

多角的なテキスト特徴量に基づくフェイク動画検出

古川 雄太¹ 駒野 雄一^{1,a)}

概要：近年、日本でも政治家のフェイク動画が選挙や世論に悪影響を及ぼす問題が深刻化している。フェイク動画は、画像や音声などの映像中の不自然なノイズに着目して検出すことが一般的であるが、AI技術が進化すると不自然なノイズが少くなり、検出が困難になる。一方、政治家などが本来の思想とは異なる発言をしているように見せかけたフェイク動画の検出には、映像中のノイズだけではなく、発言を文字起こししたテキスト情報も利用できる可能性がある。本研究では、フェイク動画の問題に対処するため、真偽対象となる動画内の発言のテキスト情報に対して言語的特徴と感情特徴に着目し、これらの特徴を多角的に分析・統合してフェイク動画を検出する手法を提案する。また、国会議事録検索システムから抽出した7人の政治家の発言を利用し、提案手法の有効性を評価する。評価の結果、現時点では提案手法の有効性は限定的ではあるが、統合値のほうが言語や感情の個々の特徴よりも分散が小さく、複数の特徴を多角的に分析・統合することで発言者の特徴を見いだせている可能性がある。データ量や粒度などを改善した更なる評価や、言語や感情に加えて思想など新たな特徴量も加えた手法の開発と評価は今後の課題である。

キーワード：フェイク動画、言語的特徴、感情、RoBERTa、LSTM、感情辞書

Fake Video Detection based on Multifaceted Text Features

YUTA FURUKAWA¹ YUICHI KOMANO^{1,a)}

Abstract: Recently, fake videos of politicians have become a serious problem, adversely affecting elections and public opinion. Fake videos are typically detected by focusing on unnatural noise(s) in images and/or audio. However, advances in AI technology have reduced this noise and fake detections become more difficult. On the other hand, textual information transcribed from statements may also be useful for detecting fake videos in which politicians make statements that differ from their true beliefs. To address the issue, this paper focuses on linguistic and emotional features of the text information in the videos, and proposes a method for detecting fake videos by analyzing and integrating these features. We also evaluate the effectiveness of the proposed method using statements by seven politicians. The evaluation results show that, although the effectiveness of the proposed method is currently limited, the integrated value has smaller variance than individual linguistic and emotional features. This means that multifaceted analysis and integration of multiple features may identify speaker characteristics. Future challenges include further evaluation with improved data volume and granularity, and the development and evaluation of methods that add new features such as thought in addition to language and emotion.

Keywords: Fake videos, linguistic features, emotions, RoBERTa, LSTM, emotion dictionary

1. はじめに

近年、生成AI技術の発展により、本物の動画と区別がつかないほど精巧なディープフェイク動画が容易に作成で

き、社会的な問題に発展しやすい状況になっている[3], [5]。ディープフェイク動画は偽情報拡散や名誉棄損、政治的混乱などの様々なリスクをもたらすため、その検出技術の開発は喫緊の課題[1]である。

従来のディープフェイク動画検出手法[7], [8], [9]では、画像や音声も含む映像中の不自然なノイズを検出し、フェ

¹ 千葉工業大学

Chiba Institute of Technology

a) yuichi.komano@p.chibakoudai.jp

イク動画を判別してきた。しかし、高度な技術で生成されたディープフェイク動画は不自然さが巧妙に隠されている場合が多く、従来の検出手法では見破ることが困難になっている。

本稿では、政治家が本来の思想とは異なる発言をしていくように見せかけたディープフェイク動画の検出手法を研究対象とする。このようなディープフェイク動画の検出には、映像中の不自然なノイズだけでなく、発言者の日頃の発言の内容や表現などのテキストの特徴も利用できると考えられる。そこで、本稿は発言者の口調や感情などテキストから得られる多角的な特徴量に基づくフェイク動画の検出手法を提案し、その有効性を実験により検証する。

1.1 関連研究

笠井ら [13] は、テキストの特徴に着目して、悪意のある切り貼り操作による動画を編集（ダイジェスト版を装って発言の趣旨を変えた動画、本稿では切り貼り動画とよぶ）したディープフェイク動画の検出手法を提案した。悪意のある切り貼り動画は、映像中の不自然なノイズが少ないため、検出が難しい。笠井らは、切り貼りの元となるオリジナルの動画と切り貼り動画からテキストを文字起こしし、(1) オリジナル動画と切り貼り動画のテキストを直接比較する、(2) オリジナル動画のテキストから作成した要約文と切り貼り動画のテキストとを比較する、(3) 切り貼り動画のテキストから質問を作成し、オリジナル動画のテキストに基づく回答と切り貼り動画のテキストとを比較する、という三つの手法を提案した。

