Deep Learning based Recommender System

Chen Si

XIAMEN UNIVERSITY

sichen@stu.xmu.edu.cn

November 7, 2018

Overview

1. DEEP LEARNING BASED RECOMMENDATION
   - MLP based Recommendation
   - Autoencoder based Recommendation
   - CNN based Recommendation
   - RNN based Recommendation
   - RBM based Recommendation
   - Neural Attention based Recommendation
   - Neural AutoRegressive based Recommendation
   - Deep Reinforcement Learning for Recommendation
   - GAN based Recommendation
2. Summary

Outline

1. DEEP LEARNING BASED RECOMMENDATION
   - MLP based Recommendation
   - Autoencoder based Recommendation
   - CNN based Recommendation
   - RNN based Recommendation
   - RBM based Recommendation
   - Neural Attention based Recommendation
   - Neural AutoRegressive based Recommendation
   - Deep Reinforcement Learning for Recommendation
   - GAN based Recommendation

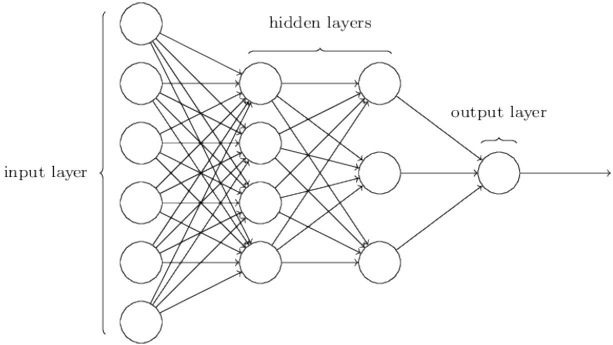
2. Summary

Multilayer Perceptron based Recommendation

1. What is Multilayer Perceptron(MLP)
2. Neural Extension of Traditional Recommendation Methods.
3. Feature Representation Learning with MLP

What is Multilayer Perceptron(MLP)

- MLP is a feed-forward neural network with multiple hidden layers

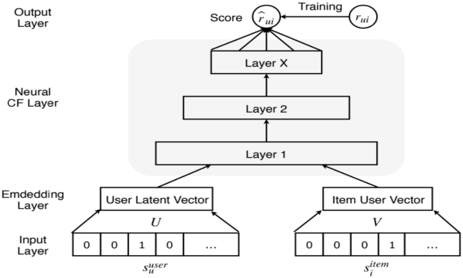


,

$$y_l = \phi(W_2 * \phi(W_1 * x))$$

Neural Extension of Traditional Methods

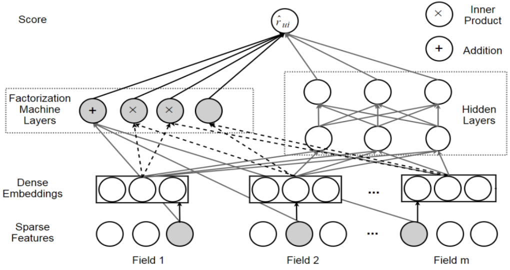
- Neural Collaborative Filtering (NCF)



$$\hat{r}_{ui} = f(U^T \cdot s_u^{user}, V^T \cdot s_i^{item} | U, V, \theta)$$

Feature Representation

- Deep Factorization Machine(DeepFM)
  - FM: linear and pairwise low-order interactions between features.
  - MLP leverages the non-linear activations and deep structure to model the high-order interactions.



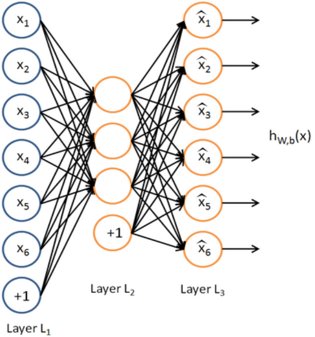
$$\hat{r}_{ui} = \sigma(y_{FM}(x) + y_{MLP}(x) + bias)$$

Autoencoder based Recommendation

- What is Autoencoder
- Autoencoder based Collaborative Filtering.
- Feature Representation Learning with Autoencoder

What is Autoencoder

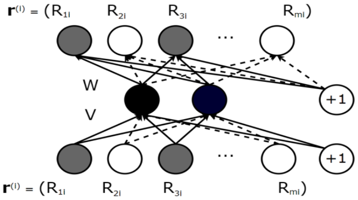
- An autoencoder neural network is an unsupervised learning algorithm, tring to learn a function $h_{W,b}(x) \approx x$.



