# ペルソナベクトルの演算に応じた新たな個性での対話応答文生成

杉本　翔† 石橋　陽一† 宮森　恒††

† 京都産業大学 コンピュータ理工学部 インテリジェントシステム学科
〒 603–8555 京都府京都市北区上賀茂本山
†† 京都産業大学大学院　〒 603–8555 京都府京都市北区上賀茂本山
E-mail: †{g1444693,g1445539,miya}@cc.kyoto-su.ac.jp

**あらまし**　対話システムが生成する応答文に特定の個性を付与することで，その個性に沿った一貫性のある内容の応答が可能となり，ユーザとより親密な人間味あふれる対話が実現できると考えられる．こうした恩恵を受けるためには，対話システムの開発者が意図した新たな個性をシステム応答に自由自在に付与できるようになることが望ましいが，現時点でそのような仕組みは実現されていない．本稿では，特定の個性を表現する「ペルソナベクトル」の演算により，開発者が意図した新たな個性での応答文を生成する手法を提案する．応答文はエンコーダ・デコーダモデルで生成され，ペルソナベクトルの学習に敵対的生成ネットワークの枠組みを導入する．各人物のペルソナベクトルを生成し応答文を出力するネットワークと，その人物による実際の応答文であるかどうかを識別するネットワークを同時に学習することで，個性の付与が可能な対話システムを構築する．実験では，ペルソナベクトルの演算の有無により，意図した個性を反映した応答文がどの程度生成できるかについて検証する．

**キーワード**　自然言語処理, 対話システム, ニューラルネットワーク, 敵対的生成ネットワーク, 個性

## 1. はじめに

近年，言葉でのコミュニケーションを行う対話システムが普及し始めている．我々の身の回りにも，スマートフォンに搭載された「Siri」や「Google Assistant」，AI スピーカーやカーナビゲーションシステムに搭載されるなど，日常的に利用可能な対話システムが様々な形で提供されている．

対話システムにおける重要な要素として,「個性」が挙げられる．対話システムが生成する応答文に特定の個性を付与することで，その個性に沿った一貫性のある内容の応答が可能となり，ユーザとより親密で人間味あふれる対話が実現できると考えられる．こうした恩恵を受けるためには，対話システムの開発者が意図した新たな個性をシステム応答に自由自在に付与できるようになることが望ましいが，現段階でそのような仕組みは実現されていない．

そこで本稿では，特定の個性を表現するペルソナベクトルの演算により，開発者が意図した新たな個性での応答文を生成する手法を提案する．なお，ここでの「個性」とは口調や応答文の内容を指し，ある発話文に対する応答文は付与されるペルソナベクトルによって変化する．また，個性の付与が可能な対話システムを構築するために，ニューラルネットワークを用いた新たな対話モデルを提案する．画像生成の分野で近年注目されている敵対的生成ネットワークの枠組みを導入し，各人物のペルソナベクトルを生成し応答文を出力するネットワークと，学習データ内にある，その人物による実際の応答文であるかどうかを識別するネットワークを同時に学習する．これにより，より精度の高いペルソナベクトルが生成されることが期待される．

評価実験では，提案する対話モデルの学習によって得られたペルソナベクトルを用いて，対話応答文に対して個性の反映が可能であるかを確認する．さらに，ペルソナベクトルの演算により新たな個性を生成し，その個性が応答文にどの程度反映されるか確認する．

## 2. 関連研究

### 2.1 対話システム

雑談自体を目的とした非タスク指向型対話システムの応答文生成には，代表的な手法として用例ベース [1–3] と生成ベース [4] [5] の 2 つが存在する．用例ベースの手法では，実際に人が対話の中で使用した発話・応答ペアの集合で構成される対話コーパスの中から，入力された発話文に最も適した応答文が選択される．そのため，応答文が文法的に破綻することは基本的には起こらないものの，必ずしも入力発話に適した応答文が対話コーパスに存在するとは限らず，応答文の自由度に制限がある．これに対して生成ベースでは，テンプレートや生成モデルから応答文そのものを生成するため，生成できる文の自由度は高い．現在，タスク指向型対話システムにおいて広く用いられているのはテンプレートに基づく手法であるが，テンプレートを全て人手で作成するコストが非常に大きいため，非タスク指向型対話システムへの適用は容易ではない．一方，生成モデルを用いた手法では，対話コーパスから学習した統計モデルやニューラルネットワークを用いて応答文を生成する．特に，近年の深層学習を用いた自然言語処理の発展に伴い，ニューラルネットワークを用いた生成モデル構築への期待は大きいと考えられる．

