

# マンガ制作支援に向けたメタデータモデルと 深層学習による構図推薦への応用

伊藤 博典<sup>†</sup> 浅野 泰仁<sup>†</sup>

<sup>†</sup> 京都大学情報学研究科 〒606-8501 京都府京都市左京区吉田本町 36-1

E-mail: †h-ito@db.soc.i.kyoto-u.ac.jp, ††asano@i.kyoto-u.ac.jp

あらまし 近年、マンガは子供だけが楽しむものではなく、幅広い世代に楽しめるものとなっている。またマンガという表現手法は、その表現の自由度の高さによる内容への親しみやすさ、伝わりやすさといった特徴を活かして、広告などのビジネスにおける利用や学習マンガなどの教育における利用など様々な分野において活用が進んでいる。しかし、マンガ制作を専門としていない一般の人がこれを活用・制作するにあたり、脚本等の内容をコマ割り・キャラクター配置・吹き出し配置などによりマンガ表現に落とし込む構図作成の技術が大きなハードルとなっている。そこで、脚本を型に合わせて書くことで自動的に構図を推薦する技術によりこのハードルを乗り越える支援を行うことで、マンガ表現の利活用がより進むと期待される。本論文では、構図と脚本の観点から機械学習を適用するためのメタデータモデルを提案するとともに、深層学習による構図推薦の手法について提案する。

キーワード コミック工学, メタデータモデル, アノテーション, 構図推薦, 深層学習

## 1. はじめに

近年、マンガは子供だけが楽しむものではなく、幅広い世代に楽しめるものとなっており、特に日本の社会においては文化としても広く浸透している。また、マンガという表現手法は、その表現の自由度の高さによるメリットを活かして、近年では商品の宣伝広告などのビジネスにおける利用や学習マンガのような教育における利用など、エンターテインメントの領域以外にも様々な分野において活用が進んでいる。これは日本国内にとどまらず、世界中においても認知や利用が進んでおり、マンガで楽しむ人、マンガでメッセージを伝える人、マンガで学ぶ人など、多様な目的のためにマンガを利用する人はますます増加している。マンガで表現するメリットとしては、次のようなものがある。

- (1) テキストのみの表現よりも目に止まりやすく、記憶に残りやすい。
- (2) ストーリーに乗せて情報を伝えるため、感情移入しやすい。
- (3) テキストのみでは伝わりにくいことを分かりやすく表現できる。

これらのメリットから、一般の人々においてもマンガの表現手法を娯楽・教育・ビジネスといった様々なシーンにおいて利用したい人々も増加している。近年はコミポ! [1] のような絵を書かなくとも、3D モデルを利用してマンガを作成することができるソフトウェアなども市販されている。しかし、マンガ表現の利用にあたって、内容への親しみやすさ、伝わりやすさ、分かり易さといったマンガの表現手法によるメリットを活用するためには、表現/伝達したい内容を書き下した脚本に応じて、マンガにおける各要素を構成する構図の技術が重要である。ここには暗黙的なものも含め専門知識が多用されており、マンガ制作を専門としていない一般の人々にとって、この部分がマンガ

制作における大きなハードルとなっているこのマンガ表現における専門知識の一例としては、場面や状況に応じたパースや人物・吹き出しなどの配置、サイジングなどがある。これらを踏まえずにマンガを制作すると、どのような状況下に登場人物があるのか、何が中心的に伝えたいものなのかといったことが、分かりにくくなってしまい、先に挙げたマンガ表現のメリットを活用できない。

そこで本研究では、マンガ表現によるメリットを活かして、効果的に内容を伝えるために、この専門知識や暗黙的な技術が用いられているマンガをもとに深層学習を行い、脚本に適した構図を生成し、推薦することで一般の人々のマンガ制作支援を行うことを目的とする。

本論文では、機械学習のためのマンガメタデータモデルを提案するとともに、これをもとに構築したアノテーションデータセット及びこれを利用したニューラル機械翻訳による構図推薦手法について提案する。

本論文の貢献としては次の3つがある。

### (1) 機械学習のためのマンガメタデータモデルの提案

機械学習により表現/伝達したい内容を書き下した脚本をもとに構図推薦を行うためには、まず学習データが必要となるが、この脚本と構図の観点から収集されたデータセットは存在していない。そこでデータセットを新たに構築するために、マンガにおいて描かれる各動作により構成される脚本とそれを表現する構図について、その構成要件を詳細に定義するとともに、それらの情報を階層的に結びつけたマンガメタデータモデルを設計した。本提案モデルの定義に当たり、はじめに必要な性質を明らかにしたモデルを作成し、これをもとに作成されたアノテーションデータを用いて、定性評価並びにインタビューを行った。[2] 本論文では、これらの実験結果のフィードバックを取り入れて完成させたメタデータモデルについて提案する。

## (2) アノテーションデータセットの構築

本提案メタデータモデルをもとに、機械学習により表現/伝達したい内容を書き下した脚本をもとに構図推薦を行うために必要となるアノテーションデータセットをクラウドソーシングによって構築した。ここでは提案メタデータモデルをもとに、クラウドワーカーに古典文学をマンガ化した三冊のマンガのコマに対して、構図アノテーション及び脚本アノテーションを付与してもらった。本データセットによりマンガの各コマにおける構図と脚本に関するアノテーションデータの組を得ることを可能にした。

