TransNFCM: Translation-Based Neural Fashion Compatibility Modeling

TaeHyung Noh

Lab Seminar

(22.04.25)

Contents

INTRODUCTION

Introduction

- **어떻게 하면 Fashion item을 잘 추천할 수 있을까?**

1. Search based → 사용자의 검색을 기반으로 추천
2. Mix and match based → 한 패션 아이템과 어울리는(compatible 한) item 추천,CCFR(cross-category fashion recommendation)

1번 방식은 이미 시각적 유사성, 관계를 학습하는 model을 통해 사용 중
2번 방식은 1번에서 사용한 것을 넘어서 **compatible한 관계를 계산**하여 modeling이 필요.

Introduction

- **어떻게 하면 Fashion Item 간의 Compatible한 관계를 계산할까?**

  - Compatible한 fashion item을 embedding하여 item간의 Euclidean distance가 가까우면 어떨까?

그러나 fashion compatibility는 단순 거리만으로 표현할 수 없는 **복잡한(다차원적인) 관계**를 가지고 있다.

Ex) Category-attribute와 같은 관계는 compatibility 계산을 복잡하게 만든다.

Introduction

- **TransNFCM(translation-based neural fashion compatibility modeling) 제안**

  - Data-dependent compatibility function 디자인

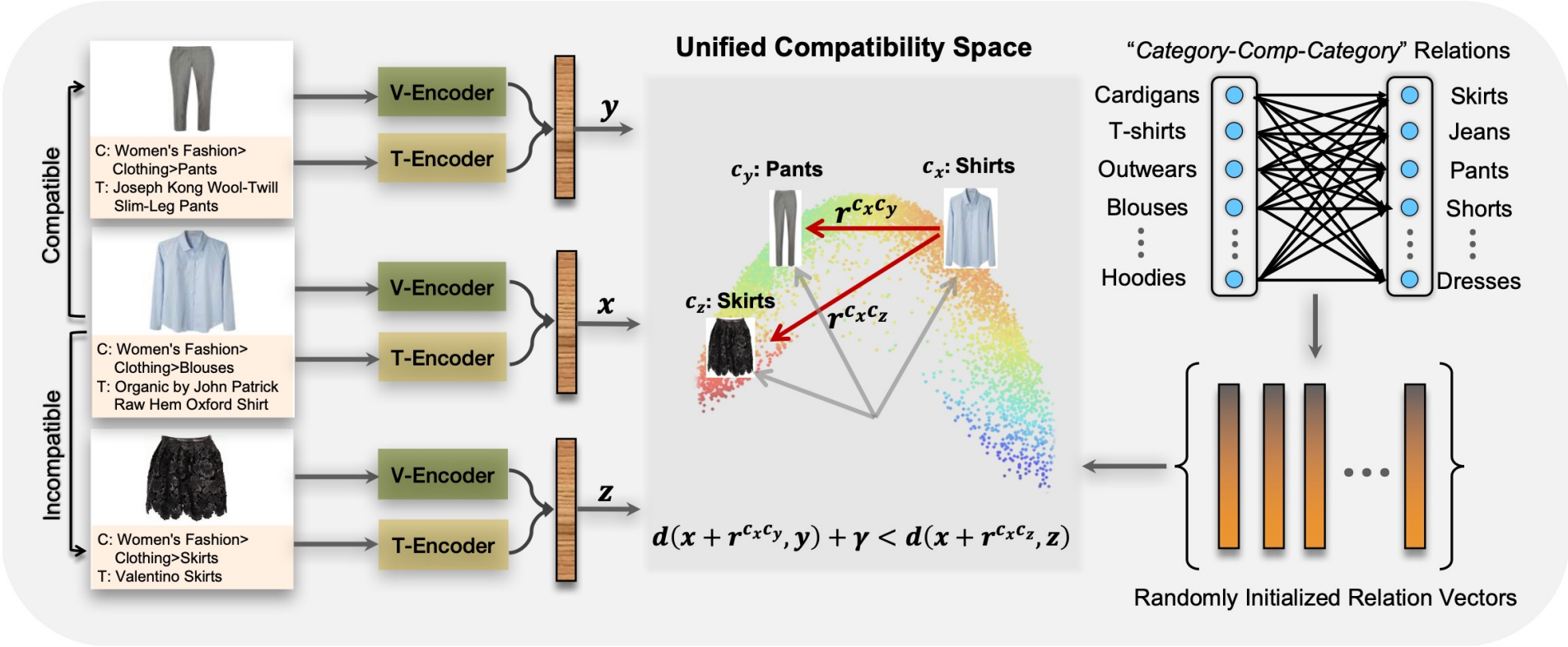
    - Fashion category 간의 관계를 고려(Category-complementary relations)

