Introduction
ooo

Related Work
oo

Model
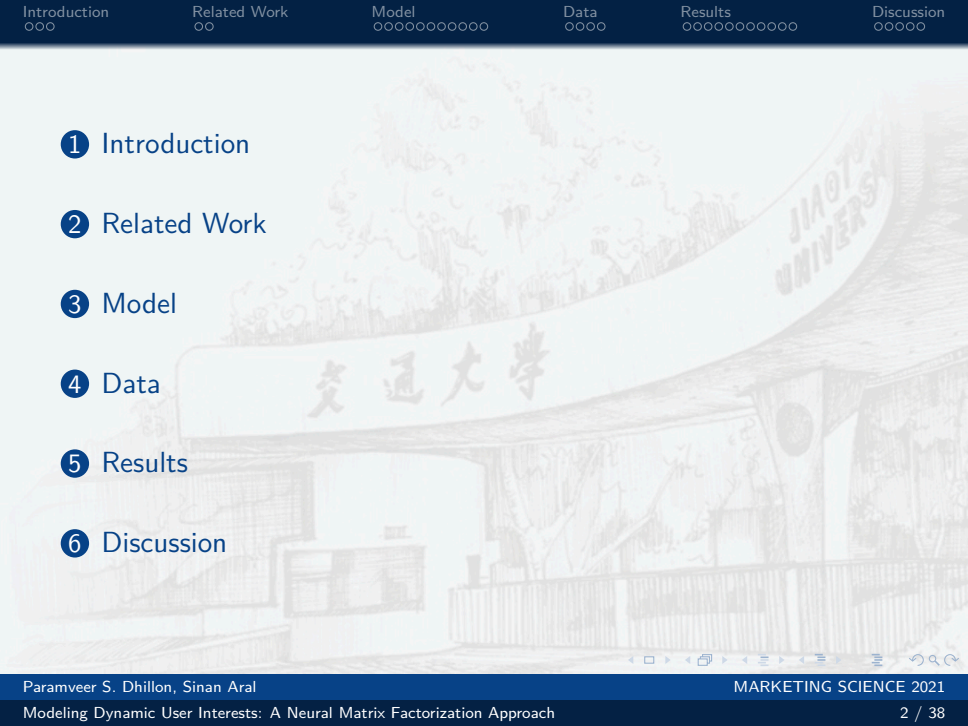ooooooooooo

Data
oooo

Results
ooooooooooo

Discussion
ooooo

Modeling Dynamic User Interests: A Neural Matrix Factorization Approach

Paramveer S. Dhillon, Sinan Aral

MARKETING SCIENCE 2021

2023 年 12 月 6 日

1 Introduction

2 Related Work

3 Model

4 Data

5 Results

6 Discussion

1 Introduction

2 Related Work
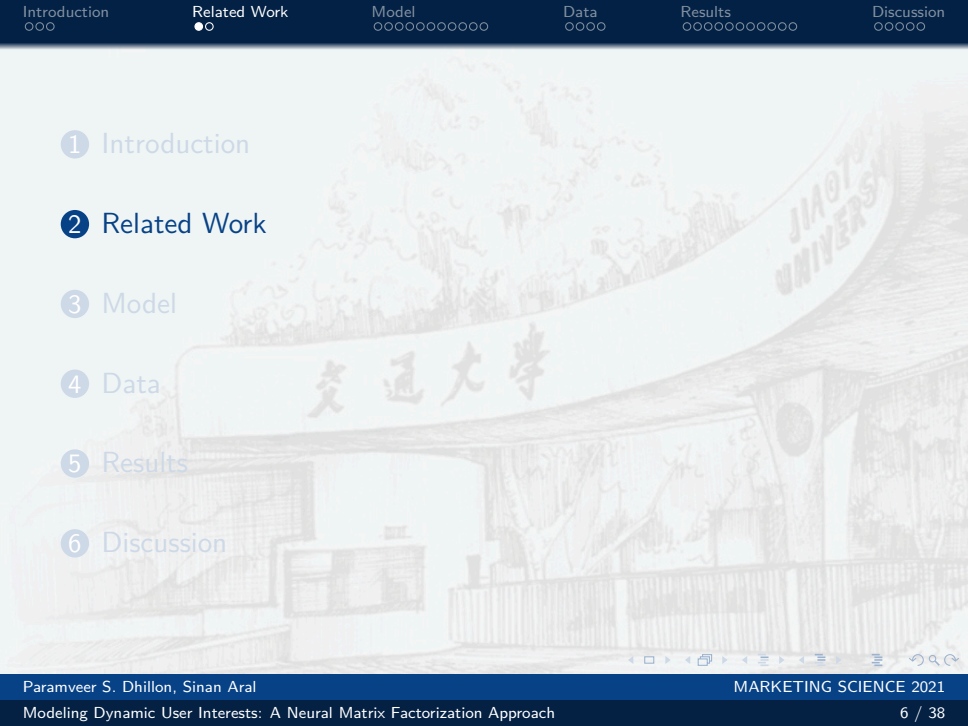
3 Model

4 Data

5 Results

6 Discussion

Background

- 消费者数据 (消费者行为数据和 UGC) 被用于营销领域的各类研究中;
- 三大挑战:
