

# 条件付き変分オートエンコーダに基づく商品画像分析モデルに関する研究

1X19C109-2 泓垂由乃  
指導教員 後藤正幸

## 1. 研究背景・目的

EC サイトで扱われる各商品には、一般に機能や色、素材などの属性、商品画像、商品説明文が用意されており、顧客はそれらを参照して購入を検討する。特に、視覚情報によるイメージは顧客の大きな判断材料になり得るため、商品画像が与える印象は重要である。そのため、近年高度に発展した深層学習系のモデルを援用して、効果的な商品画像を生成する試みも行なわれている。例えば、従来の分析手法 [1] を用いれば、具体的な属性を複数指定した商品画像を生成することができる。しかし、顧客ニーズの中には具体的な属性だけでは表現しきれない抽象的なニーズが存在するが、従来手法では顧客の望む抽象的なイメージを再現できない可能性がある。ここで、商品説明文には「かわいい」「上品」などの商品に対する抽象的なイメージを表す単語が含まれている場合が多い。これらのことから、商品説明文に表現されるような各商品の抽象的な要素が商品画像に与える影響のモデル化ができれば、顧客のニーズにより合致する商品画像の生成が可能になると期待できる。さらに企業にとっても、新たに生成された商品画像はニーズ発見の一助となる可能性がある。

そこで本研究では、トピックモデルである Latent Dirichlet Allocation [2] (以下、LDA) と属性を考慮した画像生成モデルである条件付き変分オートエンコーダ [3] (以下、CVAE) を組み合わせた、商品に対する抽象的なイメージを活用した画像生成手法を提案する。具体的には、LDA によって商品説明文から抽出した潜在トピックを属性として扱い、CVAE の条件として用いることによって、説明文に含まれる抽象的な情報を活用した商品画像生成を可能とする。本研究では、生花商品および生花と食品・飲料品のセット商品を扱う某 EC サイト (以下、EC サイト A) のデータに対して本提案モデルを適用し、その有効性を検証する。

## 2. 関連研究

### 2.1. Latent Dirichlet Allocation (LDA)

LDA [2] は、自然言語処理の分野で用いられるトピックモデルの一つである。LDA では文書  $d$  を  $N_d$  個の単語  $\{w_{d,n}\}_{n=1}^{N_d}$  の集合として捉え、各単語  $w_{d,n}$  が潜在トピック  $z_{d,n}$  に基づいて確率的に生成されると仮定している。

潜在トピック  $z_{d,n}$  は、各文書のトピック分布  $\theta_d$  に従って生成される。 $\theta_d$  は、予め設定されたトピック数である  $K$  次元のパラメータ  $\alpha$  を持つディリクレ分布に従って生成されるものとする。また、トピック  $k$  の単語分布  $\phi_k$  は総単語数  $V$  次元のパラメータ  $\beta$  を持つディリクレ分布に従って生成される。LDA のグラフィカルモデルを図 1 に示す。

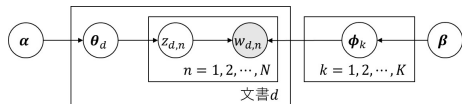


図 1: LDA のグラフィカルモデル

### 2.2. 条件付き変分オートエンコーダ (CVAE)

CVAE [3] は、パラメータ  $\phi$  を持つエンコーダ関数  $q_\phi$  とパラメータ  $\theta$  を持つデコーダ関数  $p_\theta$  で構成された、深層生成モデルの一種である。エンコーダでは、観測データ  $x$  と属性情報  $y$  を基に  $z$  の確率分布を推論し、その確率分布からサンプリングした  $z$  と属性情報をデコーダに入力することでデータの生成を行う。全体のモデル概要を図 2 に示す。

