



中國人民大學  
RENMIN UNIVERSITY OF CHINA

# 硕士学位论文

THESIS OF MASTER DEGREE

论文题目：在线患者社区用户同伴支持及其对持续参与影响研究

(英文)：Research on User's Peer Support and Its Impact on  
Continuous Participation in Online Patient Community

作者：王配配

指导教师：左美云

2020 年 6 月 5 日

# 中国人民大学

## 硕士学位论文

在线患者社区用户同伴支持及其对持续参与影响

(中文题目)

研究

---

Research on User's Peer Support and Its Impact on

(英文题目)

Continuous Participation in Online Patient Community

---

作者学号： 2017100962

---

作者姓名： 王配配

---

所在学院： 信息学院

---

专业名称： 管理科学与工程

---

导师姓名： 左美云

---

论文主题词： 在线患者社区；同伴支持；文本分  
(3-5个) 类；持续参与

---

论文提交日期： 5/6-2020

---

# 独创性声明

本人郑重声明：所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得中国人民大学或其他教育机构的学位或证书所使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

论文作者（签名）：王配配 日期：27/5-2020

## 关于论文使用授权的说明

本人完全了解中国人民大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

论文作者（签名）：王配配 日期：27/5-2020

指导教师（签名）：王美云 日期：8/5-2020

## 摘要

随着 Web2.0 的快速发展、公民健康观念的转变和健康意识的增强，在线健康社区开始广泛被公众所认可接受。在线健康社区作为线下医疗资源的补充，为实体医疗体系注入了新活力，在一定程度上缓解了我国医疗资源发展不平衡不充分的现状，促进了优质医疗资源的下沉。在线健康社区的长久繁荣是实现线上线下医疗资源的协调发展，为患者提供更优质医疗服务重要保障。但目前在线健康社区普遍面临着用户流失严重的问题，阻碍了在线健康社区的发展。因此，理解用户在社区内的参与行为以及影响其持续参与的因素，是学术界需要关注的问题。

本文以患者为主要用户的在线健康社区，即在线患者社区，为研究对象，其内部聚集了大量具有相似疾病经历的患者，这些患者可以通过互动交流来为彼此提供支持，这种支持也被称为同伴支持。本文从同伴支持的视角出发，研究了用户同伴支持行为类别的自动识别模型，以及同伴支持对用户持续参与的影响效果。

首先，在同伴支持分类研究中，本文为每类同伴支持分别构建分类模型，将多标签分类转换为二元分类，应用传统机器学习算法和深度学习算法实现文本分类，并将 doc2vec 模型和 LDA (Latent Dirichlet Allocation, 潜在狄利克雷分布) 模型训练出的文本特征向量和人工提取出的特征向量相拼接作为传统机器学习算法的输入，doc2vec 模型训练出的文本特征向量作为深度学习算法的输入。通过对比不同模型分类效果，研究发现深度学习算法 CNN (Convolutional Neural Networks) 在各类同伴支持分类任务中均有较好的表现。

其次，在同伴支持对用户持续参与的影响研究中，本文主要探讨了同伴支持行为对持续参与的作用，以及同伴支持类型在同伴支持行为和持续参与关系中所起的作用。研究基于同伴支持分类研究所建立的分类模型构建研究变量，并利用 Cox 比例风险回归模型评估自变量对用户参与的影响程度。研究证实了寻求同伴支持、提供同伴支持和接收同伴支持对持续参与的促进作用，还发现了在不同的同伴支持行为中不同同伴支持类型对持续参与的作用会有所不同。

本文不仅构建出了中文在线患者社区环境下的同伴支持分类模型，还从同伴支持角度探讨了持续参与的影响因素。理论上，本文的研究丰富了文本分类、在线医疗和持续参与相关的研究，研究结果为今后的相关研究提供了一定的理论基础和方法指导。实践上，本文的研究结果为在线患者社区管理者提高用户持续参与水平提供了策略参考，帮助其减少用户的流失，提高社区活跃度，促进社区的健康发展，使用户能够从社区中获得更多的益处。

**关键词：** 在线患者社区； 同伴支持； 文本分类； 持续参与



## Abstract

With the rapid development of web2.0, the change of citizens' health related attitude and enhancement of health awareness, the online health community started to be widely accepted by the public. As a supplement to offline medical resources, online health communities has injected fresh vitality into the physical medical system. To some extent, it alleviates the unbalanced and inadequate development of medical resources in China, and promotes the sinking of high-quality medical resources. The long-term prosperity of the online health communities is an important guarantee for achieving the coordinated development of online and offline medical resources and providing patients with better medical services. However, the online health communities are generally facing serious problems of user churn, which hinders the development of online health communities. Therefore, understanding the behaviour of users in community and factors that affect users continued participation are issues that academics should focus on. Although there has been related studies, it is still insufficient.

This article focuses on online patient community, which is a type of online health community and has a large number of patients with similar disease experiences. These patients can support each other through interactions and communications, which is called peer support. From the perspective of peer support, this paper studies the automatic classification model of peer support and the effect of peer support on user continuous participation.

Firstly, in the research of peer support classification, this paper separately builds a classification model for each type of peer support, which converts multi-label classification to binary classification. Moreover, this research applies traditional machine learning algorithms and deep learning algorithms to implement text classification. The text feature vector trained by doc2vec model and LDA (latent Dirichlet allocation) model is spliced with the artificial feature vector as the input of traditional machine learning algorithm, and the text feature vector trained by doc2vec model is the input of deep learning algorithm.. By comparing the classification performance of different models, the research finds that the deep learning algorithm CNN (Convolutional Neural Networks) performs well in different peer support classification tasks.

Secondly, in the study of the impact of peer support on user continuous participation, this article mainly examines the role of peer support behaviours on

continuous participation, and the effect of peer support types on the relationship between peer support behaviour and continuous participation. Research constructs variables based on the classification model established by the peer support classification research, and utilizes the Cox proportional hazard regression model to evaluate the independent variables' effect on user continued participation. Research confirms the positive role of seeking peer support, providing peer support, and receiving peer support on continuous participation, and also finds that different types of peer support have different effects on the relationship between peer support behaviours and continuous participation.

This article not only builds a peer support classification model in the context of Chinese online patient communities, but also explores the influencing factors of continuous participation from the perspective of peer support. Theoretically, this research enriches the related research on text classification, online healthcare and continuous participation. The research results provide a certain theoretical basis and method guidance for future related research. In practice, the findings provide strategic references for online patient community managers to improve the level of continuous participation, reduce user churn, increase community activity, promote the healthy development of the community, and enable users to gain more benefits from the community.

**Keywords:** Online Patient Communities, Peer Support, Text Classification, Continued Participation

## 目录

|   |           |
|---|-----------|
| 摘要.....                                   | I         |
| Abstract.....                             | III       |
| <b>第 1 章 绪论</b> .....                     | <b>1</b>  |
| 1.1 研究背景.....                             | 1         |
| 1.2 研究意义.....                             | 3         |
| 1.3 研究框架及创新点.....                         | 4         |
| 1.4 技术路线.....                             | 6         |
| <b>第 2 章 文献综述</b> .....                   | <b>7</b>  |
| 2.1 在线患者社区相关研究.....                       | 7         |
| 2.2 同伴支持相关研究.....                         | 9         |
| 2.3 文本分类的相关技术.....                        | 13        |
| 2.4 在线患者社区用户持续参与研究.....                   | 25        |
| <b>第 3 章 在线患者社区同伴支持分类研究</b> .....         | <b>30</b> |
| 3.1 分类框架.....                             | 30        |
| 3.2 研究设计.....                             | 32        |
| 3.3 研究结果.....                             | 38        |
| 3.4 本章小结.....                             | 40        |
| <b>第 4 章 在线患者社区用户同伴支持对持续参与的影响分析</b> ..... | <b>41</b> |
| 4.1 研究模型和假设.....                          | 41        |
| 4.2 研究设计.....                             | 45        |
| 4.3 研究结果.....                             | 49        |
| 4.4 讨论.....                               | 54        |
| 4.5 本章小结.....                             | 56        |
| <b>第 5 章 结论</b> .....                     | <b>57</b> |
| 参考文献.....                                 | 60        |



|         |    |
|---------|----|
| 附录..... | 68 |
| 致谢..... | 72 |

# 第 1 章 绪论

本章将首先介绍本文的研究背景，从而引出本文的研究问题—在线患者社区用户同伴支持及其对持续参与的影响研究。接下来，将从理论意义和实践意义两个方面论述该研究的研究意义，介绍本文的研究框架及创新点。最后将展示本文的研究路线图。

## 1.1 研究背景

随着互联网和信息技术的发展，国家健康促进行动的推进，以及国民健康意识和素养的普遍提升，互联网已成为人们获取信息的重要渠道。2018 年 12 月 15 日，中国卫生信息与健康医疗大数据学会家庭健康专委会、中国家庭报社、微医集团联合发布了《中国家庭健康大数据报告(2018)》，报告显示，互联网在家庭健康中扮演着重要的角色，老年人也开始更多地寻求在线就医。在发现自己或家人身体不适，如果是非紧急重症情况，有 71.5%的人选择“先去网上查一下症状”，有 30%的人选择先在网上咨询医生。

人们在遇到健康问题，通常会先通过网络查找相关疾病或医疗信息，然后去实体医院“看病”。但由于医生工作负担繁重，难以为不同患者提供详尽的个性化咨询、诊断和其他健康服务，难免会出现“排队三小时，看病三分钟”的窘境，有时医生也未能提供有效的评估和治疗，且医生预约困难，周期较长，部分患者面临着财务压力，家人和朋友所能提供的线下支持有限<sup>[1]</sup>。因此，人们在就诊后仍然会通过互联网寻求进一步的健康解决途径，建立除医院就医之外的其他信息渠道，广泛了解病情和治疗等相关信息。

在此背景下，随着社交媒体的蓬勃发展，在线健康社区应运而生，它可以很好地实现“互联网”+健康教育的目的。在线健康社区的兴起为患者提供了一个获取医疗资源并分享知识、经验和情感的开放式平台，使患者能够通过积极参与在线讨论和互动获取健康知识和支持，在疾病治疗中发挥重要作用。在线健康社区按照互动主体的不同，可以分为三种：以医生和患者为主体的在线医患社区，例如“好大夫在线”；以患者为主体的在线患者社区，例如“PatientsLikeMe”和“甜蜜家园”；以医生为主体的在线医生社区，例如“丁香园”。本文主要关注以患者为主体的在线患者社区，它拥有许多潜在的吸引力：“坦率（对问题的反应不那么苛刻和直接），负面判断减少，回报支持的义务减少，关系依赖性减少，更多即时获得支持的能力，更强大的网络专业知识，污名管理，亲密关系，

访问权限，更具表现力的沟通和匿名性”<sup>[2]</sup>，这些优点吸引了大量的病患聚集到在线患者平台。

在线患者社区中，用户可以通过与有相似健康问题的同伴进行在线互动，以寻求、接受和提供同伴支持，充分发挥群体智慧的优势。在集体智慧的作用下，患者群体可能显现出比单个患者，甚至专业医疗工作者更大的智慧，更好地满足患者对健康知识的需求。在线患者社区中同伴支持的益处是多方面的。首先，同伴支持能给患者带来好处：他们通过分享情感、阅读他人的经历、建立社交关系以及获取与疾病管理和治疗相关的信息和技能的过程感受到赋能<sup>[3]</sup>。通过同伴间疾病自我管理经验和教训的分享，可以增强彼此信心<sup>[4]</sup>，使患者恢复并获得正常感。其次，同伴支持对在线患者社区也有益处：对于在线患者社区，其繁荣的本质是社区用户参与，用户参与的数量和频次将直接影响平台的活跃程度<sup>[5]</sup>，只有用户持续参与同伴支持，在社区内主动寻求和提供支持，才能保证在线患者社区长久而有活力地发展下去。

同伴支持是护理专业领域的一个重要概念。研究已经证实了同伴支持在健康领域的作用，目前在慢性病管理方面也得到了广泛应用。世界卫生组织在 2007 年提出“糖尿病同伴支持计划”（peer support programmes in diabetes），我国多个地区在近年来也逐渐开展同伴支持教育。例如，香港利用网络平台，经由专业医生培训的糖尿病领袖（同伴支持者），通过自动电话教育系统丰富患者知识；安徽让生活在同一社区的糖尿病患者组成同伴支持小组，定期聚会，这显著改善了患者心理状况，正反馈作用于患者的健康和行为。可以看出，同伴支持在慢性病防控和康复过程中扮演着重要角色。慢性病患者往往需要长期不断地努力，控制病情。因此，他们需要同伴支持，特别是从“过来人”的经验中获得生活上所需的帮助，这是慢性病患者长期维持健康的关键。在线患者社区中的同伴支持不仅为慢性病患者提供了疾病和治疗的相关知识，也为其长期防控提供了精神上的助力。

虽然在线健康社区已成为当代很多病患的选择之一，但仍有很多社区面临着用户流失严重或社区活跃度低的问题，严重制约了社区的发展。在社区用户不断外流，社区活跃度日益降低的情况下，新用户以及留存用户的信息收益和心理收益也将受到一定损失。因此，了解用户持续参与在线健康社区的相关因素无论对社区管理者还是社区成员都有重要意义。现有研究多是从主观的认知角度探讨影响用户持续参与在线社区的因素，本研究则尝试从客观的互动角度出发，以慢性病在线患者社区为数据来源，基于当前在线患者社区和同伴支持的研究，在同伴支持和持续参与等理论的驱动下，实证分析不同支持行为以及不同类型的同伴支持对用户持续参与在线患者社区的影响。由于要实现对同伴支持进行分类，考虑到数据量较大，人工分类会消耗大量的时间，所以本研究还将探讨实现同伴支持自

动分类的方法。研究成果将为学术界和医疗保健行业深入理解用户的在线同伴支持行为、同伴支持对用户持续使用的影响以及在线患者社区推荐系统的设计提供一定的借鉴价值。

## 1.2 研究意义

在线患者社区的本质是为病患用户提供疾病相关的沟通和交流平台，而其用户的特殊性——患有相似的疾病，决定了平台内部同伴支持行为的普遍性和重要性，识别不同类型的同伴支持行为以及分析不同类型同伴支持行为对平台持续参与可能产生的作用，对于平台和患者都有重要意义。本文的研究意义主要体现在理论意义和实践意义两个方面。

### 1.2.1 理论意义

1、预期得到的同伴支持分类模型，将有助于丰富同伴支持文本分类的研究。虽然目前文本分类算法是研究热点，但具体到同伴支持分类的情境，关注者却是比较少。此外，以往的部分研究采用人工标注的手工分类方法对同伴支持进行分类<sup>[1][6]</sup>，但随着用户产生数据的激增，手工分类已经不能满足高效分类的需求，需辅之以新的技术，实现分类的自主化，提升分类的效率。如何将成熟的文本分类算法应用到同伴支持文本分类情境下，并取得有效的分类效果，是学术界有待深入的话题。因此，本文以成熟的分类算法为基础，结合在线健康社区同伴支持情境的特殊性，实现对同伴支持进行分类，有助于丰富文本分类在同伴支持情境下的应用研究。

2、预期成果将丰富了持续参与理论的相关内容，为持续参与研究提供了新的视角。用户的持续参与是保证在线健康社区繁荣发展的基础，因此对于在线健康社区用户持续参与的研究也是学术界的重点关注。学者们对持续参与的研究多以持续参与意愿作为研究对象，采用问卷调查或访谈的方式获取数据，以验证模型。且研究的视角多是从期望确认理论和技术接受模型等出发，鲜有探讨同伴支持与用户持续参与之间的关系。本文从同伴支持的视角出发，以在线健康社区实证数据为依据，探讨同伴支持与用户持续使用之间的关联性，丰富了对在线平台持续参与的研究，为持续参与理论拓展了新的视角。

### 1.2.2 实践意义

1、对于在线患者社区的管理者，本文的研究结果可以为在线患者社区的管理者提供借鉴。首先，同伴支持分类算法的实现和准确率的提高可以为在线患者社

区的管理者提供了推荐的新角度和算法基础。用户可能会对不同类型的支持感兴趣。所以同伴支持分类问题的研究可以更精准地为用户推荐准确的感兴趣内容。此外,根据同伴支持分类结果,还可实现不同类型同伴支持用户之间的有效匹配,助力用户问题得到更好的解决。其次,不同类型同伴支持对用户持续使用的作用结果可以帮助在线患者社区管理者设计有针对性的引导机制,通过对同伴支持行为的调节,间接实现对用户持续使用行为的刺激。

2、对于在线患者社区的用户,本文的研究对其有重要的应用意义。同伴支持行为分类的实现有助于用户更好地获得同伴支持,以及更有效地提供同伴支持,提高在线患者社区在用户自我健康管理过程中的融入度,为用户提供除专业医生、亲朋好友之外的有效支持,同时,通过用户对在线患者社区的持续参与以增加其通过在线患者社区改善健康的可能性。

### 1.3 研究框架及创新点

本文的研究内容涉及两部分:一是实现对在线患者社区同伴支持的自动分类,二是研究同伴支持行为对在线患者社区持续参与的影响。总体的研究框架图如图 1-1 所示。

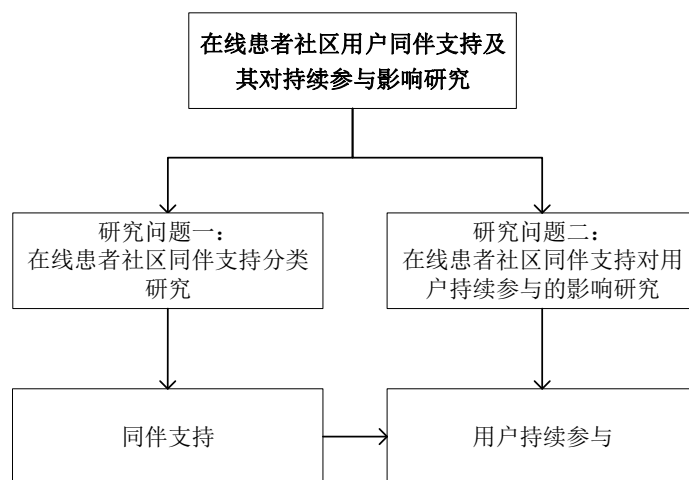


图 1-1 在线患者社区同伴支持行为研究框架

#### 研究一：在线患者社区同伴支持分类研究

研究一是对在线患者社区同伴支持行为的分类研究。首先,从糖尿病在线患者社区—甜蜜家园中采集用户发帖信息,对数据进行预处理,采用不同的文本特征表示方法生成文本特征向量;随后,将多标签分类转化为二元分类,为每个类型的同伴支持分别构建分类器,对比传统的机器学习分类算法和深度学习算法,得到每个类型同伴支持的最优分类器。

## 研究二：在线患者社区同伴支持对用户持续参与的影响分析

根据研究一同伴支持分类所构建出的最优分类器，对用户每个发帖信息所属的同伴支持类别进行分类，以构建研究模型中的相关变量。采用生存分析方法——Cox 比例风险回归模型探讨不同同伴支持行为（寻求同伴支持、提供同伴支持和接收同伴支持）对用户持续参与的作用，以及不同支持类型（信息支持、情感支持、评价支持、陪伴支持）在同伴支持行为和持续参与关系之间的作用。

与以往研究相比，本文研究的创新之处在于：

（1）本文构建了中文在线患者社区环境下的同伴支持自动分类模型。以往对于同伴支持相关类型的自动分类研究，多采用传统机器学习算法。本文的研究不仅探讨了利用传统的机器学习分类算法为不同类型的同伴支持构造分类器，还对比了采用深度学习算法在进行同伴支持分类时的效果，对同伴支持分类相关文献做出了重要补充。此外，文章还采用 doc2vec 模型和 LDA 模型训练出的文本特征向量和人工提取出的特征向量相拼接的方式构建文本特征向量，以提高传统机器学习分类算法的同伴支持分类效果，为未来的文本特征向量表示提供了新的思路。

（2）本文从同伴支持的角度，基于客观互动数据探讨了影响在线患者社区用户持续参与的因素。之前有关用户持续参与影响因素的研究多是从主观认知角度出发，缺乏从客观互动角度出发的相关探讨。且对于不同支持类型对用户持续参与的作用，以往研究往往仅关注用户接收到的支持，对于用户寻求、提供和接收支持不同行为下的支持类型与持续参与之间的关系缺乏关注。本文以在线患者社区内用户的客观互动数据为基础，不仅研究了用户不同同伴支持行为与持续参与之间的关系，还探讨了在不同同伴支持行为下，不同支持类型对持续参与的作用，丰富了对在线平台持续参与的研究，为持续参与理论拓展了新的视角。

## 1.4 技术路线

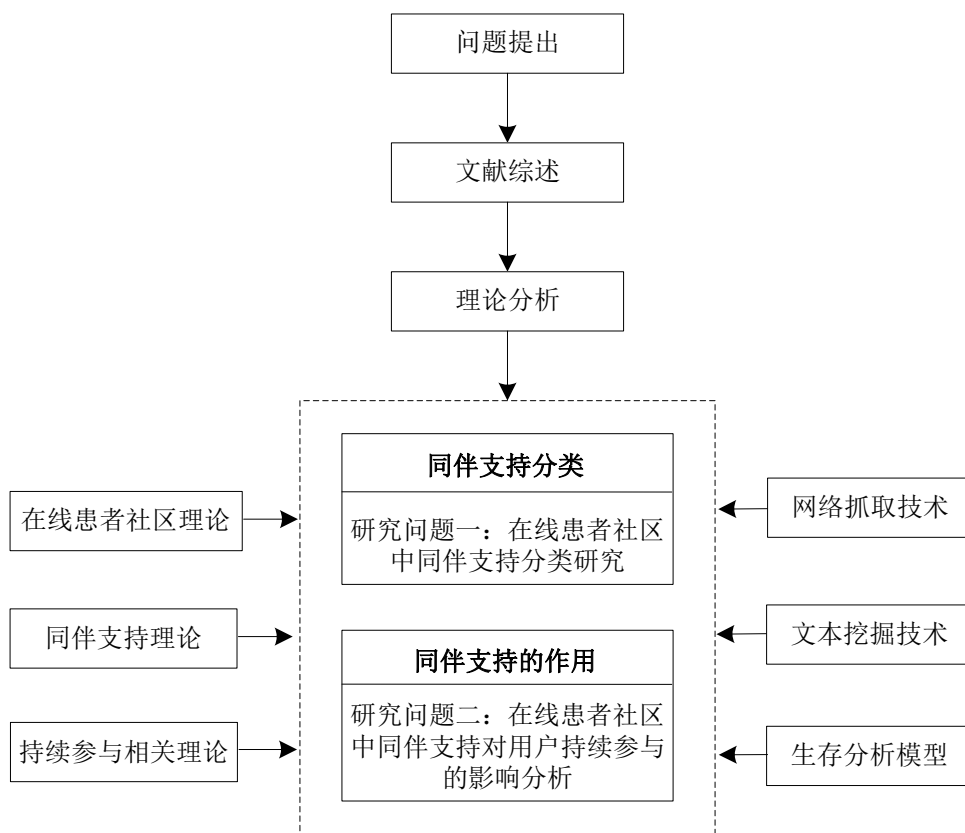


图 1-2 技术路线图

本文的技术路线图如图 1-2 所示。研究中主要采取的技术方法是对数据进行采集的网络抓取技术，对同伴支持进行分类的文本挖掘技术，以及在对同伴支持对持续参与影响研究中采用的生存分析方法。研究的理论基础主要是在线患者社区相关理论、同伴支持理论和持续参与相关理论。

## 第 2 章 文献综述

本章将对相关研究文献进行综述。章节安排如下，首先对在线患者社区相关研究进行总结。其次，对同伴支持相关文献进行回顾。接下来，将对文本分类的相关技术进行介绍。最后，对在线患者社区用户持续参与的相关研究进行分析总结。

### 2.1 在线患者社区相关研究

在线患者社区是以患者为中心的在线健康社区，为用户提供了一个可以与具有相似健康问题的病患进行交流和沟通的便捷渠道。目前的在线患者社区集成了社交媒体和健康管理的功能，用户既可以进行社交互动，例如加好友、聊天、发帖等，又可以通过在线患者社区提供的健康管理工具进行自我健康管理，例如血糖、血压等管理工具。在线患者社区按照疾病覆盖范围可以分为专一型和多元型，专一型的在线患者社区只针对特定的疾病，例如，甜蜜家园只针对糖尿病，是糖友的专有论坛；多元型的在线患者社区疾病覆盖范围广，针对多种疾病，一般会按照疾病类型形成多个同伴支持小组，例如 PatientsLikeMe 包含了多个不同疾病的在线支持小组，用户在注册后可以选择加入相应疾病的组内。在线患者社区按照有的在线患者社区拥有其专用网站，例如，甜蜜家园等；而有的则是依靠社交媒体发展起来的，例如 Facebook、百度贴吧、qq 群等。

