构建电商端到端人工智能解决方案

2018年9月25日

洪亮劫
Etsy数据科学主管

JD.Com
洪亮劼

• Etsy数据科学主管 – (2016 至今)
  主管团队开发搜索、推荐和计算广告学的解决方案
  团队分布在纽约、旧金山和多伦多

• 雅虎研究院高级研发经理 - (2013 – 2016)
  管理个性化推荐和移动搜索机器学习的解决方案

• Learning Within-Session Budgets from Browsing Trajectories for Item Recommendations (RecSys 2018)
• Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)
• Buzzsaw: A System for High Speed Feature Engineering (SysML 2018)
• Ensemble Learning based CTR Prediction Solution (AdKDD 2017/KDD 2017)
• Multi-modal Deep-learning based Search Solution (KDD 2016)
议程

• Etsy的简要介绍

• 人工智能在电子商务领域的挑战

• 构建端到端人工智能解决方案
Etsy的简要介绍
Etsy – 一个全球的电商平台
Etsy – 一个全球的电商平台

What can you sell on Etsy?

Handmade Goods

Vintage

Craft Supplies

(20 years or older)
统计数字

190万 活跃卖家
3170万 活跃买家
30亿 年GMV
4500多万 商品

Photo by Kirsty Lyn Jameson
SELLER
Pursues craft, grows business

ETSY
Invests in the platform and delivers a global base of buyers

ETSY
Facilitates the transaction

BUYER
Funds unique goods that are hard to find elsewhere

Etsy Empowerment Loop
人工智能在电子商务领域的挑战
人工智能在电子商务领域的挑战

通用搜索
人工智能在电子商务领域的挑战
人工智能在电子商务领域的挑战

通用搜索

• 传统信息检索技术（1950 – 1990）
  TF-IDF, BM25, Language Models

• 排序学习（Learning to Rank）（2000 – 2010）
  RankSVM, GBDT, LambdaMART

• 基于深度学习的排序学习（2013 – 今）
  DSSM, DESM, IRGAN
人工智能在电子商务领域的挑战

通用搜索

- Cranfield范式（Paradigm）和测试数据集（Test Collections）（1950 – 今）
  TREC (1992 – Present)
  Microsoft Learning To Rank (2007 – 2009)
  Yahoo Learning To Rank Challenge (2011)

- 理解隐回馈（Implicit Feedback）和相关性（Relevance）之间的关系（2000 – 今）
  Thorsten Joachims的工作
  Eugene Agichtein的工作
人工智能在电子商务领域的挑战