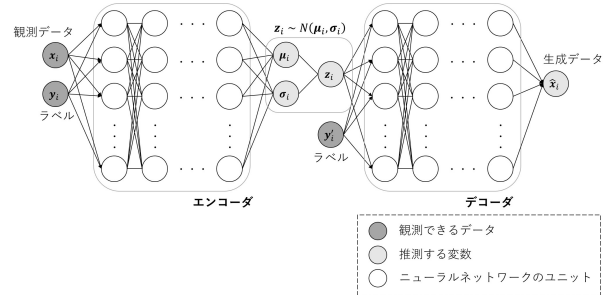


図 2: CVAE のネットワーク構造

モデル学習時には、式 (1) に示す変分下界を最大化するようにエンコーダとデコーダのパラメータが最適化される。

$$\mathcal{L}(\theta, \phi; \mathbf{x}, \mathbf{y}) =$$

$$\mathbb{E}_{q_\phi(z|\mathbf{y}, \mathbf{x})} [\log p_\theta(\mathbf{x} | \mathbf{y}, z) + \log p_\theta(\mathbf{y}) + \log p(z)] - \mathbb{E}_{q_\phi(z|\mathbf{y}, \mathbf{x})} [\log q_\phi(z | \mathbf{y}, \mathbf{x})] \quad (1)$$

CVAE のエンコーダで得られる潜在変数ベクトル  $z$  は、観測データの属性情報に依存しない特徴となっている。そのため、エンコーダに入力した属性情報とは異なる属性情報をデコーダに入力した場合、元の特徴を維持したまま属性のみを変更したデータを生成することができる。

## 3. 提案モデル

### 3.1. 着想

本研究で対象とする EC サイト A の各商品には、メイン花材や商品形態 (花束、アレンジメントなど) などの属性、商品画像、および商品説明文が設定されている。顧客は、メイン花材や価格帯などの条件を指定して商品の絞り込み検索を行うことができ、絞り込んだ商品の中から商品画像や商品説明文を参考に購入商品を決定する。実際の商品は EC サイト A の加盟店によって商品画像を参考に作成されるため、商品画像が顧客の満足度に及ぼす影響は大きいと考えられる。また、各商品に対応する説明文には、「華やか」「人気」「安心できるデザイン」など、主観的・抽象的な情報も含まれている。これらはメイン花材や商品形態などの属性のみでは判断できない情報であり、顧客の抽象的なニーズを表現するものであると言える。商品説明文が既存の商品画像の抽象的なイメージを正しく表現できていると仮定すれば、商品説明文から抽象的な情報を含んだトピックを抽出し、そのトピックを変更した画像を生成することで、変更元の画像を顧客ニーズ

に合致するよう更新した商品画像の生成も可能になると考えられる。

### 3.2. 提案手法

本提案手法ではまず商品説明文に LDA を適用し、商品ごとのトピック分布を推定する。そして、所属確率が最大となるトピックをその商品のトピック属性とみなし、CVAE の属性情報  $y$  として用いることで既存の商品画像とともに CVAE を学習する。学習済みの CVAE においてトピック属性を変更させることで、商品画像の特徴を保持しながらも、イメージを変えた新たな商品画像を生成することが可能になる。提案モデルの概要図を図 3 に示す。

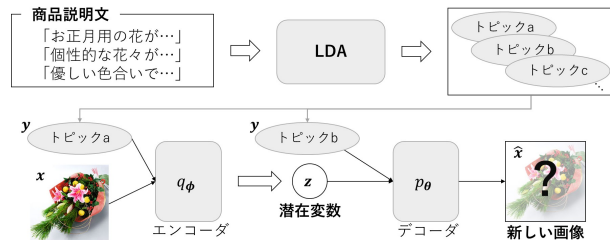


図 3: 提案モデルの概要図

## 4. 実データ分析

本提案モデルの有効性を検証するため、EC サイト A に掲載されている実データを用いて画像生成の実験を行なう。

### 4.1. 実験条件

本実験では、 $200 \times 200 \times 3$  に調整した画像特徴量を CVAE の学習データとして用いる。データ数は 4,382 枚であり、LDA によるトピック分析では、これらに対応する商品説明文のうち、重複を除いた 3,196 種類を用いて 6 つのトピックを抽出した。CVAE ではエポック数を 1,000、バッチサイズを 100 とし、潜在変数の次元数は 64 とした。