本稿では個性を付与した様々な応答文を生成するために，生成ベースでの対話システムを実装する．生成ベースの対話シス

テムにおいて用いられる代表的な手法として，Sutskever らによって提案された Sequence to Sequence モデル (seq2seq) [6] が挙げられる．文章を単語系列として入出力するモデルで，対話システムや翻訳タスクなど幅広く用いられている．seq2seq では，可変長の入力単語系列 $\mathbf{x} = \{x_1, \cdots, x_{T_x}\}$ が与えられると，可変長の出力単語系列 $\mathbf{y} = \{y_1, \cdots, y_{T_y}\}$ を返す．例えば，図 1 のように A，B，C という単語系列が入力されると，W，X，Y，Z という単語系列が順に出力される．

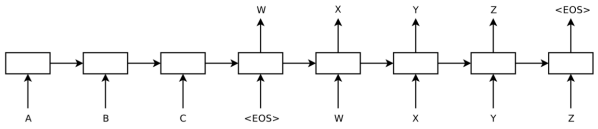


図 1　seq2seq モデル [6]

具体的には，入力単語系列から固定長の中間ベクトル $v$ を生成する Encoder と，そこから可変長の単語列を出力する Decoder で構成され，それぞれ再帰的ニューラルネットワーク (Recurrent Neural Network; RNN) [7] [8] の仕組みが応用された長・短期記憶ユニット (Long Short-Term Memory; LSTM) [9] [10] で実装される．時刻 $t$ において単語 $x_t$ が入力されるたびに式 1 により隠れ層 $h_t$ の状態を更新し，式 2 により単語 $y_t$ を出力する．ここで，$W^{\mathrm{hx}}$ は入力層から隠れ層への重み行列，$W^{\mathrm{hh}}$ は隠れ層から隠れ層への重み行列，$W^{\mathrm{yh}}$ は隠れ層から出力層への重み行列を示す．

$$h_t = \mathrm{sigm}(W^{\mathrm{hx}} x_t + W^{\mathrm{hh}} h_{t-1}) \tag{1}$$

$$y_t = W^{\mathrm{yh}} h_t \tag{2}$$

さらに，LSTM では，直前の隠れ層で与えられる中間ベクトル $v$ と，それまでに出力された単語系列 $y_1, y_2, \cdots, y_{t-1}$ が与えられた際の次の出力単語 $y_t$ の条件確率 $p(y_t|v, y_1, y_2, \cdots, y_{t-1})$ を用いて，式 3 に従って，入力単語系列 $x$ から出力単語系列 $y$ の条件付き確率 $p(y|x)$ を算出する．

$$p(y|x) = \prod_{t=1}^{T_y} p(y_t|v, y_1, y_2, \cdots, y_{t-1}) \tag{3}$$

### 2.2　対話システムの出力応答文への個性付与

Li ら [11] は，先述した seq2seq を応用し，特定の個性を表現するベクトルを元に応答文を制御する Speaker Model を提案している．Speaker Model では，seq2seq での Decoder において，単語を出力するたびに応答者の個性を表現するベクトルを入力として与える．大量の発話者情報付きの訓練データが必要である点や，気分や感情などは考慮できていない点など課題は残されているものの，出力応答文への個性付与が実現されている．モデルは，式 4 の条件付き確率が最大になるように学習する．ここで，$uid$ は応答文の発話者に割り当てられた ID，$P_{uid}$ はその発話者の個性を表現するベクトルを示す．

$$p(y|x) = \prod_{t=1}^{T_y} p(y_t|v, y_1, y_2, \cdots, y_{t-1}, P_{uid}) \tag{4}$$