如图 2-1 所示，与面对面的线下交流相比，在线患者社区具有四个方面独特的优势：

(1) 摆脱了时间和空间的限制。这一特质使得不同地域的人能够在互联网上实现异步通信，为时间和空间上局限的个体提供了方便<sup>[7]</sup>，支持群体的扩大也为患者提供了更多疾病管理的可借鉴经验和治愈的可能性；

(2) 提供了匿名性。匿名性一方面消除了因疾病的特殊性可能带来的歧视，另一方面，也保护了用户的隐私，可以促进用户对疾病经历和健康管理经验的分享，降低其对于丢面子的担忧<sup>[8]</sup>，更放心地寻求帮助；

(3) 基于显式文本。显式文本可以更清楚地表达社区内用户的意图，促进用户间的交流；

(4) 异步通信。在线患者社区中，由于计算机会话的异步性质，人们也可以花费足够的时间来思考语句内容的选择<sup>[9]</sup>，从而提供更周到的支持。



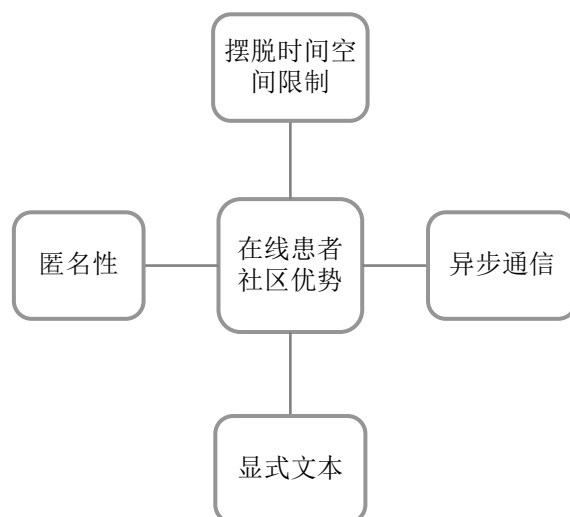


图 2-1 在线患者社区优势

在线患者社区可以被视为一个复杂的系统，其中包含用户、信息、社区三个重要的要素<sup>[10]</sup>，三者之间通过相互作用、相互依存形成稳定的交互系统，如图 2-2 所示。在该系统内，用户是行为主体，主要包含患者、家属、医疗工作人员等，社区为用户产生、传播和使用不同类型的信息提供了环境，这些信息是用户行为的一种记录，反映了用户的需求、认知和情感。同时，用户所产生的大量信息在社区内的积累和聚合，也为信息服务和知识发现提供了基础。

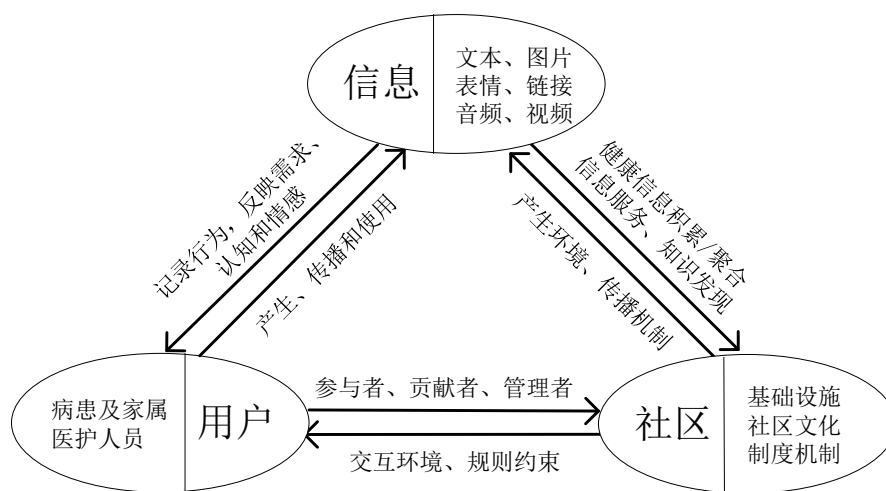


图 2-2 在线患者社区要素间关系<sup>[10]</sup>

从在线患者社区的三要素视角出发，现有研究多集中在以下几个方面：①针对用户与用户之间的交互，现有研究多探讨用户之间的社交关系网络的特点以及成因<sup>[11][12]</sup>；②针对用户和信息之间的交互，现有研究多探讨用户的健康信息行为，包括健康信息行为产生的影响因素<sup>[13][14]</sup>以及健康信息行为的分类<sup>[6][15]</sup>；③针对用户和社区之间的交互，主要集中在用户对在线患者社区的采纳和持续使用研究方面<sup>[16][17][18]</sup>；④针对信息这一要素，现有研究主要集中在对信息主题内容以及

信息质量的探讨和分析<sup>[1]</sup>；⑤针对社区这一要素，现有研究主要集中在对社区运作模式以及社区价值的分析<sup>[19]</sup>。

从研究方法角度来看，现有研究所采用的技术方法也越来越多元化。例如，在获取数据方面，既有传统的问卷调查或访谈数据，也有采用爬虫技术获得的客观网络数据；在进行数据分析过程中，既有结构方程模型、回归分析等定量方法，也有内容分析法、扎根理论等定性分析方法；且随着大数据、机器学习等技术的快速发展，在线患者社区的研究中也融入了文本挖掘、社会网络分析等方法。

综合来看，目前对在线患者社区的研究所涵盖的范围已比较广泛，已经形成了相对比较完善的研究体系和理论体系，但整体来说还处于探索阶段<sup>[10]</sup>，对于一些问题的研究还有待继续深入和丰富。本文聚焦用户之间、用户和信息之间以及用户和社区之间的交互，采用文本挖掘技术和生存分析模型，深入探讨用户之间所形成的同伴支持在信息这一载体上的分类问题，以及其所形成的不同类型的同伴支持对用户和社区之间的持续参与行为的影响，有助于丰富在线患者社区的研究，更深层次的理解用户之间所形成的同伴支持，以及它对用户持续参与的影响。

## 2.2 同伴支持相关研究

“同伴支持（peer support）”一词来源于护理领域，是社交关系概念的一部分<sup>[20]</sup>，也是社会支持的一种。社会支持是指使个体相信自己被关心、被爱护、被尊重、被当成具有相互责任的共同体成员的各类支持行为的总称，是一种具有帮助意图的行为。与社会支持相似，同伴支持也是一种强调正向互助的模式。

CL Dennis 基于“同伴（peer）”和“支持（support）”的语言学定义，将同伴支持定义为平等的个体所给予的帮助和鼓励<sup>[20]</sup>。具体到医疗健康领域，同伴支持是指通过多样化的形式使具有相似疾病、身体状况或经历的病患，彼此之间提供实质性帮助及生活实践、社会和情感方面的支持<sup>[21]</sup>。从最初的减少青少年吸烟的健康项目，到艾滋病预防等教育项目，再到肿瘤、糖尿病、精神病等慢性疾病健康领域，同伴支持的应用范围越来越广泛，效果也逐渐在实践中得到认可，形式越来越多样化。同伴支持具有多样性，可以在多种不同的环境下（诊所、社区组织或互联网等）通过各种提供者（专业组织或志愿者组织）提供多种模式的交互（一对一、互助小组或基于互联网）。

在不同的同伴支持结构中，个体的参与程度会有所不同，在高度正规化的同伴支持小组中，个体的参与程度会相对高于非正规化的同伴支持小组，例如，线下正规化支持小组的成员参与度会普遍高于线上非正规化同伴支持小组成员的参与度。但尽管个体参与程度有高有低，现有研究已经证实同伴支持不仅对患者

健康状况具有正向的促进作用<sup>[22]</sup>，还可以加强人际网络中的支持性关系，提高支持者的自我效能感，增强其应对疾病和自我发现的能力<sup>[23]</sup>。

同伴支持系统形成的要素包括相似的经历、利益共享、自然的社会支持、自愿参与和接受帮助以及同伴对服务的控制<sup>[23]</sup>。其中，相似的经历是指同伴需要有相似的疾病经历，可以加强彼此间的信任；利益共享是指在获得其他人帮助的同时也要为他人提供帮助，这是同伴支持系统成立的前提条件；自然的社会支持类似于经验学习要素，即参与者可以从活动中受益，而不受同伴社会地位的影响；自愿参与和接受帮助是获益的重要影响因素，那些非自愿参与的支持者很难从中获益；同伴对服务控制是和谐同伴关系的基础。

### 2.2.1 同伴支持与在线患者社区

从同伴支持系统形成的要素层面看，在线患者社区是一个典型的同伴支持系统。在线患者社区的内部成员具有相似的疾病经历，他们有自发性的发帖、回帖等健康信息行为。但不同的是，在线患者社区的同伴支持中，病患对是否寻求或提供支持有较高的自由度，所以并不强调对服务的控制。相似的是，自愿参与和接受帮助不仅是同伴支持系统形成的要素，也是在线患者社区长久发展的要素。所以对同伴支持的研究也有助于为在线患者社区的发展提供见解。

在线患者社区中的同伴支持为成员提供了有关健康和疾病管理的信息、情感和工具支持<sup>[24]</sup>。此外，在线寻找同伴的信息和支持已成为一种有效且经济的解决方案，特别是当专家的反馈成本高且难以扩展时<sup>[25]</sup>。在线患者社区中，患者可以分享他们健康状况管理策略的细节，描述个人疾病轨迹，并就疾病管理达成某种共识<sup>[26]</sup>。与其他主题的在线论坛不同，之前的研究发现，在线患者社区重视意见和经验的多样性，而不是达成共识<sup>[24]</sup>。在线患者社区中，人们经常分享不同的个人经历以及很有价值的详细建议和支持，这是医疗专业人员所不能提供的<sup>[27]</sup>。

目前，类似在线患者社区的同伴支持系统主要应用于慢性疾病管理方面。慢性病是全球第一位的疾病，其治疗过程复杂且漫长，治愈可能性较小，需要患者长期关注自己的身体变化，并进行有效控制。因此，慢性病控制的核心策略是自我管理<sup>[28]</sup>。但患者在长期自我管理过程中，由于医生时间的有限，只能提供短时间的病情咨询指导，不可能全程关注患者的疾病变化情况。而疾病的复杂多变性也意味着患者可能会遇到诸多问题，因此，仅靠专业的医疗工作人员已经不能满足慢性病患者服务需求。同伴支持作为社会支持的一种，在慢性病自我管理领域发挥了重要作用。例如，同伴支持有助于提高糖尿病患者的糖尿病知识水平和自我管理能力和<sup>[29]</sup>，提高心理疾病患者的自我效能感<sup>[30]</sup>，帮助其恢复社会功能。在线患者社区主要通过同伴之间的经验分享、问题解答和情感支持等行为来影响生

活方式、科学治疗，从而实现健康的改善。本文主要关注由糖尿病患者构成的在线患者社区。

糖尿病是全球最主要的慢性病之一，糖尿病及其所带来的并发症严重影响着患者的身体健康和生活质量，使其成为全球第五大死因（世界卫生组织，2006）。中国已经成为糖尿病的重灾区，糖尿病患者人数高居世界榜首，成人患病率高达11.6%<sup>[31]</sup>，每十个成年人中就有一位糖尿病患者，糖尿病成为我国最为重要的公共卫生问题之一。对于糖尿病患者来说，需要每天监测血糖水平和估计营养的摄入量，1型糖尿病患者还需要注射胰岛素。糖尿病的治疗需要患者付出长久的努力。随着社交媒体的出现，糖尿病患者越来越多地在各种形式的在线社区中寻求糖尿病患者同伴的信息和情感支持<sup>[17]</sup>，推动了在线糖尿病社区数量和规模的增长。因此，在线糖尿病社区为研究在线患者社区的持续参与提供了环境基础。

对于在线患者社区中的同伴支持研究，现有文献多将在线患者社区中互助行为看作社会支持来进行研究。国内外研究多集中在以下两个方面：①社会支持作为前置因素，对用户知识分享等社区参与行为的影响<sup>[32]</sup>；②社会支持对用户健康状况的作用效果分析<sup>[33]</sup>。除上述两个方面的内容外，还涉及对在线患者社区社会支持的分类研究<sup>[1][6]</sup>。

### 2.2.2 同伴支持分类相关研究

Dennis 通过对文献的深入研究，在同伴干预描述中识别出了三类关键的同伴支持：情感支持（emotional support）、信息支持（informational support）和评价支持（appraisal support）<sup>[20]</sup>。其中，情感支持是通过分享令人心情波动的事件来传递的，内容包含鼓励、关怀、信任、喜欢或爱心，可以帮助患者减轻压力水平，使其感到关怀和温暖。例如表达对患者的理解、与其同在的同情、情况会变好的鼓励等；信息支持是信息和指导的传递，由同伴提供疾病治疗、康复以及生活方面的信息或指导，给予忠告或提供协助解决问题的建议；评价支持，也称肯定支持，包括与自我评估和对情感、认知和行为的适当性表达肯定有关的信息交流<sup>[34]</sup>。具体的机制包括激励方面，例如鼓励其坚持解决问题，保证努力将取得积极成果，帮助其抵御挫折以及保持乐观沟通<sup>[35]</sup>。根据 Dennis 对评价支持的定义，可以看出评价支持兼有信息支持和情感支持的特点，所以在实践中评价支持和另外两种支持之间的区分度并不高，例如对他人治疗方案表示肯定，在表达评价支持的同时也传递了信息支持，鼓励他人坚持锻炼在传递评价支持的同时也表达了情感支持。这些不同的支持类型可能在不同的阶段有不同的作用，例如，针对新冠肺炎确诊患者最初会涉及更多的情感支持，而对已经进入稳定阶段的患者，则可能会涉及较多的信息支持，同时还有情感支持和评价支持的强化。

对于同伴支持分类的研究,学者较少使用同伴支持一词,研究涉及比较多的是对社会支持分类的研究。同伴支持也是社会支持的一种,因此,本部分内容在进行文献回顾时将上升至社会支持,对其分类研究进行归纳总结。国内外学者对社会支持分类做了一系列探索,在社会支持类型划分方面,学者们一致认为社会支持包括信息支持和情感支持<sup>[1][6][36][37][38]</sup>,除关注较多的信息支持和情感支持外,也有部分学者在研究中纳入了陪伴支持(companion support)<sup>[38][39]</sup>,也称为网络支持(network support)或离题支持(off-topic support),它包括聊天、幽默以及与健康问题无关的线下活动和日常生活的讨论。此外,还包括有形支持(tangible support),也称工具支持(instrument support),指现实世界中的支持实践<sup>[38]</sup>,例如提供药物购买帮助等。但在虚拟社区中,有形支持非常少见,比较常见的是信息支持、情感支持和陪伴支持<sup>[40]</sup>。综合 Dennis 对同伴支持分类以及文献中对社会支持分类,信息支持和情感支持是普遍包含在内的。考虑到评价支持与信息支持、情感支持之间存在交叉,有形支持在虚拟社区中的稀缺性,以及陪伴支持在虚拟社区中的常见性,本文的研究将在线患者社区中的同伴支持分为信息支持、情感支持和陪伴支持三种类型。

关于同伴支持和社会支持分类方法,可以大致分为两类:一种是采用传统的质性分析方法。GUI 等采用扎根理论对某孕妇在线论坛的帖子进行分析,归纳得出孕妇在论坛中寻求同伴支持中占比较高的支持类型是诸如知识建议等的信息支持、诸如肯定的评价支持和情绪支持<sup>[1]</sup>;Rodgers 等利用内容分析法对乳腺癌医疗社区的文本信息进行分析,发现该患者获得信息支持、彼此进行情感交流以及和其他患者相互鼓励后,负面情绪会得到明显改善,并且更有信心应对后续的治疗<sup>[41]</sup>。虽然传统的质性分析技术可以从自然文本科学地提炼出支持类型,但随着在线文本数据的爆炸式增长,单纯依靠人工标注的方法不免要付出巨大的时间成本和人工成本。

另一类是应用机器学习方法进行文本分类。和传统的质性分析技术相比,机器学习方法可以节约大量的人工成本和时间成本,提高了文本分类的效率。Zhang 等采用贝叶斯分类方法对在线健康社区的用户发帖意图(寻求社会支持、提供社会支持、接受社会支持)和社会支持类型(信息支持和养育支持)进行分类,并对比了标题、帖子内容和评论的特征对帖子分类效果的作用,发现标题是用户发帖意图分类最重要的特征<sup>[6]</sup>。Chen 采用机器学习方法对在线健康社区社会支持进行分类,利用词嵌入特征、词袋特征、语言特征作为深度学习神经网络的输入,研究发现,word2vec 能更好地捕获在线社会支持分类的重要特征<sup>[42]</sup>。

通过对社会支持和同伴支持分类的文献回顾,现有的研究不足之处主要在于缺少对中文文本的社会支持分类研究。现有研究主要集中在对英文文本社会支持的分类,但中文文本和英文文本在词法和语法方面具有很大的差异,英文文本的

社会支持分类方法并不能完全适应中文文本。因此，需要对中文文本的社会支持分类做进一步的探索，建立自动化的分类器，提高分类的效率和准确率。

## 2.3 文本分类的相关技术

文本分类包括以下几个基本的流程：数据预处理、文本特征表示、分类器选择、模型训练和性能评估，本节重点回顾和介绍文本特征表示、分类模型以及性能评估指标三个模块的内容。

### 2.3.1 文本特征表示

在进行诸如文本分类的文本数据分析中，需要将非结构化的文本数据表示为计算机能够理解的数学形式，即将文本信息数字化。其中最简单的方法是独热编码方法 `one hot representation`，它将所有出现过的词语组合在一个向量中，词向量长度为词语的个数，若句子中出现该词语，则词向量对应位置标为 1，未出现则为 0。因此 `one hot representation` 所生成的词向量非常稀疏，且词与词之间相互独立，不利于文本分类。在本节中，文章重点介绍基于神经网络的 `word2vec` 词向量模型和 `doc2vec` 词向量模型，以及基于分布概率的 LDA 文档主题生成模型。

#### (1) `word2vec` 词向量模型

`word2vec` 是一种基于三层神经网络的自特征抽取的词向量表示方法，由 Mikolov 等人于 2013 年提出<sup>[43]</sup>。其基本思想是通过训练将每个词映射成低维实数向量，通过词在向量空间上的距离来判断词之间的语义相似度，向量空间上距离越近的词语义相似度越高。`word2vec` 可以高效地产生词向量，在文本分类、情感分析等领域都得到了广泛应用。

`word2vec` 包含两种模型：`CBOW` (Continuous Bag of Words) 模型和 `Skip-gram` 模型，`CBOW` 模型是根据上下文预测当前词语，`Skip-gram` 模型是根据当前词语预测上下文。由于 `word2vec` 模型仅是单个词语的向量表示，因此结果不能直接用于文档（或消息）级别的文本分类。表示文档的典型做法是将文档中所有词向量的平均值作为文档的嵌入表示。但是，这种聚合并不能保证对相似文档的适当表示<sup>[42]</sup>，且忽略了单词之间的排列顺序。

## (2) doc2vec 文档向量模型

doc2vec 是在 word2vec 词向量模型基础上得到的，它将 word2vec 方法扩展到了文档、段落或句子的连续表示中，由 Le 和 Mikolov 于 2014 年提出<sup>[44]</sup>。doc2vec 和 word2vec 类似，不同的是 doc2vec 增加了一个段落向量，即每个句子都有一个唯一的 id，并将该 id 映射为向量。段落向量和词向量连接到一起，作为神经网络的输入。

doc2vec 也包含两种方法：(1)DM(distributed memory)模型，类似于 word2vec 中的 CBOW 模型，使用文档/段落向量和上下文词向量来预测某一单词；(2)DBOW(distributed bag of words)模型，类似于 word2vec 中的 skip-gram 模型，使用文档/段落向量预测一组随机单词的概率。如图 2-3 和图 2-4 所示。

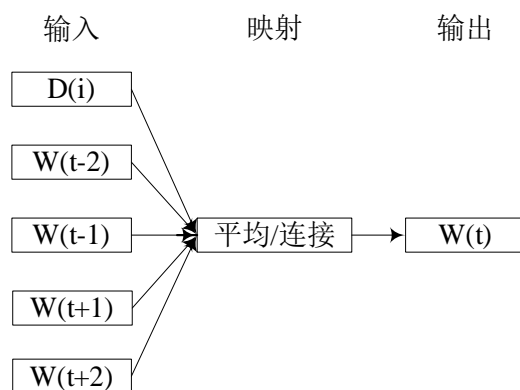


图 2-3 DM 模型图

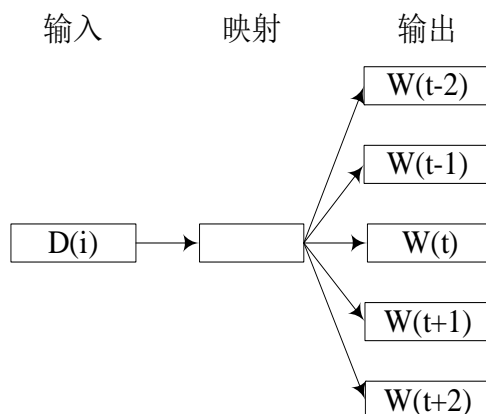


图 2-4 DBOW 模型图

## (3) LDA 文档主题生成模型

LDA (Latent Dirichlet Allocation)，即潜在狄利克雷分布，是一种文档主题生成模型，由 Blei 等人于 2003 年提出<sup>[45]</sup>，用以推测文档的主题分布，是最流行的主题模型方法之一，常用来进行文本分类。它的基本思想是将文档表示为潜在主

题上的随机混合，其中每一个主题特征都由词的概率分布所表示。LDA 包含文档、主题、词三层结构，文档由不同的主题组成，主题由不同的词组成，文档—主题分布、主题—词分布均服从多项式分布。

LDA 生成模型指的是文档的生成过程，一篇文档可以看作有限个词语的有序组合，文档中的每一个词都是通过“以一定的概率选择某一个主题，并在这个主题下以一定的概率选择某一个词”得到的，图模型如图 2-5 所示。首先，假设语料库有  $M$  篇文档，每篇文档有  $N$  个词，共涉及  $K$  个主题。

- i. 从参数为  $\alpha$  的狄利克雷分布中采样生成文档  $m$  的主题分布  $\theta_m$ ;
- ii. 对于文档  $m$  的第  $n$  个词，从多项分布  $\theta_m$  中采样生成其对应的主题  $z_{m,n}$ ,  $z_{m,n} \in [1, K]$ ;
- iii. 从参数为  $\beta$  的狄利克雷分布中采样生成主题  $z_{m,n}$  的词分布  $\varphi_{z_{m,n}}$ ;
- iv. 从词分布  $\varphi_{z_{m,n}}$  中采样生成文档  $m$  的第  $n$  个词  $w_{m,n}$

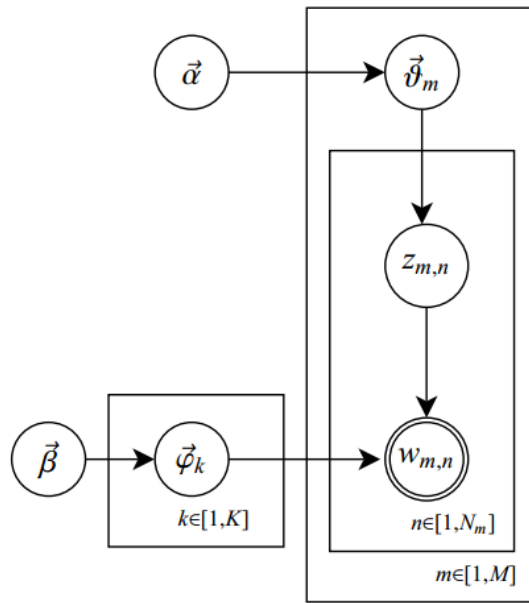


图 2-5 LDA 图模型结构

在进行 LDA 模型构建过程中，对于一篇文档  $m$ ， $w_{m,n}$  是已知的变量， $\alpha$  和  $\beta$  是根据经验给定的先验参数，所要学习进行参数估计的则是隐变量  $\theta_m$  和  $\varphi_{z_{m,n}}$ ，根据图模型，所有变量的联合分布为：

$$p(\overrightarrow{\theta}_m, \overrightarrow{z}_m, \overrightarrow{w}_m, \varphi | \vec{\alpha}, \vec{\beta}) = p(\overrightarrow{\theta}_m | \vec{\alpha}) \prod_{n=1}^{N_m} p(\overrightarrow{z}_{m,n} | \overrightarrow{\theta}_m) p(w_{m,n} | \varphi_{z_{m,n}}) p(\varphi | \vec{\beta})$$



