Graph-based Recommender System

For Bipartite Network

XIN SUN

Heidelberg University
Institute of Mathematics & IWR
ethan.sun921107@gmail.com

June 26, 2020

Outline

1. Introduction
2. Recommender System
3. Baseline System
4. Experiment
5. Conclusion

Outline

1 Introduction

2 Recommender System

3 Baseline System

4 Experiment

5 Conclusion

Inspiration

- Inspiration
  - Recommender system - frequently used in social media field
  - Traditional approaches not suitable for exploding data
- Target
  - Reformulate structure data into graph or network
  - To predict the links between two nodes as Recommendation

Outline

1. Introduction
2. Recommender System
3. Baseline System
4. Experiment
5. Conclusion

Recommender System

Frequently used approaches in Recommender System:

- Collaborative filtering
- Content-based filtering
- Model-based filtering
- Hybrid recommender systems

Outline

1. Introduction
2. Recommender System
3. Baseline System
4. Experiment
5. Conclusion

Baseline System

- Content-based filtering approach
- User-based filtering approach

Baseline System

Content-based filtering approach

- The first neighbors of each movie node point out the users who rated / watched this movie and conversely, the movies that were rated by users are placed in the first neighbors of each user node

	userId	rec_movie1	rec_movie2	rec_movie3	rec_movie4	rec_movie5	rec_movie6
0	1	[Toy Story (1995)]	[Bug's Life, A (1998)]	[E.T. the Extra-Terrestrial (1982)]	[Toy Story (1995)]	[American Beauty (1999)]	[Shakespeare in Love (1998)]
1	20	[Star Wars: Episode IV - A New Hope (1977)]	[Star Wars: Episode V - The Empire Strikes Bac...	[Star Wars: Episode IV - A New Hope (1977)]	[Star Wars: Episode V - The Empire Strikes Bac...	[L.A. Confidential (1997)]	[2001: A Space Odyssey (1968)]
2	100	[Star Wars: Episode IV - A New Hope (1977)]	[Star Wars: Episode V - The Empire Strikes Bac...	[Star Wars: Episode IV - A New Hope (1977)]	[Star Wars: Episode V - The Empire Strikes Bac...	[American Beauty (1999)]	[Schindler's List (1993)]
3	500	[American Beauty (1999)]	[Shakespeare in Love (1998)]	[E.T. the Extra-Terrestrial (1982)]	[Toy Story (1995)]	[Toy Story (1995)]	[Bug's Life, A (1998)]
4	1000	[American Beauty (1999)]	[Schindler's List (1993)]	[Schindler's List (1993)]	[Star Wars: Episode V - The Empire Strikes Bac...	[Star Wars: Episode IV - A New Hope (1977)]	[Star Wars: Episode V - The Empire Strikes Bac...

Figure: Recommended movies by content-based filtering

Outline

1. Introduction
2. Recommender System
3. Baseline System
4. Experiment
5. Conclusion

Experiment

Inspiration for Graph approach:

- Traditional collaborating approaches not suitable for exploding data, especially sparse data
- More consider similarity relations between item-item or item-user

Experiment

Concrete experiment steps:

- Dataset Description
- Graph Construction
- Graph Analysis
- Graph Embeddings
- Neural Network for Link prediction
- Result Analysis

Experiment - Dataset Description

The dataset consists of 3,900 movies, 6,040 users who rated movies in 2000, 1,000,209 ratings from the respective users.

Dataset	Attributes
Movies	movieID, title, genres
Users	userID, movieID, rating, timestamp
Ratings	UserID, age, gender, occupation, zipcode

Figure: MovieLens dataset attributes

Experiment - Graph Construction

Reformulate raw structure data into a bipartite graph:

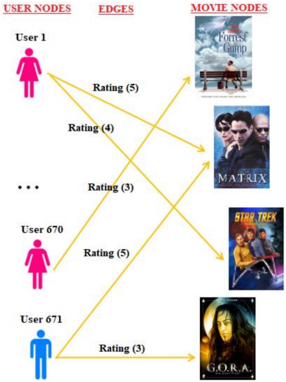

Figure: Bipartite Graph G is composed of a pair of sets (V, E)

Experiment - Graph Analysis

- Degree for all User and Movie nodes
- Degree centrality for all User and Movie nodes
- Clustering coefficients for all User and Movie nodes
- Preferential attachment for all User and Movie nodes
- Community (cluster) for all User and Movie nodes
- Most Favourite Movie Category for different User occupations
- . . . . . .

Experiment - Graph Embeddings

- Inspiration
  - Learning latent representations of vertices in a network considering similarity relations
- Approaches
  - Random Walk
  - Language model: Skip-gram
  - DeepWalk

Experiment - Graph Embeddings

Random Walk

- Assumption: Adjacent nodes are similar and should have similar embeddings
- Goal: To discover neighborhoods in the network and extract sequences from a graph

Experiment - Graph Embeddings

Random Walk Example

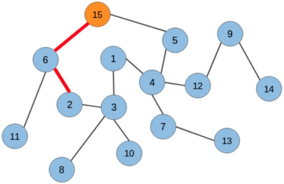

Figure: Random Walk start Node 15 with two edges

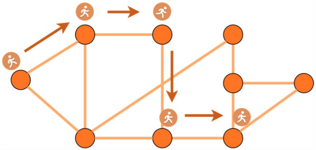

Figure: Random walk with length 5

Experiment - Graph Embeddings

Skip-gram Assumption

- Words that occur in the same context tend to have close meaning - their embeddings should be close to each other