しかし、笠井らの手法はオリジナルの動画を必要とするため、オリジナルの動画が入手困難である場合や、複数の動画が切り貼りされるなどしてオリジナルの動画を特定できない場合には、適用することができない。

Mittal ら [8] は、複数の特徴量を利用してディープフェイク動画を検出する手法として、深層学習技術を用いたオーディオ・ビジュアルディープフェイク動画検出手法を提案した。Mittal らの手法は、動画内の音声と視覚（顔）の両方の情報源から感情的な特徴を抽出してそれらの類似性を分析し、ディープフェイク動画を検出する。

しかし、Mittal らの手法では、前述の切り張り動画のようにそれぞれの時点で動画内の音声と視覚（顔）の特徴に差が生じない場合には、ディープフェイク動画を検出することはできない。

1.2 本稿の貢献

本稿の貢献は以下のとおり。

- 本稿は、動画内の発言者の口調と表現の特徴に着目して、多角的なテキスト特徴量に基づくディープフェイク動画の検出手法を提案する。我々が提案する手法は、笠井ら [13] の手法とは異なり、比較対象となるオリジ

ナルの動画は不要である。そのため、オリジナルの動画を入手することができないような場合でも、ディープフェイク動画を検出することができる。

- 我々の提案では、多角的なテキスト特徴量として、発言者の口調を中心とする言語的特徴モデルと、発言者が用いる単語に含まれる感情の時系列データである感情軌跡モデルを利用する。提案手法では、発言者ごとにこれらの特徴量を評価するモデルを事前に学習して作成する。そして、ディープフェイクが疑われる検査データが与えられたら、それぞれのモデルでその発言者らしさを評価して、得られた評価値を統合することでフェイクであるか否かを判定する。
- 提案手法の有効性を実験により検証する。実験では、国会議事録などから 7 人の政治家の発言を抽出し、政治家ごとに言語的特徴モデルと感情モデルをそれぞれ作成する。そして、それぞれの政治家の学習には用いていない発言を検査対象として、それらの発言が正しい発言者によるものであると分類されることを確認する。
- 実験の結果を踏まえて、提案手法の有効性を議論する。用いることができる特徴量としては、本稿で用いる口調と表現のほかに、思想や信念などが挙げられる。これらを学習してフェイク検出に応用するためには多量のデータを準備することなどが課題であるが、本稿の発展として今後取り組んでゆく予定である。

本稿の構成は以下のとおり。2 節では、準備として本稿で使用する要素技術を説明する。3 節では、本稿で提案する“多角的なテキスト特徴量に基づくフェイク動画識別手法”について、構成のアイデアと具体的な構成を説明する。4 節では、提案した識別手法の有効性を確認するための評価実験の環境と、実験結果を紹介する。そして、5 節で実験結果を考察し、最後に 6 節でまとめを述べる。

2. 準備

本節では、提案手法で使用する RoBERTa、LSTM、感情辞書について説明する。

2.1 RoBERTa

我々の提案手法では、Transformer[10] のアーキテクチャに基づいた事前学習済み言語モデルである RoBERTa[6] を使用する。

RoBERTa は、オリジナルの BERT[2] を改良したモデルであり、より大規模なデータセットでの事前学習、より長い事前学習時間の確保、および動的なマスキング手法を採用している点が特徴である。

2.2 LSTM

我々の提案手法では、テキストの時系列的な感情の「軌跡」

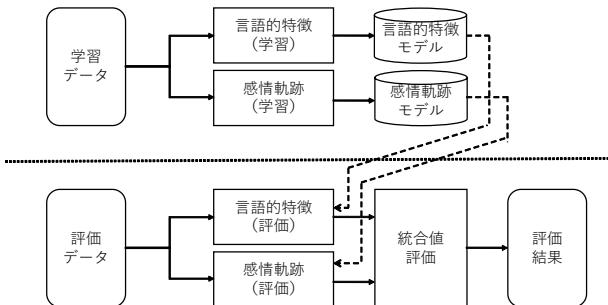


図 1 提案手法の全体像

を捉えるため、長・短期記憶（Long Short Term Memory, LSTM [4]）ネットワークを使用する。

LSTM は、時系列データを扱うリカレントニューラルネットワーク（RNN）の一種であり、勾配消失問題を軽減するために、入力ゲート、忘却ゲート、出力ゲートをもつことが特徴である。

