**摘要**

随着社会的发展，推荐系统正在社会中变得越来越重要，这里设计了一个多引擎的推荐系统，并就不同引擎的效果和效率，以及多引擎融合推荐的效果和效率做了分析

（中文摘要是论文内容的总结概括，应简要说明论文的主要内容和观点，一般在300以内。摘要内容的字体为宋体小四号，1.5倍行间距。）

**关键词：**推荐系统，协同过滤，多引擎，×××

（另起一行，顶格写。3～5个关键词，不同关键词之间用逗号分隔。）

**Abstract**

××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××（英文摘要应与中文摘要内容相对应。英文摘要内容的字体、字号为Times New Roman小四号，1.5倍行间距，）

**Key words**: ×××××××××××××××（另起一行，顶格写。不同关键词之间用逗号空格分隔。）

**目录（示例）**

**第一部分毕业设计**

[**1.** **引言 1**](#_Toc479953242)

[1.1 研究背景与意义 1](#_Toc479953243)

[1.2 国内外研究现状 1](#_Toc479953244)

[1.3 主要设计任务及预期目标（指标） 1](#_Toc479953245)

[**2.** **系统设计方案 1**](#_Toc479953246)

[2.1总体设计方案 1](#_Toc479953247)

[2.2关键问题解决 1](#_Toc479953248)

[2.3设计结果呈现 1](#_Toc479953249)

[**3.** **系统测试、验证和结果分析 1**](#_Toc479953250)

[3.1测试、验证方法 1](#_Toc479953251)

[3.2测试、验证结果 1](#_Toc479953252)

[3.3测试、验证结果的分析 2](#_Toc479953253)

[**4.** **总结 2**](#_Toc479953254)

[4.1本设计的特色与总结 2](#_Toc479953255)

[4.2设计中出现的问题及思考 2](#_Toc479953256)

[4.3展望 2](#_Toc479953257)

[**参考文献 2**](#_Toc479953258)

[**附件 4**](#_Toc479953259)

[**作者简历 5**](#_Toc479953259)

[**《浙江大学本科生毕业设计任务书》 （可不编页码）**](#_Toc479953259)

[**《浙江大学本科生毕业设计考核表》 （可不编页码）**](#_Toc479953259)

**第二部分 文献综述和开题报告**

[**文献综述和开题报告封面 （可不编页码）**](#_Toc479953259)

[**指导老师对文献综述和开题报告具体内容要求 （可不编页码）**](#_Toc479953259)

[**目录 （可不编页码）**](#_Toc479953242)

[一、文献综述 1](#_Toc479953242)

[二、开题报告 3](#_Toc479953242)

[三、外文翻译 5](#_Toc479953242)

[四、外文原文 （可不编页码）](#_Toc479953242)

[**《浙江大学本科生文献综述和开题报告考核表》 （可不编页码）**](#_Toc479953259)

1. **引言**

**1.1 研究背景与意义**

在现今知识过载这样的一个时代背景下，每个人都被淹没在信息的海洋里，如何为用户找到其感兴趣的内容，对用户进行个性化的推荐，已经成了一个重要的课题。

推荐系统作为一种有效的信息过滤手段，是当前解决信息过载问题及实现个性化信息服务的有效方法之一。尤其是在网络和电子商务环境下，得到了广泛的应用，带动了一大批中小型公司的快速成长：无论是淘宝、亚马逊的商品推荐，还是今日头条等新型内容网站的新闻推荐，都因为推荐系统得快速成熟而受益，从之前的千人一面做到了千人千面，极大地提高了点击率和转化率；

同时企业通过推荐系统，还可以有效保留用户，重建客户关系，提高用户的忠诚度，甚至可以根据企业需求，从企业的众多客户中识别最有价值和最有潜力的客户，向企业推荐客户。

推荐系统作为一门充满商业价值的学科，从诞生以来就得到了飞速的发展。下面简单讲解一下推荐系统研究的阶段[1]：

第一阶段：推荐系统形成的初期阶段。该阶段是面向系统的探索阶段，还没有出现大规模的商业化应用，但是基础理论已经提出。协同过滤、自动推荐、推荐系统等名词与想法在该阶段开始出现。

