本語 Pretrained モデル”， http://nlp.ist.i.kyoto-u.ac.jp/index.php?BERT%E6%97，参照 2019.10.15.

図一覧

表一覧