2.3 感情辞書

我々の提案手法では、テキストにあらわれる感情の起伏を捉えるため、事前に構築された感情辞書 [12] を使用する。

この感情辞書は、日本語の感情を表す言葉を特定のカテゴリに分類したリストである。具体的には、愛情、忍耐、驚き、悲哀、といった 68 種類の感情カテゴリと、それに紐づく言葉が含まれている。

提案手法では、これらの言葉の出現の頻度と順序に基づいて、テキストの感情スコアを算出する。これにより、単一のテキストの感情だけでなく、時系列的な感情の変化を分析するための特徴量を生成することができる。

3. 提案手法

1 節で説明したように、本稿は政治家が本来の思想とは異なる発言をしているかのように見せかけたディープフェイク動画の検出手法について検討する。本節では、提案手法の全体像と構成要素について説明する。

3.1 多角的なテキスト特徴量に基づくフェイク動画検出

図 1 に提案するディープフェイク動画の検出手法の全体像を示す。提案する検出手法は、言語的特徴を学習・評価する機能と感情特徴を学習・評価する機能、それらの特徴を踏まえて分析した評価結果から統合値を計算する統合値評価の機能をもつ。

提案手法では、言語的特徴モデルと感情特徴モデルをそれぞれ事前に学習する。そして、評価対象のデータが与えられたら、それぞれのモデルを用いて評価結果を算出し、それら評価結果から計算される統合値を利用してフェイクを検出す。

3.1.1 言語的特徴モデル

話者固有の語彙や、言い回し、構文といった言語的な

スタイルを捉えるため、事前学習済み言語モデルである RoBERTa をファインチューニングして、言語的特徴モデルを構築する。

ベースモデルには、nlp-waseda/roberta-base-japanese[11] を採用する。このモデルは大規模な日本語コーパスで事前学習されており、日本語の文脈理解に優れた性能をもつ。事前学習済みモデルの知識を活用しつつ、ベースモデルの最終層には、話者のクラス数（4 節の評価実験では 7）に対応する出力をもつ全結合層を追加し、モデルを学習させる。

評価実験では、入力として句点までの文章（4 節の sentence に該当する）を受け取り、7 名の話者それぞれに対するクラス確率を 7 次元のベクトルで出力する。モデルの汎化性能を堅牢にするため、k 分割交差検証を用いて学習と評価を行う。

3.1.2 感情軌跡モデル

言語的なスタイルに加えて、発言中の感情も話者を特徴づける重要な要素であると仮定し、感情を捉えるための複合モデルを構築する。本モデルでは、テキストエンコーダとしての RoBERTa、時系列特徴抽出器としての LSTM (Long Short-Term Memory)、そして感情の情報源としての感情辞書を組み合わせた独自のモデルを構築する。

感情軌跡は、一連の文章からなるセッション単位で分析する。例えば、国会答弁において、ある質問に対する回答をはじめてから終わる（話者が切り替わる）までの一連の文章をセッションとする（4 節の session に該当する）。セッション単位で分析することで、同一の感情の文章を続けて発言する、感情をすぐに抑制する、などの発言者の特徴が捉えられるのではないかと期待した。

入力されたセッションテキストを文章単位で分割し、各文章に含まれる感情語の出現頻度から感情ベクトルを算出する。これは、事前に定義された感情辞書に基づいて行われる。文章単位で得られた感情ベクトルの時系列は、文章ごとの感情の起伏を反映するため、ノイズが含まれる可能性がある。そこで、時系列データの短絡的な変動を滑らかにし、より感情の移動を捉るために、移動平均を適用する。これにより、ばらつきの少ない感情の時系列データが生成される。平滑化された感情の時系列データを、LSTM に入力する。これにより、発言全体を通じた感情の変化や時間的関係のノイズを抑えた形で学習できる。

感情特徴とは独立して、セッションテキスト全体を RoBERTa に入力し、テキストがもつ意味や文脈を数値化された特徴量ベクトルに変換する。RoBERTa から得られた特徴量ベクトル、LSTM から得られた特徴量ベクトルを連結し、これを最終的な特徴量として全結合層に入力することで、話者に対するクラス確率を出力する。

