

MO HU REN ZHI CHAO TU  
YU DUO GUAN XI SHU JU WA JUE

# 模糊认知超图 与多关系数据挖掘

李慧 著



中国出版集团公司  
现代教育出版社

本书受北京市自然科学基金项目“针对影视数据的个性化推荐关键技术研究”（项目编号：9164028）和北京联合大学专著出版资助项目资助

# 模糊认知超图与多关系 数据挖掘

李 慧 著



中国出版集团公司  
现代教育出版社

---

## 图书在版编目 (CIP) 数据

模糊认知超图与多关系数据挖掘/李慧著. -- 北京:  
现代教育出版社, 2017. 7

ISBN 978-7-5106-5568-5

I. ①模… II. ①李… III. ①模糊图象 IV.

①TP75

中国版本图书馆 CIP 数据核字 (2017) 第 171681 号

---

## 模糊认知超图与多关系数据挖掘

---

策 划 庞 强 刘 媛  
著 者 李 慧  
责任编辑 刘小华  
封面设计 宋晓璐·贝壳悦读

---

出版发行 现代教育出版社  
地 址 北京市朝阳区安华里 504 号 E 座  
邮 编 100011  
电 话 (010) 64244927  
传 真 (010) 64251256

---

印 刷 北京市金星印务有限公司  
开 本 710mm×1000mm 1/16  
印 张 10.75  
字 数 151 千字  
版 次 2017 年 7 月第 1 版  
印 次 2017 年 7 月第 1 次印刷  
书 号 ISBN 978-7-5106-5568-5  
定 价 38.00 元

---

版权所有 违者必究

## | 序 言 |

本书所进行的研究涉及基于模糊认知图的数据挖掘方法研究与多关系数据的知识发现两大新兴领域，属于两个领域的交叉领域研究。本书首先总结了数据挖掘技术，然后对模糊认知超图模型的两大理论基础——模糊认知图和超图进行了说明，尤其对基于模糊认知图的数据挖掘技术的优点与缺陷进行分析；然后将有向超图引入模糊认知图概念中，使用有向超图描述多关系数据中的数据关系；然后在上述两大理论基础之上，建立模糊认知超图挖掘模型（Fuzzy Cognitive HyperMap-FCHM），使之满足复杂系统多关系数据源中不同对象间的复杂关系分析；并对模糊认知超图的知识推理方法和数据聚类方法，进行说明；最后将模糊认知超图应用于多关系数据挖掘，对多关系数据进行数据特征分析与描述，尤其是应用于社区网络发现等复杂系统的模型建模与推理中。

本书的主要研究工作可以表示如下：

1. 总结了基于模糊认知图的数据挖掘和知识发现方法，分析了模糊认知图应用于多关系数据挖掘的不足，提出将有向超图引入模糊认知图的方法，结合两者尝试建立模糊认知超图模型（FCHM）。
2. 分析与讨论了模糊认知超图的结构、形式化描述方法和推理机制。论文通过概念集节点之间的关系直观地表现出基于模糊认知超

图的因果知识表示,通过网络各概念节点之间的连接关系及其推理模拟对应系统的动态行为,通过在模糊认知超图网络中前向概念节点对后向概念节点的影响关系,实现基于模糊认知超图的知识推理。

3. 在充分分析模糊认知超图特点的基础上,引入了动态因果关系,以图论为理论研究支撑,以模糊认知超图模型的网络结构为研究核心,提出了因果链的求解算法。该算法借助图论中超图的邻接矩阵及可达矩阵的性质以及关系,利用模糊认知超图中概念节点间的可达性,实现对所有因果链的搜索及分析。

4. 针对复杂模糊认知超图,提出了基于强连通和基于频繁项集算法的模糊认知超图分解方法。该方法借鉴软件工程中模块化的思想,将一个包含大量节点且关系复杂的模糊认知超图分解为多个较小的强连通模块;根据模糊认知超图的特点建立分解准则,并在分解准则的指导下,将模糊认知超图分解成相对独立的多个子图。

5. 基于模糊认知超图模型,提出适应于社区发现的层次聚类方法。充分利用模糊认知超图的模型特点和推理机制,通过超边相似性定义和计算实现超边的融合,并通过划分密度来衡量聚类的质量。该层次聚类方法为社会网络中的社区网络发现提供了技术支持。