第二阶段：推荐系统商业应用的出现。推荐系统首先应用在了电子商务领域，其中最有名的是Amazon的电子商务推荐系统，其大大提高了Amazon的营业额——据统计，推荐系统的贡献率在20%~30%之间。另一个成功的应用是Facebook的广告，Facebook的广告推荐系统可以根据个人资料、用户朋友感兴趣的广告等对个人提供精准的广告推销。

第三阶段：研究大爆发，新型算法不断涌现阶段。来自数据挖掘、人工智能、信息检索、安全与隐私以及商业与营销等各个领域的研究，都为推荐系统提供了新的分析和方法。同时在不同的应用场景下，基于上下文的推荐、基于社交网络的推荐、基于人口统计学数据的推荐、基于关联的推荐等新场景下的推荐系统开始不停地涌现；在应用领域，电子商务，在线教育[9,10]，商业旅游等都有所涉及；至于到今天面对着海量的数据，同样出现了大数据环境下的推荐系统，其靠着分布式的思想与技术来处理着海量的数据，提供了更加快速与准确的推荐。

随着时代不断向前，推荐系统在越来越多领域大放光彩，如Amazon的个性化产品推荐、Netflix的视频推荐、Pandora的音乐推荐、Facebook的好友推荐和Google Reader的个性化阅读等[2]。但是面对着新的应用场景与海量的数据，以及一些仍然没有得到完全解决的历史问题，推荐系统仍然有许多困难等待着有志之人来挑战。

* 1. **国内外研究现状**

国内外研究现状主要分为三个方向来阐述：

算法研究

应用场景

推荐策略

首先，在推荐系统所使用的算法上，主流的算法分为：随机推荐、最热门推荐、基于邻域的推荐算法、隐语义模型以及基于图的推荐算法。XXX

随机推荐是最简单的一种推荐算法，即给定一个用户，将目标物品列表中的随机一个物品推荐给他。这种算法在早期使用较多，目前多用于冷启动物品的推荐以及对新加入用户推荐物品，同时在一些没有复杂的推荐系统的网站上进行新闻资讯的推荐。

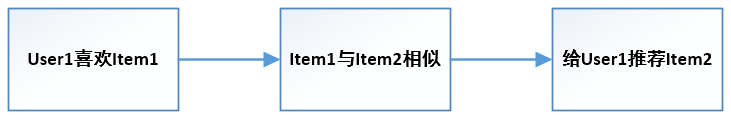
最热门推荐是指，将用户对物品的行为，包括浏览、购买、打分等，按照一定的时间跨度进行统计，然后将其中最热门的物品推荐给用户。这种算法虽然原始简单，但仍是目前最普遍使用的算法之一，在各大网站上都可以看到今日排行、最多话题、十大热搜等推荐模块。

基于邻域的推荐算法比较复杂，主要分为两种：基于用户的协同过滤算法和基于物品的协同过滤算法。

其中，基于用户的协同过滤算法是指根据用户的行为计算用户之间的相似度，然后将与目标用户最相似的几个用户感兴趣的物品推荐给目标用户。



而基于物品的协同过滤算法则需要先计算物品之间的相似度，然后将与目标用户产生行为的物品相似度高的物品推荐给目标用户。



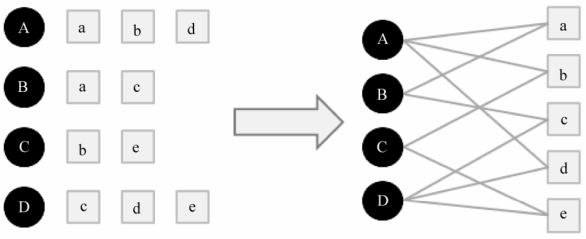
隐语义模型（LFM，Latent Factor Model）又称为隐含类别模型（Latent Class Model）、话题模型（Topic Model）、 矩阵分解模型（Matrix Factorization Model）等，其都是在同一思想体系的不同扩展，其基本思想就是通过隐含的特征联系用户和物品。