通用推荐
传统推荐系统

• 基于内容的推荐系统（1990 – 2000）
  TF-IDF, PLSA

• 基于相似相关的推荐系统（2005 – 2013）
  Matrix Factorization, Neighborhood Methods

• 基于深度学习的推荐系统（2013 – 今）
  DNN, RNN, Deep RL
人工智能在电子商务领域的挑战

传统推荐系统

- 静态数据集（1990–今）
  - MovieLens
  - Netflix
  - Yahoo Music

- 评分预测、评分排序（2005–今）
  - RMSE，NDCG，MAP
人工智能在电子商务领域的挑战
人工智能在电子商务领域的挑战

电商推荐

Recommendations for You, Liangjie

Our picks for Liangjie  See more
人工智能在电子商务领域的挑战

挑战一 相关度
人工智能在电子商务领域的挑战

挑战一 相关度
人工智能在电子商务领域的挑战

挑战二 用户满意度
人工智能在电子商务领域的挑战

挑战二 用户满意度
人工智能在电子商务领域的挑战

挑战三 发现（Discovery）
人工智能在电子商务领域的挑战

电商人工智能 – 挑战

• 相关性

• 用户满意度

• 发现
构建端到端人工智能解决方案
构建端到端人工智能解决方案

简单的人工智能解决方案

• 构建数据集

• 挑选算法

• 上线测试
构建端到端人工智能解决方案

简单的人工智能解决方案

• 构建数据集
• 挑选算法
• 上线测试

Not Work
构建端到端人工智能解决方案

线下建模
- RMSE
- AUC
- MAP
- NDCG

在线测试
- 点击率
- 转化率
- 驻留时间
- 会话级收入

长期指标
- 日活用户数
- 月活用户数
- 长期收入
构建端到端人工智能解决方案

线下建模
- RMSE
- AUC
- MAP
- NDCG

在线测试
- 点击率
- 转化率
- 驻留时间
- 会话级收入

长期指标
- 日活用户数
- 月活用户数
- 长期收入

不匹配
构建端到端人工智能解决方案

线下建模
- RMSE
- AUC
- MAP
- NDCG

在线测试
- 点击率
- 转化率
- 驻留时间
- 会话级收入

长期指标
- 日活用户数
- 月活用户数
- 长期收入
构建端到端人工智能解决方案

端到端的人工智能解决方案

• 理想方案
  能否直接从线下建模优化在线测试目标
  能否直接从线下建模优化长期目标
  能否直接优化在线测试目标
  能否直接从在线测试目标优化长期目标
  能否直接优化长期目标

• 近似方案
  能否探究线下建模和在线测试指标的关系
  能否探究在线测试指标和产品长期指标的关系

• 迭代前进
构建端到端人工智能解决方案

线下建模 → 在线测试 → 长期指标

- 日活用户数
- 月活用户数
- 长期收入
构建端到端人工智能解决方案

端到端的人工智能解决方案

• 理想方案
  能否直接从线下建模优化在线测试目标
  能否直接从线下建模优化长期目标
  **能否直接优化在线测试目标**
  能否直接从在线测试目标优化长期目标
  能否直接优化长期目标

• 近似方案
  能否探究线下建模和在线测试指标的关系
  能否探究在线测试指标和产品长期指标的关系

• 迭代前进
构建端到端人工智能解决方案

端到端的人工智能解决方案

• 理想方案
  能否直接从线下建模优化在线测试目标
  能否直接从线下建模优化长期目标
  能否直接优化在线测试目标
  能否直接从在线测试目标优化长期目标
  能否直接优化长期目标

• 近似方案
  能否探究线下建模和在线测试指标的关系
  能否探究在线测试指标和产品长期指标的关系

• 迭代前进
构建端到端人工智能解决方案

线下建模和在线测试指标的关系

• Etsy推荐模块
构建端到端人工智能解决方案

线下建模和在线测试指标的关系

- Etsy推荐模块
  - 线下模型指标AUC
  - 在线测试指标 — 模块级收入
  - 在线测试指标 — 会话级收入
构建端到端人工智能解决方案

线下建模和在线测试指标的关系

- Etsy推荐模块
  线下模型指标AUC
  在线测试指标 - 模块级收入
  在线测试指标 - 会话级收入

- 结论
  首页模块：AUC和会话级收入的相关系数为0.34
  商品页推荐模块：AUC和会话级收入的相关系数为0.025
  购物车页推荐模块：AUC和会话级收入的相关系数为0.022

  首页模块：AUC和模块级收入的相关系数为0.56
  商品页推荐模块：AUC和模块级收入的相关系数为0.59
  购物车页推荐模块：AUC和模块级收入的相关系数为-0.388
构建端到端人工智能解决方案

线下建模和在线测试指标的关系

- Etsy搜索模块
  - 线下两个模型NDCG的差值
  - 在线测试指标 - 会话级收入的差值
构建端到端人工智能解决方案

线下建模和在线测试指标的关系

- **Etsy搜索模块**
  - 线下两个模型NDCG的**差值**
  - 在线测试指标—会话级收入的**差值**

- **结论**
  - 线下两个模型NDCG的**差值**和会话级收入的**差值**：相关系数 0.21
  - 线下每1%d因为模型带来的NDCG提升，都会带来$331,217.04的会话级收入
构建端到端人工智能解决方案

端到端的人工智能解决方案

• 理想方案
  能否直接从线下建模优化在线测试目标
  能否直接从线下建模优化长期目标
  能否直接优化在线测试目标
  能否直接从在线测试目标优化长期目标
  能否直接优化长期目标

• 近似方案
  能否探究线下建模和在线测试指标的关系
  能否探究在线测试指标和产品长期指标的关系

• 迭代前进
构建端到端人工智能解决方案
Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

- **Liang Wu**, PhD Student from Arizona State University
- **Diane Hu**, Manager of Data Science at Etsy
- **Liangjie Hong**, Head of Data Science at Etsy
Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