  - Encoder을 통해 visual, text feature을 뽑아 패션 item을 하나의 vector space로 embedding 하고 vector translation을 통해 item간의 관계 표현

Fashion category pair c_x, c_y, 에 대해 $x + r^{c_x c_y} \approx y$ 을 만족할 수 있도록 학습

Data-dependent distance function → $P\big((x, y) \in \mathcal{P}\big) \propto -d(x + r^{c_x c_y}, y)$

Introduction

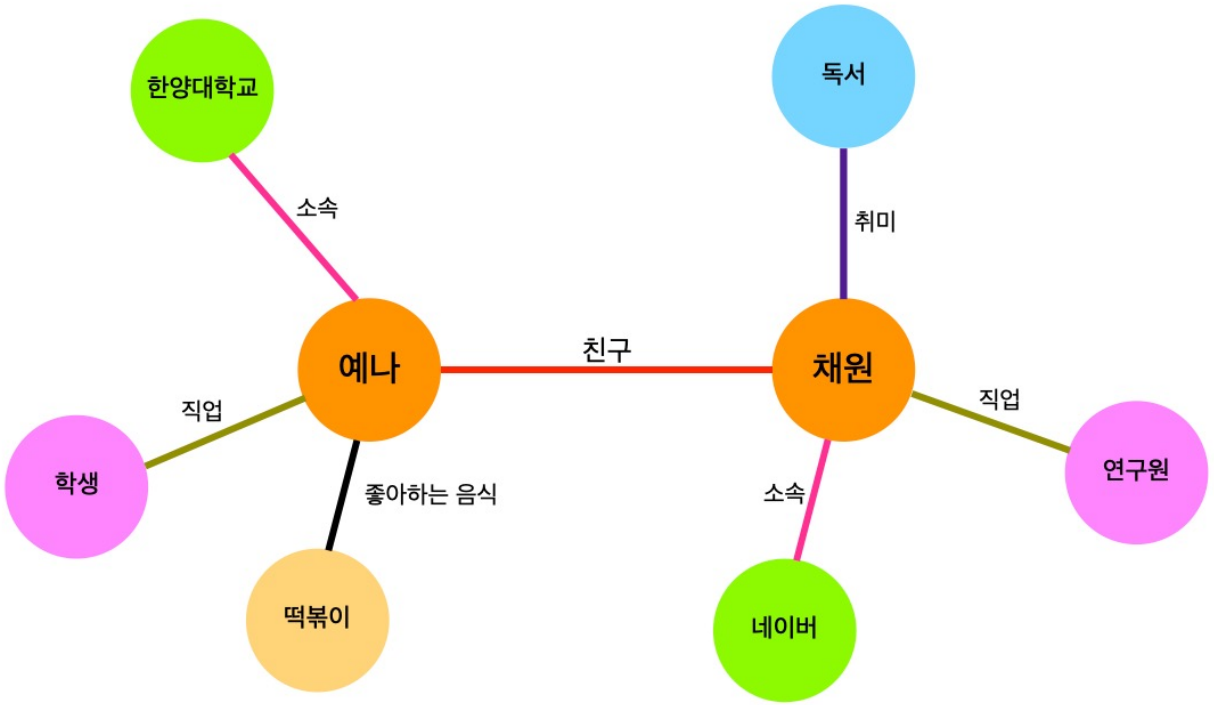


Background

- **Fashion compatibility**
  - Fashion recommendation을 위해 compatibility를 연구
  - 웹 상에서 compatible한 item을 크롤링(Amazon에서 한 item과 같이 나오는 '다른 사람들이 같이 본, 구매한 item')하여 분석
  - 단순 visual feature만 뽑아 embedding하고, compatible한 item 간의 거리를 좁히는 방식으로 modeling
  - Fashion item 간의 관계를 정의하여 modeling에 사용하고, text feature도 사용하기 시작
  - 본 연구의 특징은 **fashion item 간의 관계를 vector space에 transform**하여 모델링, **Global 한 특징과 category-specific 한 특징**까지 고려

