Recommender System and Data Analysis

Boyoung Kim

Seoul National University

March 3, 2017
Outline

1 Collaborative Filtering Using Auto-encoder
   - Introduction
   - AutoRec (Suvash Sedhain, et al., 2015)
   - Iterative method using Auto-encoder
   - Experimental results

2 Recommendation Based On Click Through Rates
   - Introduction
   - Likelihood based approach
   - Experimental results
Outline

1. Collaborative Filtering Using Auto-encoder
   - Introduction
   - AutoRec (Suvash Sedhain, et al., 2015)
   - Iterative method using Auto-encoder
   - Experimental results

2. Recommendation Based On Click Through Rates
   - Introduction
   - Likelihood based approach
   - Experimental results
Introduction

- Some Deep Learning models can be used to model tabular data, such as user’s ratings of movies.

Jester 5k dataset

- 5000 users and 100 jokes.
- Rating between -10.00 and 10.00
- Users give the score for a few jokes.

<table>
<thead>
<tr>
<th></th>
<th>j1</th>
<th>j2</th>
<th>j3</th>
<th>j4</th>
<th>j5</th>
<th>j6</th>
<th>j7</th>
<th>j8</th>
<th>j9</th>
<th>j10</th>
<th>j11</th>
<th>j12</th>
</tr>
</thead>
<tbody>
<tr>
<td>u883</td>
<td>-0.49</td>
<td>-4.27</td>
<td>0.29</td>
<td>1.89</td>
<td>-1.12</td>
<td>0.39</td>
<td>-2.52</td>
<td>-3.01</td>
<td>-0.78</td>
<td>-4.03</td>
<td>1.75</td>
<td>-0.63</td>
</tr>
<tr>
<td>u14061</td>
<td>0.49</td>
<td>2.91</td>
<td>-2.62</td>
<td>NA</td>
<td>-0.19</td>
<td>-0.15</td>
<td>-6.75</td>
<td>2.14</td>
<td>NA</td>
<td>-3.20</td>
<td>5.44</td>
<td>-4.03</td>
</tr>
<tr>
<td>u6974</td>
<td>0.49</td>
<td>-6.12</td>
<td>-7.04</td>
<td>-6.50</td>
<td>8.06</td>
<td>-6.65</td>
<td>5.78</td>
<td>3.45</td>
<td>-2.86</td>
<td>7.96</td>
<td>-3.35</td>
<td>-4.81</td>
</tr>
<tr>
<td>u5081</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2.48</td>
<td>5.97</td>
<td>-8.69</td>
<td>-5.97</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>u3328</td>
<td>NA</td>
<td>-2.28</td>
<td>NA</td>
<td>NA</td>
<td>-7.62</td>
<td>-7.04</td>
<td>-2.38</td>
<td>-7.52</td>
<td>NA</td>
<td>-0.58</td>
<td>-3.01</td>
<td>-3.98</td>
</tr>
<tr>
<td>u8103</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>-3.59</td>
<td>NA</td>
<td>4.56</td>
<td>-0.49</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
Deep Neural Network (Deep learning)

Figure 1: Classification example of deep neural network
Auto-encoder

- **Unsupervised learning version of Neural Network.**
- AE can be used for dimensionality reduction of high-dimensional data.
- AE generate a hidden representation from an input, and reconstruct the output as the input from the hidden representation.
- Setting the target values to be equal to the input: $\hat{x} \approx x$. 
Auto-encoder

Figure 2: Architecture of autoencoder
Auto-encoder

- Suppose that $n$ inputs and $F$ hidden units.
- Then the hidden and output units are:

$$h_j = g\left(\sum_{i=1}^{n} V_{ij}x_i + a_j\right) \text{ for } j = 1, \cdots, F$$

$$\hat{x}_i = f\left(\sum_{j=1}^{F} W_{ij}h_j + b_i\right) \text{ for } i = 1, \cdots, n$$

where $a \in \mathbb{R}^F$ and $b \in \mathbb{R}^n$ are bias vectors,
$V \in \mathbb{R}^{n \times F}$ and $W \in \mathbb{R}^{n \times F}$ are weight matrices and
$f(\cdot)$ and $g(\cdot)$ are activation functions (eg, $f(x) = 1/(1 + e^{-x})$).
AutoRec : Autoencoders Meet Collaborative Filtering (Suvash Sedhain, et al., 2015)

- We use different Autoencoder for each user.
- Item-based AutoRec use Autoencoder for each item.

