

# 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に関する研究

情報数理応用研究

5220C013-1 北里礼  
指導教員 後藤正幸

## A Method for Automatic Generation of New Product Images of Fresh Flowers Based on CVAE Co-learning of Purchase History and Image Information

Aya Kitasato

### 1. 研究背景と目的

本研究では、購買履歴データを用いた売上向上のためのマーケティング施策に着目し、生花を販売する某 EC サイト（以下 EC サイト A）を対象事例とする。生花商品は、色やサイズ、花の種類、アレンジメントの形状（バスケットやブーケなど）といった属性の組み合わせによって構成されている。顧客は商品の属性を細かに指定した個別注文はできず、商品画像や花材、サイズなどを参考にサイト上の出品アイテムの中から見栄えを重視して商品を選択する。また、EC サイトの商品の購買理由は誕生日や母の日などのイベントにおける贈呈がメインであり、イベント毎に好まれる生花商品は異なるという特徴がある。そのため現状の EC サイトでは、企業側の経験的なノウハウに基づき、イベントごとのダイレクトメールの送付や特設ページの開設、商品の開発などを通して、イベントと連動した販売促進策が講じられている。ここで、より顧客の嗜好にマッチした施策立案のためには、顧客属性とイベントの組み合わせ毎に、好まれるアイテムの特徴を把握することが望まれる。従来の分析手法を用いず、購買履歴データから各顧客がどのような既存商品を好んでいるかを明らかにすることができ、既存商品を用いた顧客の嗜好分析が可能である [1],[2]。しかし、既存の生花商品は約 1000 種類と、一般的な本や食品といった日用品などと比べてアイテム数がかなり少なく、限られた既存商品のみを用いて顧客の嗜好を把握するには限界がある。一方、生花商品は、様々な機能を有する日用品と比べ、色やサイズなど、画像によって商品イメージが捉えやすいという特徴がある。そのため、既存商品の購買履歴と商品画像のデータから、各顧客が好むであろう生花商品を仮想商品イメージとして生成できれば、これらの多数の仮想商品イメージを用いて、各顧客属性のイベント毎の嗜好を詳しく分析できると考えられる。加えてこの分析結果は、新商品開発やイベント毎のターゲティング方法などにも利用可能である。また、これらの仮想商品イメージは、加盟店とのコミュニケーションツールとしての活用など、様々なマーケティング活動での有用性も期待できる。

そこで本研究では、画像の付属情報を考慮した画像生成に適する Conditional Variational Autoencoder[3] を応用して、「イベントと顧客属性を考慮した個人の嗜好に基づく仮想商品画像の生成モデル」を構築し、生成した仮想商品イメージを活用した顧客の嗜好分析手法を提案することを目的とする。具体的には、顧客属性とイベントの両方を付属情報とするため、マルチラベルの CVAE を導入し、顧客属性とイベ

ントのマルチラベルのもとで仮想商品画像を生成する。なお、顧客の嗜好は多様であり、ある商品は、誕生日や母の日、そして 20 代から 40 代の女性に購買される、などといったように、同じ商品であっても様々な顧客属性により様々なイベントで購買される。そのため、過去の研究 [4] のように正解ラベルが一意の問題設定ではモデルの学習がうまくいかないという問題が発生する。そこで、本研究では正解ラベルが複数存在しても安定した学習が可能となる手法を提案し、イベントや顧客属性毎に好まれる仮想商品イメージを生成可能なモデルを構築する。本研究では、この提案手法を、EC サイト A から提供された生花商品の購買履歴の実データセットに適用し、提案手法の有効性を実用的な観点から明らかにする。

### 2. 準備

#### 2.1. 対象の EC サイト

本研究で対象とする EC サイトは、生花もしくは生花と食品・飲料のセット商品を取り扱っている。サイト上では、イベントや価格帯、メインに入れたい花の種類などで生花商品を絞り込み検索することができ、価格帯だけ選択し、予算に合わせて加盟店側へアレンジメントをまかせることも可能である。商品は、ピンクバラのアレンジメントやトルコキキョウの花束などメインの花の種類と商品の大体の形まではサイト上で選択できるが、その他の細かいカスタマイズはできず、サイト上の商品のイメージ画像に添った商品を加盟店が作成し配送する。よってイメージ画像は商品の実物感を顧客が知るための手がかりとなり、顧客は購買において商品画像を特に重要視すると考えられる。