Background

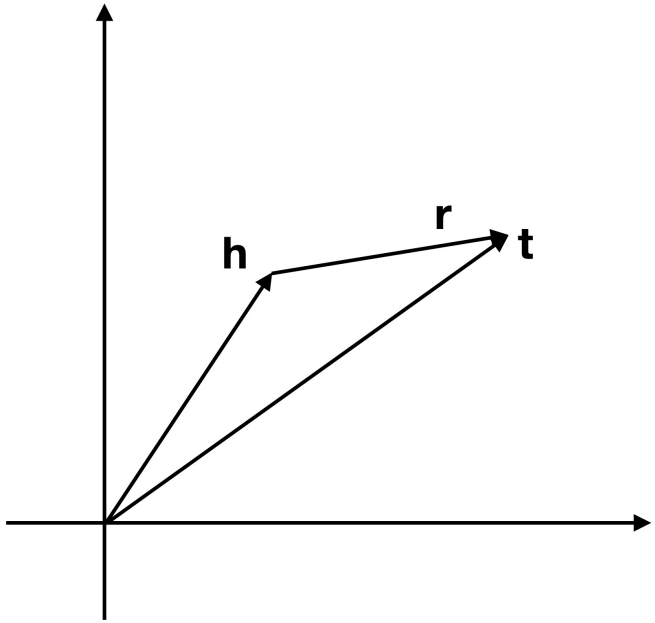
- **Knowledge Graph**
  - 지식을 그래프 형태로 나타낸 것.
  - Entity, Relationship 으로 구성



Background

- **Knowledge Graph Embedding**
  - 그래프 내부의 의미를 어떻게 추출할까? -> graph embedding
  - Embedding을 통해서 entity를 vector화 할 수 있고, vector간 다양한 연산을 할 수 있다.
  - Direct graph 데이터는 Triplet 형식으로 나타낼 수 있다.

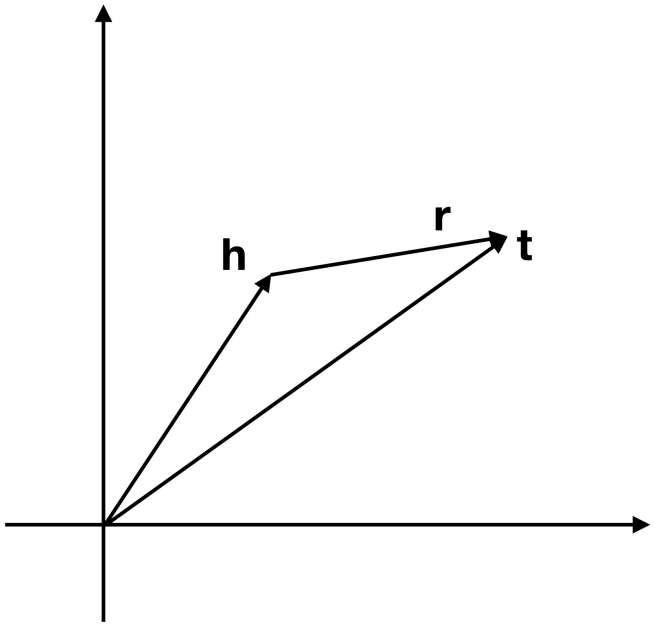
    - (head, tail, relation) 와 같은 구조



Background

- **TransE(Translating embeddings for modeling multi-relational data)**
  - H, R을 어떻게 embedding해야 T를 찾을 수 있을까??
  - TransE 방식은 h, r, t가 있을 때 R^d차원으로 나타내는 것
  - $f_r(h,t) = \|h + r\ - t\|_2^2 \rightarrow 0$ 에 가까워지면 좋은 목적 함수
  - $L(h,r,t) = \max(0, d_{pos} - d_{neg} + margin)$

    → Loss 함수, positive sample과의 distance가 negative sample과의 distance보다 가까워야 한다.



head + relation = tail

Contributions

- Fashion compatibility model인 TransNFCM 개발
- Category-comp-category를 vector space 에 translation하는 방식을 통해 category 관계를 compatibility modeling에 반영
- Fashion item에서 feature encoder을 통해 multimodal 한 fashion item의 특성 반영
- 두개에 dataset에 실험을 진행하여 TransNFCM의 성능 입증

METHODOLOGY

Methodology: Multimodal Item Encoder

- **Multimodal Item Encoder**
  - 온라인 상의 Fashion item은 multimodal data로 표현 되어있다.
  - 본 연구에서는 V-Encoder와 T-Encoder을 설계하여 multimodal data의 특성을 모두 활용할 수 있도록 한다.

