MLA'20 - The 18th China Symposium on Machine Learning and Applications
2020/11/06-08, Nanjing, China

University of Science and Technology of China

COLUMBIA UNIVERSITY

JDAI Research

Learning the Compositional Visual Coherence for Complementary Recommendations

Zhi Li[1], Bo Wu[2], Qi Liu[1;3;*], LikangWu[3], Hongke Zhao[4], Tao Mei[5]

[1]Anhui Province Key Laboratory of Big Data Analysis and Application, School of Data Science, University of Science and Technology of China (USTC); [2]Columbia University; [3]School of Computer Science and Technology, USTC; [4]The College of Management and Economics, Tianjin University; [5]JD AI Research

{zhili03, wulk}@mail.ustc.edu.cn; bo.wu@columbia.edu, qiliuql@ustc.edu.cn, hongke@tju.edu.cn, tmei@jd.com

Abstract

Complementary recommendations, which aim at providing users product suggestions that are supplementary and compatible with their obtained items, have become a hot topic in both academia and industry in recent years. Existing work mainly focused on modeling the co-purchased relations between two items, but the compositional associations of item collections are largely unexplored. Actually, when a user chooses the complementary items for the purchased products, it is intuitive that she will consider the visual semantic coherence (such as color collocations, texture compatibilities) in addition to global impressions. Towards this end, in this paper, we propose a novel **Content Attentive Neural Network (CANN)** to model the comprehensive compositional coherence on both global contents and semantic contents. Specifically, we first propose a *Global Coherence Learning (GCL)* module based on multi-heads attention to model the global compositional coherence. Then, we generate the semantic-focal representations from different semantic regions and design a *Focal Coherence Learning (FCL)* module to learn the focal compositional coherence from different semantic-focal representations. Finally, we optimize the CANN in a novel compositional optimization strategy. Extensive experiments on the large-scale real-world data clearly demonstrate the effectiveness of CANN compared with several state-of-the-art methods.

Background

- **Recommender systems** are those techniques that support users in the various decision-making process and catch their interest among the overloaded information.
- For enhancing user satisfaction and recommendation performances, it is an indispensable part to understand **how products relate to each other** in recommender systems.
- **Complementary recommendations,** which aim at exploring item compatible associations to enhance the qualities of each item or another, have become a hot topic in both academia and industry in recent years.

Motivation

- **Compositional Coherence** on both global visual content and semantic visual content are important for a visually-aware complementary recommender system.
  - ✓ **Global Coherence:** the compositional relationships of complementary items via global visual content;
  - ✓ **Semantic-focal Coherence:** the compositional relationships of complementary items via semantic visual content (such as color-focal, texture-focal and hybrid-focal).

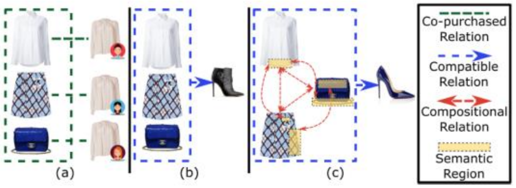
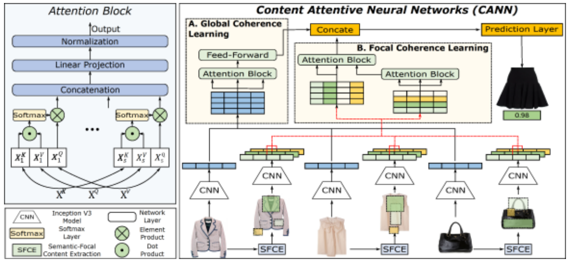


Figure 1: Illustration of complementary recommendations. (a) Recommendations based on co-purchased relations. (b) Recommendations based on compatible relations. (c) Recommendations based on compositional coherence.

Methodology: *framework*

- *Method Overview*
- **Content Attentive Neural Network (CANN)**
  - ✓ *Global Coherence Learning*
  - ✓ *Focal Coherence Learning*



Methodology: *Global Coherence Learning*

- *Image Feature Extractor:* **Inception-V3**
- Global Coherence Learning:
  - ✓ **Multi-head Attention Block**
  - ✓ **Attention Stack**

$$\alpha^s_{i,j} = \frac{W^q_s x^s_i \cdot (W^k_s x^s_j)^T}{\sqrt{d_s}}, \widehat{\alpha}^s_{i,j} = \frac{\exp(\alpha^s_{i,j})}{\sum_{j=1}^{k}\exp(\alpha^s_{i,j})}$$

$$M^{(b)}_a = f_a(h^{b-1}) \oplus h^{b-1},$$

$$h^{(b)} = BN(W_{bn} f_n(M^{(b)}_a)) + b_{bn}),$$

Methodology: *Focal Coherence Learning*

- *Three Semantic-Focal Contents:* Generating the semantic region based on the color or textual similarity computing
  - ✓ **Color-Focal Contents:** semantic regions in similar color
  - ✓ **Texture-Focal Contents:** semantic regions in similar texture
  - ✓ **Hybrid-Focal Contents:** semantic regions in similar color and texture
- *Hierarchical Attention Module:* Model the compositional relations in two aspects
  - ✓ **Semantic-specific attention**
  - ✓ **Cross-semantic attention**
  - ✓ **Final semantic-focal representation**

