

令和 3 年 6 月 28 日現在

機関番号：31304

研究種目：若手研究

研究期間：2018～2020

課題番号：18K13240

研究課題名（和文）敵対生成ネットワークに着眼した小論文・レポート自動評価の研究

研究課題名（英文）Study on Automatic Review of Essay Based on Generative Adversarial Networks

研究代表者

岩田 一樹 (Iwata, Kazuki)

東北福祉大学・総合マネジメント学部・講師

研究者番号：20515457

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究では、小論文やレポートの自動評価を目標に、その基礎研究として、人の手を返さない教師なし学習の一つである敵対生成ネットワークに着目し、それから得られる識別器による文書分類タスクを対象に研究を行った。その結果、敵対学習から得られる識別器によって文書分類が可能であること、および、識別器と敵対し、文を生成する生成器の性能が識別器の識別性能に大きく寄与することが明らかになった。今後はこの知見、および、新たなアルゴリズムも加えてさらに高い精度での分類を目指していく。

研究成果の学術的意義や社会的意義

本研究は、近年、小論文やレポートによる評価が多様化される中で、機械学習の手法を通してそれらの自動評価システムの構築を目指したものである。研究期間を通して、教師なし学習である敵対生成ネットワークによる識別器の作成が可能であることを見出すことができた。大量のデータを扱うことが可能な教師なし学習による自動評価の構築可能性はビッグデータの利用によって更に性能を向上させることによって、採点者の負担軽減、ならびに、執筆者の学習機会の増加に繋がるものと考えられる。

研究成果の概要（英文）：In this study, we focused on adversarial generation networks, one of the unsupervised learning methods without human intervention, as basic research for the automatic evaluation of essays and reports. In addition, we researched document classification tasks using discriminators obtained from adversarial networks. As a result, we found that the performance of discriminators obtained from adversarial learning can classify documents depends on the generator's performance.

In the future, we will add this knowledge and a new algorithm to achieve even higher accuracy in classification.

研究分野：教育工学

キーワード：機械学習 自然言語処理 敵対生成ネットワーク 自動化

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

敵対生成ネットワーク (Generative Adversarial Network : GAN) は 2014 年に I. Goodfellow が考案した生成モデルである[1]。この生成モデルは、訓練データのみを与える「教師なし学習」の一種であり、訓練データの特徴を反映したデータを機械に生成させるもので、画像データを対象として提案され、それまで不可能だった鮮明な画像の自動生成に成功した。

GAN では、生成器と識別器の 2 つの機構を敵対させ、それら 2 つを競わせながら学習を進める。具体的には、生成器は与えられた訓練データを元にそれらの特徴を有するデータの生成を目指して学習を進める。一方、識別器には生成器が自動生成したデータと人が生成した訓練データを同時に与え、与えられたデータが生成器による生成データか、訓練データかを正確に鑑別できるように学習を進める。その結果、最終的に、生成器は訓練データの特徴を捉えた訓練データらしいデータを生成し、識別器は生成データがどのくらい訓練データらしいかを高精度で評価できるようになる。このモデルは画像のみでなく、文章や音楽などの系列データにも適用されるようになり、現在も盛んに研究されている生成モデルの 1 つである。

一方、近年の大学教育においてはアクティブ・ラーニングが重要視されるようになり、小論文やレポートを評価課題とすることが多くなった。そして、それらの課題を評価するにあたって、採点者の時間的負担の軽減などを目的に、自動評価システムの開発が行われた[2]。これらの自動評価システムは、提出された文章を入力として、その評価結果の出力をタスクとする。これらシステムは、既存の「文章」と人が行ったその文章の「評価」の組を訓練データとして与え、その訓練データを元に機械が学習し、未知のデータ (レポートや小論文) の評価に適用したり、採点者がヒューリスティックに重要と判断する基準を特徴量として、それを基に自動評価したり、または、これら 2 つの評価を組み合わせたといったアプローチが取られている。訓練データとして「入力データ」とその「正解」の組で与える学習は「教師あり学習」と呼ばれ、機械学習において重要な 1 分野を築いている。しかしながら、教師あり学習においては、訓練データを作成する際に 大量の文章を人力によって評価する必要がある点、 人力で評価を行うため、訓練データの「正解」に評価ゆらぎが包含される可能性がある。特に、学習に数十万～数百万、または、それ以上の数の訓練データを学習に利用するようになった今日、 の問題は大きな課題である。一方、ヒューリスティックに特徴量を設定する場合、その根拠が経験的なものであるため、その特徴が正しく評価の基準を捉えているかを確認するのが困難であると同時に、経験の共有が難しいために定義が困難な論理性的の評価は不可能といえるのが現状である。