  - filling the blanks of the interaction matrix directly in the reconstruction layer.
  - using autoencoder to learn lower-dimensional feature representations at the bottleneck layer;

Autoencoder based Collaborative Filtering.

- AutoRec
  - takes user partial vectors $r^{(u)}$ or item partial vectors $r^{(i)}$ as input, and aims to reconstruct them in the output layer.
  - Two variants: item-based AutoRec (I-AutoRec) and user-based AutoRec

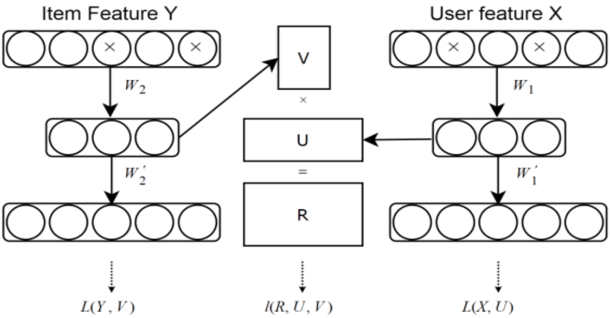


$$argmin_\theta = \sum_{i=1}^{N} \left\| r^i - h(r^i;\theta) \right\|_O^2 + \lambda \times reg$$

here $\|\cdot\|_O^2$ means that it only considers observed ratings

Feature Representation Learning with Autoencoder

- general framework to build hybrid collaborative models.



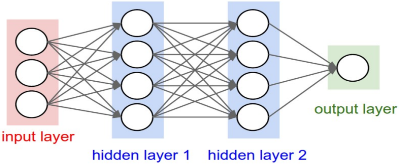
$$arg_{U,V} min\ l(R,U,V) + \beta(U^2 + V^2) + \gamma L(X,U) + \delta L(Y,V)$$

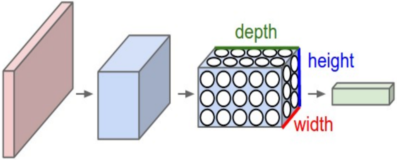
Convolutional Neural Networks based Recommendation

- What is Convolutional Neural Networks(CNN)
- Feature Representation Learning with CNNs.
- CNNs based Collaborative filtering.
- Graph CNNs for Recommendation.

What is Convolutional Neural Networks(CNN)

- a special kind of feed-forward neural network with convolution layers and pooling operations.





  - 默认输入是图像，把特定的性质编码入网络结构，使前馈函数更加有效率，并减少了大量参数。
  - CNNs are powerful in processing unstructured multimedia data with convolution and pool operations.

Feature Representation Learning with CNNs

- Most of the CNNs based recommendation models utilize CNNs for feature extraction.
  - Image Feature Extraction.
  - Text Feature Extraction
    - model user behaviors and item properties from review texts
    - alleviates the sparsity problem and enhances the model interpretability
  - Audio and Video Feature Extraction.
    - content based model can alleviate the cold start problem (music has not been consumed)

CNNs based Collaborative filtering

- ConvNCF
  - use outer product instead of dot product to model the user item interaction patterns.
  - CNNs are applied over the result of outer product and could capture the high-order correlations among embeddings dimensions.