## (3) 深層学習を用いた構図推薦手法の提案

本研究では、マンガ制作における表現/伝えたい内容を書き下した脚本に応じた構図を生成する問題を、可変長の単語ラベル列からなる入力に対して対応する可変長の単語ラベルを適切に出力する翻訳の問題と同等に捉えた。そこで入力文を描きたい内容を書き下した脚本アノテーションとし、出力文をこれに対応した適切な構図を示した構図アノテーションとして、ニューラル機械翻訳モデル [3] を適用することで、構図情報を生成し構図推薦を行った。提案手法では、上記のアノテーションデータセットによってモデルを訓練し、この訓練済みモデルによりユーザがコマ内で表現したい内容を指定の形式で脚本アノテーションとして入力することで、それに適した構図アノテーションを得ることができる。本論文では、提案手法の定量評価により提案メタデータモデルの機械学習への応用可能性を、また定性評価により提案手法が素人によるマンガ制作の支援を行えることを示した。

本論文の構成は以下の通りである。第 2 章では、関連研究としてマンガを対象としたメタデータモデルの諸研究とニューラル機械翻訳の概要について説明する。第 3 章では、機械学習のためのマンガメタデータモデルについて述べ、第 4 章では、機械学習に利用する提案モデルをもとに構築したデータセットについて説明する。第 5 章では、ニューラル機械翻訳を用いた構図推薦手法について説明し、第 6 章では、この構図推薦手法に対して行った評価実験について説明する。第 7 章では、研究全体のまとめと今後の展望について述べる。

## 2. 関連研究

本章では、はじめにマンガを対象としたメタデータモデルの先行研究について紹介し、次に構図推薦に利用したニューラル機械翻訳の概要について説明する。

マンガのメタデータモデルに関する研究としては、Morozumi らの研究 [4] や Fujimoto らの研究 [8] が挙げられる。

Morozumi ら [4] は、ネットワーク環境におけるマンガの発見やアクセス、再利用のためのメタデータフレームワークを提案している。このフレームワークでは、出版された具体物としてのマンガに関する書誌記述、マンガにおける表現やストーリーに関する構造記述、マンガに登場する主要人物や動作主体、コンセプトに関するオントロジー記述の 3 つの観点から議論し、それぞれ「FRBR [5]」、「TV-Anytime [6]」、「知的実体を示す辞書項目 [7]」を基礎として拡張したモデルを提案している。

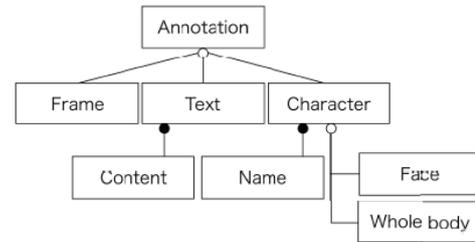


図 1 Fujimoto らのメタデータモデル

Morozumi らの手法では、デジタルマンガを主な対象として、先の 3 つの観点から個別の表現に依らず、内容に関する要素をマンガの構造に従って抽象化された実体として記述するためのメタデータフレームワークを定義している。そのため、書誌情報やキャラクタ、動作、話数、シリーズといった観点からインターネット上におけるマンガ内外における検索やプロによるマンガ制作における再利用に向けたモデルとなっているが、各情報を横断的に結びつけるアノテーションの困難性やそのコストの大きさが問題であり、またマンガ内の各オブジェクトの存在する領域についての情報を持たないため、構図を対象とした機械学習への適用は難しい。

本研究では、マンガ制作に詳しくないユーザを対象として、表現したい内容を書き下した脚本からマンガ構図を自動生成することでマンガ制作支援を行うことを目標としており、その内容を効果的に伝えるためにアングルやショットなどの構図において重要となる要素に着目したメタデータモデルを提案する。また提案メタデータモデルにおいては、話数やシリーズといった大きな粒度のストーリー構造ではなく、マンガ制作の基となる脚本から動作を用いて、マンガ内で描かれる内容を表現できるようにした点も本研究は Morozumi らの研究と異なる。

Fujimoto ら [8] は、マンガ画像における登場人物や文字の認識やテキストと画像からなるマルチモーダルな表現に対応した検索に関する技術を発展させることを目標として、学術利用可能な大規模で高品質なオープンマンガデータセットを構築するとともに、このデータセットのアノテーションに用いるメタデータモデルについて提案した。

Fujimoto らのメタデータモデルは、長方形の囲いを基準として、コマ、登場人物、テキストの 3 つから定義され、各要素の存在する画像中の空間座標、登場人物の名前や顔、全体像、台詞の内容のみから成る簡潔なモデルである。図 1 にこのメタデータモデルを示す。

このモデルは、キャラクタや吹き出しといった観点からの画像的な特徴抽出やその画像認識など機械学習に向けたモデルであり、そのアノテーションのコストも小さく、データセット作成に向けたものとなっているが、マンガ内で表現されている内容については表現できない点やマンガ表現における特徴の一つであるオノマトペや吹き出しの形といった要素が不足している点が、構図推薦において必要となる構図と脚本の関連についての学習において問題である。