对于隐变量 $\theta$ 和 $\varphi$ 的参数估计，目前通常采用 Gibbs 抽样算法。通过训练 LDA 主题模型，最终可以得到文档—主题矩阵 $Z_M$ ， $Z_{M,K}$ 为文本 M 对应的主题概率向量。

$$Z_M = \begin{bmatrix} Z_{1,1} & Z_{1,2} & \cdots & Z_{1,K} \\ Z_{2,1} & Z_{2,2} & \cdots & Z_{2,K} \\ \vdots & \vdots & \vdots & \vdots \\ Z_{M,1} & Z_{M,2} & \cdots & Z_{M,K} \end{bmatrix}$$

### 2.3.2 文本分类模型

#### (1) k 近邻模型

k 近邻模型 (k-Nearest Neighbor, kNN) 是文本挖掘分类算法中最基础和最简单的模型之一。所谓“近朱者赤，近墨者黑”，kNN 的工作机制很简单，即给定一个测试样本，如果在特征空间内该测试样本的 k 个最邻近训练样本中的大多数属于某一类，则该测试样本也被划分为这个类别。可以看出，kNN 是基于实例的学习，它并没有显式的训练过程，而是在收到测试样本后才开始处理，是“懒惰学习”的典型代表。

kNN 算法的计算过程描述如下：

- i. 确定参数 k 的取值
- ii. 计算测试样本与各训练样本之间的距离
- iii. 按照距离大小选择 k 个最近邻样本
- iv. 确定 k 个最近邻样本所属的类别
- v. 按照“投票法”，选择 k 个最近邻样本中出现次数最多的类别作为测试样本的预测结果

在 kNN 分类过程中，最重要的就是参数 k 的确定。如果 k 值取值过小，则预测结果容易受到噪声的影响。如果 k 值取值过大，则预测的误差就会变大。虽然 kNN 分类算法简单易实现，但由于需要在特征空间内计算样本间的距离，所以对于大规模数据，kNN 的计算量很大，比较耗时。对于样本不均衡的数据，kNN 的分类容易受样本多的类别的影响，将少数类别中的样本误判为出现次数多的类别，效果并不好。改进的做法是根据距离调整权重<sup>[46]</sup>，或根据类别样本量赋予不同的权重等<sup>[47]</sup>。

## (2) 逻辑回归模型

逻辑回归模型 (Logistics Regression, LR) 是一种经典的二分类机器学习算法, 在生产实践中应用广泛, 也是自然语言处理常用的算法之一。它是在线性回归基础上得到的, 线性回归的基础形式是:

$$z = w_0x_0 + w_1x_1 + \dots + w_nx_n$$

逻辑回归将线性函数的结果映射到了 sigmoid 函数中, 从而得到逻辑回归的基本表示形式:

$$h_w(x) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-w^T x}}$$

其中,  $h_w(x)$  表示具有特征  $x$  的样本所属类别的概率, 介于 (0,1) 之间, 中间值为 0.5。如果是二分类问题, 假设有 0、1 两种类别, 若  $h_w(x) < 0.5$ , 则该样本属于类别 0, 如果  $h_w(x) > 0.5$ , 则该样本属于类别 1。可以得到:

$$p(y = 1|x; w) = h_w(x)$$

$$p(y = 0|x; w) = 1 - h_w(x)$$

$$p(y|x; w) = h_w(x)^y (1 - h_w(x))^{1-y}$$

其中,  $y$  是样本的预测值。在  $h_w(x)$  中  $w$  是未知的参数, 需要进行参数估计, 一般采用最大似然估计的方法进行求解, 找到一组  $w$  值, 使  $h_w(x)$  正确的概率最大。

## (3) 决策树模型

决策树 (Decision Tree, DT) 是分类算法中应用最广泛的算法之一, 和人类决策的过程非常相似, 因此具有较好的可解释性, 对噪声数据具有很好的适应性。决策树是一种树形结构, 由节点和有向边组成。节点包含两种类型: 叶节点和内部节点, 每个叶节点表示一种类别, 每个内部节点表示一个属性。

决策树分类的过程就是自上而下逐层建树的过程。一个决策树包含一个根节点, 若干个内部节点和叶节点。在决策树中, 根节点对应的是样本集合, 每个内部节点对应一个属性测试, 每个叶节点对应一个类别, 从根节点到内部节点的有向边代表了一条测试路径, 根据属性测试结果, 样本被划分至相应的类别, 即连接至相应的叶节点, 如图 2-6 所示。

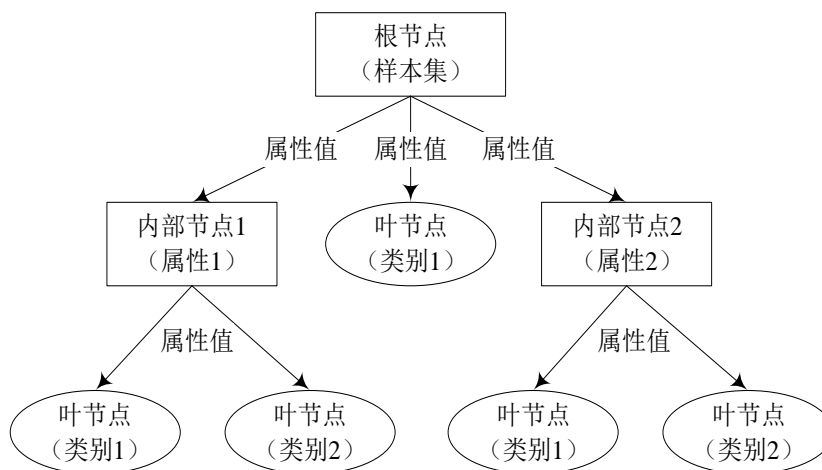


图 2-6 决策树模型

决策树的构建包含三个流程：特征选择、决策树生成和决策树剪枝。特征选择通常基于信息增益或信息增益率。决策树的生成有三种方法：ID3、C4.5 和 Cart 树，其中 ID3 是根据信息增益选择构建决策树的分支节点，依次递归建树；而 C4.5 则在 ID3 基础上，根据信息增益率划分决策树，改进了信息增益偏向特征值较多的不足；Cart 树根据 Gini 系数划分决策树。决策树考虑了所有样本集，很多分支可能反映的是数据中的噪声或离群点，可能会出现噪声，因此，需要对决策树进行剪枝，剪掉一些信息量少的分支，以避免过拟合。

#### (4) 朴素贝叶斯模型

朴素贝叶斯 (Naïve Bayes, NB) 是文本分类中常用的方法之一，该方法的基础是贝叶斯定理，方法成立的前提是特征之间必须独立，所以称其为朴素贝叶斯。根据贝叶斯定理，在事件 B 已经发生的情况下，事件 A 发生的概率  $P(A|B)$  为：

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

在实际的应用过程中，一个事件的发生可能受多个因素的影响，因此，假设 B 事件发生的影响因素有 n 个，分别为  $b_1, b_2, b_3, \dots, b_n$ ，则：

$$P(A|b_1, b_2, b_3, \dots, b_n) = \frac{P(b_1, b_2, b_3, \dots, b_n|A)P(A)}{P(b_1, b_2, b_3, \dots, b_n)}$$

根据链式法则，可知：

$$P(b_1, b_2, b_3, \dots, b_n|A) = P(b_1|A)P(b_2|A, b_1)P(b_3|A, b_1, b_2) \cdots P(b_n|A, b_1, b_2, b_3, \dots, b_{n-1})$$

若  $b_1, b_2, b_3, \dots, b_n$  之间相互独立，则：

$$P(A|b_1, b_2, b_3, \dots, b_n) = \frac{\prod_{i=1}^n P(b_i|A) P(A)}{P(b_1, b_2, b_3, \dots, b_n)}$$

在文本分类中，将上式中的 A 视为分类的类别 C， $b_i$  视为特征 F，则可得到朴素贝叶斯分类器的模型公式为：

$$P(C|F_1, F_2, F_3, \dots, F_n) = \frac{\prod_{i=1}^n P(F_i|C)P(C)}{P(F_1, F_2, F_3, \dots, F_n)} \quad \textcircled{1}$$

基于上述公式，假设文本分类结果包含 k 个类别： $C_1, C_2, \dots, C_k$ ，分类测试的特征包含 n 个  $F_1, F_2, \dots, F_n$ ，给定文本 d，朴素贝叶斯分类的基本流程为将文本 d 的所有特征值代入公式①，分别计算文本 d 属于类别  $C_1, C_2, \dots, C_k$  的概率，概率值最大的类别就是文本 d 的最终预测结果。

### (5) 支持向量机模型

支持向量机模型（Support Vector Machine, SVM）是一种针对二分类任务的分类模型，其基本思想是基于训练集在样本空间中寻找到一个超平面对样本进行划分，使得超平面与每个类中距离最近的样本距离最大，即满足“间隔”最大化。这样的超平面所产生的分类结果最具鲁棒性，泛化能力最强。在二维空间中，超平面是一条直线；在三维空间中，超平面是一个平面。

给定样本集  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ，在样本空间中，超平面的划分可以用下式表示：

$$\mathbf{w}^T \mathbf{x} + b = 0$$

其中  $\mathbf{w} = (w_1; w_2; \dots, w_n)$  是法向量，决定了超平面的方向； $b$  是位移项，代表了超平面与原点之间的距离。样本到超平面的距离为：

$$r = \frac{|\mathbf{w}^T \mathbf{x} + b|}{\|\mathbf{w}\|}$$

假设样本  $y$  的取值为 -1 或 +1，超平面可以对样本正确分类，令：

$$\begin{cases} \mathbf{w}^T \mathbf{x}_i + b \geq +1, & y_i = +1 \\ \mathbf{w}^T \mathbf{x}_i + b \leq -1, & y_i = -1 \end{cases}$$

距离超平面最近的几个样本点使上式中等号成立，这些样本被称为“支持向量”，两个类别的支持向量到超平面的距离之和为  $\frac{2}{\|\mathbf{w}\|}$ ，这段距离被称之为“间隔”，如图 2-7 所示。

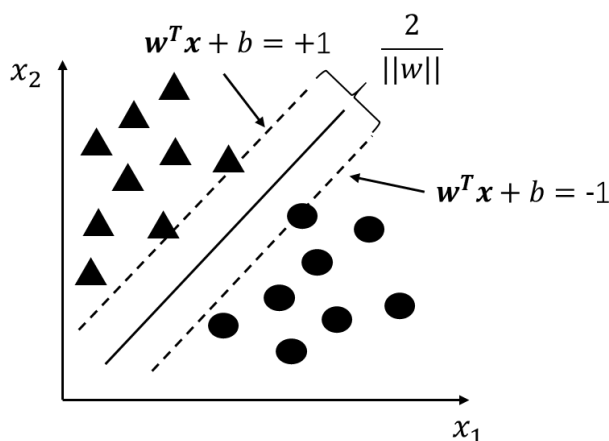


图 2-7 支持向量

支持向量机的目的就是找到“间隔”最大的超平面，即 $\frac{2}{\|w\|}$ 最大化，等价于 $\|w\|^2$ 最小化，满足要求的样本点就是支持向量，位于最大间隔边界上。支持向量机训练出的最终模型只与支持向量有关，和其他样本点无关。SVM 模型适用于解决二分类问题，对多分类问题效果不好，适用于小样本、高维、非线性问题，具有较高的泛化性能<sup>[48]</sup>。

## (6) 卷积神经网络模型

卷积神经网络（Convolutional Neural Networks, CNN）是一种前馈神经网络（Feedforward Neural Networks, FNN），是深度学习算法代表的一种，在自然语言处理领域也有广泛的应用。在文本分类中，卷积神经网络的卷积层和池化层可以实现对文本特征的抽取，准确提炼文本主题，实现较好的分类效果。卷积神经网络包括输入层、隐含层和输出层，其中隐含层又包括卷积层、池化层和全连接层。下面对 CNN 主要组成部分在文本分类任务中的应用进行重点介绍。

### 1) 输入层（Input Layer）

在文本分类任务中，CNN 的输入层是句子中词语对应的词向量序列，若句子包含的词长度是 $n$ ，词向量维度为 $k$ ，则输入层输入的是 $n \times k$ 的矩阵。词向量可以是静态的，也可以是动态的。所谓静态词向量是指词向量在模型训练过程中固定不变，动态词向量是指词向量随着模型的训练不断进行优化。

### 2) 卷积层（Convolution Layer）

在输入文本词向量后，通过卷积层的卷积运算得到文本的特征。卷积运算就是将词向量矩阵和卷积核矩阵进行相乘后再相加，其中，卷积核的宽度与词向量的维度保持一致，卷积核的高度无特定要求，可自行设置，一般设置为 2/3/4/5。卷积操作后通过激活函数的非线性映射可以得到对应的特征，卷积操作兼顾了词义、词序和词语的上下文信息，可以实现对文本的特征映射。

### 池化层（Pooling Layer）

池化层的功能是对卷积层输出的特征向量进行池化，降低特征维数，减少参数数量，可以缩短训练时间，并控制过拟合。池化层最常用的方法是 **max-pooling**，即对特征向量  $C$  保留值最大的特征，舍弃其他特征，得到一维的向量。**max-pooling** 虽然保留了最强特征，降低了特征维度，但却丢失了文本的位置信息和特征的强度信息。为了解决上述问题，Kalchbrenner 等提出了 **k-max-pooling** 方法，即保留前  $k$  个值最大的特征，并保留这些特征在原特征向量  $C$  中的先后顺序<sup>[49]</sup>。**k-max-pooling** 保留了部分位置信息和特征强度信息。

### 3) 全连接层（Fully Connected Layer）

全连接层是 CNN 的“分类器”，它将池化层的输出通过全连接的方式连接一个 **Softmax** 激活函数，通过激活函数得到属于每个类别的概率。**Softmax** 函数需要根据任务类型进行设置。对于二分类任务，可以使用 **sigmoid** 作为激活函数；对于多分类任务，可以设置 **categorical\_crossentropy** 作为激活函数。

在全连接层还采用了 **dropout** 技术，在模型训练过程中随机使某些节点不工作，防止模型的过拟合，并通过损失函数对模型训练结果进行优化，生成最优的分类模型。

## （7）循环神经网络模型

循环神经网络模型（Recurrent Neural Networks, RNN）也是一种深度学习模型，是在前馈神经网络 FNN 的基础上发展而来，其算法的主要目的是处理可变量长度的时序数据<sup>[50]</sup>。RNN 模型拥有短期记忆能力，在自然语言处理领域有广泛的应用场景。

与 FNN 模型和 CNN 模型不同的是，在 FNN 模型和 CNN 模型中，隐藏层神经元的连接是单向的，当前层的输出只与其输入即上一层的输出有关，而同一层之间的神经元是不相连的。但在实际任务中，网络的输出不仅与其输入有关，也可能和自身在前面时刻的输出有关。例如，在处理文本任务预测下一个词时，由于句子中词之间是相互联系的，所以下一个词不仅和当前输入的词有关，也和上一个词有关。RNN 通过引入定向循环结构，能够解决信息之间相互关联的情况，在 RNN 中，神经元不仅可以接收上一层神经元的输出信息，还可以接收自身上一时刻的信息，形成定向循环的网络结构，如图 2-8 所示。

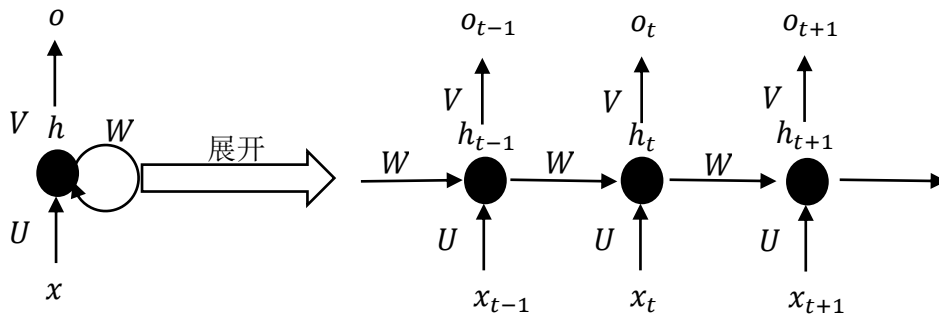


图 2-8 RNN 网络结构

在图 2-8 中，当前隐藏层的输出  $h_t$  取决于其输入  $x_t$  以及前一时间隐藏层的输出  $h_{t-1}$ 。  $U$ 、 $V$ 、 $W$  是层共享权重矩阵，  $h_i$  表示  $i$  时刻隐藏层的输出，  $o_i$  表示  $i$  时刻输出层的输出。

$$h_t = \begin{cases} 0, & t = 0 \\ f(U * x_t + W * h_{t-1}), & t \neq 0 \end{cases}$$

$$o_t = g(V * h_t)$$

$f(\cdot)$  和  $g(\cdot)$  为激活函数。在文本分类任务中，通常是使用一个 RNN 将输入的文本序列映射为固定长度的向量，然后将向量输入到 softmax 层以进行分类<sup>[51]</sup>。

RNN 的训练算法是基于时间的反向传播算法 (Back Propagation Through Time, BPTT)，BPTT 采用交叉熵函数作为损失函数，并根据损失函数利用梯度下降算法 SGD 来求解最优参数。

但 RNN 存在一个问题，对于长序列数据，在训练过程中，梯度矢量的分量会呈指数级增长或衰减<sup>[52][53]</sup>。梯度爆炸和梯度消失的问题使 RNN 模型难以学习序列中的长距离相关性。

针对 RNN 梯度爆炸和梯度消失问题，Hochreiter 和 Schmidhuber 提出了长短期记忆神经网络 (Long Short Term Memory, LSTM)<sup>[53]</sup>。LSTM 在隐藏层内部维护一个单独的记忆单元  $c_t$ ，记忆单元包含三个“门”，分别是输入门  $i_t$ 、遗忘门  $f_t$ 、输出门  $o_t$ <sup>[54]</sup>。输入门  $i_t$  决定了保留多少当前时刻的输入信息到记忆单元中，遗忘门  $f_t$  决定了保留多少上一时刻记忆单元信息到记忆单元中，输出门  $o_t$  决定了当前时刻记忆单元信息输出的多少。通过激活函数变换，输入门  $i_t$ 、遗忘门  $f_t$ 、输出门  $o_t$  三者的值都属于  $[0,1]$ 。

在 LSTM 基础上，Cho 等提出了门限循环单元 (Gated Recurrent Unit, GRU)<sup>[55]</sup>。和 LSTM 相似，GRU 也有调节内部信息流动的控制单元，但没有单独的记忆单元。GRU 有重置门  $r_t$  和更新门  $z_t$  两个门，重置门决定了舍弃多少前一时刻的状态信息到当前时刻的记忆单元，更新门决定了保留多少前一时刻状态信息到当前时刻的记忆单元。

综上所述，文本分类模型主要包括传统的基于统计学习的机器学习模型和基于神经网络的深度学习模型。在文本分类领域，应用较多的机器学习模型主要是 k 近邻模型、逻辑回归模型、决策树模型、朴素贝叶斯模型和支持向量机模型，应用较广泛的深度学习模型是卷积神经网络模型和循环神经网络模型。在对上述模型进行介绍的基础上，本节还总结对比了这些模型的优缺点，如表 2-1 所示。

表 2-1 文本分类模型优缺点对比

| 模型       | 优点                                   | 缺点                                   |
|----------|--------------------------------------|--------------------------------------|
| k 近邻模型   | 简单易实现；<br>支持增量式运算                    | 对大规模数据计算量大；<br>对样本平衡性要求较高；<br>可解释性不强 |
| 逻辑回归模型   | 计算代价低；<br>易于实现和解释                    | 分类精度不高；<br>不易做增量处理                   |
| 决策树模型    | 简单易理解；<br>适合高维数据                     | 不支持增量式运算；<br>容易过拟合                   |
| 朴素贝叶斯模型  | 简单速度快；<br>所需估计参数少                    | 样本属性独立假设不易满足                         |
| 支持向量机模型  | 解决小样本机器学习问题；<br>泛化性能高；<br>解决非线性和高维问题 | 核函数选择困难；<br>只适合二分类问题                 |
| 卷积神经网络模型 | 适合短文本分类；<br>可以自动提取特征                 | 需要调参；<br>可解释性差                       |
| 循环神经网络模型 | 使用文本序列信息；<br>输入长度可变                  | 训练速度慢；<br>可解释性差                      |

### 2.3.3 文本分类评估指标

文本分类任务中常用的性能评估指标有准确率(accuracy)、精确率(precision)、召回率(recall)、F1 值(F1-score)以及 AUC (Area under curve)，其中 F1 值是精确率和召回率的调和平均值。因此，接下来将主要介绍准确率、F1 值和 AUC。

#### (1) 准确率

准确率是分类问题中最原始的评价指标，准确率的含义是预测正确的结果占总样本的百分比，其公式如下：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

其中，TP (True Positive)，即真正例，表示被模型预测为正的正样本数量；FP (False Positive)，即假正例，表示被模型预测为正的负样本数量；FN (False Negative)，即假负例，表示被模型预测为负的正样本数量；TN (True Negative)，即真负例，表示被模型预测为负的负样本数量。



但准确率指标有一个缺陷，在样本数据类别不平衡时，准确率容易受到样本数据多的类别的分类结果影响，所以在这种情况下准确率并不能客观评价算法的优劣。

## (2) F1 值

F1 值 (F1-Score) 是精确率 (Precision) 和召回率 (Recall) 指标的加权调和平均。精确率又称查准率，指在所有被预测为正的样本中实际为正的样本的概率，其公式如下：

$$Precision = \frac{TP}{TP + FP}$$

召回率又叫查全率，指在实际为正的样本中被预测为正样本的概率，其公式如下：

$$Recall = \frac{TP}{TP + FN}$$

可以看出，有时精确率的升高，随之而来的是召回率的下降。在同伴支持分类中，需要两者兼顾，因此考虑采用 F1 分数，即精确率和召回率的调和加权平均，其公式如下：

$$\frac{1}{F_1} = \frac{1}{2} \left( \frac{1}{Precision} + \frac{1}{Recall} \right)$$

$$F_1 = \frac{2 \times TP}{\text{样例总数} + TP - TN}$$

## (3) AUC

AUC (Area Under Curve)，又称曲线下面积，是 ROC 曲线下方那部分面积的大小，表示正例排在负例前面的概率。ROC 中包含两个重要的指标：真正率 TPR 和假正率 FPR，其曲线的横坐标为假正率 (FPR)，纵坐标为真正率 (TPR)。其中真正率召回率的含义相同，指在实际为正的样本中被预测为正样本的概率，其公式如下：

$$TPR = \frac{TP}{TP + FN}$$

假正率表示实际为负的样本中被预测为正的样本数量，其公式如下：

$$FPR = \frac{FP}{TN + FP}$$

真正率和假正率都是针对单一类别的预测结果而言的，因此对样本是否均衡并不敏感。在一个分类任务中，负样本误判的越少，正样本召回的越多，表示该

模型的性能越好。即 TPR 越大, FPR 越小, ROC 曲线越陡, 则模型的性能越好。ROC 曲线越陡, 则其曲线下方的面积越大, 即 AUC 越大。因此, AUC 越大, 模型的性能越好。

## 2.4 在线患者社区用户持续参与研究

在线患者社区已经成为一种深受患者喜爱、易于获取的信息来源<sup>[17]</sup>, 特别是对于慢性病患者。大约四分之一患有慢性病(如高血压、糖尿病、心脏病和癌症)的互联网用户会使用在线社区<sup>[56]</sup>。研究发现参与在线健康社区可以提高用户在线寻求信息的能力、参与医疗保健的自我效能、情绪健康和生活质量<sup>[57]</sup>。同时, 有研究表明这些益处都伴随着持续使用而发生<sup>[58]</sup>。持续参与是持续使用的增强, 持续使用包含了潜水行为, 即只浏览而不产生内容, 而持续参与则表示用户持续性地产生内容, 参与平台活动。用户的持续参与是在线社区成功的关键, 但留住用户并保持活跃一直是在线社区面临的主要挑战。患者会在确诊后立即参与在线健康社区, 但其积极参与时间往往不到一年<sup>[59]</sup>, 具有较大的动态性和不稳定性。因此, 对在线患者社区用户持续参与影响因素的研究无论是对患者还是在线患者社区都是至关重要的。