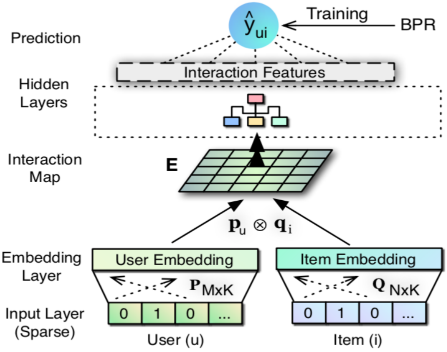


Figure 1: Outer Product-based NCF framework

CNNs based Collaborative filtering

- ConvNCF
  - 外积：交互矩阵融合了每个维度下特征之间的关系（传统 CF ：主对角线求和），能刻画特征维度之间的高阶关系。
  - 卷积：后一层的每一个元素都是由前一层的 4 个元素计算得来的，可以认为是一个 4 阶关系的刻画。直到最后的输出层，降到 1×1 后，即包含了特征每一个维度之间的交互信息。CNN 比 MLP 更容易泛化和建立更深的网络。

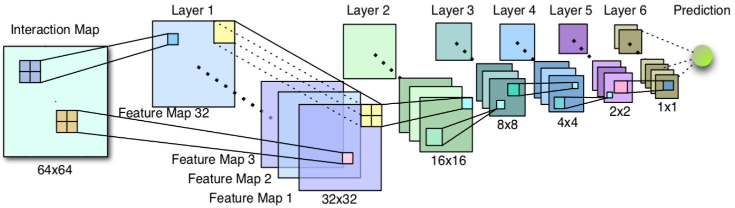


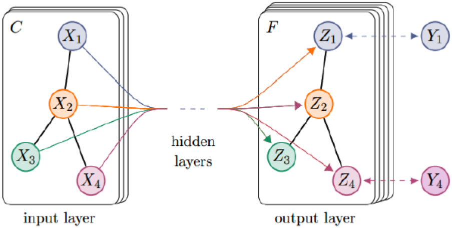
Figure 2: An example of the architecture of our ConvNCF model that has 6 convolution layers with embedding size 64.

Graph CNNs for Recommendation.

- Interactions in recommendation area can also be viewed as bipartite graph,thus the recommendation problem can be considered as a link prediction task with graph CNNs.

- GCN: Graph Convolution Network
  - Set $K = 2$ and view the center node as one of its neighbor, resulting in only one free parameter for each convolution filter
  - Offer an explanation of feature diffusion over graph

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \qquad H^{(0)} = X$$

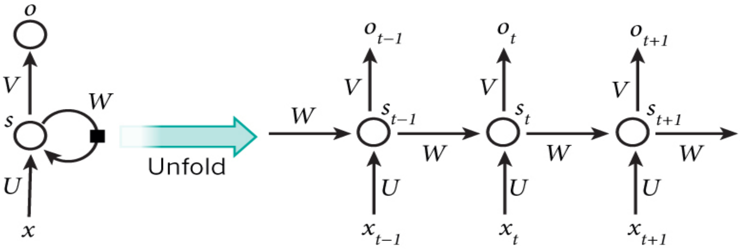


Recurrent Neural Networks based Recommendation

- What is Recurrent Neural Networks(RNN)
- Session-based Recommendation without User Identifier
- Sequential Recommendation with User Identifier
- Feature Representation Learning with RNNs

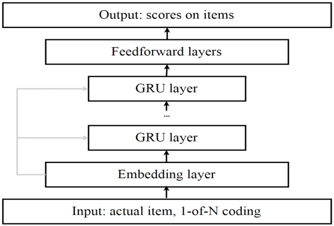
What is Recurrent Neural Networks(RNN)

- RNN is suitable for modelling sequential data. Unlike feedforward neural network, there are loops and memories in RNN to remember former computations.
- RNN 的结构不同于 MLP，输入层与来自序列中上一元素隐层的信号共同作用到当前的隐藏层



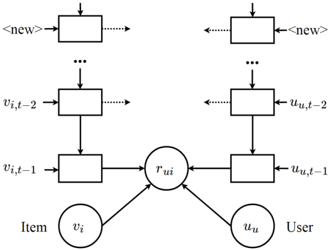
Session-based Recommendation without User Identifier

- GRU4Rec
  1. 输入:用户的行为序列: $[x_1, x_2, x_3, .., x_N]$(1-of-N encoding,或者再过一个Embedding层)
  2. 过若干层的GRU(核心的序列化建模)
  3. Feedforward网络转换
  4. 对下一个目标x_{N+1}进行预测



Sequential Recommendation with User Identifier

- Recurrent Recommender Network (RRN)
  - modelling the seasonal evolution of items and changes of user preferences over time
  - uses two LSTM networks to model dynamic user state u_{ut} and item state v_{it} .
  - incorporates stationary latent attributes such as user long-term interests and item static features: u_u and v_i.



