

令和元年6月26日現在

機関番号：14501

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00476

研究課題名(和文) 記述特徴に基づいた授業レポート用の盗用発見への深層学習の適用

研究課題名(英文) An Application of Deep Learning to detect Plagiarisms in Assigned Reports based on the Style Model

研究代表者

村尾 元 (Murao, Hajime)

神戸大学・国際文化学研究科・教授

研究者番号：70273761

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：本研究では、授業における課題レポートの盗用を発見に深層学習を適用を試みた。まず、あらかじめ作成者が明らかな複数のレポートから、句読点の利用法など、レポート作成者の記述上の特徴を抽出、これを深層学習の一種であるVRAEを用いて学習する。VRAEは、レポート作成者の記述上の特徴を潜在空間に写像する。新しいレポートが得られる度に、これをVRAEに入力し、潜在変数を得る。この潜在空間上における分布を調べることで、そのレポートがどの作者によるものかを推定する。実験の結果、提案手法により、レポートの記述上の特徴から作者を推定できる可能性が示された。

研究成果の学術的意義や社会的意義

本研究により、授業の課題レポートのように、同じテーマについて書かれており、内容に基づいた比較による盗用発見が困難な場合においても、盗用の可能性をコンピュータにより指摘することができるようになり、教員の負担軽減につながる。また、本手法を従来研究されてきた、内容に基づいた盗用発見と組み合わせることにより、さらに精度を高めることが可能となり、より広範な盗用発見に適用できる可能性がある。

研究成果の概要(英文)：In this study, we have applied a Variational Recurrent Auto-Encoder (VRAE) to detect plagiarisms in assigned reports. VRAE is preliminarily trained by using "expression features" of reports submitted by students. Here, "expression features" are a number of punctuations, the location of new-lines, etc. VRAE learn a map from the expression features of each students' report onto the latent space. Then, whenever a student submits a new report, we extract the expression features from it and input to the trained VRAE, and estimate the author of the report by checking where the report will map on the latent space. We test the proposed method and validate possibility.

研究分野：社会システム科学

キーワード：盗用発見 機械学習 深層学習

1. 研究開始当初の背景

コンピュータとインターネットの特に検索技術の発展とともに、学校のレポート課題におけるコピーすなわち盗用は常に問題となっている。授業の効果を上げるためには、提出されたレポートにおける盗用を発見し、適正に評価することが欠かせないが、大量のレポートを対象とした盗用発見の試みは教員への負担が大きい。したがって、適正な評価の実施と、それに関わる教員の負担軽減のためにも、コンピュータによる盗用発見の支援は必要不可欠である。

一方、社会的、経済的な影響も大きいことから、文書の盗用発見に対するコンピュータによる支援については、これまでも様々な研究がなされている。それらは大きく2つの方法に分けることができる。1つは、Webや他の文書などの外部コンテンツとの一致検出に基づく方法（太田ら、CaNexus社のEve2など）であり、もう1つは、同一文書内における文言の一貫性に基づく方法（iParadigm社のTurnitinや高橋ら、渡邊らなど）である。しかし、そのいずれもが、単語や文を比較し、その一致や特異性について調べるものである。

知的財産保護の観点から言えば、文章そのものとその伝える内容の盗用が大きな問題となるため、これらの手法が有用であることは疑いようもない。しかし、授業のレポート課題では、同一または限られた少数のテーマに基づいてレポートを作成するため、使用する単語や文は類似する場合も多い。そもそも、利用する単語や文がある程度指定されている場合すらある。また、授業で作成するレポートは、単語や文の比較からは出典を特定できないほど短いものも多い。つまり、授業で作成されたレポートに対して、既存の文書盗用発見手法の適用は困難である。

申請者らはこれまでの機械学習に関する研究の一部として、プログラミング授業におけるソースコードの盗用や課題レポートにおける盗用の発見手法を提案してきた。提案手法では、空白や改行、文の区切りの使い方や、コメントの形式といった、従来法では有効に利用されてこなかった「書き方の特徴」に着目し、この出現確率を、確率モデルの一つである隠れマルコフモデル（以下 HMM）を用いてモデル化した。ここでは、HMMの構造（状態の設定や状態遷移グラフの構造）によっては一定の有効性を示すことができた。逆に言えば、提案手法を実際の授業で利用するには、課題レポートの書式などに合わせて、事前に HMM の構造を適切に決定する必要がある。

