Restricted Boltzmann Machine based Recommender System

Deep Learning based Recommender System: A Survey and New Perspectives

Zhang *et al.*2017

Kuhwan Jeong [1]

[1]Department of Statistics, Seoul National University

January, 2018

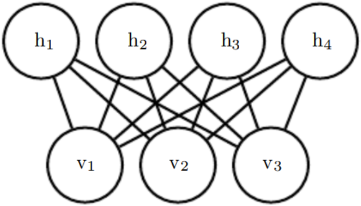
Related Papers

1. Salakhutdinov, Ruslan, Andriy Mnih, and Geoffrey Hinton. "Restricted Boltzmann machines for collaborative filtering." *Proceedings of the 24th international conference on Machine learning*. ACM, 2007.

2. Georgiev, Kostadin, and Preslav Nakov. "A non-iid framework for collaborative filtering with restricted boltzmann machines." *International Conference on Machine Learning*. 2013.

3. Liu, Xiaomeng, et al. "Item Category Aware Conditional Restricted Boltzmann Machine Based Recommendation." *International Conference on Neural Information Processing*. Springer, Cham, 2015.

Restricted Boltzmann Machines



- The restricted Boltzmann machine (RBM) is an undirected graphical model based on a bipartite graph, with visible units in one part of the graph and hidden units in the other part.
- The joint probability distribution is

$$P(\mathbf{v} = v, \mathbf{h} = h) = \frac{\exp\left(b^T v + c^T h + v^T W h\right)}{Z}$$

  where $Z = \sum_v \sum_h \exp\left(b^T v + c^T h + v^T W h\right)$.
- The marginal probability distribution is

$$P(\mathbf{v} = v) = \frac{\sum_h \exp\left(b^T v + c^T h + v^T W h\right)}{Z}.$$

Restricted Boltzmann Machines

- Let $\theta = (W, b, c)$.

- The gradient of the marginal log-likelihood with respect to θ is

$$\begin{aligned} &\nabla_\theta \log p(v;\theta) \\ = \; &\mathbb{E}_{h\sim p(\mathbf{h}|\mathbf{v}=v)} \nabla_\theta (b^T v + c^T h + v^T W h) - \mathbb{E}_{(v',h')\sim p(\mathbf{v},\mathbf{h})} \nabla_\theta (b^T v' + c^T h' + {v'}^T W h'). \end{aligned}$$

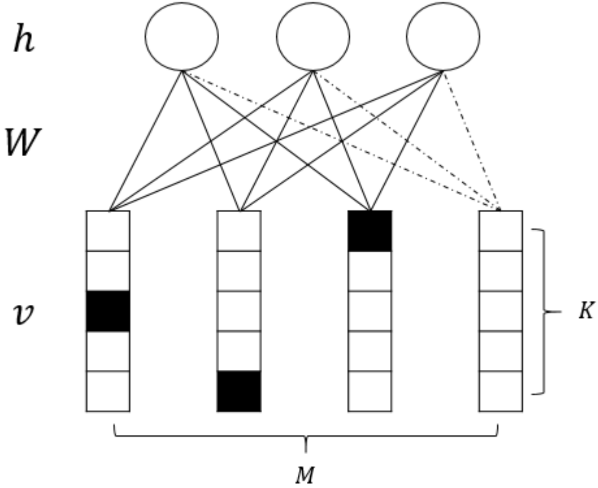
Restricted Boltzmann Machines

Algorithm 18.2 The contrastive divergence algorithm, using gradient ascent as the optimization procedure

Set ϵ, the step size, to a small positive number.
Set k, the number of Gibbs steps, high enough to allow a Markov chain sampling from $p(\mathbf{x};\boldsymbol{\theta})$ to mix when initialized from p_{data}. Perhaps 1–20 to train an RBM on a small image patch.
while not converged **do**
  Sample a minibatch of m examples $\{\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(m)}\}$ from the training set
  $\mathbf{g} \leftarrow \frac{1}{m}\sum_{i=1}^{m}\nabla_{\boldsymbol{\theta}}\log\tilde{p}(\mathbf{x}^{(i)};\boldsymbol{\theta})$.
  for $i = 1$ to m **do**
    $\tilde{\mathbf{x}}^{(i)} \leftarrow \mathbf{x}^{(i)}$.
  end for
  for $i = 1$ to k **do**
    for $j = 1$ to m **do**
      $\tilde{\mathbf{x}}^{(j)} \leftarrow \text{gibbs\_update}(\tilde{\mathbf{x}}^{(j)})$.
    end for
  end for
  $\mathbf{g} \leftarrow \mathbf{g} - \frac{1}{m}\sum_{i=1}^{m}\nabla_{\boldsymbol{\theta}}\log\tilde{p}(\tilde{\mathbf{x}}^{(i)};\boldsymbol{\theta})$.
  $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \epsilon\mathbf{g}$.
end while

RBM-CF

- Suppose we have M movies, N users, and integer rating values from 1 to K.