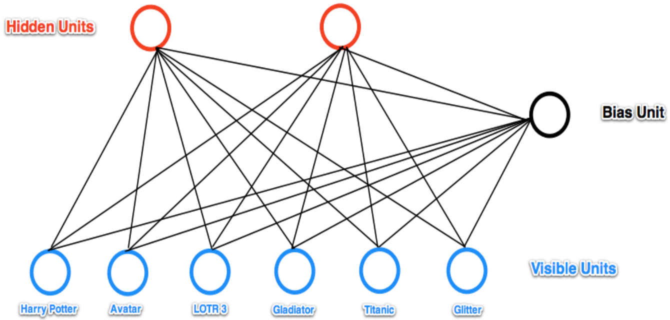
$$\hat{r}_{ui|t} = f(u_{ut}, v_{it}, u_u, v_i)$$

Feature Representation Learning with RNNs

- For side information with sequential patterns
  - learn representations of evolution and co-evolution of user and item features.
  - encode the text sequences into latent factor model.
  - learn more expressive aggregation for user browsing history
  - predicting ratings as well as generating textual tips for users simultaneously
  - two sub-networks to modelling user static features (with MLP) and user temporal features (with RNNs).

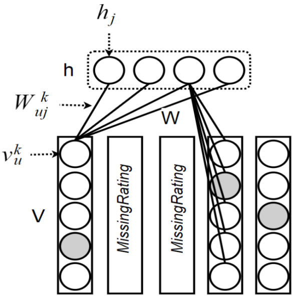
- What is Restricted Boltzmann Machine(RBM)
- Restricted Boltzmann Machine based Recommendation

- RBM is a two layer neural network consisting of a visible layer and a hidden layer.
  - 限制在同一层的神经元之间不会相互连接，而不在同一层的神经元之间会相互连接，连接是双向的以及对称的。这意味着在网络进行训练以及使用时信息会在两个方向上流动，而且两个方向上的权值是相同的。
  - 可见变量和隐藏变量都是二元变量，亦即其状态取0,1



Restricted Boltzmann Machine based Recommendation

- RBM-CF
  - the first recommendation model that built on neural networks.
  - user/item-based RBM-CF :given user’ s/item’s rating is clamped on the visible layer.
  - 假设有m个电影, 则使用m个softmax单元来作为可见单元来构造RBM.如果一个用户没有对第j个电影评分, 则该用户的RBM中不存在第j个softmax单元.



Algorithm 4.4 RBM - Making recommendations

Inputs: a user u, an movie i

Outputs: an estimation of $R(u, i)$

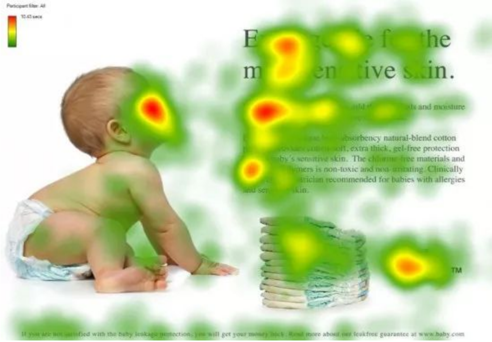
1. Clamp the ratings of u over the softmax units of the RBM.
2. Compute $\boxed{\hat{p}_j = p(h_j = 1|V)}$ for all hidden units j.
3. Compute $\boxed{p(v_i^k = 1|\hat{p}) = \frac{\exp(b_i^k + \sum_{j=1}^{F} \hat{p}_j W_{ij}^k)}{\sum_{l=1}^{K} \exp(b_i^l + \sum_{j=1}^{F} \hat{p}_j W_{ij}^l)} \text{ for } k = 1, ..., K.}$
4. Take the expectation as the prediction, i.e., $R(u, i) = \sum_{k=1}^{K} p(v_i^k = 1|\hat{p})k$.

