

# A Joint Learning Framework for Restaurant Survival Prediction and Explanation

陆文胜

## 摘要

互联网的蓬勃发展和最近深度学习技术的突破为电子商务的人工智能打开了一扇新的大门，趋势从使用流动性和盈利能力等少数财务因素演变为使用更先进的人工智能技术来处理复杂和多模态数据。在本文中，我们解决了餐馆生存预测的实际问题。我们认为，传统的方法忽略了两个方面，这对任务非常有帮助:1) 对客户评论建模，2) 联合考虑状态预测和结果解释。因此，我们提出了一种基于用户-餐厅互动和用户文本评论的多模态数据的可解释餐厅生存预测的联合学习框架。此外，我们设计了一个图神经网络来捕获高阶交互，并设计了一个共同注意机制来从嘈杂的文本评论中捕获最有信息量和最有意义的信号。我们在两个数据集上的结果表明，与 SOTA 技术相比，我们有了显著和一致的改进 (平均预测提高了 6.8%，解释提高了 45.3%)。

**关键词:** 可解释推荐; 餐厅生存分析

## 1 引言

企业生存预测是管理和财务文献中的一个热门话题。传统的研究方法在很大程度上依赖于财务因素 (例如: 流动性, 偿付能力和盈利能力)<sup>[1-3]</sup>。然而, 有两个显著的缺点:1) 由于隐私问题, 商店/公司的财务因素很难获得;2) 同时, 财务因素是一种宏观指标, 只能粗略地反映企业的状况。随着信息技术的发展, 许多与餐厅相关的数据可以在网上收集。例如, 人们可以在餐厅消费后签到, 他们可以通过在线评论平台 (如 Yelp.com) 分享评论, 以显示他们如何/为什么喜欢这家餐厅。此外, 评论包含了用户在细粒度层面上的信息反馈。更重要的是, 这些反馈深刻反映了餐厅的经营状况, 进而有助于产生可解释的预测原因。最近的一些研究工作也证实了这一点, 使用在线评论来了解业务绩效是一个新兴趋势<sup>[4-5]</sup>。深度学习的最新进展有各种模型, 用于研究不同类型任务的审查和交互, 例如推荐<sup>[6]</sup>, 假新闻检测<sup>[7-8]</sup>, 评分预测<sup>[9]</sup>, 但很少关注餐厅生存分析的应用。在本文中, 我们提出了一种新的联合学习框架来解决具有挑战性的可解释餐厅生存预测任务。我们的模型由两个必修模块组成: 用于选择有价值评论文本的共同注意网络和用于学习用户-餐厅图上高阶交互的图神经网络。具体来说, 利用共同注意机制来选择有意义的评论文本, 这是一个特征选择和学习的过程。用户-物品交互图可以揭示用户 (或物品) 之间偏好的相似性。因此, 构建我们编码高阶关系的图神经网络, 可以通过对用户和餐厅之间的高阶交互进行建模来增强声誉用户和优质餐厅的表示, 这是我们建模例证的关键。仅仅预测餐厅未来的生存状况是不够的。对于企业来说, 理解他们未来为什么会繁荣或倒闭也是至关重要的。幸运的是, 我们可以利用 NLP 模型对大量用户评论进行编码, 并输出一些解释, 就像文档摘要过程一样。为此, 我们联合训练了生存预测和解释任务。预测任务预测餐厅未来的状态, 解释任务生成一些可解释的文本, 为餐厅的管理提供有信息量的总结。我们将统一框架命名为餐厅生存预测和解释 (RSPE)。通过实验, 我们发现, 与其他几个竞争基准相比, RSPE 显著提高了这两个任务的性能。我们的框架的主要贡献如下:

1) 我们提出了一种新的联合学习框架, 用于预测餐厅的生存, 并通过评论和互动生成总结文本。

2) 我们在 RSPE 中设计了两个关键组件, 即共同关注组件 (用于挖掘高质量和信息丰富的评论) 和图表示组件 (用于编码用户-餐厅图上的高阶交互)。

3) 我们在大众点评和 Yelp 数据集上进行了广泛的实验。我们的模型在预测和解释任务上都显著优于所有 SOTA 方法, 在预测任务上平均提高了 6.8%, 在解释任务上平均提高了 45.3%。