在线患者社区的持续参与影响因素区别于在线患者社区采纳使用的影响因素, 现有研究探索了在线社区用户参与意向或参与行为的影响因素, 这些因素包括信息收益、关系收益(例如归属感)、自我效能等<sup>[60][61]</sup>, 但用户参与不等于持续参与, 用户初次参与的影响因素可能和持续参与有所不同, 初次参与后所出现的因素可能会对后续的持续参与产生影响<sup>[62]</sup>。例如, 在在线患者社区中, 用户的首次参与可能是因为咨询疾病的某个相关知识, 即获得信息收益, 但在该问题解决后用户是否还能持续使用? 这可能与用户在咨询过程中所感知到的其他因素有关, 例如用户满意度<sup>[63]</sup>。因此, 也有研究探讨了在线社区用户持续参与的影响因素, 这些研究多认为用户持续参与的主要驱动因素是社交互动和从互动中所获得的收益<sup>[63]</sup>。

表 2-2 通过识别研究的理论视角、自变量、因变量、主要发现, 总结了近年来国内外在线社区用户持续参与影响因素的相关文献。研究多从主观的认知角度出发, 以期望确认理论、技术接受模型和社会交换理论等为基础, 通过构念开发, 以问卷调查的方式获取数据, 了解信息系统相关属性以及用户的感知获益或价值等对持续参与的影响。此外, 也涉及从社会支持视角出发, 或以问卷调查形式获取数据, 或以实证数据为基础, 对不同类型社会支持, 例如信息支持、情感支持等, 对持续参与作用的探究。

表 2-2 在线社区用户持续参与相关研究

| 研究   | 理论基础                 | 自变量                           | 因变量                     | 主要发现   |
|------|----------------------|-------------------------------|-------------------------|--|
| [16] | 社会支持                 | 信息支持、情感支持                     | 在线健康社区承诺                | 情感支持与持续参与正相关，信息支持与持续参与负相关。                                 |
| [64] | 社会交换理论               | 声誉提升、互惠、学习、获取知识的能力            | 在线问答社区满意度、持续贡献知识意愿      | 声誉提升、学习和获取知识的能力显著影响满意度，通过影响满意度间接影响持续贡献知识意向。                |
| [63] |                      | 社区回复                          | 在线创新社区持续参与              | 社区回复对持续参与有积极作用。  |
| [65] | 技术接受模型、理性行为理论        | 技术平台因素、内容因素、组织组成因素            | 在线健康社区持续使用              | 信赖的他人推荐、内容有用性以及大、多样、支持性社区环境与持续使用正相关。                       |
| [66] | 使用与满足理论              | 娱乐、社交、信息寻求和自我展示               | 社交媒体感知互动性、满意度、持续使用      | 娱乐、社交、信息寻求和自我展示对互动积极有影响；感知互动对满意度有积极影响；满意度、感知互动性是持续使用的决定因素。 |
| [67] | 社会支持、信息系统成功模型        | 系统质量、信息质量、服务质量，评价支持、情感支持、尊重支持 | 在线健康社区用户满意度、归属感、忠诚度     | 系统质量、信息质量对满意度有积极作用，评价、情感支持对归属感有积极作用，满意度和归属感对忠诚度有积极作用。      |
| [17] | 自我决定理论               | 自主、能力、关系的满足                   | 在线健康社区持续使用              | 自主、能力、关系的满足促进持续使用，其满足机制是寻求、提供支持和支持交换等方式。                   |
| [38] | 社会支持                 | 寻求、提供、接收信息、情感支持，陪伴支持          | 在线健康社区持续参与              | 提供信息支持与持续参与正相关，寻求、接收信息支持与其负相关；陪伴支持与其正相关；寻求情感支持与其正相关        |
| [68] | 期望确认理论和技术接受模型（采纳后模型） | 感知有用性、感知易用性、期望确认              | 健康软件满意度、持续使用意愿          | 感知有用性、感知易用性、期望确认和满意度与持续使用意愿密切相关。                           |
| [69] | 技术接受模型、期望确认理论        | 感知易用性、感知有用性、期望确认、感知愉悦、感知信任    | 在线问答社区用户满意度、持续使用意向      | 感知愉悦和感知信任对持续使用行为正向影响最大，期望确认和感知易用性次之，感知有用性影响较小。             |
| [70] | 期望确认理论和技术接受模型（采纳后模型） | 期望确认、感知有用性、感知交互性、感知的风险        | 在线医疗社区满意度、持续使用意愿、持续使用行为 | 绩效期望的确认与感知互动性会显著影响满意度，进而影响持续使用意愿。                          |

|      |        |                                      |                   |                                     |
|------|--------|--------------------------------------|-------------------|-------------------------------------|
| [32] | 社会交换理论 | 收益：情感支持、信息支持、归属支持；<br>成本：时间经历、知识权力丧失 | 在线健康社区信任、持续知识贡献意愿 | 持续知识贡献意愿源于对社区的信任，情感支持和归属支持对信任有积极影响。 |
|------|--------|--------------------------------------|-------------------|-------------------------------------|

期望确认理论最初是营销学领域用来研究消费者购买后的行为，它认为个体消费后的行为主要取决于其最初使用物品或服务后期望得到确认的程度，在信息系统领域也被用来研究用户采纳后行为。该理论主要关注期望确认和满意度之间的关系，强调期望确认和满意之间的正相关，用户期望确认水平越高，其满意度越高，持续参与的意愿也会随之增加<sup>[68][69][70]</sup>。期望确认理论的模型如图 2-9 所示。

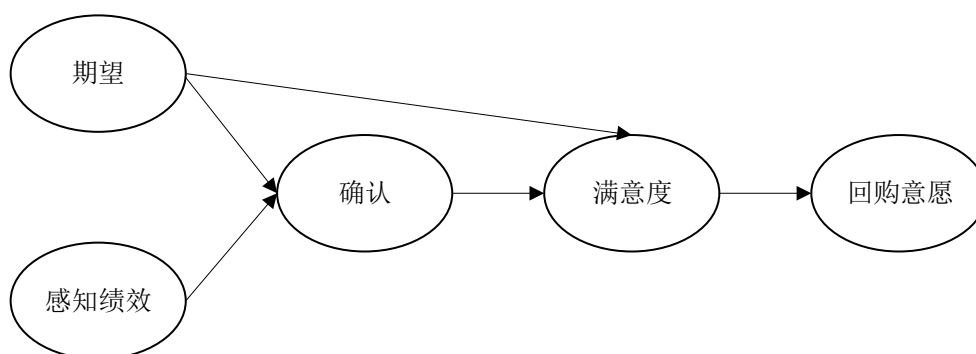


图 2-9 期望确认理论模型图

技术接受模型主要关注感知易用性和感知有用性对信息系统使用相关态度的影响，感知易用性指用户使用信息系统所要付出的努力程度，感知有用性指用户使用信息系统所带来的工作绩效提升程度。感知易用性和感知有用性通过影响用户的态度进而影响行为意向，最终影响实际的使用行为。技术接受模型的结构如图 2-10 所示。研究表明，感知有用性和感知易用性对用户的持续使用意愿均有正向作用<sup>[68][69]</sup>。

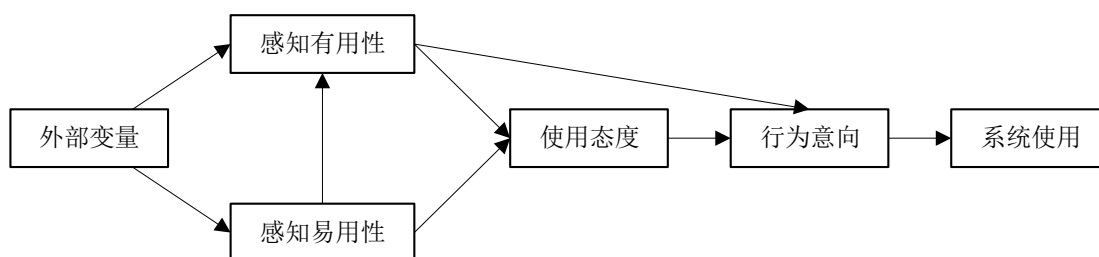


图 2-10 技术接受模型图

Bhattacharjee 结合期望确认理论和技术接受模型，提出了采纳后模型，以解释信息系统持续使用行为<sup>[71]</sup>。采纳后模型保留期望确认理论中的确认和满意度，以及结束接受模型中感知有用性，作为模型架构的组成部分，模型结构图如 2-11

所示。研究表明，期望确认、感知有用性和满意度与持续使用意愿之间的关系密切<sup>[68]</sup>。

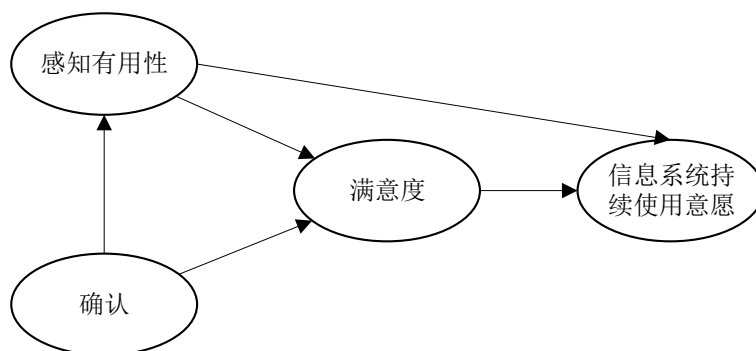


图 2-11 采纳后模型图

社会交换理论将人的社会活动看作交换行为，重视成本和收益的比较，人们实施交换行为的原则是成本最小化和收益最大化，降低成本和提高收益能够促进用户的参与行为<sup>[32]</sup>。例如，用户从社区参与中所获得收益，包括声誉的提升、学习和获取知识的能力会显著影响用户的满意度，通过影响满意度间接影响用户的持续性行为意向<sup>[64]</sup>。用户从社区中获得的社会支持收益会通过信任对用户的持续性行为意向产生促进作用<sup>[32]</sup>。而用户参与社区所要付出的成本，包括时间、精力等，则会对用户的持续性行为意向产生负向作用<sup>[32]</sup>。

社会支持是一种具有帮助意图的支持性行为，支持的获得可能来自家人、朋友、邻居、组织等。在线患者社区中，支持主要是来自和自己具有相似疾病经历的同伴，认为自己有较强的社会支持的个体可能持有相对较强的自我效能感，这对个体的行为变化和变化的维持有积极影响。研究表明，社会支持对患者疾病治疗的坚持以及健康状况改善有积极作用<sup>[37]</sup>。在线健康社区中，社会支持大体上可以分为信息支持、情感支持和陪伴支持三类<sup>[40]</sup>。对于社会支持对在线社区用户持续参与行为的影响，学者们也做了一系列探索。

目前，学术界普遍认为情感支持是用户持续参与的主要动机。Wang 和 Kraut 等研究了用户的持续使用长度与其在在线癌症社区中获得的社会支持的类型和数量之间的关系，发现用户接受的情感支持越多，离开的风险就越低<sup>[16]</sup>。宋晓龙通过对比信息支持、情感支持和伴随支持对持续参与竞争活动的作用，发现情感支持对用户的持续参与有积极影响，且作用强度最大。几项定性研究也支持了这一发现，表明参与者留下的主要原因是分享眼泪和快乐，逃避社会隔离，与同伴保持友谊<sup>[41][72]</sup>。而与情感支持相比，信息支持在用户使用在线健康社区方面的作用尚未有较为统一的认知。Wang 和 Kraut 等认为，信息支持仅满足成员的短期信息需求，其对在线癌症社区用户的承诺有负向作用<sup>[16]</sup>。Wang 和 Zhao 等同样发现乳腺癌社区中用户接收到的信息支持越多，其持续参与的可能性反而越小<sup>[38]</sup>。而宋晓龙在对在线减重社区持续竞争行为的研究中则发现信息支持对用户

持续参与竞争有促进作用<sup>[11]</sup>。Massimi 等在研究中认为, 尽管用户在情感上已经准备离开, 但一些参与者还是会留下来寻求社区提供的信息支持。而且有的用户虽然一度退出社区, 但再次遇到医疗问题时又会重新活跃起来<sup>[18]</sup>。对于陪伴支持, Wang 和 Zhao 等的研究表明, 陪伴支持与用户持续参与正相关, 通过对比信息支持、情感支持和陪伴支持三者对持续参与的作用强度, 发现陪伴支持对持续参与的保护作用最强<sup>[38]</sup>。

总的来说, 虽然情感支持在用户持续参与在线健康社区方面的作用已经得到了证实, 但对于同伴支持所包含的信息支持以及陪伴支持的作用尚未形成统一的认识。对于同伴支持系统来说, 包含三种不同的角色: 支持寻求者、支持提供者和支持接收者, 支持寻求者通过在社区发帖以发起不同类型的求助, 支持提供者或基于支持寻求者的求助提供支持, 或自发帖无针对性个体地提供支持。相应地, 支持寻求者在接收到他人的支持后成为支持接收者。因此, 一个用户可能会产生三种支持行为: 寻求支持行为、提供支持行为、接收支持行为。但目前对于这三种不同支持行为对于持续参与的影响尚未探讨清楚。因此, 有必要结合不同支持类型以及不同支持行为, 探讨其对用户持续参与在线患者社区的作用。

## 第3章 在线患者社区同伴支持分类研究

本章将传统机器学习算法和深度学习算法应用于在线患者社区中的同伴支持分类。在文本特征表示部分，本研究应用 doc2vec 模型、LDA 模型和人工提取的方法来表示每个文本的特征向量，doc2vec 模型是 word2vec 模型的扩展<sup>[43]</sup>。由 word2vec 和 doc2vec 训练的单词空间模型已被证明能够捕获语义信息，并在各种自然语言处理任务中表现良好，如情感分析、新闻分类和社会支持分类等。但对于其在中文情境下的同伴支持分类还未有相关的研究，且对于 LDA、doc2vec 和人工特征相结合后的表现缺少相关探索。本研究在文本特征表示和分类模型结合部分做了新的尝试。

本章将对在线患者社区同伴支持研究的分类框架和研究设计进行详细介绍，通过对比不同分类模型的性能评估结果，为不同类型的同伴支持构建出最优的分类器。本章将按照以下组织顺序展开叙述，首先介绍本研究的分类框架，其次，介绍同伴支持分类模型构建的实现过程，最后展示并讨论不同类型同伴支持分类模型的性能评估结果。

### 3.1 分类框架

本文的在线患者社区同伴支持分类研究旨在为每一类同伴支持构造出分类性能最优的分类器，将同伴支持行为（寻求支持、提供支持、接收支持）和同伴支持类型（信息支持、情感支持、陪伴支持）进行组合，最终，同伴支持包括：寻求信息支持、寻求情感支持、寻求陪伴支持、提供信息支持、提供情感支持、提供陪伴支持和接收信息支持、接收情感支持、接收陪伴支持九种类型。

由于接收支持类型就是用户寻求支持的主题帖，即由某一用户发出讨论帖后，下方提供支持的回复帖即为所属类型，因此不必为接收支持单独构造分类器。同时，陪伴支持是对日常生活等与疾病无关问题的讨论，用户在发起陪伴支持相关话题讨论时，期待有更多的人参与讨论，既是在寻求陪伴支持，同时也为他人提供了陪伴支持，用户在他人帖子下回复的陪伴支持则代表了用户在提供陪伴支持。所以对寻求陪伴支持和提供同伴支持可以通过帖子是主题帖还是回复帖进行区分，若帖子是主题帖，则既属于寻求陪伴支持，也属于提供陪伴支持，若帖子是回复帖，则属于提供陪伴支持。所以，在同伴支持分类研究中，将为寻求信息支持、寻求情感支持、提供信息支持、提供情感支持和陪伴支持五种类型分别建立分类器。

同伴支持分类框架如图 3-1 所示，共包含五层：数据采集、数据预处理、文本特征表示、文本分类和性能评估。首先，从在线患者社区中采集文本信息，对

获取到的文本信息进行文本分词和去停用词预处理操作，并从这些预处理后的文本数据中通过训练或人工提取得到 doc2vec 的词向量表示、LDA 文档主题向量和语言特征向量。doc2vec 是词粒度层面的特征表示，LDA 是文本粒度层面的特征表示，doc2vec 词向量、LDA 文档主题向量和语言特征向量的结合可以表达更丰富的语句特征。由于传统机器学习算法对特征工程有较高的要求，因此上述三种特征向量都将作为传统机器学习分类模型的输入，基础的 doc2vec 词向量将作为深度学习分类模型的输入。最终，将通过 Accuracy、F1 分数和 AUC 对分类性能的测量来评估不同分类训练算法的性能。



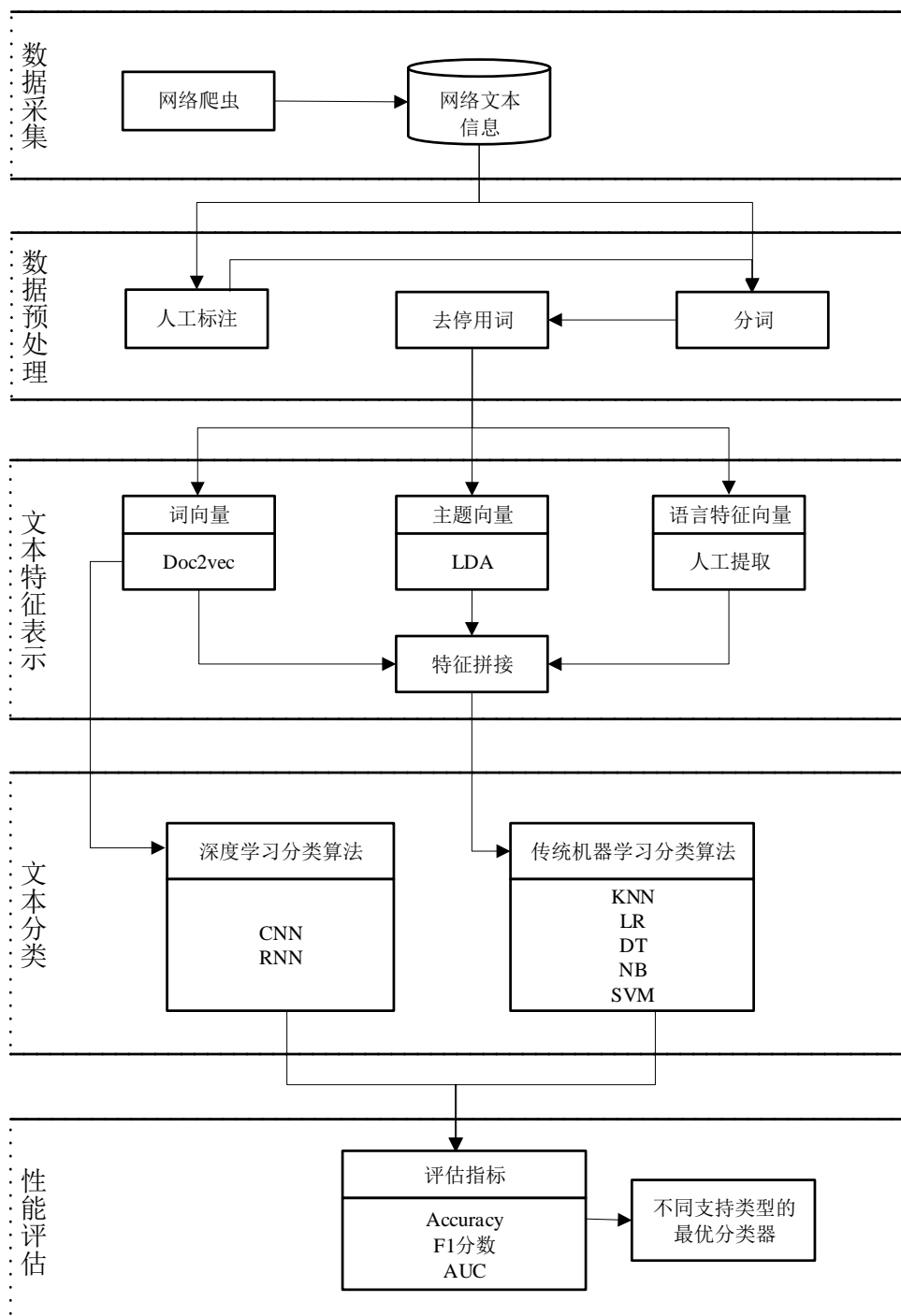


图 3-1 在线患者社区同伴支持分类框架

## 3.2 研究设计

### 3.2.1 数据采集

本文的实验数据来自于国内最大的糖尿病网站“甜蜜家园” (<http://bbs.tnbz.com/>)，该网站为糖尿病患者提供了一个咨询交流的平台，用户

在注册后可以通过发帖、加好友等方式进行交流。网站主要包括疾病相关的糖尿病知识交流专区和非疾病相关的糖友交流区、糖友生活区。由于本研究是同伴支持对持续参与影响研究的前置研究，所以从用户出发，对每个用户的所有发帖数据进行采集。通过 python 编写网络爬虫对该网站上的 2009-2019 年间 298375 个用户的发帖信息进行采集，共采集到 134589 个主题帖，主题帖下方的回复帖 1168103 个，共计 1302692 个帖子。采集字段和数据示例如表 3-1 所示，数据包含帖子 ID、用户 ID、发帖时间和帖子内容四个 ID，其中主题帖和其回复贴共享一个帖子 ID，用户 ID 即发帖者身份的唯一标识。在本研究中，我们所要使用的文本分类数据是帖子内容这一字段所包含的数据。

表 3-1 数据采集结果示例

| 帖子 ID  | 用户 ID  | 发帖时间              | 帖子内容               |
|--------|--------|-------------------|--------------------|
| 365892 | 188445 | 2014-1-1 08:07:40 | 新的一年，让我们一起努力，万马奔腾。 |
| 365892 | 227546 | 2014-1-1 08:19:33 | 新年快乐,,,,,          |

### 3.2.2 数据预处理

在进行文本特征表示前，要先对文本数据进行预处理，对不规则的成分进行筛选和过滤，以方便后续的文本的特征表示。在数据预处理阶段，主要包含三个操作，一是文本分词，二是去除停用词，三是人工标注。研究表明，词粒度的特征表示优于字粒度的特征表示，字粒度的特征表示忽略了文本的词序信息和上下文信息，因此，在进行文本特征表示前，有必要对文本进行分词处理。停用词(Stop Words)是对文本分类结果无意义的词，通常是一些高频的代词、介词或特殊符号等，需要对这些词进行过滤。同时，由于本研究的目标是实现同伴支持分类，为了训练文本分类算法，首先要对文本所属的同伴支持类型进行标注。

对于分词处理，本研究选择开源分词工具结巴(jieba)分词对文本数据进行分词。结巴分词包含三种模式：精确模式、全模式和搜索引擎模式，精确模式尝试将句子精确地切分开，比较适合文本分析任务；全模式将句子中所有可以组成词语的词扫描出来；搜索引擎模式是在精确分词的基础上，对长词再次进行切分，比较适合搜索引擎，比较适合搜索引擎分词。由于本研究是对在线患者社区中的用户发帖文本进行分类，因此本研究选用精确模式来实现对文本的分词。

对于去除停用词，本研究维护了一个包含 1890 个停用词的停用词表，包含一些无意义的介词、代词和特殊符号等。通过停用词和文本中分词后词语的比对，过滤掉文本中的停用词。

在人工标注阶段，本研究从 1302692 个数据集中随机抽取了 2668 个帖子作为人工标注的数据集，邀请在线健康领域内三位专家对样本数据进行标注。在标注开始前，每位专家会收到一份指导性文档，文档描述了标注任务要求，以及上述五类同伴支持的定义和具体示例（具体内容见附录 1）。在接受培训指导后，三位专家根据指导文档和自身的领域知识，对 2668 个帖子进行标注，决定每个帖子属于哪一类或哪几类同伴支持。例如，一个帖子可能既属于寻求信息支持，也属于寻求情感支持。最后，根据三位专家的标注结果决定帖子所属的支持类型。具体而言，对于某一个帖子，如果三位专家标注结果一致，则取三位专家共同的标注结果，如果三位专家标注不一致，则取多数标注类型作为该贴子的最终标注结果。使用 Fleiss Kappa 系数对数据标注结果一致性进行检验，Fleiss Kappa 系数为 0.87，表明数据标注结果达到了较高的一致性。最终的标注结果如表 3-2 所示，人工标注数据将作为分类模型的训练集和测试集。