$f_V(v_x)$ → item x 를 표현하는 image v_x를 feature space $\mathcal{R}^d$에 뿌리는 V-Encoder함수

$f_T(t_x)$ → item x 을 설명하는 text t_x를 feature space $\mathcal{R}^d$에 뿌리는 T-Encoder함수

Methodology: Fashion compatibility Modeling

- Fashion compatibility modeling
  - 이전 연구에서는 다양한 방법으로 compatibility 계산

1. Inner-product
2. Euclidean distance
3. Probabilistic mixtures of multiple distance
4. Conditional similarity

Methodology: Fashion compatibility Modeling

- **Fashion compatibility modeling**
  - **기호 설명**

1. (x, y) : compatible 한 item x, y (e.g., Shirt, Pants)
2. $\mathbb{x} \in \mathcal{R}^D, \mathbb{y} \in \mathcal{R}^D$: x, y 의 feature vector
3. $P\big((x, y) \in \mathcal{P}\big)$: x, y 가 서로 compatible할 확률
4. $d(x\ , y)$: x, y 사이의 거리

Methodology: Fashion compatibility Modeling

- **Fashion compatibility modeling**
  - **Inner product**

$$P\big((x, y) \in \mathcal{P}\big) \propto \mathbb{x}^T \mathbb{y}$$

(x, y) 의 feature vector 간의 내적이 compatible할 확률에 비례

내적의 의미

1. 크기가 1인 벡터가 있을 때 두 벡터가 평행하면 내적 값은 1
2. 두 벡터가 수직일 경우 내적 값은 0
3. 두 벡터가 이루는 각도가 0 ~ 90일 경우 내적 값은 0 ~ 1 사이 값
4. 결국 두 벡터의 방향이 같을 수록(벡터가 닮을 수록) 내적 값이 커진다.

Methodology: Fashion compatibility Modeling

- **Fashion compatibility modeling**
  - **Euclidean distance**

$$d(x, y) = \|\mathbb{x} - \mathbb{y}\|_2^2 = \|\mathbb{x}\|_2^2 + \|\mathbb{y}\|_2^2 - 2\mathbb{x}^T\mathbb{y}$$

$$P\big((x, y) \in \mathcal{P}\big) \propto -d(x, y)$$

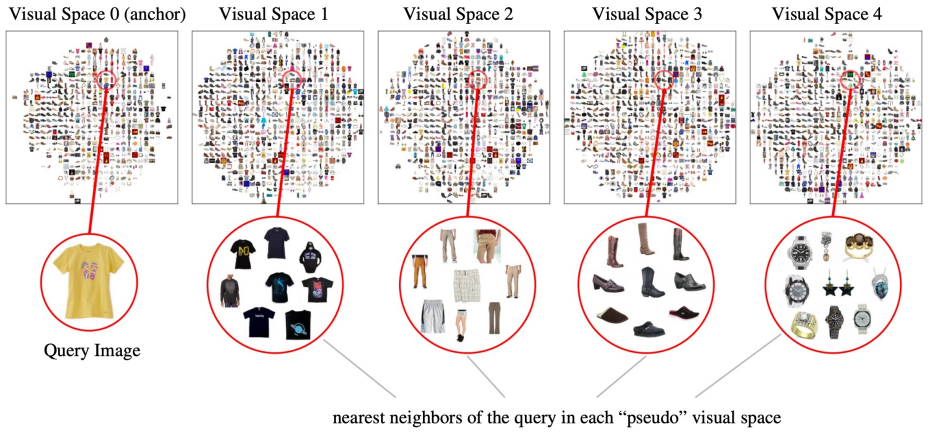
(x, y) 의 feature vector 간의 거리가 가까울수록 compatible

Methodology: Fashion compatibility Modeling

- **Probabilistic mixtures of multiple distances**
  - **Mahalanobis distance 사용**

$$d(x,y) = \sum_k P(k|x,y)d_k(x,y) = \sum_k P(k|x,y) \left\| \mathrm{E}_0^T \mathbb{x} - \mathrm{E}_k^T \mathbb{y} \right\|_2^2$$

$d_k(x,y)$ 를 구할 때 mahalanobis distance 사용 → x, y 가 얼마나 같이 발생하기 힘든 값인지 수치화 하는 방법, 서로 비슷하다면 distance 값이 작다.



Methodology: Fashion compatibility Modeling

- **Probabilistic mixtures of multiple distances**
  - **Conditional similarity**

$$d\left(\mathbb{x}_i^{(u)}, \mathbb{x}_j^{(v)}, \mathbb{w}^{(u,v)}\right) = \left\| f(\mathbb{x}_i^{(u)}; \theta) \odot \mathbb{w}^{(u,v)} - f(\mathbb{x}_j^{(v)}; \theta) \odot \mathbb{w}^{(u,v)} \right\|_2^2$$

$$= (f(\mathbb{x}_i^{(u)}; \theta) \odot \mathbb{w}^{(u,v)})^2 + (f(\mathbb{x}_j^{(v)}; \theta) \odot \mathbb{w}^{(u,v)})^2 - 2\left(f\left(\mathbb{x}_j^{(v)}; \theta\right) \odot \mathbb{w}^{(u,v)}\right)\left(f\left(\mathbb{x}_i^{(u)}; \theta\right) \odot \mathbb{w}^{(u,v)}\right)$$

Type u와 v의 관계 벡터인 $\mathbb{w}^{(u,v)}$을 사용하여 type간의 특성을 반영하였다.