## 2 相关工作

### 2.1 餐厅生存分析

商店生存分析是财务和营销领域的一个重要而实用的研究课题, 它提供了对商店财务、营销策略和管理的深刻见解<sup>r11, r12</sup>[10-11]。传统上, 研究人员通常利用餐厅财务因素建立线性预测模型, 该模型敏感且难以获得。随着在线服务的发展, 研究者发现用户生成内容 (User Generated Content, UGC), 如来自 Yelp.com 或 Dianping.com 的文本评论, 包含了大量的信息, 涵盖了商店 (本文中的餐馆) 的各个方面。利用异构 UGC 可以有效提高餐厅生存预测模型的性能<sup>[12]</sup>。然而, 这组方法的主要弱点有三方面:1) 他们使用了传统的 NLP 模型, 如 LDA、bag-of-words 或 word2vec;2) 没有考虑顾客与餐厅之间的互动图;3) 他们没有明确地减少原始 UGC 中的噪声信息。

### 2.2 预训练模型

预训练模型在自然语言处理领域得到了广泛的应用。它在具有自监督学习任务的大规模开放域数据集上进行训练, 将公共语言知识编码到模型中。训练良好的模型可以使用少量标记数据进行微调, 以在给定的目标任务上表现良好。例如, BERT<sup>[13]</sup>是一个多层双向变压器编码器, 并使用屏蔽语言模型 (MLM) 和下句预测 (NSP) 来捕获单词和句子级表示。UniLM<sup>[14]</sup>是基于 Bert 的, 它在单向、双向和序列到序列预测等 NLP 任务上取得了巨大成功。此外, 一些研究<sup>[15]</sup>表明, 预训练的模型能够捕获层次结构敏感和语法依赖, 这有利于下游的 NLP 任务。

### 2.3 图表征学习

图神经网络 (GNN) 是我们框架中的一个关键组件。GNN 以消息传递的方式融合图上的自信息和邻域信息来表示节点。例如, LightGCN<sup>[16]</sup>通过去除转换层和非线性激活函数, 简化了经典的 GCN<sup>[17]</sup>和 NGCF<sup>[6]</sup>, 并使用平均池聚集器来融合邻居信息。它处理齐次图。异构图神经网络模型 (HetGNN)<sup>[18]</sup>考虑了每个节点的异构结构 (图) 信息以及异构内容信息。一些研究<sup>[19]</sup>已经表明, 高阶相互作用的存在可能会对网络系统的动力学产生重大影响。因此, 我们认为有必要对用户-餐厅图的高阶交互进行编码, 以更好地模拟现有文献忽略的用户偏好和餐厅状态。

## 3 本文方法

### 3.1 本文方法概述

在本节中, 我们将介绍我们的模型, 一个用于餐厅生存预测和解释的联合学习框架, 如图1所示。RSPE 由四个模块组成, 包括输入模块、共同注意模块、图形表示模块和联合学习模块。我们将在以下部分介绍详细信息。