#### 2.2. 関連研究

山極ら [1] は、本研究対象である EC サイトの購買履歴データを用いた研究を行っている。生花 EC サイトは、日用品と比較し個々の顧客の購買回数が非常に少なく共起情報によるモデル化が困難であるため、同一顧客との共起でなく、各商品が購入される状況から商品の類似度を測り商品推薦に活かす方法を提案している。この問題設定では、類似した既存商品を推薦すること、或いはその類似した商品たちの分析から商品開発を行なうことが目的となり、商品の情報は購買履歴に基づく単価や生花の種類になる。一方、本研究は、商品情報でなく色味や型などの情報を得られる画像に着目し、顧客属性の嗜好を反映した仮想商品画像を生成することで、画像を基に商品開発やマーケティング方法の改善を行なうという全く異なったアプローチとなる。

購買分析モデルの分野では、Conditioned Variational

Autoencoder と呼ばれる、エンコーダのみにラベルを入力したモデルが提案されている [2]。このモデルでは、入力データはアイテムへの評価値、ラベルは商品のジャンルなどを補足情報として学習し、評価値の高い商品をジャンルなどの制約に合わせて推薦することを目的としている。また、マルチラベリングの CVAE の研究では、マルチラベルを 1 度に学習する負担を軽減し、データ数を確保するため各ラベルを順次学習する手法 [4] などがある。

### 3. 提案モデル

本研究では、付属情報としてのラベルと入力データから画像生成を学習できる CVAE の適用を考える。そのため、イベントと顧客属性によって商品の嗜好が変化することを考慮し、イベントと顧客属性を商品の付属情報として組み合わせてマルチラベル  $\mathbf{y}$  とする。そして、学習させる生花商品の画像データを入力  $\mathbf{x}$  として入力して画像生成を学習後、ラベルを変更することで様々な顧客属性とイベントのラベルを持つ仮想商品画像を出力する。本研究では、入力  $\mathbf{x}$  が画像であることからニューラルネットワーク構造として Convolutional Neural Network[5] を用いる。その際、エンコーダではフィルタを重み  $\varphi$ 、デコーダでは重みを  $\theta$  として、目的関数 (1) を最大化するように更新することで学習し、デコーダで画像を再構成する。ただし、全データ数を  $N (1 \leq n \leq N)$  とし、 $n$  はデータ番号である。なお、データ番号が異なっても同じ画像が使われる可能性がある点に注意する。購買履歴であれば、データ 1 つは 1 回の購買にあたるため、データ番号が異なれば別の時点での購買となり、データ内で何度も同じ商品が出現することがある。

$$\begin{aligned} \mathcal{L}(\mathbf{x}_n, \mathbf{y}_n; \theta, \varphi) = & \mathbb{E}_{q_\varphi(z_n|\mathbf{x}_n, \mathbf{y}_n)} [\log p_\theta(\mathbf{x}_n|\mathbf{y}_n, z_n)] \\ & - KL[q_\varphi(z_n|\mathbf{x}_n, \mathbf{y}_n) || p_\theta(z_n)] \\ & + \log p_\theta(\mathbf{y}_n) \end{aligned} \quad (1)$$

また、今回の商品画像のラベルは様々な顧客の購入履歴によって決定されている。今回のような、複数の正解ラベルをもつ商品がある場合、単純な CVAE の適用では学習が成功しない。また、イベント、性別、年齢のうち、イベントは他の 2 つの要素よりも商品に対して一意に決定される割合が高く、イベントの特徴が重点的に学習されてしまう。そこで、複数の正解を許容し、かつ学習の重みが偏らないように各ラベルの重みを調整できるラベリング方法が必要となる。

ここで、商品数を  $M$ 、イベントタイプの数  $U$ 、ジェンダタイプ数を  $G$ 、年代タイプ数を  $A$  とする。このとき  $M$  個の入力データ行列を  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_m, \dots, \mathbf{x}_M)$  とする。その際、その要素は画像が 1 ピクセルにつき、RGB を表す 0 から 1 までの間の値 (RGB の定義域である 0 から 255 の整数値を、255 で割った値) をもっているため、1 枚の画像の縦のピクセル数を  $V$ 、横のピクセル数を  $H$  とすると  $\mathbf{x}_M \in [0, 1]^{V \times H \times 3}$  と定義される。次に本研究のマルチラベルの行列  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_m, \dots, \mathbf{y}_M)$  の詳細の構成を説明する。まず、横軸を購買履歴のデータ番号でなく、商品数  $M$ 、縦軸を各ラベルの種類とした行列で、各種類の商品の累積購

	U				G			A		
$\mathbf{y}_1$	0.8	0.1	...	0	0.8	0.2	0.1	0.2	...	0
$\mathbf{y}_2$	0.1	0.2	...	0	0.1	0.9	0	0	...	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathbf{y}_M$	0	0.2	...	0.1	0.8	0.2	0.2	0	...	0.4

