

**本 科 毕 业 设 计（论文）**



**题目: 异质信息网络中的相似性推荐算法研究**

**姓 名 王睿嘉**

**学 院 计算机学院**

**专 业 计算机科学与技术**

**班 级 2015211307**

**学 号 2015211906**

**班内序号 31**

**指导教师 石川**

**2019 年 5 月**

**异质信息网络中的相似性推荐算法研究**

**摘 要**

随着大数据时代的来临，互联网信息爆炸式增长，而推荐是解决信息过载的有效方法，已广泛应用至电子商务和互联网服务中。作为重要的数据挖掘任务，推荐系统包括许多对象类型及对象间的丰富交互关系，可以自然而然地利用异质信息网络（Heterogeneous Information Network）建模。而异质信息网络优越的信息整合能力及丰富的语义信息，有潜力产生更准确的推荐。但是，现有异质信息网络的推荐算法大多聚焦于用户和物品间的交互信息，即购买历史，而这在实际应用中仅能反映用户偏好和物品特性的一个方面。因此，为深入挖掘不同方面的语义关系相似性，Neural network based Aspect-level Collaborative Filtering model (NeuACF)设计了基于元路径的推荐算法，使得推荐模型融合多方面语义信息，得到更加全面精确的结果。

本文首先介绍选题背景及相关技术，从宏观角度回顾推荐领域的经典算法。其次，深入分析NeuACF的具体结构,提出相应改进模型——Aspect-level Collaborative Filtering model based on Metric Learning（ACFML）。具体地，针对用户和物品的交互可能性利用点积度量，可能导致相似性传递特征被破坏的问题，引入度量学习，利用距离定义的metric捕捉数据间关系，实现更全面的信息传递；针对关注于评分信息绝对数值的point-wise损失函数，提出使用基于负采样的pair-wise损失函数，刻画对于Top-N推荐更为重要的相对交互可能性，产生排序更合理的推荐列表。最后，本文在推荐数据集Amazon和MovieLens上，与经典算法进行了大量对比实验，验证了所提模型的优越性。

**关键词** 推荐 异质信息网络 度量学习

**Research on Similarity Recommendation Algorithm in Heterogeneous Information Network**

**ABSTRACT**

Recommendation is an effective way to address information overload and is widely used in e-commerce and Internet services. As an important data mining task, the recommendation system includes many object types and rich interactions between objects, which can be naturally modeled using the Heterogeneous Information Network (HIN). The superior information integration capability and rich semantic information of HIN further empower more exact recommendation. However, the existing HIN recommendation algorithms mainly focus on the interaction information between users and items, that is, purchase history. And this can only reflect one aspect of users’ taste and the characteristics of items in practical applications. Therefore, in order to dig deeper into the semantic similarity of various aspects, the Neural network based Aspect-level Collaborative Filtering model (NeuACF) designed a recommendation algorithm based on the meta-path, which makes the recommendation model fuse multiple aspects of semantic information and obtain more comprehensive and accurate results.

This paper first introduces the background of the topic and related technologies, and reviews the classic algorithms in the recommendation field from a macro perspective. Secondly, we deeply analyze the specific model structure of NeuACF, and propose the corresponding improved model, Aspect-level Collaborative Filtering model based on Metric Learning (ACFML). Specifically, concerning the dot product, which is used for measuring user and item interaction possibility and leads to the destruction of similarity transfer characteristics, we introduce metric learning to use distance defined metric to capture data relationships, and achieve comprehensive information transfer; in terms of the point-wise loss function, which focus on the absolute value of ratings, we propose to use the pair-wise loss function based on negative sampling to characterize the relative interaction possibilities which are important for Top-N recommendation, and work out a more reasonable recommendation list. Finally, this paper has carried out extensive comparative experiments with the classical algorithms such as NeuACF on the recommended datasets Amazon and MovieLens, and verifies the superiority of the proposed model.

**KEY WORDS** recommendation heterogeneous information network metric learning

**目 录**

[第一章 引言 1](#_Toc9434785)

[1.1 课题背景 1](#_Toc9434786)

[1.1.1 研究意义 1](#_Toc9434787)

[1.1.2 研究现状 1](#_Toc9434788)

[1.2 课题任务 2](#_Toc9434789)

[1.2.1 课题内容 2](#_Toc9434790)

[1.2.2 本人承担任务 2](#_Toc9434791)

[1.3 论文结构 3](#_Toc9434792)

[第二章 相关技术介绍 5](#_Toc9434793)

[2.1 异质信息网络 5](#_Toc9434794)

[2.1.1 基本定义 5](#_Toc9434795)

[2.1.2 元路径 6](#_Toc9434796)

[2.2 注意力机制 7](#_Toc9434797)

[2.3 度量学习 10](#_Toc9434798)

[2.3.1 形式化定义 10](#_Toc9434799)

[2.3.2 kNN的度量学习 11](#_Toc9434800)

[2.4 本章小结 12](#_Toc9434801)

[第三章 基于异质信息网络的推荐模型ACFML详解 14](#_Toc9434802)

[3.1 基础模型NeuACF 14](#_Toc9434803)

[3.1.1 模型框架 14](#_Toc9434804)

[3.1.2 方面级相似性矩阵计算 15](#_Toc9434805)

[3.1.3 方面级隐因子学习 15](#_Toc9434806)

[3.1.4 基于注意力机制的方面级隐因子融合 16](#_Toc9434807)

[3.1.5 模型优化 16](#_Toc9434808)

[3.2 改进模型ACFML 17](#_Toc9434809)

[3.2.1 NeuACF问题分析 17](#_Toc9434810)

[3.2.2 改进方案 18](#_Toc9434811)

[3.2.3 总体模型 19](#_Toc9434812)

[3.3 本章小结 19](#_Toc9434813)

[第四章 实验 20](#_Toc9434814)

[4.1 实验设置 20](#_Toc9434815)

[4.1.1 数据集 20](#_Toc9434816)

[4.1.2 评价指标 20](#_Toc9434817)

[4.1.3 Baselines 21](#_Toc9434818)

[4.1.4 实现细节 22](#_Toc9434819)

[4.2 实验结果 22](#_Toc9434820)

[4.2.1 推荐性能分析 23](#_Toc9434821)

[4.2.2 margin大小对性能的影响 23](#_Toc9434822)

[4.2.3 负采样数量对性能的影响 24](#_Toc9434823)

[4.2.4 隐因子数量对性能的影响 25](#_Toc9434824)

[4.3 本章小结 25](#_Toc9434825)

[第五章 结束语 26](#_Toc9434826)

[5.1 论文工作总结 26](#_Toc9434827)

[5.2 问题和展望 26](#_Toc9434828)

[5.2.1 问题 26](#_Toc9434829)

[5.2.2 展望 26](#_Toc9434830)

[参考文献 28](#_Toc9434831)

[致 谢 30](#_Toc9434832)

[外 文 资 料 32](#_Toc9434833)

[外 文 译 文 50](#_Toc9434834)

# 引言

## 1.1 课题背景

### 1.1.1 研究意义

随着大数据时代的来临，互联网信息爆炸式增长，而推荐是解决信息过载的有效方法，已广泛应用至电子商务和互联网服务中。作为重要的数据挖掘任务，推荐系统包括许多对象类型（如电影推荐中的用户、电影、演员和导演等）及对象间的丰富交互关系（如观看信息、社交信息和属性信息等）。因此，融合更多信息来进行混合推荐是解决数据稀疏性的有效途径。

近年来，大量涌现的社会媒体网站包含许多不同类型对象间的复杂交互，将这些对象建模为同质网络是很困难的。作为半结构化的表示方法，异质信息网络（Heterogeneous Information Network），即多种类型节点和链路的建模方式，可以有效建模和处理复杂多样的数据。其次，与同质网络相比，异质网络不仅可以融合更多类型的对象及其复杂交互关系，而且可以对齐多个社交网络平台的信息。最后，异质信息网络包含丰富的语义，将导致更细微的知识发现。因此，异质信息网络作为有效的信息融合方式可以用于建模推荐系统中的对象和关系，可能产生更准确的推荐结果。

### 1.1.2 研究现状

传统的推荐系统通常仅利用用户-物品评分反馈信息进行推荐。协同过滤[1]是最普及的技术之一，其基本假设是：有着相似购买历史的用户会倾向于购买相似的物品。为捕捉相似用户的购买偏好，隐因子模型（如矩阵分解）[2,3]应运而生，并在推荐系统中表现出令人瞩目的性能。具体地，它将用户-物品评分矩阵分解为两个低秩的用户因子和物品因子，然后利用隐因子进行预测。由于隐因子模型极易受数据稀疏性影响，随着社交媒体的普及，越来越多的研究人员通过整合辅助信息扩展隐因子模型，如异质网络推荐系统[4]。

近年来，异质信息网络，即多种类型节点和链路的建模方式，在数据挖掘领域大放异彩。作为重要的数据挖掘任务，推荐系统包括许多对象类型（如电影推荐中的用户、电影、演员和导演等）及对象间的丰富交互关系（如观看信息、社交信息和属性信息等），可以自然而然地利用异质信息网络建模。而异质信息网络优越的信息综合能力及丰富的语义信息使得更准确的推荐成为可能。

元路径（meta-path），即连接两类对象的关系组合，被广泛用于异质信息网络中的语义信息建模和关系抽取。Shi等人[5]实现了一个基于语义的推荐系统HeteRecom，利用元路径的语义信息来评估电影间的相似性。此外，考虑属性值，如链接上的评分，他们进一步将推荐系统建模为加权异质信息网络，并提出基于语义路径的个性化推荐方法SemRec[6]。为充分利用关系异质性，Yu等人[7]引入基于元路径的隐特征来表示沿不同类型路径的用户和物品间的连接，然后利用贝叶斯排名优化技术在全局和个性化水平上定义推荐模型。同样基于元路径，NeuACF[8]利用深度神经网络模拟和融合不同方面级别的隐因子，从不同方面表示用户偏好和物品特征，并自动确定其重要性。