それに対し、訓練データに「正解」を与えない機械学習のことを「教師なし学習」と呼び、教師なし学習においては、訓練データは「入力データ」のみで、その「正解」にあたるものを有さない。したがって、訓練データに対する人力による評価が不要であるため、教師あり学習よりも大きな訓練データを比較的容易に取得することが可能となり、上記 の問題を克服でき、人が評価を行わないので の問題は生じないメリットを有している。

### 2. 研究の目的

この様な研究背景に対し、教師なし学習である GAN を利用することで上記の 2 つの問題を解決し、また同時に、従来と異なる機械のみの判断をベースにした小論文やレポートの自動評価・添削の実現化を目指し、GAN をベースとした小論文やレポートの自動評価の研究・開発が本研究の目的である。この開発によって、採点側は、評価の不公平性を低減できると共に時間的負担の軽減が可能となり、その分を他の教育や研究に当てることが可能になると期待できる。また、一

方、評価される側は自動で評価を受けられるようになることで、自らが執筆した小論文やレポートに対するフィードバックを受ける学習機会が増加すると考えられる。この目的の達成に向け、

- (1) 生成器のニューラルネットワーク構造の検討
- (2) 日本語の文書を対象とする GAN の検討
- (3) GAN による学習した識別器の性能評価

の 3 つの項目を研究期間中に行った。

### 3. 研究の方法

#### (1) 生成器のニューラルネットワーク構造の検討 [ 3 ]

GAN は教師あり学習が内包する問題を回避できる利点がある一方で、学習が不安定である。すなわち、Generator と Discriminator を敵対させながら相互に強化されていく中で、片方が先に強くなり過ぎると、学習が進まなくなってしまう。加えて、文章生成を行う場合は、画像と異なり、“文法”(系列)を踏まえる必要がある点も課題となる。そのため、小論文やレポートらしい文章を生成するのに適している Long Short-Term Memory (LSTM) のネットワーク構造の検討を行った。

#### (2) 日本語の文書を対象とする GAN の検討 [ 4 ]

文書を生成する GAN には、SeqGAN、TextGAN、LeakGAN など多くの種類がある。GAN による文書生成の研究対象は英語や中国語であり、日本語を対象とした研究は少ない。他の言語よりも単語の語順の自由度が高い日本語において文書生成を行うに当たり、適切な GAN を検討するために、いくつかの GAN および Maximum Likelihood Estimation (MLE) を用いて文書生成を行い評価した。なお、評価指標としては Negative Log-Likelihood Loss (NLL)、BLEU、Embedding Similarity (EmbSim) を用いた。なお、NLL は 0 に近いほど、BLUE は 1 に近いほど、EmbSim は 0 に近いほど、性能が高い。

#### (3) GAN による学習した識別器の性能評価 [ 5 ]

GAN の訓練によって得られた識別器による文書識別性能の検討を行った。具体的には、夏目漱石の文を学習データに用い、GAN により学習させた後に、学習データに用いた以外の夏目漱石の文、太宰治の文、芥川龍之介の文の 3 種類を用いて、夏目漱石の文とそれ以外を分離できるのが実験を行った。この実験の意図は、レポートの評価を“レポートらしさ”で行うことを想定し、レポートの文中でそれらしいと識別される文の割合が反映すると推察しているからである。つまり、夏目漱石らしい文章を、文中で夏目漱石の文と識別される文の数で評価するために、1 文が夏目漱石らしいかを機械が識別できるかを検討するということである。

### 4. 研究成果

#### (1) 生成器のニューラルネットワーク構造の検討

図 1 はいくつかのニューラルネットワーク構造における学習曲線を示す。横軸が epoch 数、縦軸が損失値であり、パネル右に示している「x」は LSTM 層の数を示している。すべての構造において 200 epochs 程度までは損失値が減少して学習が進んでいる。なお、機械学習においては、この損失値を 0、または、最小化する方向に学習を進めていく。しかし、200 epochs 以降については、いくつかの構造において損失値が減少せずに増加してしまい学習が破綻してしまっているものや、0 に収束せずに学習が進まなくなっているものがある。なお、ネットワー

