一般社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

HCGシンポジウム2018
HUMAN COMMUNICATION GROUP SYMPOSIUM 2018
HCG2018-B-5-2

Fine-Tuned VGG モデルを用いた表紙画像からのコミックの内容推定

朴　　炳宣†　　松下　光範†

† 関西大学大学院　〒 569–1052 大阪府高槻市霊仙寺町 2–1–1
E-mail: †{k281401,t080164}@kansai-u.ac.jp

あらまし　従来，コミックの内容を把握するためのリソースには，コミックそのものやレビューから抽出できる情報が用いられてきた．しかし，これらのリソースはテキストや画像が混交しているために詳細な分析が難しく，抽出できる情報に偏りがあるといった問題があった．この問題を解消するため，本稿ではコミックの表紙に着目し，表紙から検索に利用可能な情報を抽出する．提案手法では，VGG-16 モデルに Fine-Tuning を行った推論モデルを用いて，表紙に描かれた作品の世界観を推定する．更に，得られた意味ベクトルやタグに基づき各コミックを関連づける．実験の結果，推論モデルの精度は 0.69 であり，関連付けられたコミックには時代や文化的背景の類似性が示唆された．

キーワード　内容推定，転移学習，コミック工学

Estimating Comic Content from The Book Cover Information Using Fine-Tuned VGG Model

Byeongseon PARK† and Mitsunori MATSUSHITA†

† Graduate School of Informatics, Kansai University　2-1-1, Reizanji-cho, Takatsuki-shi, Osaka, 569–1052 Japan
E-mail: †{k281401,t080164}@kansai-u.ac.jp

Abstract　Resources for extracting the contents information of existing comics were only the comics itself and review. However, these resources have drawbacks that they can not sufficiently extract information necessary for searching, and contain a lot of unnecessary information. To solve the problem, this paper proposes an estimation method by using the book cover images of comics as an additional resource. The proposed method estimates the “Background” of comics drawn on the book cover of comics from fine-tuned VGG-16 model, then associates comics based on the semantic vectors and tags obtained. Our experiment revealed that the accuracy of the model was 0.69, and unity in the comics related were observed by this method.

Key words　Content Estimation, Transfer Learning, Comic Computing

1. はじめに

出版月報 2018 年 2 月号（公益財団法人全国出版協会・出版科学研究所発行）によれば，2017 年度に出版された新刊コミックの点数は 12,461 点であった．こうした莫大な量のコミックの中からユーザの興味に

沿った内容の作品を探し出すことは困難である．このような問題を解決するために，我々はコミック探索システム [1] を開発した．本システムは各コミックの内容を把握し，ユーザの興味に応じてそれらを取捨選択し提示することで，ユーザの情報アクセスを支援する．このシステムを用いることで，ユーザは自らの嗜好に合致した作品にアクセスができるようになった．しかし，現状では内容に基づく検索を実現するための内容情報を十分に抽出することは難しい．そこで，本研究では，このような問題を解決するためにコミックの内容情報を簡易的に把握する情報源として表紙を用いる．

コミックの表紙は，ユーザは読んだことのない本を購入する際，重要な役割を持つ．なぜならコミックの表紙は，それを見た人がコミックの内容を一目で理解するための高度な情報が含んでいるためである．例えば，あるユーザがコミックを選ぶために図 1 にある読んだことのない 3 つの作品の表紙を見ているとする．ユーザは図 1a と図 1b のコミックの内容を知らなくとも，表紙の中のキャラクターの服装や持ち物 (e.g., 図 1a の場合は “ローブ” と “剣”), 図 1b の場合は “制服” と “スーツ”) を見ることで，二つのコミックが違う時代や文化に基づく題材（以下，世界観とする）を扱っていることが明確にわかる．一方で，図 1c の表紙は，図 1b と題材は異なるものの、似ている世界観に基づいていることが分かる．このように，ユーザーは表紙に描かれた情報を読み取ることで，本文の内容を推測し，自分の趣向に合うコミックの選別を行う情報として用いる．