许多方法利用异质信息网络来融合各种信息。结合不同上下文，Jamali等人[9]提出基于上下文的矩阵分解模型，考虑每个实体的一般隐因子和依赖于上下文的隐因子。使用用户隐式反馈数据，Yu等人[7]进一步解决全局和个性化实体推荐问题。根据相关的兴趣组，Ren等人[10]提出基于簇的引用推荐框架来预测每个查询在书目网络中的引用。同样，Wu等人[11]结合兴趣组信息，利用图表摘要和基于内容的聚类方法来进行媒体推荐。此外，利用多种类型的关系，Luo等人[12]提出一种社交协同过滤算法。

总的来说，大多数异质信息网络的相似性推荐算法集中在简单的网络结构（如二分或星型模式）和粗糙的语义建模（如元路径）上。但很多实际数据，特别是电子商务数据，所构建的异质信息网络包含很多节点和边类型，没有简单的网络模式。这种无模式或丰富模式的异质信息网络带来了许多新的研究问题，如多类型对象关系的管理及元路径的自动产生等，与此同时也亟需更强大的语义建模和分析方法挖掘丰富细致的语义信息。

## 1.2 课题任务

### 1.2.1 课题内容

本课题的主要内容是改进基于异质信息网络的推荐算法，使得推荐模型更好地融合多种语义信息，得到更加全面准确的结果。

为此，首先阅读异质信息网络综述，了解异质信息网络中的关键理论及相似性推荐算法的研究现状。建立了宏观概念后，有选择地调研相关论文，具体学习部分相似性推荐算法。在此基础上，对比分析现有的相似性推荐算法，选取合适的算法并改进。最终，完成相应改进模型的编码工作，并进行实验验证及结果评估。

### 1.2.2 本人承担任务

首先，了解课题背景及意义，调研相似性推荐算法，完成开题工作。

其次，基于对现有算法的分析，选择NeuACF作为基本模型，并剖析其存在的问题。具体地，NeuACF中，用户和物品的交互可能性是利用embedding间的点积来度量的。由于点积本身所固有的缺陷——不满足三角不等式，可能导致相似性的传递特征被破坏，从而限制模型效果的提升。与此同时，NeuACF使用point-wise损失函数，关注于评分信息的绝对数值。但对于Top-N推荐而言，相对的交互可能性更为重要。因而通过修改损失函数显式捕捉排序信息，可能产生更为精准的推荐列表。

针对分析所得结论，进一步完成相应改进方案ACFML的详细设计。具体地，引入metric learning，利用距离定义的metric捕捉数据间关系，弥补点积的不足，从而实现更全方位的信息传递。与此同时，使用基于负采样的pair-wise损失函数替换point-wise损失函数，产生排序更合理的推荐列表。

确定改进方案后，熟悉数据特点及相应算法框架。编码实现模型，完成中期报告及答辩。

最终，基于经典数据集进行实验验证，并对推荐结果进行分析评估，完成并提交毕业设计论文、相关文档及源程序，进行答辩。

## 1.3 论文结构

本文除引言外，后续内容可分为四个章节：

第二章主要介绍模型相关的背景技术，包括异质信息网络的关键概念、注意力机制的核心内容及度量学习的基础知识。

第三章主要介绍模型的详细方案。首先，阐述NeuACF的具体框架。其次，分析NeuACF存在的问题与不足，及其可能对模型性能产生的不良影响。最后，介绍如何结合先进方法设计改进模型ACFML，及其背后相应的理论原因。

第四章主要介绍在经典数据集Amazon和MovieLens上所进行的对比实验，并对所得推荐结果进行解释分析，总结模型的优点及问题。

第五章，作为最后一个章节，对本文进行回顾，综合评价模型的创新点及进一步的优化方向，并对异质信息网络推荐算法的未来发展进行展望。

# 第二章 相关技术介绍

## 2.1 异质信息网络

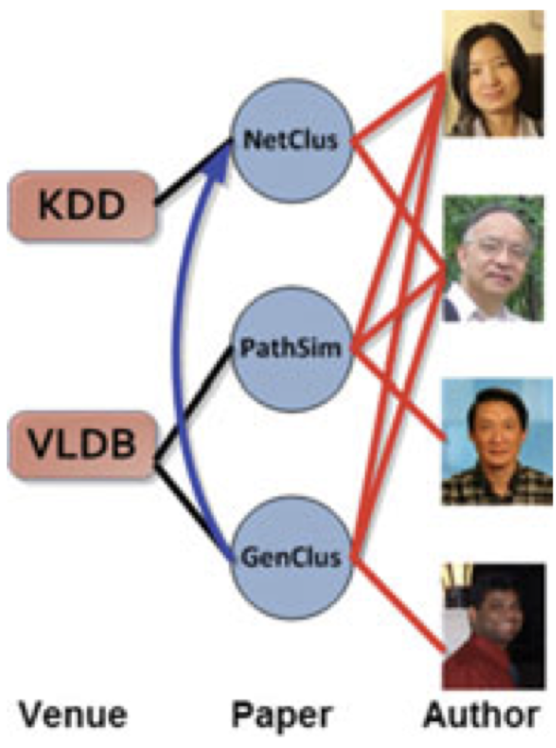
### 2.1.1 基本定义

众所周知，大多数实际系统通常由许许多多交互的、多类别的部件组成，如社交活动、计算机系统和生物网络等。在这些系统中，交互的部件构成信息网络。而无处不在的信息网络，形成了现代信息基础设施的重要组成部分，从而吸引了研究人员的广泛关注。一个信息网络代表现实世界中对象及其之间交互的抽象，正式定义如下：

**定义2.1**（信息网络[13,14])：信息网络定义为一个有向图,并具有对象类型的映射函数和链路类型的映射函数。每个对象属于对象类型集合中的某特定对象类型，每个链路属于关系类型集合中的某特定关系类型。如果两链路属于同一关系类型，那么这两个链路的始末对象类型分别相同。

为明确区分信息网络中的对象和关系类型，提出了异质/同质信息网络的概念，定义如下：

**定义2.2**（异质/同质信息网络）：若信息网络的对象类型数或关系类型数，那么称为异质信息网络；否则，称为同质信息网络。

（a）网络实例 （b）网络架构

图2-1 文献数据的异质信息网络[19]

图2-1是文献数据的异质信息网络示例[13]。一个文献信息网络，如从DBLP1中抽取的涉及计算机科学研究人员的文献网络，是典型的异质网络，它包括三种类型的实体：论文、会议和作者。对于每篇论文，都有到相应作者集合和会议的链路，这些链路属于链路类型的集合。

为更好地理解复杂异质信息网络中的对象和链路类型，有必要提供元层次（即模式层次）的网络描述。因此，进一步以网络模式来描述网络元结构：

**定义2.3**（网络模式[13, 14]）：网络模式记为，是信息网络的元结构，并有对象类型映射和链路类型映射。网络模式是定义在对象类型集合上的有向图，以上的关系为边。

异质信息网络的网络模式强调了对象集合与对象间关系类型的限定，这些限定使异质信息网络半结构化，并指导网络的语义探索。遵循某网络模式的信息网络称为网络模式的实例。对于从对象类型到对象类型的链路类型，即，和分别是链路类型的源对象类型和目标对象类型，可分别记为和。自然地，有逆关系。通常，不等于，除非是对称的。

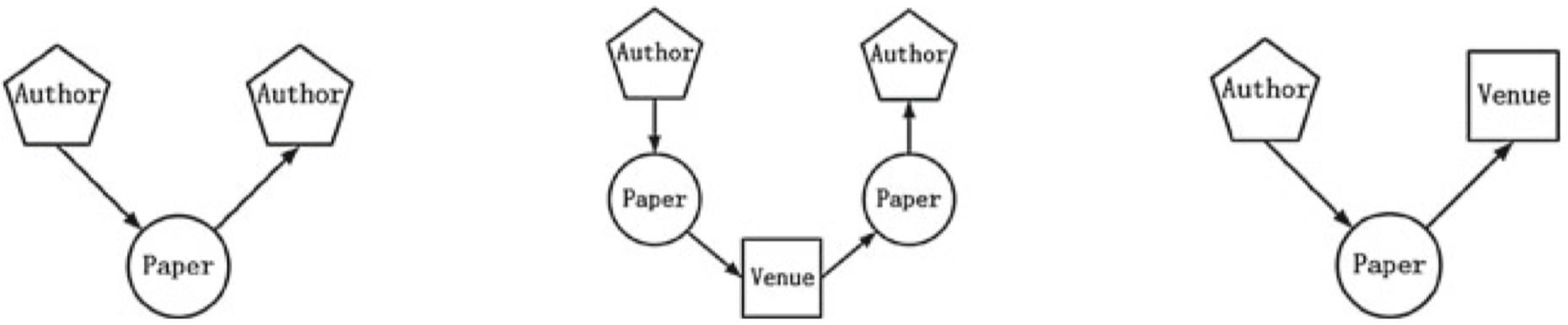
如上所述，图2-1（a）展示了文献数据中真实对象及其之间的联系，图2-1（b）展示了文献数据相应网络模式，且图2-1（a）是图2-1（b）的网络实例。在此示例中，包含三种类型对象：论文（），作者（）和会议（）。链路连接不同类型的对象，由两种对象类型间的关系定义。 例如，作者和论文间的链路表示写或被写关系，而会议和论文间的链路表示发表或发表于关系。

### 2.1.2 元路径

与同质网络不同，异质网络中两对象可通过不同的路径连接，这些路径具有不同的实际意义，从而以元路径分类：

**定义2.4**（元路径[15]）：元路径是在模式上定义的路径，记为。同时，定义对象间的复合关系，其中表示关系上的合成运算。

简单起见，若同一组对象类型间没有多种关系类型，可以使用对象类型表示元路径：。例如，图2-1（a）中作者在会议上发表论文，可以利用长度为2的元路径，或简写来描述。类似地，对象和间的具体路径是相关路径的路径实例。如果在中，对于每个，都有，且每个链路都属于关系，则记。当由元路径定义的关系是对称时（即等于），元路径是对称路径，如和。



