

Factorization Machines

<https://ieeexplore.ieee.org/document/5694074>

0. Introduction

- 추천 시스템에서 사용자-아이템 간 상호작용을 모델링하는 방법 소개
- 기존 행렬 분해(Matrix Factorization)는 2차 상호작용만 표현 가능하다는 한계 있음
- Factorization Machines(FM)는 고차원 희소 데이터에서도 효과적으로 상호작용을 학습할 수 있도록 설계됨

1. Overview

Feature vector \mathbf{x}																	Target y					
$\mathbf{x}^{(1)}$	1	0	0	...	1	0	0	0	...	0.3	0.3	0.3	0	...	13	0	0	0	0	...	5	$y^{(1)}$
$\mathbf{x}^{(2)}$	1	0	0	...	0	1	0	0	...	0.3	0.3	0.3	0	...	14	1	0	0	0	...	3	$y^{(2)}$
$\mathbf{x}^{(3)}$	1	0	0	...	0	0	1	0	...	0.3	0.3	0.3	0	...	16	0	1	0	0	...	1	$y^{(2)}$
$\mathbf{x}^{(4)}$	0	1	0	...	0	0	1	0	...	0	0	0.5	0.5	...	5	0	0	0	0	...	4	$y^{(3)}$
$\mathbf{x}^{(5)}$	0	1	0	...	0	0	0	1	...	0	0	0.5	0.5	...	8	0	0	1	0	...	5	$y^{(4)}$
$\mathbf{x}^{(6)}$	0	0	1	...	1	0	0	0	...	0.5	0	0.5	0	...	9	0	0	0	0	...	1	$y^{(5)}$
$\mathbf{x}^{(7)}$	0	0	1	...	0	0	1	0	...	0.5	0	0.5	0	...	12	1	0	0	0	...	5	$y^{(6)}$
	A	B	C	...	TI	NH	SW	ST	...	TI	NH	SW	ST	...	Time	TI	NH	SW	ST	...		
	User				Movie					Other Movies rated						Last Movie rated						

- FM은 일반화된 선형 모델에 잠재 벡터를 도입하여 특징 간의 상호작용을 자동으로 학습
- 데이터가 희소(sparse)해도 안정적으로 작동
- 추천 시스템, 클릭 예측, 랭킹 문제 등 다양한 응용 가능

2. Challenges

- 고차원 데이터에서 전통적 방법은 차원의 저주 발생
- Sparse feature가 많아 학습이 불안정해질 수 있음
- 단순 행렬 분해는 새로운 feature 조합을 잘 반영하지 못함

3. Method

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$

두 번째 항: 1차 특징 가중치

세 번째 항: 잠재 벡터 내적을 통한 2차 상호작용

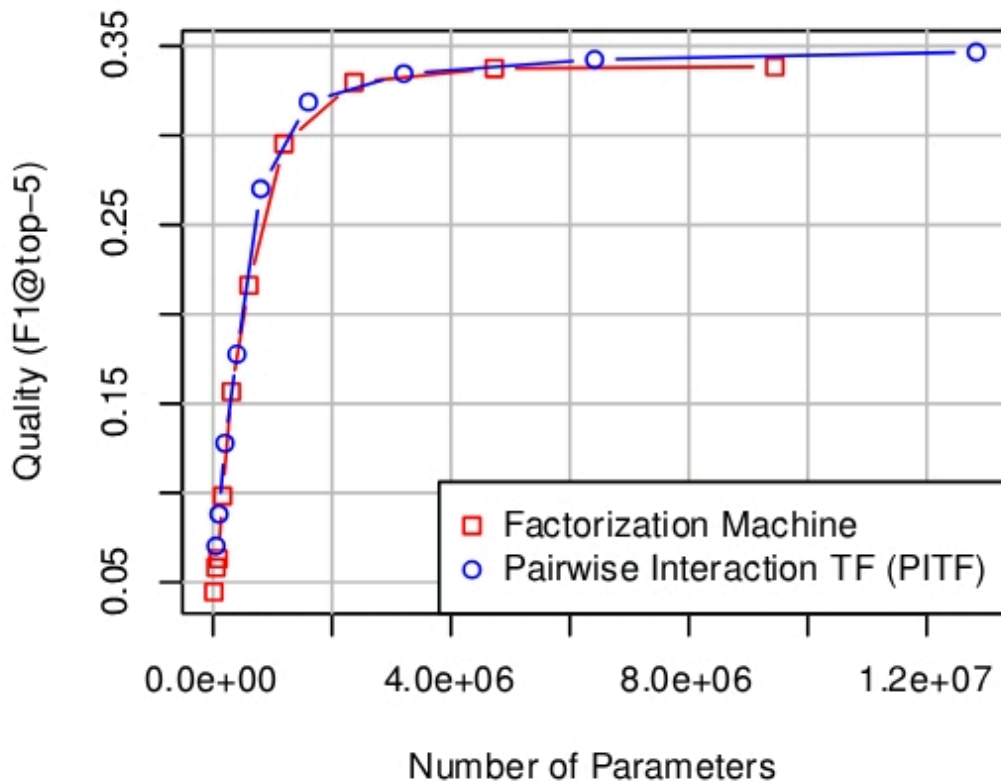
이렇게 설계함으로써 일반적인 회귀모델 + 행렬 분해의 장점을 동시에 가짐

4. Experiments

- 데이터셋: MovieLens, 클릭 예측 데이터 사용
- baseline: Logistic Regression, SVM, Matrix Factorization 비교

5. Results

ECML Discovery Challenge 2009, Task 2



- sparse 데이터셋에서 FM이 consistently 높은 성능 보임
- feature 조합을 명시적으로 하지 않아도 다양한 interaction 포착 → feature engineering 비용 절감
- 추천 시스템 외 CTR(Click-Through Rate) 예측, 태그 추천 등 다목적 활용 가능성 입증

6. Insight

- FM은 희소 데이터와 고차원 feature 문제를 동시에 해결
- 단순한 구조로 다양한 영역에 적용 가능
- 향후 확장 연구로 DeepFM, Neural FM 같은 신경망 결합 모델 등장 기반 마련