対話システムに個性を付与する試みは，他にも様々な手法で行われている [12] [13]．しかし，これらの研究では個性の学習に着目しており，新たな個性は生成していない点が本稿との違いである．また，濱田らの研究 [14] において，Speaker Model で用いられる個性を表現するベクトルの変換により目的とする個性を生成する試みがなされているが，訓練データとして大量の発話文に対して人手で個性を付与した応答文を作成する必要があった．本稿では，ソーシャルネットワーキングサービス (SNS) から得られる対話データと，学習したい発話者情報を保持した対話データを合わせて訓練データを作成することで，訓練データ作成のコストを削減する．

### 2.3　敵対的生成ネットワーク

画像処理の分野では，画像生成において敵対的生成ネットワーク (Generative Adversarial Nets; GAN) [15–17] を用いることで，他の従来手法より自然な画像が生成できることが示されている．敵対的生成ネットワークでは，画像を生成する Generator と，訓練データ中の実際に撮影された画像かモデルが生成した合成画像かを判定する Discriminator の 2 つのニューラルネットワークを交互に競い合うように学習することで，より自然な画像を生成できるようになる．

本稿では，この敵対的生成ネットワークの枠組みをペルソナベクトルの学習に用いることで，従来手法で得られるペルソナベクトルよりも，より的確に個性を反映できることを示す．

## 3.　提 案 手 法

本稿では，特定の個性を表現するペルソナベクトルの演算により，開発者が意図した新たな個性での応答文を生成する手法を提案する．さらに，ペルソナベクトルの学習に敵対的生成ネットワークの枠組みを導入し，従来手法と比べてより的確に個性を学習できることを目指す．

### 3.1　学習モデル

本モデルは，Li らが提案する Speaker Model [11] を元に構築され，新たにペルソナベクトルを生成する Generator と，出力応答文を判定する Discriminator が加えられる．図 2 に，提案する学習モデルの概要図を示す．

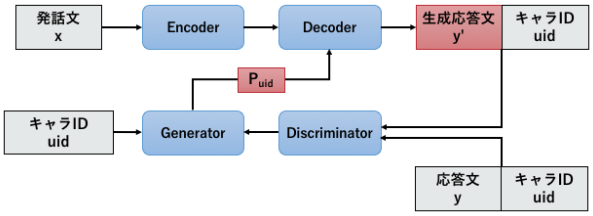


図 2　学習モデル

応答文の生成部は，入力である発話文 $x$ から中間表現 $v$ を生成する Encoder，その中間表現 $v$ とペルソナベクトル $P_{uid}$ の入力から応答文 $y'$ を生成する Decoder で構成される．式 5 の条件付き対数尤度が最大化するように学習する．

$$p(y|x) = \prod_{t=1}^{T_y} p(y_t|v, y_1, y_2, \cdots, y_{t-1}, P_{uid}) \qquad (5)$$

ここで，$uid$ は応答文の発話者に割り当てられた ID を示す．

Generator では，訓練データ内の人物を示す通し番号 $uid$ が入力として与えられ，特定の個性を表現するペルソナベクトル $P_{uid}$ を出力する．Discriminator では，生成応答文 $y'$ と訓練データ内の実際の応答文 $y$ のいずれかと，$uid$ が入力として与えられ，その応答文が実際の人物 $uid$ による生成文であるかどうかを判定する．それぞれのネットワークは式 6 の通りに学習する．

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[logD(x|uid)] + \mathbb{E}_{z\sim p_z(z)}[log(1 - D(S(G(uid))))] \qquad (6)$$