(a) APA (b) APVPA (c) APV

图2-2 文献数据的元路径示例[19]

|  |  |  |
| --- | --- | --- |
| 路径实例 | 元路径 | 实际意义 |
| 孙-NetClus-韩 | 作者-论文-作者 | 作者合作同一篇论文 |
| 孙-PathSim-VLDB-PathSim-韩  孙-PathSim-VLDB-GenClus-阿加沃尔 | 作者-论文-会议-论文-作者 | 作者在同一会议发表论文 |
| 孙-NetClus-KDD  孙-PathSim-VLDB | 作者-论文-会议 | 作者在一个会议上发表论文 |

表2-1 关于文献数据的元路径示例及其实际意义

以图2-2所示的元路径为例，作者可以通过相应元路径连接，如“作者-论文-作者”路径，“作者-论文-会议-论文-作者”路径等。此外，表2-1说明了元路径的路径实例和相应语义。显然，这些路径的语义不同。路径指作者合作了一篇论文（即合作作者关系），而路径指作者在同一会议发表论文。

元路径包含丰富语义是异质网络的一个重要特征。基于不同的元路径，对象具有不同的连接关系和相应的路径语义，这可能会对数据挖掘任务的结果产生影响。例如，基于不同元路径评估作者间的相似性得分是不同的[15]。根据路径，共同发表论文的作者更相似，而根据路径，在同一会议发表论文的作者将更相似。作为一个独特的特征和有效的语义捕捉工具，元路径已被广泛使用在异质网络的信息挖掘中，如相似性度量[16, 15]、聚类[17]和分类[18]等。

## 2.2 注意力机制

在机器翻译领域，原始的方法大都利用数理统计进行分析和处理。近些年来，随着AlphaGo的出现，除了在游戏的人工智能领域，深度学习在计算机视觉和自然语言处理等领域也展现出了强大的效用。随着深度学习的进一步发展，Seq2Seq的训练和翻译模式开始进入人们的视野。除此之外，改进端到端的网络结构，弥补其已有的缺陷也是非常必要的。在此情形下，基于循环神经网络（Recurrent Neural Network,RNN）的注意力机制[20]被提出。因其显著的效果提升和端到端的高效训练方式，席卷人工智能各界，并快速应用至机器学习众多任务中。本文同样利用了注意力机制融合模型模块，因此简要介绍注意力机制的核心理论。

机器翻译，即输入某种语言中的一句话，输出目标语言相对应的语句。例如，将德语中的一段话翻译成合适的英语。之前的机器翻译模型通常是encoder-decoder结构，即encoder读取输入句子，将其转换为定长向量，然后decoder再将这个向量翻译成目标语言的对应文字。通常encoder及decoder均采用RNN结构，如长短期记忆网络（Long Short-Term Memory，LSTM）或门控循环单元（Gated Recurrent Unit，GRU）等。如图2-3所示，利用encoder将输入语句编码为最后一层中的隐向量，并将其作为decoder初始的隐向量，最终解码成目标语言中的文字。

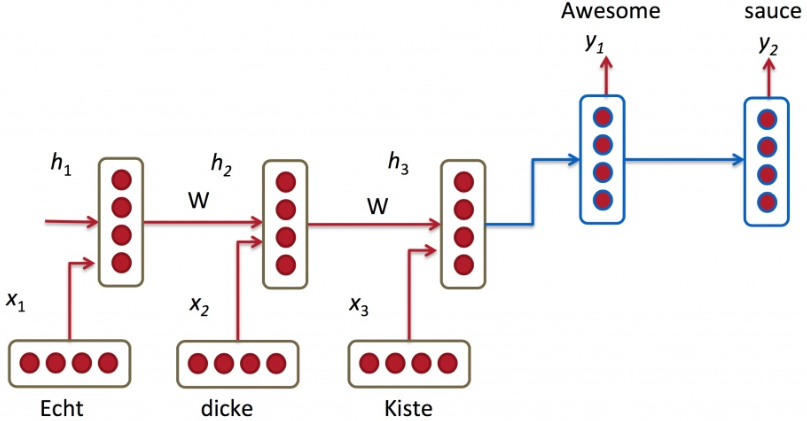


图2-3 encoder-decoder结构示例

但这个结构存在问题一些问题。其中，较为突出的问题是RNN机制中存在长程梯度消失，即对于较长的句子，很难将输入序列转化为定长向量而保存所有有效信息。因此，随所需翻译句子的长度的增加，这种结构的效果会显著下降。

而注意力机制，正是解决长序列到定长向量转化过程中造成部分信息损失的问题。通俗理解，注意力机制跟人类自身翻译文章的思路类似，即将注意力放于翻译部分对应的上下文。相应地，注意力模型中，会聚焦于源语句中相对应的几个词语来翻译当前的词语，并结合之前的已翻译部分得出后续翻译。如图2-4所示，当翻译英文“knowledge”时，只需将注意力放在源中文语句中“知识”的部分，当翻译英文“power”时，只需将注意力集中在中文“力量”部分。这样，当decoder预测目标语句时，就可以考虑到encoder的所有信息，而不仅局限于原模型中最顶层的定长隐向量，因此不会丧失长程信息。

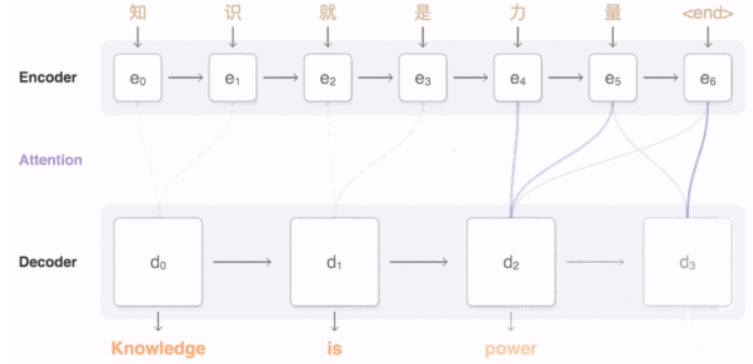


图2-4 基于注意力机制的机器翻译示例

以上是直观理解，接下来详细解释注意力机制在数学上对应哪些运算。

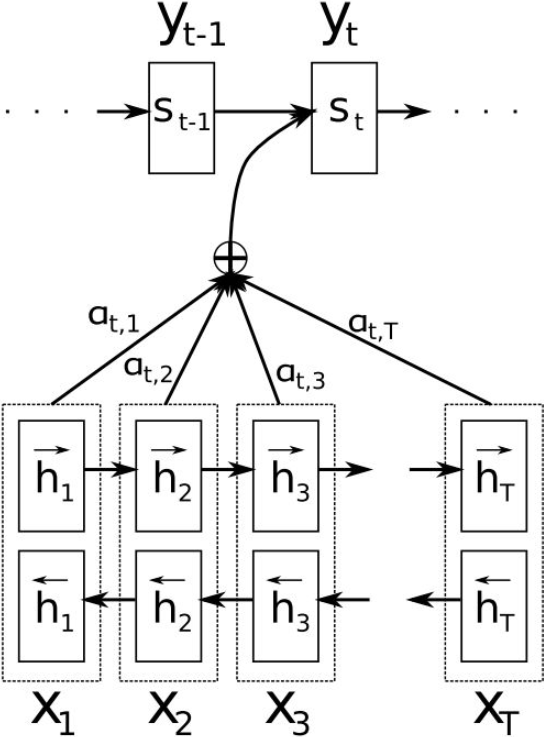


图2-5 注意力机制详解

首先，利用RNN结构得到encoder的隐状态。假设当前decoder的隐状态是，可以计算每一个输入位置与当前输出位置的关联性，相应向量形式为：

其中，是一种相关性运算符。例如，常见的有点乘形式、加权点乘、加和等。

然后，通过softmax将归一，得到注意力分布：

展开形式为：

利用可以加权求和得到相应的上下文向量：

由此，可以计算decoder的下一个隐状态：

以及该位置的输出：

上述过程的关键操作是计算encoder与decoder隐状态之间的关联性权重，得到注意力的分布。从而，对于当前输出位置比较重要的输入位置，相应地在预测时会占较大的比重。

通过注意力机制的引入，打破了只能利用encoder最终的单一向量的限制，使得模型可以集中在所有对于下一个目标单词重要的输入信息上，极大地改善模型效果。除此之外，通过观察注意力权重矩阵的变化，可以更好地知道哪部分翻译对应哪部分原文字，给模型提供更好的解释性，如图2-5所示：

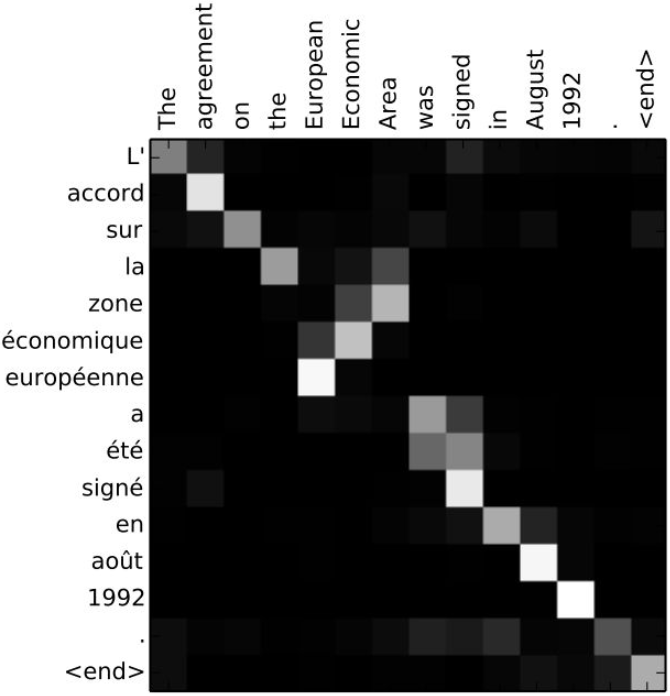


图2-5 注意力分布示例

## 2.3 度量学习

引入度量学习，是模型改进的出发点。因此，本节概述度量学习的基础知识，从而便于后续改进模型的阐述与理解。

距离，是许多机器学习算法的核心概念，包括K-nearest neighbors（kNN）、K-means和SVM等。度量学习，产生以距离为关键的度量指标，从而捕获数据间的关系。在许多机器学习任务中得到应用，包括图像分类和文档检索等[21,22,23,24]。