本书针对模糊认知超图的不同角度和层次进行了一一探讨,首先总结了前人的研究成果,然后对模糊认知超图做了多方面的研究和创新,最后对模糊认知超图技术的研究与发展前景进行了展望。

本书中所提出的主要创新点可以描述如下:

(1) 首先分析了模糊认知图和超图的特点,在两者的基础上,提出了模糊认知超图的理论模型,并对模糊认知超图模型进行了结构化分析。

(2) 然后在模糊认知超图中,针对概念节点间的因果影响进行分析,并通过关联矩阵的乘积运算,来模拟节点间因果关系影响的传递

过程。通过可达矩阵的运算来度量各个超边上节点间的因果影响程度，并基于该度量方法，求解模糊认知超图中的因果链。

(3) 针对适用于复杂系统的模糊认知超图，提出了基于强连通理论的模糊认知超图分解算法，并基于频繁项集，对模糊认知超图进行模块化分解。

(4) 以社会网络中的社区发现为应用背景，研究了通过模糊认知超图对社会网络的建模方法，提出了基于模糊认知超图的网络层次聚类方法，实现社区网络中的社区发现算法。实验表明，该方法具备便于使用、准确高效的特点。

本书可以作为基于模糊认知超图的数据挖掘研究方面的参考书，还可以作为数据挖掘领域的技术研究人员是提高性参考书，对于图论以及超图方面也具有一定的参考与实用意义。

本书的出版得到了北京市自然科学基金项目“针对影视数据的个性化推荐关键技术研究”（项目编号：9164028）和北京联合大学专著出版资助项目的资助，在此表示衷心的感谢！本书内容的基础是作者本人前期研究过程中所取得的成果，感谢导师杨炳儒教授对本人的悉心指导，感谢本书作者原博士研究单位北京科技大学计算机与通信工程学院的培养，感谢曾为本书做出贡献的同学、同事、朋友。本书的出版得到了北京联合大学各级领导的关心、支持和帮助，书中参考了国内外许多专家、学者的论著，在此一并致以衷心的感谢。

由于本书所涉及的内容中，一些理论方法和技术还在继续研究之中，由于水平有限，书中错漏之处难免，欢迎读者批评指正。

作者

2017年6月

# | 目 录 |

<b>第一章 数据挖掘技术概论</b> .....	1
1.1 数据分析与数据挖掘 .....	2
1.2 复杂系统与多关系网络 .....	4
1.3 多关系数据挖掘 .....	4
1.4 多关系数据挖掘的特别之处 .....	6
<b>第二章 模糊认知超图的理论基础——模糊认知图</b> .....	8
2.1 模糊认知图模型 .....	9
2.2 模糊认知图的学习方法 .....	12
2.3 模糊认知图的领域应用 .....	18
2.4 模糊认知图的挑战 .....	23
<b>第三章 模糊认知超图的理论基础——超图</b> .....	25
3.1 超图模型 .....	26
3.2 模糊认知图与超图的结合可能性分析 .....	30
<b>第四章 模糊认知图的演化模型和算法改进</b> .....	33
4.1 模糊认知图的演化模型 .....	34
4.2 模糊认知图的学习算法改进 .....	45

第五章 模糊认知超图的结构与表示 .....	49
5.1 模糊认知超图的结构 .....	51
5.2 模糊认知超图的形式化表示 .....	53
5.3 模糊认知超图中的最短路径 .....	55
5.4 模糊认知超图的最短路径算法 .....	56
第六章 模糊认知超图的因果链与推理机制 .....	60
6.1 模糊认知超图中因果影响的表示及计算 .....	62
6.2 模糊认知超图的因果链 .....	65
6.3 模糊认知超图推理机制 .....	69
第七章 模糊认知超图的分解与层次聚类 .....	81
7.1 模糊认知超图的分解 .....	82
7.2 模糊认知超图的层次聚类方法 .....	93
第八章 超边聚合与节点重要性评估 .....	97
8.1 模糊认知超图中超边相似度计算 .....	98
8.2 超边的聚合过程 .....	105
8.3 聚类树密度划分 .....	107
8.4 节点重要性评估方法 .....	109
第九章 多关系网络的社区发现 .....	126
9.1 多关系网络的社区发现 .....	128
9.2 社区节点质量评估 .....	132
9.3 社区发现实例分析 .....	137
第十章 总结和展望 .....	142
参考文献 .....	145