表 3-2 人工标注结果

|      | 寻求信息支持 | 寻求情感支持 | 提供信息支持 | 提供情感支持 | 陪伴支持 |
|------|--------|--------|--------|--------|------|
| 帖子数量 | 1046   | 345    | 1219   | 298    | 921  |
| 占比   | 39%    | 13%    | 46%    | 11%    | 35%  |

### 3.2.3 文本特征表示

文本特征表示是对文本的向量化建模，可以用来表示文本中表征性强的价值特征，有助于提高文本分类的效果。本研究的文本特征表示包含三个部分：词向量表示、主题向量表示和语言特征向量表示。

文本的词向量表示从词粒度层面出发，挖掘文本的词义表达。本文选择 doc2vec 中的 DBOW 模型进行训练，学习得到词向量的表示方法。doc2vec 是 word2vec 在文档级别上的延伸，相较于 word2vec 取文档中词语的词向量均值作为文档的向量表示，doc2vec 直接得出的文档向量表示方式更恰当。之所以选择 DBOW 方法，是因为用户发帖数据多为短文本，上下文信息缺失，DM 模型利用上下文预测词语出现概率的方式显然不适合，而 DBOW 模型采用跳跃组合的方式的学习词向量的表示规则，更适合短文本的需要<sup>[73]</sup>。本研究将所有帖子数据作为训练集，将向量维度设置为 200，对 DBOW 模型进行训练优化。DBOW 模型不仅可以训练出词向量，也可以训练出句子向量。对于新的句子文本，其向量表示可以通过预测得出，第  $i$  个文档的向量表示形式为  $D_{(DBOW,i)} = (w_{i,1}, w_{i,2}, \dots, w_{i,200})$ 。

文本的主题向量表示从文本粒度层面出发,挖掘文本深层次的语义特征。本文采用 LDA 主题模型进行训练,得到文本在主题上的分布式向量表示。对于 LDA 主题模型的训练,比较重要的是确定主题数  $K$ 。本文采用主题相关度 (topic coherence) 和困惑度 (perplexity) 作为确定主题数的标准,主题相关度通过测量主题中高权重词语之间的语义相似度来对单个主题评分,这些度量有助于区分语义上可解释的主题和统计推断的主题<sup>[74]</sup>。研究表明,主题相关度和人工判断结果非常吻合。其表达式如下:

$$coherence(V) = \sum_{(v_i, v_j) \in V} \log \frac{D(v_i, v_j) + \epsilon}{D(v_j)}$$

其中,  $V = (V_1, V_2, \dots, V_n)$  表示某一主题中主题词的集合,  $v_i$  表示  $V$  中的第  $i$  个主题词,  $D(x, y)$  表示包含词  $x$  和  $y$  的文档数量,  $D(x)$  表示包含  $x$  的文档数量,  $\epsilon$  表示保证计算结果返回实数的平滑因子。主题相关度越高,表示 LDA 模型训练得到的主题解释性越强,语义更连贯,也更有意义。

困惑度表示对于一篇文档  $d$ , LDA 主题模型对于文档  $d$  所属主题的不确定程度,困惑度越小,表示聚类效果越好,一般主题数越多,困惑度越小,但也容易导致过拟合。困惑度的计算公式如下:

$$perplexity(D) = \exp\left(-\frac{\sum \log p(w)}{\sum_{d=1}^m N_d}\right)$$

其中,  $N_d$  表示文档  $d$  包含的词语个数,所以分母部分表示测试集中所有词语的个数之和。 $p(w)$  指测试集中每个词语出现的概率,其计算公式为:

$$p(w) = p(z|d) * p(w|z)$$

$p(z|d)$  表示某个文档在  $z$  中每个主题上出现的概率,  $p(w|z)$  表示某个主题中每个词语出现的概率。

本文从原始数据集中随机抽取 13233 个文本数据作为 LDA 模型训练的数据集,计算主题相关度以确定最佳的主题数  $K$ ,主题数—主题相关度曲线以及主题数—困惑度曲线分别如图 3-2 和图 3-3 所示。从图 3-2 中可以看出,主题相关度随着主题数的变化而发生变化。随着主题数的增加,主题相关度在一开始逐渐增大,在  $K=45$  时达到最大值,此后虽有下降,但总体取值稳定在 0.50~0.55 之间。根据图 3-3,困惑度随着主题数的增加而不断下降,且下降趋势逐渐变缓,在主题数为 45 时困惑度虽然并未达到最低,但下降趋势已经逐渐缓慢。因此,结合主题相关度和困惑度曲线,最佳主题数取值为 45。

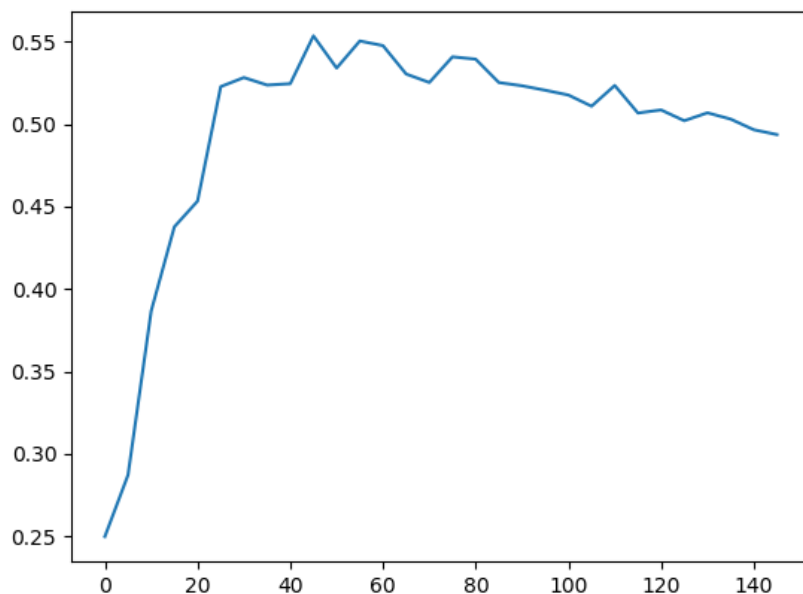


图 3-2 主题数—主题相关度曲线

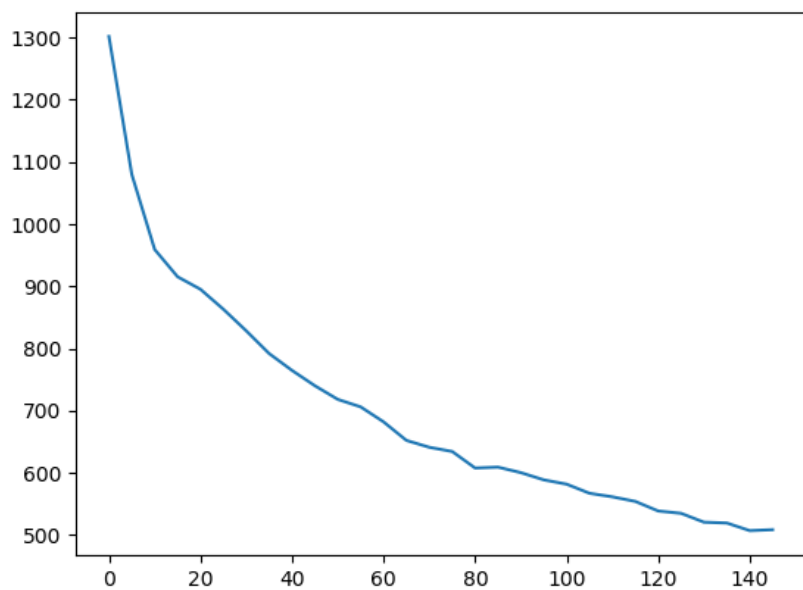


图 3-3 主题数—困惑度曲线

在确定最佳主题数  $K=45$  后，将  $K=45$  代入 LDA 主题模型进行主题聚类，生成 50 个主题的特征词及其对应的权重。在 LDA 模型生成后，对于文本  $i$ ，可以利用加载出的 LDA 模型预测出该文本属于每个主题的概率，从而生成文本—主题概率向量  $D_{(LDA,i)} = (z_{i,1}, z_{i,2}, \dots, z_{i,45})$ 。

语言特征向量主要是从文本表达的层面出发，在此分类任务情境下，经过观察不同种类同伴支持之间的特点所得到的。本文经过人工筛选共抽取出了三个语言特征向量，分别是是否包含“？”、是否包含“！”和帖子的类型：原始帖或回复

贴。若文本中包含“?”，则该文本更倾向于在寻求支持；若文本中包含“!”，则该文本的情感色彩更强烈，更倾向于情感支持；若文本属于原始贴，即用户主动发起的帖子，这类帖子更多的是在寻求支持或是离题类的陪伴支持，若文本属于回复贴，则这类帖子更多的是在提供支持或是离题类的陪伴支持。上述三种特征的向量化表示均采用 0-1 的编码方式，若包含，则标为 1，反之则为 0，若文本属于原始贴，则标为 0，原始贴则标为 1。通过对上述三类特征的向量化表示，可以得到文本*i*的语言特征向量 $D_{(LF,i)} = (c_{i,1}, c_{i,2}, c_{i,3})$ 。

最终，对 DBOW 词向量、LDA 主题向量和语言特征向量进行向量拼接，得到新的特征向量表示 $D_{(NEW,i)}$ 。

$$\begin{aligned} D_{(LF,i)} &= D_{(DBOW,i)} \oplus D_{(LDA,i)} \oplus D_{(LF,i)} \\ &= (w_{i,1}, w_{i,2}, \dots, w_{i,200}, z_{i,1}, z_{i,2}, \dots, z_{i,45}, c_{i,1}, c_{i,2}, c_{i,3}) \end{aligned}$$

$D_{(NEW,i)}$ 的维度为上述三类向量维度之和 248，对上述三种向量拼接有助于提高分类的性能。

### 3.2.4 文本分类模型

在处理完上述所有步骤后，为寻求信息支持（SIS）、寻求情感支持（SES）、提供信息支持（PIS）、提供情感支持（SES）、陪伴支持（COM）五类同伴支持分别建立其专属的分类器。由于存在一个文本属于多种同伴支持的情况，只采用一个多分类器会对结果造成影响，因此本研究将其转化为二元分类问题，为每类同伴支持建立二元分类器。

为了确定每类同伴支持的最优分类器，在文本分类器的选择上，本文既选取了传统的机器学习分类模型，也选取了深度学习模型。文中的传统机器学习分类模型包括 k 近邻模型（k-Nearest Neighbor, KNN）、逻辑回归模型（Logistics Regression, LR）、决策树模型（Decision Tree, DT）、朴素贝叶斯模型（Naïve Bayes, NB）和支持向量机模型（Support Vector Machine, SVM）。在支持向量机模型中，选择基于径向基的核函数和线性核函数作为训练模型。深度学习分类模型包括卷积神经网络模型（Convolutional Neural Networks, CNN）和循环神经网络模型（Recurrent Neural Networks, RNN），RNN 模型的核选择 GRU 核。

在分类器的输入数据方面，传统的机器学习模型和深度学习模型的输入数据有所不同，传统机器学习模型的输入数据为词向量、主题向量和语言特征向量拼接后所形成的新的向量，而深度学习模型则是将训练好的词向量作为输入，并在模型训练过程中不断迭代优化。最终所形成的分类模型如表 3-3 所示。

表 3-3 同伴支持分类模型

| 文本特征表示                     | 分类器 |
|----------------------------|-----|
| DBOW 词向量、LDA 主题向量和语言特征向量拼接 | KNN |
|                            | LR  |
|                            | DT  |
|                            | NB  |
|                            | SVM |
| DBOW 词向量                   | CNN |
|                            | RNN |

本文选取准确率、F1 分数和 AUC 三个指标作为文本分类性能评估的指标。下面将以寻求信息支持分类器的建立为例，对分类模型的建立进行简单的概述。首先，将人工标注后产生的数据集按照 8:2 的比例拆分为训练集和测试集，由于寻求情感支持和提供情感支持的样本数较少，所以在进行模型训练时采用上采样的方法增加样本数量，降低数据的不平衡性。其次，利用训练好的 DBOW 模型和 LDA 模型，分别为训练集和测试集构造词向量表示和主题向量表示，同时根据语言特征向量构建规则，产生语言特征向量。最后将产生的文本特征表示向量作为分类器的输入数据来进行分类器的训练和优化，并利用 Accuracy、F1 分数和 AUC 对模型的性能进行评估。

### 3.3 研究结果

通过对模型参数的调整和优化，得到每类同伴支持上每种分类模型的性能评估结果如表 3-4 所示。根据表 3-4，CNN 在 SIS、SES、PIS、PES 和 COM 五种同伴支持上的性能评估结果都相对优于其他分类器。对于 SIS，CNN 在三个评价指标上均取得了最高值，在传统机器学习算法中，SVM\_L SVC（线性核的支持向量机）表现最佳；对于 SES，CNN 在 Accuracy 评价指标上虽然值不是最高的，但其在 AUC 评价指标上表现却是最优的，在传统机器学习算法中，表现相对较好的也是 SVM\_L SVC；对于 PIS，性能表现最佳的同样是 CNN 算法，CNN 算法在三个评价指标上都取得了最大值，在传统机器学习算法中，表现最佳的是 NB 算法；对于 PES，总体表现最优的是 CNN，和 SES 相同，CNN 在 Accuracy 和 F1 两个评价指标上虽然得分不是最高，但在 AUC 这一评价指标上表现最好，在传统机器学习中，分类效果整体表现较好的是 SVM\_L SVC；对于 COM，CNN 同样表现最佳，在三个评价指标上均取得了较好的结果，在传统机器学习中，表现比较好的是 KNN。

表 3-4 同伴支持分类性能评估结果

|          |          | SIS         | SES         | PIS         | PES         | COM         |
|----------|----------|-------------|-------------|-------------|-------------|-------------|
| CNN      | Accuracy | <b>0.86</b> | <b>0.89</b> | <b>0.74</b> | <b>0.89</b> | <b>0.83</b> |
|          | F1       | <b>0.85</b> | <b>0.88</b> | <b>0.74</b> | <b>0.88</b> | <b>0.83</b> |
|          | AUC      | <b>0.84</b> | <b>0.65</b> | <b>0.74</b> | <b>0.68</b> | <b>0.80</b> |
| RNN      | Accuracy | 0.76        | 0.82        | 0.71        | 0.88        | 0.74        |
|          | F1       | 0.76        | 0.84        | 0.71        | 0.89        | 0.75        |
|          | AUC      | 0.71        | 0.58        | 0.68        | 0.62        | 0.73        |
| KNN      | Accuracy | 0.82        | 0.91        | 0.68        | 0.91        | 0.71        |
|          | F1       | 0.80        | 0.87        | 0.67        | 0.88        | 0.69        |
|          | AUC      | 0.72        | 0.53        | 0.63        | 0.50        | 0.60        |
| DTC      | Accuracy | 0.75        | 0.82        | 0.68        | 0.84        | 0.68        |
|          | F1       | 0.75        | 0.83        | 0.66        | 0.82        | 0.67        |
|          | AUC      | 0.70        | 0.52        | 0.64        | 0.53        | 0.60        |
| LR       | Accuracy | 0.83        | 0.90        | 0.67        | 0.91        | 0.62        |
|          | F1       | 0.82        | 0.87        | 0.67        | 0.89        | 0.64        |
|          | AUC      | 0.73        | 0.53        | 0.65        | 0.51        | 0.62        |
| NB       | Accuracy | 0.67        | 0.69        | 0.69        | 0.61        | 0.64        |
|          | F1       | 0.68        | 0.75        | 0.69        | 0.71        | 0.65        |
|          | AUC      | 0.61        | 0.55        | 0.70        | 0.57        | 0.62        |
| SVM_SVC  | Accuracy | 0.80        | 0.90        | 0.69        | 0.93        | 0.72        |
|          | F1       | 0.79        | 0.86        | 0.48        | 0.89        | 0.63        |
|          | AUC      | 0.72        | 0.50        | 0.50        | 0.5         | 0.52        |
| SVM_LSVC | Accuracy | 0.83        | 0.90        | 0.66        | 0.86        | 0.70        |
|          | F1       | 0.83        | 0.86        | 0.66        | 0.87        | 0.65        |
|          | AUC      | 0.79        | 0.56        | 0.70        | 0.57        | 0.55        |

在得到每类同伴支持的最佳分类模型后，对采集到文本数据进行预测，所得到的预测结果如表 3-5 所示，其中帖子数量最多的类型的是提供信息支持，其次是陪伴支持，最少的是寻求情感支持。用户主动发起的帖子多是在寻求支持，而回复贴多是在提供支持，所以提供的数量显然会多于需求支持的数量，这和本文的预测结果相一致。同时，从预测结果可以看出，在线患者社区内部，用户发帖涉及更多的还是信息支持，多是讨论病情相关的信息，而较少涉及情感方面的表达，寻求情感安慰的需求较少。这和其他疾病类型的在线患者社区有所不同，例如，在抑郁症论坛和癌症论坛中，表达情感支持的帖子数量多于表达信息支持的帖子数量<sup>[42]</sup>。在糖尿病在线患者社区中，除疾病信息的交流外，用户更多的是交流一些和糖尿病无关的离题信息，例如，分享生活日常：美食、风景等，或发一些节日祝福等，通过这种和谐的交流互动提升个人在社区内部的活跃度。



表 3-5 同伴支持分类预测结果

|      | 寻求信息支持 | 寻求情感支持 | 提供信息支持 | 提供情感支持 | 陪伴支持   |
|------|--------|--------|--------|--------|--------|
| 帖子数量 | 246221 | 60056  | 755923 | 186992 | 450599 |
| 占比   | 19%    | 5%     | 58%    | 14%    | 35%    |

### 3.4 本章小结

本章首先构建了在线患者社区同伴支持分类的分类框架，并对分类模型的构建流程做出了详细阐述，描述了如何采集数据和对采集到的数据进行人工标注、对样本数据进行预处理、文本特征表示以及文本分类模型选择、性能评估指标选择的详细过程。最终，通过实验结果表明，CNN 在五种类型的同伴支持分类中均表现良好，能够取得较高的准确性，且分类性能稳定。导致 CNN 算法性能优于传统机器学习算法的原因是传统机器学习算法高度依赖特征工程，用户发帖内容及表达方式的多样性，导致准确提取各类同伴支持特征困难，而 CNN 算法可以实现特征的自动提取，并在学习过程中不断优化特征的表达。章节最后还利用训练好的分类模型对采集到的帖子文本进行了预测，发现发帖最多的同伴支持类型是提供信息支持，其次是陪伴支持。

## 第4章 在线患者社区用户同伴支持对持续参与的影响分析

在线患者社区聚集了大量拥有共同兴趣或相似疾病的用户，在在线患者社区平台上，这些用户可以提问或回答问题、分享个人经验、提供或接受支持，以及交换医疗保健知识<sup>[75]</sup>。研究表明，用户获益都与持续参与社区活动有关，但在实践中，大量在线患者社区用户随着时间的推移慢慢退出，或是社区活动参与慢慢减少。实证研究显示，大量参与者会在他们有可能为团体贡献利益或获得利益之前退出参与。这种用户的高度流失，不仅使得流失用户未能从在线患者社区中获得预期的益处，对社区内其他用户也会产生一定影响，同时也为在线患者社区的发展带来了严重阻碍<sup>[76]</sup>。因此，如何使得用户持续参与在线患者社区是一项巨大的挑战，探讨在线患者社区内部用户持续参与的相关影响因素可以为解决上述问题提供可行的方案。

在线患者社区中，用户可以通过发帖、回帖和交友的方式进行社交互动。在社交互动过程中，用户可以向其同伴寻求支持，也可以为他人提供支持，自身也可能接收到来自他人的支持，支持类型包括健康知识和建议相关的信息支持、表达情绪相关的情感支持<sup>[1]</sup>和与疾病无关的陪伴支持（也称离题支持）<sup>[15]</sup>。这种由相似疾病或经历的用户产生的相互支持称之为同伴支持，同伴支持不仅可以促进成员之间信任的建立，也会提升用户对社区的归属感，从而增强用户持续参与的可能性。至于同伴支持是如何对在线患者社区用户的持续参与产生影响，不同类型的同伴支持行为对用户的持续参与影响是否有差异，学术界尚未有统一的认知。

为分析不同类型的同伴支持对在线患者社区用户持续参与的影响，本文以国内最大的糖尿病在线患者社区为数据收集来源，从中采集用户的基础资料数据及发帖数据，基于研究一所建立的分类模型，识别用户发帖信息所属的支持类型，并利用Cox比例风险模型评估不同类型同伴支持对用户流失这一事件的重要性。本章将按照以下组织顺序展开叙述，首先根据现有理论和研究提出本研究的研究模型和假设；其次，介绍本研究的研究设计；接下来，对研究结果进行讨论；最后，得出研究结论，并探讨研究意义、研究的局限性和未来的研究方向。

### 4.1 研究模型和假设

图4-1展示了文章的研究模型，模型将同伴支持和社区持续参与联系起来，重点探讨在线患者社区中同伴支持行为（寻求支持、提供支持、接收支持）对用户持续参与的作用，以及不同类型的同伴支持（信息支持、情感支持、陪伴支持）在同伴支持行为和用户持续参与关系之间的作用。

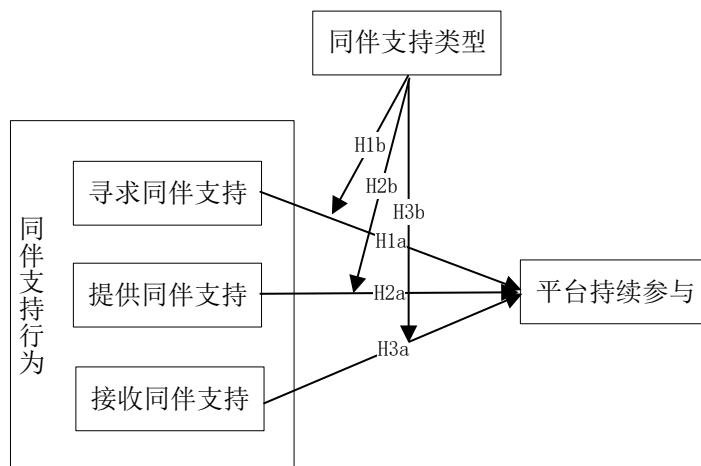


图 4-1 研究模型

### 4.1.1 寻求同伴支持和持续参与

在线患者社区聚集了大量有相似苦难和忧虑的病患用户，寻求和交换同伴支持已经成为社区内最普遍的活动，支持交换进一步存成了庞大的寻求和共享支持关系网络<sup>[2]</sup>。拥有更多寻求支持行为的用户，更容易和其他用户产生互动，建立寻求—帮助的同伴支持关系网络<sup>[77]</sup>。在同伴支持关系网络中接收到更多来自同伴的知识、经验、情感等方面的帮助，增强自身健康管理的能力，疏散心理情感方面的压力，使得用户所感知到的收益增加。根据 Mo 和 Coulson 的研究，用户感知到的益处与在线患者社区的持续使用相关<sup>[78]</sup>。因此，用户的寻求支持行为越多，其在社区持续参与的可能性越高。另一方面，由于拥有更多寻求支持行为的用户更易形成同伴支持关系网络，关系的形成伴随着互惠性原则，意味着用户接收到更多同伴支持的同时也会在未来向他人提供帮助支持，从而促成用户在社区内部的持续参与。

研究表明，用户的持续性行为与用户初次参与后所浮现出的因素有关<sup>[62]</sup>。用户前一次寻求支持的体验对用户是否继续寻求支持将会产生一定影响，在多数情况下，只有用户对之前的寻求支持行为所接收到的帮助感到满意，用户才会有下次的寻求支持行为。所以，寻求支持行为数量的多少在一定程度上代表了用户对寻求支持所接收到的帮助的满意程度，更多的寻求支持行为表示用户对其寻求支持所获得帮助的更高满意度。根据信息系统持续使用理论，满意度会对用户的持续使用意愿产生正向作用，而持续使用意向同时也会对用户的实际持续使用行为产生影响。因此，从这个角度来看，用户的寻求支持行为越多，表示用户对在该在线患者社区的满意度越高，更容易产生持续参与行为。根据 Yang 等的研究，愿意分享自身经验以寻求支持或使他人受益的成员拥有更高的生存概率<sup>[80]</sup>。因此，本文提出以下假设：