首先，从上面关于推荐系统的本质的讨论可以看出，推荐系统的本质就是通过不同的方式来联系用户和物品，即可以抽象为通过特征集来联系用户和物品。但是，一般的推荐引擎其特征都是人为规定好的，比如用户的相似度、物品的相似度、用户的社交关系等。隐语义模型的不同是，其特征的生成是一个通过机器学习自动生成的黑箱过程，虽然有很好的推荐结果，但是不具有可解释性。

虽然隐语义模型中的特征不可解释，但我们仍然可以在此处做一个简单的类比：假设有100个人来自10个学校，不同学校的学生校服不同，校歌也不同，如果我们需要通过校服来推荐校歌，那么，此处机器自动学习到的隐含特征就可能是学校。需要注意的是，虽然对人来说校服对应校歌很简单，但对机器来说，他们都是平等的数据，校服是否对应校歌要根据训练结果来决定；其次，特征的数量也未必是10个，可能指定了是5个或者20个，到时候特征和结果也都会不一样。

XXX（隐语义的算法推导）

基于图的推荐算法是指，将用户、物品以及其他特征（比如标签或者用户的社交关系）都表现在同一张图上，通过在该关系图上的随机游走，通过设定每一步决策的概率和随机试验的次数，来决定给用户推荐什么物品。该算法的优点是思想简单，不需要写复杂的算法，缺点是大部分图算法所共有的效率低下问题。



以上讲解的都是最基础的算法思想，在推荐系统的不断发展之中，也出现了很多对算法的改进方式：

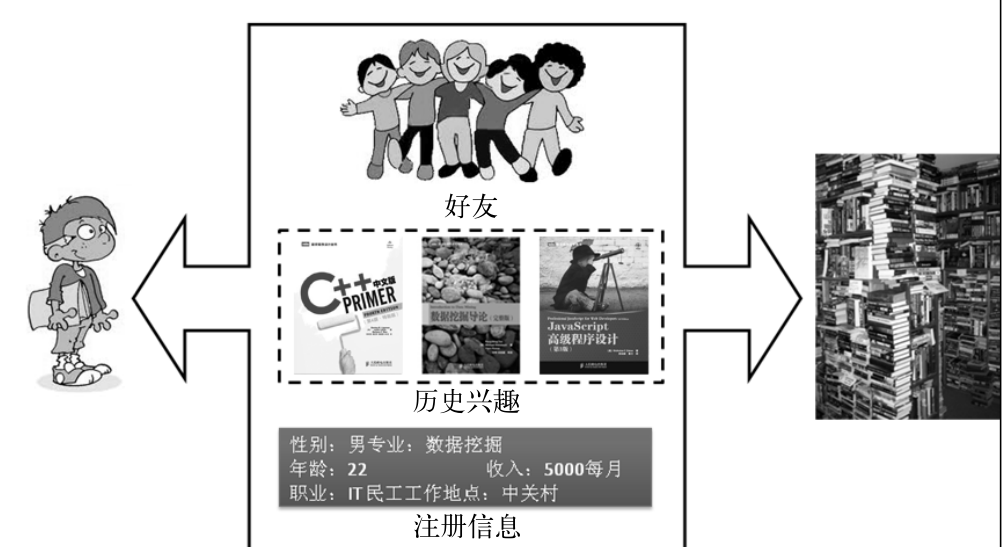
随着推荐系统的发展和社会的进步，推荐系统在不同的应用场景也大放光彩。虽然算法的思想是一致的，但在不同的应用领域，推荐系统的具体设计和具体算法却又有些不同。以下简单介绍一下利用用户标签、利用时间信息、利用社交网络的推荐系统。

