**SIFN: A Sentiment-aware Interactive Fusion Network for Review-based Item Recommendation**

Kai Zhang¹, Hao Qian³, Qi Liu¹,², Zhiqiang Zhang³, Jun Zhou³, Jianhui Ma¹, Enhong Chen¹,²

¹ Anhui Province Key Lab of Big Data Analysis and Application, School of Data Science
² School of Computer Science and Technology, University of Science and Technology of China
³ Ant Group, Hangzhou, China
Review-based Item Recommendation?

• E-commerce platforms (e.g., Amazon, Alibaba) allow users to post their reviews towards products. The reviews may contain the opinions of users and the features of the items.

• Value of the reviews:
  • User-generated content
  • Carrying explanation
  • Rich textual information
Outline

1. Background
   a. Formal definition & previous methods
   b. Some existing problems
2. Our method: Sentiment-aware Interactive Fusion Network
3. Experiments
4. Conclusion
Outline

1. **Background**
   a. Formal definition & previous methods
   b. Some existing problems
2. **Our method: Sentiment-aware Interactive Fusion Network**
3. Experiments
4. Conclusion
Background – formal definition

To begin with, we first define the review-based item recommendation. It estimates the probability that a user rating at the candidate items based on the input feature representation.

predicts the rating $R_{u,i}$ of item $i$ by user $u$. General schema:

$$R_{u,i} = f(user\_reviews, item\_reviews)$$

- CNN / RNN
- Capsule network
- Attention methods
- Local & Global Info
- Aspect info
- ...
Background – previous methods

**CNN, RNN based-methods**

- ConvMF [D.H. Kim et al. Recsys’16]
  - Convolutional Matrix Factorization
  - Utilizing CNN to extract item latent features from item reviews

- Parallel CNN model: DeepCoNN [L. Zhang et al. WSDM’17]
  - Deep Cooperative Neural Networks
  - Jointly modeling user & item reviews; Dual CNN

- TransNet [R. Catherine et al. Recsys’17]
Background – previous methods

Deep Models with Attention for review-based recommendation

- **D-Attn** [S. Seo et al. Recsys’17]
  - Dual Local and Global attention for selective features
  - Combining local and global attention on review text

- **NARRE** [C. Chen et al. WWW’18]
  - Predicting the rating and learning the usefulness of reviews simultaneously

- **MPCN** [Tay et al. KDD’18], **CARL** [Wu et al. TOIS’19], **DAML** [Liu et al. KDD’19]
Background – shortcomings

Although these works achieved significant performance improvement, they still suffer from two intrinsic issues:

1. largely ignoring the explicit sentiment polarity of reviews
2. neglect the personalized interaction of reviews with user/item

implicitly mining the semantic information and interactions of reviews may lead to sub-optimal prediction because the reviews’ sentiment label has not been applied to the training process
Outline

1. Background
   a. Formal definition & previous methods
   b. Some existing problems
2. Our method: Sentiment-aware Interactive Fusion Network
3. Experiments
4. Conclusion
Our solution – Sentiment-aware Interactive Fusion Network

Exploring the explicit sentiment and personalized interaction:

1. There exists a large amount of sentiment labels which carry the user attitudes and preferences (i.e., which kinds of item user may like or dislike)
2. it is necessary to model the interactions between each review and user/item features.

E.g., sentiment for each review

E.g., weigh and fuse
Our solution – **Sentiment-aware Interactive Fusion Network**

Overall architecture of **SIFN**:

1. An Encoding Module (BERT embedding + Sentiment Learner)
2. Sentiment Prediction (MLP + softmax)
3. Rating Prediction (Rating Learner: aggregation + fusion)
Our solution – **Sentiment-aware Interactive Fusion Network**

---

Overall architecture of SIFN:

1. An Encoding Module (BERT embedding + **Sentiment Learner**)

2. Sentiment Prediction (MLP + softmax)

3. Rating Prediction (**Rating Learner**: aggregation + fusion)

\[
\alpha_i = \frac{\exp(\tanh(W_a e_i^U + b_a))}{\sum_{i=1}^{l} \exp(\tanh(W_a e_i^U + b_a))}
\]

\[
s_U = \sum_{i=1}^{l} \alpha_i e_i^U
\]

\[
f = d^U W_f d^I
\]

\[
p = (d^U + e_{id}^U) \odot (d^I + e_{id}^I) + W_f + b
\]
Outline

1. Background
   a. Formal definition & previous methods
   b. Some existing problems
2. Our method: Sentiment-aware Interactive Fusion Network
3. Experiments
4. Conclusion
Experiments

Dataset

✓ Benchmark Amazon dataset:
  • Music Instruments, Office Products, Digital Music, Tools, Video Games;

Baseline methods

① MF-based methods: PMF; ConvMF+
② Neural-based methods: DeepCoNN; D-Attn; NARRE; CARP

