Movie recommendation based on users rating

Bachir El Khadir

May 5, 2016

Idea

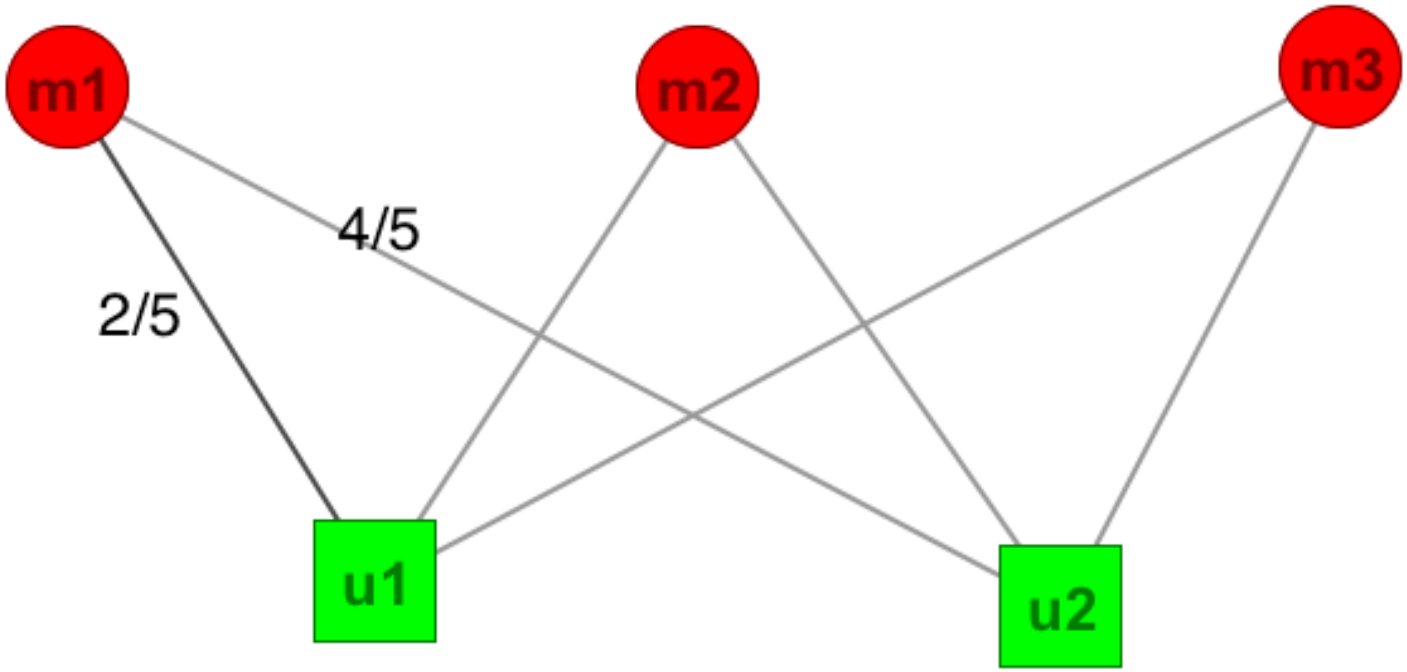
- Make movie recommendations for users.
- Problem of interest to many companies: Netflix, Youtube ...
- New approch based on graph modeling and natural language processing.

Movielens Dataset

users	movies	rating
0	1	5
2	9	3
2	4	4
3	10	3

- $80K$ rows
- $1.6K$ movies, ~ 50 ratings per movie.
- 943 users, ~ 80 ratings per user.

Modeling



$$d(m_1, m_2) = ?$$

Tools

- Language: Python.
- Pandas for data manipulation.
- Word2Vec toolkit for statistical modeling.

Generate a random walk on the graph

- Each user u implies a distribution over movies:
  - $\mathbb{P}_u(m) \propto$ rating of u to the movie m
- Each movie m implies a distribution over users:
  - $\mathbb{Q}_m(u) \propto$ rating of u to the movie m

- Start with an empty list of movies $L = [\,]$.
- Loop
  - Select a random user u
  - Sample $m \sim \mathbb{P}_u$
  - `L.append(m)`.
  - sample $u' \sim \mathbb{Q}_m$
  - $u \leftarrow u'$