How to Optimize Gross-Merchandise-Value (GMV)?
Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

How to Optimize *Gross-Merchandise-Value* (GMV)?

\[
GMV = \sum_{s \in S} \sum_{i^s} \text{Price}(i^s) \cdot Pr(\Phi = 1 | i^s, q^s) \cdot \text{Price of } i^s \cdot \text{Prob of purchase}
\]
Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

Purchase Decision Process

![Diagram showing search page and product page](image)
Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

• Click Decision(s) from Search-Result-Page (SERP)

• Purchase Decision(s) from Listing Page

\[
Pr(\Phi = 1|i, q) = Pr(\Psi = 1|i, q) \cdot \underbrace{Pr(\Phi = 1|\Psi = 1, i, q)}_{\text{click model}} \cdot \underbrace{Pr(\Psi = 1|i, q)}_{\text{purchase model}}
\]
Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

- Click Decision(s) from Search-Result-Page (SERP)
- Purchase Decision(s) from Listing Page

\[
NDCG_K(q) = N_{\text{max}}^{-1} \sum_{r=0}^{K-1} \frac{2^I(r^{-1})}{\log(1+r)},
\]
Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

• Click Decision(s) from Search-Result-Page (SERP)

• Purchase Decision(s) from Listing Page

\[
NDCG_K(q) = N_{max}^{-1} \sum_{r=0}^{K-1} \frac{2^l(r^{-1})}{\log(1+r)}
\]

• \( I \) is transformed from \textit{empirical GMV}.

• \( r \) is approximated by the product of a click model and a purchase model where the click model is a RankNet model and the purchase model is \textit{price-weighted} logistic regression.
Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

<table>
<thead>
<tr>
<th>Relevance</th>
<th>Low Level</th>
<th>Sum of TF</th>
<th>Sum of Log TF</th>
<th>Sum of Normalized TF</th>
<th>Sum of IDF</th>
<th>Sum of Log IDF</th>
<th>Sum of ICF</th>
<th>Sum of TF-IDF</th>
<th>Sum of Log TF-IDF</th>
<th>TF-Log IDF</th>
<th>Length</th>
<th>Log Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High Level</td>
<td>BM25</td>
<td>Log BM25</td>
<td>LM_DJR</td>
<td>LM_JM</td>
<td>LM_ABS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revenue</td>
<td></td>
<td>Price</td>
<td>Price – Cat. Mean</td>
<td>(Price – Cat. Mean)/Cat. Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sessions</th>
<th>Queries</th>
<th>Items</th>
<th>Avg. Items per Session</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>334,931</td>
<td>239,928</td>
<td>6,347,251</td>
<td>19.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Buyers</th>
<th>Sellers</th>
<th>Avg. Items per Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>631,778</td>
<td>270,239</td>
<td>550,025</td>
<td>26.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Click</th>
<th>RankNet [2]</th>
<th>RNet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RankBoost [8]</td>
<td>RBoost</td>
</tr>
<tr>
<td></td>
<td>AdaRank [36]</td>
<td>ARank</td>
</tr>
<tr>
<td></td>
<td>LambdaRank [3]</td>
<td>LRank</td>
</tr>
<tr>
<td></td>
<td>ListNet [4]</td>
<td>LNet</td>
</tr>
<tr>
<td></td>
<td>MART [10]</td>
<td>MART</td>
</tr>
<tr>
<td></td>
<td>LambdaMART [35]</td>
<td>LMART</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Purchase</th>
<th>SVM [5]</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Logistic Regression [25]</td>
<td>LR</td>
</tr>
<tr>
<td></td>
<td>Random Forest [19]</td>
<td>RM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Both</th>
<th>Weighted Purchase [41]</th>
<th>WT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LMART+RM</td>
<td>LMRM</td>
</tr>
<tr>
<td></td>
<td>LETORIF</td>
<td>LETORIF</td>
</tr>
</tbody>
</table>
Figure 2: Position distribution of items being purchased in the top 4 spots of a search result page. The first position achieves the most purchases, while nearly 70% of purchases are in the lower positions.
构建端到端人工智能解决方案