図 1: 提案するラベル行列  $\mathbf{Y}$  のイメージ

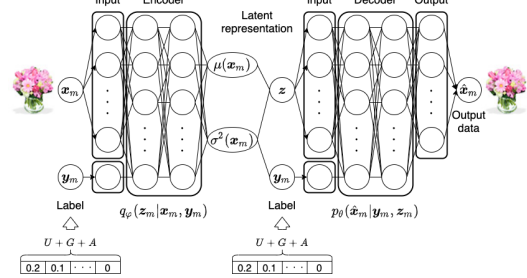


図 2: 学習の提案手法

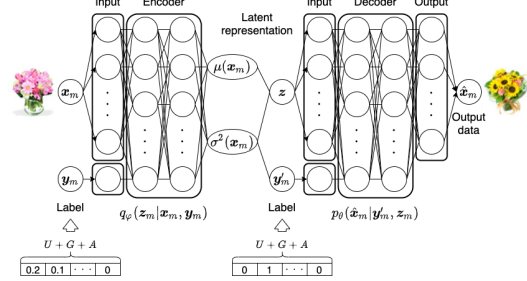


図 3: 学習後の提案手法

入数をとる。さらにそれらを各ラベルごとに行で正規化したベクトルを  $\mathbf{y}_{m1} = (y_{m,1,1}, \dots, y_{m,1,U})$ ,  $\mathbf{y}_{m2} = (y_{m,2,1}, \dots, y_{m,2,G})$ ,  $\mathbf{y}_{m3} = (y_{m,3,1}, \dots, y_{m,3,A})$  とし、これらを連結したベクトル  $\mathbf{y}_m$  を入力とする。以上のように構成されるマルチラベル行列  $\mathbf{Y}$  のイメージを図 1 に示す。このような手順によりラベル間の重みに大きな差が生まれず、すべてのラベルを偏りなく学習することができる。生成モデルを学習する際には、図 2 のように 0 から 1 の連続値を持つ  $\mathbf{y}_m$  と  $\mathbf{x}_m$  を入力し、同じ  $\mathbf{y}_m$  とエンコーダで生成した潜在変数  $z_m$  をデコーダで入力する。

学習後、分析したいラベルを入力し仮想商品画像を生成する際、エンコーダは学習時と同様に連続値で  $\mathbf{y}_m$  と  $\mathbf{x}_m$  を入力し、図 3 のように  $z_m$  を得る。一方、デコーダでは、分析したい性別、年齢、イベントラベルに応じて、 $\mathbf{y}'_m \in \{0, 1\}^{U+G+A}$  を入力して商品画像を得る。この商品画像を目視、もしくは画像の RGB 情報などによって分析することで、各属性・イベントでの商品の嗜好を捉えターゲティング方法の改善、或いは商品開発などに活用する。

## 4. 実データ分析

### 4.1. 分析条件

提案モデルの有効性を実証するために実データを用いて画像生成実験を行った。以下にその分析とデータの条件を示す。

表 1: Data Conditions.

購買履歴データ収集期間	2018/08-2019/09	
購買履歴データ数	(about)400,000	
生花商品画像数	1,211	
画像のサイズ	200 × 200	
画像のラベル数	イベント	34
	性別	2
	年代	6

表 2: Experimental Conditions

エポック数	1,000
バッチサイズ	100
潜在次元数	64

表 3: CNN Conditions

フィルター 1 のサイズ	3 × 3 × 3
フィルター 1 の数	32
フィルター 2 のサイズ	3 × 3 × 32
フィルター 2 の数	64
ストライド幅	2, 2



図 4: 商品 C のオリジナル画像

また、本研究の提案モデルの性能を確認するために、“学習が収束するかどうか”と“学習によって画像が適切に再構成されるかどうか”の双方の観点から分析を行なった。その結果、1,000 エポックで損失関数が約-3,500 で学習が収束したと判断できた。また、再構成された 1211 枚の商品画像すべてを元の画像と目視で比較したところ、すべての画像が再構成されたこと、すなわち 100%再構成に成功したことを確認した。これは、生成モデルが正常に学習データ画像の特徴をモデル化できたことを示しており、このモデルの学習プロセスは成功しているといえる。