私たちはこのような表紙の役割を用いることにより，ユーザが読んだことのないコミックを探す際に必要な内容情報をより容易に取得できる点について着目した．本稿では，コミックの表紙に含まれた情報を抽出し，類似した世界観を持つ同士を関連づける手法について提案する．提案手法では，コミックの表紙から VGG-16 モデル [11] に基づく転移学習を行うことで表紙に描かれた情報 (e.g., 服装, 持ち物, 舞台) を取得する．また，得られた情報から得られた意味ベクトルやタグに基づき，各コミックのを関連づける．

2. 表紙情報

読者がコミックの表紙から読み取ることができる情報は下記のように大別できる．

絵柄
　作者の外的表現記法
ストーリー
　表紙に描かれた人物の行動や関係性による簡略的で暗示的シナリオ
世界観
　作品の基盤となる時代や文化

まず “絵柄” 情報は，作者がどのような記法を用いて作品を表現しているかに関する情報を指す．例えば、図 1c は他の作品よりも使われている色の種類が少ない上にデフォルメ化された人物の書き方をしていることがわかる．絵柄情報は読者がまだ読んだことのないコミックへの第 1 印象を決定づける上で最も直感的な領域に関わる役割を持つ．また，“ストーリー” 情報は，表紙によって表現された短編的かつ暗示的な本編の内容に関する情報を指す．例えば，図 1b と 1c の作品のキャラクターは，温厚な表情を浮かべている上に両隣に並んでいることから，友好的な関係が伺える．一方で図 1a のキャラクターは剣を持って緊張している様子から，女性を守っているよう見える．このように，読者は表紙に含まれた情報からコミックのストーリーを短編的に推測できる．最後に，“世界観” 情報は作品のストーリーの題材となる時代的文化的背景を指す．例えば，図 1a のキャラクターはローブをつけていたり，剣を持っているなど，西洋の中世時代の文化に類似した服装をしている．一方で図 1b と 1c のキャラクターは現代の制服やスーツ，T シャツなど，現代の文化に近い服装をしている．これらの要素によって，ユーザは内容を知らないのコミックのであっても，そのコミックのストーリーが描かれる舞台を推測し，作品の世界観を捉えることができる．

このような表紙に含まれた情報を全て活用することで，読者のコミック検索に用いる内容情報をより容易に抽出することが可能となる．本稿では，上記の 3 つの要素の中でも世界観に関する情報を抽出し，コミックを関連づける手法について考察する．世界観情報は他の要素よりも「知らない作品を探す」という状況において，情報量の個人差が少ない要素であると考えられるため，検索システムへの有効な活用が期待できる．

(注1)：©Yuzuru Shimazaki, Kodansha Ltd.
(注2)：©Ken Yagami, Kadokawa Publishing Ltd.
(注3)：©Yuka Kuniki, Takeshobo Ltd.

(a) 征神記ヴァルナス (注1)

(b) ありさ [2] (注2)

(c) エヴリデイおさかなちゃん (注3)

図 1: コミックの表紙の例

3. 関連研究

コミックにおける情報抽出に関する試みとして，直接的抽出手法と間接的抽出手法が存在する．Rigaud らは，コミックに含まれた要素を理解できるエキスパートシステムを提案した [2]．彼らのエキスパートシステムは、簡単な要素から複雑な要素に移って、画像の内容を徐々に理解するために、低レベル（画像処理）と繰り返し相互作用することで，コミックの画像からコマや吹き出し，キャラクターの同定を実現している．Rigaud らの他にも，コミックに含まれた要素を同定するための手法に関する試みが存在する [3]〜[5]．しかし，これらの研究は，まだコミックに描かれた要素の同定にのみ注目しており，コミックの内容を把握するための段階には至っていない．そのため，現状ではコミックの内容を検索に応用可能な情報 (e.g., ストーリーの要約文, ストーリーに基づく情報の構造化) として活用することは困難である．