    - 高维稀疏的非结构化数据, 使用传统方法进行统计推断往往比较困难;
    - 数据生成过程本身的非平稳性 (动态性);
    - 不同于市场营销研究中常见的购买数据建模研究, 用户内容消费数据的建模更加复杂, 内容消费的种类总是在不断增加, 而且几乎没有重复的"动机"。
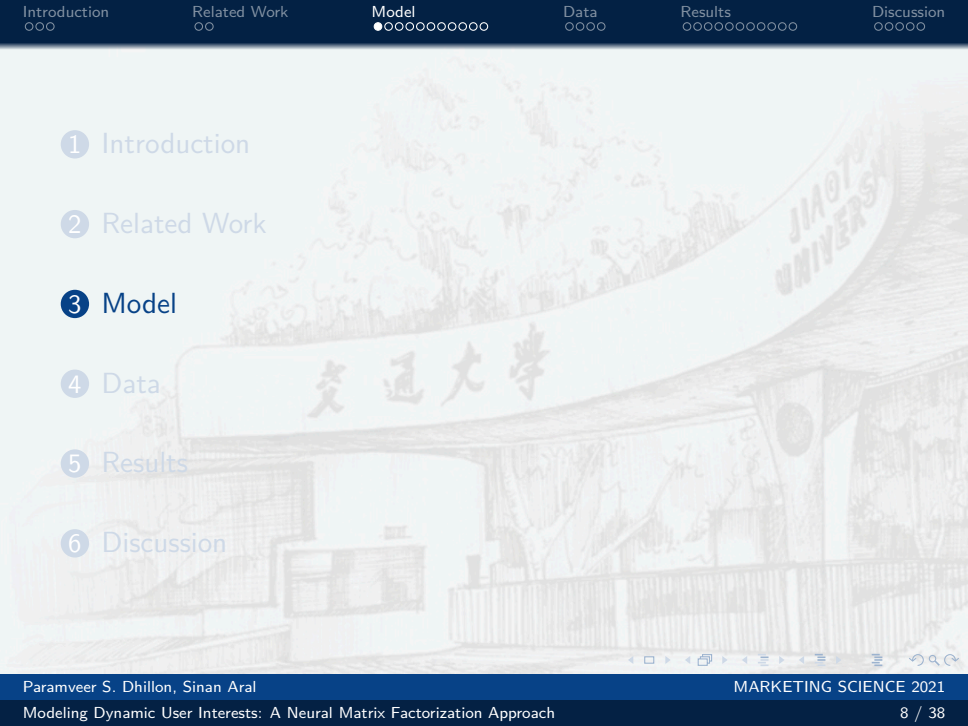
Research Question and Objective

⊙ 本文旨在解决上述三大挑战，提出了一种基于神经网络的矩阵分解模型，用于建模用户的动态兴趣;

⊙ 模型结果具有可解释性，同时还保留了神经网络的灵活性。

Introduction
○○○

Related Work
●○

Model
○○○○○○○○○○○

Data
○○○○

Results
○○○○○○○○○○○

Discussion
○○○○○

1 Introduction

2 Related Work

3 Model

4 Data

5 Results

6 Discussion

Introduction
ooo
Related Work
o●
Model
ooooooooooo
Data
oooo
Results
ooooooooooo
Discussion
ooooo

