

論文題目

ファッション分野における
画像と感性語、文章の関連性分析
及び提案システム構築

A Relevance Anylysis between Images and
Affective Words, Sentences in a Fashion Field
And Building on Recommendation Systems

指導教授

萩原 将文 教授

学習指導副主任

西 宏章 教授

慶應義塾大学 理工学部 情報工学科

平成 29 年度

学籍番号 61402068

井上 幹晴

目 次

第 1 章	あらまし	1
第 2 章	はじめに	2

第 1 章

あらまし

本論文では、ファッション分野における画像と感性語、文章の関連性を分析し、また自動コーディネートシステムを提案する。近年、Convolutional Neural Network(CNN) は画像分類を始めとする様々なタスクで高精度を記録しては画像分類を始めとする様々なタスクで高精度を記録してきた。そこで本研究では、ファッション分野における画像を CNN に入力することで感性語や文章と対応付け、その関連性を分析した。さらにその関連性を用いた周辺システムを考案した。

本研究では、CNN モデルとして AlexNet と VGG-19 モデルを使用した。これらのモデルに Large-scale Fashion データセット (DeepFashion) を用いて事前学習させ、さらに表現力を高めるため、実際のデータセットを用いてファインチューニングも行った。感性語はキャプションから抽出した単語とし、教師ラベルとして CNN を学習させた。

本論文では、アイテムごと、またアイテムの組み合わせ (コーディネート) の画像に対し、CNN を通して出力される感性語がふさわしいか評価実験を行った。また反対に、ユーザが欲しいアイテムやコーディネートに関する文を入力した時、ふさわしいアイテムやコーディネート画像が出力されるかについても実験した。その結果、CNN を通して正しく画像と感性語や文章が対応づけされていることが確認された。さらに本論文ではこの関連性を用いてユーザが持つアイテムに似合う別のアイテムとコーディネートするシステムを構築した。

第 2 章

はじめに

1. ファッションについて

ファッションというのは、自己表現の一つである。近年では店舗に足を運ばなくても商品を購入できる通信販売が人気を博している。ZOZOTOWN を運営するスタートトゥデイ（株）は営業利益率が前年比 11.6% と大幅に上昇している。

2. システムの必要性

近年では Siri 等の音声による検索システムが開発されている。音声検索では口語の文章を分析する必要がある。本研究では感性語を画像と文章の媒介として用いることで口語文における検索精度向上の可能性を模索した。

通販サイトは実際に試着できるわけではないので、コーディネートまでには分からない。購入した商品が自分の持っている服に合わないという場合が多々ある。ユーザの所持品を考慮する必要性あり。

3. CNN の能力について

画像分類タスクを始めとする様々なタスクで突出した精度を誇る。ILSVRC 2015 では人間を超えたと言われている。

4. word2vec の能力について

word2vec は単語を分散表現にすることで、単語同士のベクトル演算を可能にする。また、 \cos 類似度をとることによって、単語間の距離を測ることも可能である。

5. 事前学習の有効性

Donahue, Razavian, Sermanet らの研究では事前学習した CNN モデルの特徴は様々な画像分類タスクに転用することができる包括的な画像表現であることを実証した。Wei らの研究では、事前学習させたモデルにさらに検証に用いるデータの一部を用いてファインチューニングさせることでさらに精度が向上することを発見した。

6. 既存研究

ファッションアイテムの検索システムとしては、アイテム画像を入力して似たアイテムの画像を表示する研究や、画像に加えて属性情報を考慮して柔軟に検索する研究が行われている。しかしこれらの検索方法は参考画像を必要としており、文章のみの入力に対応できない。そこで本研究では文章のみでアイテム画像を検索するシステムを構築する。また、アイテムを組み合わせでコーディネートを作成するシステムが開発されているが、感性語を考慮していない。そこで本研究では感性語を考慮した新たなコーディネートシステムを提案する。

7. システムの概要

- (a) 既存の CNN モデルを DeepFashion データセットのカテゴリ分類タスクで事前学習。その後 SENSY (株) から提供された検証用データの一部を用いてファインチューニング。
- (b) アイテム画像を CNN に入力し、感性語を取得。逆に文章を入力し、アイテム画像を取得。
- (c) アイテムの組み合わせに対しても、画像から感性語を取得。逆に文章からアイテムの組み合わせ画像を取得。
- (d) 画像と感性語の関連性に着目し、アイテム画像から似合う他のファッションアイテムの組み合わせを提示。

8. 評価実験

画像から感性語を取得するタスクでは、出力した感性語が実際のキャプションに含まれている割合を検証。また主観評価実験で、得られた感性語が画像にふさわしいかどうかを検証。文章から画像を取得するタスクでは、主観評価実験で入力文にふさわしい画像が出力されているかを検証。提案システムに関しては、提案されたコーディネートが自然かどうかを検証。

9. 論文の構成第2章では提案システムについて、第3章では評価実験、第4章で今後の展望、第5章で結論を述べる。