3.1.3 統合判定部

最終的な話者識別の評価は、3.1.1 節で述べた言語的特徴

モデルと 3.1.2 節で述べた感情軌跡モデルの出力を統合することによって得られる。

フェイクを判定する動画からテキストを文字起こしし、句点を区切りとして複数の文に分割する。各文に対して、学習済みの言語的特徴モデルと感情軌跡モデルをそれぞれ適用する。各モデルは入力された文が話者に対応するクラスのそれぞれに所属する確率を算出する。

そして、得られた 2 つの確率を統合し、単一の統合確率を生成する。本稿では、統合手法として相加平均と相乗平均の 2 つを検討し、それぞれ評価する。

4. 評価実験

本節では、提案手法の有効性を検証するための評価実験の環境と実験結果について説明する。

4.1 実験環境

実験には以下の環境を用いた。

- CPU: AMD Ryzen5 4500 6-Core Processor (3.60 GHz)
- GPU: NVIDIA GeForce RTX 3060
- RAM: 16.0 GB
- OS: Windows 11
- プログラミング言語: Python 3.11.9

実験に用いるデータは、国会議事録検索システム [14] から、検索用 API を使用して以下の 7 名の発言のテキストデータを取得した：安倍晋三氏、石破茂氏、山口那津男氏、野田佳彦氏、蓮舫氏、玉木雄一郎氏、志位和夫氏。

実験に用いたデータサイズは以下のとおりである。

- RoBERTa モデル（単一）
 - 総サイズ：22MB
 - 総件数：112,000sentence (16,000sentence/人)
- RoBERTa + LSTM + 感情辞書 + 移動平均モデル
 - 総サイズ：14.2MB
 - 総件数：7,000session (1000session/人)

4.2 実験結果

本節では、3.1 節に示した二つの特徴量のそれぞれでの識別結果と、それらを統合した本稿の提案手法の識別結果をまとめた。本研究は政治家の人格的側面の分析を含み得るため、本節以降では政治家の氏名は伏せて識別結果を報告して議論する。表 1 にそれぞれの識別結果を示す。

4.2.1 言語的特徴モデル

対応する識別結果は、表 1 の 2 列目から 5 列目までとなる。B (平均特徴量: 0.81, 平均順位: 1.27) と G (平均特徴量: 0.75, 平均順位: 1.54) に対しては、他の政治家と比べて高い性能を示した。

特に B に対しては、7 名の中で平均特徴量が最も高く、平均順位が最も低くなかった。この結果から、B の発言には

言語的特徴に独自性があり、識別しやすかったと考えられる。

4.2.2 感情軌跡モデル

対応する識別結果は、表 1 の 6 列目から 9 列目までとなる。D (平均特徴量: 0.62, 平均順位: 1.96) と G (平均特徴量: 0.65, 平均順位: 1.77) に対して、良好な性能を示している。この結果から、D と G の発言には識別可能な感情的パターンが含まれていると考えられる。

一方、B (平均特徴量: 0.21, 平均順位: 2.87) に対しては、著しく性能が低い結果となった。言語的特徴モデルで最も識別しやすかった B が、感情特徴モデルでは識別が困難で合ったことから、B の言語スタイルには独自性があるが、感情表現は抑制的あるいはパターン化されていないことを示しており興味深い。

C (平均順位: 3.26, 順位分散: 4.354) と F (平均順位: 3.12, 順位分散: 4.9) に対しては、平均順位が 3 位以下と低く、分散も極めて大きい。この結果から、感情軌跡からは C と F に対して識別に有効な特徴を抽出できなかったことを意味している。

4.2.3 統合判定

対応する識別結果は、表 1 の 10 列目から 13 列目（相加平均を用いた統合評価）、14 列目から 17 列目（相乗平均を用いた統合評価）となる。

統合モデル（相乗平均）は、個々のモデルと比較して順位の分散を抑制する傾向がみられた。例えば、G の平均順位は、言語 (1.54) → 感情 (1.77) → 相加平均 (1.58) → 相乗平均 (1.55) と推移し、順位の分散も、言語 (1.516) → 感情 (1.884) → 相加平均 (1.559) → 相乗平均 (1.506) と推移して、どちらも最終的に最も低い値となった。これは、二つのモデルからの情報を統合することで、より安定した予測が可能になることを示している。