提案メタデータモデルでは、マンガにおける構図を表現する

ために必要となる各要素と内容を示す脚本を表現するために必要となる各要素を定義するとともに、これらの関連についても定義した。また、構図において重要な小道具や背景などの場面状況を認識するために必要な要素についても新たに定義した。

次にニューラル機械翻訳の概要について述べる。NMTとは、機械翻訳にニューラルネットワークの計算モデルを適用したものであり、2014年に登場して以降、それまでの標準であった統計的機械翻訳(SMT)から翻訳精度を飛躍的に向上させた技術である。SMTでは、単語の翻訳確率や並べ替えの確率などの翻訳に必要な知識を対訳コーパスから統計的な情報として学習し、辞書や言語モデルといった独立した翻訳知識を用意したり、翻訳時に考慮する素性を設計する必要があるのに対し、NMTでは1つのニューラルネットワークを用意するだけで訓練も翻訳もすべて同じ枠組みで行うことができ、対訳文を与えるだけでニューラルネットワークが翻訳に必要な情報を自動的に学習し、入力から出力までが単一のモデルで完結する。

ここでは、ニューラル機械翻訳の代表的な仕組みについてSeq2Seq[9]を例に概説する。seq2seqは大きく分けて3つのパーツで構成されている。1つ目は、入力文をベクトル表現に変換するエンコーダ、2つ目は出力すべき単語を決定する際に入力文のどこに注目するかをコントロールするアテンション機構、3つ目は入力文とアテンションの情報を基に出力文を生成するデコーダである。

エンコーダでは、入力文中の各単語を分散表現と呼ばれる数百次元からなる実数値ベクトルに変換するembeddingを行う。この分散表現では単語の意味のようなものがそれぞれ表現される。

アテンション機構は、embeddingされた入力文とデコーダの内部状態を判断材料として、次の単語を訳出する際に注目すべき箇所を判断する役割を担う。ここで注目する単語は決定的に1つに決めるのではなく、入力文のすべての単語に対して確率的に注目し、入力文の各単語の分散表現とアテンションを掛けた結果を足し合わせることで、入力分散表現の重み付き和を計算し、これをデコーダで出力する際の判断材料として利用している。

デコーダは、アテンションの情報と1つ前に出力した単語の情報を入力として受け取り、次の単語を出力する。デコーダでは、各単語の出力確率をもとにそれぞれ最も確率の高い単語を出力することで翻訳を行う。以上がseq2seqの仕組みの概略である。

本研究ではコマに描きたい内容を書き下した脚本をもとに適した構図の推薦を行う手法を提案するが、このニューラル機械翻訳(NMT)においてCNN[10]の進化系であるDenseNet[11]を活用したモデル[12]によって脚本アノテーションと構図アノテーションの関連を学習させることによって行った。

### 3. 機械学習のためのマンガメタデータモデル

本章では、マンガに対して機械学習を脚本と構図の観点から適用するために新たに定義したマンガメタデータモデルについて提案する。提案モデルは、脚本をもとにしたマンガの構図推

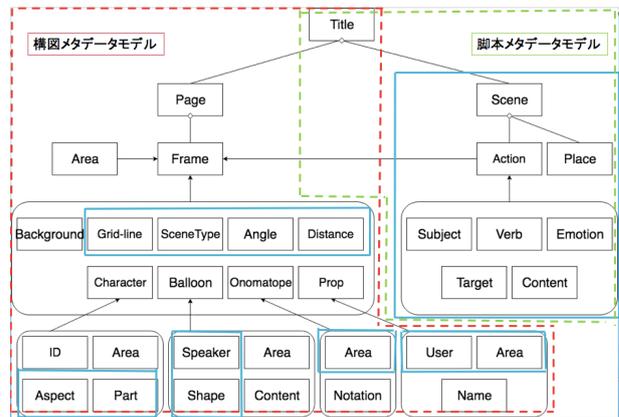


図2 提案マンガメタデータモデルの全体図

薦のための深層学習モデルの訓練に用いるアノテーションデータセット構築の基盤として利用する目的で定義した。

そのため、提案モデルはアノテーションデータ作成の上で重要となる時間的なコストを踏まえて、構図と脚本について関連性も含めて表現する上で必要となる要素をそれぞれ定義した。このモデルは先行研究[2]のモデルに対して、実験結果をフィードバックして定義したものである。

本研究では、「マンガ原作の書き方」[13]に基づいて、「脚本」を「だれが、何に、どんなことを、どんな気持ちでしているか」の観点から記述された複数の動作から構成されるものと定義し、「構図」については「イラスト・マンガ書き方ナビ」[14]から「一つのコマにおいて、視点や物体の大きさ、配置などから描かれるキャラクタや吹き出しなどをどのように表現するか」について示すものとして各要素について定義した。

本論文で提案するマンガメタデータモデルの全体像を図2に示す。

ここでは、図2の赤点線で囲まれた実体集合を構図メタデータモデル、緑点線で囲まれた実体集合を脚本メタデータモデルとして定義した。また青実線で囲まれた実体は、既存のマンガメタデータモデルでは定義されていない実体であり、本研究において脚本を表現し、構図において重要となる要素を表現するために新たに定義したものである。