ク構造の詳細については[3]を参照のこと。訓練用のデータは夏目漱石著『坊ちゃん』を用いた。

この研究からは Embedded Layer と呼ばれる LSTM 層と入力層の間に配置する全結合層を含んだ構造において学習が安定しやすいことが見出された。この結果は文書分類などのタスクにおいて、Embedded Layer を含んだ構造の方が性能は高いことが知られており、妥当といえる。また、同時に、LSTM 層の数が少ない方が、学習が安定しやすいことも見出せた。

この様な日本語の文書を対象として、ネットワーク構造を系統的に扱った報告は少なく、見出された学習の困難さは語順の自由度が高い日本語において旧来のネットワーク構造による文書生成が難しい可能性を示唆している。

(1)で得られた知見を元にして、後の研究においては Embedded Layer を含む 1 層の LSTM 構造を生成器として研究を実施した。

## (2) 日本語の文書を対象とする GAN の検討

図 2 は、教師あり学習の文書生成モデルであり GAN の生成器と同じ構造を有した MLE、SeqGAN、および、LeakGAN 上段から NLL、BLUE、EmbSim を学習回数に対してプロットしたものである。縦に引かれた破線はプレトレーニングと呼ばれる学習期間であり、生成器と識別器の学習の進み具合を調整するために、この間は生成器のみを学習させる。その後、100 epochs は生成器と識別器による敵対学習を実施した。3 つの指標は、NLL において、敵対学習開始から SeqGAN の NLL が敵対学習時にやや増加したが、その他の指標に大きな違いは見出せなかった。表 1~3 は MLE、SeqGAN、LeakGAN において生成された文書である。文法に正しくない箇所もあるが、ある意味の通じる程度の文章が生成できていることがわかる。

日本語における GAN による文書生成は俳句などに限られており、通常の文書を扱った研究は少ないことから、日本語においても GAN による文章生成が可能であることを確認できたことは成果に値する。