**H1a:** 在线患者社区中, 用户的寻求同伴支持行为越多, 其持续参与的可能性越高。

在寻求支持中, 包含三种不同的类型: 寻求信息支持、寻求情感支持、寻求陪伴支持。用户一般会在遇到疾病上的专业问题时发帖寻求信息支持, 期望他人能够提供相关的知识解答。研究表明, 信息支持对用户持续参与具有促进作用<sup>[11][16]</sup>。对于情感支持, 它产生的情境一般是用户初期确诊或是病情恶化等带来了消极情绪, 从而急于向他人倾诉自己的不幸, 获取他人的同情和关怀, 寻求心理上的安慰; 也可能是用户健康状况得到改善等带来了积极情绪, 向他人分享自己的喜悦, 寻求鼓励或称赞。以往研究显示, 情感支持能够增强用户和其他成员之间的关系, 促进用户的持续参与<sup>[16]</sup>。陪伴支持则多是用户发起的关于某一话题的闲聊, 或是对日常生活发生的某件事的分享, 和情感支持相同, 陪伴支持同样可以促进成员之间关系的建立<sup>[79]</sup>, 且有研究发现陪伴支持对于保持用户持续参与的作用更强<sup>[38]</sup>。

对于伴随终生的慢性疾病, 患者需要长期坚持治疗, 在这个过程中可能会遇到较多的专业性问题, 寻求信息支持为问题的解决提供了一条有效的路径, 寻求信息支持占比的提升在一定程度上表明用户对其问题解决的高满意度, 更能体现用户对社区专业性的认可。另外, 结合上一章的研究发现, 糖尿病在线患者社区中信息支持的帖子最多, 其次是陪伴支持, 可以合理推测多数用户参与在线患者社区的主要目的是解决疾病治疗过程中遇到的专业性问题以及寻求陪伴支持, 且用户在寻求信息支持和接收信息支持过程中也能感受到其他同伴的陪伴, 而寻求情感支持则是其参与的次要动机。因此, 本文提出以下假设:

**H1b:** 在线患者社区中, 用户寻求信息支持和寻求陪伴支持行为对持续参与的作用大于寻求情感支持行为对持续参与的作用。

#### 4.1.2 提供同伴支持和持续参与

在线患者社区成功的一个重要因素是有活力, 即社区内用户积极参与讨论, 不仅需要积极的话题发起者, 也需要活跃的话题讨论者。可以说, 活跃的个体成员共同定义了在线社区的成功。在线患者社区中, 不仅有寻求支持的患者和家庭成员, 也有提供支持的个体。用户既可以利用发帖功能分享健康管理知识或经验提供支持, 也可以在他人主题帖下有针对性地回帖以帮助。通过提供健康信息或情感等方面的支持, 成员和其他用户之间建立支持提供网络关系。用户为其他个体提供的支持数量越多, 被他人认可、重视和支持的可能性就越大<sup>[77]</sup>, 特别是对于那些对所讨论主题提供关键见解支持的人。同时, 其对社区的贡献度也可能越高, 所积累的社会资本相应也会更高<sup>[81]</sup>, 拥有更大的权力和影响力<sup>[82]</sup>。研

究表明,社会资本对用户的持续参与意愿有显著的正向作用<sup>[83][84]</sup>。因此,用户更多的提供支持行为可能与其持续参与正相关。

另一方面,用户通过为他人提供帮助支持,也可能从他人的经历或错误中学到新的知识,提升自身解决问题的能力<sup>[77][81]</sup>。同时,为了向他人提供更好的支持,用户也可能会在提供支持之前对寻求帮助者的问题进行全面了解,加强自身有关的知识储备。因此,在为他人提供健康知识、答疑解惑的过程中,用户的自我健康管理能力可能在无形中得到提升。用户在提供支持过程中所获得的社会资本以及其自身知识储备和问题解决能力的提升,反过来又会促进用户继续提供支持,也将使其获得更多其他用户提供的支持<sup>[81]</sup>,从而提升用户的持续参与意愿。因此,本文提出以下假设:

**H2a:** 在线患者社区中,用户的提供同伴支持行为越多,其持续参与的可能性越高。

用户提供的支持同样包含提供信息支持、提供情感支持和提供陪伴支持三种。信息支持由于涉及到的疾病知识较多,所以对提供支持者的专业性会有较高要求。用户提供信息支持的数量越多,表明用户的专业性越高,对疾病有丰富的知识储备,比较有可能发展为社区内的意见领袖,有较高的声望,从而在社区内拥有较高的社会资本,不会轻易离开社区<sup>[64]</sup>。对于情感支持和陪伴支持来说,用户提供这两类支持并不需要具备较强的专业知识,提供情感支持可能只是很简单的一句形式上的话语,例如,“加油,你很棒!”,用户从中感知到的益处较少。而用户在参与某个非疾病话题的讨论,提供陪伴支持的同时,也可以感受到其他用户的陪伴支持,可以增强用户感知到的互动性,促进其持续参与意愿。因此,本文提出以下假设:

**H2b:** 在线患者社区中,用户提供信息支持和提供陪伴支持行为对持续参与的作用大于提供情感支持对持续参与的作用。

### 4.1.3 接收同伴支持和持续参与

在对在线孕产论坛的研究中,Kouper指出提供和接收建议是该社区内社交互动的重要类型<sup>[85]</sup>。对于其他在线患者社区,同样如此,只有产生支持互换,社区才能正常运转<sup>[63]</sup>。用户在社区内部发帖寻求帮助,其他用户能在理解其需求后提供相应的帮助,寻求帮助者能够接收到他人的支持,这是社区运转的理想状态。用户在社区内发帖寻求帮助时往往期望得到回应<sup>[86]</sup>,获得有效的支持。通过接收支持,用户可以从社区获得信息、情感等方面的益处,例如,健康状况或信息的评估支持,困难应对技巧的支持,心理转换的安慰支持等。当用户接收到想要的支持后,其需求就会得到满足,从而增强其持续参与的意愿<sup>[87]</sup>。研究发现,新用户首次发帖中收到支持后更容易产生第二次发帖行为<sup>[88]</sup>。

接收支持不仅可以满足用户的需求，也会提高用户对其他成员的依恋感，促进成员之间信任关系的建立<sup>[89]</sup>。研究已经证实了信任和持续使用之间的存在着正向联系<sup>[90]</sup>。此外，接收支持也会在一定程度上增强用户对社区的归属感。社会支持的一个重要产物就是归属感，研究发现，用户获得的支持会对用户的社区归属感产生正向作用<sup>[91]</sup>，用户接收到的支持越多，这种情感上的依恋可能会越深，使得其越发认同社区中的行为，认可和依赖社区中的其他成员，归属感愈加深厚<sup>[67]</sup>。社区归属感的加深有利于培养用户对社区的忠诚度，用户更容易将最初的信任和满足逐渐转变为长久的忠诚。因此，本文提出以下假设：

**H3a:** 在线患者社区中，用户的接收同伴支持行为越多，其持续参与的可能性越高。

用户接收到的支持一般是他人对其寻求支持的回应。用户在寻求信息支持后，收到来自其他用户通过分享个人经验或建议而提供的信息支持，包括身体症状和疾病治疗等方面建议<sup>[32]</sup>。信息支持不仅可以帮助用户改善健康状况<sup>[92]</sup>，还能提高用户对社区合作意识的感知水平<sup>[63]</sup>。用户获得的支持中信息支持越多，不仅对在线患者社区中成员的健康知识水平会更加信任<sup>[32]</sup>，学习和获取健康知识的能力也会越高，这两者都会对用户的持续参与意向产生影响<sup>[64]</sup>。鉴于用户参与在线患者社区的首要目的是寻求信息支持，所以，接收到的信息支持越多，用户需求得到满足的可能性越大，对社区的感知有用性以及满意度也会越高。用户在社区内获得的情感支持和陪伴支持可能比其他地方获得的更有益<sup>[16]</sup>，因为这些用户有着相似的疾病经历，更能产生同理心，能够更有效地为其他用户提供情感和陪伴上的支持。接收情感支持和陪伴支持能够促进接收支持的个体和提供支持个体关系的建立，提高用户对社区的认同感。此外，陪伴支持讨论了与健康无关的主题，包括问候和聊天等，此类帖子在社区中提供了一种友好的氛围<sup>[75]</sup>。研究表明，陪伴支持是绑定用户并保持其持续参与的一种很好的方式<sup>[38]</sup>。因此，本文提出以下假设：

**H3b:** 在线患者社区中，用户接收信息支持和接收陪伴支持行为对持续参与的作用大于接收情感支持对持续参与的作用。

## 4.2 研究设计

本章研究主要探索在线患者社区中用户同伴支持行为和持续参与之间的关系，包括不同的同伴支持行为对持续参与的作用，以及不同类型的同伴支持对用户持续参与的作用。研究的数据来源是糖尿病在线论坛“甜蜜家园”，在经过数据处理和变量构建后，采用生存分析模型 Cox 比例风险模型对假设进行验证，所使用的软件是 SPSS 26。

### 4.2.1 数据收集

本章的研究背景延续上一章的内容，仍选取的是糖尿病在线论坛“甜蜜家园”，除上一章采集到的论坛发帖数据外，还采集了用户的基本资料数据，包括用户的年龄、好友数量、注册时间、最后登录时间等，共包含 298375 条用户数据。从采集到的用户数据中随机抽取 18015 个发帖数量大于 0 的用户作为模型估计的样本数据，因为没有公开的发帖数据就无法估计用户的同伴支持数量。将这些用户的帖子信息输入到上一章建立好的分类模型中，对帖子所属的同伴支持类型进行预测，并根据用户基本资料数据和帖子预测结果进行变量构建。由于 Cox 比例风险模型是使用协变量的均值来进行参数估计，所以为了排除值过大或过小对模型结果造成影响，研究剔除了协变量值明显偏离其他样本观测值的样本。最终得到 17672 个用户的样本数据。

### 4.2.2 研究变量

在得到样本数据后，对研究模型中的涉及到的变量进行构建。研究变量包括因变量、控制变量和解释变量。

#### (1) 因变量

**参与时间：**研究的因变量是参与时间，是一个数值型变量。本研究将用户在社区第一次发帖的时间视为参与的开始时间，最后一次发帖时间视为参与的结束时间，参与时间的长度为两者之差，单位用周表示。

**是否停止参与：**停止参与是一个 0-1 二元变量。如果用户在数据收集结束前的 12 周内未发布任何帖子，则认为该用户已停止参与<sup>[16]</sup>，并将该变量标记为 1，停止参与是本研究的感兴趣事件。如果用户在数据收集结束前的 12 周内仍有帖子发布，则认为该用户仍在持续参与，并将该变量标记为 0，属于生存分析中的截尾数据，在做参数估计时，这部分样本数据会被删除。

#### (2) 控制变量

**好友数：**好友数表示用户在社区内的好友数量，是一个数值型变量，在社区内，有近 12% 的用户有好友。

### (3) 分层变量

**年龄层：**在用户的基本资料中，有些用户会提供出生日期，根据其出生日期可以计算出该用户在停止参与时的年龄，样本数据中有 11% 的用户提供了出生日期。对年龄进行编码，将年龄在 50 岁以下的编码为 0，50 岁及以上的编码为 1，年龄未知的编码为 2。由于很多样本数据缺失年龄值，所以在进行整体模型分析时，把用户所属的年龄层作为分层变量而不是控制变量输入 Cox 回归模型。

### (4) 解释变量

**寻求同伴支持：**表示用户发布的寻求同伴支持的帖子数量，是用户寻求信息支持、情感支持、陪伴支持的帖子总和。

**提供同伴支持：**表示用户发布的提供同伴支持的帖子数量，是用户提供信息支持、情感支持、陪伴支持的帖子总和。

**接收同伴支持：**表示用户接收到的同伴支持的帖子数量，是用户发起的帖子下所收到的提供同伴支持的帖子数量之和。

**寻求信息支持：**表示用户寻求信息支持的帖子数量在其寻求同伴支持帖子数量中的占比。

**寻求情感支持：**表示用户寻求情感支持的帖子数量在其寻求同伴支持帖子数量中的占比。

**寻求陪伴支持：**表示用户寻求陪伴支持的帖子数量在其寻求同伴支持帖子数量中的占比。虽然在上一章的研究中并未区分寻求陪伴支持和提供陪伴支持，但在本章的研究中，将用户发起的陪伴支持定义为寻求陪伴支持。

**提供信息支持：**表示用户提供信息支持的帖子数量在其提供同伴支持帖子数量中的占比。

**提供情感支持：**表示用户提供情感支持的帖子数量在其提供同伴支持帖子数量中的占比。

**提供陪伴支持：**表示用户提供陪伴支持的帖子数量在其提供同伴支持帖子数量中的占比。本研究将用户在他人主题帖下回复的陪伴支持定义为用户是在提供陪伴支持。

**接收信息支持：**表示用户接收到的信息支持的帖子数量在其接收同伴支持帖子数量中的占比。用户接收到的信息支持的帖子数量是用户发布的寻求同伴支持的主题帖下回复贴中提供信息支持的数量。

**接收情感支持：**表示用户接收到的情感支持的帖子数量在其接收同伴支持帖子数量中的占比。用户接收到的情感支持的帖子数量是用户发布的寻求同伴支持的主题帖下回复贴中提供情感支持的数量。



**接收陪伴支持：**表示用户接收到的陪伴支持的帖子数量在其接收同伴支持帖子数量中的占比。用户接收到的陪伴支持的帖子数量是用户发布的寻求同伴支持的主题帖下回复贴中提供陪伴支持的数量。

模型中的变量汇总及编码情况如表 4-1 所示。

表 4-1 模型中研究变量汇总

|            | 变量     | 解释                            |
|------------|--------|-------------------------------|
| <b>ST</b>  | 生存时间   | 用户首次发帖时间和最后一次发帖时间的差值          |
| <b>SPS</b> | 寻求同伴支持 | 用户寻求同伴支持的帖子总量                 |
| <b>PPS</b> | 提供同伴支持 | 用户提供同伴支持的帖子总量                 |
| <b>RPS</b> | 接收同伴支持 | 用户接收到的同伴支持帖子总量                |
| <b>SIS</b> | 寻求信息支持 | 用户寻求信息支持的帖子用户在寻求同伴支持的帖子中的占比   |
| <b>SES</b> | 寻求情感支持 | 用户寻求情感支持的帖子在寻求同伴支持的帖子中的占比     |
| <b>SCS</b> | 寻求陪伴支持 | 用户寻求陪伴支持的帖子在寻求同伴支持的帖子中的占比     |
| <b>PIS</b> | 提供信息支持 | 用户提供信息支持的帖子在提供同伴支持的帖子中的占比     |
| <b>PES</b> | 提供情感支持 | 用户提供情感支持的帖子在提供同伴支持的帖子中的占比     |
| <b>PCS</b> | 提供陪伴支持 | 用户提供陪伴支持的帖子在提供同伴支持的帖子中的占比     |
| <b>RIS</b> | 接收信息支持 | 用户收到的提供信息支持的帖子在收到的同伴支持的帖子中的占比 |
| <b>RES</b> | 接收情感支持 | 用户收到的提供情感支持的帖子在收到的同伴支持的帖子中的占比 |
| <b>RCS</b> | 接收陪伴支持 | 用户收到的提供陪伴支持的帖子在收到的同伴支持的帖子中的占比 |
| <b>FRI</b> | 好友数量   | 用户的好友数量                       |
| <b>AGE</b> | 年龄层    | 用户所属的年龄层                      |

### 4.2.3 研究方法

本章的研究采用生存分析方法进行假设验证，它是一种处理、分析生存数据的方法，又叫时间—效应分析（Time-Effect Analysis），广泛应用于医学、社会科学等领域，例如患者治疗后的生存时间分析等。在本研究中，感兴趣的事件是用户不再参与社区活动的时间，即用户不再发帖的时间，通过对这一事件的观察，可以得到用户的生存时间。

生存时间可以分为两种类型：完全数据（complete data）和截尾数据（censored data），完全数据是指观察对象在观察期内出现感兴趣的事件，此时用户的生存时间是完整的，截尾数据则是指观察对象在观察期内未出现感兴趣的事件，所记录到的生存时间信息是不完整的。对于本研究来说，在观察时刻，即数据采集时刻之前，某些用户可能还未流失，因此产生了截尾数据。本研究之所以采用生存分析方法，是因为线性回归分析没有考虑到截尾数据，会产生偏差估计。因此本研究采用 Cox 比例风险回归模型（Cox's proportional hazard regression model），简称 Cox 回归模型，来研究影响用户生存时间和生存率的因素。

Cox 回归模型主要用于慢性疾病的预后分析，但同时也是在线社区用户持续参与影响因素分析的常用手段。该方法兼有非参数法和参数法的特点，其主要优点是不需要考虑生存时间的分布特征，能够有效地利用截尾数据，是一种半参数的多因素分析方法。Cox 回归模型的基本形式如下：

$$h(t, X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)$$

其中：

(1)  $h(t, X)$  表示具有协变量  $X$  的个体在时刻  $t$  的风险率，即在  $t$  时刻存活， $t \sim t + \Delta t$  时刻内流失的概率，是一种瞬时流失率。

(2)  $h_0(t)$  是基础风险率，表示所有因素都为 0 时的流失风险率，不受解释变量的影响，假定它与  $h(t, X)$  是比例关系。它是未知的，无法对其进行参数估计，是模型的非参数部分。

(3)  $(\beta_1, \beta_2, \dots, \beta_n)$  为模型的偏回归系数，是需要进行估计的参数， $\beta_i$  表示其他协变量不变时， $X_i$  每变化一个单位，流失的风险率的自然对数变化  $\beta_i$  个单位。当  $\beta_i > 0$  时，因素  $X_i$  为危险因素；当  $\beta_i < 0$  时，因素  $X_i$  为保护因素；当  $\beta_i = 0$  时，因素  $X_i$  为无关因素。 $(\beta_1, \beta_2, \dots, \beta_n)$  可以根据样本的实际观察值，利用偏似然函数进行参数估计，是模型的参数部分。

具体到本章的研究，Cox 回归模型方程如下：

$$h(ST, X) = h_0(ST) \exp(\beta_1 * SPS + \beta_2 * PPS + \beta_3 * RPS + \beta_4 * SIS + \beta_5 * SES + \beta_6 * SCS + \beta_7 * PIS + \beta_8 * PES + \beta_9 * PCS + \beta_{10} * RIS + \beta_{11} * RES + \beta_{12} * RCS + \beta_{13} * FRI)$$

其中， $\beta_1$  到  $\beta_{13}$  为模型中需要进行估计的参数，代表了其对应变量对生存时间作用的大小。

### 4.3 研究结果

本章的研究使用 SPSS 26 对 Cox 回归模型进行参数估计。首先，对解释变量和控制变量进行标准化处理，以消除变量量纲不同所带来的偏差。表 4-2 展示了

模型中变量的描述性统计特征以及变量之间的相关性。从表 4-2 可以看到除 RPS 和 SPS 的相关系数大于 0.8 外，其他变量之间的相关系数均在 0.8 以下。因此，为了消除 RPS 和 SPS 多重共线性对模型参数估计的影响，在进行 Cox 回归分析时采用逐步回归的方法消除多重共线性的影响。

表 4-2 模型变量描述性统计和相关性

| 变量  | 均值    | 标准差    | 1        | 2        | 3       | 4        | 5        | 6        | 7        |
|-----|-------|--------|----------|----------|---------|----------|----------|----------|----------|
| SPS | 5.83  | 9.173  | 1        |          |         |          |          |          |          |
| PPS | 13.95 | 25.589 | 0.775**  | 1        |         |          |          |          |          |
| RPS | 24.02 | 41.417 | 0.856**  | 0.648**  | 1       |          |          |          |          |
| SIS | 0.51  | 0.41   | 0.251**  | 0.159**  | 0.164** | 1        |          |          |          |
| SES | 0.10  | 0.20   | 0.153**  | 0.152**  | 0.096** | -0.132** | 1        |          |          |
| SCS | 0.20  | 0.33   | -0.042** | -0.045** | 0.056** | -0.445** | -0.160** | 1        |          |
| PIS | 0.51  | 0.35   | 0.164**  | 0.128**  | 0.148** | .109**   | 0.007    | -0.292** | 1        |
| PES | 0.10  | 0.16   | 0.038**  | 0.096**  | 0.037** | -0.116** | 0.108**  | 0.020**  | -0.206** |
| PCS | 0.24  | 0.26   | 0.026**  | 0.076**  | 0.020** | -0.104** | 0.045**  | 0.021**  | -0.378** |
| RIS | 0.36  | 0.33   | 0.342**  | 0.171**  | 0.403** | 0.421**  | 0.032**  | -0.042** | 0.189**  |
| RES | 0.084 | 0.11   | 0.185**  | 0.103**  | 0.295** | -0.009   | 0.074**  | 0.349**  | -0.128** |
| RCS | 0.21  | 0.24   | 0.060**  | 0.016*   | 0.153** | -0.144** | -0.060** | 0.631**  | -0.263** |
| FRI | 0.24  | 0.91   | 0.333**  | 0.424**  | 0.260** | 0.074**  | 0.077**  | -0.035** | 0.047**  |

表4-2 (续表)

| 变量  | 均值    | 标准差  | 8        | 9        | 10      | 11      | 12     | 13 |
|-----|-------|------|----------|----------|---------|---------|--------|----|
| PES | 0.10  | 0.16 | 1        |          |         |         |        |    |
| PCS | 0.24  | 0.26 | 0.013    | 1        |         |         |        |    |
| RIS | 0.36  | 0.33 | -0.094** | -0.160** | 1       |         |        |    |
| RES | 0.084 | 0.11 | 0.142**  | -0.024** | 0.238** | 1       |        |    |
| RCS | 0.21  | 0.24 | -0.018*  | 0.076**  | 0.082** | 0.384** | 1      |    |
| FRI | 0.24  | 0.91 | 0.045**  | 0.048**  | 0.008   | 0.016*  | -0.012 | 1  |

在 0.01 级别（双尾），相关性显著。\*\*

在 0.05 级别（双尾），相关性显著。\*

研究采用逐步回归的方法进行生存分析，表 4-3 展现了 Cox 回归方法对模型参数估计的最终结果。本研究分别选取生存时间为 1 年以内、2 年以内和 3 年以内的样本数据进行参数估计。模型 1、模型 2 和模型 3 分别为生存时间为 1 年以内、2 年以内和 3 年以内的参数估计结果。

表 4-3 生存分析参数估计结果

| 变量  | 模型 1        |             | 模型 2        |             | 模型 3        |             |
|-----|-------------|-------------|-------------|-------------|-------------|-------------|
|     | 风险系数<br>(B) | 风险率<br>(HR) | 风险系数<br>(B) | 风险率<br>(HR) | 风险系数<br>(B) | 风险率<br>(HR) |
| SPS | -0.088**    | 0.916       | -0.089**    | 0.915       | -0.090**    | 0.914       |
| PPS | -0.109**    | 0.896       | -0.105**    | 0.900       | -0.095**    | 0.909       |
| RPS | -0.057**    | 0.944       | -0.050**    | 0.951       | -0.053**    | 0.948       |
| SIS | -0.233**    | 0.792       | -0.239**    | 0.788       | -0.249**    | 0.780       |
| SES | -0.111**    | 0.895       | -0.114**    | 0.892       | -0.121**    | 0.886       |
| SCS | -0.136**    | 0.873       | -0.142**    | 0.868       | -0.144**    | 0.866       |
| PIS | -0.291**    | 0.747       | -0.282**    | 0.754       | -0.294**    | 0.746       |
| PES | -0.129**    | 0.879       | -0.127**    | 0.881       | -0.126**    | 0.882       |
| PCS | -0.204**    | 0.815       | -0.204**    | 0.816       | -0.204**    | 0.815       |
| RIS | 0.044**     | 1.045       | 0.044**     | 1.045       | 0.053**     | 1.054       |
| RES | -0.004      | 0.996       | -0.004      | 0.996       | 0.005       | 1.005       |
| RCS | 0.060**     | 1.062       | 0.043**     | 1.044       | 0.041**     | 1.042       |
| FRI | -0.065**    | 0.937       | -0.065**    | 0.937       | -0.066**    | 0.936       |

在 0.01 级别（双尾），相关性显著。\*\*

在 0.05 级别（双尾），相关性显著。\*

根据表 4-3 结果中的第一行，SPS 在模型 1、模型 2 和模型 3 中的风险系数均小于 0，且显著，停止参与的风险率均小于 1（以模型 2 为例： $B = -0.089^{***}$ ,  $HR = 0.915$ ），说明寻求同伴支持是用户生存即持续参与的保护因素，用户寻求更多的同伴支持可以显著提升用户持续参与的可能性。以模型 2 为例，用户每多寻求一个标准差的同伴支持，其持续参与的可能性会增加 8.5%。因此，假设 H1a 得到支持，且随着时间的增加，寻求同伴支持的作用效果逐渐增大。