言語的特徴に高い性能を示していた B の識別においては、平均順位は相加平均 (1.51)、相乗平均 (1.61) となっただ。性能の低い感情モデルの結果を統合したことで、言語 (1.27) よりも順位の平均が悪化したものの、大きな性能の低下は避けられた。特に、順位の分散は、相加平均 (0.82) と相乗平均 (1.02) であり、感情 (2.35) に比べて大幅に改善した。

全体的に、相加平均と相乗平均に大きな性能の差はみられなかった。しかし、最も安定した性能を示した G のケースでは、相乗平均が最も良い平均順位 (1.55) と最も小さな分散 (1.506) を実現した。相乗平均は、モデルの一方でも確信度が低い場合には結果が大きく下がり、両方のモデルがともに「確からしい」と判断した場合をより強く評価する効果がある。そのため、この効果が安定性の向上に寄与したと考えられる。

表1 統合判定での実験結果

政治家	言語的特徴モデル				感情軌跡モデル				統合判定(相加平均)				統合判定(相乗平均)			
	特微量		順位		特微量		順位		統合量		順位		統合量		順位	
	平均	分散	平均	分散	平均	分散	平均	分散	平均	分散	平均	分散	平均	分散	平均	分散
A	0.68	0.149	1.54	1.172	0.34	0.18	2.61	3.012	0.51	0.106	1.78	1.334	0.35	0.154	1.87	1.644
B	0.81	0.109	1.27	0.526	0.21	0.115	2.87	2.35	0.52	0.07	1.51	0.82	0.3	0.119	1.61	1.02
C	0.62	0.157	1.7	1.421	0.31	0.176	3.26	4.354	0.48	0.115	1.95	1.898	0.35	0.153	2.13	2.46
D	0.6	0.174	1.86	2.057	0.62	0.201	1.96	2.526	0.61	0.147	1.81	2.056	0.54	0.183	1.79	2.077
E	0.7	0.138	1.48	1.029	0.41	0.186	2.28	2.1	0.56	0.112	1.7	1.326	0.44	0.161	1.65	1.282
F	0.59	0.168	1.76	1.53	0.37	0.196	3.12	4.9	0.49	0.12	1.87	1.489	0.35	0.164	2.32	3.435
G	0.75	0.143	1.54	1.516	0.65	0.19	1.77	1.884	0.7	0.13	1.58	1.559	0.64	0.174	1.55	1.506

5. 考察

言語的特徴モデルを用いた識別は、独自のスローガンや言い回しをもつBのような話者に対しては有効となる一方で、より一般的、あるいは多様な表現方法を用いるDやFのような話者に対しては現時点では効果が限定的であることが明らかになった。特に、Dに対しての性能の不安定さは、Dが演説の状況に応じて多様なスタイルを取りことを反映している可能性がある。

感情軌跡モデルは、DやFのような特定の話者に対しては有効な特徴を捉えたが、有効となる範囲は言語的特徴モデルよりも限定的であった。しかし、Bの結果は、言語的特徴と感情軌跡は話者を識別する上で補完的な情報を提供しうることを示している。すなわち、単一のモデルでは捉えきれない話者の多面的な特徴を捉えるために、モデルの統合が有効である可能性が示唆される。

実際、本実験において、単一のモデルでは、特定の話者に対する識別性能が不安定になる、あるいは識別自体が困難であるという課題がみられたが、二つのモデルの出力を統合したところ多くの話者で予測順位の分散が低下する傾向が確認された。これは、統合によって予測の安定性が向上したことを示している。

しかし、現時点では、統合による性能向上には限界もみられた。特に、いずれの単一モデルでも識別が困難であったC、D、Fなどに対しては、統合後も平均順位や特微量の平均値に顕著な改善はみられなかった。本稿で用いた相加平均や相乗平均を利用した統合は、性能の安定化には寄与したが、識別困難なクラスに対する根本的な識別能力の向上には至らなかった。

相乗平均を用いた統合手法は、一部の話者において、平均順位と分散の両方で最もバランスの取れた結果を示した。これは、相乗平均が両モデルから一貫して高い評価を得た予測を重視する特性をもつためと考えられる。一方で、この特性は片方のモデルの性能が極端に低い場合(Bに対する感情軌跡モデル)、全体の評価を過度に引き下げる可能性もある。したがって、相乗平均が全てのケースで最適な