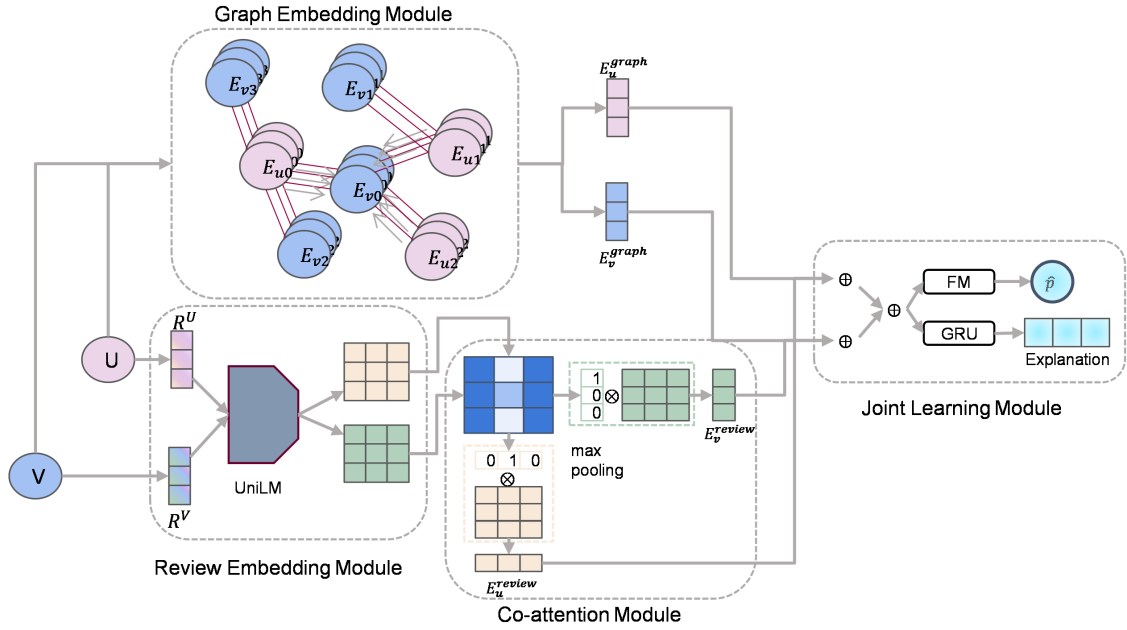


图 1: 方法模型图

### 3.2 输入模块

输入模块的功能是对输入特征进行编码，输入包括两类序列: 餐厅的评论  $(R_{v,1}^{(V)}, \dots, R_{v,l}^{(V)})$ ，相关用户评论  $(R_{u,1}^{(U)}, \dots, R_{u,n}^{(U)})$ ,  $u_i \in U_v$ 。每个序列都包含一个评审列表。这个模块将审查编码为嵌入表示。每篇复习都由一系列的句子组成。我们使用 UniLM<sup>[14]</sup>将每个句子转换为  $d$  维的嵌入表示  $z \in \mathbb{R}^d$ ，因为 UniLM 是在大规模无监督数据集上进行预训练的，通过我们的实验，我们发现它比 BERT 更好。给定一条评论  $R_{V,i}^{(V)}$ ，其嵌入向量  $r_{u,i} = \sum_{z \in \mathbb{R}_{u,i}} z$  用该评论中句子嵌入的平均值表示。此外，我们使用嵌入查找操作，从每个用户和餐厅的 ID 中获得可训练的嵌入向量表示，这将用作图形表示模块的输入。

### 3.3 协注意力模块

直觉很简单，但很有力。每个用户由其写的所有评论表示，餐厅由属于该餐厅的所有评论表示。共同关注模块的目标是从用户/餐厅的评论集合中选择高质量的评论，最后将评论嵌入合并到一个用户/餐厅嵌入中。

#### 3.3.1 亲和矩阵

给定用户评论嵌入  $\mathbf{a}_i (a_i \in \mathbb{R}^{l \times d})$  和餐厅评论嵌入  $\mathbf{b}_j (b_j \in \mathbb{R}^{l \times d})$ ，亲和矩阵计算公式为:

$$M_{i,j} = f(a_i)^T \mathbf{A} f(b_j)$$

其中  $\mathbf{A} \in \mathbb{R}^{d \times d}$  为权重矩阵， $f(\cdot)$  为前馈神经网络。

#### 3.3.2 最大值池化函数

我们使用  $\text{argmax}$  来获得矩阵的每一行和每一列的最大值，然后分别对审查  $a_i$  和  $b_j$  进行加权。计算过程如下:

$$\zeta_i = (\text{Gumbel max}_{\text{col}}(M))^\top \mathbf{a}_i$$

$$\eta_j = (\text{Gumbel max}_{\text{row}}(M))^\top \mathbf{b}_j$$

其中  $\zeta_i$  和  $\eta_j$  表示用户和餐厅的共同注意向量。 $Gumbel()$  是直通 Gumbel softmax<sup>[20]</sup>，由于  $\arg\max$  函数不可微，我们使用  $Gumbel()$  返回一个离散向量，并将非归一化向量  $\mathbf{e} = (e_1, e_2, \dots, e_d)$  转化为概率分布：

$$s_i = \frac{\exp\left(\frac{e_i + g_i}{\tau}\right)}{\sum_{j=1}^d \exp\left(\frac{e_j + g_j}{\tau}\right)},$$