本研究では、この問題を解決するために深層学習を利用する。深層学習は、事前に問題固有の知識なしに入力文章の特徴を学習することができる。申請者らも SNS 上のテキストから都市の特徴を抽出するためなどに利用しており、同一の学生の課題レポートから共通の書き方の特徴を抽出するために利用できると考えた。

2. 研究の目的

本研究の目的は、レポートにおける盗用発見システムの構築である。すなわち、課題として学生から提出されたレポートを対象として、盗用の可能性を推定し、これを教員に提示するシステムの試作である。これにより、教員の負担を軽減しつつ、適正な採点および授業の実施を支援する。

本研究の中心的な課題は、授業のレポート課題のように、単一もしくは非常に限定されたテーマに基づいて書かれた自然言語の文書から盗用の可能性を推定する手法の開発である。このような文章では、その内容の類似性が盗用の根拠とならず、また、レポートが短いために文章やその内容上の類似性を発見することが困難となる。このような場合にも利用できるような盗用発見手法の開発を行う。

本研究ではこの目的のために、句読点や空白、改行の使用法といった、レポート作成者の記述時の“くせ”とも言えるような書き方の特徴を利用して、作成者の書き方のモデルを作成する。疑わしいレポートが、実際にその作成者によって書かれたものか、それとも盗用であるかは、このモデルへの合致度合いに基づいて判定する。

また、このアルゴリズムの実用性を検討するため、授業における課題レポートの盗用可能性を提示するシステムを試作し、このシステムを申請者が担当している実際の授業で利用し、評価する。

3. 研究の方法

3.1 概要

本研究では、まず、(1)学習用の文章から作成者の書き方の特徴を抽出する。これを文章の規格化と呼ぶ。この規格化された文章を用いて作成者の書き方のモデルを作成する。続いて、(2)作成した書き方モデルに、テスト用の文章を規格化したものを入力してその作者を推定、実際の作者であるかどうかを調べる。

3.2 文章の規格化

まず、形態素解析器を用いて、文章を形態素に分割する。続いて、句読点および空白、改行、括弧といった記号以外の形態素を、対応する品詞を表す記号に置き換える。これにより、例えば「今日は、とても天気がよく、気持ちが良いです。」という文章は「<名><助>、<副><名><助><形>、<名><助><形><助動>。」というように書き方の特徴のみを残した文章に変換される。

3.3 Word2Vec

本研究では規格化された文章から作者の書き方モデルを作成するために深層学習を用いる。深層学習で学習するために、先に規格化した文章を、Word2Vecを用いて数値の列に変換する。Word2Vecは類似したコンテキストで利用される単語を、距離の近いベクトルで表現する。

3.4 深層学習 VRAE

本研究で提案する盗用発見手法では、まず、書き方の特徴からレポート作成者の書き方のモデルを作成する。そして、新たなレポートが得られる度に、そのレポートの作成者とされる人物のモデルと照らし合わせ、レポートが実際にモデルの対象となっている作成者によって書かれたものであるかどうかを判定する。

書き方モデルでは、例えば、特定の位置に空白を幾つ入れるか、句読点の比率はどれくらいか、改行の頻度はどれくらいか、といった記述上の特徴をモデル化する。この目的では、幾つかの確率モデルを利用してきたが、本研究では、深層学習、中でも、時系列データの特徴抽出能力に優れた VRAE (Variational Recurrent Auto-Encoder) を利用する。

VRAEはAE (Auto-Encoder) と呼ばれる深層学習の一種である。図1に示すように、AEは、構造的には符号化ネットワークと復号化ネットワークという2つのニューラルネットワークを、少数のニューロンで構成される潜在層で接続したものである。これを、入力信号を同時に教師信号とする、すなわち、入力されるデータをそのまま出力するように学習する。これにより、データセットに含まれる共通の情報は2つのニューラルネットワークの結合荷重として獲得され、潜在層には、データセットを復号化するために必要な特徴量が得られる。この特徴量を潜在変数と呼ぶ。

VRAEは次の二点についてAEを拡張したものである。まず、潜在変数の確率分布を得られるように、潜在層の出力をガウス分布としている。また、時系列データを扱うように、符号化ネットワークと復号化ネットワークに再帰型ニューラルネットワーク (Recurrent Neural Network) を用いている。