Related Work

- ⊙ 消费者行为建模
  - 使用 Topic Model 对用户消费的文本数据进行建模，从而生成用户画像 (Trusov et al. 2016).
- ⊙ 消费者偏好演变及其对各种营销变量敏感性
  - 指数平滑模拟品牌偏好的演变 (Guadagni and Little 1983);
  - 高斯过程模拟消费者偏好的动态变化 (Dew et al. 2020);
  - 使用粒子滤波对广告和销售等其他营销变量之间的非线性关系进行建模 (Bruce 2008)
- ⊙ 市场营销领域使用机器学习方法研究客户兴趣 (Netzer et al. 2012, Tirunillai and Tellis 2014, Bĺuschken and Allenby 2016, Liu and Toubia 2018, Timoshenko and Hauser 2019)
- ⊙ 矩阵分解

1 Introduction

2 Related Work

3 Model

4 Data

5 Results

6 Discussion

Introduction
○○○

Related Work
○○

Model
○●○○○○○○○○○○

Data
○○○○

Results
○○○○○○○○○○○

Discussion
○○○○○

Notation

- ⊚ n users, T time slices;

- ⊚ Content consumed by user i in time period t and user i's unique identity:

$$\mathbf{z}_i^t = \left[\, \mathbf{x}_i^t \, ; \, \mathbf{a}_i \, \right]$$

$p \times 1$ one-hot vector $n \times 1$ vector

- ⊚ All content consumed by all users during the whole time slices:

$$\{\mathbf{Z}\}^{t=1:\tau}$$

$n \times (n + p)$

Introduction
000

Related Work
00

Model
00●●●●00000

Data
0000

Results
00000000000

Discussion
00000

Matrix Factorization for Modeling Users' Content Interests I

⊙ Approximate the data matrix using the user and content factors

$$z_{ij}^t \approx {v_j}^\top u_i^t$$

$K \times (n+p)$ item vector $K \times n$ user vector

⊙ Loss function $\mathcal{L}(\cdot)$

$$\left(U^t,\, V \right) = \underset{U^t,\, V}{\operatorname{argmin}} \mathcal{L}\left(Z^t,\, V^\top U^t \right)$$

$$= \left\| Z^t - V^\top U^t \right\|_2^2$$

Matrix Factorization for Modeling Users' Content Interests II

⊙ Probabilistic form

$$p\left(Z^t \mid U^t, V, \sigma^2\right) = \prod_{i=1}^{n} \prod_{j=1}^{p+n} \mathcal{N}\left(z_{ij}^t \mid v_j^\top u_i^t, \sigma^2\right),$$

$$p\left(U \mid \sigma_u^2\right) = \prod_{i=1}^{n} \mathcal{N}\left(u_i^t \mid 0, \sigma_u^2\right),$$

$$p\left(V \mid \sigma_v^2\right) = \prod_{j=1}^{p+n} \mathcal{N}\left(v_j \mid 0, \sigma_v^2\right).$$

Matrix Factorization for Modeling Users' Content Interests III

$$
\left(U^t, V\right) = \operatorname*{argmin}_{U^t, V} \sum_{i=1}^{n} \sum_{j=1}^{p+n} \left\| z_{ij}^t - v_j^\top u_i^t \right\|_2^2 + \lambda_U \sum_{i=1}^{n} \left\| u_i^t \right\|_2^2
$$

$$
\underset{\sigma^2/\sigma_V^2}{\uparrow} + \lambda_V \sum_{j=1}^{p+n} \left\| v_j \right\|_{2'}^2
$$

σ^2/σ_U^2

⊚ 扩展了基础的矩阵分解模型
  • 加入非线性因素;
  • 考虑用户潜在因子的状态依赖关系;
  • 将矩阵分解框架应用于文本建模任务中。

Introduction
ooo

Related Work
oo

Model
ooooo●●oooooo

Data
oooo

Results
ooooooooooooo

Discussion
ooooo

Matrix Factorization for Modeling Users' Content Interests IV

g(·) encodes the neural network parameterized by Θ

$$\left[\{U\}^{t=1:\tau}, V, \Theta\right] = \operatorname*{argmin}_{\{U\}^{t=1:\tau}, V, \Theta} \sum_{i=1}^{n} \sum_{j=1}^{p+n} \sum_{t=1}^{\tau} \left\| z_{ij}^t - \mathrm{g}\left(v_j^\top u_i^t; \Theta\right) \right\|_2^2$$