其中  $\tau$  是温度参数， $g_i$  是 Gumbel 噪声。在前馈过程中， $s_i$  将被转换成独热编码  $k_i$ ，我们使用  $Gumbel(s) = k$  去表示：

$$k_i = \begin{cases} 1, & i = \arg \max_j (s_j) \\ 0, & \text{otherwise} \end{cases}$$

### 3.4 图表征模块

餐馆并不是孤立的。有时候，如果只看点评内容，我们无法理解为什么一家餐厅会如此受欢迎。影响一家餐厅经营状况的因素有很多，比如附近的竞争对手、社会潮流等等。为了对餐厅的全球环境进行建模，我们构建了一个二部图，其中节点是餐厅和用户，边缘是用户与餐厅的交互。由于 gnn 在从图结构数据中学习有用信息方面表现出了巨大的优势<sup>[21-22]</sup>，在本节中，我们介绍了一个图表示模块来学习用户-餐厅交互图上有意义的模式，这样餐厅的信息就可以通过他们的社区来增强。它源于这样一个想法，即用户和餐厅之间的特定互动可以揭示餐厅的生存。

#### 3.4.1 节点嵌入

我们通过用户 ID 和餐厅 ID 得到可训练的嵌入向量，分别用  $p_u^{(0)}$  和  $q_v^{(0)}$  表示。

#### 3.4.2 高阶邻居聚合

嵌入在传播层中不同阶的邻居节点对目标节点的影响不同。通过堆叠多个传播层，我们可以探索高级连接信息并增强表示。根据传播规则，我们得到目标节点相邻的一阶、二阶、三阶传播层的邻居节点，传播层嵌入计算如下：

$$\begin{aligned} p_u^{(t+1)} &= \sum_{v \in S_u} \frac{1}{\sqrt{|S_u|}\sqrt{|S_v|}} p_v^{(t)}, \\ p_v^{(t+1)} &= \sum_{u \in S_v} \frac{1}{\sqrt{|S_v|}\sqrt{|S_u|}} p_u^{(t)}, \end{aligned}$$

其中  $p_u^{(t)}$  和  $p_v^{(t)}$  分别表示用户  $u$  和餐厅  $v$  在  $t$  层传播后的嵌入， $S_u$  和  $S_v$  表示用户  $u$  和餐厅  $v$  的第一跳邻居。为了避免较大的嵌入规模，卷积节点的每一层都需要正则化。然后对得到的传播嵌入层进行聚合，得到最终的目标节点嵌入。计算过程如下：

$$p_u = \sum_{t=0}^T \alpha_t p_u^{(t)}, p_v = \sum_{k=0}^T \alpha_k p_v^{(k)}$$

其中  $\alpha_t$  表示第  $t$  个 ( $t = 0, 1, 2, 3$ ) 层嵌入的权值。对于每个餐厅  $v$ ，都会有很多用户评论。因此，我们使用平均池来聚合所有评论过餐厅  $v$  的用户  $u \in S_v$  的向量表示，其表达式为：

$$\mathbf{p}_{S_v} = \sum_{u \in S_v} \mathbf{p}_u$$

### 3.5 联合学习损失函数定义

联合学习是一种利用相关任务训练信号中的域信息作为归纳偏差来提高泛化的归纳迁移方法。由于预测和解释任务是关联的，我们在一个统一的框架中联合训练它们，以获得更好的泛化性能。我们在共同关注模块和图形表示模块中聚合了用户和餐厅的嵌入。公式如下：

$$\mathbf{q}_u = \zeta_i + \mathbf{p}_{S_v}, \mathbf{q}_v = \eta_j + \mathbf{p}_v$$

#### 3.5.1 评分预测任务

因式分解机<sup>[23]</sup>有助于提取最本质的潜在或隐藏特征，从而解决分类问题。公式如下：

$$\mathbf{q}_u = \zeta_i + \mathbf{p}_{S_v}, f(\mathbf{q}) = \mathbf{b} + \sum_{i=1}^n \mathbf{w}_i \mathbf{q}_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{h}_i, \mathbf{h}_j \rangle \mathbf{q}_i \mathbf{q}_j$$

