Graph Convolution Matrix Completion for Yelp Rating Prediction

Ming-Chun Wu, Haonan Li, Runsheng Liu
Department of Statistics

Introduction

- We aim to built a large-scale recommendation system using graph structured data
- We analysis the Yelp dataset and build predictive models for review ratings using recently developed Graph Convolution Matrix Completion (GCMC)
- The project provide an end-to-end pipeline: from data cleaning, pre-processing, model implementations to training and testing, from scratch in Tensorflow and Spark

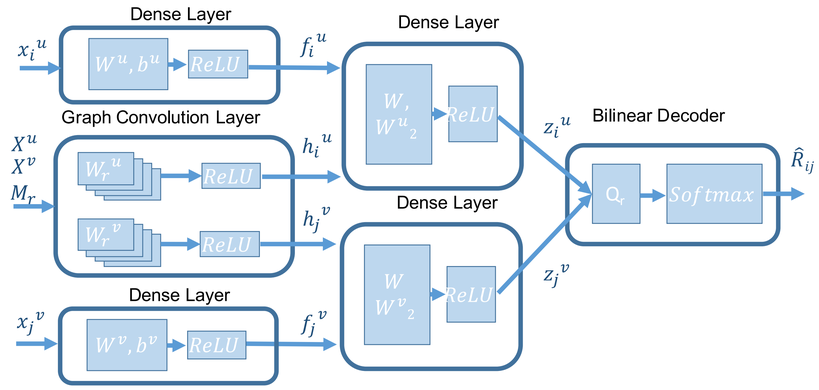
The Yelp Dataset

- Yelp is a popular social network where users can leave reviews of businesses
- The dataset contains:
  - 1637138 users and 192609 business items with 6685900 ratings
  - 1637138×192609 rating matrix with 6685900 known entries in values of $\{1, 2, 3, 4, 5\}$
  - There are 175 raw item features like `Music.dj`, `Alcohol` and 83 raw user features like `review count`

Challenges

- Very sparse data set: 0.006% ratings; Traditional examples like MovieLens are around 1.34%
- Need to use as much useful information as possible to do prediction: how to utilize graph structured information?
- Building an end-to-end data pipeline to train neural networks is challenging!

Graph Convolution Matrix Completion

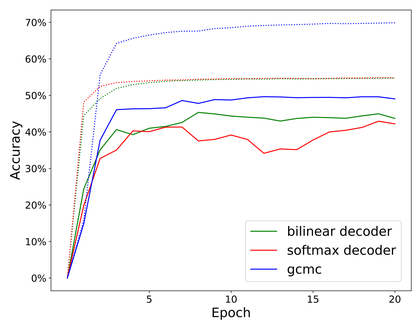


- Given a user and its neighbors N, the Graph Convolution Layer is: $h = \sigma(\frac{1}{N}\sum_{j\in\mathcal{N}} W^{(1)}_{graph} x_j)$
- Graph embedding z: $z = \sigma(W_1^{(2)}h + W_2^{(2)}f), f = \sigma(W^{(1)}_{dense}x + b)$
- Bilinear decoder: given trainable coefficients Q_r, the rating of user i on item j is predicted by

$$p(\hat{R}_{ij} = r) = \frac{\exp[(z_i^u)^T Q_r z_j^v]}{\sum_{r=1}^{K} \exp[(z_i^u)^T Q_r z_j^v]} \quad (1)$$

Training/Validation

- GCMC (blue line) outperforms models without graph convolution layers in both training and validation



Test Set

- Benchmarks: Spark matrix factorization, neural networks without graph convolution layers
- GCMC beats all benchmarks in popular performance measures: accuracy and MSE.
- GCMC is time consuming: due to graph structured training batches instead of computing gradient and back propagation

Method	Accuracy	MSE	Time
GCMC	49.625	2.702	25h
bilinear	44.255	3.019	2.5h
softmax	43.004	3.318	2.4h
Spark MF	27.515	2.703	41s

Implementations

- Train/Validation/Test: $85\%/5\%/10\%$
- Batch-normalization, l_2 regularization with scale 0.001, dropout with probability 0.7
- Cross entropy loss function; Adam optimizer with learning rate 0.01
- batch size 20000 and 20 epochs
- Relu activation; Hidden units: $256(128)$ for feature dense layers, graph convolution layers and embedding layers of item(user) related features.

Conclusions

- We provide a pioneer study of GCMC on the very challenging Yelp dataset:
  - The original studies [1], [2], are focus on small and less challenging datasets (at most 10M ratings).
  - The Yelp dataset is much more sparse (0.006% ratings) compared to MovieLens used in the original studies (1.34%)
- Our study shows the effectiveness of using graph structured data:
  - We directly compare GCMC to controlled groups, benchmarks without accessing to graph information. The result suggests the importance of graph structured data in prediction.
- Future Works:
  - Add more useful features like review texts into the model
  - As suggested by [3], we can use the GANs training to boost the performance of GCMC

References

[1] Thomas N. Kipf, Max Welling. *Semi-Supervised Classification with Graph Convolutional Networks*. In ICLR, 2017.
[2] Rianne van den Berg, Thomas N. Kipf, Max Welling. *Graph Convolutional Matrix Completion*. KDD́18 Deep Learning Day.
[3]graphgan Wang Hongwei, Wang Jia, Wang Jialin, Zhao Miao, Zhang Weinan, Zhang Fuzheng, Xie Xing, and Guo Minyi. *Graphgan: Graph representation learning with generative adversarial nets*. In AAAI, 2018.

Contact

Ming-Chun Wu, Haonan Li, Runsheng Liu
{ chunwu, lihaonan, rsliu }@uw.edu
Department of Statistics
University of Washington