$$+ \lambda_U \sum_{i=1}^{n} \sum_{t=1}^{\tau} \left\| u_i^t \right\|_2^2 + \lambda_V \sum_{j=1}^{p+n} \|v_j\|_2^2$$

$$u_i^t = f\left(u_i^{t-1}\right)$$

Introduction
○○○

Related Work
○○

Model
○○○○○○●○○○○○

Data
○○○○

Results
○○○○○○○○○○○

Discussion
○○○○○

Neural Network Architecture I

$$
[\{U\}^{t=1:\tau}, V, \Theta] = \underset{\{U\}^{t=1:\tau}, V, \Theta}{\mathrm{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{p+n} \sum_{t=1}^{\tau} \left\| z_{ij}^{t} - \mathrm{g} \left(v_j^{\top} u_i^{t}; \Theta \right) \right\|_2^2
$$

$\mathrm{g}(\cdot)$ encodes the neural network parameterized by Θ

$$
+ \lambda_U \sum_{i=1}^{n} \sum_{t=1}^{\tau} \left\| u_i^{t} \right\|_2^2 + \lambda_V \sum_{j=1}^{p+n} \| v_j \|_2^2
$$

$u_i^{t} = f\left(u_i^{t-1} \right)$

Introduction
○○○

Related Work
○○

Model
○○○○○○○●●●●●●

Data
○○○○

Results
○○○○○○○○○○○

Discussion
○○○○○

Neural Network Architecture II

◎ Embedding the Input Data

$$\{x_i^t, a_i\} \rightarrow \left\{ \underbrace{E_x}_{d \times p} x_i^t, \underbrace{E_a}_{d \times n} a_i \right\}$$

pretrained embeddings   model parameter

◎ Estimating a Nonlinear Hidden State for Each User

$$\ell_i^t = \underset{\text{ReLU}}{\sigma_1} \left(\underbrace{W_\ell}_{d \times 2d} \cdot \left[E_x x_i^t; E_a a_i \right] \right)$$

Introduction
○○○

Related Work
○○

Model
○○○○○○○●●●●●●

Data
○○○○

Results
○○○○○○○○○○○

Discussion
○○○○○

Neural Network Architecture III

⊚ Incorporating Dynamics by Combining a User's Current and Previous Hidden States

$$u_i^t = \sigma_2 \left(W_u \, \ell_i^t + W_r \, u_i^{t-1} \right)$$

softmax

$K \times K$

$K \times d$

$$u_i^t = \alpha \cdot \left[\sigma_2 \left(W_u \ell_i^t + W_r u_i^{t-1} \right) \right] + (1 - \alpha) \cdot u_i^{t-1}$$

smoothing hyperparameter

⊚ Combining the User and Content Factors (Encoder-decoder architecture)