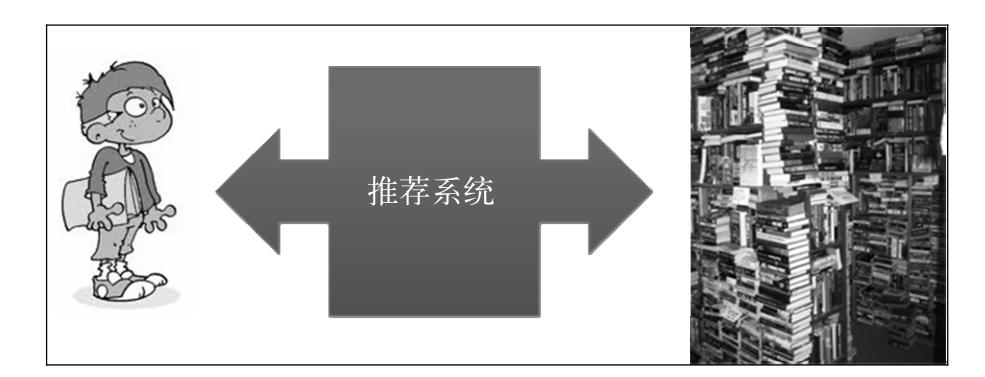
利用用户标签的推荐系统，引入的一个新的变量就是标签，而通过标签这一个新的变量，能在以下几个方向上改进算法：通过标签计算用户相似度，然后通过用户相似度推荐物品；通过标签计算物品相似度，然后通过用户相似度推荐物品；通过用户常用的标签来给用户推荐该标签下的物品；通过用户所产生行为的物品上的标签，给用户推荐该标签下的物品。

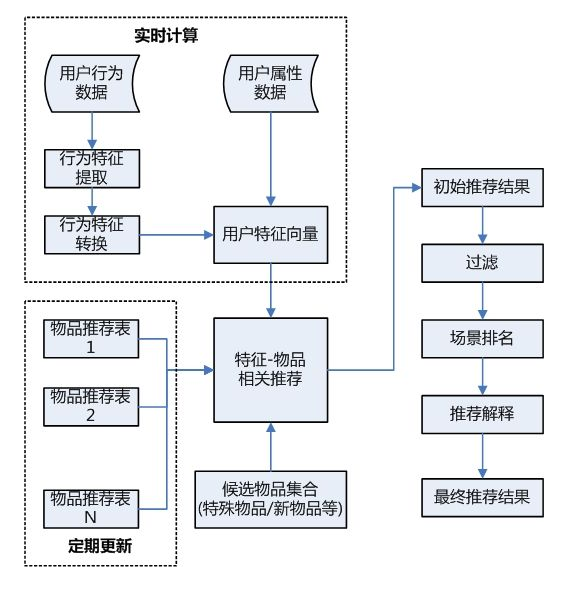
利用时间信息的推荐系统，引入了时间这个新变量，其考虑了用户或者物品在时间中的改变情况，比如用户的兴趣会变，对某个物品的评价会随着时代的改变而改变。通过引入时间这一个新变量，有以下几种方式改良算法：加入时间的考虑计算用户的相似度，然后根据用户相似度推荐物品；加入时间的考虑计算物品的相似度，然后根据物品相似度推荐；再利用用户相似度推荐时，加大相似用户最近行为的权重；再利用物品相似度推荐时，加大用户最近喜欢的物品的权重；以及上面几种方式的组和改进方式。

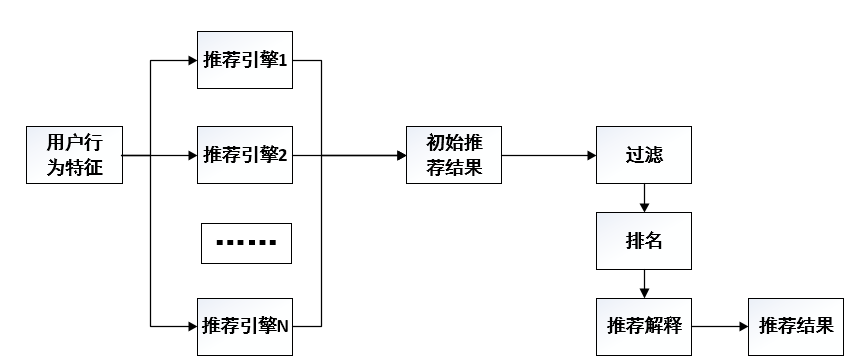
利用社交网络的推荐系统，引入了用户之间的社交关系这一新变量。而用户的社交关系能在以下几个方向上改良算法：通过用户的社交关系重新计算用户之间的相似度，再利用用户相似度进行推荐；XXX