提供同伴支持（PPS）假设（H1b）提出，在线患者社区中，用户的提供同伴支持行为越多，其持续参与的可能性越高。表 4-3 结果中的第二行为该假设提供了支持，PPS 的风险系数在三个模型中均小于 0 且显著，停止参与的风险率也小于 1（以模型 2 为例： $B = -0.105^{***}$ ,  $HR = 0.900$ ）。这意味着当用户提供支持行为的数量上升时，用户持续参与的可能性也会升高，用户每多提供一个标准差的同伴支持，持续参与的可能性就会上升 10%。因此，提供同伴支持对用户的持续参与有正向作用，假设 H2a 成立，但随着时间的增加，提供同伴支持的作用效果逐渐减小。

接收同伴支持（RPS）假设（H1c）提出，在线患者社区中，用户的接收同伴支持行为越多，其持续参与的可能性越高。表 4-3 结果中的第三行支持该假设的

成立, RPS 的风险系数在三个模型中均小于 0 且显著, 停止参与的风险率也小于 1 (以模型 2 为例:  $B = -0.050^{***}$ ,  $HR = 0.951$ ), 表明用户接收到的同伴支持数量越多, 其持续参与的可能性就越高。用户每多接收到一个标准差的同伴支持, 持续参与的概率就会增加 4.9%。因此, 接收同伴支持对用户的持续参与有正向作用, 假设 H3a 成立。

假设 H1b 提出, 在线患者社区中, 用户寻求信息支持 (SIS) 和寻求陪伴支持 (SCS) 行为对持续参与的作用大于寻求情感支持 (SES) 对持续参与的作用。根据表 4-3 结果中的第四行、第五行和第六行所示, SIS、SES 和 SCS 三者的风险系数均显著小于 0, 风险率小于 1 (以模型 2 为例, SIS:  $B = -0.239^{***}$ ,  $HR = 0.788$ , SES:  $B = -0.114^{***}$ ,  $HR = 0.892$ , SCS:  $B = -0.142^{***}$ ,  $HR = 0.868$ ), SIS 和 SCS 的风险率小于 SES 的风险率。在寻求同伴支持中, 每增加一个标准差的寻求信息支持和寻求同伴支持占比, 用户持续参与的概率分别可以提升 21.2% 和 13.2%, 每增加一个标准差的寻求情感支持占比, 用户持续参与的概率可以提升 10.8%, 小于寻求信息支持和寻求同伴支持的概率提升值, 即寻求信息支持对持续参与的作用大于寻求情感支持和寻求陪伴支持对持续参与的作用。为验证这种作用差异是否显著, 以模型 2 数据为例, 本文采用 Chin (2003) 提出的方法对其进行检验<sup>[93]</sup> (具体方程见附录 2), 检验结果如表 4-4 所示, t 值均大于 1.65, 说明各个因素对持续参与的作用差异显著。因此, 假设 H1b 成立。

表 4-4 寻求同伴支持中各因素作用差异显著性检验结果

|      | 作用路径   | PC(B)    | SE    | t 值      |
|------|--------|----------|-------|----------|
| 对比 1 | SIS→ST | -0.239** | 0.011 | 1141.992 |
|      | SES→ST | -0.114** | 0.009 |          |
| 对比 2 | SIS→ST | -0.239** | 0.011 | 739.608  |
|      | SCS→ST | -0.142** | 0.013 |          |
| 对比 3 | SES→ST | -0.114** | 0.009 | 229.941  |
|      | SCS→ST | -0.142** | 0.013 |          |

假设 H2b 提出, 在线患者社区中, 用户提供信息支持 (PIS) 和提供陪伴支持 (PCS) 行为对持续参与的作用大于提供情感支持 (PES) 对持续参与的作用。根据表 4-3 结果中的第七行、第八行和第九行所示, PIS、PES 和 PCS 三者的风险系数均显著小于 0, 风险率小于 1 (以模型 2 为例, PIS:  $B = -0.282^{***}$ ,  $HR = 0.754$ , PES:  $B = -0.127^{***}$ ,  $HR = 0.881$ , SCS:  $B = -0.204^{***}$ ,  $HR = 0.816$ ), SIS 和 SCS 的风险率明显小于 SES 风险率。在提供同伴支持中, 提供信息支持占比和提供陪伴支持占比每增加一个标准差, 用户持续参与的概率分别可以提升 24.6% 和 18.4%, 每增加一个标准差的提供情感支持占比, 用户持续参与可能性增加 11.9%, 小于提供信息支持和提供陪伴支持对持续参与的贡献, 即提供信息支持和提供陪伴支持对持续参与的作用大于提供情感支持对持续参与的作用。为

验证这种作用差异是否显著，同样采用 Chin (2003) 提出的方法对其进行检验，检验结果如表 4-5 所示，t 值均大于 1.65。因此，假设 H2b 成立。

表 4-5 提供同伴支持中各因素作用差异显著性检验结果

|      | 路径     | PC(B)    | SE   | t 值      |
|------|--------|----------|------|----------|
| 对比 1 | PIS→ST | -0.282** | .010 | 1495.964 |
|      | PES→ST | -0.127** | .009 |          |
| 对比 2 | PIS→ST | -0.282** | .010 | 752.808  |
|      | PCS→ST | -0.204** | .009 |          |
| 对比 3 | PES→ST | -0.127** | .009 | 785.529  |
|      | PCS→ST | -0.204** | .009 |          |

假设 H3b 提出，在线患者社区中，用户接收信息支持和接收陪伴支持行为对持续参与的作用大于接收情感支持对持续参与的作用。根据表 4-3 结果中的第十行、第十一行和第十二行所示，RIS 和 RCS 的风险系数大于 0 且显著，风险率大于 1，RES 的风险系数虽然小于 0，但并不显著（以模型 2 为例，RIS:  $B = -0.044^{***}, HR = 1.045$ , RES:  $B = -0.004, HR = 0.996$ ,  $B = -0.043^{***}, HR = 1.044$ ），表明 RIS 和 RCS 是持续参与的危险因素，在接收同伴支持中，接收信息支持占比和接收同伴支持占比每增加一个标准差，用户持续参与的可能性分别降低 4.5%和 4.4%，而 RES 则对持续参与无显著作用，研究结果和假设不匹配。因此，假设 H3b 不成立。

除上述的解释变量外，本文还探究了控制变量 FRI（好友数）对持续参与的作用，研究结果如表 4-3 最后一行所示，FRI 的风险系数显著小于 0，风险率大于 1（以模型 2 为例，FRI:  $B = -0.065^{***}, HR = 0.937$ ）。结果表明 FRI 也是持续参与的保护因素，用户的好友数每增加 1 个标准差，其持续参与的可能性就会增加 6.3%，即用户的好友数越多，持续参与的可能性就越大。因此，好友数对持续参与有正向作用。

年龄作为 Cox 回归模型的分组变量，并未出现在参数估计结果中，不同年龄层生存时间在 2 年内的生存分析函数图如图 4-2 所示，图中年龄层 0 表示年龄在 50 岁以下，1 表示年龄在 50 岁及以上，2 表示年龄未知，可以看出中老年用户的累积生存率普遍高于年轻用户。为了验证年龄层对用户持续参与造成的差异是否显著，本文做了高低年龄组 0 和 1 的方差分析，结果显示  $sig = 0.035$ ，差异显著。因此，中老年用户的持续参与概率显著高于年轻用户的持续参与概率。相较于年轻用户，中老年用户一旦参与进在线患者社区，更不容易流失。

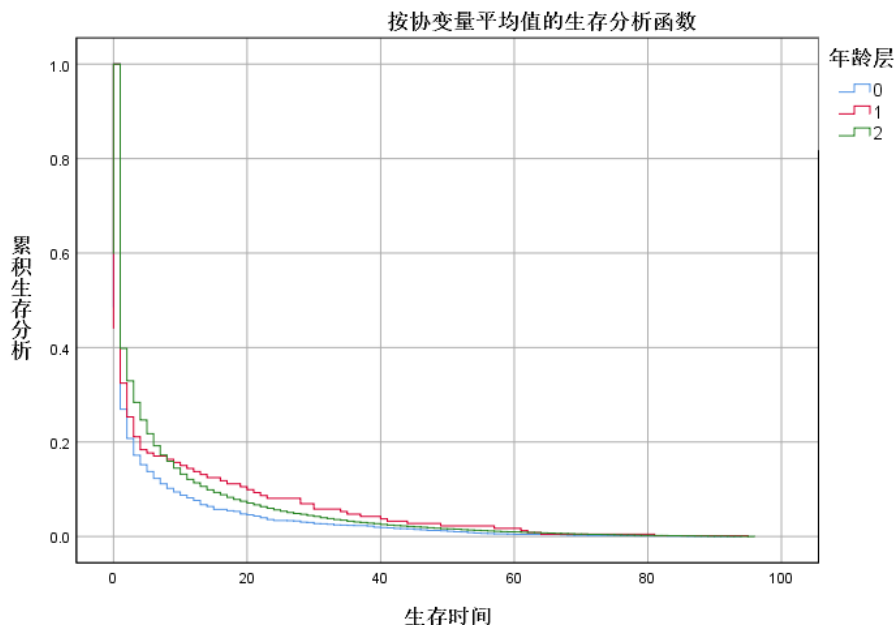


图 4-2 生存分析函数图

## 4.4 讨论

在本章研究中，调查了在线患者社区中，同伴支持行为对持续参与的作用，以及同伴支持类型在同伴支持行为和持续参与关系中所起的作用。研究假设寻求同伴支持、提供同伴支持和接收同伴支持行为均对用户的持续参与有积极的影响；且在寻求同伴支持行为中，寻求信息支持和寻求陪伴支持对用户持续参与的作用强度大于寻求情感支持对用户持续参与的作用强度；在提供同伴支持行为中，提供信息支持和提供陪伴支持对持续参与的作用大于提供情感支持对持续参与的作用；在接收同伴支持行为中，接收信息支持和接收陪伴支持对持续参与的影响大于接收情感支持对持续参与的影响。

利用从糖尿病在线患者社区收集到的数据信息，研究建立了实证研究模型来探索同伴支持行为和同伴支持类型与持续参与之间的关系，并利用 Cox 回归方法对模型中的变量进行参数估计，进而验证假设。研究结果证实了不同的同伴支持行为和持续参与之间的关系，主要有以下几个发现：

### (1) 同伴支持行为和持续参与

研究结果表明，寻求同伴支持对持续参与有正向作用，用户寻求同伴支持行为越多，用户持续参与的可能性越大，停止参与的风险越低。寻求同伴支持是用户主观能动性体现，只有用户主动寻求支持，才有可能接收到有效帮助，解决自身面临的难题，感受到社区的关怀和有用性，对社区的信任又会反向促进用户的持续参与意向<sup>[32]</sup>。

提供同伴支持同样对持续参与有积极作用,用户提供的同伴支持越多,用户持续参与的可能性越高,肯定了利他主义对持续参与的积极作用<sup>[94]</sup>。和寻求同伴支持一样,提供同伴支持也是一种主动性行为,用户通过提供支持可以获得他人的感谢和认可,满足其自我实现需求,提高声誉及社区内的社会地位,声誉和社会地位的提升能够显著提高用户的持续参与意愿,侧面印证了声誉和地位的提升与持续性行为意愿之间的相关性<sup>[64]</sup>。

接收同伴支持对持续参与也有显著的促进作用,这和 Zhang 等的研究结论一致<sup>[63]</sup>。和寻求同伴支持、提供同伴支持不同的是,接收同伴支持是一种被动性行为,用户在疾病治疗过程中可能遇到各种各样的问题,在社区内获得的健康知识、专业建议和情感慰藉等都会促进用户对社区的信任和认同,增强用户的持续参与意愿。而通过对比寻求同伴支持、提供同伴支持和接收同伴支持三者的风险系数,发现提供同伴支持对持续参与的作用强度大于其它两个,可以认为在在线患者社区中,利他主义和声誉的提升是其保持持续参与的强效驱动力。

## (2) 同伴支持类型和持续参与

研究表明了在线患者社区的寻求同伴支持中,信息支持、情感支持和陪伴支持占比的提升都能促进用户的持续参与,且寻求信息支持和寻求同伴支持占比的提升对持续参与的作用强度高于寻求情感支持。在提供同伴支持中,提供的信息支持、情感支持和陪伴支持越多,用户持续参与的可能性越高,且提供信息支持和提供陪伴支持对持续参与的作用强度高于提供情感支持。上述发现打破了以往情感支持对持续参与作用强度最大的普遍认知<sup>[11][16]</sup>,表明了慢性病在线患者社区中信息支持的重要性,支持了 Budak 和 Agrawal 的发现,即信息支持比情感支持更重要<sup>[95]</sup>。

但在接收同伴支持行为中,研究结果显示接收到的信息支持越多,用户持续参与的可能性反而越低,虽然和研究假设相悖,但却和 Wang 和 Zhao 等的研究结论相一致。和信息支持一样,陪伴支持也有同样的负向效果,而接收到的情感支持则对用户持续参与无显著作用。对于接收信息支持对持续参与的负向影响,一种可能的解释是,虽然糖尿病治疗是一项长久的工作,但不排除一部分用户只是为了解决某一个特定的问题而注册发帖寻求信息支持,对于新用户,他们更可能接收到来自其他用户的支持<sup>[81]</sup>,在接收到有用的信息支持后,这些用户可能会立即离开社区;另一种可能的解释是,由于社区内同伴用户的健康素养参差不齐,所以提供的信息支持有用性水平也有高有低。用户接收到的信息支持越多,接触到低质量信息支持的概率也就越大,这些低质量的信息支持可能会削弱用户持续参与的意愿。对于接收陪伴支持对用户持续参与的负向作用,可能的原因是陪伴支持是与疾病无关的离题信息,而对于接收陪伴支持的用户来说,更期待获得的是与其发起话题相关的支持信息,而不是无用的闲聊。



### (3) 好友数量、年龄和持续参与

最后，研究还发现好友数量对用户的持续参与有正向的作用，表明用户的社交网络关系越多，用户离开的可能性就越小。好友数量也是社会资本的一种体现，研究再次证实了社会资本在用户持续参与中发挥着重要作用。此外，研究也探索了年龄和持续参与之间的关系，和以往对中老年用户不擅长使用互联网的刻板印象有差异的是，相较于年轻用户群体，中老年用户群体更容易持续参与在线患者社区，可能在糖尿病这个慢性病领域，中老年用户相对而言拥有更多的治疗经验和相关知识，因此，他们可能会为其他同伴提供较多的支持，拥有较高的社会资本。

## 4.5 本章小结

本章从模型和假设建立、研究设计、研究结果和讨论四个方面进行阐述。以糖尿病在线患者社区为研究背景，通过对社区内用户发帖数据及个人资料数据的收集，构建出相应的研究变量，利用生存分析方法对假设进行验证，分析并讨论了同伴支持与持续参与之间的关系。研究结果显示，寻求同伴支持、提供同伴支持和接收支持对持续参与均有积极作用，在寻求同伴支持和提供同伴支持中，信息支持和同伴支持对持续参与的作用均大于情感支持，在接收同伴支持中，信息支持和陪伴支持对持续参与有负向影响，情感支持对持续参与无显著作用。

## 第5章 结论

在线患者社区的本质是为病患用户提供疾病相关的沟通和交流平台，而其用户的特殊性——患有相似的疾病，决定了平台内部同伴支持行为的普遍性和重要性，识别不同类型的同伴支持行为以及分析不同类型同伴支持行为对平台持续参与可能产生的作用，对于平台和患者都有重要意义。文章以文本分类、在线社区、医疗保健和同伴支持领域相关文献为基础，基于大型糖尿病在线患者平台的用户活动数据，对上述两个问题展开了实证研究。在不同类型同伴支持行为识别中，本文为每一类同伴支持分别建立分类模型，通过对比传统机器学习和深度学习不同模型的实验评估结果，挑选最优的分类模型。在同伴支持对平台持续参与的作用分析中，本文基于构建好的分类器对用户的发帖行为进行分类，以构建研究所需要的变量，并利用生存分析方法对研究模型进行假设验证。文章的研究成果主要体现在以下几个方面：

(1) 本文建立了中文在线患者社区情境下的同伴支持分类模型，丰富了同伴支持分类的研究。探讨了有效进行同伴支持分类的新方法，为以后相似的文本分类任务提供了算法选择方向。以往研究多采用传统机器学习算法对社会支持进行分类<sup>[61][38][39]</sup>，本文的研究结果表明了深度学习算法在同伴支持分类任务上的优越性，doc2vec 文本特征表示方法和深度学习算法相结合能够在同伴支持分类任务中取得相对优异的效果。文章所提出的同伴支持分类框架和研究结果将为在线健康社区的研究和实践提供重要意义，同样的方法也可以应用或扩展到其他在线环境，例如在线知识社区，对其中参与者知识问答帖子的类型进行挖掘。

(2) 研究从同伴支持视角探讨了在线医疗社区用户持续参与的影响因素，对在线患者社区持续参与相关的文献做出了贡献，丰富了对持续参与影响因素的理解。用户的持续参与是保证在线患者社区繁荣发展的基础，文献多是从主观的认知角度出发探索影响用户持续参与的因素<sup>[32][64][68]</sup>，鲜有探讨客观的同伴支持互动与持续性行为之间的关系。本研究从同伴支持的视角出发，以在线患者社区实证数据为基础，探讨了同伴支持与用户持续参与之间的关系，丰富了对在线平台持续参与的研究，为持续参与研究拓展了新的视角。同时，本研究利用机器学习分类模型实现对同伴支持的自动分类，并根据分类结果构造研究变量，为未来的研究中变量的构造开拓了思路。

(3) 从理论视角出发，本研究模型由同伴支持行为以及同伴支持类型对持续参与的影响组成，目前结合同伴支持行为和同伴支持类型的研究还比较少，即便是对同伴支持类型和持续参与关系的研究，也多是接收不同类型的同伴支持这一角度出发<sup>[16]</sup>。本研究证实了寻求同伴支持、提供同伴支持和接收同伴支持均会对持续参与产生促进作用，还探究了在不同的同伴支持行为中支持类型对持续参

与的作用。研究发现在寻求同伴支持和提供同伴支持中，信息支持、情感支持和陪伴支持都会对持续参与产生积极影响，且信息支持和陪伴支持的作用大于情感支持。而在接收同伴支持中，信息支持和陪伴支持对持续参与产生的是负向的影响，情感支持无显著作用。本研究更深入地分析了同伴支持和持续参与之间的关系，为理解两者之间的关系提供了更全面细化的见解。

(4) 研究还探讨了好友数以及年龄对持续参与的影响，结果显示好友数和年龄层对用户的持续参与都有促进作用。好友数量对用户持续参与的积极作用佐证了社会网络关系数量对持续参与的积极影响。至于年龄层对用户持续参与的正向作用，表明了中老年用户的持续参与可能性高于年轻用户，这一发现与早期的老年人不愿使用新技术的结论有所不同。这些结果可以为研究人员在技术和健康综合环境中探索用户行为提供新的见解。

本文的研究可以为在线患者社区的管理者提供一定的启示，帮助管理者更有效地促进用户的持续参与，提高社区内部的活力，维持社区的稳定发展。根据研究结果，本文为在线患者社区运营提出以下建议：

(1) 研究结果表明寻求同伴支持、提供同伴支持和接收同伴支持均能提高用户持续参与的可能性，且提供同伴支持对持续参与的作用强度更大。根据此结论，在线患者社区可以设置激励机制，鼓励用户主动寻求支持以及为其他同伴提供支持。例如，社区可以对寻求支持和提供支持的用户奖励不同的积分，加速用户的身份升级。尤其针对提供同伴支持的行为，社区可以采用更多元的形式表示支持。例如，创建乐于助人排行榜，实时刷新记录，刺激用户积极为他人提供帮助。

(2) 研究发现信息支持、陪伴支持和情感支持在寻求同伴支持和提供同伴支持中对用户持续参与均有积极作用，信息支持作用强度最大。据此，社区可以设置健康知识交流、闲谈、情感交流三个不同的版块，鼓励用户在版块内积极交流。在健康知识交流版块对帖子加标签以优化用户对知识的搜索，促进健康信息知识的交流互换。但另一方面，研究还发现接收到的同伴支持中，信息支持和陪伴支持占比越高，用户持续参与概率越低。因此，社区还要注重用户提供信息支持的质量，可以根据该提供支持性帖子所获得的点赞数量对帖子进行加精或置顶等操作，强调对帖子质量的看重程度，引导用户提供高质量的信息支持，同时减少用户在主题帖下对无关话题的讨论。

(3) 根据研究结果中的好友数量会对用户持续参与产生正向作用，社区可以鼓励用户之间的加好友操作，定期向用户推荐具有相似兴趣的个体，或者根据用户的同伴支持行为为其推荐相应的个体。例如，针对经常寻求信息支持的个体，可以为其推荐经常提供信息支持的个体。此外，在线患者社区也应该更加关注年轻群体，如何留住年轻用户是像糖尿病这类慢性疾病在线患者社区值得思考的一个问题。

虽然本研究具有一定的理论价值和实践价值，但仍存在一定的局限性：

(1) 对于同伴支持分类效果仍有提升的空间，未来可以更有针对性的对此类文本分类任务进行算法层面的改进，采用前沿的多标签分类算法，提升分类效率。

(2) 本文的研究仅基于从“甜蜜家园”健康社区收集的数据，可能产生研究结果仅适用这类在线患者社区的情况。例如，慢性病的在线患者社区可能与心理疾病、急性疾病或综合性的在线患者社区有所不同，未来的研究有必要测试研究结果在其他在线患者社区中的有效性，以便对研究发现进行适当的扩展和应用。

(3) 由于采集到的数据限制，无法获得用户详细的个人资料数据，研究的控制变量仅纳入了好友数量。但可能还存在其他可能对用户持续参与产生影响的因素，例如用户的性别、学历、疾病严重程度等。此外，用户在交互过程中的响应时间、支持质量等都可能对用户持续参与产生影响。未来的研究可以将这些因素纳入模型中加以控制，提高模型的准确性。

(4) 在接收到的同伴支持中，研究是采用用户寻求支持下方回复的提供支持帖子数量作为接收同伴支持行为的数量。但在实际中，用户也可能通过浏览他人的帖子收到支持，本研究可能低估了用户收到的支持数量。未来在可能的条件下，可以结合用户的浏览数据深入研究用户接收同伴支持和持续参与之间的关系。

(5) 研究只是从同伴支持行为数据出发探讨了同伴支持和持续参与之间的关系，但关于同伴支持对用户持续参与的影响路径尚不清晰，未来可以将传统的持续参与影响因素构念，例如满意度、信任等，和同伴支持行为数据相结合，从认知层面更深入地理解同伴支持对持续参与的影响路径。此外，另一个有趣的研究方向是利用用户产生的纵向行为数据，可以探索用户行为状态随时间转移改变的规律，全面理解用户的行为规律，有针对性的预防用户的流失。

## 参考文献

- [1] Gui X, Chen Y, Kou Y, et al. Investigating support seeking from peers for pregnancy in online health communities[J]. Proceedings of the ACM on Human-Computer Interaction, 2017, 1(CSCW): 1-19.
- [2] Walther J B, Boyd S. Attraction to computer-mediated social support[J]. Communication technology and society: Audience adoption and uses, 2002, 153188: 50-88.
- [3] Chung J E. Social interaction in online support groups: Preference for online social interaction over offline social interaction[J]. Computers in Human Behavior, 2013, 29(4): 1408-1414.
- [4] 张雨薇, 蒋晓梦, 王婧, et al. 同伴支持在糖尿病患者健康教育中的研究现状综述[J]. 考试周刊, 2016, 000(064):194-194,97.
- [5] Zhou J, Zuo M, Yu Y, et al. How fundamental and supplemental interactions affect users' knowledge sharing in virtual communities? A social cognitive perspective[J]. Internet Research, 2014.
- [6] Zhang M, Yang C C. Classifying user intention and social support types in online healthcare discussions[C]// IEEE International Conference on Healthcare Informatics. 2015.
- [7] Eichhorn K C. Soliciting and providing social support over the Internet: An investigation of online eating disorder support groups[J]. Journal of Computer-Mediated Communication, 2008, 14(1): 67-78.
- [8] Duthler K W. The politeness of requests made via email and voicemail: Support for the hyperpersonal model[J]. Journal of computer-mediated communication, 2006, 11(2): 500-521.
- [9] Walther J B. Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction[J]. Communication research, 1996, 23(1): 3-43.
- [10] 赵栋祥.国内在线健康社区研究现状综述[J].图书情报工作,2018,62(09):134-142.
- [11] 宋晓龙.在线健康社区的病患用户社交关系及竞争行为研究[D].黑龙江:哈尔滨工业大学,2015
- [12] Xitong Guo, Shanshan Guo, Douglas Vogel, Yijun Li. Online Healthcare Community Interaction Dynamics[J]. Journal of Management Science and Engineering, Volume 1, Issue 1, 2016, Pages 58-74.