Neural Attention based Recommendation

- What is Neural Attention.
- Recommendation with Vanilla Attention
- Recommendation with Co-Attention

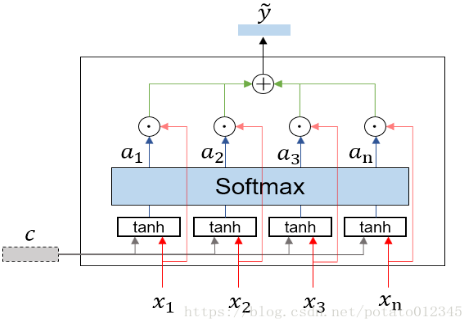
What is Neural Attention

- Attention mechanism is motivated by human visual attention.
- 核心目标:从众多信息中选择出对当前任务目标更关键的信息。

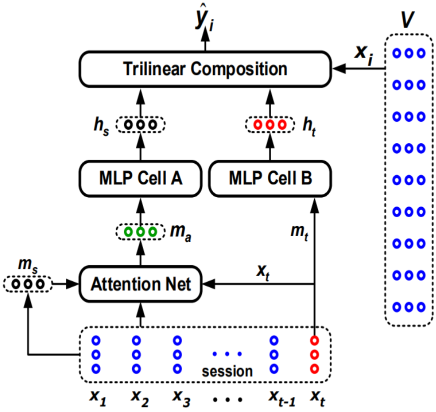


What is Neural Attention

- attention model learns to attend to the input with attention scores (the heart of neural attention models).

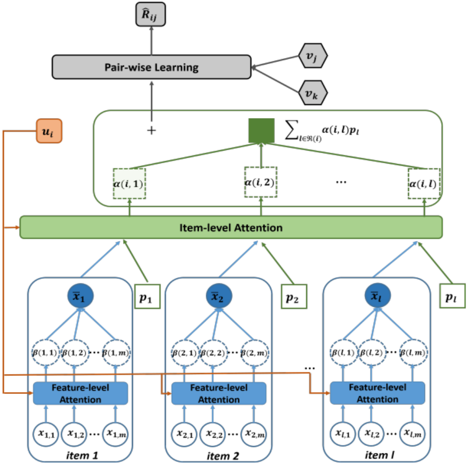


- 基于记忆优先级的短序列推荐
  - 对session内前n-1个全局商品用attention建模得到一个全局表达，并输入MLP



Recommendation with Vanilla Attention

- Attentive Collaborative Filtering(ACF)
  - select items from implicit that are representative to user preferences and then aggregate them to characterize users.



- item-level attention

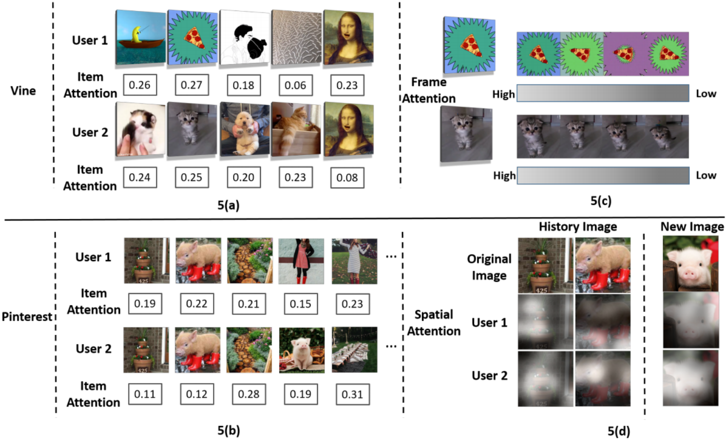
$$u_i + \sum_{l \in R(i)} \alpha(i,l) p_l$$