ここで，D は Discriminator，G は Generator，S は seq2seq における Decoder を示す．

### 3.2 ペルソナベクトル

ペルソナベクトルとは，特定の個性を表現する分散表現である．本稿では，ペルソナベクトルの演算により，意図した個性を表現するベクトルを得られることを示す．Mikolov ら [18] の研究では，単語を意味ベクトルで表現した場合，例えば，フランスとパリやドイツとベルリンといった，国と首都の関係が同一ベクトルでの加算演算で構成されていることを明らかにした．さらに，Speaker Model [11] で得られる個性を表現したベクトルに関しても，単語の意味ベクトルと同様に演算が可能であることが濱田ら [14] の研究によって示されている．そこで，濱田らの演算手法を元に，式 7 の通りにペルソナベクトルを演算する．

$$p_{uid}^{CONV} = p_{uid} + \alpha(p^{TARGET} - p^{SOURCE}) \qquad (7)$$

個性を付与したいベクトル $p_{uid}$ に，目的のベクトル $p^{TARGET}$ と個性の付与されていないことを表現したベクトル $p^{SOURCE}$ の差分を足し合わせることで，目的のペルソナベクトル $p_{uid}^{CONV}$ を得る．なお，$\alpha$ は個性付与の強弱を任意に指定できる値である．

## 4. 評価実験

実験 1 では，ペルソナベクトルにより個性を反映した応答文が生成されているかどうか確認するために，提案手法とベースライン手法による応答文を比較評価する．また，実験 2 では，ペルソナベクトルの演算による新たな個性を反映した応答文が生成されるかどうかを確認するために，ペルソナベクトルの演算で新たな個性を段階的に付与し，個性反映の様子を調査する．

### 4.1 実験設定

#### 4.1.1 データセット

モデルの学習に使用する対話データは，発話文・応答文・応答文の発話者名の 3 つの要素を含む日本語のデータである．seq2seq を用いた文生成には大規模な訓練データが必要であるが，全て発話者の情報を保持した対話データを大量に収集するのは困難なため，発話者の情報の持たないデータも学習に使用した (表 1)．発話者の情報を保持した対話データは実際の作品中のセリフから，保持しない対話データは Twitter(注1) から収集した．

評価を行うために使用するテストデータとしては，訓練データ内に存在しない発話文を 100 件を用いた．

表 1　訓練データの概要

| | |
|---|---|
| 発話者の情報あり | 69,421 対話 |
| 発話者の情報無し | 329,005 対話 |
| 合計 | 398,426 対話 |

#### 4.1.2 学習設定

Encoder の隠れ層の次元数は 1000，Decoder の隠れ層の次元数を 1000 に設定する．最適化アルゴリズムには Adam [19] を使用する．

### 4.2 実　験　1

提案手法，及び，個性付与における先行研究である Speaker Model(従来手法) のそれぞれで学習したペルソナベクトルを用いて，発話文を入力として与え，出力される応答文を評価する．各手法において，学習データに存在する個性を学習し，モデルに入力された新たな発話文に対しても任意の個性を正しく反映することが出来るのかどうかを比較する．生成された各応答文に対して，以下の 3 項目をそれぞれ 2 段階で評価する．

- 文法的に正しいか
- 目的の個性が反映されているといえるか
- 対話の応答文として正しいか

テストデータ 100 件に対して出力された応答文を評価した結果を表 2 に示す．また，提案手法と従来手法によって実際に生成された応答文の例を表 3 に示す．

(注1)：http://twitter.com

表 2　実験 1 における評価結果

| | 提案手法 | | | | 従来手法 | | | |
|---|---|---|---|---|---|---|---|---|
| | 個性 A | 個性 B | 個性 C | 合計 | 個性 A | 個性 B | 個性 C | 合計 |
| 文法的に正しいか | 74% | 58% | 71% | 67.7% | 60% | 62% | 74% | 65.3% |
| 個性が反映されているか | 65% | 52% | 65% | 60.7% | 59% | 55% | 69% | 61.0% |
| 応答文として正しいか | 49% | 36% | 27% | 37.3% | 29% | 26% | 19% | 24.7% |