## 5. 分析結果

### 5.1. マルチラベルによる画像生成分析結果

本研究の主目的である「仮想商品イメージを活用した顧客の嗜好分析」のため、本節ではある商品画像をベースにイベントや年代、性別を変化させた生成画像を得る実験を行なう。そして本節では、結果の例として、イベント={お誕生日:0.67, その他:0.33}, 年代={30代:0.67, 40代:0.33}, 性別={女性:1.0} のラベルを持つ図 4 の商品 C の画像をベースにし (なお、0.67 などは商品全体の購買に対する各ラベルの割合を示す。女性:1.0 は全て女性となる)、イベントと性別を変化させた結果のみを示す。

#### 5.1.1. イベントの変化による画像分析結果

本項では、顧客が各イベントでどの様な商品を好むかを検証するため、購買された商品 C の画像をベースとし、イベントのみ変化させた際の商品画像を生成する実験を行なう。ここではその検証結果を示すため、一例として年代が 20 代・40 代で性別は女性のクリスマス、お盆、ハッピーローズ (EC サイト A のイベント) での商品画像の結果を図 5 に示す。

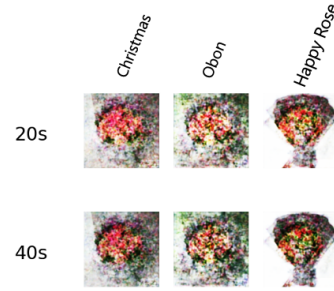


図 5: イベントの変更による生成結果

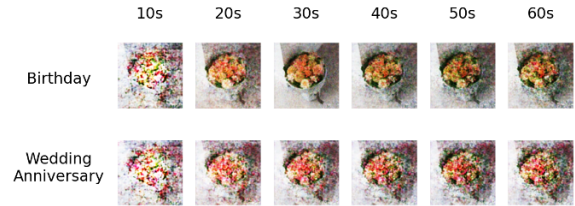


図 6: 年代の変更による生成結果

この結果、例えばクリスマスでは、マゼンタのような明るい赤い花が中心となったラウンドブーケの形、お供えでは白が多くなりラウンドブーケの形になるなどの変化が確認できる。すなわち、イベントの変化に応じて画像も変化する。そのため、クリスマスでは明るめの花束に注力して商品開発を行うといったような生成画像を基に商品開発をすることや、生成された仮想商品画像を用いて加盟店にそのイメージを共有することなどが可能と考えられる。

#### 5.1.2. 年代の変化による画像分析結果

本項では、ある商品をベースとして、各年代が好むような商品画像を生成した場合、どのような変化があるかを分析して、新たな商品開発やターゲティングなどに役立てることを想定し、年代のラベルのみを変えて画像を生成する実験を行った。図 6 に女性に対し、20 から 60 代までの画像の生成結果を示す。ここから若年層は、誕生日ではピンクや明るいオレンジが多く、結婚記念日は白色やピンクが多くなるのがわかる。また誕生日と結婚記念日の両方で、高齢になるにつれ暗めの緑が増え落ち着いた色合いの生花が好まれることもわかる。この様に、年代による商品の好みの変化が本モデルによって確認できる。そのため、各年代で好まれる新商品開発、或いは年代で細分化して訴求点を再定義するといったターゲティング方法の改善などに活用できる結果が得られたといえる。

#### 5.1.3. 変化のない画像分析結果

入力データによっては、5.1.1 項と同じ手順でラベルを変化させても生成画像が全て同じになってしまうケースが確認できた。ここでは、例として女性ラベルで、すべての年代の退院祝いの生成結果を図 7 に示す。図から情報を読み取れない同じ画像がしか生成されていないことがわかる。また、考察のため退院祝いの購買の、約 40 万件中の割合を分析すると、0.0014% となり非常に少なかった。購買の割合が 0.0005% の出産祝いについても、画像生成うまくいかないことが確認

