

CNN による画像形状特徴を利用した推薦システム

Recommender System using Image shape Feature for CNN

段玉鋒^{*1}
Yufeng Duan

佐賀亮介^{*1}
Ryosuke Saga

^{*1} 大阪府立大学院人間社会システム研究科
Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University

PMF (Probabilistic Matrix Factorization) is a well-known approach of recommending systems. It has achieved wide use not only in research but also in industry. In recent years, this algorithm has been combined with some side information, and the accuracy of recommendation has been improved. In this paper, we propose a novel probabilistic model using image shape features that integrates convolutional neural network (CNN) into probabilistic matrix factorization (PMF)

1. はじめに

推薦システムは情報過多において、ユーザが好むと思われる問題を解決する手段である。特に、PMF(Probabilistic Matrix Factorization) [Salakhutdinov 08]はよく使われる推薦手法であり、広く使用されている。

一方、昨今では、ユーザ評価だけでなくテキストや画像のクラス情報を組み合わされたものが提案されている。例えば、Kimらは、ConvMF という畳み込みニューラルネットワーク(CNN)を確率的行列分解(PMF)に統合した新しい推薦モデルを提案している[Kim 16]。ConvMF は文章のコンテスト情報をキャプチャし更に推薦精度を向上させる。しかしながら、デザインを重要視している服や靴など、ファッションに関する服飾製品においては、テキストや画像のカテゴリだけでは十分でなく、画像に表れている商品の形状を考慮する必要がある。そして、現状、この画像形状の特徴を利用している推薦システムが存在しない。

本論文では、CNNを用いた画像形状特徴をPMFに統合し、デザインを重要視している商品に向けた推薦システムを提案する。そのための、画像形状するモデルを提案する。このモデルをISFMFと呼ぶ。

2. 関連研究

2.1 PMF

Salakhutdinovらは、PMF (Probabilistic Matrix Factorization) という情報推薦手法を提案している。確率行列分解とは、ユーザ数 N 、アイテム数 M 、整数 D と採点行列 $R \in \mathbb{R}^{N \times M}$ を仮定し、この R を $D \times N$ のユーザ行列 U と $D \times M$ のアイテム行列 V に行列分解することである。ユーザ i によるアイテム j の予測採点値 $\hat{r}_{ij}$ は、ユーザ行列 u_i とアイテム行列 v_j の内積である (i.e. $\hat{r}_{ij} = u_i^T v_j$)。最適な行列 U と行列 V を見つける方法は実際の採点値と予測の採点値の残差平方和からなる損失関数 E を最小にするであり、式(1)で表される。

$$\min E(U, V) = \sum_i \sum_j \frac{\lambda_j}{2} (r_{ij} - u_i^T v_j)^2 + \frac{\lambda_u}{2} \sum_i \|u_i\|^2 + \frac{\lambda_v}{2} \sum_j \|v_j\|^2 \quad (1)$$

2.2 DeepContour

Shen らは、DeepContour という輪郭検出手法を提案している[Shen 15]。教師あり学習で CNN に色々な画像の形状の特

徴を学習させ、図1のように形状のみを抽出した画像を生成できる。そして、Positive-Sharing Loss function という関数を開発した。この関数で画像のある所は輪郭かどうかを判断できる手法である。

3. 提案手法

3.1 ISFMF の確率モデル

図2はCNNをPMFに統合するISFMFの確率モデルの概要を示したものである。ユーザ数 N 、アイテム数 M と次元数 D があり、採点行列 $R \in \mathbb{R}^{N \times M}$ を仮定する。次に、採点行列 R が再構築できるユーザ潜在行列 $U \in \mathbb{R}^{D \times N}$ とアイテム潜在行列 $V \in \mathbb{R}^{D \times M}$ を求める。

しかし、従来のPMFのアイテム潜在行列と異なり、我々のアイテム潜在行列は三つの変数から生成されると仮定する。 S (アイテム j の画像形状特徴)、 W (W はCNNの全ての重みを示す変数で、 w_d は各重み) と σ_v (潜在行列 V のガウス雑音) である。従って、ISFMFの損失関数 E は式(2)で表される。



図1: DeepContour

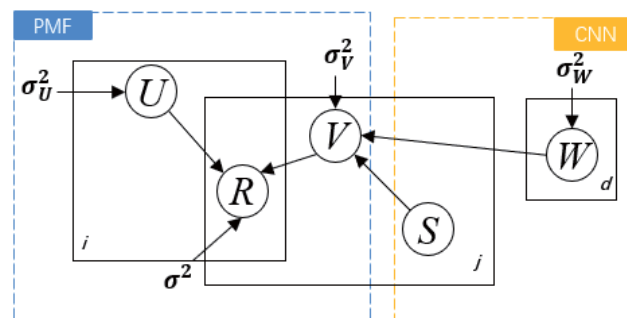


図2: ISFMF の確率モデル

連絡先: 佐賀亮介, 大阪府立大学院人間社会システム研究科, 599-8531, 大阪府堺市 中区 学園町 1-1, saga@cs.osakafu-u.ac.jp

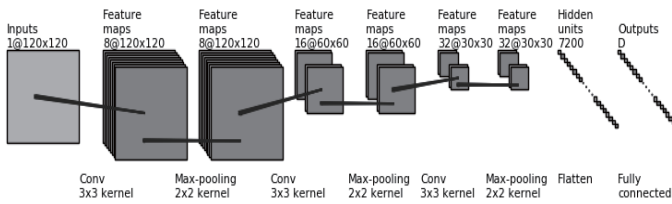


図 3: CNN のアーキテクチャ

$$\min E(U, V, W) = \sum_i \sum_j \frac{I_{ij}}{2} (r_{ij} - u_i^T v_j)^2 + \frac{\lambda_U}{2} \sum_i \|u_i\|^2 + \frac{\lambda_V}{2} \sum_j \|v_j - \text{cnn}(W, S_j)\|^2 + \frac{\lambda_W}{2} \sum_d \|w_d\|^2 \quad (2)$$