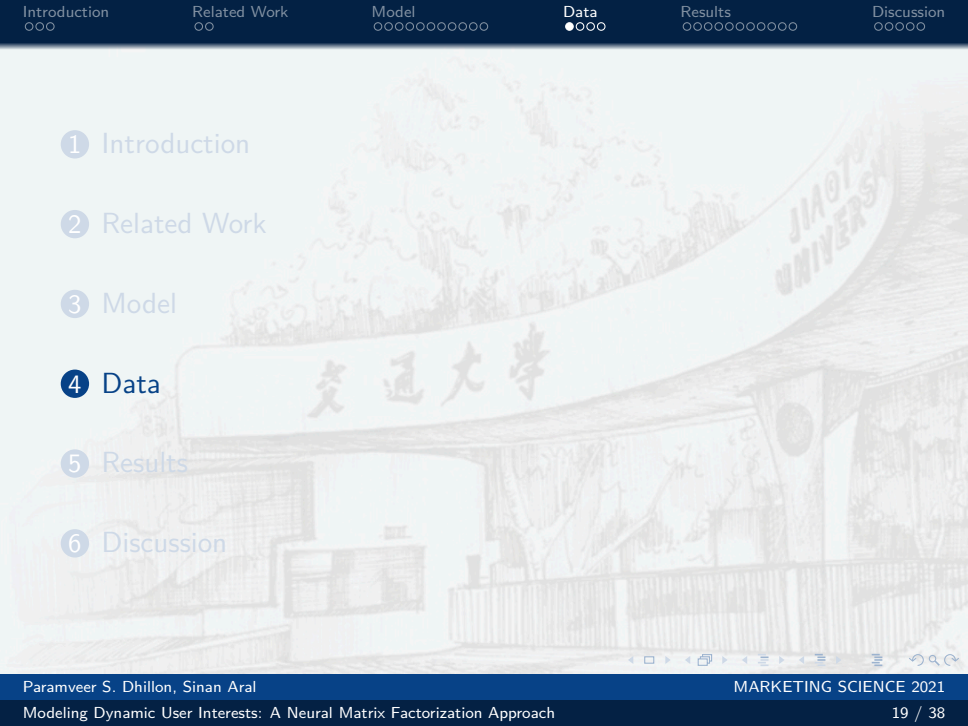$$r_i^t = V^\top u_i^t$$

$K \times d$

Neural Network Architecture IV

⊙ Minimizing the Loss Function

$$\left[\{U\}^{t=1:\tau}, V\right] = \operatorname*{argmin}_{U^{t=1:\tau}, V} \sum_{i=1}^{n} \sum_{t=1}^{\tau} \left\| E_x x_i^t - r_i^t \right\|_2^2$$

⊙ Hyperparameters: $K = 30, \alpha = 0.5, d = 300$, random uniform
  initialization.
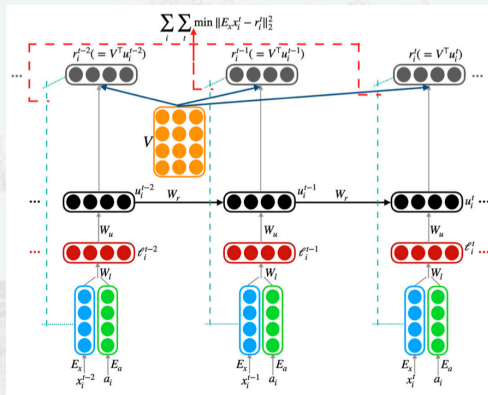
Introduction
ooo

Related Work
oo

Model
oooooooo●●●●●

Data
oooo

Results
oooooooooooo

Discussion
ooooo

Neural Network Architecture V

1 Introduction

2 Related Work

3 Model

4 Data

5 Results

6 Discussion

Data I

- ⊙ 《波士顿环球报》2014 年 2 月 1 日-2019 年 5 月 13 日的用户点击流数据;
- ⊙ 阅读的文章、阅读文章的时长、订阅状态、访问者的人口统计数据: 如地区代码、邮政编码、设备类型 (手机或台式机)、操作系统和国家;
- ⊙ Week level analysis;
- ⊙ Visit the website at least five different times;
- ⊙ 500000 unique visitors over 276 weeks, 5610008 non-zero person-week observations;
- ⊙ Use headlines, 135861569 tokens, 85228 unique tokens;

Data II

Table 1. Summary Statistics of the Visitation and Reading Behavior of the Visitors to the Globe Website

|	Min.	Median	Mean	Max.
Visits per week	1	1	1.64	626
News articles per week	1	1	3.83	1,400
Number of active weeks	5	8	12.40	264

Note. Our data set consists of only those users who were active in at least five different weeks during our observation period.