具体地，给定对象集合，已知某些对象是“相似的”或“不相似的”，度量学习的目标是学习一种符合相似和不相似关系的距离度量，即构建指标使得相似对间距离较小，不相似对间距离较大。

从数学的角度而言，度量指标需要满足几个条件，而三角不等式是其中最重要的一个[25]。具有三角不等式性质的度量指标可以表明，对于任意三个对象，任何两个成对距离的总和应大于或等于剩余的成对距离。这意味着，给定信息“x与y和z都相似”，度量指标不仅会拉近上述的两对关系，还会将剩余的关系对拉近。因此，可以将此视为相似性的传播过程，即度量指标将已知的相似性信息传播到关系未知的对象间。

### 2.3.1 形式化定义

设是输入空间上的数据集合。度量学习中的标签信息以成对的形式体现，包括已知相似对集合：

和已知不相似对集合：

基础的度量学习方法，构建以马氏距离为核心的度量指标：

其中，是半正定矩阵[25]。也就是说，将输入投影到新的空间，且在新空间中的欧氏距离服从给定的相似性约束。有许多种方法来构建这样的度量指标，最基础的做法是解决以下凸优化问题：

现有的先进方法倾向于使用非线性变换，如核函数或神经网络，来提高度量指标精度[26，27，28，29]。

### 2.3.2 kNN的度量学习

上述全局优化方法的本质是将所有相似对拉近，而将不相似对分开。一方面，这种优化并非始终可行。另一方面，Weinberger等人研究表明，若度量学习用于kNN，那么只需一个度量指标就可以让每个对象的kNN与该对象共享同类标签[27]，这个目标更易实现。本文中的改进模型也采用这种度量学习的概念，因为推荐的目标同样是找到每个用户的kNN物品。

具体地，给定输入，最接近的数据点称为的目标邻居。为了便于理解，可以认为的目标邻居建立了其余不同标记的输入不应入侵的边界。假设将侵入边界的不同标记的输入称为入侵者。那么，此时度量学习的目标就是构建一种最小化入侵者数量的指标[27]。

最经典的此类模型是large margin nearest neighbor（LMNN）[27]，它使用两个损失函数完成上述目标。即LMNN定义拉损失函数，将输入的目标邻居拉近：

其中，表示是的目标邻居。此外，LMNN进一步定义推损失函数，推动入侵者离开目标邻居，并在kNN决策边界周围保持较大距离：

如果和属于同一标签，则，否则，且是标准铰链损失。在此基础上，LMNN的总体损失函数是和的加权求和。

## 2.4 本章小结

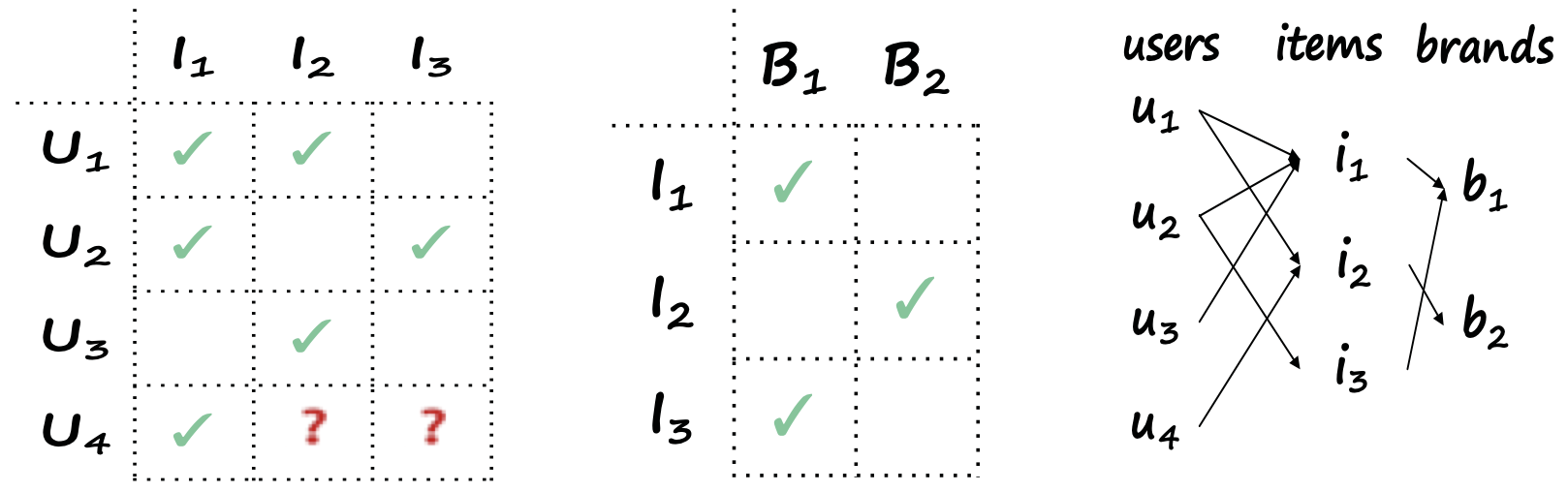
本章主要介绍了模型所依赖的相关背景技术。首先，解释了异质信息网络的基本定义及其关键语义挖掘工具——元路径。其次，阐述了注意力机制的核心理论。最后，说明了度量学习的相关概念，为后续模型的提出做好铺垫。

# 第三章 基于异质信息网络的推荐模型ACFML详解

## 3.1 基础模型NeuACF

经广泛调研后，本文选择NeuACF[8]作为基础模型，并加以改进。因此，本节具体介绍NeuACF的基本思想及模型结构。

虽然现有的隐因子模型在推荐领域表现出了强劲性能，但它们通常只挖掘购买历史的信息，而这仅从一个方面反映用户偏好和物品特征。但是，用户和物品的隐因子源于实际应用的不同方面。这些方面，从不同角度反映用户偏好和物品特征。因此，隐因子模型应利用来自不同方面的信息。以图3-1为例，进一步说明多方面信息的重要性。如果仅利用图3-1（a）中的交互矩阵（即购买历史），会推断出用户将购买物品和。然而，若考虑到图3-1（b）所示的品牌信息时，可以发现物品可能更受用户喜爱，因为物品和属于同一品牌。



（a）用户-物品交互矩阵 （b）物品-品牌关系矩阵 （c）异质信息网络

图3-1 用户-物品方面级交互关系示例[8]

尽管全面利用各方面信息是极有价值的，但面临两个严峻挑战。其一，如何提取不同方面级的信息。其二，如何从不同方面学习和融合隐因子。即使能够提取出不同方面的特征，学习相应的隐因子并实现有效融合仍不是一件易事。虽然可以利用矩阵分解学习隐因子，但它只能学习“浅层”的隐因子。

为解决上述挑战，基于深度神经网络的NeuACF应运而生，可以有效提取和融合不同方面级的隐因子。具体地，推荐系统中不同类型的对象和其之间的交互被建模为异质信息网络，并利用元路径[15]从不同方面提取用户和物品特征。如图3-1（c）所示，可以通过用户-物品-用户路径，从购买历史角度提取用户的隐特征。此外，进一步利用深度神经网络捕捉复杂的非线性关系，为用户和物品学习不同方面级的隐因子，并基于注意力机制实现有效融合，进行Top-N推荐。

### 3.1.1 模型框架

NeuACF模型包含三个主要步骤。首先，基于推荐系统中丰富的用户-物品交互信息构造异质信息网络，并在不同元路径下计算不同方面级的相似性矩阵，以反映用户和物品的不同方面级特征。接下来，利用深度神经网络，以所得相似性矩阵为输入，分别学习方面级隐因子。最后，结合注意力机制融合方面级隐因子，获得用户和物品的最终表示。接下来，详细说明以上三个步骤。

### 3.1.2 方面级相似性矩阵计算

给定元路径，选择相似性矩阵提取方面级特征。原因有二，其一，相似性度量可以减弱噪声；其二，相似性数值在之间，便于隐因子学习。因此，使用PathSim[15]计算方面级相似性矩阵，如基于元路径User-Item-Brand-Item-User（）和Item-Brand-Item（）可以提取品牌方面特征。

