Multi-view Machines

Bokai Cao, Hucheng Zhou, Guoqiang Li and Philip S. Yu
Multi-view data is prevalent and important in web applications.

- Web images on Instagram
  - Visual information
  - Textual tags
- Web documents on Wikipedia
  - URL
  - Words on the page
- Web search on Bing
  - User profile
  - Ad information
  - Query description

Multi-view Data
Multi-view Data

Multi-view data is prevalent and important in web applications.

- Web documents on Wikipedia
  - URL
  - Words on the page
- Web images on Instagram
  - Visual information
  - Textual tags
- Web search on Bing
  - User profile
  - Ad information
  - Query description
Support Vector Machines (SVMs)

[Vapnik 1995] The linear SVM model (and LR etc.) is limited to the first-order interactions. Polynomial SVMs are likely to degenerate into linear SVMs on sparse datasets.
Support Tensor Machines (STMs)

[Cao et al. 2014] The STM model explores only the highest-order interactions.
Factorization Machines (FMbs)

[Rendle 2010] The FM model uses separate parameters to approximate interactions in different orders.
Multi-view Machines (MVMs)

MVMs are a general predictor that can work with different loss functions for a variety of machine learning tasks.

MVMs include the full-order interactions between multi-view features in a compact model.

MVMs jointly factorize the interaction parameters in different orders to learn a consistent representation under sparsity.
Experiments

• Datasets
  • MovieLens: users (138,493), movies (27,278) and implicit feedback (27,278)
  • BingAds: queries (958,426), ads (1,935,510) and impressions (18)

• Baselines
  • Linear regression/logistic regression (LR)
  • Tensor factorization (TF)
  • Multi-view FMs (mvFM, mvFM-3d, mvFM-reg)

• Implementation
  • Gradient descent with AdaGrad
  • GraphX in Spark with iterative forward and backward steps
Experiments

Table 5: Prediction accuracy. $\downarrow$ indicates the smaller the value the better the performance; $\uparrow$ indicates the larger the value the better the performance.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>MOVIELENS (RMSE) $\downarrow$</th>
<th>BINGADS (AUC) $\uparrow$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVM</td>
<td>0.8376</td>
<td>0.7917</td>
</tr>
<tr>
<td>FM</td>
<td>0.8681</td>
<td>0.7872</td>
</tr>
<tr>
<td>mvFM</td>
<td>0.8447</td>
<td>0.7729</td>
</tr>
<tr>
<td>mvFM-3D</td>
<td>0.9060</td>
<td>0.7201</td>
</tr>
<tr>
<td>mvFM-REG</td>
<td>0.9807</td>
<td>0.6947</td>
</tr>
<tr>
<td>TF</td>
<td>0.8572</td>
<td>0.6645</td>
</tr>
<tr>
<td>LR</td>
<td>1.0017</td>
<td>0.7450</td>
</tr>
</tbody>
</table>
Acknowledgements

• WSDM Student Scholarship

• SIGIR Student Travel Grant

• We also thank Dinggang Shen for his discussions, Chunhui Zhang from the Bing Ads team for providing the dataset and Bo Zhao for helping with the experiments.