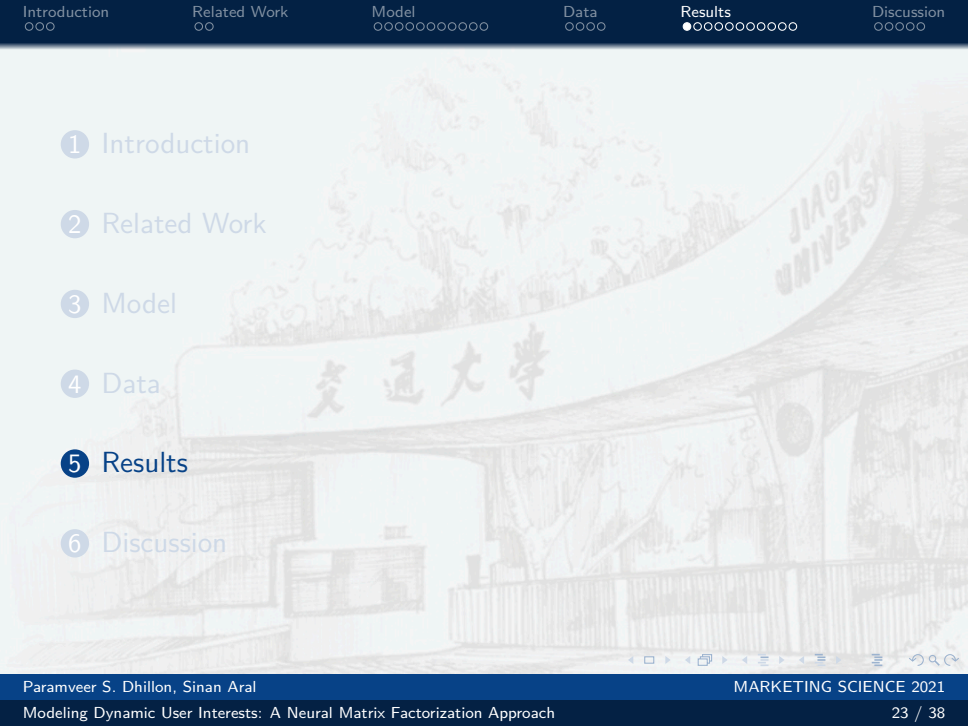**Figure 1.** (Color online) Frequency Distribution of User Activity
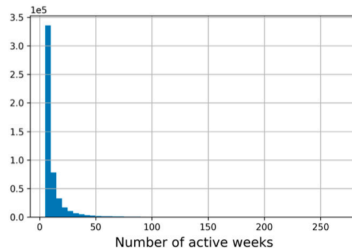
Data III

Figure 2. (Color online) Plot Showing the Prevalence of Words in the News Stories Consumed by Users in Each 52-Week (One Year) Period Starting February 2014
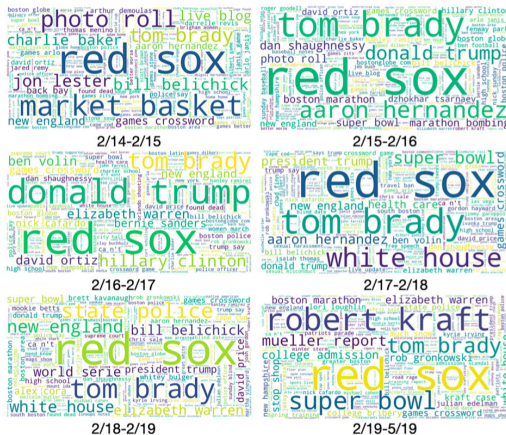
Note. Larger font size indicates the higher prevalence of those terms in users' consumption patterns.
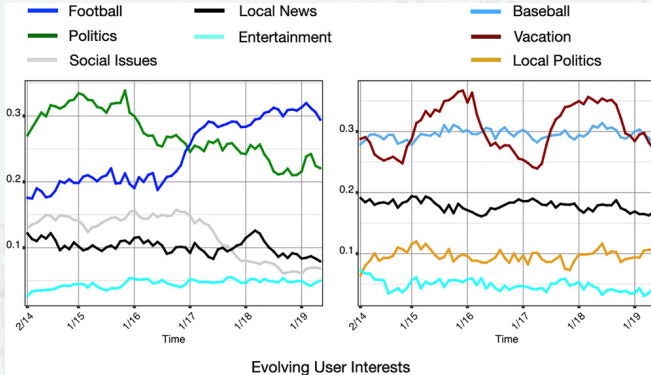
1 Introduction

2 Related Work

3 Model

4 Data

5 Results

6 Discussion

Visualizing Trajectories of User Interests U^t

Compute the nearest neighbors of each row of V from the word embedding matrix E_x.

Introduction
ooo

Related Work
oo

Model
ooooooooooo

Data
oooo

Results
ooooooooooo

Discussion
ooooo

Stable Interests

Various User Personas with Stable Interests

Introduction
○○○

Related Work
○○

Model
○○○○○○○○○○○

Data
○○○○

Results
○○○●○○○○○○○

Discussion
○○○○○

Evolving Interests

Evolving User Interests

Introduction
000

Related Work
00

Model
00000000000

Data
0000

Results
00000●00000

Discussion
00000

Crowdsourced Evaluation of Content Attributes I

⊙ Word intrusion task

⊙ $K = \{10, 30, 50, 100\}$

$$\text{Mean Precision }_k = \sum_{s=1}^{S} \frac{1\left(i_{k,s} = w_k\right)}{S}$$