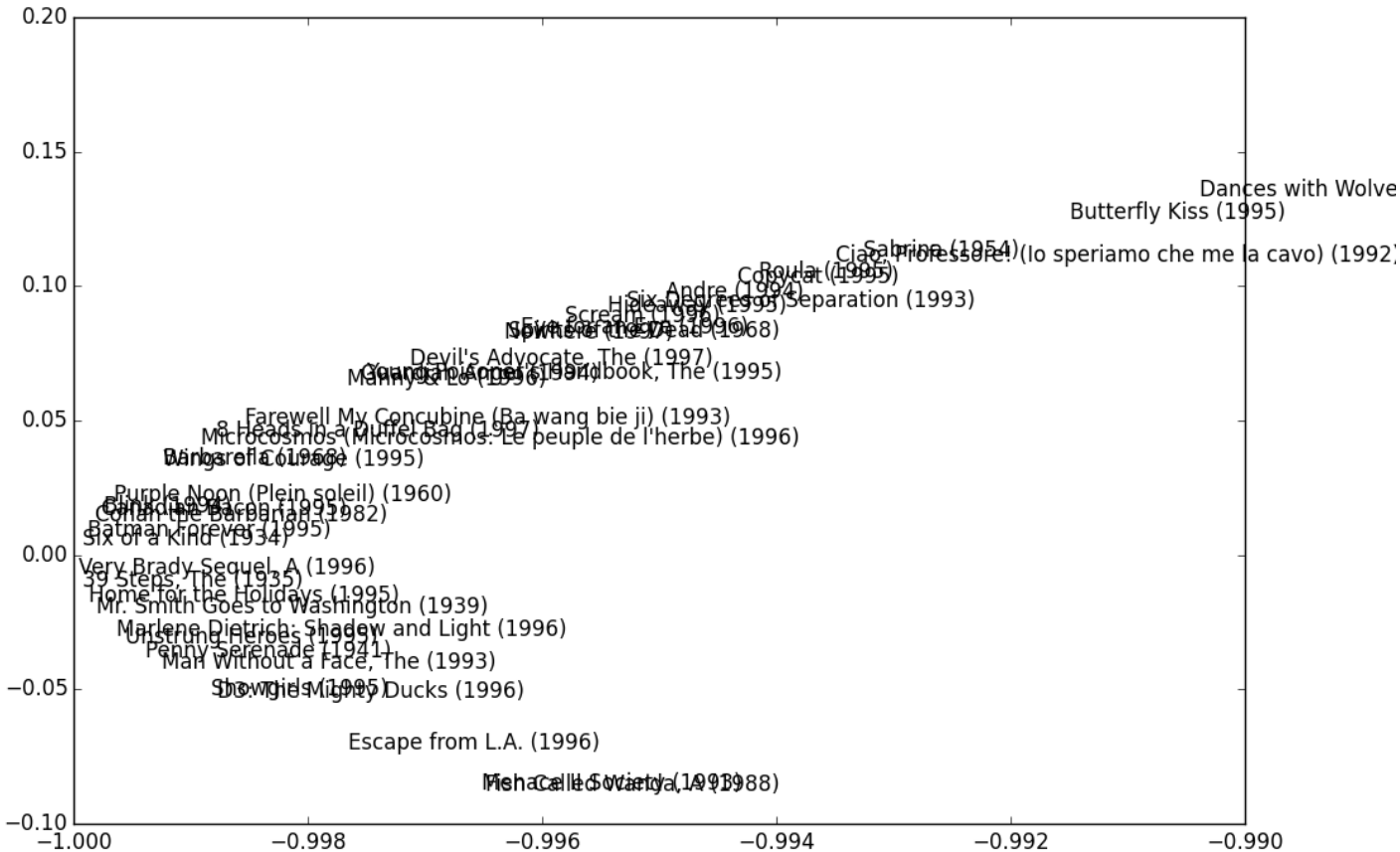
$$L = \{493,\ 526,\ 526,\ 7\ ...\}, \#L = 100K$$

Word2Vec

- $\mathbb{R}^d$ Embedding
- $$\arg\max_{\theta} \prod_{\text{movie} \in \text{Walk}} \prod_{\text{context} \in C(\text{movie})} p(\text{context}|\text{word}; \theta)$$
- $$p(c|w; \theta) = \frac{\exp(\langle v_c, v_w \rangle)}{\exp(\sum_{c'} \langle v_{c'}, v_w \rangle)}$$

Result

$$d = 2$$


0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−1.000
−0.998
−0.996
−0.994
−0.992
−0.990
Dances with Wolve
Butterfly Kiss (1995)
Separation (1993)
Devil's Advocate, The (1997)
Farewell My Concubine (Ba wang bie ji) (1993)
Purple Noon (Plein soleil) (1960)
Six of a Kind (1934)
Very Brady Sequel, A (1996)
39 Steps, The (1935)
Mr. Smith Goes to Washington (1939)
Man Without a Face, The (1993)
Escape from L.A. (1996)