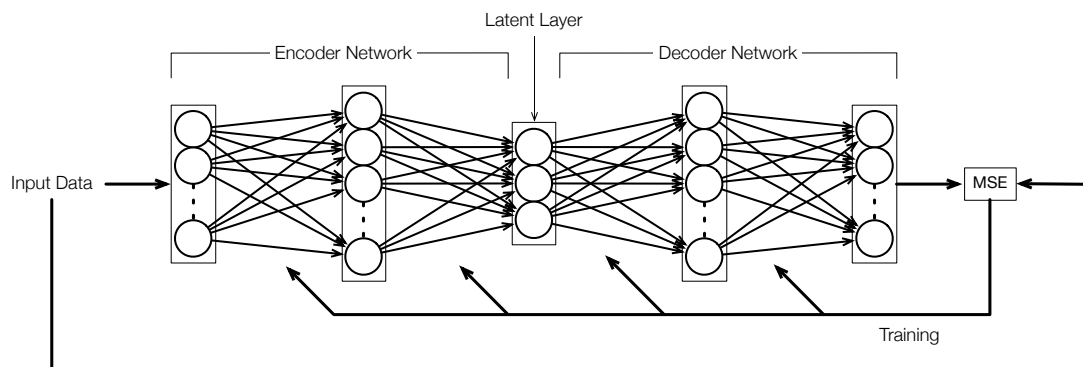


図 1 Auto-Encoder の構造

3.5 盗用発見システムの構築

盗用発見システムのプロトタイプを Web アプリケーションとして構築する。学習済みの VRAE を組み込み、Web 経由で提出されたレポートについて、盗用の可能性に関する情報を確認することができる。すなわち、学生 A が提出したレポートが、どの学生の書き方モデルと合致するか、学生 A の書き方モデルとどの程度合致するのかが提示される。具体的には、合致度は潜在変数空間上における、距離で示される。

4. 研究成果

4.1 VRAE による学習

文章からの書き方モデルの生成に先立って、まず、VRAE の学習パラメータを変更しながら学習を繰り返し、適切なパラメータの推定を行った。実験では、学生 7 人分、1 人 5 件程度、

合計 40 件の課題レポートのデータを利用した。8 割にあたる 32 件のランダムに選んだデータを学習用、残りの 8 件分のデータをテスト用とした。これに対して、中間層のニューロン数 (パラメータ H) と、中間層の数 (パラメータ L), 潜在変数の数 (パラメータ Z) および学習係数 (パラメータ B) を変えながら学習を行った。