<table>
<thead>
<tr>
<th>Category</th>
<th>Method</th>
<th>Click NDCG@5</th>
<th>Purchase NDCG@5</th>
<th>Revenue NDCG@5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Train</td>
<td>Vali</td>
<td>Test</td>
</tr>
<tr>
<td>Click</td>
<td>RNet</td>
<td>0.1743</td>
<td>0.1731</td>
<td>0.1378**</td>
</tr>
<tr>
<td></td>
<td>RBoost</td>
<td>0.2150</td>
<td>0.1768</td>
<td>0.1323**</td>
</tr>
<tr>
<td></td>
<td>ARank</td>
<td>0.1718</td>
<td>0.1711</td>
<td>0.1351**</td>
</tr>
<tr>
<td></td>
<td>LRank</td>
<td>0.1694</td>
<td>0.1688</td>
<td>0.1360**</td>
</tr>
<tr>
<td></td>
<td>LNet</td>
<td>0.1665</td>
<td>0.1703</td>
<td>0.1355**</td>
</tr>
<tr>
<td></td>
<td>MART</td>
<td>0.2700</td>
<td>0.1758</td>
<td>0.1380**</td>
</tr>
<tr>
<td></td>
<td>LMART</td>
<td>0.3056</td>
<td>0.1777</td>
<td>0.1412 **</td>
</tr>
<tr>
<td>Purchase</td>
<td>SVM</td>
<td>0.1785</td>
<td>0.1772</td>
<td>0.1336**</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.1978</td>
<td>0.1739</td>
<td>0.1310**</td>
</tr>
<tr>
<td></td>
<td>RM</td>
<td>0.3359</td>
<td>0.1698</td>
<td>0.1363**</td>
</tr>
<tr>
<td>Both</td>
<td>WT</td>
<td>0.1970</td>
<td>0.1682</td>
<td>0.1334**</td>
</tr>
<tr>
<td></td>
<td>LMRM</td>
<td>0.2943</td>
<td>0.2597</td>
<td>0.1354**</td>
</tr>
<tr>
<td></td>
<td>LETORIF</td>
<td>0.1765</td>
<td>0.1550</td>
<td>0.1351**</td>
</tr>
</tbody>
</table>

Symbol * indicates that the method is outperformed by the best one by 0.05 statistical significance level. ** indicates 0.01.
## 构建端到端人工智能解决方案