Experiment - Graph Embeddings

Skip-gram Approach

- Given a specific word in the middle of a sentence (the input word), look at the words nearby and pick one at random. The network is going to tell us the probability for every word in our vocabulary of being the "nearby word" that we chose

Experiment - Graph Embeddings

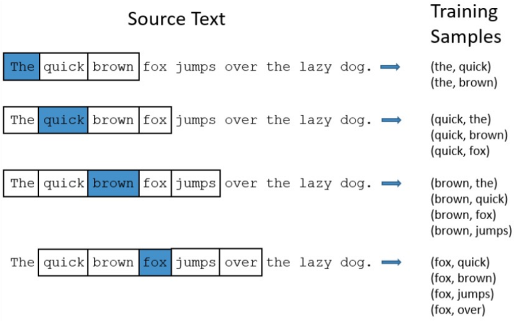

Figure: Training example of Skip-gram model

Experiment - Graph Embeddings

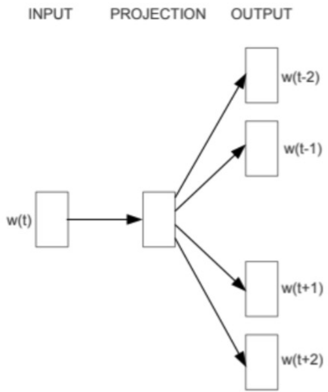

Figure: Architecture of Skip-gram model

Experiment - Graph Embeddings

Adopt Skip-gram to networks

- Node sequences in networks as word sequences in text
- Maximize similarity of embeddings of nodes that occur on same walks

Experiment - Graph Embeddings - DeepWalk

Algorithm 1 DEEPWALK(G, w, d, γ, t)

Input: graph $G(V, E)$
  window size w
  embedding size d
  walks per vertex γ
  walk length t
Output: matrix of vertex representations $\Phi \in \mathbb{R}^{|V|\times d}$

Initialization: Sample Φ from $\mathcal{U}^{|V|\times d}$
Build a binary Tree T from V
for $i = 0$ to γ **do**
  $\mathcal{O} = \text{Shuffle}(V)$
  for each $v_i \in \mathcal{O}$ **do**
    $\mathcal{W}_{v_i} = RandomWalk(G, v_i, \text{t})$
    SkipGram(Φ, $\mathcal{W}_{v_i}$, w)
  end for
end for

Figure: Concrete Steps of DeepWalk Algorithm

Experiment - Graph Embeddings

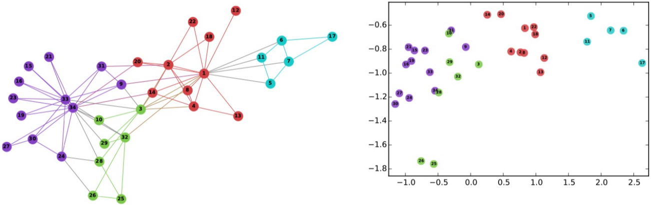

Figure: DeepWalk Embedding result Visualisation

Experiment - Neural Network Construction

- MLP target
  - Reformulate Movie Recommendation as Link prediction task
- Inputs
  - Embeddings of all users and movies nodes by DeepWalk
  - Graph data attributes from graph analysis result as features
- Output
  - Whether a link existing between a user node and a movie node $==$ Whether a user will give high rating score or like the given specific movie

Experiment - Result Analysis

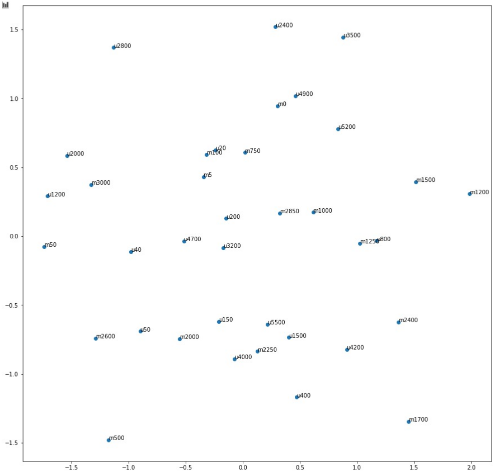

Figure: User and Movie nodes embeddings by DeepWalk

Experiment - Result Analysis

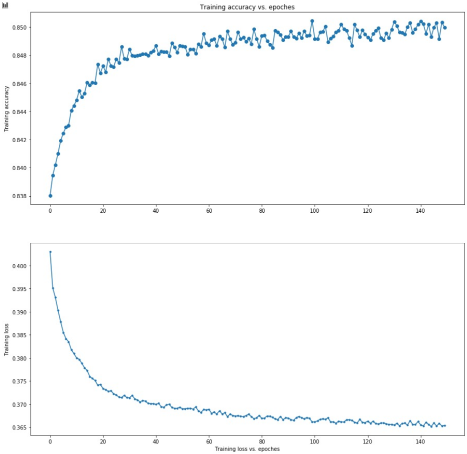

Figure: Loss and Accuracy plot

Outline

1 Introduction

2 Recommender System

3 Baseline System

4 Experiment

5 Conclusion

Conclusion

- Thanks to DeepWalk, we can reflect neighborhood relations and local structures in the network to node representations.
- MLP classifier is beneficial from more graph attributes as additional features achieving from graph analysis.

Questions

Questions

Reference

- [1] Bryan Perozzi et al. (2014) DeepWalk: Online Learning of Social Representations.
- [2] Göksu Tüysüzoğlu and Zerrin Işık. (2018) A Hybrid Movie Recommendation System UsingGraph-Based Approach.
- [3] Rıza Özçelik (2019) An Intuitive Explanation of DeepWalk.