各実体についての詳細及びアノテーションにおける基準は先行研究[2]と同様であるため、本章では新たに追加したOnomatope実体を具体化する「Area」と「Notation」実体の詳細についてのみ説明する。

先行研究[2]では、Onomatope実体は表記のみを記述するものとなっていたが、評価実験の結果からマンガでの自然さに大きな影響を与えること、マンガ制作経験者と未経験者間で最も差が出る要素であったことから、未経験者のマンガ制作支援においては、オノマトペを配置やサイズを指定する必要があると考え、Onomatope実体を具体化するArea実体とNotation実体を新たに加えた。このArea実体は、コマ内において対象のオノマトペの始点座標とそのサイズに関する情報を持っており、Notation実体はそのオノマトペの表記についての情報を持つ。



図 3 アノテーションツールの作業画面

#### 4. アノテーションデータセット

本章では、提案メタデータモデルをもとに構築したアノテーションデータセットについて述べる。

本研究では、脚本をもとにした構図推薦によるマンガ制作支援を目的としており、この構図推薦に利用する深層学習では、マンガ内で表現される内容を書き下した脚本を構造化して記述した脚本アノテーションとマンガ内で描かれているキャラクターなどの各要素について構造化して記述した構図アノテーションが組み合わさったデータセットが必要となる。しかし、先行研究のメタデータモデルとそのデータセットではこれらをカバーしたものは存在しない。そのため、提案マンガメタデータモデルをもとに、コマごとにマンガ内で表現される内容を描かれている動作を中心に脚本アノテーションを、構図を表現するキャラクターなどの各要素について構図アノテーションを付与したデータセットを構築した。

このアノテーションデータセットを構築するにあたり、アノテーションを付与する対象のマンガとして、「坊っちゃん」[16]、「銀河鉄道の夜」[17]、「蟹工船」[18]の三冊を選定した。この理由としては、将来的に原作からの脚本アノテーションの自動生成を見据え、古典文学を原作としてマンガ化されたものを選定した。

このマンガ三冊に対して、クラウドソーシングサービスである Lancers [15] において募集したマンガに親しみのある日本語話者にアノテーションデータを作成してもらった。また、このデータセットの構築に際し、広くアノテータを募集し、アノテーションデータを収集することを目的として、提案メタデータモデルを網羅した Web ベースのアノテーションツールを作成した。このツールの作業画面を図 3 に示す。

ここで今回構築したデータセットにおける各要素のデータ量について、表 1 に示す。

表 1 アノテーションデータセットの各項目のデータ量

	Page	Frame	Character	Balloon
データ量	556	2545	3782	2771
	Onoma	Prop	Action	
データ量	1048	1239	4784	

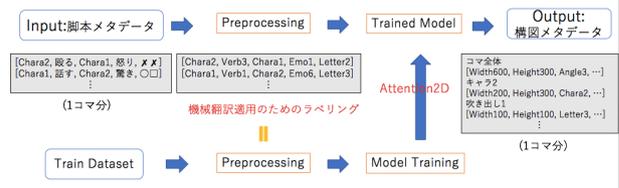


図 4 構図推薦の全体の流れ

#### 5. 深層学習を用いたマンガ構図推薦

本章では、提案メタデータモデルを元に構築したアノテーションデータセットで訓練した深層学習を利用した、脚本をもとにしたマンガ構図推薦手法について提案する。

本研究において取り組んだマンガ構図推薦の問題は、入力として可変長のマルチラベルを受け取り、それに対応する可変長のマルチラベルを予測して出力する問題として捉えることができ、これと同種の問題として機械翻訳のタスクがある。機械翻訳のタスクは、可変長の翻訳元文の単語列を受け取り、それに対応する可変長の目的言語の単語列を出力する問題であり、我々は翻訳元文として脚本アノテーションを、目的言語として構図アノテーションを与えることで、同様の手法で脚本を元にしたマンガ構図推薦の問題を解けると考えた。

関連研究においても述べたように、近年の機械翻訳の分野においてはニューラルネットワークを利用したニューラル機械翻訳が発展しており、これにより翻訳精度は年々目覚ましいスピードで向上している。そして、最近のニューラル機械翻訳における精度向上の一因として、画像処理の領域で目覚ましい成果を上げた畳み込みの仕組みの適用がある。

本研究では、この畳み込みの手法をニューラル機械翻訳に適用し、IWSLT2014 の翻訳タスクに置いて SoTA を達成した Attention2D [12] と呼ばれるモデルを用いて構図推薦を行った。ここで、この Attention2D を用いた構図推薦手法の全体の流れを図 4 に示す。

ここでは、提案手法の入出力の詳細について 5.1 節で説明し、提案手法において用いた Attention2D の概要について 5.2 節で説明する。また本提案手法では、複数のマンガのアノテーションデータを横断的に学習で利用できるようにし、またデータを集約してラベルの種類を減らすことで学習を上手く進めることを目的として、いくつかの前処理を行った。その詳細については、5.3 節で述べる。