## | 图 目 录 |

图 2.1	简单模糊认知图 .....	10
图 2.2	模糊认知图 .....	13
图 3.1	一个简单的有向超图 .....	28
图 4.1	规则模糊认知图 .....	35
图 4.2	FGCM 中三种相关关系 .....	36
图 4.3	I-FCM-II 中节点的相关关系 .....	38
图 4.4	模糊认知网络的交互操作过程 .....	41
图 4.5	FTCM 时间延迟与在虚节点中单位时间转换 .....	42
图 5.1	一个简单的模糊认知超图 .....	52
图 6.1	节点之间的影响关系 .....	64
图 6.2	模糊认知超图的简单示意图 .....	67
图 6.3	模糊认知超图推理过程示意图 .....	71
图 6.4	一个简单的模糊认知超图模型示例 .....	75
图 6.5	论文撰写合作网络的模糊认知图表示 .....	77
图 6.6	执行 Add 操作后的模糊认知超图模型 .....	79
图 6.7	论文合作撰写网的对偶图 .....	80
图 7.1	社会网络中社区发现的层次聚类算法示意图 .....	95
图 8.1	两条超边不完全相交 .....	98

图 8.2	超边相交图 .....	99
图 8.3	层次聚类实例 .....	107
图 8.4	网络中每个节点重要性计算流程图 .....	115
图 8.5	节点多属性到节点质量的映射过程 .....	116
图 9.1	网络的社区结构,阴影对应社区 .....	128
图 9.2	模糊认知超图层次聚类流程图 .....	130
图 9.3	节点与超边的相关性 .....	131
图 9.4	一个社区划分密度计算的示意图 .....	131
图 9.5	影响因子与势熵的关系变化曲线 .....	134
图 9.6	各节点删除后网络势的变化 .....	135
图 9.7	两种方法的节点排序 .....	136
图 9.8	HMETIS 方法的结果 .....	139
图 9.9	层次方法的结果 .....	140
图 9.10	层次聚类密度划分变化曲线 .....	140

数据挖掘技术是计算机科学的一个重要分支，它主要研究如何从大量的数据中提取有用的信息。数据挖掘技术的应用非常广泛，包括市场营销、金融、医疗、教育等领域。

数据挖掘技术可以分为数据预处理、数据分类、数据聚类、数据关联分析、数据异常检测等。数据预处理是数据挖掘的第一步，它包括数据清洗、数据集成、数据变换和数据归约。数据分类是根据数据的特征将数据分成不同的类别。数据聚类是根据数据的相似性将数据分成不同的簇。数据关联分析是发现数据中的关联规则。数据异常检测是发现数据中的异常点。

数据挖掘技术的应用非常广泛，包括市场营销、金融、医疗、教育等领域。

数据挖掘技术是计算机科学的一个重要分支，它主要研究如何从大量的数据中提取有用的信息。

## 第一章 数据挖掘技术概论

数据挖掘技术是计算机科学的一个重要分支，它主要研究如何从大量的数据中提取有用的信息。

数据挖掘技术的应用非常广泛，包括市场营销、金融、医疗、教育等领域。

数据挖掘技术可以分为数据预处理、数据分类、数据聚类、数据关联分析、数据异常检测等。

数据预处理是数据挖掘的第一步，它包括数据清洗、数据集成、数据变换和数据归约。

数据分类是根据数据的特征将数据分成不同的类别。

数据聚类是根据数据的相似性将数据分成不同的簇。

数据关联分析是发现数据中的关联规则。

数据异常检测是发现数据中的异常点。

数据挖掘技术的应用非常广泛，包括市场营销、金融、医疗、教育等领域。

数据挖掘技术是计算机科学的一个重要分支，它主要研究如何从大量的数据中提取有用的信息。

数据挖掘技术的应用非常广泛，包括市场营销、金融、医疗、教育等领域。

数据挖掘技术可以分为数据预处理、数据分类、数据聚类、数据关联分析、数据异常检测等。

数据预处理是数据挖掘的第一步，它包括数据清洗、数据集成、数据变换和数据归约。

随着社会的进步和信息技术的发展，信息系统在各行业、各领域快速拓展。系统采集、处理、积累的数据越来越多，数据量增速越来越快，以至于“海量、爆炸性增长”不足以形容数据增长的速度。数据已经渗透到各个行业，如何从大数据中分析并萃取有效信息已经成为亟须解决的问题之一。