- [13]Zhang X, Liu S, Deng Z, et al. Knowledge sharing motivations in online health communities: A comparative study of health professionals and normal users[J]. *Computers in Human Behavior*, 2017.
- [14]张敏,马臻,张艳.在线健康社区用户隐性知识显性化行为的扎根分析[J].*图书馆学研究*,2019(01):53-61+93.
- [15]周军杰. 用户在线参与的行为类型——基于在线健康社区的质性分析[J]. *管理案例研究与评论*, 2016, 9(2): 173-184.
- [16]Wang Y C, Kraut R, Levine J M. To stay or leave? The relationship of emotional and informational support to commitment in online health support groups[C]//*Proceedings of the ACM 2012 conference on computer supported cooperative work*. 2012: 833-842.
- [17]Zhang Y. Understanding the sustained use of online health communities from a self-determination perspective[J]. *Journal of the Association for Information Science and Technology*, 2016, 67(12): 2842-2857.
- [18]Massimi M, Bender J L, Witteman H O, et al. Life transitions and online health communities: reflecting on adoption, use, and disengagement[C]//*Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 2014: 1491-1501.
- [19]Mein Goh J, Gao G, Agarwal R. The creation of social value: Can an online health community reduce rural–urban health disparities?[J]. *Mis Quarterly*, 2016, 40(1).
- [20]Dennis C . Peer support within a health care context: A concept analysis[J]. *International Journal of Nursing Studies*, 2003, 40(3):321-332.
- [21]黄琦, 林泽娜, 官玲燕, et al. 同伴支持在老年 2 型糖尿病患者健康教育中的实施[J]. *护理学报*, 2013(13):70-72.
- [22]Lakey B, Cohen S. Social support and theory[J]. *Social support measurement and intervention: A guide for health and social scientists*, 2000, 29.
- [23]王华, 王静. 同伴支持与心理健康研究[J]. *湖南第一师范学院学报*, 2009, 009(004):139-141.
- [24]Mamykina L, Nakikj D, Elhadad N. Collective sensemaking in online health forums[C]//*Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015: 3217-3226.
- [25]Mamykina L, Smyth T N, Dimond J P, et al. Learning from the crowd: Observational learning in crowdsourcing communities[C]//*Proceedings of the*

- 2016 CHI Conference on Human Factors in Computing Systems. 2016: 2635-2644.
- [26] Huh J, Ackerman M S. Collaborative help in chronic disease management: supporting individualized problems[C]//Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. 2012: 853-862.
- [27] Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians[J]. Journal of medical Internet research, 2011, 13(3): e62.
- [28] Epping-Jordan J E, Pruitt S D, Bengoa R, et al. Improving the quality of health care for chronic conditions[J]. BMJ Quality & Safety, 2004, 13(4): 299-305.
- [29] 郁花, 魏圆满. 社区糖尿病患者同伴支持干预的应用效果分析[J]. 中国保健营养, 2019, 29(14):280-281.
- [30] Bracke P, Christiaens W, Verhaeghe M. Self-esteem, self-efficacy, and the balance of peer support among persons with chronic mental health problems[J]. Journal of Applied Social Psychology, 2008, 38(2): 436-459.
- [31] 宁光, 毕宇芳. 中国成人糖尿病流行与控制现状[C]// 营养与糖尿病并发症——达能营养中心第十六届学术会议论文集. 2013.
- [32] 张星, 吴忱, 夏火松, 赵越. 在线健康社区用户持续知识贡献意愿的影响因素研究——以社会交换理论为视角[J]. 医学信息学杂志, 2019, 40(03):2-9.
- [33] Na Liu, Yu Tong, Hock Chuan Chan. Dual effects of social support seeking in patient-centric online healthcare communities: A longitudinal study[J]. Information & Management, 2020.
- [34] Kahn R L. Conboys over the life course: Attachment roles and social support[J]. Life-span development and behavior, 1980, 3: 253-286.
- [35] Wills T A. Supportive functions of interpersonal relationships[J]. 1985.
- [36] Yu B , Gerido L , He Z . Exploring text classification of social support in online health communities for people who are D/deaf and hard of hearing[J]. Proceedings of the Association for Information Science and Technology, 2017, 54(1):840-841.
- [37] Ballantine P W , Stephenson R J . Help me, I'm fat! Social support in online weight loss networks[J]. journal of consumer behaviour, 2011, 10(6):332-337.
- [38] Wang X, Zhao K, Street N. Analyzing and predicting user participations in online health communities: a social support perspective[J]. Journal of medical Internet research, 2017, 19(4): e130.

- [39] 吴江, 侯绍新, 靳萌萌, et al. 基于 LDA 模型特征选择的在线医疗社区文本分类及用户聚类研究[J]. 情报学报, 2017, v.36(11):89-97.
- [40] Zhang M, Yang C C, Gong X. Social Support and Exchange Patterns in an Online Smoking Cessation Intervention Program[C]// IEEE International Conference on Healthcare Informatics. IEEE, 2013.
- [41] Rodgers S, Chen Q M. Internet community group participation: psychosocial benefits for women with breast cancer[J]. Journal of Computer-Mediated Communication, 2005, 10(4).
- [42] Chen L. A Classification Framework for Online Social Support Using Deep Learning[C]// International Conference on Human-Computer Interaction 2019. Springer, Cham, 2019: 178-188.
- [43] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Advances in neural information processing systems. 2013: 3111-3119.
- [44] Le Q, Mikolov T. Distributed representations of sentences and documents[C]//International conference on machine learning. 2014: 1188-1196.
- [45] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of machine Learning research, 2003, 3(Jan): 993-1022.
- [46] Dudani S A. The distance-weighted k-nearest-neighbor rule[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1976 (4): 325-327.
- [47] LI Y, ZHOU R, LIAO W, et al. Case K-Nearest Neighbors method Retrieval Model with Rule-Based Reasoning [J]. Mechanical Science and Technology, 2004, 11.
- [48] 王亚露.基于支持向量机分类器的睡眠分期算法研究[D].浙江:浙江大学,2019.
- [49] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences[J]. arXiv preprint arXiv:1404.2188, 2014.
- [50] Ronan C, Jason W, Léon B, et al. Natural language processing from scratch[J]. Journal of Machine Learning Research (JMLR), 2011, 12: 2493-2537
- [51] Chung J, Gulcehre C, Cho K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555, 2014.
- [52] Hochreiter S, Bengio Y, Frasconi P, et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies[J]. 2001.
- [53] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.



- [54] Liu P, Qiu X, Huang X. Recurrent neural network for text classification with multi-task learning[J]. arXiv preprint arXiv:1605.05101, 2016.
- [55] Chung J, Gulcehre C, Cho K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555, 2014.
- [56] Fox S. The social life of health information, 2011[M]. Washington, DC: Pew Internet & American Life Project, 2011.
- [57] Welbourne J L, Blanchard A L, Wadsworth M B. Motivations in virtual health communities and their relationship to community, connectedness and stress[J]. Computers in Human Behavior, 2013, 29(1): 129-139.
- [58] Mo P K H , Coulson N S . Empowering processes in online support groups among people living with HIV/AIDS: A comparative analysis of ‘lurkers’ and ‘posters’[M]. Elsevier Science Publishers B. V. 2010.
- [59] Chomutare T, Årsand E, Hartvigsen G. Temporal community structure patterns in diabetes social networks[C]//2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE, 2012: 745-750.
- [60] Ridings C M, Gefen D. Virtual community attraction: Why people hang out online[J]. Journal of Computer-mediated communication, 2004, 10(1): JCMC10110.
- [61] 张敏, 唐国庆, 张艳. 虚拟社区用户健康知识共享行为的扎根分析[J]. 情报理论与实践, 2018, 41(7): 93.
- [62] Karahanna E, Straub D W, Chervany N L. Information technology adoption across time: a cross-sectional comparison of pre-adoption and post-adoption beliefs[J]. MIS quarterly, 1999: 183-213.
- [63] Zhang, Chen, et al. Research Note: Continued Participation in Online Innovation Communities: Does Community Response Matter Equally for Everyone?[J]. Information Systems Research, vol. 24, no. 4, 2013, pp. 1112–1130.
- [64] 金晓玲, 汤振亚, 周中允, et al. 用户为什么在问答社区中持续贡献知识?: 积分等级的调节作用[J]. 管理评论, 2013(12):140-148.
- [65] Zhang, Yan. A qualitative investigation of factors enabling sustained use of online health communities[J]. Proceedings of the American Society for Information Science & Technology, 2014.
- [66] Hsu M H , Chang C M , Lin H C , et al. Determinants of continued use of social media: The perspectives of uses and gratifications theory and perceived interactivity[J]. Information Research, 2015, 20(2).

- [67]张星, 陈星, 夏火松, 等. 在线健康社区中用户忠诚度的影响因素研究: 从信息系统成功与社会支持的角度[J]. 情报科学, 2016, 34(3): 133-138.
- [68]Cho J. The impact of post-adoption beliefs on the continued use of health apps[J]. International journal of medical informatics, 2016, 87(MAR.):75-83.
- [69]王哲. 社会化问答社区知乎的用户持续使用行为影响因素研究[J]. 情报科学, 2017, 035(001):78-83,143.
- [70]龙天悦. 在线医疗社区的持续使用行为及其对医患关系影响研究[D].安徽:合肥工业大学,2017.
- [71]Bhattacharjee A. Understanding information systems continuance: an expectation-confirmation model[J]. MIS quarterly, 2001: 351-370.
- [72]Massimi M, Bender J L, Witteman H O, et al. Life transitions and online health communities: reflecting on adoption, use, and disengagement[C]//Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. 2014: 1491-1501.
- [73]张小川, 余林峰, 桑瑞婷, et al. 融合 CNN 和 LDA 的短文本分类研究[J]. 软件工程, 2018, v.21; No.228(06):21-25.
- [74]Stevens K, Kegelmeyer P, Andrzejewski D, et al. Exploring topic coherence over many models and many topics[C]//Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics, 2012: 952-961.
- [75]Shang Y. Users' Participation Motivation and Behavior Patterns in Online Health Community: A Game Theory Viewpoint[C]//Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019.
- [76]Nijland N, van Gemert-Pijnen J E W C, Kelders S M, et al. Factors influencing the use of a Web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study[J]. Journal of medical Internet research, 2011, 13(3): e71.
- [77]Zhang X, Venkatesh V. A nomological network of knowledge management system use: Antecedents and consequences[J]. MIS quarterly, 2017, 41(4): 1275-1306.
- [78]Mo P K H, Coulson N S. Empowering processes in online support groups among people living with HIV/AIDS: A comparative analysis of 'lurkers' and 'posters'[J]. Computers in Human Behavior, 2010, 26(5): 1183-1193.
- [79]Ren Y, Kraut R, Kiesler S. Applying common identity and bond theory to design of online communities[J]. Organization studies, 2007, 28(3): 377-408.

- [80] Yang D, Kraut R E, Smith T, et al. Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities[C]//Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019: 1-14.
- [81] 刘璇, 汪林威, 李嘉, 等. 在线健康社区中用户回帖行为影响机理研究[J]. 管理科学, 2017, 30(1): 62-72.
- [82] Sykes T A, Venkatesh V, Johnson J L. Enterprise system implementation and employee job performance: Understanding the role of advice networks[J]. MIS quarterly, 2014, 38(1): 51-72.
- [83] 陈明红. 学术虚拟社区用户持续知识共享的意愿研究[J]. 情报资料工作, 2015, 36(1): 41-47.
- [84] Guo W, Liang R Y, Wang L, et al. Exploring sustained participation in firm-hosted communities in China: the effects of social capital and active degree[J]. Behaviour & Information Technology, 2017, 36(3): 223-242.
- [85] Kouper I. The pragmatics of peer advice in a LiveJournal community[J]. Language@ internet, 2010, 7(1).
- [86] David Gefen C M R. Implementation team responsiveness and user evaluation of customer relationship management: A quasi-experimental design study of social exchange theory[J]. Journal of management information systems, 2002, 19(1): 47-69.
- [87] Chen I Y L. The factors influencing members' continuance intentions in professional virtual communities—a longitudinal study[J]. Journal of Information science, 2007, 33(4): 451-467.
- [88] Joyce E, Kraut R E. Predicting continued participation in newsgroups[J]. Journal of Computer-Mediated Communication, 2006, 11(3): 723-747.
- [89] David Gefen C M R. Implementation team responsiveness and user evaluation of customer relationship management: A quasi-experimental design study of social exchange theory[J]. Journal of management information systems, 2002, 19(1): 47-69.
- [90] Akter S, Ray P, D'Ambra J. Continuance of mHealth services at the bottom of the pyramid: the roles of service quality and trust[J]. Electronic Markets, 2013, 23(1): 29-47.
- [91] Lin H F. Determinants of successful virtual communities: Contributions from system characteristics and social factors[J]. Information & Management, 2008, 45(8): 522-527.

- [92] Helgeson, V. S. & Cohen, S. Long-term effects of educational and peer discussion group interventions on adjustment to breast cancer[J]. *Health Psychology*, 2001, 20(5), 387-392.
- [93] Chin, Wynne. A permutation procedure for multi-group comparison of PLS models. *PLS and Related Methods: Proceedings of the PLS'03 International Symposium*, 2003, 33-43.
- [94] 张星, 吴忱, 夏火松, et al. 基于 S-O-R 模型的在线健康社区知识共享行为影响因素研究[J]. *现代情报*, 2018(8):18-26.
- [95] Budak C, Agrawal R. On participation in group chats on twitter[J]. In *Proceedings of the 22nd international conference on World Wide Web*, pages 165–176. International World Wide Web Conferences Steering Committee, 2013.

## 附录

### 附录 1：同伴支持分类标注指导

您好！您的任务是对来自线上糖尿病社区病人所发的帖子进行人工标注，这些帖子既包含原始帖，也包含原始贴下方的回复贴。对于每个帖子，您需要根据定义判断原始帖和回复贴属于下面同伴支持中的哪一类，一个帖子可能同时属于多个类别。若该帖子属于某一类同伴支持，则在对应表格中标 1，否则标 0。帖子分类如下：

- (1) 寻求信息支持
- (2) 提供信息支持
- (3) 寻求情感支持
- (4) 提供情感支持
- (5) 陪伴支持

#### 一、信息支持

信息支持是信息和指导的传递。由同伴提供疾病治疗、康复以及生活方面的信息或指导，给予忠告或提供协助解决问题的建议。帖子的内容与治疗或症状的建议、推荐、教学和个人经验有关，其中可能包括患者的医疗信息。这种支持可以帮助用户更好地了解如何处理自己的状况。

##### 1. 寻求信息支持

用户发帖寻求信息支持。例如，用户发帖询问疾病专业知识、治疗的药物、当前状态的疑惑等。示例如下：

- (1) 3/25 号确诊的，发现空腹 11.45 糖化 11.5 ，6/6 号开始不吃药， 9/7 号正好 3 个月了，去复查了下，结果如下，请懂的人看下， 停药之前也有测有糖化 6.1，这次糖化 6.0 本来每天现在也差不多坚持自己检测，感觉空腹跟餐后都还坚持的不错，怎么糖化没怎么下去。还有甲功能 8 项里面那个抗体怎么那么高？
- (2) 这个月的定期到医院检测此空腹血糖 6.27.。顿时感觉晴空霹雳，一直控糖控制得比较稳妥，想请教下大神原因哦，是不是头天睡觉睡得比较晚得原因影响力吗（0：30 多才睡的）！
- (3) 请问下大家有没有慢行胃炎治好的呀？

##### 2. 提供信息支持

用户发帖提供信息支持，例如有关疾病/治疗经历自我描述、给他人推荐/建议。示例如下：

- (1) 晚餐加点儿药试试

- (2) 随着年龄增长，人体对糖尿病并发症的承受能力每况愈下。美国“每日健康”网站援引医学专家的观点称，七八十岁以上的高龄糖友在对抗糖尿病时会面临六大挑战。1.低血糖。哈佛医学院副教授迈哈·孟希表示，相比高血糖，极低的血糖水平对老年人来说更加危险。应对措施就是简化治疗方案。孟希博士 2016 年发表的一篇论文表明，如果注射胰岛素的 70 岁以上糖友将每天多次注射改为每天仅注射一次，低血糖的风险会显著减少……
- (3) 你用同一滴血两根试纸测，数值就一样了，如果是同一个手指不同的 2 滴血，数值不一样的，还有每个手指的数值不一样的，还有左手右手的数值也不一样的

## 二、情感支持

情感支持是通过分享令人心情波动的事件来传递的。内容包含照顾、鼓励、关怀、信任、喜欢或爱心，可以帮助患者减轻压力水平，使其感到关怀和温暖。例如表达对患者的理解、与其同在的同情、情况会变好的鼓励。

### 1. 寻求情感支持

用户发帖寻求情感支持。例如，用户发帖描述无助/兴奋的心情、寻求共鸣等。示例如下：

- (1) 糖龄 14 年了，老大姑娘今年十岁了，老二月初也到来了，还是个小棉袄，以后要热死我的节奏啊
- (2) 昨天去省城医院开了基础胰岛素，让我每天打 8 个单位，今天一天心里都不安宁，不知该怎么下手？朋友再三告戒不痛的，我始终下不了手！我快疯了！觉得负能量的请绕行，谢谢！
- (3) 8 个月了，心态像过山车一样，再加上生活的各种压力，每天都恍恍惚惚，吃饭还没有俺闺女吃得多，天天噩梦，醒来浑身痛，二十年后这病能治愈吗？谁有缓解的好方法？说说

### 2. 提供情感支持

用户发帖提供情感支持，例如，用户发帖对他人表示祝福。示例如下：

- (1) 会的，放心好了别多想。糖人一样是健康的人，只是血糖控制得好没什么的
- (2) 我理解你的心理 很简单 但是下不了手
- (3) 三年的路过，加油

## 三、陪伴支持

陪伴支持有助于增强小组成员的社交网络和社区意识。陪伴支持包括闲聊、幽默/调侃、离题事件和团体感（加入或回应团体）。用户分享日常事件和想法，例如烹饪、园艺、旅行、天气、假日、生日、周年纪念日、游戏等。

示例如下：

- (1) 1.性别：男  
2.年龄：32  
3.糖尿病类型：医生暂时判断是 2 型  
4.患病日期：2019.9.21 确诊  
5.治疗方式：胰岛素泵  
6.出生地：安徽  
7.目前居住地：北京  
8.兴趣特长：上网  
9.注册原因：寻求帮助  
10.你是从什么地方知道甜蜜家园的：水木 BBS  
11.你对甜蜜家园发展的意见和建议：无  
12.自我介绍和留言（包括 QQ 等联系方式）：
- (2) 好长时间没吃饺子，今晚吃的是猪肉韭菜陷，包了 17 个，你们有吃 17 个的吗，刚才吃了我儿子的一个，吃了老公的一个，等一下在吃 17 个
- (3) 中秋节快乐小伙伴们！生活需要的仪式感。摆拍下，只敢吃一小口。怕升糖，哈哈哈
- (4) 可以考虑东北避暑，我能说我今天穿了条裙子。冻腿么

需要注意的是，有些帖子可能同时属于多个类型的同伴支持。例如，一个帖子可能既寻求信息支持也寻求情感支持，或提供情感支持的同时也在寻求信息支持，等等。示例如下：

(1) 寻求信息支持&寻求情感支持

夏天的时候每天散步，餐后一般都控制在 5.8 以下。在 8 月检查了一次后，糖化 4.6，C 肽 0.89，胰岛素 2.87，葡萄糖 4.5。停止打针 1 个月。期间血糖一直控制的很好，基本都在 7 以下。之后 9 月末，我觉得有必要恢复胰岛素的注射，于是就重新打针了。

但是入冬之后，因为天气太冷，我又对减重过于着迷，于是买了动感单车在家锻炼。到了 11 月的时候，糖化升到了 5.2，C 肽降到了 0.67，胰岛素降到了 2.23。每天锻炼坚持的很苦，单车调的档位很高，结果，体重也没控制住。好不容易降到了 55.5 之后，最近又恢复到了 58。当然这是我刻意的，因为之前几个月减重饮水很少，尿素都从 8.9 涨到了 9.5，很多时候尿液都很浓了。所以，我最近开始了大量补水。

现在真的不知道该咋个办了。胰岛素每天打了是早 2 晚 2，结果还没前段时间 1+1 的时候餐 2 控制的好，饮食没多大改变。

不知道怎么办了。

求安慰。

(2) 提供情感支持&寻求信息支持

先恭喜你，然后想请教，你只打针吗 配别的口服药吗，这些年体重增加的明显吗，我很想知道，请回复 谢谢。

(3) 提供信息支持&提供情感支持

糖友要知道血糖控制好能减少或延缓并发症的发生发展，血糖如果不控制，波动过大就会导致机体损害，会加速并发症的发生。不能说知道了最终结果，我们就什么也做不了了，50岁得并发症和70岁得并发症，差距还是很大的，希望糖友不要过度悲观。

请根据定义以及上下文信息，利用您已有的知识判断所给帖子属于哪一类同伴支持。有任何问题，请随时和我联系，谢谢！

## 附录 2：作用差异显著性检验公式

本文使用 Chin (2003) 提出的方法检验两个系数之间的统计差异，它是对传统 t 检验的改编。计算公式如下：

$$t = \frac{PC_1 - PC_2}{\sqrt{(SE_1^2 + SE_2^2)/N}}$$

其中， $t$ 是 $t$ 值； $N$ 是样本数量； $SE_1$ 和 $SE_2$ 是指每个路径的标准差； $PC_1$ 和 $PC_2$ 是对应的路径系数。



## 致谢

寒来暑往，从本科到研究生，在人大度过的七年美好时光转眼即逝。回首往事，不禁感慨万千。人大七年的求学生涯，占据我近三分之一的生命旅程，对我产生了巨大的影响，也在我人生中留下浓墨重彩的一笔。在这里，我有幸结识了一群意气风发、热情洋溢的同学，接触到了诸多年轻有为、德高望重的老师。值此论文完成之际，谨向所有在我求学过程中给予关心和帮助的老师、同学以及家人表示衷心的感谢。

念及师恩，感其恩重。感谢我的导师左美云教授，感谢左老师三年来的悉心指导和关怀照顾。从入学时文献阅读的引导，到组会上科研进展的监督和指导，再到论文写作过程中的指点和建议，左老师为我的学业提供了莫大的帮助。左老师是一位治学严谨的学者，大到研究问题的意义，小到论文标点的符号，都力求完美。在左老师的影响下，我也逐渐养成了良好的写作习惯。此外，左老师还是一位严以律己的奋斗者，有着优秀的时间管理概念，我所看到的永远是他饱满的精神状态，这也激励着我不断奋进。虽然左老师在治学上对我们要求严格，但在生活中他也是一位平易近人的益友，会指导我们的为人处世，和我们分享读过的书，叮嘱我们的安全和健康。感谢左老师这三年来的谆谆教诲和殷切关怀，这些指导将使我受益终生。

感谢我的论文开题组老师—余艳老师、张文平老师和杜玮老师对我的论文提出的宝贵意见和建议，为我后续研究工作的开展进一步指明了方向。感谢论文答辩委员会老师—许伟老师、余艳老师、张新勇老师和杜玮老师对我的论文提出的宝贵修改意见，使我的研究不断完善，帮助我更好地完成了毕业论文的研究工作。

感谢我的同学们。感谢我的实验室师兄、师姐和师妹们陪我度过了三年美好的研究生时光。感谢我的师姐商丽丽和师妹雷东荧对我的毕业论文研究提供的帮助，感谢同门师兄、师姐和师妹们在我的毕业研究过程中提出的宝贵意见和建议。感谢 2017 级信息硕士班全体同学带来的美好回忆。感谢我的室友阿努罕，良好的宿舍氛围为我提供了一个解压舒适的环境。

感谢我的家人一直以来对我的支持、关心和照顾。今年特殊的新冠肺炎疫情阻隔了我回校的脚步，但也使得我和家人有了更长的相处时间，感谢家人为我创造的舒适环境帮助我更好地完成毕业论文的写作。衷心感谢父母一直以来对我的包容和鼓励，无论我做出什么决定都对我表示支持，给了我莫大的前行勇气。