##### 5.1 入出力

提案手法では、入力として一コマで表現したい内容を脚本メタデータモデルの Action 実体に従って書き下した文を与える。ここでは、一行が Action 実体に従って書かれた一つの動作を示し、表現したい内容を表現する動作の数に応じて行数が増える。Action 実体の表現形式としては、["Subject", "Verb", "Target", "Emotion", "Content"] の形式であり、各動作をこのフォーマットに従って記述したものを入力として与える。例えば、「Chara2 が怒って Chara1 を○○と言いながら殴る」と

いった内容を記述する場合には、[Chara2, 殴る, Chara1, 怒り, ○○]といった形の入力となる。

また、出力ではこの入力された構図アノテーションに応じて、構図メタデータモデルの Area, Grid-Line, Scene Type, Angle, Distance, Character, Balloon, Onomatope, Prop の各実体に従った形で表現された構図アノテーションが生成される。この出力では、Area, Grid-Line, Scene Type, Angle, Distance のコマ全体に関連する実体の情報についてはどの入力に対しても必ず一つずつ含んでおり、Character, Balloon, Onomatope, Prop の各要素に関する実体の情報については、入力に応じてその個数が増える。ここで、コマ全体に関連する実体の情報は、["Frame Width", "Frame Height", "Grid-Line", "Scen Type", "Angle", "Distance"] の形式で表現される。またキャラクタについての情報は各キャラクタごとに、["Width", "Height", "Start Point", "Character ID", "Part", "Aspect"], 吹き出しについての情報は各吹き出しごとに、["Width", "Height", "Start Point", "Speker ID", "Shape", "Content"] の形で表現される。そしてオノマトペについての情報は各オノマトペごとに、["Width", "Height", "Start Point", "Notation"], 小道具についての情報は各小道具ごとに、["Width", "Height", "Start Point", "Prop Name", "User ID"] の形で表現される。Start Point はコマの始点座標からの距離座標を示すものであり、各 ID は登場人物に割り当てられた ID を示す。

## 5.2 Attention2D

本節では、本研究において構図推薦に利用した Attention2D [12] について概説する。このモデルは、従来の Seq2Seq に代表されるようなエンコーダとデコーダを用いたニューラル機械翻訳のモデルと比較して、大きく異なる 2 つの特徴を持っている。

1 つ目は、従来のニューラル機械翻訳ではエンコーダとデコーダの 2 つを用いる構造がデファクト・スタンダードになっているのに対し、このモデルでは CNN の進化系である DenseNet [11] を用いることでデコーダの機構だけで機械翻訳を実現している。この機構によって訓練時の生成においてすべての単語をまとめて生成できるため、学習時間が少なく済むといったメリットが有る。

2 つ目は、このモデルはアテンション機構を持たず、入力・出力の系列をまとめて二次元の CNN を用いて畳み込むことで、アテンション機構を擬似的に実現しているという特徴がある。従来のニューラル機械翻訳では、アテンションは最初にエンコーダで計算したものを繰り返し用いているため、デコーダの出力に応じてアテンションの特徴量を変更することはできない。しかし、このモデルではデコーダで生成されたラベルによって Attention の特徴量を再計算するため、全体でより自然な文章が生成されやすいといった特徴がある。

今回、アノテーションデータセットを用いてマンガ構図推薦を行うに当たり、この Attention2D を活用する理由として上記の特徴に係る 2 つの理由がある。

1 つ目の理由は、ラベルの出現位置の情報を入力として組み込むことである。今回提案する構図推薦手法では、脚本ア

ノテーションと構図アノテーションの関係をコマ単位で学習させる。ここでは脚本アノテーションをソース文、構図アノテーションをターゲット文としており、それぞれの一文が各コマの持つ脚本情報と構図情報を持っている。そのため、ソース文はコマ内で表現されている各行動についての情報を連結した形で、ターゲット文はコマの構図を表現する各要素が連結した形で記述されている。つまり、今回のデータにおいてはソース文では各動作ごとに、ターゲット文はコマ、キャラなどの各要素の単位で順番が固定であるため、出現位置の情報がモデルに組み込まれることで予測精度が向上すると考えられる。

2 つ目の理由は、Attention2D では各ラベルを生成する際に、それまでの出力に応じたアテンションの値を利用できることである。これまでの Seq2Seq を始めとするアテンション機構では、生成を行う前の段階でアテンションの値の計算はすでに完了しており、生成済みの単語に関係なく、これを各デコーダに入力として使いまわして与えるのが一般的な形であった。しかし、この仕組みではそれまでに生成された単語を考慮した上で関連性の高い単語に着目することができないため、文章全体の自然さを保つことが難しい。それに対して Attention2D では、生成済みのラベルを考慮した Attention の値を用いるため、全体で自然な文章が生成されやすい。今回の出力である構図アノテーションに置いては、例えば出力されたコマのサイズに応じてキャラのサイズが変わる、キャラの位置に応じてセリフの位置が変化するという形で、前後のレイヤー間のラベルに強い相関があるため、この Attention2D の特徴が有用であると考えた。

また本モデルの利用においては、いくつかのハイパーパラメータの設定が必要となるが、Elbayad らが論文中で各ハイパーパラメータの探索結果を提示しており、本研究ではこの中で最も高い翻訳精度を達成した、embedding size = 128, layer num = 24, growth rate = 3 を構図推薦の学習において採用した。