表 3　生成された応答文の例

| 入力文 | | あけましておめでとうございます。今年もよろしくお願い致します。 |
|---|---|---|
| 提案手法 | 個性あり A | よろしくお願いします! |
| | 個性あり B | バッチリお会いに…! |
| | 個性あり C | よろしくお願いします! |
| 従来手法 | 個性あり A | お、ヒッコリー |
| | 個性あり B | あら!シンジ君……ありがとう! |
| | 個性あり C | わあ! |
| 入力文 | | おはようございます！今日も一日頑張りましょう |
| 提案手法 | 個性あり A | おはよう、閣下ちゃん。 |
| | 個性あり B | おはよう、海の支度だよ。 |
| | 個性あり C | おはよう、今日も一日頑張りましょう! |
| 従来手法 | 個性あり A | おはよう |
| | 個性あり B | あなたアキ、待機しているのね、おは ok。 |
| | 個性あり C | ルイくん。 |
| 入力文 | | よろしくお願いします ♡ |
| 提案手法 | 個性あり A | よろしくお願いします! |
| | 個性あり B | よろしくお願いします! |
| | 個性あり C | よろしくお願いします! |
| 従来手法 | 個性あり A | ううっうぅぅ |
| | 個性あり B | わ、碇司令をメンテナンスをを祈ります。 |
| | 個性あり C | かばんちゃん、わたし達、パークの公演を見て、わたし達 |
| 入力文 | | 最近ふとした時に考え込んだり悩んだりすることが多いのが自分でわかる、疲れた。 |
| 提案手法 | 個性あり A | どこのどい、こそ、みんなの本音が、大きいんだよ |
| | 個性あり B | うわ、無事心配させてしまってごめんなさい。 |
| | 個性あり C | うわ、無事 |
| 従来手法 | 個性あり A | シャア、僕は、ミライさんはジオンの坊やは、あなたの |
| | 個性あり B | ……シンジ君、あなたの顔色のために、あなたの命令… |
| | 個性あり C | …あ、… |

### 4.3 実　験　2

実験 1 における個性 A，B に対して，それぞれ個性 C を付与する．演算には式 7 を使用し，$\alpha$ の値は 0〜1.0 まで 0.1 刻みで設定する．この時に生成される応答文が，演算によって生成された新たな個性が付与されていることを確認する．個性 A に個性 C を付与した結果が表 4，個性 B に個性 C を付与した結果が表 5 の通りである．

表 4　個性 A に個性 C を付与した結果

| 入力文 | | 函館着いた!ただいま!帰宅するまでが遠足です! |
|---|---|---|
| $\alpha$ | 0 | おつかれさん |
| | 0.1 | おつかれさん |
| | 0.2 | おつかれさん! |
| | 0.3 | おつかれさまでした。 |
| | 0.4 | 今は、まだそちらからしているのか? |
| | 0.5 | それは大変でしたね |
| | 0.6 | それは大変でしたね |
| | 0.7 | それは大変でしたね |
| | 0.8 | それは大変 |
| | 0.9 | それは大変 |
| | 1.0 | それは大変 |

表 5　個性 B に個性 C を付与した結果

| 入力文 | | 世界史頑張ったし微積も頑張ったから今日は行ける |
|---|---|---|
| $\alpha$ | 0 | ? |
| | 0.1 | ?!なんで? |
| | 0.2 | ?!なんで? |
| | 0.3 | ?!なんで? |
| | 0.4 | ?!なんで? |
| | 0.5 | 大丈夫? |
| | 0.6 | 大丈夫? |
| | 0.7 | 大丈夫?早く教えてくれれば。 |
| | 0.8 | 大丈夫?早く教えてくれれば。 |
| | 0.9 | 大丈夫?ちゃんと調整してたのに… |
| | 1.0 | 大丈夫?ちゃんと調整してたのに… |

## 5. 考　　察

実験 1 では，提案手法が入力された発話文に対して，一定の精度で個性を付与した応答文を出力できることが確認された．特に，各個性において評価の差は見られるものの，合計値では文法，及び応答文としての精度は従来手法を上回る結果となった．また，個性の反映に関しては従来手法とほぼ変わらない結果となった．表 3 の各手法で得られたペルソナベクトルによる応答文を確認すると，従来手法はその人物が使用する人名などの固有名詞や口調が多く含まれているのに対し，提案手法ではそういった単語が比較的少ない．このことから，従来手法の方がより学習対象の人物が使用する語彙集合を出力に反映し，個性の表現が出来ているといえる．提案手法でもより深く個性を表現するために，モデルの学習回数を増やしたり，モデルの改良などを行う余地がある．個性の付与により対話の応答文としての評価が下がる可能性もあるため，両方のバランスを取りつつ適切に個性を表現しなければならない．