Mean Squared Error (MSE) as the evaluation metric

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Users</th>
<th># Items</th>
<th># Ratings</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Music Instruments</td>
<td>1,429</td>
<td>900</td>
<td>10,261</td>
<td>0.798%</td>
</tr>
<tr>
<td>Office Products</td>
<td>4,905</td>
<td>2,420</td>
<td>53,228</td>
<td>0.448%</td>
</tr>
<tr>
<td>Digital Music</td>
<td>5,540</td>
<td>3,568</td>
<td>64,664</td>
<td>0.327%</td>
</tr>
<tr>
<td>Tools</td>
<td>16,638</td>
<td>10,217</td>
<td>134,345</td>
<td>0.079%</td>
</tr>
<tr>
<td>Video Games</td>
<td>24,303</td>
<td>10,672</td>
<td>213,577</td>
<td>0.089%</td>
</tr>
</tbody>
</table>
Experiments

<table>
<thead>
<tr>
<th>Methods</th>
<th>Music Instruments</th>
<th>Office Products</th>
<th>Digital Music</th>
<th>Tools</th>
<th>Video Games</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMF [8]</td>
<td>1.398(+45.7%)</td>
<td>1.092(+35.7%)</td>
<td>1.206(+33.7%)</td>
<td>1.566(+39.3%)</td>
<td>1.672(+37.4%)</td>
<td>1.386(+38.6%)</td>
</tr>
<tr>
<td>ConvMF+ [6]</td>
<td>0.991(+23.4%)</td>
<td>0.960(+26.9%)</td>
<td>1.084(+26.3%)</td>
<td>1.240(+23.4%)</td>
<td>1.449(+27.7%)</td>
<td>1.145(+25.7%)</td>
</tr>
<tr>
<td>DeepCoNN [18]</td>
<td>0.814(+6.76%)</td>
<td>0.860(+18.4%)</td>
<td>1.058(+24.5%)</td>
<td>1.061(+10.5%)</td>
<td>1.145(+8.56%)</td>
<td>0.988(+13.9%)</td>
</tr>
<tr>
<td>D-Attn [12]</td>
<td>0.982(+22.7%)</td>
<td>0.825(+14.9%)</td>
<td>0.911(+12.3%)</td>
<td>1.043(+8.92%)</td>
<td>1.144(+8.48%)</td>
<td>0.981(+13.3%)</td>
</tr>
<tr>
<td>NARRE [2]</td>
<td>0.803(+5.48%)</td>
<td>0.848(+17.2%)</td>
<td>0.898(+11.0%)</td>
<td>1.029(+7.68%)</td>
<td>1.129(+7.26%)</td>
<td>0.941(+9.6%)</td>
</tr>
<tr>
<td>CARP [7]</td>
<td>0.773(+1.81%)</td>
<td>0.719(+2.36%)</td>
<td>0.820(+2.56%)</td>
<td>0.960(+1.04%)</td>
<td>1.084(+3.41%)</td>
<td>0.872(+2.4%)</td>
</tr>
<tr>
<td><strong>SIFN</strong></td>
<td><strong>0.759</strong></td>
<td><strong>0.702</strong></td>
<td><strong>0.799</strong></td>
<td><strong>0.950</strong></td>
<td><strong>1.047</strong></td>
<td><strong>0.851</strong></td>
</tr>
</tbody>
</table>

Overall performance

1. **MF-based** methods (e.g., PMF) consistently fall behind other methods
2. Neural-based methods (e.g., D-Attn) **outperform** MF-based ones by a large margin
3. our proposed **SIFN** model still outperforms CARP by **1.81%~3.41%**, which shows the superiority of the well designed interactive& Fusion module
Experiments

- **SIFN_sa**: replaces the sentence attention with a simple average sum pooling over all the reviews.
- **SIFN_fn**: removes the fusion network so that user and item features are disentangled without explicit interactions.
- **SIFN_in**: replaces the interactive network with commonly used Factorization Machine (FM) [11] to estimate the ratings.
- **SIFN_w2v**: replaces the BERT encoding of text reviews with commonly used pre-trained word embedding GloVe [9].
- **SIFN_sp**: removes the sentiment prediction task so that the model focuses on user-item rating prediction.

- the performance of **SIFN_sa** drops as it just assumes every review contributes equally for user-item rating.
- the performance of **SIFN_fn** also declines since the interactions of user and item reviews are not fully exploited.
- **SIFN_in** also declines because it is inefficient to perform second-order operations in the same space with FM.
- Not surprisingly, without BERT encoding, **SIFN_w2v** is incapable of representing deep semantic of reviews.
- Without sentiment prediction in **SIFN_sp**, there is no supervision towards attending sentiment-aware words in reviews, which are vital for the rating prediction.
Case study

1. In review (a), SIFN aligns the sentiment words, e.g., “hated” and “cheap”, with a rating of 1.08, which is consistent with actual value

2. In review (b), SIFN accurately predicts a rating of 4.98 by extracting the sentiment words, e.g., “love” and “perfectly”, while SIFN_sp is incapable of achieving this
Outline

1. Background
   a. Formal definition & previous methods
   b. Some existing problems
2. Our method: Sentiment-aware Interactive Fusion Network
3. Experiments
4. Conclusion
Conclusion

1. We highlight the explicit sentiment polarity in each review, and focus on modeling the multiple feature interactions between each review and user/item.

2. We propose a novel Sentiment-aware Interactive Fusion Network (SIFN) model with two main components, Sentiment Leaner and Rating Learner.

3. We conduct extensive experiments on five datasets so that the results demonstrate the effectiveness of our proposed method.
References


Thanks!