### 3.1.3 方面级隐因子学习

已计算出不同方面级用户-用户和物品-物品相似性矩阵，接下来需要利用它们学习相应隐因子，架构如图3-2所示：

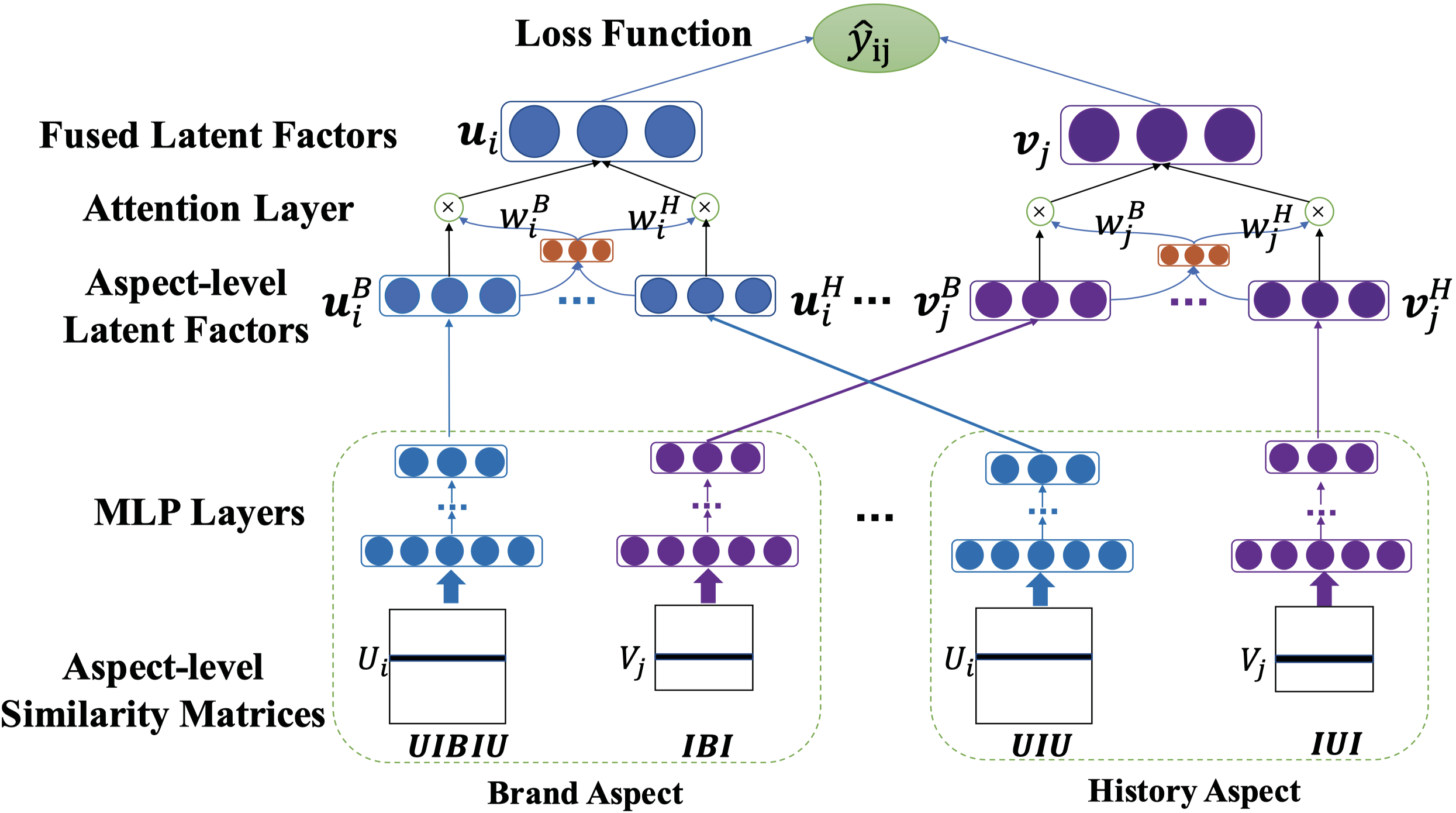


图3-2 NeuACF中的深度神经网络架构

具体地，以方面级相似性矩阵作为多层感知器（Multi-Layer Perceptron，MLP）的输入。相应地，方面级隐因子为其输出。以元路径下的用户相似性矩阵为例，用户以维的向量表示，该向量反映用户与所有其他用户的相似性。其中，是用户总数。MLP将用户的初始相似性向量投影到低维方面级隐因子。在MLP的每层中，输入向量被投影为新空间中的一个向量。形式化如下，给定初始输入向量和隐藏层，其中代表第层，可以通过以下多层映射来学习最终的方面级隐因子：

其中，和分别是第层的权重矩阵和偏置，是隐藏层的激活函数。从图2-4中的架构可以看出，对于用户和物品的每个方面级相似性矩阵，均存在相应的MLP学习该方面级隐因子。

### 3.1.4 基于注意力机制的方面级隐因子融合

已获得用户和物品的方面级隐因子后，接下来需要将它们融合形成最终的表示。比较简单的方法是直接拼接所有方面级隐因子，或者平均所有方面级隐因子。但是，这两种方法都没有区分不同方面的重要性。这显然是不合理的，因为并非所有方面对于推荐都是有意义的。因此，考虑利用先进的注意力机制融合方面级隐因子。

注意力机制在多种机器学习任务中都表现出了优越的特性，如机器翻译[30]。基于注意力机制，可以为所有方面级隐因子分配合适的权重：较高（较低）权重反映相应方面对于推荐是信息量充足（信息量匮乏）的。具体地，利用用户的品牌方面隐因子，通过双层注意力网络计算相应权重：

其中，是权重矩阵，是偏置。

利用Softmax函数归一化，从而获得各方面级隐因子的最终权重，即方面对用户最终表示的贡献：

其中，是方面集合。

获得用户的所有方面级隐因子的权重后，可以计算最终表示：

### 3.1.5 模型优化

将Top-N推荐建模为分类问题，该问题预测用户与物品间的交互概率。为确保输出值是概率，需要将输出约束在内。使用Logistic函数作为输出层的激活函数，根据下式计算用户和物品间的交互概率：

其中，和分别是用户和物品的最终表示。

在所有训练集上，似然函数是：

其中，和分别是正负实例集合，是参数集合。

利用似然函数的负对数形式，得到point-wise损失函数：

其中，是实例的ground truth，是预测结果。上式即为模型的目标函数，可以通过随机梯度下降或其变种技术来优化。

## 3.2 改进模型ACFML

### 3.2.1 NeuACF问题分析

NeuACF中，用户和物品的交互可能性是利用embedding间的点积度量的。由于点积本身所固有的缺陷——不满足三角不等式，可能导致相似性的传递特征被破坏，从而限制模型效果的提升。与此同时，NeuACF使用point-wise损失函数，关注于评分信息的绝对数值。但对于Top-N推荐而言，相对交互可能性更为重要。接下来，本节具体分析以上两点问题可能对于模型效果产生的不良影响。

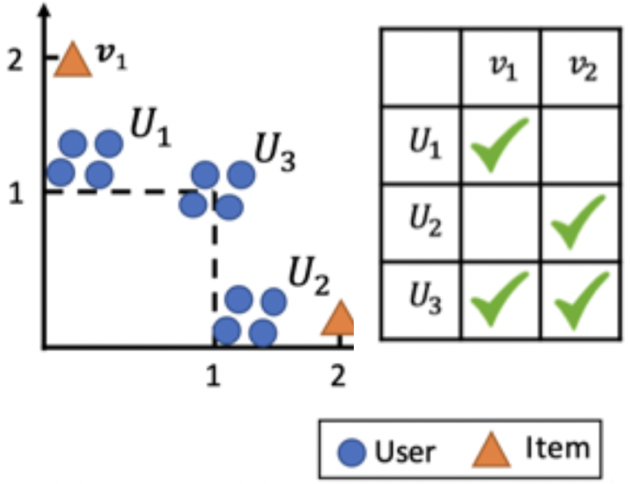


图3-3 矩阵分解稳定解示意图

首先，如图3-3所示，、和是三个规模相同的用户组。其中，喜欢物品,喜欢物品,喜欢物品和。图3-3左半部分展示了在矩阵分解框架下的一组稳定解，即当用户喜欢物品时，用户和物品间的点积等于2，否则点积等于0。可以观察到，虽然用户同时喜欢物品和，但是物品和间的点积等于0。这一点违反了三角不等式，即相似对和间的相似关系无法传播到。

这种相似性传播被打破的情况，可能造成两种后果。第一，尽管矩阵分解框架在两个轴上捕获了最突出的隐因子，但是并没有反映用户的反馈信息中包含的更细粒度偏好。第二，隐因子不能准确捕获用户-用户或物品-物品间的相似性，而这对于缓解协同过滤的数据稀疏性问题至关重要。

其次，point-wise损失函数作为预测结果与ground truth的交叉熵，更为关注分数的绝对数值。而这种损失函数形式对于推荐任务而言，存在两点不足。第一，推荐产生的结果往往是一个物品列表，即Top-N推荐，因此追求的是排序的准确性，而非精确打分数值，只需相对打分即可。第二，point-wise损失函数没有建模预测物品列表中的排序信息，可能会无意中过分强调那些不重要的物品，即排序结果后部，用户并不喜爱的物品。因此，Top-N推荐效果可能受到限制。

### 3.2.2 改进方案

针对已分析的两点问题，本节阐述具体的改进方案及相应理论基础。

首先，相似性的传递与协同过滤技术密切相关。因为，在协同过滤中，也是通过已知的某些用户-物品对间关系，来推断其他未知关系的用户-物品对。正如上节所分析的那样，因三角不等式的性质未被满足，使用用户和物品间embedding的点积来捕捉已有的评分信息，并非是一种度量学习方式。从而，导致相似性的传递过程可能被打破。

为此，引入距离的概念。距离，是许多机器学习算法的核心概念，如K-nearest neighbor（kNN）、K-means等。度量学习利用距离定义的度量指标捕捉数据间的关系，从而使得相似对间距离较近，不相似对间距离较远。而度量学习所需满足的重要性质之一——三角不等式，自然地蕴含相似性的传递过程，即将已知的相似性信息通过距离传递给未知的点对，这与协同过滤的本质思想不谋而合。因此，本文考虑以用户-物品度量指标替换NeuACF中的点积估计形式，在反映用户偏好的同时，捕捉用户-用户及物品-物品间的相似性，实现更全方位的信息传递。具体地，对于已知相似关系的用户-物品对集合，度量指标将会使得中相应对的embedding距离较近，而不在中的点对距离较远。由于满足三角不等式，在此过程中会使得喜爱相同物品的用户聚成簇，且被相同用户喜爱的物品聚成簇。最终，对于用户的kNN物品将是：该用户曾喜爱的物品或具有相同口味的相似用户曾喜爱的物品。换句话说，通过学习满足已知相似信息的度量指标，不仅可以将这些信息传播到其他未知的用户-物品对，而且还可以传播给无法直接观察的用户-用户和物品-物品对。