$$V' = AV = \hat{A}_C V_S = \hat{A}_C \hat{A}_S V,$$

Methodology: *Optimization Strategy*

- *Loss Function:*

$$Pr(\widehat{x}|P) = \frac{\exp(\widehat{x} \cdot x_c)}{\sum_{x_c \in \mathcal{N}} \exp(\widehat{x} \cdot x_c)}$$

$$L(P, \mathcal{N}; \theta) = -\frac{1}{|\mathcal{N}|}\sum_{t=1}^{|\mathcal{N}|} log Pr(\widehat{x}|P)$$

- Compositional Optimization Strategy
  - ✓ Random choose the prediction items in the sets
  - ✓ Other samples in mini-batch as training negative candidates

Algorithm 1 Compositional Optimization Strategy

Input: Initialization model $f(P_i, C; \theta)$; The length of the seed collection k; The complementary item database S; The number of epochs T; The size of batch m
Parameter: Model parameter θ

```
for i = 1, 2, 3, ..., T do
    Random sample m seed collections O ∈ S
    Initial input mini-batch Input as ∅
    for O_i in Batch do
        Random choose an item p from the collection O_i
        Generate the seed collection P_i ← Mask(O_i, p)
        if |P_i| < k then
            Add an padding to the left of P_i until |P_i| = k
        end if
        Generate the input mini-batch Input ← Input ∪ P_i
    end for
    Build the training candidates N ← ∀x ∈ Input
    Update the model θ ← SGD(f(Input, C; θ), θ)
end for
```

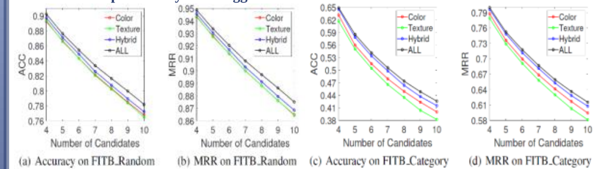
Experiments

- *Datasets: Polyvore (FITB_Random, FITB_Category)*
- *Metrics: Accuracy, MRR*
- *Comparison Methods:*
  - SetRNN
  - SiameseNet
  - VSE
  - Bi-LSTM
  - CSN-Best
  - NGNN
- *Variant Implements of Our Proposed CANN*
  - CANN-G
  - CANN-F
  - CANN

Approaches	FITB_Random		FITB_Category	
	Accuracy	MRR	Accuracy	MRR
SetRNN	29.6%	48.1%	28.7%	46.1%
SiameseNet	52.2%	71.6%	54.0%	72.8%
VSE	29.2%	49.1%	30.2%	53.2%
Bi-LSTM	83.6%	91.1%	58.2%	75.7%
CSN-Best	58.9%	76.1%	56.1%	74.2%
NGNN	87.3%	93.2%	57.3%	74.9%
CANN-G	88.8%	94.1%	62.4%	78.1%
CANN-F	71.9%	84.1%	56.7%	74.7%
CANN	**90.7%**	**95.1%**	**66.5%**	**80.9%**

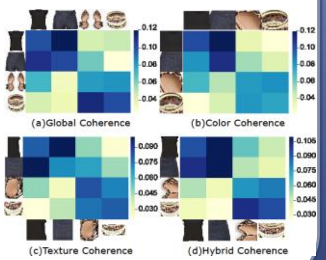
Recommendation Performances

- *Overview Results:*
  - CANN outperforms all the compared methods in both datasets, which indicates the superiority of our proposed model for content-based complementary recommendations.
  - It is advisable to model the compositional coherence of items on both global and semantic-focal contents.
- *Ablation Study*
  - CANN with all the semantic-focal contents has outperformed others, which clearly demonstrate the **effectiveness** of all components in our proposed CANN.
  - CANN with all semantic-focal contents outperforms other single semantic-focal models on FITB_Category with a larger margin than on FITB_Random. These observations imply that semantic-focal contents can help the model to better understand the **item compositional relationships** and generate **the best-matched complementary item suggestions.**



Ablation Study on Focal Coherence in Different Candidate Settings

- *Visualization of the Attention*
  - The coherence scores between the t-shirt and shorts are higher than others in all coherence spaces.
  - Scores between shoes and bracelet are also quite high. The shoes and bracelet are similar in **leopard print style**.
  - Our proposed CANN can provide a good way to capture the **visual coherence** for the complementary items from both global and semantic-focal views



Visualization of the Attention Mechanism

Anhui Province Key Laboratory of Big Data Analysis and Application, University of Science and Technology of China (USTC)