数据挖掘是一种通过分析数据从而在大量数据中寻找未知规律的技术，该研究领域在三十余年的发展历程中，众多该领域专家进行了大量的研究工作，在理论和应用两个方面都取得了突破性的进展。但是面对当前越来越复杂的数据内容和系统应用，迫切需要类似于关系模式、数据库管理系统等方面的理论和方法的指导。

需求牵引与市场推动是永恒的，数据挖掘的焦点不再是追求提高传统算法的稳定性与有效性，或算法的时间和空间效率。数据挖掘的环境已经扩展为复杂数据环境，如对文本数据、视频图像数据、声音数据、图形数据以及综合多媒体数据的挖掘。数据挖掘的模型将发展为能够处理比较复杂或结构独特的数据，并且需要一些更新更好的分析和建模的方法。

## 1.1 数据分析与数据挖掘

数据分析是指用适当的统计分析方法对收集来的大量数据进行分析，提取有用信息和形成结论而对数据加以详细研究和概括总结的过

程。这一过程也是质量管理体系的支持过程。在实用中，数据分析可帮助人们做出判断，以便采取适当行动。

数据分析过程的主要活动由识别信息需求、收集数据、分析数据、评价并改进数据分析的有效性组成。典型的数据分析可能包含以下三步：

1. 探索性数据分析：当数据刚取得时，可能杂乱无章，看不出规律，通过作图、造表、用各种形式的方程拟合，计算某些特征量等手段探索规律性的可能形式，即往什么方向和用何种方式去寻找和揭示隐含在数据中的规律性。

2. 模型选定分析：在探索性分析的基础上提出一类或几类可能的模型，然后通过进一步的分析从中挑选一定的模型。

3. 推断分析：通常使用数理统计方法对所定模型或估计的可靠程度和精确程度做出推断。

数据挖掘一般是指从大量的数据中通过算法搜索隐藏于其中的信息的过程。数据挖掘通常与计算机科学有关，并通过统计、在线分析处理、情报检索、机器学习、专家系统（依靠过去的经验法则）和模式识别等诸多方法来实现上述目标。

数据挖掘利用了来自如下一些领域的思想：（1）来自统计学的抽样、估计和假设检验；（2）人工智能、模式识别和机器学习的搜索算法、建模技术和学习理论。数据挖掘也迅速地接纳了来自其他领域的思想，这些领域包括最优化、进化计算、信息论、信号处理、可视化和信息检索。一些其他领域也起到重要的支撑作用。特别地，需要数据库系统提供有效的存储、索引和查询处理支持。源于高性能（并行）计算的技术在处理海量数据集方面常常是重要的。分布式技术也能帮助处理海量数据，并且当数据不能集中到一起处理时更是至关重要。

## 1.2 复杂系统与多关系网络

为了描述现实生活中越来越复杂的关系网络，满足和方便社会网络中的数据挖掘，研究者们构建了大量的系统用来表达对象间的关系以及关系之间的相关程度，如运输、通信、金融、生物信息等系统。生物信息中的神经系统即可以看作是一个大型的复杂关系网络，大量神经元细胞通过神经纤维相互连接而组成一个网络，各个网络节点之间存在错综复杂的多对多关系，类似的网络还有气象系统、社会关系网络等。

复杂关系网络系统往往由大量复杂的结构化对象组成，该网络系统由许多“概念节点/顶点”与连接节点之间的“边”组成。其中节点用来表示现实系统中不同的实体、事件，而实体、事件之间的关系则用连接网络节点之间的边来表示。随着网络的规模越来越大，网络中的概念节点越来越多，而概念节点间的连接关系也越来越复杂，复杂的系统甚至可能涉及成千上万个概念节点和非常复杂的概念节点间关系。

## 1.3 多关系数据挖掘

目前的研究中，对社会网络的分析大多是采用静态的网络来模拟，这样难以描述网络的动态变化特征，从而较少进行相关的研究。随着社会网络应用重要性越来越大，对于社会网络群体中的动态交互