形式化而言，将用户的embedding记为，物品的embedding记为。以它们之间的欧几里得距离作为度量指标，修改NeuACF打分函数：

从而满足用户对不同物品的相对喜爱程度，即喜爱的物品将比不喜爱的物品更接近该用户。

其次，既然已经在模型中引入了度量学习，那么相应地需要使用适合于度量学习的损失函数，即pair-wise损失函数。其输入是一个三元组，是目标对象，是正样本，即同类对象，是负样本，即异类对象。对应于本文的推荐任务，也就是，是用户，是用户喜爱的物品，是用户不喜爱或不清楚喜好关系的物品。有如下不等式：

其中，表示某种度量指标下的距离数值，threshold表示特定阈值。该不等式本质上定义了正负样本之间的距离关系，即正样本间的距离与阈值之和会小于负样本间的距离。因此，基于此种三元组输入和不等式理论，相应的pair-wise损失函数为：

其中，是正样本集合。因此，最终的优化目标是拉近正样本的距离，推远负样本(i, k)的距离。显然，只有括号内公式的值大于0时，才可以计算误差，并依据反向传播调节网络参数。

### 3.2.3 总体模型

保留NeuACF的模型框架，即方面级相似性矩阵计算、方面级隐因子学习和基于注意力机制的方面级隐因子融合，只替换整个模型的优化模块。

具体地，以欧几里得距离定义的metric衡量用户和物品的交互可能性：

其中，和分别是用户和物品的最终表示。Metric数值越小，交互可能性越大；反之，越小。

基于已定义的打分函数和负采样技术，得到pair-wise损失函数：

其中，是正样本集合，margin表示特定阈值。上式即为模型的目标函数，可以通过随机梯度下降或其变种技术来优化。

## 3.3 本章小结

本章详细介绍了模型各模块的组成。首先，阐述了基础模型NeuACF的具体框架。其次，分析了NeuACF存在的两点问题，即点积和point-wise度量方式可能对模型性能造成的不良后果。最后，引入度量学习和pair-wise损失函数的理论，阐明改进方案ACFML，形成最终的模型结构。

# 第四章 实验

## 4.1 实验设置

### 4.1.1 数据集

本文利用两个公开的数据集评价模型性能，即亚马逊数据集和不同规模的MovieLens数据集。数据集的网络模式如图4-1所示，具体统计信息如表4-1所示。

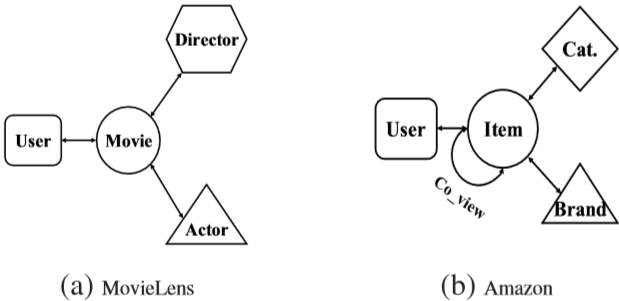


图4-1 实验数据集的网络模式

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Datasets** | **#users** | **#items** | **#ratings** | **#density** |
| **ML100K** | 943 | 1682 | 100,000 | 6.304% |
| **Amazon** | 3532 | 3105 | 57,104 | 0.521% |

表4-1 实验数据集的统计信息

MovieLens-100K（ML100K）2：MovieLens数据集广泛用于电影推荐中。本文使用ML100K版本。为构建异质信息网络，需要对原有数据集进行处理，即对于每部电影，从IMDb中爬取相应导演和演员。

Amazon3：该数据集包含亚马逊电商网站中的用户评分数据。在实验中，本文选择电子类商品进行模型评估。

公开数据集MovieLens和Amazon所包含的用户-物品交互信息为1至5的相应评分，因而直接建模会得到带权异质信息网络。但就本文所提模型而言，不考虑权重对于欧几里得距离度量指标的影响，因此将其统一处理为1。也就是说，若存在评分信息，相应用户-物品有连边；若不存在评分信息，则相应用户-物品无连边。

### 4.1.2 评价指标

本文采用leave-one-out法对模型性能进行评估。具体地，选择一个用户已评分的物品作为正样本，随机选择99个未被该用户评分的物品作为负样本。对于该用户而言，模型将对此100个物品进行排序，产生相应的推荐列表。为与baselines进行公平比较，本文对所有baselines测试集中每个（用户，物品）正样本，使用相同的负样本集合。

在此基础之上，通过命中率（Hit Ratio ，HR）和归一化折损累计增益（Normalized Discounted Cumulative Gain，NDCG）作为具体评价指标。

对于HR而言：

其中，是对测试集中所有用户来说，产生的相应推荐列表中出现正样本的用户。

而NDCG通常用于评价搜索算法，后续用于推荐算法的评价。首先介绍其基本思想，即高关联度的结果比一般关联度的结果更影响最终的指标数值，且高关联度的结果出现在列表更靠前的位置时，指标数值会越高。具体地，先从累计增益（Cumulative Gain，CG）开始理解。CG只考虑了关联度，而没有考虑到所处列表位置的影响。它实际上是搜索结果相关性分数的总和：

其中，代表位置上搜索结果的相关性分数。

进一步，出现了CG的改进版本DCG（Discounted Cumulative Gain），即在每个CG的数值上除以一个折损值。其目的在于，让排名越靠前的结果越影响最后的评价指标，换句话说，排序越靠后，价值越低。对于位置上的搜索结果，其价值是，那么产生的效益就为：

但是，由于检索词的不同，返回的搜索结果数量是不一致的。而DCG是一个累加值，无法针对两种不同的搜索结果进行比较。因此，需要归一化处理，最终出现了NDCG：

其中，IDCG表示理想情况下的最大DCG值。

实际应用到本文所采用的leave-one-out法，NDCG的具体形式为：

其中，是正样本在推荐列表中所处的排序位置。

与此同时，由于效率原因，本文并不产生全部的物品推荐列表，而是利用Top-K排序结果。其中，。

### 4.1.3 Baselines

本文共利用8个基线进行对比实验，包括基础模型NeuACF、两个基本方法（ItemPop和ItemKNN）、两个矩阵分解方法（Matrix Factorization，MF[2]和element-wise Alternating Least Squares，eALS[31]）、一个基于pair-wise损失函数的排序方法（Bayesian Personalized Ranking，BPR[32]）和两个基于神经网络的方法（Deep Matrix Factorization Models, DMF[33]和Neural networkbased Collaborative Filtering，NCF[34]），概要介绍如下：

NeuACF，将推荐系统建模为异质信息网络，利用元路径捕捉用户和物品不同方面级隐因子，并实现高效融合。

ItemPop，物品按照受欢迎程度，即与所有用户的交互总次数，排序推荐。

ItemKNN，基于物品的标准协同过滤方法。

MF, 矩阵分解框架中最具有代表性的隐因子模型。

eALS，用于推荐的矩阵分解算法，基于element-wise交替最小二乘技术，有效优化具有可变加权缺失数据的矩阵分解模型。

BPR，利用贝叶斯后验估计，实现个性化排序，并进一步使用pair-wise损失函数优化矩阵分解模型。

DMF，利用深度神经网络的矩阵分解模型，将用户和物品映射到共同的低维向量空间。

NCF，基于神经网络的协同过滤算法，结合了矩阵分解的线性和深度神经网络的非线性，用于对用户-物品隐结构进行建模。

### 4.1.4 实现细节

本文基于TensorFlow深度学习框架实现所提模型，并对所有数据集，调节相应超参数，使得模型达到最佳效果。对于神经网络，MLP层数在集合中调节，每个隐藏层具有600个隐藏单元。潜在因子维数在集合中调节。batch大小在集合中调节，学习率在集合中调节。在训练模型时，对于每个正样本，负采样的负样本数量在集合中调节,且度量中的margin在集合中调节。模型用Xavier初始化器随机初始化模型参数，并使用Adam作为优化器优化模型。表4-2说明了实验中所用的元路径及相应方面。与此同时，对于所有baselines，根据相应文献，设定其最佳参数。

|  |  |  |  |
| --- | --- | --- | --- |
| **数据集** | **方面** | **元路径** | |
| **用户** | **电影/物品** |
| **MovieLens** | 历史 | UMU | MUM |
| 导演 | UMDMU | MDM |
| 演员 | UMAMU | MAM |
| **Amazon** | 历史 | UIU | IUI |
| 品牌 | UIBIU | IBI |
| 种类 | UICIU | ICI |
| 共同浏览 | UIVIU | IVI |

表4-2 实验所用元路径及其对应方面

## 4.2 实验结果

### 4.2.1 推荐性能分析

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Datasets** | **Metrics** | **ItemPop** | **ItemKNN** | **MF** | **eALS** | **BPR** | **DMF** | **NCF** | **NeuACF** | **ACFML** |
| **ML100K** | HR@5 | 0.2331 | 0.3072 | 0.3634 | 0.3698 | 0.3984 | 0.2983 | 0.3942 | **0.4097** | 0.3648 |
| NDCG@5 | 0.1392 | 0.1667 | 0.2021 | 0.2201 | 0.2315 | 0.1787 | 0.2357 | **0.2505** | 0.2351 |
| HR@10 | 0.3498 | 0.4891 | 0.5437 | 0.5638 | **0.5914** | 0.4494 | 0.5766 | 0.5846 | 0.5483 |
| NDCG@10 | 0.1764 | 0.2283 | 0.2605 | 0.2819 | 0.2733 | 0.2269 | 0.2745 | 0.2868 | **0.2915** |
| HR@15 | 0.4866 | 0.6094 | 0.6338 | 0.6529 | 0.6741 | 0.5373 | 0.6635 | **0.6813** | 0.6713 |
| NDCG@15 | 0.2124 | 0.2576 | 0.2843 | 0.3056 | 0.3149 | 0.2502 | 0.3175 | 0.3218 | **0.3249** |
| HR@20 | 0.4725 | 0.6656 | 0.7144 | 0.7155 | 0.7388 | 0.6019 | 0.7324 | 0.7464 | **0.7572** |
| NDCG@20 | 0.2326 | 0.2708 | 0.3034 | 0.3204 | 0.3302 | 0.2551 | 0.3338 | **0.3469** | 0.3442 |
| **Amazon** | HR@5 | 0.2412 | 0.1897 | 0.3027 | 0.3063 | 0.3296 | 0.2693 | 0.3117 | 0.3268 | **0.3595** |
| NDCG@5 | 0.1642 | 0.1279 | 0.2068 | 0.2049 | 0.2254 | 0.1848 | 0.2141 | 0.2232 | **0.2319** |
| HR@10 | 0.3576 | 0.3126 | 0.4278 | 0.4287 | 0.4657 | 0.3715 | 0.4309 | 0.4686 | **0.5313** |
| NDCG@10 | 0.2016 | 0.1672 | 0.2471 | 0.2441 | 0.2693 | 0.2179 | 0.2524 | 0.2683 | **0.2873** |
| HR@15 | 0.4408 | 0.3901 | 0.5054 | 0.5065 | 0.5467 | 0.4328 | 0.5258 | 0.5591 | **0.6554** |
| NDCG@15 | 0.2236 | 0.1877 | 0.2676 | 0.2647 | 0.2908 | 0.2332 | 0.2774 | 0.2924 | **0.3202** |
| HR@20 | 0.4997 | 0.4431 | 0.5680 | 0.5702 | 0.6141 | 0.4850 | 0.5897 | 0.6257 | **0.7370** |
| NDCG@20 | 0.2375 | 0.2002 | 0.2824 | 0.2797 | 0.3067 | 0.2458 | 0.2925 | 0.3080 | **0.3396** |