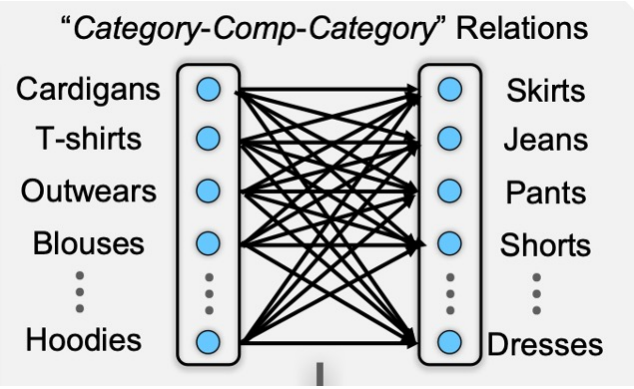
Methodology: Fashion compatibility Modeling

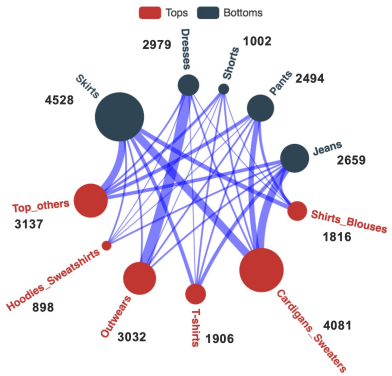
- **이전 모델링의 문제점**
  - 내적, 유클리드 거리 계산의 경우 데이터의 특성을 반영 X, Global한 개념만 compatibility에 반영 된다.()
  - Mahalanobis distance 은 데이터의 특성을 반영하면서, Global한 개념도 반영함, 그러나 M+2 projection을 최적화 해야 하는 문제가 있다.
  - Conditional similarity의 경우 처음으로 type간의(category) 특성을 반영, 그러나 global한 개념이 반영되지 않았다.

이 논문의 compatibility 계산은 **Global한 개념 + category의 특징**을 포함하여 진행

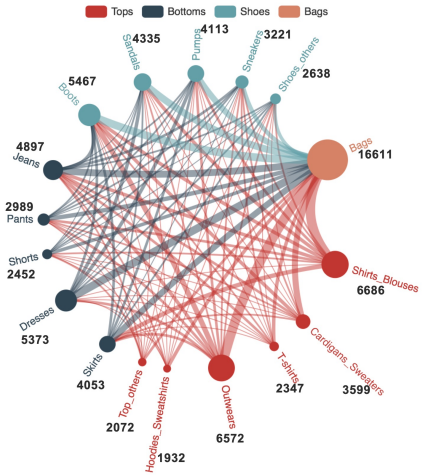
Methodology: Fashion compatibility Modeling

- **Translation-based Compatibility Modeling**
  - 이전 연구는 global notion을 학습하기 위한 single-relational data modeling 진행
  - 그러나 compatibility는 multi-dimensional 한 개념
  - 드레스, 코트... 와 같은 다양한 카테고리가 있고, A와 B 아이템이 주어졌을 때 전문가들은 아이템이 어울리는지를 결정 내릴 때 attribute(색상, 길이...)를 고려한다.
  - Single-relational data modeling → multi-relational data modeling
  - Category-complementary relationship을 정의(category-complementary-category)





(a) FashionVC



(b) PolyvoreMaryland

Methodology: Fashion compatibility Modeling

- **Translation-based Compatibility Modeling**

$$d(\mathbb{x} + r^{C_x C_y}, \mathbb{y}) = \left\| \mathbb{x} + r^{C_x C_y} - \mathbb{y} \right\|_2^2$$

$$= \|\mathbb{x}\|_2^2 + \|\mathbb{y}\|_2^2 + \left\| r^{C_x C_y} \right\|_2^2,$$

$$\underbrace{-2\mathbb{x}^T\mathbb{y}}_{\text{global}} \underbrace{- 2(\mathbb{y} - \mathbb{x})^T r^{C_x C_y}}_{\text{Category-specific}}$$