- component-level attention

$$\bar{x}_l = \sum_{m=1}^{|\{x_{l*}\}|} \beta(i,l,m) \cdot x_{lm}$$

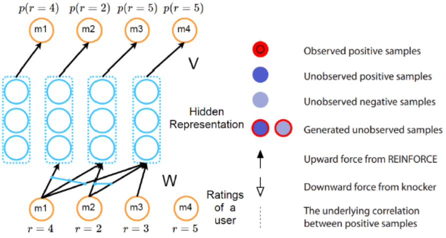
Recommendation with Vanilla Attention

- Attentive Collaborative Filtering(ACF)
  - Attention Visualization



Neural AutoRegressive based Recommendation

- tractable的分布估计器，它是RBM的理想的替代品
- NADE based collaborative filtering model (CF-NADE)
  - models the distribution of user ratings.
  - 有4部电影：m1（评分为4），m2（评分为2），m3（评分为3），m4（评分为5）。 CF-NADE利用链式法则得到的评分向量r的联合概率



(a)

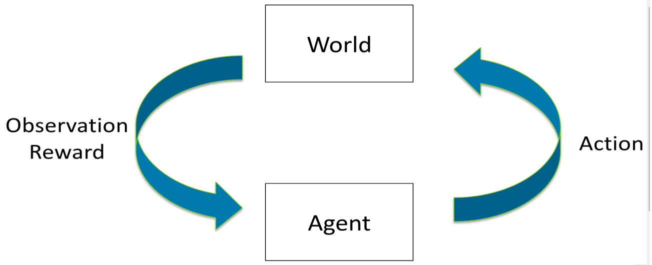
$$p(r) = \Sigma_{i=1}^{4} p(r_{m_i} | r_{m_{<i}})$$

Deep Reinforcement Learning for Recommendation

- What is Deep Reinforcement Learning (DRL)
- Deep Reinforcement Learning for Recommendation

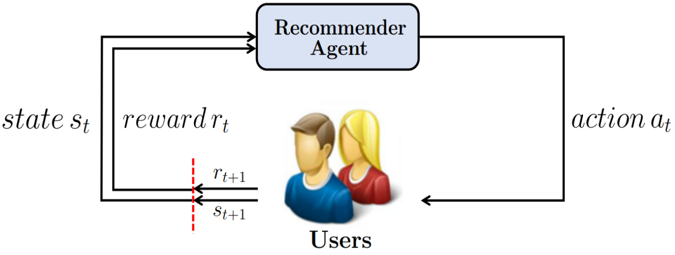
What is Deep Reinforcement Learning (DRL)

- Reinforcement Learning:Learn to make good sequences of decisions
  - Repeated Interactions with World
  - Reward for Sequence of Decisions
  - Repeated Interactions with World
- trial-and-error paradigm
- components: agents, environments, states, actions and rewards.



Deep Reinforcement Learning for Recommendation

- recommender agent (RA) interacts with environment E (or users) by sequentially choosing recommendation items over a sequence of time steps, so as to maximize its cumulative reward.



Adversarial Network based Recommendation

- What is Adversarial Network (AN)
- Adversarial Network based Recommendation

What is Adversarial Network (AN)

- Adversarial Networks (AN) is a generative neural network which consists of a discriminator and a generator.
- They trained simultaneously by competing with each other in a min-max game framework.

Adversarial Network based Recommendation

- IRGAN
  - 在信息检索上有两个思维方式，即生成式检索和判别式检索:
  - Generative retrieval assumes that there is an underlying generative process between documents and queries, and retrieval tasks can be achieved by generating relevant document given a query.
  - Discriminative retrieval learns to predict the relevance score given labelled relevant query-document pairs.
  - minimax game : generative retrieval aims to generate relevant documents similar to ground truth to fool the discriminative retrieval model.

Outline

1 DEEP LEARNING BASED RECOMMENDATION

2 Summary

Why Deep Neural Networks for Recommendation

- Nonlinear Transformation
  - 捕获非线性和非平凡的用户-物品关系
- Representation Learning
  - 捕获数据本身的复杂联系(上下文，文本和视觉信息)
- Sequence Modelling
- Flexibility

The End