## 5.3 前処理

そして、今回データセットを深層学習の訓練に用いるに際して、アノテーションデータセットのデータをそのまま用いるとデータ量に対してラベルの種類が多すぎるため、学習が進まない点や各キャラクタ ID が示すキャラクタがマンガに応じて異なるため、複数のマンガを横断的に学習に利用できないといった問題があった。そこで本提案手法では、複数のマンガのアノテーションデータを横断的に学習で利用できるようにし、またデータを集約してラベルの種類を減らすことで学習を上手く進めることを目的として、いくつかの前処理を行った。まず、複数のマンガのアノテーションデータを横断的に扱う上で最も問題となる同じキャラクタ ID が示すキャラクタがマンガに応じて変化する点に対しては、各マンガの脚本アノテーションデータにおけるキャラクタの出現頻度をもとにキャラクタの ID に対して重要度付けを行い、それをもとにラベル付を行う前処理によって、横断的にキャラクタを扱えるようにした。そして、データ量が少ない中で学習を上手く進めるために、脚本アノテーションデータにおける動作を示す動詞については、

表 2 WER の結果

	全体	コマ	キャラ	セリフ	オノマトベ	小道具
WER	21.7%	17.5%	29.1%	20.0%	16.4%	21.5%

Wikipedia で学習済みの Word2Vec [19] を利用して各動詞をベクトル化し、そのベクトル化された各動詞を Kmeans [20] によって 20 クラスに分類し、そのクラスに応じたラベル付を行ったほか、吹き出しやオノマトベの文字を含む実体については、文字数に応じて 3 文字ごとにクラス分けを行い、そのクラスに応じてラベル付を行った。また構図アノテーションデータ特有の各要素の領域に関する情報については、それぞれ 25 ピクセルごとに近い方に寄せてラベル付することで、ラベルの集約を行った。これらの前処理によって、ニューラル機械翻訳の学習モデルを脚本・構図アノテーションデータセットに対して適用できるようにしている。

## 6. 評価実験

本章では、提案マンガメタデータモデルをもとに提案した構図推薦手法に対して行った定量評価と定性評価について説明する。本研究では、Attention2D の学習自体がどの程度うまく行えたかを確認するために、定量評価として Word Error Rate (WER) による精度評価を行った。この実験詳細と結果について、6.1 節において説明する。また、提案手法の構図推薦により未経験者のマンガ制作を支援できるか確認するために、実際に生成された構図アノテーションをもとに 4 名の被験者にコミ Po! を用いてマンガを制作してもらい、それを別の 6 名の被験者に分かりやすさと自然さの観点及びコマ同士の比較により評価してもらった。この実験詳細と結果について、6.2 節において説明する。

### 6.1 定量評価

本研究の構図推薦手法に利用した Attention2D のモデルによるアノテーションデータセットの学習がどの程度上手く行えたかを確認するために、Word Error Rate (WER) による精度評価を行った。この評価実験では、学習に利用していない 100 コマ分のテストデータを用いて行った。この WER は機械翻訳の評価指標としてよく用いられる指標で、以下のように定義される。

$$WER = \frac{S + D + I}{N} \quad (1)$$

ここで、S, D, I は、それぞれ出力文を正解文と比較したときに、正解文に近づけるために置換した単語の数、削除した単語の数、挿入した単語の数を表し、N は正解文の総単語数を示す。

本実験では、コマごとの構図に関する全要素に対して、及びコマごとの構図に関する各要素に対して WER を算出し、その平均を求めた。この結果を表 2 に示す。

上記の結果は、近年のニューラル機械翻訳による WER と比較するとやや高い結果となったが、実用できる範囲内の精度が得られたと考える。また、今回近年のニューラル機械翻訳による WER よりやや高い値となった理由としては、データセットの量的不足が大きいと考える。ニューラル機械翻訳においてよく

用いられるデータセットとして WMT'14 (De - EN) があるが、これは約 500 万ペアの文から構成されている。ニューラル機械翻訳の分野においては、これ以上の量を含んだデータセットが用いられることが多いが、これらのデータセットではその分未知語を含め単語数が多い。それに対し、我々のデータセットはペアの数は少ないものの、単語数としては固定のラベル数しかなく、またこれも提案手法の前処理において数をできる限り小さくしているため、大幅に WER が上昇しなかったと考える。また、実際の誤予測を詳細に見ると、各要素の持つラベルの位置や個数といった観点でのミスはなく、同じ属性のラベル内での予測誤りが多く見受けられたため、データセットの量を増やしていくことで WER は改善されるものと考えられる。また今回の WER による評価では、キャラクタにおける値が一番悪かったが、これは一部の出力において同じキャラクタを示す羅列を上限まで繰り返し生成していたものがあつたことが原因である。今回のモデルでは、全体に対する生成ラベルの上限は組み込んでいたが、要素単位での生成上限を設ける機構がなかったため、この機構を設けることによって WER は改善されると考える。