<table>
<thead>
<tr>
<th>Category</th>
<th>Method</th>
<th>Rev@1</th>
<th>Rev@2</th>
<th>Rev@3</th>
<th>Rev@4</th>
<th>Rev@5</th>
<th>Rev@6</th>
<th>Rev@7</th>
<th>Rev@8</th>
<th>Rev@9</th>
<th>Rev@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Click</td>
<td>RNet</td>
<td>4.47</td>
<td>4.69</td>
<td>4.89</td>
<td>4.91</td>
<td>5.06</td>
<td>5.23</td>
<td>5.21</td>
<td>5.33</td>
<td>5.46</td>
<td>5.55</td>
</tr>
<tr>
<td></td>
<td>RBoost</td>
<td>4.57</td>
<td>4.69</td>
<td>4.69</td>
<td>4.76</td>
<td>4.97</td>
<td>5.17</td>
<td>5.23</td>
<td>5.36</td>
<td>5.49</td>
<td>5.57</td>
</tr>
<tr>
<td></td>
<td>ARank</td>
<td>4.37</td>
<td>4.66</td>
<td>4.76</td>
<td>4.90</td>
<td>5.06</td>
<td>5.20</td>
<td>5.33</td>
<td>5.47</td>
<td>5.59</td>
<td>5.67</td>
</tr>
<tr>
<td></td>
<td>LRanck</td>
<td>4.38</td>
<td>4.61</td>
<td>4.74</td>
<td>4.86</td>
<td>5.07</td>
<td>5.25</td>
<td>5.42</td>
<td>5.42</td>
<td>5.67</td>
<td>5.78</td>
</tr>
<tr>
<td></td>
<td>LNet</td>
<td>4.30</td>
<td>4.59</td>
<td>4.78</td>
<td>4.99</td>
<td>5.16</td>
<td>5.35</td>
<td>5.49</td>
<td>5.61</td>
<td>5.63</td>
<td>5.63</td>
</tr>
<tr>
<td></td>
<td>MART</td>
<td>4.62</td>
<td>4.72</td>
<td>4.86</td>
<td>5.04</td>
<td>5.26</td>
<td>5.47</td>
<td>5.47</td>
<td>5.64</td>
<td>5.74</td>
<td>5.86</td>
</tr>
<tr>
<td></td>
<td>LMART</td>
<td>4.46</td>
<td>4.54</td>
<td>4.73</td>
<td>5.10</td>
<td>5.31</td>
<td>5.56</td>
<td>5.75</td>
<td>5.90</td>
<td>6.01</td>
<td>6.14</td>
</tr>
<tr>
<td>Purchase</td>
<td>SVM</td>
<td>4.41</td>
<td>4.54</td>
<td>4.76</td>
<td>4.77</td>
<td>4.95</td>
<td>5.16</td>
<td>5.34</td>
<td>5.50</td>
<td>5.64</td>
<td>5.77</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>4.29</td>
<td>4.65</td>
<td>4.65</td>
<td>4.69</td>
<td>4.74</td>
<td>4.81</td>
<td>4.94</td>
<td>4.97</td>
<td>5.11</td>
<td>5.11</td>
</tr>
<tr>
<td></td>
<td>RM</td>
<td>4.52</td>
<td>4.82</td>
<td>4.86</td>
<td>5.02</td>
<td>5.18</td>
<td>5.33</td>
<td>5.50</td>
<td>5.66</td>
<td>5.79</td>
<td>5.92</td>
</tr>
<tr>
<td>Both</td>
<td>WT</td>
<td>4.52</td>
<td>4.69</td>
<td>4.80</td>
<td>4.85</td>
<td>5.01</td>
<td>5.07</td>
<td>5.23</td>
<td>5.32</td>
<td>5.35</td>
<td>5.41</td>
</tr>
<tr>
<td></td>
<td>LMRM</td>
<td>4.42</td>
<td>4.50</td>
<td>4.72</td>
<td>5.08</td>
<td>5.23</td>
<td>5.41</td>
<td>5.57</td>
<td>5.60</td>
<td>5.73</td>
<td>5.85</td>
</tr>
<tr>
<td></td>
<td>LETORIF</td>
<td>4.58</td>
<td>4.90</td>
<td>5.08</td>
<td>5.47</td>
<td>5.64</td>
<td>5.85</td>
<td>6.02</td>
<td>6.19</td>
<td>6.40</td>
<td>6.54</td>
</tr>
</tbody>
</table>

Symbol * indicates that the method is outperformed by the best one by 0.05 statistical significance level, ** indicates 0.01.
构建端到端人工智能解决方案

- A simplified 2-Stage model deployed into recommendation, improved GMV +0.8%.
- A weighted purchase model deployed into search ranking, improved GMV +0.9%.
- An extended candidate selection deployed into search ranking, improved GMV +2.0%.
- A model heavily utilizing historical information deployed into search ranking, improved GMV +0.7%.
构建端到端人工智能解决方案

端到端的人工智能解决方案

- 理想方案
  能否直接从线下建模优化在线测试目标
  能否直接从线下建模优化长期目标
  **能否直接优化在线测试目标**
  能否直接从在线测试目标优化长期目标
  能否直接优化长期目标

- 近似方案
  能否探究线下建模和在线测试指标的关系
  能否探究在线测试指标和产品长期指标的关系

- 迭代前进
A Sequential Test for Selecting the Better Variant

- **Phyllis Ju**, PhD Student from Harvard University
- **Diane Hu**, Manager of Data Science at Etsy
- **Adam Henderson**, Staff Data Scientist at Etsy
- **Liangjie Hong**, Head of Data Science at Etsy
A Sequential Test for Selecting the Better Variant

• Stop Early

• Dynamic Allocation
构建端到端人工智能解决方案

端到端的人工智能解决方案

• 理想方案
  能否直接从线下建模优化在线测试目标
  能否直接从线下建模优化长期目标
  **能否直接优化在线测试目标**
  能否直接从在线测试目标优化长期目标
  能否直接优化长期目标

• 近似方案
  能否探究线下建模和在线测试指标的关系
  能否探究在线测试指标和产品长期指标的关系

• 迭代前进
议程

- Etsy的简要介绍
- 人工智能在电子商务领域的挑战
- 构建端到端人工智能解决方案
Thanks!