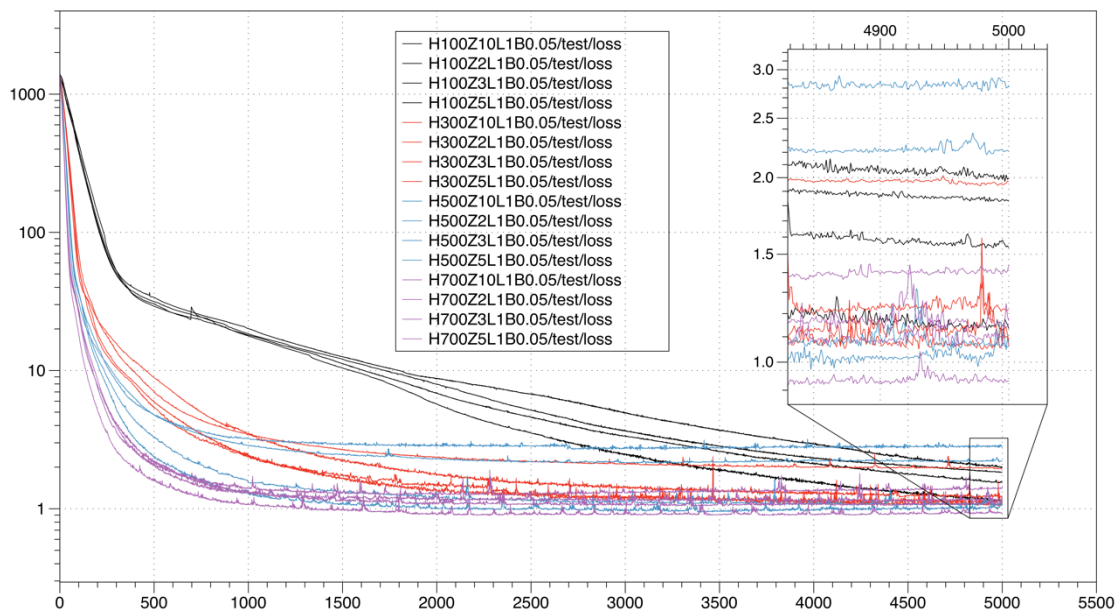


図 3 パラメータ H と Z を変更しながら行った実験結果

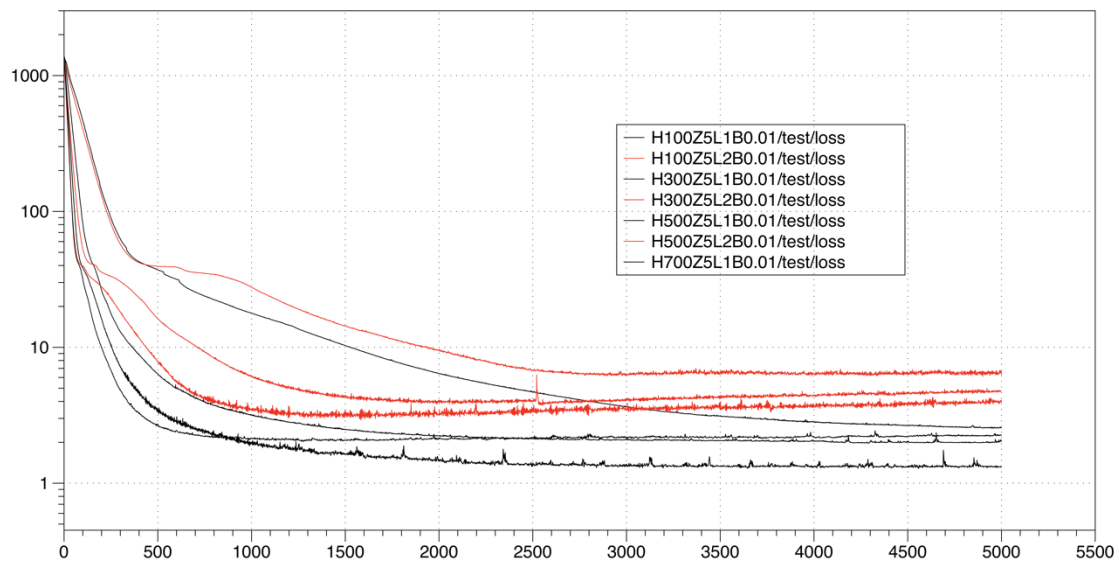


図 2 パラメータ L を変更しながら行った実験結果

図 2 に中間層ニューロン数 H および潜在変数の数 Z を、図 3 に中間層数 L を変えながら行った実験の結果を示す。図では、5,000 エポック学習させた際に、各エポックの学習後にテストデータを入力した際の誤差をプロットした。この値が小さくなれば比較的学習がうまくいっていると言える。図 2 より、中間層ニューロン数については、 $H=300$ 以上であればさほど学習結果に差がないこと、潜在変数の数 Z は影響が少ないことが分かる。また図 3 より、中間層の数 L については、2 層 ($L=2$) よりも明らかに 1 層 ($L=1$) の方が成績が良い。ここには示していないが、他の実験データより、潜在変数の数 Z は 3 以上であれば差がないこと、学習係数 B については 0.01 または 0.1 程度が良いが大きく学習性能に差はないことが分かっている。

4.2 潜在変数の抽出

前節の結果より、中間層ニューロン数 $H=700$ 、中間層数 $L=1$ 、潜在変数の数 $Z=5$ 、学習係数 $B=0.01$ として学習後の VRAE に学習済みのデータを入力し、潜在変数の分布を調べた。 $Z=5$ 、すなわち 5 次元の潜在変数を、主成分分析を用いて 2 次元に変換し、プロットした様子を図 4 に示す。図では、学習に用いた学生を A1~C3 とラベル付けし、マーカと色を分けて示した。

図より、レポートの作成者ごとに特徴が分布してプロットされており、潜在変数空間上で、書き方の特徴を用いて、それぞれのレポート作成者の書き方がモデル化できていることが分か

