Cold Start Similar Artists Ranking with Gravity-Inspired Graph Autoencoders

MITON Times
12. 10. 2021
Radek Bartyzal

Similar Items Ranking as a Directed Link Prediction Task

- graph nodes = items
  - have feature vector
  - have node embedding
- edges = relation = fans of this item also like this item = weighted
- edges are directed = random reggae band -> Bob Marley but not the other way around

- cold start item = new node not connected to other ones
- clothes are bi-directional I think

Graph Auto Encoders

- train on reconstruction loss
- 2 layer:

$$\tilde{Z} = \tilde{A}\mathrm{ReLU}(\tilde{A}XW^{(0)})W^{(1)}.$$

- A = n x n adjancency matrix, A_tilde = outdegree normalized A
- X = n x f = content feature embeddings of n items
- Z = n x d = n embeddings of n items = latent representation
- A can be reconstructed from Z
- at each layer, each node averages the representations from its neighbors (and itself), with a ReLU activation

Graph Variational Auto Encoders

- just have 2 networks predicting μ and σ matrices =
  - μ_i and σ_i for each node
- node_i is sampled from N(μ_i, σ_i)

Gravity inspired GAE

- each node has mass
- nodes accelerate faster toward more massive nodes = directed edge weights
- Z = n x (d+1)
  - +1 = mass value
- decoding A from Z according to Newton formulas:
- G = accel. constant
- m_i = mass of node i

$$\begin{aligned}\hat{A}_{ij} &= \sigma(\log a_{i\rightarrow j}) = \sigma(\log \frac{Gm_j}{\|z_i - z_j\|_2^2}) \\ &= \sigma(\underbrace{\log Gm_j}_{\text{denoted } \tilde{m}_j} - \log \|z_i - z_j\|_2^2).\end{aligned}$$

The elephant in the room = O(dn^2)

- optimization is O(dn^2) = whole A times d dim. embeddings
- yes they approximate the whole NxN matrix for the experiments
- BUT you can approximate the losses by decoding random sub-graphs = O(n)
- = FastGAE

Cold start

- m new items with content features of dim f:

- A = (n+m) x (n+m) with new rows, cols full of zeroes
- X = (n+m) x f with new rows = content features
- do a forward pass calculating latent representations of new nodes => Z
- reconstruct A from new Z = predicting new edges

- top-k most similar items of i will correspond to the k nodes j with highest estimated weights A_ij

Usage at Deezer

- 70M items
- 16M users
- tested on a dataset with 24K items
- K = 20 = each item points to top 20 other items
- embeddings recomputed **weekly …**
- item content features:
  - 32 dim **genre** vector, 300 genres, calculated by factorizing a co-occurrence matrix based on listening usages with SVD
  - 20 dim **one-hot country** vector
  - 4 dim **mood** vector = average and standard deviations of the *valence* and *arousal* scores
    - calculated by DNN from audio data of the song
  - actual app uses more content features extracted from audio/text for each item

Table 1: Cold start similar artists ranking: performances on test set.

Methods ($d = 32$)	Recall@K (in %) $K = 20$	Recall@K (in %) $K = 100$	Recall@K (in %) $K = 200$	MAP@K (in %) $K = 20$	MAP@K (in %) $K = 100$	MAP@K (in %) $K = 200$	NDCG@K (in %) $K = 20$	NDCG@K (in %) $K = 100$	NDCG@K (in %) $K = 200$
Popularity	0.02	0.44	1.38	<0.01	0.03	0.12	0.01	0.17	0.44
Popularity by country	2.76	12.38	18.98	0.80	3.58	6.14	2.14	6.41	8.76
In-degree	0.91	3.43	6.85	0.15	0.39	0.86	0.67	1.69	2.80
In-degree by country	5.46	16.82	23.52	2.09	5.43	7.73	5.00	10.19	12.64
K-NN on x_i	4.41	13.54	19.80	1.14	3.38	5.39	4.29	8.83	11.22
K-NN + Popularity	5.73	15.87	19.83	1.66	4.32	5.74	4.86	10.03	11.76
K-NN + In-degree	7.49	17.29	18.76	2.78	5.60	6.18	7.41	12.48	13.14
SVD + DNN	6.42 ± 0.96	21.83 ± 1.21	35.01 ± 1.41	2.25 ± 0.67	6.36 ± 1.19	11.52 ± 1.98	6.05 ± 0.75	12.91 ± 0.92	17.89 ± 0.95
STAR-GCN	10.03 ± 0.56	31.45 ± 1.09	43.92 ± 1.10	3.10 ± 0.32	10.64 ± 0.54	16.62 ± 0.68	10.07 ± 0.40	21.17 ± 0.69	25.99 ± 0.75
DropoutNet	12.96 ± 0.54	37.59 ± 0.76	49.93 ± 0.82	4.18 ± 0.30	13.61 ± 0.55	20.12 ± 0.67	13.12 ± 0.68	25.61 ± 0.72	30.52 ± 0.78
DEAL	12.80 ± 0.52	37.98 ± 0.59	50.75 ± 0.72	4.15 ± 0.25	14.01 ± 0.44	20.92 ± 0.54	12.78 ± 0.53	25.70 ± 0.62	30.69 ± 0.70
Graph AE	7.30 ± 0.51	25.92 ± 0.95	40.37 ± 1.11	2.81 ± 0.29	7.97 ± 0.47	14.24 ± 0.67	6.32 ± 0.39	15.54 ± 0.66	20.94 ± 0.72
Graph VAE	10.01 ± 0.52	34.00 ± 1.06	49.72 ± 1.14	3.53 ± 0.27	11.68 ± 0.52	19.46 ± 0.70	10.09 ± 0.58	21.37 ± 0.73	27.31 ± 0.75
Sour.-Targ. Graph AE	12.21 ± 1.30	39.52 ± 3.53	56.25 ± 3.57	4.62 ± 0.81	14.67 ± 2.33	23.60 ± 2.85	12.42 ± 1.39	25.45 ± 3.37	31.80 ± 3.38
Sour.-Targ. Graph VAE	13.52 ± 0.64	42.68 ± 0.69	59.51 ± 0.76	5.19 ± 0.31	16.07 ± 0.40	25.48 ± 0.55	13.60 ± 0.73	27.81 ± 0.56	34.19 ± 0.59
Gravity Graph AE	**18.33 ± 0.45**	**52.26 ± 0.90**	**67.85 ± 0.98**	**6.64 ± 0.25**	**21.19 ± 0.55**	**30.67 ± 0.68**	**18.64 ± 0.47**	**35.77 ± 0.66**	**41.42 ± 0.68**
Gravity Graph VAE	16.59 ± 0.50	49.51 ± 0.78	65.70 ± 0.75	5.66 ± 0.35	19.07 ± 0.57	28.66 ± 0.59	16.74 ± 0.55	33.34 ± 0.66	39.29 ± 0.64

Graph Autoencoders for Directed Link Prediction

- Gravity GAE Code: https://github.com/deezer/gravity_graph_autoencoders
- Gravity GAE Paper: https://arxiv.org/abs/1905.09570

- this code: https://github.com/deezer/similar_artists_ranking
- this paper: https://dl.acm.org/doi/pdf/10.1145/3460231.3474252