TEM: Tree-enhanced Embedding Model for Explainable Recommendation

Xiang Wang¹, Xiangnan He¹, Fuli Feng¹, Liqiang Nie², Tat-Seng Chua¹

¹ National University of Singapore
² Shandong University

Work presented in WWW 2018, Lyon, France
Value of Recommender Systems (RS)

- **Amazon**: 35% sales from recommendations
- **Netflix**: 80% TV shows discovered are recommended
- **Google News**: RS generates 38% more click-through
Our Goal:

- **Explainable**: be transparent in generating a recommendation & can identify the key rules for a prediction
- **Accurate**: achieve the same level or comparable performance as embedding-based methods
Embedding-based Models: Learn latent factors for each feature (IDs & side Info)

Matrix Factorization (MF)
- **Input:** user ID, item ID
- **Interaction:** Inner Product
  \[
  \hat{y}_{MF}(u, i) = b_0 + b_u + b_i + p_u^T q_i
  \]

Factorization Machine (FM)
- **Input:** user ID, item ID, side features ID
- **Interaction:** Element-wise Product
  \[
  \hat{y}_{FM}(x) = w_0 + \sum_{t=1}^{n} w_t x_t + \sum_{t=1}^{n} \sum_{j=t+1}^{n} v_t^T v_j \cdot x_t x_j.
  \]

Neural Network Methods
- NCF, Deep Crossing, Wide&Deep, DIN, NFM
Explainable Recommendation

Black-box Model

Recommended List:
- $\hat{y}_{u3} = 0.9$
- $\hat{y}_{u1} = 0.7$
- $\hat{y}_{u5} = 0.3$
- $\hat{y}_{u2} = 0.2$

Recommended Reason
Users of <City: SG, Age: 18-25, Style: Sci-Fi> like items of <Price: $$, Attribute: Science> ...

End User/ Data Scientist
- Why did RS recommend it?
- Why not something else?
- How do I correct an error?

End User/ Data Scientist
- I understand why
- I understand why not
- I know how to correct RS

Transparency, Trust, Explainability, Scrutability
**Cross Feature**: combinatorial feature that crosses (or multiplies) multiple individual input features.

**Why?**
- Higher-order feature interactions: e.g., $[\text{Age} \times \text{Occupation} \times \text{Gender}]$
- Explicit decision rules

Users of $<\text{Gender}=\text{female} & \text{Age}=20-25 & \text{Income Level}=$8,000> tend to adopt items of $<\text{Color}=\text{Pink} & \text{Product}=\text{Apple}>$. 
Tree-based Methods

- **Decision Tree (DT):**
  - Each node splits a feature variable into two decision edges based on a value.
  - A path from the root to a leaf -> a decision rule (like a cross feature).
  - The leaf node -> the prediction value.

- **Forest (ensemble of trees):**
  - Since a single tree may not be expressive enough, a typical way is to build a forest, i.e., an ensemble of additive trees.

\[
\hat{y}_{GBDT}(x) = \sum_{s=1}^{S} \hat{y}_{DT_s}(x),
\]

Prediction of the s-th tree

Leaf node \(v_{L_2}\) represents \([x_0 < a_0] \& [x_3 \geq a_3] \& [x_2 \neq a_2]\)
### Tree-based vs. Embedding-based Model

<table>
<thead>
<tr>
<th>Tree-based Model (e.g., GBDT)</th>
<th>Embedding-based Model (e.g., DNN, FM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Strong at continuous features</td>
<td>+ Strong at categorical features</td>
</tr>
<tr>
<td>+ Explainable</td>
<td>- Blackbox</td>
</tr>
<tr>
<td>+ Low serving cost</td>
<td>- High serving cost</td>
</tr>
<tr>
<td>- Weak generalization ability to unseen feature combinations.</td>
<td>+ Strong generalization ability to unseen feature combinations.</td>
</tr>
</tbody>
</table>

Why not combining the strengths of the two types of models?
Concrete Reasons: Explicitly discover effective cross features from rich side information of users & items

Explicit Decision Process: Estimate user-item matching score in an explainable way
Constructing Cross Features

- **Traditional Solution**: manually cross all values of feature variables
- **Our Solution**: GBDT -> automatically identify useful cross features
- **We build GBDT on user attributes and item attributes.**

$$q = GBDT(x|Q) = [Q_1(x), \cdots, Q_S(x)]$$

Explicit Cross Features with easy-to-comprehend semantics!
Primary Consideration: seamlessly integrate cross features with embedding-based CF