る。

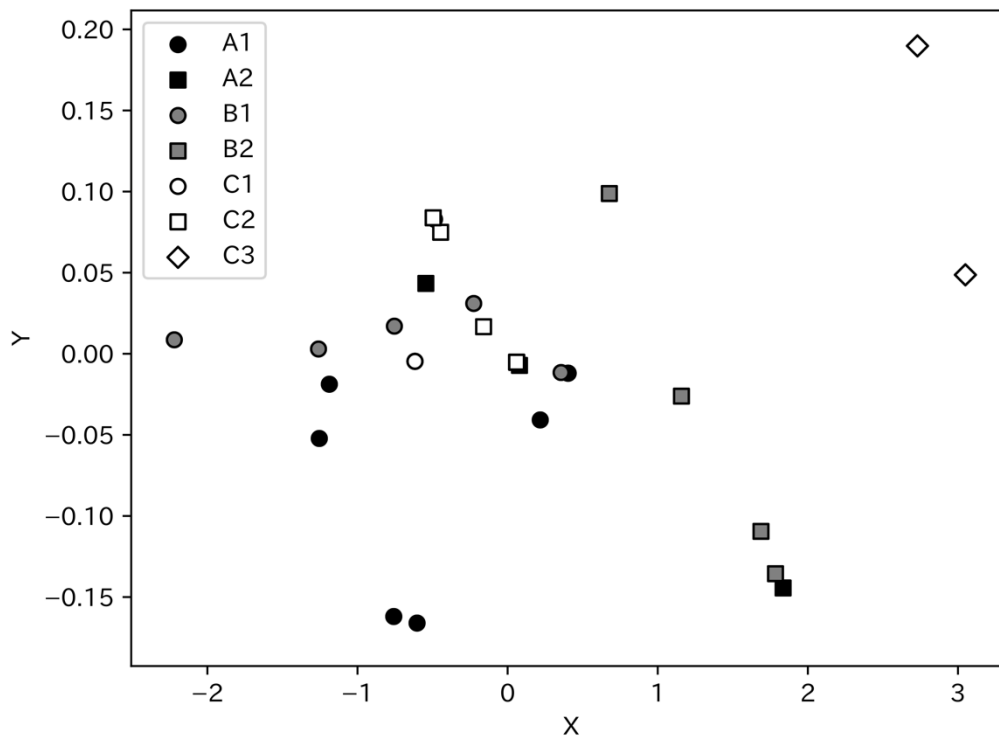


図 4 潜在変数の様子

4.3 盗用発見システムのテスト

学習済みの VRAE を組み込んだ盗用発見システムのプロトタイプを Web アプリケーションとして構築し、テストを行った。Web ブラウザから、アップロードされたレポートについて、文書の正規化、Word2Vec によるベクトルへの変換を行い、VRAE を用いて潜在変数に変換、すでに学習済みの書き方モデルとの比較を行った。潜在空間における距離の計算に先立って、可視化して確認したところ、学習済みの書き方モデルとの比較によって、レポートの作成者が推定できる可能性が示された。ただし、有意性については、今後、複数名による実験とアンケート等による分析が必要だと考えられる。

5. 関連する発表論文等

〔雑誌論文〕 (計 2 件)

(1) Fan Zhang and Hajime Murao. Analysis of reading experience with heart rate variability and galvanic skin response sensor. ICIC Express Letters, Part B: Applications, 10(2):129-136, 2 2019.

(2) Fan Zhang and Hajime Murao. Advantages and disadvantages of biofeedback on playing puzzle games. ICIC Express Letters, Part B: Applications, 8(2):263-268, 2 2017.

〔学会発表〕 (計 4 件)

(1) Fan Zhang and Hajime Murao. Analysis of reading experience with heart rate variability and galvanic skin response sensor. In The 13th International Conference on Innovative Computing, Information and Control (ICICIC2018), 8 2018.

(2) Megumi Kawata and Hajime Murao. Study on kinect-based sonification system for blind spot warning. In IEEE International Conference on Information, Communication and Engineering (ICICE 2017), 11 2017.

(3) Hajime Murao and Kazuaki Nonohara. A study on finding key movements to discriminate expert and novice by using a kinect-like 3d sensor. In IEEE International Conference on

Information, Communication and Engineering (ICICE 2017), 11 2017.

(4) Fan Zhang and Hajime Murao. A preliminary study to develop a lightweight biofeedback device for reading activity. In IEEE International Conference on Information, Communication and Engineering (ICICE 2017), 11 2017.

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計 0 件）

○取得状況（計 0 件）

〔その他〕

ホームページ等

6. 研究組織

(1)研究分担者

なし

(2)研究協力者

なし

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。