RBM-CF

- The joint probability distribution is

$$P(\mathbf{v} = v, \mathbf{h} = h) = \frac{1}{Z} \exp\left(\sum_{i \in \mathcal{O}} \sum_{k=1}^{K} b_{ik} v_{ik} + \sum_{j=1}^{H} c_j h_j + \sum_{i \in \mathcal{O}} \sum_{k=1}^{K} \sum_{j=1}^{H} W_{ikj} v_{ik} h_j \right).$$

- The gradients of the marginal log-likelihood are

$$\begin{aligned}
\nabla_{W_{ikj}} \log p(v) &= \mathbb{E}_{h_j \sim p(h_j | \mathbf{v} = v)} v_{ik} h_j - \mathbb{E}_{(v', h') \sim p(\mathbf{v}, \mathbf{h})} v'_{ik} h'_j, \\
\nabla_{b_{ik}} \log p(v) &= v_{ik} - \mathbb{E}_{(v', h') \sim p(\mathbf{v}, \mathbf{h})} v'_{ik}, \\
\nabla_{c_j} \log p(v) &= \mathbb{E}_{h_j \sim p(h_j | \mathbf{v} = v)} h_j - \mathbb{E}_{(v', h') \sim p(\mathbf{v}, \mathbf{h})} h'_j.
\end{aligned}$$

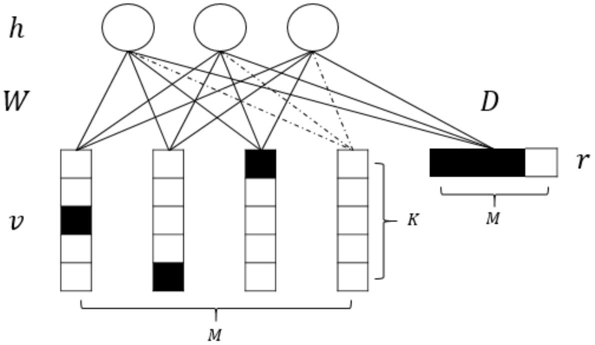
Conditional RBM

- Let $\mathbf{r} \in \{0, 1\}^M$ be a binary vector of length M, indicating which movies the user rated.

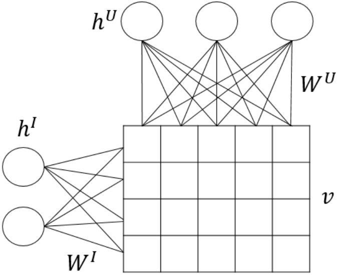
- Given $\mathbf{r} = r$, the joint probability distribution is

$$P(\mathbf{v} = v, \mathbf{h} = h | \mathbf{r} = r)$$

$$= \frac{1}{Z} \exp\left(\sum_{i \in \mathcal{O}} \sum_{k=1}^{K} b_{ik} v_{ik} + \sum_{j=1}^{H} c_j h_j + \sum_{i \in \mathcal{O}} \sum_{k=1}^{K} \sum_{j=1}^{H} W_{ikj} v_{ik} h_j + \sum_{i=1}^{M} \sum_{j=1}^{H} D_{ij} r_i h_j \right).$$



Hybrid RBM

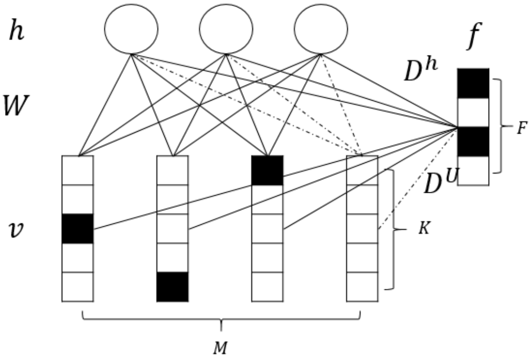


$$P(h^U_{ip} = 1 | \mathbf{v}_{i,:} = v_{i,:}) = \sigma\left(b^U_p + \sum_{j \in \mathcal{O}_{i,:}} W^U_{jp} v_{ij}\right)$$

$$P(h^I_{jq} = 1 | \mathbf{v}_{:,j} = v_{:,j}) = \sigma\left(b^I_q + \sum_{i \in \mathcal{O}_{:,j}} W^I_{iq} v_{ij}\right)$$

$$v_{ij} = \frac{1}{2}\left[c^U_i + \sum_{p=1}^{H^U} W^U_{jp} h^U_{ip} + c^I_j + \sum_{q=1}^{H^I} W^I_{iq} h^I_{jq}\right]$$

Item Category aware Conditional RBM



$$P(\mathbf{v} = v, \mathbf{h} = h | \mathbf{f} = f)$$

$$= \frac{1}{Z} \exp \left(\sum_{i \in \mathcal{O}} \sum_{k=1}^{K} b_{ik} v_{ik} + \sum_{j=1}^{H} c_j h_j + \sum_{i \in \mathcal{O}} \sum_{k=1}^{K} \sum_{j=1}^{H} W_{ikj} v_{ik} h_j + \sum_{j=1}^{H} \sum_{q=1}^{F} D_{jq}^{H} h_j f_q + \sum_{i \in \mathcal{O}} \sum_{k=1}^{K} \sum_{q=1}^{F} D_{ikq}^{U} v_{ik} f_q \right) .$$