表4-3 不同方法的HR@K和NDCG@K指标对比

表4-3展示了不同方法的实验结果。其中，峰值性能用粗体标出。可以看出，本文所提模型并非在所有数据集和指标上都能达到最佳性能。具体地，相比较Amazon数据集，ML100K数据集更加稠密，评分信息较多。因其数据特点，baselines中基于评分信息的方法都有较好的表现。而对于Amazon数据集而言，数据较稀疏。在这种情况下，本文所引入的基于欧几里得距离的度量指标凸显其相应作用，即不仅刻画了显式的用户-物品间关系，也捕捉了隐含的更细粒度的用户-用户和物品-物品相似性，从而有效缓解数据稀疏问题。与此同时，本文所提模型在NDCG指标上更具优势，也进一步说明了pair-wise损失函数相比point-wise损失函数，显式约束了排序信息，可以得到排序更为合理的推荐列表。

观察baselines表现，可以看到，NeuACF在ML100K和Amazon数据集的多数指标上都能达到峰值性能，说明NeuACF作为基于异质信息网络的最新模型，在推荐领域仍具有不俗的表现。究其原因，通过利用异质信息网络这一强有力工具，NeuACF基于不同元路径学习到不同方面级隐因子，从多角度提供了用户和物品的特征，从而得到全面的表示学习结果。

### 4.2.2 margin大小对性能的影响

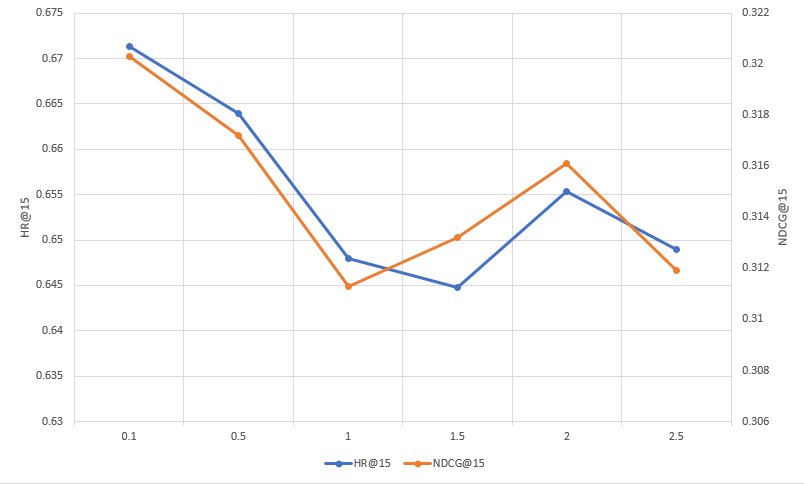


图4-2 不同margin对应的推荐性能

在基于欧几里何距离的度量指标中，margin的大小，可能会对推荐性能产生重要影响。因此，本文具体研究了所提模型的margin大小变动时，推荐性能的相应变化趋势。由于超参数量较多，本文采取控制变量法，即固定MLP层数为2，batch大小为1024，隐因子数量为32，学习率为0.00005，负采样数量为10，使得margin大小从0.1至2.5变动。因时间限制，本文选择在ML100K数据集上进行实验，其他数据集上的结论应近似，结果如图4-2所示。可以看出，随margin的增加，推荐性能虽稍有波动，但总体呈下降趋势。究其原因，可能是在数据集规模较大的情况下，大的margin可能导致收敛速度降低，甚至难以收敛。

### 4.2.3 负采样数量对性能的影响

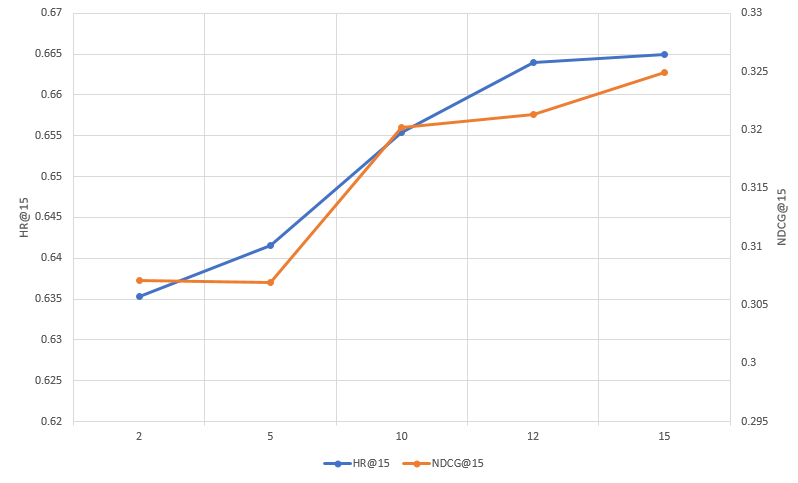


图4-3 不同负采样数量对应的推荐性能

在基于负采样技术的pair-wise损失函数中，负采样的数量，可能对推荐性能产生重要影响。因此，本文具体研究了所提模型的负采样数量变动时，推荐性能的相应变化趋势。由于超参数量较多，本文采取控制变量法，即固定MLP层数为2，batch大小为1024，隐因子数量为32，学习率为0.00005，margin大小为2.0，使得负采样数量从2至15变动。因时间限制，本文选择在ML100K数据集上进行实验，其他数据集上的结论应近似，结果如图4-3所示。可以看出，随负采样数量增加，推荐性能逐渐增加，并慢慢趋于稳定。因此，为了平衡性能与效率，负采样数量可以在15左右选择。

### 4.2.4 隐因子数量对性能的影响

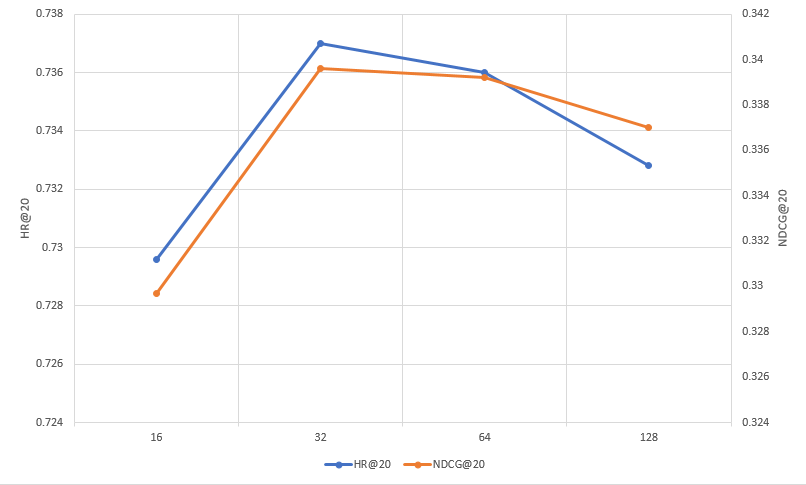


图4-4 不同隐因子数量对应的推荐性能

在隐因子模型中，隐因子的数量，即embedding维度，可能对推荐性能产生重要影响。因此，本文具体研究了所提模型在MLP最后一层所学到的隐因子的数量变动时，推荐性能的相应变化趋势。由于超参数量较多，本文采取控制变量法，即固定MLP层数为2，batch大小为1024，学习率为0.00005，负样本数量为10，margin大小为2.0，使得隐因子维数自16至256变动。因时间限制，本文选择在ML100K数据集上进行实验，其他数据集上的结论应近似，结果如图4-4所示。可以看出，随隐因子数量增加，推荐性能首先增加，峰值性能在32-64左右达到。此时，如果维度进一步增大，推荐性能会下降，可能产生了过拟合。

## 4.3 本章小结

本章详细介绍了实验的相关内容。

对于实验部分，首先说明了本文所利用的两个经典数据集MovieLens和Amazon的数据内容及其相应统计信息。其次，介绍了用于评价模型性能的两个指标HR和NDCG及其相应计算方法。然后，列举了与本文所提模型ACFML相对比的8个baselines，并概括其核心理论。最后，详细说明了对比实验的结果和关键超参数对模型性能的影响，依据实验结果进行解释分析，总结所提模型的优点及劣势。

# 第五章 结束语

## 5.1 论文工作总结

本文通过选题背景介绍，从宏观角度回顾了推荐领域的经典算法。进一步地，深入分析了基础模型NeuACF的具体结构，并论述了其存在的两点问题，给出相应改进方案ACFML。具体地，引入metric learning和pair-wise损失函数，产生排序更合理的推荐列表。最终通过编码实现改进模型，在推荐数据集Amazon和MovieLens上，与经典算法进行了大量对比实验，验证了所提模型的优越性。