其中  $\mathbf{q}_i \in \mathbb{R}^d$  是  $\mathbf{q} = [\mathbf{q}_u, \mathbf{q}_v]$  的第  $i$  项， $\mathbf{b} \in \mathbb{R}^d$  是偏差， $\mathbf{w}_i \in \mathbb{R}^d$  和  $\mathbf{h}_i \in \mathbb{R}^k$  是要学习的参数。损失函数使用 sigmoid 交叉熵：

$$L_p = \frac{1}{2|\Theta|} \sum_{(u,v) \in \Theta} (-[\mathbf{y} \log(\hat{\mathbf{y}}) + (1 - \mathbf{y}) \log(1 - \hat{\mathbf{y}})])$$

其中  $\mathbf{y}$  为真值标签， $\Theta$  为训练集。

#### 3.5.2 解释生成任务

由于门控循环单元 (gate Recurrent Unit, GRU)<sup>[24]</sup>在生成中表现良好，因此我们选择它来完成解释任务。

GRU 的详细信息如下：首先，计算初始隐藏状态  $\mathbf{h}_0$ ：

$$\mathbf{h}_0 = \tanh(\mathbf{w}^1 \mathbf{q}_u + \mathbf{w}^2 \mathbf{q}_v + \mathbf{w}^3 \hat{\mathbf{y}} + \mathbf{b}_e)$$

其中  $\mathbf{w}^1, \mathbf{w}^2, \mathbf{w}^3$  为要学习的参数。 $\mathbf{b}_e$  是偏置。

当前  $t$  状态与上一个  $t-1$  状态相关：

$$\mathbf{h}_t = GRU(\mathbf{h}_{t-1}, \mathbf{w}_t)$$

其中  $\mathbf{w}_t$  是  $t$  时刻生成的单词。

最后的输出层生成  $t$  时刻隐态词的分布  $d_t$ ：

$$\eta_t = O(\mathbf{w}^4 \mathbf{h}_{t-1} + \mathbf{b}_r)$$

其中  $\mathbf{w}^4$  是要学习的参数， $\mathbf{b}_r \in \mathbb{R}^{|\mathcal{V}| \times l_d}$  是偏置。 $|\mathcal{V}|$  是词汇量大小， $O()$  是 softmax 函数。然后，我们使用波束搜索来选择生成的最佳文本  $\mathbf{Y}$ 。

我们希望最大限度地提高基本事实文本的概率。因此，解释任务的损失函数为：

$$L_g = \frac{1}{|\Theta|} \sum_{(u,v) \in \Theta} \sum_{t=1}^T \left( -\log \eta_{t, \hat{l}_t} \right),$$

其中  $\hat{l}_t$  是  $t$  时刻的真实文本的单词。

### 3.5.3 多任务损失

通过共享相关任务之间的表示，我们将两个任务的三个损失函数聚合起来进行优化：

$$\mathcal{L} = \lambda_1 L_p + \lambda_2 L_g + \lambda_3 \|\Psi\|_2^2$$

其中  $\lambda_\xi (\xi = 1, 2, 3)$  是控制不同损失函数权重的超参数。 $\Psi$  表示可训练参数的集合。

## 4 复现细节

### 4.1 与已有开源代码对比

本人复现的代码参考了原论文中的代码：<https://github.com/Complex-data/RSPE>。基于其项目的结构进行改进和融合，其中包括模型的改进、评价指标的改进和新增、数据集处理。

### 4.2 实验环境搭建

本实验运行在 GPU V100 和 CentOS 7 服务器上。代码使用 Tensorflow1.5、python3.6 实现。对于一个有 10251 个餐厅、448325 条评论的数据集，其训练时间大约为 30 小时。