### 6.2 定性評価

本研究の提案構図推薦手法を利用することで、どの程度素人を含めた制作者のマンガ制作を支援できるかを評価するため、以下のような定性評価を行った。今回の評価実験では、マンガ制作経験者 2 名 A, B と未経験者 2 名 C, D に各コマに用意されたアノテーションをもとに作成してもらった。ここでは、アノテーションデータセットに含まれている 3 冊のマンガからそれぞれ 10 コマ、計 30 コマを利用し、この 30 コマを 15 コマずつの 2 つのグループ X, Y に分けた。そして、制作者を A と C からなる Team1 と B と D からなる Team2 の 2 チームに分け、Team1 には X グループのコマのそれまでの状況説明と脚本アノテーション、及び Y グループのコマのそれまでの状況説明と脚本アノテーションに加えて、脚本アノテーションから提案手法により生成された構図アノテーションを割当て、もう一方の Team2 には X と Y で与えるものを入れ替えたものを割り当てた。そして、各チームの制作者それぞれがコミ Po! を用いて、状況描写と脚本アノテーションのみの場合はそれをもとに想像力だけで、構図アノテーションがある場合については基本的にはそれに従って作成してもらった。ただし、今回の提案手法の目的はマンガの制作支援であるため、構図アノテーションに従う上で明らかに表現がおかしくなるものについては修正、削除を、表現が足りないものについては追加を許可している。

次に、上記のようにして作成された合計 120 コマに対して、別の被験者 6 名によりそれぞれ 3 つの評価を行ってもらった。この評価においては、それぞれ制作者が経験者か未経験者かは伏せて行った。1 つ目の評価実験では、制作された各コマが状況描写と比較して内容をわかりやすく表現できているかを ○, △, X の三段階で評価してもらった。2 つ目の評価実験では、制作された各コマが自然かどうかを自然、どちらともいえない、不自然の三段階で評価してもらった。3 つ目の評価実験では、異なる制作者による同じ状況描写について表現したコマを比較

表 3 Team1 により制作されたコマに対する分かりやすさの評価

	A(WithC)	A(Without)	C(WithC)	C(Without)
○	80	82	60	46
△	7	7	20	31
X	3	1	10	13

表 4 Team2 により制作されたコマに対する分かりやすさの評価

	B(WithC)	B(Without)	D(WithC)	D(Without)
○	68	69	64	58
△	19	20	23	16
X	3	1	3	6

表 5 Team1 により制作されたコマに対する自然さの評価

	A(WithC)	A(Without)	C(WithC)	C(Without)
自然	70	77	38	24
どちらでも	18	12	43	49
不自然	2	1	9	17

表 6 Team2 により制作されたコマに対する自然さの評価

	B(WithC)	B(Without)	D(WithC)	D(Without)
自然	56	65	47	40
どちらでも	32	23	35	38
不自然	2	2	8	12

表 7 グループ X のコマにおける制作者ごとの比較評価結果

	A(WithC)	B(Without)	C(WithC)	D(Without)
A(WithC)	-	45	63	50
B(Without)	45	-	61	53
C(WithC)	26	29	-	40
D(Without)	38	36	50	-

表 8 グループ Y のコマにおける制作者ごとの比較評価結果

	A(Without)	B(WithC)	C(Without)	D(WithC)
A(Without)	-	55	82	49
B(WithC)	35	-	67	51
C(Without)	7	22	-	30
D(WithC)	40	38	59	-

し、分かりやすさと自然さを総合的に加味してどちらがよいか選択してもらった。

1 つ目の実験の結果について、それぞれの作者について構図推薦ありとなしに分けて、獲得した評価値の数を表 3 と表 4 に示す。2 つ目の実験の結果について、それぞれの作者について構図推薦ありとなしに分けて、獲得した評価値の数を表 5 と表 6 に示す。3 つ目の実験の結果について、それぞれの作者の組み合わせについて構図推薦ありとなしに分けた場合のより良いと評価された数を表 7 と表 8 に示す。これらの表 7 と表 8 では、ある行の作者の制作したコマが各列の作者の制作したものと比較してより良いと評価された数を示しているここで表中の (WithC) は構図推薦ありを意味し、(Without) は構図推薦なしを意味する。

上記の表 3 と表 4 より、分かりやすさの観点については、A、B のマンガ経験者の場合は、構図推薦をベースとして制作されたコマについても、構図推薦なしで制作されたものと同等の評価を得ており、またマンガ制作経験のない C、D については、構図推薦をベースとして制作されたコマのほうが高い評価をより多く得ている。また表 5 と表 6 より、自然さの観点については、A、B のマンガ経験者については、構図推薦をベースとして制作されたコマのほうが、構図推薦なしで制作されたコマよりも評価がやや低いものの、分かりやすさと同様にマンガ制作経験のない C、D については、構図推薦をベースとして制作されたコマのほうが高い評価をより多く得ている。

そして、表 7 と表 8 の結果から、制作されたコマを比較するとマンガ制作経験者 A、B の場合は、構図推薦なしで自作したもののほうが良い出来のものが多いことが見て取れる。しかし、未経験者の C、D の場合では、構図推薦をベースとして制作した場合の方が比較においてよい評価を多く得た。以上より、品質としてはマンガ経験者の作成するものにやや劣るものの、未経験者のマンガ制作の手助けとして十分に役立つことが示せた。