Introduction
○○○

Related Work
○○

Model
○○○○○○○○○○○

Data
○○○○

Results
○○○○●○○○○○○

Discussion
○○○○○

Crowdsourced Evaluation of Content Attributes II

Figure 7. (Color online) Crowdsourced Mean Model Precision for Different Number of Attributes

Introduction
○○○

Related Work
○○

Model
○○○○○○○○○○○

Data
○○○○

Results
○○○○○○●○○○○○

Discussion
○○○○○
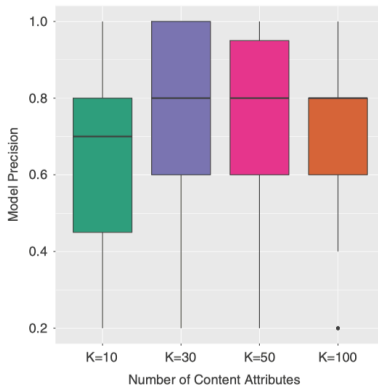
Evaluating the Predictive Quality

⊙ Mean precision at K (MP@K)

$t = \tau - a$ representation $\qquad$ $t = \tau$ representation

$$\text{Mean Precision} = \sum_{i=1}^{n} \frac{1\left(\text{NN}_1\left(r_i^{\tau-a}\right) = c_i^{\tau}\right)}{n}$$

nearest neighbor function

⊙ Real-valued similarity score $s\left(r_i^{\tau-a}, c_i^{\tau}\right)$.

Evaluating the Predictive Quality

Table 2. Results on the Task of Retrieving the Nearest Neighbor, That Is, MP@1

Method	$a = 1$ Mean precision	$a = 2$ Mean precision	$a = 3$ Mean precision
Weighted average of sections	3.8	2.2	1.4
LDA	10.4	7.8	6.4
LDA-GPDH	12.2	10.7	8.7
DTM	14.9	12.6	10.9
Our approach	17.1	15.6	13.2

Notes. (1) Mean precision represents the fraction of users whose nearest neighbor was retrieved correctly. Please refer to Equation (11). (2) Precision numbers are multiplied by 100 to standardize them. (3) Table shows training set accuracy. (4) Model hyperparameters were tuned on the validation data set. The models are estimated on data up until a previous time periods. The prediction is always made on content consumption in the final τ^{th} period.

Table 3. Results Showing Cosine Similarity Between Embeddings of Users and the Content They Consumed

Method	$a = 1$ Similarity ($\mu \pm \sigma$)	$a = 2$ Similarity ($\mu \pm \sigma$)	$a = 3$ Similarity ($\mu \pm \sigma$)
Weighted average of sections	42.9 ± 10.6	40.1 ± 9.4	38.7 ± 9.2
LDA	55.4 ± 5.1	52.9 ± 4.6	50.1 ± 5.4
DTM	64.6 ± 2.1	61.2 ± 2.9	58.6 ± 4.3
LDA-GPDH	62.8 ± 3.0	61.0 ± 2.4	59.9 ± 4.0
Our approach	71.3 ± 3.3	69.4 ± 3.9	67.0 ± 3.6

Notes. (1) Similarity represents the cosine similarity. (2) Similarity numbers are multiplied by 100 to standardize them. (3) Table shows training set accuracy. (4) Model hyperparameters were tuned on the validation data set. The models are estimated on data up until a previous time periods. The prediction is always made on content consumption in the final τ^{th} period.

Robustness Tests

Table 4. Table Showing the Impact of Hyperparameter Choice on the Validation Set Accuracy

Hyperparameters	Mean nearest neighbor precision				
	$\alpha = 0.10$	$\alpha = 0.25$	$\alpha = 0.50$	$\alpha = 0.75$	$\alpha = 0.90$
K = 10	12.9	14.1	15.2	15.0	13.6
K = 30	12.1	15.8	18.4	16.6	14.4
K = 50	11.2	15.3	17.7	15.2	14.1
K = 100	13.4	16.2	17.5	16.8	14.5