このような現状の代案として，Park らはコミックの検索に応用可能な内容情報のリソースとしてレビューを用いることで，レビュー作成者による作品のあらすじや評価のような間接的な内容情報を抽出することを試みた [1]．これにより，ユーザはコミックの内容や評価に基づき趣向に適したコミックを検索することができる．しかし，レビューは作成者によってレビューに含まれる情報量が大きく異なる場合が多い上に，必ずしもコミックの内容に直接関わる情報 (e.g., ストーリーやキャラクターの説明) を詳細に含めているとは限らない．

これらのような点を踏まえ，本研究ではコミックの内容情報の新たなリソースとして表紙を用いる．2. 章で述べたように，表紙にはコミックの内容の概要を把握できるための情報が集約されている．つまり，1 枚の絵に対する分析を行うだけでストーリーの概要を把握できる上に，レビューのような間接的情報よりもノイズが少ないといった長点がある考えられる．よって，本稿では表紙に含まれた情報の中でも検索への有効な活用が期待できる世界観情報の推定を行う．

4. 提案手法

本研究では,「コミックの表紙に表現された世界観には，あるテーマ (e.g., 家族, 恋愛, SF) を持っている」という仮説を元に，表紙に含まれた情報の抽出を行った．図 1 のように，表紙に含まれた世界観は，キャラクターの衣服や持ち物，ストーリーの舞台によって表現される．これらの要素の組み合わせを把握することにより，各コミックがどのようなテーマを持っているかを推定できると考える．

さらに我々は，抽出した情報からトピックの推定を行った．トピックとは，スポーツ・政治・音楽といったテキストの内容を指す．また，あるテキストに含まれたトピックが単一であるとは限らず，多重で存在する場合が存在する [8]．このようなトピックの特徴を踏まえ，表紙の中のトピックを推定するアプローチとして用いる．そして，トピックの分布パターンが類似するコミック同士は世界観が類似していると考える．

また，コミックの表紙は各作品ごとに 1 つの表紙とは限らない．本稿では，あるコミックの表紙が複数存在する場合，すべての表紙に存在するタグを一つの作品を表す集合とし，作品単位でトピックを推定する．さらに表紙の数の異なる二冊のコミックのトピックの推移を比較可能な関数に基づいて，コミック同士の世界観の類似度を算出することとした．

4.1 コミックの表紙からの情報抽出

4.1.1 モデル設計

人によって描かれた画像に対する情報抽出の成果として，Saito らによる Illustration2Vec [12] が挙げられる．Saito らは画像の意味をベクトル化して解釈し，

意味に基づいた画像検索を実現するために， Network In Network (NIN) モデル [10] と VGG [11] モデルを組み合わせた deep convolutional neural networks (CNNs) [6] モデルを提案した．Saito らの方法では、入力されたイラストを 4,096 次元のベクトルに写像し，イラストを表す特徴空間を構築するために，イラストからバイナリ属性（タグ）を予測するための CNN モデルを訓練する．

Saito らが主に学習対象とした画像は Web 上に自由に投稿されたイラストであるが，これらの画像は本稿の対象となるコミックの表紙の特徴 (e.g., 文字が含まれている, 余白が存在する, 色が単調) と大きく異なるものが多い．そのため Saito らのモデルを再学習する必要があるが，CNN の特性上，学習には大量のデータが必要となる．しかし，コミックの表紙は Saito らが学習データ（インターネット上のイラスト）とは異なり，限られた数しか存在しない．

そこで，本稿では Pre-trained network を用いた Fine-tuning を行う．Fine-tuning とは，Pre-trained network の識別層だけを対象タスクのものに付け替え，その他の階層は学習済みのパラメータを初期値として用い，誤差逆伝搬法による学習を進める方法であり，Pre-trained network の応用方法として広く使われている．Fine-tuning は Pre-trained network の豊富な特徴量を用いることで，少ないデータからの再学習であっても，高い精度を得られることという長点を持つ．