### 4.3 创新点

1) 将语言预训练模型替换为基于句子相似度预训练任务的 Bert 模型，使用 Bert-whitening 将句子的标准嵌入降低维度。修改后输入模块图2所示。

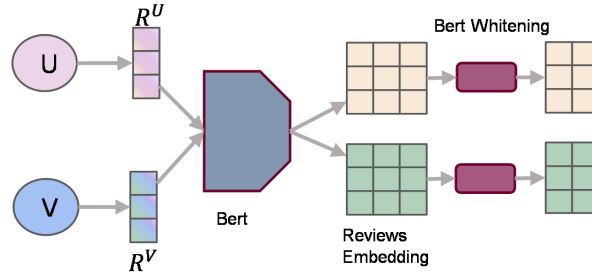


图 2: 输入模块

2) 协注意力 (co-attention) 模块仅仅依靠句子表征间的内积的大小作为权重，输入 Gumble softmax 转换为接近独热码的权重向量。这种方法太过于简单粗暴，因此本人复现版本里将协注意力模块替换为一个基于 attention 的 retrieval 网络作为选择模块。

3) 额外增加 Bert similarity 评价指标，用于评估生成的解释文本与真实解释文本相比的质量。

## 5 实验结果分析

### 5.1 评价指标

在我们的实验中，我们使用 AUC<sup>[25]</sup> 来评估预测任务。BLEU<sup>[26]</sup> 和 ROUGE<sup>[2]</sup> 被用来评价解释任务。ROUGE 的评价基于文本中 n-grams 的共现信息。ROUGE-N (N=1,2) 主要计算 N-grams。ROUGE-L 是通过匹配最长公共子序列来计算的。BLEU 和 ROUGE 值越大说明解释能力越强。

5.2 预测任务结果

预测生成任务的结果如表1所示。

表 1: Results on the prediction task of improved RSPE(RSPE+).

Datasets	Dianping-SH	Dianping-BJ	Dianping-GZ	Yelp-NV	Yelp-AZ	Yelp-ON
AUC	0.8994	0.9073	0.9096	0.9521	0.9171	0.9379

5.3 解释任务结果

解释生成任务的结果如表2所示。

表 2: Results on the explanantion task of improved RSPE(RSPE+).

Metrics-Datasets	BLEU	R1	R2	RL	BS
Dianping-SH	1.866	7.79	2.47	7.49	64.5
Dianping-BJ	2.025	5.43	1.94	5.72	66.4
Dianping-GZ	2.474	10.22	4.66	10.27	68.0
Yelp-NV	2.226	29.57	12.24	23.97	54.4
Yelp-AZ	2.400	31.64	14.81	25.73	55.8
Yelp-ON	1.826	30.50	12.69	23.28	58.2

5.4 案例分析

我们的 RSPE 在中文和英文数据集中的文本中都表现得更好，因为我们利用图卷积神经网络来增强用户-餐厅交互的隐藏协作信号建模，使模型能够学习口碑评论来提高解释文本的质量。这一观察结果与上述结果一致。进一步验证了通过在建模过程中加入图结构，我们的模型可以学习用户与餐厅之间的交互信息，从而为餐厅生存生成信息性的文本表达式。解释生成的个别案例如表3所示。

表 3: Explanations generated by improved RSPE(RSPE+).

Case 1 - Truth	味道不错，可以打 9 折, 位置也好，一鱼两吃上面的面条好吃。另外有没有童鞋可以告诉俺怎样才能上两星呢，谢谢。3333333
RSPE+	味道很好，味道还可以，味道还可以，味道还可以，每次都是吃鱼的味道味道还是不错的，脆脆的。
Case 2 - Truth	餐厅不大但和精致，尤其是景观一流，广州夜景尽收眼底，很棒。再就是服务了，没得说，在这里吃饭真感觉到舒适。
RSPE+	餐厅不大真的很棒。尤其是精致，尤其是精致，但味道都是没得说。
Case 3 - Truth	早茶品种少了些，而且价格比较贵。像鱼片粥都要 18 元一碗，结果鱼片还有土腥味。
RSPE+	早茶品种少了些，味道还不错。鱼片粥一般般，其他的品种一般般啦。

6 总结与展望

在这篇论文中，我们解决了餐馆生存的问题，这是社会公益的一个基本任务。与高度依赖敏感财务指标的传统方法不同，我们使用深度学习技术从大量 UGC 中挖掘有用的信号。我们是第一个同时进行未来状态预测和解释的联合框架。我们的模型有两个关键组件，即图形表示模块和共同注意模块。我们在两个数据集上进行了广泛的实验。结果表明，我们提出的模型在预测和解释任务上都达到了 SOTA 性能。

参考文献

[1] ZIMAN J M. Reliable knowledge: An exploration of the grounds for belief in science[M]. Cambridge University Press, 1991.