統合手法であるとはいいきれない。

以上のことから、本稿で提案した多角的な特微量に基づくアプローチは、単一モデルの不安定さを緩和するという点で一定の有効性を示したが、識別困難なクラスへの対応および各話者の特性に応じた最適な統合手法の選定は、さらなる検討が必要である。

6. まとめ

本稿は、多角的なテキスト特微量に基づくフェイク動画の検出手法として、話者の言語的特徴と感情軌跡という二つの側面に注目したアプローチを提案した。言語的特徴モデルを構築するために、話者固有の語彙や文体を捉るために、事前学習済み言語モデルRoBERTaを採用した。また、感情軌跡モデルを構築するために、発言中の感情の推移を捉えるための感情辞書とLSTMを組み合わせた複合モデルを採用した。

実験の結果、言語的特徴と感情軌跡が、話者識別のための相補的な情報源として機能することが確認できた。単一のモデルでは特定の話者の識別が困難であったり、予測が不安定になったりする傾向がみられたが、両モデルの出力を統合することで予測の安定性が向上する傾向がみられた。この結果は、単一の特微量に依存するのではなく、複数の特微量元素を組み合わせる本研究のアプローチの有効性を支持するものである。

しかし、本研究には課題も残されている。現状の二つのモデルでは、依然として識別が困難な話者が存在しており、特に言語的特徴が類似していると考えられる話者間では分類が誤っていた。そこで、今後の展望として、第三のモデルとして「政治思想モデル」を導入する予定である。このモデルは、話者固有の語彙や言い回し、構文などの言語的特徴や、話者固有の感情とは異なり、より根源的に話者の発言を特徴づけるものである。

ところが、個人の識別に利用するための政治思想モデルを作成するためには、より多くのデータを用いた学習が必要となる。そこで、個々の話者をクラスとするのではなく、まずは話者が所属する「政党」をクラスとした「政党思想モデル」を導入する予定である。個人の政治思想モデルや

政党思想モデルを用いることで、発言の政治的・社会的な思想や政策的な立場を特徴量として捉えて、話者の識別が可能になると考えられる。

また、本稿の評価実験では、発言者の識別の性能を評価していた。今後は、本稿の提案手法および政治（政党）思想モデルを組み合わせた手法で、本来の思想とは異なる発言をしたかのようなディープフェイク動画を検出できることも検証してゆく。

謝辞: 本研究は JSPS 科研費 JP24K14951 の助成を受けたものです。

参考文献

- [1] Robert Chesney and Danielle Citron. Deep fakes: A looming challenge for privacy, democracy, and national security. *Available at SSRN 3213954*, 2019.
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding, 2019.
- [3] 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 27*, 2014.
- [4] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [5] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.
- [6] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Li, Alexander Palmer, Marc Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.
- [7] Asad Malik, Minoru Kurabayashi, Sani M. Abdullahi, and Ahmad Neyaz Khan. Deepfake detection for human face images and videos: A survey. *IEEE Access*, 10:18757–18775, 2022.
- [8] Trisha Mittal, Uttaran Bhattacharya, Rohan Chandra, Aniket Bera, and Dinesh Manocha. Emotions don't lie: An audio-visual deepfake detection method using affective cues. In Chang Wen Chen, Rita Cucchiara, Xian-Sheng Hua, Guo-Jun Qi, Elisa Ricci, Zhengyou Zhang, and Roger Zimmermann, editors, *MM '20: The 28th ACM International Conference on Multimedia, Virtual Event / Seattle, WA, USA, October 12-16, 2020*, pages 2823–2832. ACM, 2020.
- [9] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Niessner. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1–11, 2019.
- [10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems 30*, 2017.
- [11] Waseda University, Kawahara Lab and Tohoku University, Imui-Suzuki Lab. nlp-waseda/roberta-base-japanese. <https://huggingface.co/nlp-waseda/roberta-base-japanese>, 2021. Accessed: 2024-05-21.
- [12] 河原修一. 感情を表す日本語の言葉. 金沢大学国語国文, 卷 23, p.74-84 (金沢大学学術情報リポジトリ収録), 1998.
- [13] 笠井健太郎, 栗林稔, 稲村勝樹, 船曳信生. 悪意のある切り貼り操作による動画編集に対する制御手法の一考察. *2024年暗号と情報セキュリティシンポジウム*, 2F3-2, 2024.
- [14] 国立国会図書館. 国会会議録検索システム, 2025.