このような特性を持つ Fine-tuning では，元のモデル自体の性能も最終的なネットワークの精度に大きな影響を与える．そのため，ImageNet [9] の 1000 クラスのデータを用いたコンペティション型ワークショップである ImageNet Large-scale Visual Recognition Challenge (ILSVRC) で発表された Oxford Visual Geometry Group の 16 層 CNN (VGG-16) [11] がよく用いられる．このような例に習い，本稿では，VGG-16 の識別層をデータ収集によって得られたタグから構成されたベクトルを出力するための識別層に置き換え，表紙の画像から再学習を行うことで，コミックの表紙に含まれた情報を抽出する．

4.1.2 データ収集

本稿では VGG-16 モデルを用いた Finue-tuning 学習を行う．学習データとして，コミックの表紙の画像と，各画像に付与されたタグを用いる．コミックの表紙における既存の情報はコミックのタイトルや著者名といった書籍情報しか存在しないため，画像に描かれた世界観に関連する情報 (e.g., キャラクターの衣服や持ち物，ストーリーの舞台) については新たにタグを付与しなければいけない．我々は任意に選定したコミック 100 作品の 536 枚の表紙に対して，下記の要素から構成されたタグを作成した．

名称　オブジェクトの名称

世界観クラス　オブジェクトの小区分

補足情報　オブジェクトの補足情報

「名称」要素とは，表紙に描かれたオブジェクトに対する名称を指す．コミックの世界観を表現するための情報を作成するにあたり，各オブジェクトの名称の統一性が失われないように,「名称」要素には，固有表現 (e.g., 地名, 商品名) や修飾語 (e.g., 色，大きさ, 形状) を全て排除した名称のみを記入した．例えば，デニムパンツのようなオブジェクトがあった場合，パンツのみ名称として記入している.「世界観クラス」とは，同じ名称を持つオブジェクトを区別するための小区分を指す．例えば，表紙に「城」が書かれいる場合,「城」は国や時代によって様式が異なる場合が存在する (e.g., ノイシュヴァンシュタイン城と大阪城)．そのため，我々はこのようなケースであってもオブジェクトを区別できるように，(1) 現代, (2) 西洋, (3) 東洋, (4) 近未来, という 4 つのクラスを「世界観クラス」として設けた．最後に,「補足情報」要素とは，オブジェクトに対する補足情報を指し,「名称」要素で省略された情報を記録するための要素である.「補足情報」要素を設けることによって,「名称」要素で省略された情報を任意に拡張することが可能となる．これらの規則に基づき作成したタグの例として，“デニムパンツ” の場合 “パンツ:1:デニム”, “大阪城” の場合，“城:3:” がある．

データ作成の結果，上記の規則に基づき作成したタグは合計 560 種類であった．学習に用いるためのデータとして，あまりに希少なタグはモデルの精度に影響を及ぼす可能性があると考えられるため，本稿では全ての表紙における出現頻度が 5 以上となる上位 100 位までのタグを使用した．

4.1.3 実　　験

データ作成後，得られたデータから VGG-16 モデルを用いた Fine-tuning を行った．学習を行う際，表紙の画像データ (合計 536 枚) の 20% (107 枚) を評価データとして用いた．CNN においてデータの数は精度と比例するため，本来用いるデータの規模は大きい