## 7. まとめと今後の課題

本論文では、機械学習に向けたマンガメタデータモデルを提案し、それを元に構築されたデータセットを利用した深層学習による構図推薦手法について提案した。提案メタデータモデルは、構図の観点からマンガ表現において内容を適切に伝えるための画像情報と、脚本の観点から表現したい内容を言語情報を結びつけ、表現したものである。ここでは、これらの情報とともに機械学習において扱えるようにした上で、データセットを構築する上で欠かせないアノテーションコストの観点も考慮し、マンガのコマにおける各要素とそのコマに対する脚本的要素を定義した。そして、提案メタデータモデルをもとにクラウドソーシングサービスを利用して、アノテーションデータセットを構築し、このデータセットをもとにニューラル機械翻訳の技術を利用して、アノテーションデータセットの各コマにおける動作を示す脚本アノテーションを入力文として、それに対応する構図アノテーションをペアとしてモデルの学習を行い、この学習済みモデルを利用した構図推薦を実現した。

この構図推薦手法の評価として、WER を用いた定量評価と推薦された構図をもとに実際に制作されたコマと想像をもとに制作されたコマの比較を行うことで本手法の定性評価を行った。WER による定量評価では、データセットの量的不足により近年の SoTA ほどの精度は達成できなかったものの現実的に応用可能なレベルの精度を得ることができた。また、定性評価においてもこの構図推薦手法を用いることで、我々が目標としていたマンガ制作未経験者によるマンガ制作の支援の目的が達成できていることを確認できた。

今後の課題としては、本論文ではデータセットの量的不足によりコマ単位での構図推薦に留まっているが、実際にマンガを制作していく上ではページ全体を考慮した上での構図を考える必要がある。そのためにはデータセットの充実が不可欠であるが、より大規模なアノテーションデータセットを構築するには

まだまだアノテーションコストが大きい。そのため、今後はマンガにおける画像認識の精度を向上させることで、マンガにおけるラベル付けを容易にし、アノテーションデータセットをより簡易に構築できるようにしていくことも考えていく必要があるだろう。

## 謝 辞

本研究は JSPS 科研費基盤研究 (C) No. 15K00423, No. 17K01590, 新学術領域研究 (研究領域提案型) No. 18H05031 の助成を受けたものです。ここで心より感謝申し上げます。

## 文 献

- [1] “コミ Po!” <https://www.comipo.com/>
- [2] 伊藤 博典, 浅野 泰仁. “マンガ制作支援に向けたメタデータモデル” 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM2018).
- [3] 中澤 敏明. “機械翻訳の新しいパラダイム:ニューラル機械翻訳の原理.” 情報管理, Vol. 60, No. 5, pp. 299-306 (2017).
- [4] Ayako Morozumi et al. “Metadata Framework for Manga: A Multi-paradigm Metadata Description Framework for Digital Comics.” Proceedings of DC-2009, pp. 61–70, 2009.
- [5] “Functional Requirements for Bibliographic Records Final Report.” International Federation of Library Associations and Institutions, 1998. <http://www.ifla.org/VII/s13/frbr/frbr.pdf>
- [6] TV-Anytime Forum. “Welcome to the TV-Anytime website.” <http://www.tv-anytime.org/>
- [7] 両角彩子, 永森光晴, 杉本重雄. “ストーリーの知的内容を表すメタデータ記述項目の提案:Wikipedia 上のマンガ・小説作品記事を対象として.” 情報処理学会研究報告, Vol. 2008-FI-92, No. 105, pp. 1-14, 2008.
- [8] Azuma Fujimoto et al. “Manga109 Dataset and Creation of Metadata.” Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding(MAMPU'16), Article No. 2, 2016.
- [9] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. “Sequence to sequence learning with neural networks.” Advances in neural information processing systems. p. 3104-3112 (2014).
- [10] Krizhevsky, A., Sutskever, I. and Hinton, G. E. “Imagenet classification with deep convolutional neural networks.” Advances in neural information processing systems. p. 1097-1105 (2012).
- [11] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q. “Densely connected convolutional networks.” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 2261-2269 (2017).
- [12] Elbayad, M., Besacier, L. and Verbeek, J. “Pervasive attention: 2d convolutional neural networks for sequence-to-sequence prediction.” arXiv preprint arXiv:1808.03867 (2018).
- [13] 大石 賢一. マンガ原作の書き方. 言視社. 2012.
- [14] CLIP STUDIO. イラスト・マンガ書き方ナビ. <https://www.clipstudio.net/oekaki/archives/1517>
- [15] “Lancers”, <https://www.lancers.jp/>
- [16] Team バンミカス, 夏目 漱石. 坊っちゃん, イースト・プレス. (2016)
- [17] バラエティ・アートワークス, 宮沢 賢治. 銀河鉄道の夜, イースト・プレス. (2007)
- [18] バラエティ・アートワークス, 小林多喜二. 蟹工船, イースト・プレス. (2007)
- [19] 鈴木 正敏. 日本語 Wikipedia エンティティベクトル. <http://www.cl.ecei.tohoku.ac.jp/m-suzuki/jawikivector>

- [20] Hartigan, J. A. and Wong, M. A. “Algorithm AS 136: A k-means clustering algorithm.” Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 28, No. 1, pp. 100-108 (1979)