模式，以及整个网络的人物关系演变轨迹的研究也越来越显示出其意义的重要性<sup>[1]</sup>，因此，对于社会网络中互动行为的动态性、多变性研究目前也已成为社会网络研究中的难点问题。

以往针对社会网络的研究中，一般使用图论和概率统计方法作为其数学基础，而其中图论恰恰最适合于表示社会网络的结构，对于复杂的社会网络，概念节点间的联系已不仅仅是两概念节点之间的关系或者一对多的关系，还存在一些多对多的关系，因此，传统意义上的图模型已经无法准确地刻画出现实世界中的社会网络特征。

如在论文和作者的关系分析系统中，假如将一位作者作为一个概念点，两位作者发表过同一篇论文作为两位作者概念间存在关系，则在简单的关系网络图中，一条连接两概念节点的边可以表示两位作者之间是否合作，也就是说这两位作者是同一篇论文的共同作者，但一条普通的边却无法表示三位作者或更多的作者一起属于一篇文章的作者<sup>[2]</sup>。同理，假如将一篇论文作为一个概念点，论文间存在共同作者作为两篇论文概念间存在关系的依据，则简单关系网络图可以使用一条边表示两论文间存在共同作者，但无法表示多篇论文间存在共同作者，或表示多篇论文作者的相似程度。

将所有论文概念点之间的关系进行数据挖掘，可进行研究领域相近或研究内容相关的论文社团发现，或者将所有作者概念点之间的关系进行数据挖掘，可进行有合作关系作者之间的社团发现。

另外，在针对超大规模的社会网络进行建模时，也同样会出现对应该网络的概念节点间相互关系错综复杂的情况，或者在一个网络中出现另一个内含子网络的问题，因此，要准确地描述该大规模社会网络，必须使用一种超越一般网络的超网络来实现。

其次，传统的基于普通图网络的人物关系分析过程只关注行动者

之间的相互连接关系，而忽略行动者的自身属性。而在实际应用中，行动者的自身属性对整个网络关系分析过程中的性能以及动态演化也具有极其重要的作用。例如，在社会网络中，财富和地位的不同，也可以推论出此人在整个网络中所起的作用可能是不同的，所以，在具体的社会网络分析中，应该结合“关系数据”和“属性数据”这两类数据进行社会网络发现。因此，要准确地描述该社会网络，需要使社会网络中的数据包含两类，分别是“属性数据”和“关系数据”。

## 1.4 多关系数据挖掘的特别之处

目前的研究中，对社会网络的分析大多是采用静态的网络来模拟，这样难以描述网络的动态变化特征，从而较少进行相关的研究。随着社会网络应用重要性越来越大，对于社会网络群体中的动态交互模式，以及整个网络的人物关系演变轨迹的研究也越来越显示出其意义的重要性，因此，对于社会网络中互动行为的动态性、多变性研究目前也已成为社会网络研究中的难点问题。

由于在描述社会人物关系的网络所体现出来的多关系、高复杂度等新特征，传统意义上的模糊认知图模型已无法准确地刻画现实世界中的社会网络关系，从而造成当前的社会网络发现过程中，对于节点重要性评估的研究和社区发现准确性的研究出现很大的不足。

本书以常用于动态系统模拟的模糊认知图为理论基础，将超图理论引入到模糊认知图中，使其既进行系统演化计算，又利用超图的多关系特点表现现实社会人物关系网络中的多关系、多属性特点，建立能够准确描述社会网络的模糊认知超图模型，从而更有效地表示社会



网络的动态性、复杂性。

在基于模糊认知超图的社区发现应用中,本书在针对社会网络建模的模糊认知超图模型基础上,使用超边作为基本单位设计出一个层次聚类方法,并使用该层次聚类方法实现社区网络中的群体发现,在不提前指定社区数目的条件下,结合模糊认知超图的动态性特征,建立能描述多关系、多属性社会网络的复杂网络模型,综合考虑模糊超图中概念节点的“关系数据”和“属性数据”,准确有效地挖掘出在社会网络模型中的关键节点。为进行社区以及重叠社区的发现提供了一种新的解决思路,对社会网络和多关系网络的数据挖掘和知识发现也具有相应的借鉴意义。