[2] LIN C Y. Rouge: A package for automatic evaluation of summaries[C]//Text summarization branches out. 2004: 74-81.

- [3] PEREIRA J M, RIBEIRO H, SILVA A, et al. To fail or not to fail: an algorithm for SME survival prediction using accounting data[J]. *The Changing Role of SMEs in Global Business: Volume II: Contextual Evolution Across Markets, Disciplines and Sectors*, 2020: 83-107.
- [4] BABIĆ ROSARIO A, SOTGIU F, DE VALCK K, et al. The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors[J]. *Journal of marketing research*, 2016, 53(3): 297-318.
- [5] KONG G, UNGER J, BAEZCONDE-GARBANATI L, et al. The associations between Yelp online reviews and vape shops closing or remaining open one year later[J]. *Tobacco prevention & cessation*, 2017, 2(Suppl).
- [6] WANG X, HE X, WANG M, et al. Neural graph collaborative filtering[C]// *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*. 2019: 165-174.
- [7] POTTHAST M, KIESEL J, REINARTZ K, et al. A stylometric inquiry into hyperpartisan and fake news [J]. *arXiv preprint arXiv:1702.05638*, 2017.
- [8] WANG W Y. "liar, liar pants on fire": A new benchmark dataset for fake news detection[J]. *arXiv preprint arXiv:1705.00648*, 2017.
- [9] TAY Y, LUU A T, HUI S C. Multi-pointer co-attention networks for recommendation[C]// *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 2018: 2309-2318.
- [10] PARSA H, SELF J T, NJITE D, et al. Why restaurants fail[J]. *Cornell Hotel and Restaurant Administration Quarterly*, 2005, 46(3): 304-322.
- [11] DU JARDIN P. Dynamics of firm financial evolution and bankruptcy prediction[J]. *Expert Systems with Applications*, 2017, 75: 25-43.
- [12] LIAN J, ZHANG F, XIE X, et al. Restaurant survival analysis with heterogeneous information[C]// *Proceedings of the 26th International Conference on World Wide Web Companion*. 2017: 993-1002.
- [13] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. *arXiv preprint arXiv:1810.04805*, 2018.
- [14] DONG L, YANG N, WANG W, et al. Unified language model pre-training for natural language understanding and generation[J]. *Advances in neural information processing systems*, 2019, 32.
- [15] LI K, FENG Y, GAO Y, et al. Hierarchical graph attention networks for semi-supervised node classification[J]. *Applied Intelligence*, 2020, 50: 3441-3451.
- [16] HE X, DENG K, WANG X, et al. Lightgcn: Simplifying and powering graph convolution network for recommendation[C]// *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*. 2020: 639-648.



- [17] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint arXiv:1609.02907, 2016.
- [18] ZHANG C, SONG D, HUANG C, et al. Heterogeneous graph neural network[C]//Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019: 793-803.
- [19] BATTISTON F, AMICO E, BARRAT A, et al. The physics of higher-order interactions in complex systems[J]. Nature Physics, 2021, 17(10): 1093-1098.
- [20] JANG E, GU S, POOLE B. Categorical reparameterization with gumbel-softmax[J]. arXiv preprint arXiv:1611.01144, 2016.
- [21] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. arXiv preprint arXiv:1710.10903, 2017.
- [22] HAMILTON W L. Graph representation learning[J]. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2020, 14(3): 1-159.
- [23] RENDLE S. Factorization machines[C]//2010 IEEE International conference on data mining. 2010: 995-1000.
- [24] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014.
- [25] HANLEY J A, MCNEIL B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve.[J]. Radiology, 1982, 143(1): 29-36.
- [26] PAPINENI K, ROUKOS S, WARD T, et al. Bleu: a method for automatic evaluation of machine translation[C]//Proceedings of the 40th annual meeting of the Association for Computational Linguistics. 2002: 311-318.