Notes. (1) Mean precision represents the fraction of users whose nearest neighbor was predicted correctly. Please refer to Equation (11). (2) Precision numbers are multiplied by 100 to standardize them.
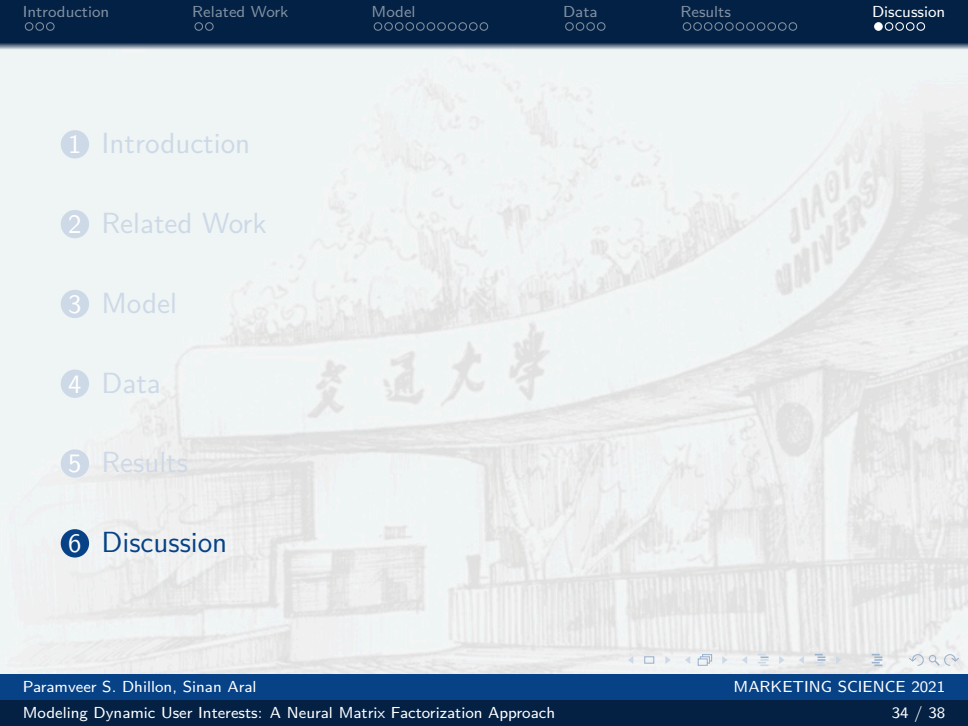
Ablation Analysis

Table 5. Table Showing the Relative Contribution of Various Components of Our Model

	Mean nearest neighbor precision		
Ablation	$a = 1$	$a = 2$	$a = 3$
No nonlinearities	11.7	8.6	6.5
No time dynamics	13.1	10.3	8.1
No exponential smoothing	15.7	12.9	10.8
Full model (Table 2)	17.1	15.6	13.2

Notes. (1) Mean precision represents the fraction of users whose nearest neighbor was predicted correctly. Please refer to Equation (11). (2) Precision numbers are multiplied by 100 to standardize them. Training set accuracy is reported. The models are estimated on data until *a* previous time periods.

Introduction
000

Related Work
00

Model
00000000000

Data
0000

Results
0000000000●

Discussion
00000

Scalability, Transferability, and Cold-Starting New Users.

- ⊙ Use transfer learning to learn representations for users with consumption traces x_i^t.
- ⊙ Freezing all the estimated parameters: $\{W_\ell, W_u, W_r, V, E_x\}$.
- ⊙ Use the observable user demographics.

1 Introduction

2 Related Work

3 Model

4 Data

5 Results

6 Discussion

Managerial Implications

- ⊙ Generating User Profiles
- ⊙ Content Categorization and Recommendation

Introduction
ooo

Related Work
oo

Model
ooooooooooo

Data
oooo

Results
ooooooooooo

Discussion
oo●oo

Conclusion and Limitations

⊙ Proposed a neural matrix factorization modeling approach to extract nonlinear patterns from text data to infer customers' evolving interests.

⊙ Model other types of high-dimensional consumption data(e.g., images, videos, and audio).

⊙ Limitations
  - Model only one kind of customer digital footprints—news consumption
    citeptrusov2016crumbs;
  - Deep learning and neural net models in marketing research is still an under explored area of study;
  - Only model the demand side and assume that consumers' consumption patterns are driven only by their consumption in previous periods.

Introduction
ooo

Related Work
oo

Model
ooooooooooo

Data
oooo

Results
ooooooooooo

Discussion
ooo●o

Code Block Example

```python
class ProdLDA(nn.Module):
    def __init__(self, vocab_size, num_topics, hidden, dropout):
        super().__init__()
```

Thanks!