### 4.2. 実験結果

#### 4.2.1. LDA によるトピック分析

表 1 に、各トピックにおいて出現率が高い特徴的な単語、およびそれらから推定したトピックごとの特徴を示す。

表 1: 商品説明文より得られたトピック

Topic	特徴的な出現単語	推定される特徴
1	北海道, 沖縄, 離島	地方名産とのセット
2	日頃, 春, 優しい	生花メインのギフト
3	コース, ママ, ボディ	母の日ギフト
4	焼, マスコット, 可愛い	子供向けギフト
5	食, 花束, ブーケ	食品とのセット
6	ばあちゃん, 世界, 梅	祖父母へのギフト

表 1 より、トピックごとの特徴として、生花商品の用途が抽出できていることがわかる。また、トピック 2 および 4 では、「優しい」「可愛い」など、見た目の印象が特徴として現れていることがわかる。

#### 4.2.2. CVAE による学習結果

ある画像のトピックを変化させた生成画像を図 4 に示す。図 4 から、元のトピック 1 における再構成は十分に行えていることがわかる。また、特にトピック 2 においては画像の鮮明度がやや上がっており、画像の鮮明度はトピックを変更した場合でもあまり劣化していないことがわかる。さらに、トピックを変化させたときには、生花の色合いもやや変

化していることがわかる。例えば、トピック 2 および 4 においては白の割合がやや多く、トピック 3 および 6 においては緑の割合がやや多くなっている。

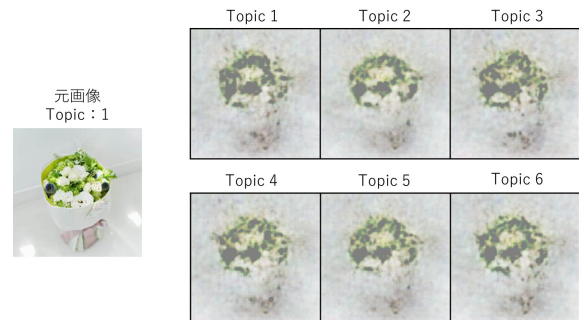


図 4: トピックの変化による生成画像の変化

## 5. 考察

本提案モデルでは、商品説明文に含まれる抽象的な情報をトピックとして抽出し、そのトピックを属性としたイメージ画像を生成した。これにより、従来のような具体的な属性ではなく、抽象的な情報に基づいたイメージ画像生成が可能となったと考えられる。しかし、図 4 からわかるように、トピックを変更した場合も顕著な画像の変化は見られない。これは、商品説明文に現れる抽象的な情報と商品画像が一意には紐づいていないことが原因であると考えられる。すなわち、同じトピックに所属する説明文が与えられている商品であっても商品画像は多種多様であり、それら全ての画像に共通する特徴量が明確には存在していないことが推察される。また、表 1 からわかるように、トピックにおける出現率の高い単語の多くが、生花商品そのものではなくセット商品についての単語であることも原因であると考えられる。したがって、トピック抽出方法の改良や、生花のみの商品とセット商品とを区別した学習により、生成画像の違いがより明確になる可能性がある。

## 6. まとめと今後の課題

本研究では、LDA と CVAE を組み合わせた、商品の抽象的な特徴を利用した画像生成モデルを提案した。提案モデルを EC サイト A の実データに適用したところ、抽象的な情報の変化に基づいて商品画像を生成することができた。

今後の課題としては、トピックの抽出方法の改良などによる生成画像の鮮明化や、他の実データへの適用などが挙げられる。

### 参考文献

- [1] 北里礼, 雲居玄道, 後藤正幸. 購買履歴と画像情報を共学習する CVAE に基づく生花の新商品イメージ自動生成手法. 情報理論とその応用シンポジウム予稿集 (CD-ROM) (SITA), 2021.
- [2] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, Vol. 3, No. Jan, pp. 993–1022, 2003.
- [3] Durk P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. *Advances in neural information processing systems*, Vol. 27, pp. 3581–3589, 2014.