Our Solution: embed them into user-item latent space

$$q = GBDT(x|Q) = [Q_1(x), \cdots, Q_3(x)].$$  
Multi-hot encoding of cross-feature ID

$$\mathcal{V} = \{q_1v_1, \cdots, q_Lv_L\}$$  
Embedding for each cross-feature ID

The correlations among cross features may be captured in the embedding space.
 prone to comprehend cross features

\[ e_{avg}(u, i, \mathcal{V}) = \frac{1}{|\mathcal{V}|} \sum_{v_l \in \mathcal{V}} w_{uil} v_l, \]

\[ e_{max}(u, i, \mathcal{V}) = \max_{v_l \in \mathcal{V}} (w_{uil} v_l), \]

- Easy-to-comprehend cross features
- Explicit contribution of each cross feature to the final prediction
**Primary Consideration:** explicit decision process & similarity-based + cross feature-based explanation mechanism

**Solution:** Simple linear regression

\[
\hat{y}_{TEM}(u, i, x) = b_0 + \sum_{t=1}^{m} b_t x_t + r_1^T (p_u \odot q_i) + r_2^T e(u, i, V)
\]

- **Similarity**
- **Cross Feature**

\[
L = \sum_{(u, i, x) \in O} -y_{ui} \log \sigma(\hat{y}_{ui}) - (1 - y_{ui}) \log (1 - \sigma(\hat{y}_{ui}))
\]

- Pointwise logloss
- Pointwise regression loss
- Pairwise Ranking loss
OUTLINE

• Introduction
• Motivation
• Tree-enhanced Embedding Model
• Experimental Results
• Conclusion
Experimental Settings

❖ Research Questions:

❖ **RQ1**: Compared with the state-of-the-art recsys methods, can TEM achieve comparable accuracy?

❖ **RQ2**: Can TEM make the recsys results easy-to-interpret by using cross features and the attention network?

❖ Tasks: Attraction Recommendation & Restaurant Recommendation

❖ Dataset: TripAdvisor

<table>
<thead>
<tr>
<th>Table 2: Statistics of the datasets.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>LON-A</td>
</tr>
<tr>
<td>NYC-R</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3: Statistics of the side information, where the dimension of each feature is shown in parentheses.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Side Information</td>
</tr>
<tr>
<td>LON-A/NYC-R User Feature</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>LON-A Attraction Feature</td>
</tr>
<tr>
<td>NYC-R Restaurant Feature</td>
</tr>
</tbody>
</table>
Baselines

- **XGBoost**: the state-of-the-art tree-based model
- **GBDT+LR [ADKDD’14]**: feeding the cross features extracted from GBDT into the logistic regression
- **GB-CENT [WWW’17]**: modeling categorical features with embedding-based model, numerical features with decision trees.
- **FM**: a generic embedding model that implicitly models all the second-order cross features
- **NFM [SIGIR’17]**: the state-of-the-art factorization model under the neural network framework

**Evaluation Protocols**:
- **logloss**: indicate the generalization ability of each model
- **ndcg@k**: reflect the top-k recommendation performance
RQ1: Overall Performance Comparison

- **Observations:**
  - TEM achieves the best performance w.r.t. logloss.
  - TEM achieves comparable ndcg@5 to NFM.

**Comparable Expressiveness & Accuracy**
Without cross feature modeling:

- All methods have worse performance
- TEM is still better than others, due to the utility of attention network
  (can learn which features are more important for a user-item prediction).
RQ2: Case Study of Explainability

We attribute the user’s preferences on *The View from the Shard* to her special interests in the item aspects of *Walk Around, Top Deck & Canary Wharf*.

TEM can provide more informative explanations based on a user’s preferred cross features.

**V130:** User Gender=Female] & [User Style=Peace and Quiet Seeker] ⇒ [Item Attribute=Sights & Landmarks] & [Item Tag=Walk Around]

**V148:** [User Age=30-40] & [User Country=USA] ⇒ [Item Tag=Top Deck & Canary Wharf]
We proposed a tree-enhanced embedding method (TEM), which seamlessly combines the generalization ability of embedding-based models with the explainability of tree-based models.

Owing to the explicit cross features from tree-based part & the easy-to-interpret attention network, the whole prediction process of our solution is transparent & self-explainable.

Future Work:
1. Jointly learn the tree-based and embedding-based
2. Relational reasoning over KG (symbolic logics) + Deep Learning
3. How to evaluate the quality of explanations?
THANK YOU

NExT research is supported by the National Research Foundation, Prime Minister’s Office, Singapore under its IRC@SG Funding Initiative.