同时，随着人们的不断研究，以及应用场景的不断开拓，推荐系统逐渐被应用在了更广阔的领域，新出现的领域有基于大数据的推荐系统，基于地理位置的推荐系统，组推荐系统等。









### 主流推荐系统

目前，主流推荐系统可以分为4类[5]：基于内容的推荐、协同过滤推荐、基于知识的推荐和组合推荐。

基于内容的推荐源于信息检索技术，不依赖于用户对项目的评价信息，侧重考察候选推荐项目与用户特征的匹配程度。

协同过滤推荐主要包括两类：

一类是基于模型的方法。此方法利用数据挖掘或者机器学习方法，在训练集上构建用户特征模型，依此进行推荐。其优点在于稳定性好；缺点是训练时间长、计算复杂性高；

另一类是启发式方法，也是目前应用最普遍的一种协同过滤推荐方法。该方法需要建立用户-项目评分矩阵，根据相似用户具有相似偏好的假设进行推荐。该方法的缺点是在冷启动情况下，即用户评分信息很少，或者没有评分信息时，无法做出有效的预测。

基于知识的推荐是一种基于特定领域规则或实例的推理方法，其优点在于不需要建立用户需求偏好模型，缺点是难以制定合理的推理规则。

组合推荐是为了克服上述各种推荐技术的弱点，对3种推荐方法的组合应用，其中，比较常见的是基于内容的推荐与协同过滤推荐的组合。

### 研究方向及进展

最近几年兴起的新型推荐系统主要有以下几类：基于社交网络的推荐[3]、基于上下文的推荐、基于标签的推荐、基于关联的推荐[4]、基于大数据的推荐[6]等；限于篇幅原因，仅重点介绍基于社交网络的推荐和基于大数据的推荐。

#### 存在的问题

* 社会化推荐系统中的冷启动问题

对于新加入的用户，可以引入用户社会关系信息，把与此用户存在社会关系的用户作为兴趣相似用户，在一定程度上缓解了推荐系统中新用户的冷启动问题；对于新加入项目，社会网络中用户之间还能够对新加入项目进行自动推荐与传播。

* 社会化推荐系统的隐私保护和安全性问题

因为社会化推荐系统中不仅涉及用户本身的属性信息，还涉及用户的社会关系信息，这些信息可能揭示用户的个人隐私。因此有些学者提出了一种具有隐私保护机制的社会化推荐系统，或者把推荐准确率和保护用户隐私进行折衷处理，在不揭示用户间敏感关系的前提下，提出个性化网络推荐方法。

* 社会化推荐系统大规模数据处理问题

社会化推荐系统不仅需要处理巨大的用户-项目关系表，同时也要处理巨大的用户间的社交网络关系，这需要消耗大量的计算资源，在实际的工业应用中，这种代价是相当昂贵的。目前已有的研究成果中涉及的数据集大多数都是小规模的或者模拟的数据集。因此，面对实际应用需求，利用社会网络信息，设计能够处理对实时性要求较高的快速推荐算法，将逐渐成为社会化推荐系统领域新的挑战。

传统问题：

* 冷启动问题：自从推荐系统诞生之初，冷启动问题就一直困扰着不同推荐系统的设计者；而如今，随着社交数据的引入，每个用户不再是孤立的个体，使得新出现的用户能迅速与已有用户联系起来，可以说已经算是比较成功的解决了用户冷启动问题；对于物品冷启动和系统冷启动，文献[11]引入了机器学习的方法来解决物品冷启动的问题，并取得了较好的效果。
* 如何处理用户多兴趣和用户兴趣转变的情况，即处理用户兴趣多样性和动态性特征问题；
* 如何劝导用户接受推荐结果，增加推荐系统的可信度与可解释性等；

新应用下的问题：

* 机器学习技术与数据挖掘技术如何在推荐系统中有效的应用；
* 如何促进推荐系统由购物助手角色到更智能的角色的转变，比如使得推荐系统可以对用户的决策进行支持等；

社会问题：

* 信任问题与隐私问题：一是用户的个人信息泄露、被推荐系统滥用，二是存在恶意推荐的可能性；
* 推荐系统与社会各方面的结合运用，比如不光要服务企业，也要服务科研机构，政府等；
* 社会分化和推荐逻辑空间的巴尔干化现象[5]：即局域性限制，在推荐系统中根据个体兴趣以及历史行为进行推荐时尤其容易导致这种局域性限制和社会分化的出现。如何防止推荐系统的“过度推荐”使得人们接触的世界越来越小，形成领域内的信息闭塞；甚至是沉迷其中的“精神鸦片”现象的形成，是信息领域推荐的一个重要课题
  1. **主要设计任务及预期目标（指标）**