- Global 한 특성 + category-specific 한 특성을 모두 반영하는 distance function 사용
- Probabilistic mixtures of multiple distances 방식에서 global한 notion을 추가 반영

Methodology: Fashion compatibility Modeling

- **Loss**
  - 학습을 위해 Incompatible한(negative) triplets를 생성
  - $(x',\ y),\ (x,\ y')$이 서로 incompatible한 item pair
  - Incompatible한 item을 random한 item을 선정하여 학습에 사용, TransE 논문에서도 같은 방식을 사용하여 진행하였기 때문에 본 논문에서도 같은 방식을 사용(figure 4에서 성능 비교를 통해 random한 방식의 우위를 입증

$$\mathcal{L} = \sum_{\mathcal{T}} \left[d(\mathbf{x} + \mathbf{r}^{c_x c_y}, \mathbf{y}) - d(\mathbf{x}' + \mathbf{r}^{c_{x'} c_{y'}}, \mathbf{y}') + \gamma \right]_+ \quad (7)$$

$$max(0, \gamma + d(\mathbb{x} + r^{C_x C_y}, \mathbb{y}) - d(\mathbb{x}' + r^{C_x C_y}, \mathbb{y}'))$$

  - Margin-based ranking criterion으로 loss 함수를 정의
  - Positive 한 아이템은 서로 가까워지고, Negative한 아이템은 서로 멀어져야 하는 loss
  - Margin 의 경우 연구자가 선택하는 parameter

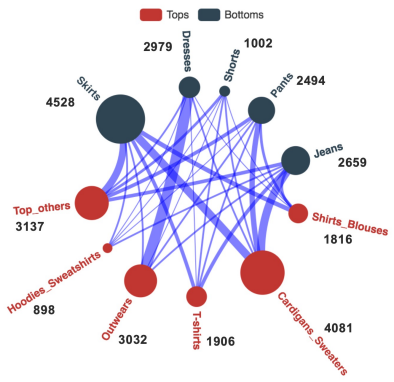
Methodology: V-Encoder, T-Encoder

- **V-Encoder:** AlexNet 사용, 마지막 FC layer을 d-dimensional embedding할 수 있도록 변경
- **T-Encoder:** CNN architecture 사용, 4개의 filter window로 preprocessing 후 300-D word2vec로 word 표현. 마지막에 d-dimensional textual embedding을 위한 max-pooling layer을 추가
- V-Encoder와 T-Encoder의 결과는 l_2 normalize 후 embedding
- 저자들은 논문에서 사용한 multimodal feature fusion strategies 말고도 score-level fusion을 TransNFCM에 사용할 수 있다고 함

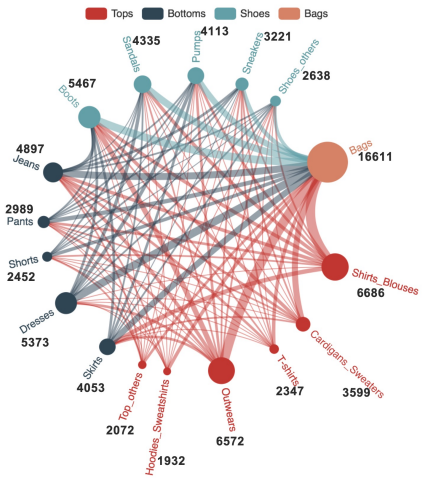
Experiments & Result

Experiments: Datasets

- **Datasets**
  - **FashionVC**
    - Top-bottom 추천을 위해 만들어진 datasets
    - Top-bottom 쌍으로 구성(top: 14,871, bottom: 13,663)
  - **PolyvoreMaryland**
    - Outfit compatibility modeling을 위해 만들어진 datasets
    - Whole outfit에 대한 compatibility 계산을 하는 것이 아니므로 5개의 item 중에서 tops, bottoms, shoes, bags 만 사용



(a) FashionVC



(b) PolyvoreMaryland

Experiments: Evaluation Protocols

- **Evaluation Protocols**
  - Compatible한 쌍 (h_i, t_{ig})에서 t_{ig}를 $\{t_{in}\}_{n=1}^{N}(N = 100)$으로 대체
  - 추천 시스템의 평가 기준인 AUC, HR@K 사용

$$\text{AUC} = \frac{1}{N|\mathcal{P}_t|} \sum_i \sum_n \delta\big(s(h_i, t_{ig}) > s(h_i, t_{in})\big) \qquad (8)$$