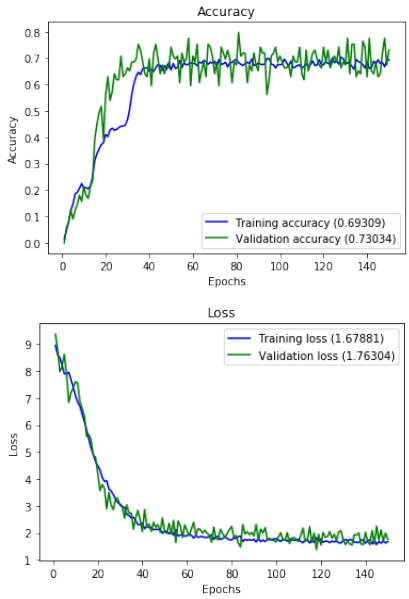

図 2: VGG-16 に基づくタグ推定学習の結果

ことが望ましい．そのため，我々は学習時に学習データのみ (1) 上下反転，(2) 左右反転，(3) 90°回転，(4) 270°回転，(5)Enhanced Edge フィルタを適用，(6) 強度な Enhanced Edge フィルタを適用，(7) Enhanced Edge フィルタによる輪郭のみを使用，といった差分データを合計 7 種類の作成し，学習に用いた．また，特定の画像やタグに偏って学習されることを防ぐために，Epoch ごとに学習データと評価データを任意に再選別した．Äś 上記の手順により学習を行った結果を Fig 2 に示す．学習の結果，モデルの精度は 0.693，loss score は 1.679 で収束した．また，学習したモデルから得られるベクトルの各タグに対する確信度の上位にあるタグと正解データを比較した結果，正解データに対する再現度は 0.918 であった．

4.2 コミックの関連付け

Saito らは CNN モデルから得られた 4,096 次元のベクトルを比較することで，イラストの意味における類似度を計算可能とした．本稿で提案したモデルを用いることで，世界観情報を表現する 100 次元のベクトルからコミックの表紙の間の類似度を測ることも可能であると考える．しかし，Saito らの手法と比べ，我々の手法で用いられたベクトルの次元数は顕著に低い．よって，ベクトル同士の類似度を比較する際，特定のタグが及ぼす影響が大きくなることにより，タグの組み合わせによる多様性が失われる恐れがある．そこで，本稿ではモデルによって得られたベクトルをタグに変換し，得られたタグから Latent Dirichlet Allocation (LDA) [7] に基づきトピック分類を行う．

LDA とは，文書中の単語は独立に出現しているのではなく，潜在的なトピックに基づいて出現するという仮定に基づいた文書生成モデルである [7]．なお，LDA では，各文書は複数のトピックで構成されており，各トピックの単語分布を合算した形で単語が生成されていると仮定しているアルゴリズムであるため，文書に含まれた複数のトピックを推定・把握する．

コミックは小説や映画など他の創作物と同様，様々な要素からストーリーが構成されるので，コミックの世界観は必ずしも一つのトピックで表現できるとは限らない．本稿では，LDA を用いてコミックの表紙に含まれた複数のトピックを推定することで，タグの種類に関わらずより容易に世界観を推定することを目指す．

LDA によるトピック分類を行うために，提案したモデルにより得られたベクトルから確信度が 0.1 以上の単語をタグとして抽出した．確信度の閾値は，全ての表紙が 1 つ以上のタグを持つ状態の数値を用いている．各表紙のタグを作品ごとに結合し，100 コミックからトピック分類を行った．また，LDA におけるトピック数 K は，トピック生成結果から単語の分布を確認し各トピックの内容が最も推測しやすいと定性的に判断した際の数値である $K = 10$ を用いている．生成されたトピックの単語の例を表 1 に示す．

4.2.1 類似度の算出

テキスト同士の類似度の算出については様々な手法が考えられるが [13]，本稿ではトピックの分布の類似性に着目しているため，バタチャリア係数 (Bhattacharyya Coefficient) [14] を用いることとした．この手法では，総頻度が正規化され，同じ数のビンに分割されたふたつのヒストグラムの類似度を，対応するビン中の頻度の積を求めることでそれらの類似度を算出する．個数 n のビンに分割されたヒストグラム P および Q の類似度 $s(P, Q)$ は，

$$s(P, Q) = \sum_{i=1}^{n} \sqrt{P_i Q_i} \tag{1}$$

となる．ここで，P_i，Q_i は各々，ヒストグラム P，Q の i 番目のビンの頻度である．本研究では，作品単位でトピックを推定し，それらにバタチャリア係数を適用して類似度を算出する．本稿で実装したプロトタイプは，瓶の数をトピック数（$K = 10$）として式 (1) で