1. **系统设计方案**

**2.1总体设计方案**

**2.2关键问题解决**

（本部分要求详细记载和描述。）

**2.3设计结果呈现**

（本部分要求详细记载和描述。）

1. **系统测试、验证与结果分析**

**3.1****测试、验证方法**

**3.2****测试、验证结果**

**3.3测试、验证结果的分析**

1. **总结**

**4.1本设计的特色与总结**

（包括：本设计的成果展示；对完成工作的总结，并以条目形式罗列出本文所实现研究成果的一些特点和优点，使人一目了然。）

* 1. **设计中出现的问题及思考**

（总结完成过程中遇到的一些问题的解决办法，最好能把一些细节再回顾一下，尽量把遇到的问题的解决办法都写出来；可能有的问题在前文中已经提到了，但仍需在这里再罗列归纳一下。）

* 1. **展望**

（总结本文尚待完善之处，展望对今后改进和完善方向的思考；接着站在更全局性的角度和高度展望一下相关研究领域今后的发展，最后说一段展望发展的概括性的话。）

**参考文献**

文献排列顺序按照在论文正文中出现的先后循序进行，并在正文中将参考文献引用的位置标注出来，两者编号应该一致。

1. Morgan R M, Hunt S D. The commitment-trust theory of relationship marketing [J]. Journal of Marketing, 1994, 58(3): 20-38.
2. 李里特, 候文义. 农业产业化和结构调整的几个关键问题[J]. 科技导报, 2002(1): 36-39.
3. Beck C A J, Sales B D. Family mediation: Facts, myths, and future prospects[M]. Washington, DC: American Psychological Association. 2001: 10-12.
4. 郭健. 哈佛大学发展史研究[M]. 石家庄:河北教育出版社, 2000: 2-3.
5. 辛希孟. 信息技术和信息服务国际研讨会论文集: A集[C]. 北京:中国社会科学出版社, 1994.
6. 张筑生. 微分半动力系统的不变集[D]. 北京: 北京大学数学系数学研究所, 1983.
7. 冯西桥. 核反应堆压力管道和压力容器的LBB分析[R]. 北京: 清华大学核能技术设计研究院, 1997.
8. 钟文发, 赵玮. 运筹学的理论和应用——中国运筹学会第五届大会论文集[C]. 西安: 西安电子科技大学出版社, 1996: 468-471.
9. Spivak G. "Can the Subaltern Speak?" [A]. In Nelson C &GrossbergL(eds.). Victory in Limbo: Imigism [C]. Urbana: University of Illinois Press, 1988: 271-313.
10. GB/T 16159-1996, 汉语拼音正词法基本规则[S].
11. 姜锡洲. 一种温热外敷药制备方案[P]. 中国专利: 881056073, 1989-07-26.
12. 王明亮.关于中国学术期刊标准化数据库系统工程的进展[M/CD]. <http://www.cajed.Cn/pub/wml.txt/980810-2.html,1998.08.16/1998-10-04>.

[13] 万锦坤. 中国大学学报论文文摘(1983-1993)英文版[DB/CD]. 北京:中国大百科全书出版社, 1996.

**本科生毕业设计任务书**

**一、题目：**

**二、指导教师对毕业设计的进度安排及任务要求：**

**起讫日期 2017年11月20日至 2018年6月8日**

**指导教师**（**签名） 职称**

**三、系或研究所审核意见:**

**负责人**（**签名）**

**毕 业 设计 考 核**

**一、指导教师对毕业设计的评语：**

**指导教师(签名）**

**年 月 日**

**二、答辩小组对毕业设计的答辩评语及总评成绩：**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **成绩**  **比例** | **文献综述**  **占（10%）** | **开题报告**  **占（15%）** | **外文翻译**  **占（5%）** | **毕业设计质量及答辩**  **占（70%）** | **总评成绩** |
| **分值** |  |  |  |  |  |

**答辩小组负责人（签名）**

**年 月 日**