**Figure 3**: User-based AutoRec model
Itarative method using Auto-encoder

- Fill in 0 or mean of ratings that users have not rated. Consider it as input.
- Get the output from the Auto-encoder model.

- Iterate until convergence
  - Fix the ratings which the users have rated and fill in the predicted values (in the previous step) that users have not rated. Consider it as input.
  - Get the output from the Auto-encoder model.
Experimental results

- Training data: 80%, Test data: 20%
- Test RMSE = \( \sqrt{\sum_{u,i \in \text{Test set}} (r_{ui} - \hat{r}_{ui})^2 / \lvert \text{Test set} \rvert} \).

Table 1: Comparison of the test RMSE

<table>
<thead>
<tr>
<th>Methods</th>
<th>test RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix Factorization</td>
<td>4.1645</td>
</tr>
<tr>
<td>Personalized</td>
<td>4.1283</td>
</tr>
<tr>
<td>U-Autorec</td>
<td>4.3570</td>
</tr>
<tr>
<td>I-Autorec</td>
<td>4.1445</td>
</tr>
<tr>
<td>Iterative method</td>
<td>4.2488</td>
</tr>
</tbody>
</table>
Outline

1. Collaborative Filtering Using Auto-encoder
   - Introduction
   - AutoRec (Suvash Sedhain, et al., 2015)
   - Iterative method using Auto-encoder
   - Experimental results

2. Recommendation Based On Click Through Rates
   - Introduction
   - Likelihood based approach
   - Experimental results
Click Trough Rates data Analysis

- The effective of a particular online ad on a particular website.
- Observe the number of clicks and exposures of the ads to users.
- Predict
  \[\text{Click through rate(\%)} = \frac{\# \text{ of clicks}}{\# \text{ of exposures}}\]
- Recommend the best ad for a specific user.
Likelihood based approach

- $n$ users, $p$ ads.
- $X_{ua}$: the number of clicks of the ad $a$ to user $u$.
- $N_{ua}$: the number of exposures of the ad $a$ to user $u$.

Likelihood based approach

- Assume $X_{ua} \sim Bin(N_{ua}, p_{ua})$.
- Log likelihood

$$l(p) = \sum_{u=1}^{n} \sum_{p=1}^{a} \{X_{ua} \log p_{ua} + (N_{ua} - X_{ua}) \log (1 - p_{ua})\}$$

- Estimate $p_{ua}$ that maximize the log likelihood.
Likelihood based approach

Propose three models for $p_{ua}$.

1) Ad-wise probability model
   - Assume that $p_{ua} = p_a$ for all $u$.
   - ML estimator: $\hat{p}_{ua} = \frac{\sum_u X_{ua}}{\sum_u N_{ua}}$.

2) Additive model
   - Assume that $\logit p_{ua} = \alpha + \beta_u + \gamma_a$.

3) Matrix factorization
   - For a pre-selected positive integer $K$, assume
     $$\logit p_{ua} = \alpha + \beta_u + \gamma_a + \sum_{k=1}^K r_{uk} \cdot q_{ka}.$$
Experimental results

**Table 2:** Predicted probability that user40 clicks the ad205 of models

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{p}_{u=40,a=205}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad-wise</td>
<td>0.00785</td>
</tr>
<tr>
<td>Additive</td>
<td>0.00567</td>
</tr>
<tr>
<td>Matrix factorization</td>
<td>0.00566</td>
</tr>
</tbody>
</table>

**Table 3:** Comparison of test log-likelihood of models

<table>
<thead>
<tr>
<th>Model</th>
<th>Test log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad-wise</td>
<td>-49489.6</td>
</tr>
<tr>
<td>Additive</td>
<td>-47066.9</td>
</tr>
<tr>
<td>Matrix factorization</td>
<td>-47033.1</td>
</tr>
</tbody>
</table>