表 1: トピック分類結果 (上位 5 語)

Topic	Words
Topic 0	Jacket:2:, Gun:2:, Combat-clothing:2:, Hair-ornament:2:Iron, Skirt:2:
Topic 1	Hat:1:, Glove:1:, Jacket:1:, Muffler:1:, Coat:1:,
Topic 2	Hachimaki:1:, Sword:3:Japanese, Uniform:1:Sailor, Ribbon:1:, Headband:1:
Topic 3	Glasses:1:, Skirt:1:, One-piece-dress:1:, Hat:1:, Uniform:1:,
Topic 4	Robe:2:, Dress:2:, T-shirt:1:, Belt:2:Leather, Chalk:2:
Topic 5	Dress:1:, Uniform:1:For-Women, Knee-socks:1:, Coat:1:, Muffler:1:
Topic 6	Dog:1:, Sword:2:, Ribbon:1:, Glasses:1:, Cloak:2:, Swimsuit:1:Bikini
Topic 7	Plush-doll:1:, Tobacco:1:, Uniform:1:, One-piece-dress:1:, School-uniform:1:
Topic 8	Gun:1:, Suit:1:, T-shirt:1:, Guitar:1:Electric, Pants:1:
Topic 9	Shirt:1:, Jacket:1:, Tie:1:, Coat:1:, Hat:1:

得られた値を類似度とし，クエリとして入力されたコミックに対して，この類似度の高いコミックから降順に出力することとした．以上の指針に基づいて，本研究では，表紙に含まれた世界観情報から作品のトピックを推定し，トピック単位のヒストグラムで表現する．

4.2.2 実　　験

作成したプロトタイプによって得られる結果を定性的に評価した例を表 2 に示す．評価の際の比較するサンプルとして，(1) LDA によるトピックのヒストグラムをバタチャリア係数によって算出した類似度，(2) モデルから得られる 100 次元のベクトル同士をコサイン類似度によって算出した類似度，(3) LDA ベースの類似度とベクトルベースの類似度の平均類似度，の 3 種類を用いた．LDA を用いた類似度からはタグ同士の関係を考慮した多様性を期待できる反面，ベクトルから得られる類似度はモデルの推論によって表現された詳細な内容を考慮した正確性を期待できる．さらに，両方の類似度の観点を考慮した柔軟な類似度として，二つの類似度の平均値を用いる．

表 2 の (A) のケースは，“ドレス” や “翼”，“花” や “ドラゴン” といった西洋の世界観が強い要素が多く含まれたコミックをクエリとしている．また，表 2 の (B) のケースでは，“ドレス” と “チェンソー” という現代の世界観が強い要素が多く含まれたコミックをクエリとしている．両サンプルにおける LDA ベースの類似度では，同じく西洋の世界観を持つ作品が上位となった反面，ベクトルベースの類似度では，それぞれの作品が世界観にバラツキが見受けられる．各作品のベクトルを確認した結果，“ドレス”，“ローブ” といったタグが局所的に類似しているということが原因であることが分かった．特に，表 2 の (A) のクエリは服装の以外にも “花” や “ドラゴン” といった要素によって世界観が複合的に形成されているため，本来世界観が近し作品同士であっても異なる単語で形成されている場合，類似度が低くなる傾向がある．一方で，LDA ベースの類似度とベクトルベースの類似度を組み合わせた複合型類似度の場合，LDA ベースの類似度では類似した作品でもベクトルベースでは類似度が大きく低下するケースが多く見られた．

5. 議　　論

5.1 表紙からの情報抽出

本稿では VGG-16 モデルから Fine-Tuning を行うことにより，精度が 0.693，正解データに対する再現度は 0.918 を示すモデルを作成することができた．これにより，世界観を推定するために必要なタグを自動的に抽出することが可能となり，より多くの作品への応用が期待できる．しかし，学習データや検証データに含まれているタグを持ちつつ，学習には用いられていない未知の表紙を現在のモデルを用いて与えた場合，正しい推定を行うことは困難であった．これは，

