

カテゴリを取得する。そして、取得したメインカテゴリのサブカテゴリを分類する CNN を選択し、入力に整形した画像を入力することでメインカテゴリ内のサブカテゴリを同様に取得する。こうして対話モジュールからユーザーが入力した服の画像がどのカテゴリであるか階層的に認識することができる。

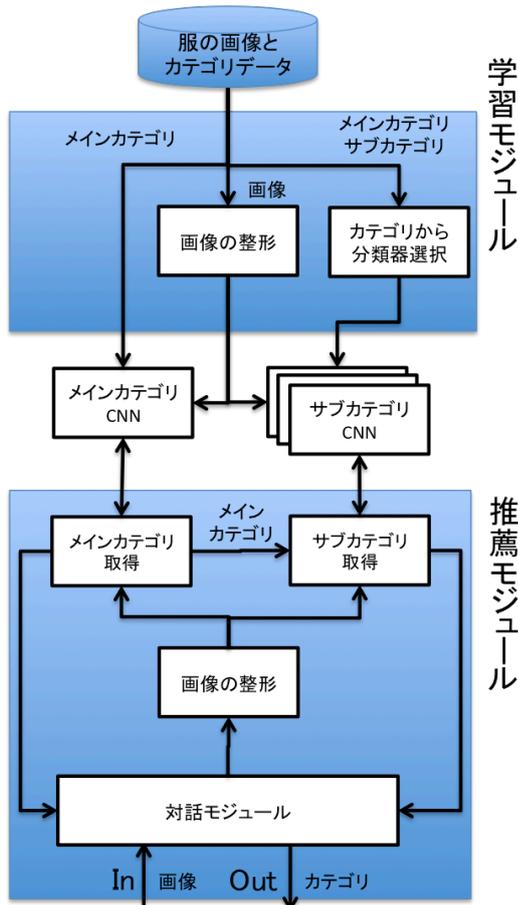


図 1: システム構成

4 節で使用したデータを説明し,5 節で画像の整形方法について説明する.6 節でカテゴリ判別方法の詳細を説明し,その結果を 7 節で考察する.

4. 服の画像データとカテゴリ

今回使用したデータは,服の画像データとそのカテゴリ 129,749 セットである.データは,tops,pants,skirt とメインカテゴリに分類されていて,更にサブカテゴリとして tops であれば cardigan,knit,hoodie など細かく分かれている.それぞれの詳細と,データ数を,表 1 に示した.データには,全身が写った画像,服だけの写真などが混ざっており,データを整形する必要がある.

5. 画像の整形

我々は,EC サイトから取得したデータを CNN の入力サイズに整形するため,opencv を用いて顔認識をし,顔がある画像であれば,顔領域の大きさの 3 倍の大きさの領域を顔領域の下から取得し,tops として抽出,pants,skirt であれば,更にその下の領域を抽出した.顔認識は,ノイズを取り除くために,opencv で

表 1: カテゴリ詳細

メインカテゴリ	サブカテゴリ	数
tops	cardigan	11380
tops	knit	6767
tops	hoodie	5107
tops	blouse	20613
tops	sweat	4885
tops	cutsew	38208
skirt	skirt	22519
skirt	denim	981
pants	pants	11475
pants	denim	7814

顔領域と判定した領域が,画像の半分より上であり,かつ,領域内の色情報を HSV 変換し,肌の色が 40 % 以上であった場合,顔であると判定した.図 2 に顔認識で得られた領域を青,tops の領域を赤,pants,skirt の領域を緑,opencv の誤認識と判定したものを水色で示した.



図 2: 領域抽出

6. 入力画像のカテゴリ判別

ここでは 5 節で整形したデータを入力することで,画像のカテゴリを判別する方法について説明する.

6.1 CNN を用いた判定

ZOZOTOWN の画像は背景やポーズが統一されていないが,画像のデータが充分あると考えたため,我々は分類器として CNN を用いた.コーディネートを詳細に推薦するために,細かいサブカテゴリで分類する実験を行った.4 節で示したように,サブカテゴリは 10 カテゴリになるため,10 クラスの分類を行った.

我々は,服のコーディネートを推薦するためには,上下どちらかであることを認識することが重要であると考え,階層的に分類する.ZOZOTOWN では大きく tops,skirt, pants のメインカテゴリに分かれているため,まず 3 クラスの分類を行った.そして,tops のサブカテゴリを詳細に分類するために 6 クラスの分類を行い,前者との分類結果を比較する.