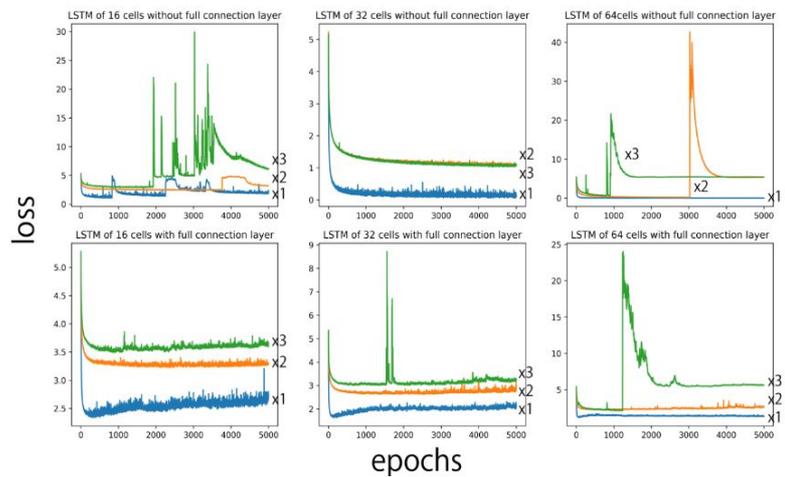


図 1 学習曲線：横軸は学習数、縦軸が損失値である。

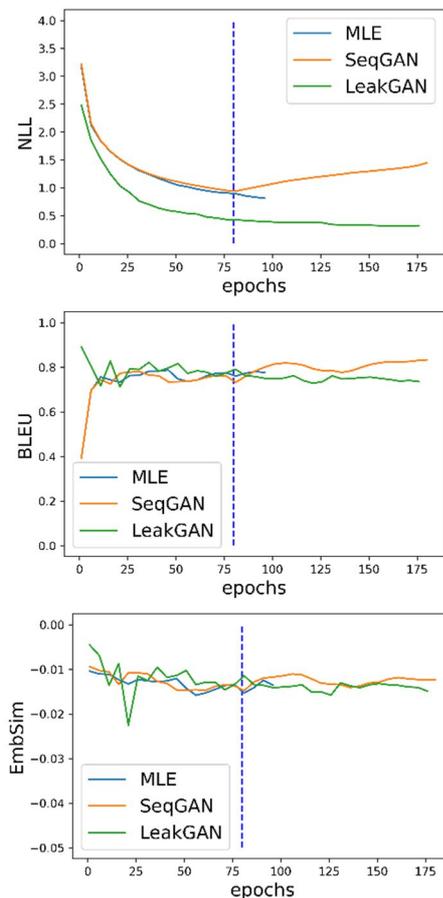


図 2 上段から NLL、BLUE、EmbSim の学習回数依存性

### (3) GAN による学習した識別器の性能評価

(2)から少なくとも、SeqGAN および LeakGAN を用いることにより、日本語文書の生成が可能であることが確認できたことから、より構造がシンプルな SeqGAN を使って GAN によって得られる識別器による文書分類を試みた。当初、この実験はレポートや小論文を訓練データに用い、それら以外のカテゴリに属する文を分類できるかを調べる予定であったが、コロナウイルス感染症拡大のため、実験に十分な訓練データを取得することができなかったことから、より、分類の困難な著者の分類が可能の実験に変更した。実験では、夏目漱石の著書に包含されている文を訓練データとして敵対学習を行い、それに得られた識別器に対して、訓練データ以外の夏目漱石の著書に包含されている文と、芥川龍之介および太宰治の著書に包含されている文を与え、筆者を漱石とそれ以外に分類可能かを検討した。

その実験から得られた結果から正解率、適合率、再現率、F 値を算出したものを表 4 に示す。また、比較のために識別器と同じ構造の TextCNN を教師あり学習で学習させた結果も同時に記載した。なお、評価時、それぞれに与えたデータは同じである。結果をみると、敵対学習から得られた識別器の性能が良くないことがわかる。この結果を受け、生成器の性能がより向上しやすいよう学習を行ったところ、識別器の性能が向上したことから、識別器の性能が出ない一因に生成器が生成する文書の未熟さがあることがわかった。したがって、生成器の性能を向上させ、より訓練データに近い文書を生成可能となれば、識別器の性能も向上すると考えられる。

表 1 MLE によって生成された文書例

1	足を覗き込んだ。
2	いよいようらなり面倒だろう。

表 2 SeqGAN によって生成された文書例

1	温泉の月給が祝詞を相手に十五灯ほど、って出来た。
2	その外によっぽど地に理窟が一本を据えた。

表 3 LeakGAN によって生成された文書例

1	おれが蜜柑の事を考えている。
2	赤シャツはおれに一銭五厘奮発させる気だから黙っていた。

表 4

	TextCNN (教師あり)	TextCNN(GAN)
正解率	0.735	0.530
適合率	0.638	0.561
再現率	0.870	0.310
F 値	0.702	0.339

#### < 引用文献 >

- [1] “Generative Adversarial Nets”, Ian J. Goodfellow et.al., NIPS 2014, 2014.
- [2] 石岡恒憲, 人工知能学会誌, Vol.23, pp.17-24, 2008.
- [3] 岩田一樹, 感性福祉研究所年報, Vol.20, pp.133-140, 2018.
- [4] 岩田一樹, 感性福祉研究所年報, Vol.21, pp.53-64, 2020.
- [5] 岩田一樹, 感性福祉研究所年報, Vol.22, pp.19-32, 2021.

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 岩田一樹	4. 巻 21
2. 論文標題 敵対生成ネットワークによる文書生成	5. 発行年 2020年
3. 雑誌名 感性福祉研究所年報	6. 最初と最後の頁 53-64
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 岩田一樹	4. 巻 20
2. 論文標題 文書生成におけるネットワーク構造依存性	5. 発行年 2018年
3. 雑誌名 感性福祉研究所年報	6. 最初と最後の頁 133-140
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 岩田一樹	4. 巻 22
2. 論文標題 敵対生成ネットワークによる文書分類	5. 発行年 2021年
3. 雑誌名 感性福祉研究所年報	6. 最初と最後の頁 19-32
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計0件

〔図書〕 計1件

1. 著者名 Kazuki Iwata, Makoto Ohuchi & Junichi Urushiyama	4. 発行年 2018年
2. 出版社 Laurea Publications	5. 総ページ数 109
3. 書名 NEW WAYS OF PROMOTING MENTAL WELL-BEING AND COGNITIVE FUNCTIONS	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------