(注4)：©Sara Yajima, Enterbrain Ltd.
(注5)：©Makoto Yukimura, Kodansha Ltd.
(注6)：©Takao Saito, Shogakukan Ltd.
(注7)：©Satoshi Ueda, Kodansha Ltd.
(注8)：©Tamaki Ishigaki, Takarajimasha Ltd.
(注9)：©Mineko Okami, Shinshokan Ltd.
(注10)：©Kami Imai, Shueisha Ltd.
(注11)：©Michiaki Watanabe, Enix Ltd.
(注12)：©Masami Tsuda, Hakusensha Ltd.
(注13)：©Yugo Kobayashi, Kodansha Ltd.
(注14)：©Fuyumi Ono, Kadokawa Ltd.
(注15)：©Tetsuya Saruwatari, Homesha Ltd.
(注16)：©Tsuyoshi Nakaima, Shogakukan Ltd.
(注17)：©Ryo Kurashina, Shueisha Ltd.
(注18)：©Motoka Murakami, Shogakukan Ltd.
(注19)：©Yasunori Mitsunaga, Kodansha Ltd.
(注20)：©Aro Hiroshi, Tokumashoten Ltd.
(注21)：©Hirokazu Shimada, Gakken Holdings Ltd.

表 2: 各手法に基づく類似度の結果

Query	Type	1st	2nd	3rd	4th	5th
(A) Dragon Knights(注9)	LDA	Tales of The Abyss(注4) 1.000	Vinland Saga(注5) 0.994	Kagegari(注6) 0.948	FLAGS(注7) 0.931	Wizardry(注8) 0.877
	Vector	NEEDLESS(注10) 0.456	Tales of The Abyss 0.435	Violinist of Hameln(注11) 0.379	Eensy-weensy Monster(注12) 0.347	Forest of water(注13) 0.304
	LDA + Vector	Tales of The Abyss 0.718	Kagegari 0.591	NEEDLESS 0.569	Vinland Saga 0.561	FLAGS 0.468
(B) Princess Resurrection(注19)	LDA	Ghost Hunt(注14) 0.999	Dog Soldier(注15) 0.994	Ucchare Goshogawara(注16) 0.926	Jyo-Oh(注17) 0.822	Musashi no Ken(注18) 0.820
	Vector	Jyo-Oh 0.875	Hunter Cats(注20) 0.869	Residents of Mister(注21) 0.163	Vinland Saga 0.116	Wizardry 0.092
	LDA + Vector	Jyo-Oh 0.843	Hunter Cats 0.773	Ghost Hunt 0.501	Dog Soldier 0.498	Ucchare Goshogawara 0.464

コミックは写真とは異なり，同じ物体であっても作者によって表現手法が全く異なることが起因すると考える．本稿と同様，人によって描かれたイラストを対象としている Saito らの illustration2vec [12] の場合，学習に用いられたのは 1,000,000 枚に及ぶ画像が用いられている．今後作者による表現手法の違いから共通点を見つけられるように，学習時のパラメータを調整や，データの拡張を行う必要がある．

5.2 コミックの関連付け

本稿では，コミックの世界観に関する情報を抽出するための意味ベクトル推定モデルから得られた推論結果に基づき，コミックの関連付けを行った．コミックの関連付けを行うための手法として，各コミックの意味ベクトルとタグ同士の類似度を計算することで，コミックの距離を把握する手法を提案した．提案手法によって関連付けられたコミックには時代や文化的背景の類似性が示唆された．また，意味ベクトルのみを考慮した類似度よりも LDA を応用した手法の方がより複雑に世界観が表現されているコミック同士を関連づけられたことが確認できた．これは，LDA を採用し