ただし、 $\lambda_U = \frac{\sigma^2}{\sigma_U^2}$, $\lambda_V = \frac{\sigma^2}{\sigma_V^2}$, $\lambda_W = \frac{\sigma^2}{\sigma_W^2}$ であり、 I_{ij} は評価値があるときに 1、それ以外は 0 をとる。また、 $\partial_U, \partial_V, \partial_W$ はそれぞれ U, V, W のガウス雑音である。式(2)を u_i と v_j それぞれ偏微分すると

$$\frac{\partial E}{\partial u_i} = (r_{i,j} - u_i^T v_j) v_j - \lambda_U u_i \quad (3)$$

$$\frac{\partial E}{\partial v_j} = (r_{i,j} - u_i^T v_j) u_i - \lambda_V (v_j - \text{cnn}(W, S_j)) \quad (4)$$

そして、確率的勾配降下法で u_i と v_j を更新し、最適なユーザ潜在行列 U とアイテム潜在行列 V を得ることができる。

3.2 ISFMF の CNN アーキテクチャ

図 3 は我々の CNN アーキテクチャである。DeepContour で処理したアイテム j の画像を入力し、次元数 D の画像形状潜在ベクトルが生成される。そして、生成されたベクトルは PMF に渡され、行列分解に組み込まれる。

4. 評価実験

4.1 実験概要・データセット

本実験では、服飾に関する商品において提案手法の性能を評価する。実験には Amazon product data [McAuley 15, He 16] の Clothing, Shoes and Jewelry の採点データと画像データを利用する。今回は服飾データのみを対象とするため、服飾以外のもの、画像がないもの、また商品説明がないものをデータセットから削除した。また、ユーザの採点数によってデータ密度が異なる三つのデータセットを作り、それらのデータセットを元に訓練データと評価データにランダムに分けた (表 1)。

- M1 : 採点数が 1 件未満のユーザを削除したデータセット
- M2 : 採点数が 2 件未満のユーザを削除したデータセット
- M3 : 採点数が 3 件未満のユーザを削除したデータセット

本実験では、訓練データと評価データを分割するために、訓練データの行列サイズは、元データの行列と同じサイズに設定している。ただし、M1 はユーザ数、アイテム数ともに巨大であり、データがあまりにも疎であり、訓練データと評価データに分割できないため、今回の実験では M1 に関しては使用しない。

また、本実験では、PMF, ConvMF と提案手法を比較した。ここで、パラメータは以下のように設定した。

- U と V の次元数 D を 50 に設定した。
- 入力画像のサイズを 120×120 に設定した。
- 共通パラメータ (λ_U, λ_V) は様々な組み合わせを事前に評価したものを用いた (表 2)。

また、評価尺度として、RMSE (root mean squared error) を用いた (式 (5))。

$$\text{RMSE} = \sqrt{\frac{\sum_{i,j}^{N,M} (r_{ij} - \hat{r}_{ij})^2}{\text{ratings}}} \quad (5)$$

4.2 結果と考察

各モデルの RMSE の結果を表 3 に示す。表 3 により、提案手法である ISFMF は全体的に最も良い精度が得られている。

5. おわりに

本稿では、CNN を用いた画像形状特徴を PMF に統合するモデルを提案した。デザインに重視するアイテムにおいて、画像形状特徴の有用性が証明された。

今後は商品のテキスト情報と画像形状特徴と一緒に利用するモデルを構築することを試みる。

参考文献

- [Kim 16] D. Kim, C. J. Oh Park, et al, Convolutional Matrix Factorization for Document Context-Aware Recommendation, ACM, pp.233-240, 2016
- [McAuley 15] J. McAuley, C. Targeet, J. Shi, A. Van den Hengel. Image-based recommendations on styles and substitutes, SIGIR, pp.43-52, 2015
- [He 16] R. He, J. McAuley, Modeling the visual evolution of fashion trends with one-class collaborative filtering, WWW, pp.507-517, 2016
- [Salakhutdinov 08] R. Salakhutdinov, A. Mnih, Probabilistic Matrix Factorization, NIPS, pp.1257-1264, 2008
- [Shen 2015] W. Shen, X. Wang, Y. Wang, X. Bai, Z. Zhang, DeepContour: A Deep Convolutional Feature Learned by Positive-Sharing Loss for Contour Detection, CVPR, pp.3982-3991

表 1: データ密度が異なる三つのデータセット

datasets	Users	Items	Ratings	density
M1	94562	22027	105994	0.0051%
M2	8222	8513	19654	0.0281%
M3	1796	4148	6802	0.0913%

表 2: 各モデルの最適な共通パラメータ

models	M2		M3	
	λ_U	λ_V	λ_U	λ_V
PMF	0.1	0.1	1	0.1
ConvMF	1	10	10	0.1
ISFMF	1	0.1	1	0.1

表 3: テストデータの RMSE

datasets	PMF	ConvMF	ISFMF
M2(70%訓練データ)	1.64	1.33	1.27
M2(80%訓練データ)	1.60	1.26	1.21
M3(70%訓練データ)	1.76	1.27	1.18
M3(80%訓練データ)	1.59	1.29	1.26