  - $\delta(a)$는 a가 true일 때 1을 return 하는 indicator 함수, $s(h_i, t_{ig}) > s(h_i, t_{in})$이면 1을 return한다.
    ($s(h_i, t_{ig}) = P((h_i, t_{ig}) \in \mathcal{P}_t)$, $|\mathcal{P}_t| = total\ number\ of\ testing\ positive\ pairs$)
  - HR@K는 K개에 예측 중에서 정답이 있으면 Hit 로 계산하는 방식

Experiments: Comparison Method

- **Siamese Network**
  - Euclidean distance 사용, contrastive loss
- **Triplet Network**
  - Euclidean distance 사용, margin-based ranking criterion
- **BPR**
  - Inner-product 사용, soft-margin based objected loss
- **Monomer**
  - Probabilistic mixtures of multiple distances 사용, same objective function
- **CSN**
  - Probabilistic mixtures of multiple distances 사용

Experiments: Comparison with another method

Table 1: Comparison on the FashionVC and PolyvoreMaryland datasets based on two metrics: AUC (%) and Hit@K (%, K∈{5, 10, 20, 40}). A larger number indicates a better result. **V** and **T** denote **V**isual modality and **T**extual modality, respectively. **V+T** denotes the fusion of visual and textual modalities. *100 negative candidates are sampled for each query during testing.* The best results are shown in boldface.

Features	Methods	FashionVC					PolyvoreMaryland				
		AUC	Hit@5	Hit@10	Hit@20	Hit@40	AUC	Hit@5	Hit@10	Hit@20	Hit@40
V	SiaNet	60.4	9.7	18.1	31.2	52.8	59.1	8.3	15.5	29.0	51.8
	Monomer	70.2	16.9	28.6	45.8	69.1	70.5	17.6	28.9	45.7	69.0
	CSN	71.6	16.7	28.4	46.7	70.8	70.2	17.3	28.4	45.1	68.4
	BPR	70.9	16.7	27.3	46.7	70.4	69.5	17.3	28.2	43.9	67.5
	TriNet	70.6	16.3	28.0	45.7	69.6	70.1	18.1	28.7	44.9	68.3
	TransNFCM	73.6	19.0	32.3	51.6	74.0	71.8	18.9	30.6	48.1	70.5
T	SiaNet	66.1	10.8	21.0	37.9	61.1	62.3	8.3	16.2	32.0	56.3
	Monomer	68.8	16.5	26.9	42.1	64.8	63.3	10.1	18.8	33.9	58.1
	CSN	67.5	11.2	22.4	41.2	64.1	63.2	8.8	17.0	32.5	57.4
	BPR	70.9	15.4	26.8	45.6	67.6	67.8	13.0	23.6	40.3	65.3
	TriNet	71.3	16.5	28.9	46.4	69.2	68.4	13.7	24.4	41.5	65.8
	TransNFCM	72.6	18.9	30.0	47.9	70.8	68.8	14.7	25.8	42.2	66.0
V+T	**TransNFCM**	76.9	23.3	38.1	57.1	77.9	74.7	21.7	34.4	52.7	75.3

Experiments: Performance Comparison and Anlysis

- **TransNFCM이 다른 method에 비해 가장 높은 성능**
  - V modality에서 좋은 성능 향상을 보임
  - T modality에서는 TriNet보다 성능 개선이 적었는데, 이는 Polyvore datasets의 text가 noise가 많고 sparse하기 때문
  - CSN과 비교했을 때, CSN은 global notion을 고려하지 않기 때문에, TransNFCM 보다 낮은 성능을 보임
  - V feature, T feature을 같이 사용하여, visual 특성은 text에 있는 noise와 sparse한 특성을 보완하고, text 특성은 visual에서 얻지 못했던 계절, style 같은 특성을 얻어 서로 보완할 수 있음

Experiments: Empirical analysis

- Global notion과 Category-specific을 같이 평가하는 것이 얼마나 효과가 있는가?