综上，本文取得的主要成果有三点：

* 1. 在基础模型NeuACF中引入度量学习的概念，以欧几里得距离指标捕捉数据间关系，弥补点积所造成的相似性传递特征丢失，实现更全方位的信息传递。
  2. 使用pair-wise损失函数替换point-wise损失函数，强调对于Top-N推荐而言的相对交互可能性，产生排序更合理的物品推荐列表。
  3. 基于推荐数据集Amazon和MovieLens，与经典算法进行对比实验，验证了模型的优越性。

此次毕业设计的全部内容是本人在石川老师指导下完成的，所包含的主要工作有：

1. 了解课题背景及意义，调研相似性推荐算法，完成开题。
2. 基于对现有算法的分析，选择NeuACF作为基本模型，并剖析其存在的问题。
3. 针对分析所得结论，进一步完成相应改进模型ACFML的详细设计。
4. 确定改进方案后，熟悉数据特点及相应算法框架。编码实现模型，进行中期检查。
5. 基于经典数据集进行实验验证，并对推荐结果进行分析评估，完成并提交毕业设计论文、相关文档及源程序。

## 5.2 问题和展望

### 5.2.1 问题

本文实现的改进模型ACFML在稀疏数据集Amazon上表现优良。但是，对于更稠密的ML100K数据集，存在收敛速度变缓且性能降低的问题。如何平衡性能和数据数量，是未来努力的方向。与此同时，在实际应用中，不仅存在用户评分这样的显式反馈信息，而且充斥着大量诸如点击、加入购物车等隐式反馈，这些信息对于刻画用户喜好同样起着关键作用。本文由于时间有限，未搜寻合适的隐式反馈数据集评估模型，这同样也是未来的工作之一。

### 5.2.2 展望

随着大数据时代的来临，网络信息快速膨胀，高效且个性化推荐的重要性愈发凸显。因此，受到了学术界和产业界的联合关注，但不得不承认推荐技术目前仍面临诸多难点，如冷启动及多模态数据的融合运用等。

与此同时，异质信息网络作为一个年轻且快速发展的研究领域，可以自然地用于推荐系统的建模。因此，展望基于异质信息网络的推荐技术，可能有如下几个研究方向：

首先，是更复杂的异质信息网络构建。已有的相关研究，通常假定异质信息网络是明确定义的，即对象和对象间的关系是已知且清晰的。然而，在实际推荐系统中，构造异质信息网络会遇到很多挑战。对于关系数据库等结构化数据，构造相应异质信息网络较容易。然而即使是这样的网络，对象和对象间的关系也可能存在噪声，如对象重名或对象间关系不完整等。对于文本、图像等非结构化数据，如何高效抽取相应对象和关系，进而建立准确完整的异质信息网络，也同样是难点。在实践中，会利用多种技术，如自然语言处理、图像处理等。

其次，是更强大的分析方法。在异质信息网络中，对象因不同的关系组织在一起。星型模式是广泛应用的异质信息网络类型。在推荐系统中，网络化数据更复杂且无规律可循。举个例子，网络中的链路会包含属性值，如评分信息，从而构成带权异质信息网络。这种丰富模式的异质信息网络，亟需强有力的语义分析工具。

最后，是更大数据的处理。为展现异质网络建模的优势，需要在更广泛的领域中设计实用的推荐算法。多样性是大数据的重要特征之一，而异质信息网络是处理多样性的有效方法。实际生活中，异质信息网络是巨大的，甚至是动态的，所以通常不能在内存中直接处理。由于用户往往只对小部分节点、链路或子网络感兴趣，因此可以从现有网络中动态地提取子网络进行分析。

前路漫漫，但机遇与挑战并存。随着国内外学者持续的关注与投入，推荐技术必能够越来越成熟，为人类带来更好的生活体验。

# 参考文献

[1] Hu Y, Koren Y, Volinsky C. Collaborative Filtering for Implicit Feedback Datasets[C]//ICDM. 2008, 8: 263-272.

[2] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems[J]. Computer, 2009 (8): 30-37.

[3] Koren Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C]//Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2008: 426-434.

[4] Shi C, Liu J, Zhuang F, et al. Integrating heterogeneous information via flexible regularization framework for recommendation[J]. Knowledge and Information Systems, 2016, 49(3): 835-859.

[5] Shi C, Zhou C, Kong X, et al. Heterecom: a semantic-based recommendation system in heterogeneous networks[C]//Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012: 1552-1555.

[6] Shi C, Zhang Z, Ji Y, et al. SemRec: a personalized semantic recommendation method based on weighted heterogeneous information networks[J]. World Wide Web, 2019, 22(1): 153-184.

[7] Yu X, Ren X, Sun Y, et al. Personalized entity recommendation: A heterogeneous information network approach[C]//Proceedings of the 7th ACM international conference on Web search and data mining. ACM, 2014: 283-292.

[8] Han X, Shi C, Wang S, et al. Aspect-Level Deep Collaborative Filtering via Heterogeneous Information Networks[C]//IJCAI. 2018: 3393-3399.

[9] Jamali M, Lakshmanan L. HeteroMF: recommendation in heterogeneous information networks using context dependent factor models[C]//Proceedings of the 22nd international conference on World Wide Web. ACM, 2013: 643-654.

[10] Ren X, Liu J, Yu X, et al. Cluscite: Effective citation recommendation by information network-based clustering[C]//Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014: 821-830.

[11] Wu J, Chen L, Yu Q, et al. Trust-aware media recommendation in heterogeneous social networks[J]. World Wide Web, 2015, 18(1): 139-157.

[12] Luo C, Pang W, Wang Z, et al. Hete-cf: Social-based collaborative filtering recommendation using heterogeneous relations[C]//2014 IEEE International Conference on Data Mining. IEEE, 2014: 917-922.

[13] Sun Y, Han J. Mining heterogeneous information networks: a structural analysis approach[J]. Acm Sigkdd Explorations Newsletter, 2013, 14(2): 20-28.

[14] Sun Y, Yu Y, Han J. Ranking-based clustering of heterogeneous information networks with star network schema[C]//Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009: 797-806.

[15] Sun Y, Han J, Yan X, et al. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks[J]. Proceedings of the VLDB Endowment, 2011, 4(11): 992-1003.

[16] Shi C, Kong X, Yu P S, et al. Relevance search in heterogeneous networks[C]//Proceedings of the 15th international conference on extending database technology. ACM, 2012: 180-191.

[17] Sun Y, Norick B, Han J, et al. Pathselclus: Integrating meta-path selection with user-guided object clustering in heterogeneous information networks[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2013, 7(3): 11.

[18] Kong X, Yu P S, Ding Y, et al. Meta path-based collective classification in heterogeneous information networks[C]//Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012: 1567-1571.

[19] Shi C, Philip S Y. Heterogeneous information network analysis and applications[M]. Springer International Publishing, 2017.

[20] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural information processing systems. 2017: 5998-6008.

[21] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1701-1708.

[22] Wan J, Wang D, Hoi S C H, et al. Deep learning for content-based image retrieval: A comprehensive study[C]//Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014: 157-166.

[23] Koestinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 2288-2295.

[24] Xu Z E, Chen M, Weinberger K Q, et al. From sBoW to dCoT marginalized encoders for text representation[C]//Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012: 1879-1884.

[25] Xing E P, Jordan M I, Russell S J, et al. Distance metric learning with application to clustering with side-information[C]//Advances in neural information processing systems. 2003: 521-528.

[26] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification[J]. Journal of Machine Learning Research, 2009, 10(Feb): 207-244.

[27] Wang J, Do H T, Woznica A, et al. Metric learning with multiple kernels[C]//Advances in neural information processing systems. 2011: 1170-1178.

[28] Cogswell M, Ahmed F, Girshick R, et al. Reducing overfitting in deep networks by decorrelating representations[J]. arXiv preprint arXiv:1511.06068, 2015.

[29] Kedem D, Tyree S, Sha F, et al. Non-linear metric learning[C]//Advances in Neural Information Processing Systems. 2012: 2573-2581.

[30] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.

[31] He X, Zhang H, Kan M Y, et al. Fast matrix factorization for online recommendation with implicit feedback[C]//Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. ACM, 2016: 549-558.

[32] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback[C]//Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence. AUAI Press, 2009: 452-461.

[33] Xue H J, Dai X, Zhang J, et al. Deep Matrix Factorization Models for Recommender Systems[C]//IJCAI. 2017: 3203-3209.

[34] He X, Liao L, Zhang H, et al. Neural collaborative filtering[C]//Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017: 173-182.

1 http://dblp.uni-trier.de/

2 https://grouplens.org/datasets/movielens/

3 http://jmcauley.ucsd.edu/data/amazon/

# 致 谢

行文至此，意味着本科的学习生活即将结束。四年的时间，说长不长，说短也不短，是旧时期的结束，也是新阶段的开始。在这匆匆的几年中，我有许多想要感谢的人，谢谢他们一路的鼓励与支持，谢谢他们如一的包容与理解，谢谢有人教我如何成长，谢谢有人教我怎样爱与被爱。

首先，我要感谢毕业设计的指导老师——石川老师，北京邮电大学计算机学院教授，也是我未来博士研究生的导师。作为我的研究生导师，石川教授在学术研究中认真严谨，对于我的毕业设计给予了很多建议与帮助，并及时指出其中存在的问题。与此同时，他也耐心引导我适应学术研究的过程，让我从最初的迷茫困惑，到现在渐渐明晰方向。

其次，我要感谢实验室的两位青年教师及师兄师姐。顺利保研后，经半年相处，我认为用大家庭来形容实验室更为准确。除了和师兄师姐的融洽交流，两位年轻老师亦师亦友，在这种和谐的环境中，无论是科研还是项目，都有事半功倍的效果。在此期间，我也逐渐明白了真正做科研应是怎样的心态与方式。

再次，我要感谢我的舍友、同学、我爱的人和爱我的人，他们为我的生活增添了许多明媚的色彩，使我从踌躇与痛苦中脱身而出，实现蜕变。

最后，感谢努力而不言败的自己。