6.2 CNN の構成

CNN の構成について説明する.3 クラス分類の CNN の構成を表 2 に示す.層種の conv,pool,fc はそれぞれ畳み込み層,プーリング層,全結合層を表し,ネットワークは 5 つの畳み込み層,3 つのプーリング層,3 つの全結合層から成る.活性化関数

表 2: CNN の構成

層種	パッチ	ストライド	出力マップサイズ	関数
data	-	-	227 × 227 × 3	-
conv1	11 × 11	4	55 × 55 × 96	ELU
pool1	3 × 3	2	27 × 27 × 96	-
conv2	5 × 5	1	27 × 27 × 256	ELU
pool2	3 × 3	2	13 × 13 × 256	-
conv3	3 × 3	1	13 × 13 × 384	ELU
conv4	3 × 3	1	13 × 13 × 384	ELU
conv5	5 × 5	1	13 × 13 × 256	ELU
pool5	3 × 3	2	6 × 6 × 256	ELU
fc6	-	-	1 × 1 × 4096	ELU
fc7	-	-	1 × 1 × 2048	ELU
fc8	-	-	1 × 1 × 3	softmax

には ELU を使い、このネットワークをランダムに初期化し、学習させる。ミニバッチサイズは 120 としたが、データに偏りが生じるため、ミニバッチにはすべてのメインカテゴリから同じだけの量を確率的に取得し、入力として使用した。

10 クラスの分類、6 クラスの分類については、3 クラス分類で用いた CNN の構成を最後の fc8 層の出力をそれぞれ 10, 6 に変更して学習させた。

7. 結果

10 クラスの分類の学習遷移を図 3 に示した。青色が学習データのエラー率の推移で、赤色がテストデータのエラー率の推移である。エラーは下がるものの、38% からエラーが下がらず、学習が進むに連れてオーバーフィッティングしている。

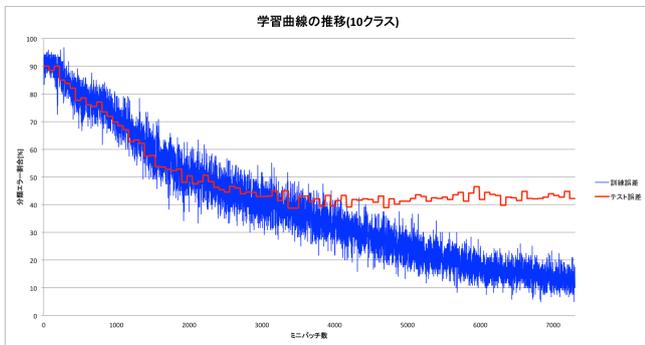


図 3: 10 クラス分類の学習曲線の推移

エラーが低くなった大きな理由は pants や skirt のメインカテゴリの分類がうまくできたためであり、tops のサブクラスについては、pants や skirt に比べて精度があまり良くなかった。また、tops のサブカテゴリである hoodie などがメインカテゴリを跨いで、pants のサブカテゴリである denim に分類されてしまう場合があった。我々の推薦システムで利用する場合、コーディネートを紹介するので、tops の画像を入力したら pants か skirt が出てくるのが望ましい。10 クラス分類の分類ミスがメインカテゴリを跨いだ場合、例えば tops の画像を入力して pants のカテゴリが出力されてしまう場合、tops のコーディネートとして tops が推薦されてしまうため、システムとしては望ましくない。

これを防ぐためにも、階層的に分類する方法の有用性があると考え。まず、メインカテゴリの 3 クラスの分類の結果を表

3 に示す。結果より、3 クラスの分類は、エラーが 10% 以下になり、精度が 90% 以上であることから、うまく学習ができていることが確認された。インターネット上のデータだけでなく、実際に町中でスマートフォンを使用して撮った画像や、服を床において撮った画像を入力することでも 90% 以上分類できることを確認した。

表 3: 3 クラス分類の結果とその正答率

メインカテゴリ	pants	skirt	tops	正答率
pants	1693	162	65	88%
skirt	94	1769	57	92%
tops	14	29	1877	98%

次に、tops 内のサブカテゴリの分類の結果を表 4 に示す。6 クラス分類は 3 クラス分類ほど精度が良好でなかった。理由としてあげられるのが、CNN では、層が深くなるに連れて畳み込み層とプリーング層によって出力される画素が荒くなるため、hoodie であれば、特徴であるフードのような局所的な情報を失ってことであると考えられる。しかし、カテゴリの階層性を用いて学習したことで、コーディネートのおすすめシステムが同じカテゴリを推薦するような明らかなエラーを防ぐことができることが示された。

表 4: 6 クラス分類の結果とその正答率

サブカテゴリ	bl	ho	sw	cu	ca	kn	正答率
blouse	205	24	27	67	87	70	43%
hoodie	9	279	55	14	90	33	58%
sweat	2	37	246	60	42	93	51%
cutsew	20	10	76	238	36	100	50%
cardigan	14	30	11	21	389	15	81%
knit	18	11	59	84	56	252	53%

8. まとめ

結果より、カテゴリを階層的に分類することで、分類精度が向上し、推薦システムへ有用性が高まることがわかった。今後の課題として、hoodie の特徴となるフードは画像に占める領域が小さく、出力層側では特徴が消えてしまっているため、局所的情報を取得する方法を考える必要がある。また、学習が明らかに過学習しているため、Batch Normalization[5] などの正規化層を入れる必要がある。

参考文献

- [1] ZOZOTOWN, <http://zozo.jp/>
- [2] 堀 和紀, 岡田 将吾, 新田 克己. オンラインファッションカタログを利用した画像とテキストからの組み合わせ推薦. JSAI, 2015.
- [3] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. ANIPS, 2012.
- [4] Yamaguchi, K. et al. Retrieving Similar Styles to Parse Clothing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.
- [5] Ioffe, S. and Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv, 2015.