た理由が顕著と現れる結果であると考えられる．一方で，両方の類似度を組み合わせた類似度により，両手法の長所を生かした結果を期待していたものの，一部の結果（表 2 の B）においては期待した成果は得られなかった．これは，上述したベクトルベースの類似度の短所が起因していると考えられるため，平均値求める際，それぞれの類似度に対して重み付けすることによって平均類似度に及ぼす影響を調整し，より正確な結果を推定できると考える．また，今回 LDA のトピック推定を行う際，設定しなければいけないハイパーパラメータの決定や類似度の算出結果に対する評価は，すべて定性的なものであった．これは，今後意味ベクトル推定モデルのデータを拡張するにあたり，大きく影響及ぼす可能性のあるため，定量的にパラメータを選定するための手法を考慮する必要がある．

6. おわりに

本研究の目的は，内容情報に基づくコミックの検索の実現である．本稿では，その実現を可能とするために，コミックから内容情報を抽出するリソースとして表紙を用いる手法について提案した．我々の提案手法では，コミックの表紙に描かれた服装や持ち物によって表現されるコミックの時代や文化的背景を推定し，推定結果に基づきコミック同士を関連づける．まず，表紙から世界観に関連する情報を抽出するために，世界観を表現する情報（e.g., 衣服，持ち物）をタグとして付与したコミックの表紙の画像 536 枚に対して VGG-16 モデルを用いて Fine-Tuning を行った．その結果，正解データに対するタグの再現度が 0.918 を示すモデルを作成した．また，世界観情報の推論結果から，意味ベクトル同士の類似度や LDA によって生成したトピックの分布による類似度を算出し，コミックの関連付けを行った．これにより，世界観が似ている作品同士を関連づけることが可能なケースを複数確認することができた．一方で，意味ベクトル推定モデルの学習時に必要なデータが不足している点や，LDA にパラメータの定量的選定方法が必要な点が今後の課題として残っている．今後は本手法の改善を行いつつ，世界観以外の表紙に含まれた情報（e.g., 絵柄，ストーリー）を抽出する手法について検討していく．

文　献

[1] B. Park, K. Okamoto, R. Yamashita and M. Matsushita. Designing a comic exploration system using a hierarchical topic classification of reviews. *Information Engineering Express International Institute of Applied Informatics*, Vol. 3, No. 2, pp. 45–57, 2017.

[2] C. Rigaud, C. GuĀlrin, D. Karatzas, J. C. Burie and J. M. Ogier. Knowledge-driven understanding of images in comic books. *IJDAR* Vol. 18, No. 3, pp. 199–221, 2015.

[3] T. Tanaka, K. Shoji, F. Toyama and J. Miyamichi. Layout analysis of Tree-Structured scene frames in comic images. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, pp. 2885–2890, 2007.

[4] K. Arai and H. Tolle. Method for automatic E-Comic scene frame extraction for reading comic on mobile devices. In *2010 Seventh International Conference on Information Technology*, pp. 370–375, 2010.

[5] W. T. Chu and W. W. Li. Manga FaceNet : Face detection in manga based on deep neural network. In *17th Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, pp. 412–415, 2017.

[6] Y. L. Cun, L. Bottou, Y. Bengio and P. Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, pp. 2278–2324, 1998.

[7] D. M. Blei, Y. Andrew and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, Vol. 3 pp. 993–1022, 2003.

[8] D. M. Blei. Probabilistic topic models. *Comunications of the ACM* Vol. 55, No. 4, pp. 77–84, 2012.

[9] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

[10] M. Lin, Q. Chen and S. Yan. Network in network. In *International Conference on Learning Representations 2014*, 2014.

[11] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations 2015*, 2015.

[12] M. Saito and Y. Matsui. Illustration2Vec: a semantic vector representation of illustrations. *SIGGRAPH ASIA 2015 Technical Briefs*, No. 5, 2015.

[13] A. Islam and D. Inkpen. Semantic text similarity using corpus-based word similarity and string similarity. *ACM Transactions on Knowledge Discovery from Data* Vol. 2, No. 2, 2008.

[14] S. H. Cha. Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of mathematical models and methods in applied sciences*, Vol.1, No.4, pp. 300–307, 2007.