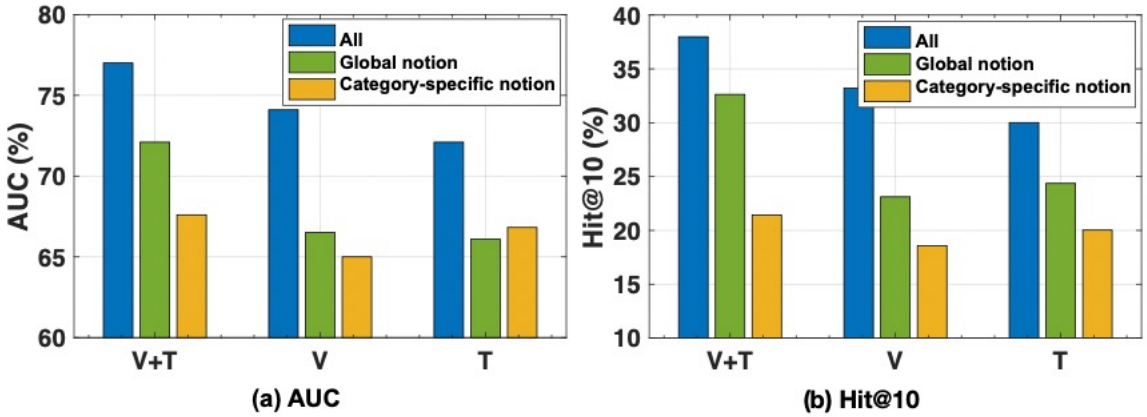


Figure 3: Effects of using different parts in Eq. (5) for compatibility computing on FashionVC. *Global notion* refers to using $\mathbf{x}^T\mathbf{y}$, *Category-specific notion* refers to using $(\mathbf{y}-\mathbf{x})^T\mathbf{r}^{c_x c_y}$, and *All* refers to using Eq. (5).

Experiments: Empirical analysis

- Target category is known

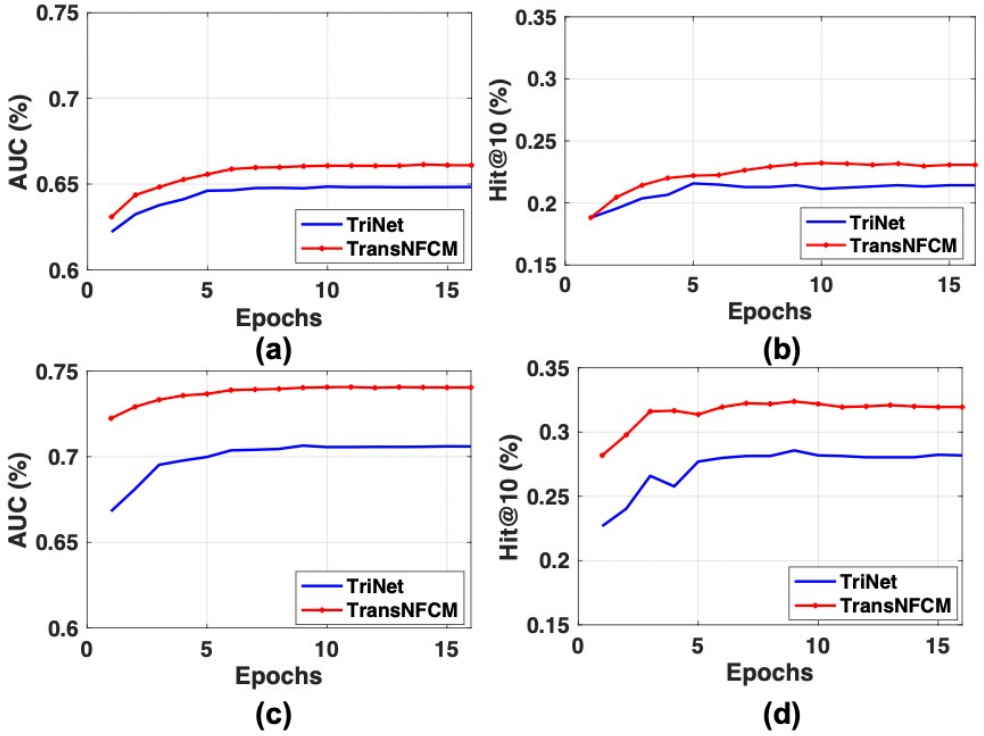


Figure 4: (a) (b) denote the comparison with TriNet when target category is known, (c) (d) denote the comparison with TriNet when target category is unknown. Experiments are conducted on FashionVC. Only visual modality is used.

DISCUSSION POINT

Discussion points

- Vector transform에서 TransE 방식만 사용하는 것이 아닌 다른 방식도 같이 사용하여 성능 비교를 진행하면 TransE를 사용하는 이유를 확인할 수 있었을 것 같습니다.

- 논문에서도 패션 전문가들이 item의 compatible을 평가할 때 attribute를 사용한다 하였는데, text data에 더 detail한 Fashion attribute 을 활용하여 modeling 되었으면 전문가의 의사결정을 도울 수 있는 modeling이 되지 않았을까 에 대한 아쉬움이 있습니다.(dataset의 한계)

- Evaluation 을 진행할 때 다른 model은 V+T에 대한 평가를 하지 않았는데, 성능을 같이 비교해 주면, TransNFCM 성능을 좀 더 확인할 수 있었을 것 같습니다.

Thank you