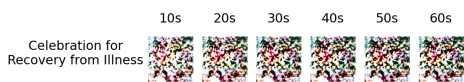


図 7: 退院祝いの年代による変化

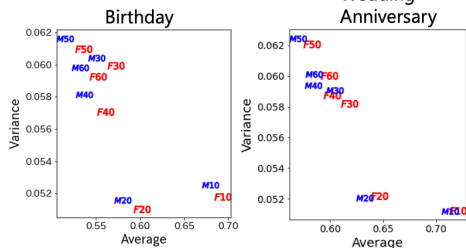


図 8: 各イベントでの輝度分析

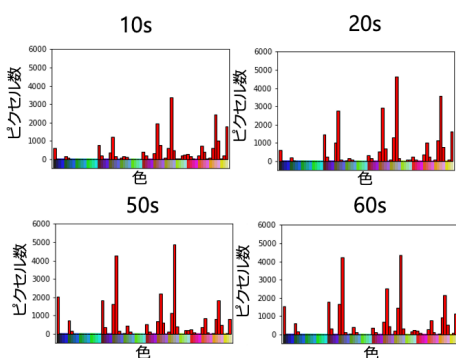


図 9: 誕生日の RGB 分析

された。以上のことから、学習データが少ない場合は画像生成は上手く行かないと考えられる。

## 5.2. 輝度・RGB 値に基づく生成画像の定量分析

イベント・性別・年代による商品の嗜好の変化についてより定量的に分析するため、分析したい組み合わせの画像を生成し、その各生成画像の輝度や RGB 値（64 色に集約）を比較した。輝度は RGB 値を用いて  $R \times 0.21 + G \times 0.72 + B \times 0.07$  で定義され、画像の明るさを表す。本節では紙面の都合上、5.1 節と同様の商品 C を対象にした輝度の結果のみを示す。分析したい生成画像の全ピクセル  $200 \times 200$  分の輝度を計算し、その平均と分散を求め定量的分析を行なった。図 8 に、誕生日と結婚記念日の輝度の結果を、赤の F を女性、青の M を男性としてその隣に年代の数字を付随して示す。この結果、両方で輝度が高い（明るい）方に若年層が分布し、高齢層は輝度の分散が高くカラーバリエーションが多いといえる。この結果は、画像比較の印象と合致しており、定量的・具体的に明るみの増減を提案できることが示唆される。

RGB 値については、10, 20, 50, 60 代の女性、誕生日の結果のみ抜粋したものを図 9 に示す。図 9 より若年層は黄色やオレンジなど右側の明るい色のピクセルが多く、高齢層は黒や深緑など左側の暗い色のピクセルが多いことがわかる。すなわち、RGB 値も輝度と同様に画像の比較結果を裏付け、より定量的・具体的に色味や明るみの増減を分析可能である。

## 6. 考察

上記の分析結果から、基本的にイベントや属性の変化によって生成画像も変化すること、そして画像比較や輝度の比較から各イベントや各顧客属性の商品の好みを定量的・定量的に分析しターゲットング方法の改善、或いは商品開発が提案ができることが明らかとなった。また、学習データの不足によって画像生成が困難な部分があるものの、損失関数の収束や画像の再構成度からこのモデルの信頼度は高く、ある程度のデータが存在すれば商品の画像生成は上手くいくと考えられる。

本研究では当初、イベント、年代、性別それぞれで、購買履歴データを軸に 1-hot ベクトルを作成し結合したマルチラベルを検討したが、この方法では学習が上手くいかなかった。具体的には、イベントを変化させるとやや画像は変化するものの、年代や性別の変更ではほぼ変化がなかった。そのため、本研究では、商品の種類を軸に各ラベルで累計購買回数をとり正規化したラベルを利用する方法を提案した。今回の実験の成功は、このラベルの作成方法が寄与しており、実際に複数の正解ラベルを許容し各ラベルの学習の重みのバランスがとれたことを示唆している。

## 7. まとめと今後の課題

本研究では CVAE を用いて「イベントと顧客属性を考慮した個人の嗜好に基づく仮想商品画像の生成モデル」を構築し、これらの仮想商品イメージを活用した顧客の嗜好分析手法の提案を行なった。そして、実データを用いて顧客属性とイベントのマルチラベルを付与した仮想商品画像の生成学習を行い、その結果生成される仮想商品イメージを用いた分析を行なった。この学習により、データが十分に存在すれば解釈可能な仮想商品画像を生成可能であり、生成画像から各顧客属性の嗜好を捉えられることを示した。この結果を用いて、今後は新商品開発やターゲットング方法の改善などへと繋げることが期待できる。今後は、既存商品と生成された仮想商品画像の定量的比較をすることなどが課題である。

### 参考文献

- [1] 山極綾子, 雲居玄道, 後藤正幸. 生花 EC サイトの購買履歴に基づく商品特性分析モデル. 第 19 回日本データベース学会年次大会 DEIM2021, 2021.
- [2] Tommaso Carraro, Mirko Polato, and Fabio Aioli. Conditioned variational autoencoder for top-n item recommendation. *ArXiv Preprint ArXiv:2004.11141*, 2020.
- [3] Diederik P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems*, pp. 3581–3589, 2014.
- [4] Hong-Min Chu, Chih-Kuan Yeh, and Yu-Chiang Frank Wang. Deep generative models for weakly-supervised multi-label classification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 400–415, 2018.
- [5] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, Vol. 1, No. 4, pp. 541–551, 1989.