なお，いずれの手法でも，文法的に正しい応答文が 60%を超えて出力できているのにも関わらず，対話の応答文としての評価は 40%にも達していない．この原因は，訓練データ数が大きく影響していると考えられる．対話の応答文としての評価を上げるためには，より大規模かつ多様な対話データが必要であると考えられる．

実験 2 では，任意の個性に別の個性を付与することで，新たな個性を持った応答文を生成できることが確認された．また，$\alpha$ 値の変化によって個性に重み付けをした合成が可能となるが，これにより出力される単語や，文の内容，口調などが徐々に変化している様子が確認された．本稿では 2 種類のペルソナベクトルの合成を検証したが，個性の差分を取り分解したり，3 種類以上のペルソナベクトルを組み合わせたりすることで，生成できる個性の幅も広がると考えられる．そのためにも，より多くの種類の個性に対応したペルソナベクトルを生成することが不可欠である．

## 6. ま　と　め

本稿では，対話において重要な要素である「個性」の付与が可能な対話システムを提案した．その上で，訓練データ中に含まれない新たな個性の生成を検証した．結果として，提案手法では一定の精度で対話応答文の出力が行えたものの，個性の反映にはまだ改善の余地が見られた．

提案手法を用いて自由自在に個性を生成するにはまだ課題が残されている．今後は，GAN の仕組みを用いて学習されたペルソナベクトルがより個性を表現できるよう，ニューラルネットワークの学習方法を改善していく予定である．また，最終的には，生成された各ペルソナベクトルを用いて，誰もが簡単に新たな個性を意図した通りに生成できるようなシステムの実装を目指す．

## 文　　献


[1] Alexander Rudnicky, Antoine Raux, Ian Lane, and Teruhisa Misu. *Situated Dialog in Speech-based Human-computer Interaction*. Springer, 2016.

[2] Alan Ritter, Colin Cherry, and William B Dolan. Data-driven response generation in social media. In *Proceedings of the conference on empirical methods in natural language processing*, pp. 583–593. Association for Computational Linguistics, 2011.

[3] Kozo Chikai and Yuki Arase. Analysis of similarity measures between short text for the ntcir-12 short text conversation task. In *NTCIR*, 2016.

[4] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*, 2015.

[5] Chen Xing, Wei Wu, Yu Wu, Jie Liu, Yalou Huang, Ming Zhou, and Wei-Ying Ma. Topic aware neural response generation. In *AAAI*, pp. 3351–3357, 2017.

[6] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pp. 3104–3112, 2014.

[7] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. Learning representations by back-propagating errors. *Cognitive modeling*, Vol. 5, No. 3, p. 1, 1988.

[8] Paul J Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, Vol. 78, No. 10, pp. 1550–1560, 1990.

[9] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.

[10] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*, 2014.

[11] Jiwei Li, Michel Galley, Chris Brockett, Georgios P Spithourakis, Jianfeng Gao, and Bill Dolan. A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155*, 2016.

[12] 赤間怜奈, 稲田和明, 小林颯介, 佐藤祥多, 乾健太郎. 転移学習を用いた対話応答のスタイル制御. 言語処理学会第 23 回年次大会発表論文集, pp. 338–341, 2017.

[13] 宮崎千明, 平野徹, 東中竜一郎, 牧野俊朗, 松尾義博, 佐藤理史. 文節機能部の確率的書き換えによる言語表現のキャラクタ性変換. 人工知能学会論文誌, Vol. 31, No. 1, pp. DSF–E_1, 2016.

[14] 濱田晃一, 藤川和樹, 小林颯介, 菊池悠太, 海野裕也, 土田正明ほか. 対話返答生成における個性の追加反映. 研究報告自然言語処理 (NL), Vol. 2017, No. 12, pp. 1–7, 2017.

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.

[16] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[17] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[18] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[19] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.